Solutions Manual

to accompany

Probability, Random Variables and Stochastic Processes

Fourth Edition

Athanasios Papoulis
Polytechnic University

S. Unnikrishna Pillai
Polytechnic University

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogotá Caracas Kuala Lumpur Lisbon London Madrid Mexico City
Milan Montreal New Delhi Santiago Seoul Singapore Sydney Taipei Toronto

McGraw-Hill Higher Education

A Division of The McGraw-Hill Companies

Solutions Manual to accompany
PROBABILITY, RANDOM VARIABLES AND STOCHASTIC PROCESSES, FOURTH EDITION
ATHANASIOS PAPOULIS

Published by McGraw-Hill Higher Education, an imprint of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020.

www.mhhe.com

2-1 We use De Morgan's law:

(a) $\overline{\bar{A}+\bar{B}} + \overline{\bar{A}+B} = AB + A\bar{B} = A(B+\bar{B}) = A$

(b) $(A+B)(\overline{AB}) = (A+B)(\bar{A}+\bar{B}) = A\bar{B} + B\bar{A}$

because $A\bar{A} = \{\emptyset\}$ $\quad B\bar{B} = \{\emptyset\}$

2-2 If $A = \{2 \le x \le 5\}$ $\quad B = \{3 \le x \le 6\}$ $\quad S = \{-\infty < x < \infty\}$ then

$A+B = \{2 \le x \le 6\}$ $\quad AB = \{3 \le x \le 5\}$

$$(A+B)(\overline{AB}) = \{2 \le x \le 6\}\ [\{x<3\} + \{x>5\}]$$
$$= \{2 \le x < 3\} + \{5 < x \le 6\}$$

2-3 If $AB = \{\emptyset\}$ then $A \subset \bar{B}$ hence

$$P(A) \le P(\bar{B})$$

2-4 (a) $P(A) = P(AB) + P(A\bar{B})$ $\qquad P(B) = P(AB) + P(\bar{A}B)$

If, therefore, $P(A) = P(B) = P(AB)$ then

$$P(A\bar{B}) = 0 \qquad P(\bar{A}B) = 0 \qquad \text{hence}$$
$$P(\bar{A}B + A\bar{B}) = P(\bar{A}B) + P(A\bar{B}) = 0$$

(b) If $P(A) = P(B) = 1$ then $1 = P(A) \le P(A+B)$ hence

$$1 = P(A+B) = P(A) + P(B) - P(AB) = 2 - P(AB)$$

This yields $P(AB) = 1$

2-5 From (2-13) it follows that

$$P(A+B+C) = P(A) + P(B+C) - P[A(B+C)]$$
$$P(B+C) = P(B) + P(C) - P(BC)$$
$$P[A(B+C)] = P(AB) + P(AC) - P(ABC)$$

because $ABAC = ABC$. Combining, we obtain the desired result.

Using induction, we can show similarly that

$$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$
$$- P(A_1A_2) - \cdots - P(A_{n-1}A_n)$$
$$+ P(A_1A_2A_3) + \cdots + P(A_{n-2}A_n)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\pm P(A_1A_2 \cdots A_n)$$

2-6 Any subset of S contains a countable number of elements, hence, it can be written as a countable union of elementary events. It is therefore an event.

2-7 Forming all unions, intersections, and complements of the sets {1} and {2,3}, we obtain the following sets:

$$\{\emptyset\},\ \{1\},\ \{4\},\ \{2,3\},\ \{1,4\},\ \{1,2,3\},\ \{2,3,4\},\ \{1,2,3,4\}$$

2-8 If $A \subset B, P(A) = 1/4$, and $P(B) = 1/3$, then

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/4}{1/3} = \frac{3}{4}$$

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

2-9

$$P(A|BC)P(B|C) = \frac{P(ABC)}{P(BC)}\frac{P(BC)}{P(C)}$$

$$= \frac{P(ABC)}{P(C)} = P(AB|C)$$

$$P(A|BC)P(B|C)P(C) = \frac{P(ABC)}{P(BC)}\frac{P(BC)}{P(C)}P(C)$$

$$= P(ABC)$$

2-10 We use induction. The formula is true for $n = 2$ because $P(A_1A_2) = P(A_2|A_1)P(A_1)$. Suppose that it is true for n. Since

$$P(A_{n+1}A_n \cdots A_1) = P(A_{n+1}|A_n \cdots A_2A_1)P(A_1 \cdots A_n)$$

we conclude that it must be true for $n+1$.

2-11 First solution. The total number of m element subsets equals $\binom{n}{m}$ (see Probl. 2-26). The total number of m element subsets containing ζ_o equals $\binom{n-1}{m-1}$. Hence

$$p = \binom{n}{m}\Big/\binom{n-1}{m-1} = \frac{m}{n}$$

Second solution. Clearly, $P\{\zeta_o|A_m\} = m/n$ is the probability that ζ_o is in a specific A_m. Hence (total probability)

$$p = \sum P\{\zeta_o|A_m\}p(A_m) = \frac{m}{n}\sum P(A_m) = \frac{m}{n}$$

where the summation is over all sets A_m.

2-12 (a) $P\{6 \le t \le 8\} = \frac{2}{10}$

(b) $P\{6 \le t \le 8 | t > 5\} = \frac{P\{6 \le t \le 8\}}{P\{t > 5\}} = \frac{2}{5}$

2-13 From (2-27) it follows that

$$P\{t_o \le t \le t_o + t_1 | t \ge t_o\} = \int_{t_o}^{t_o + t_1} \alpha(t)dt \Big/ \int_{t_o}^{\infty} \alpha(t)dt$$

$$P\{t \le t_1\} = \int_0^{t_1} \alpha(t)dt$$

Equating the two sides and setting $t_1 = t_0 + \Delta t$ we obtain

$$\alpha(t_o) \Big/ \int_{t_o}^{\infty} \alpha(t)dt = \alpha(0)$$

for every t_o. Hence,

$$-\ln \int_{t_o}^{\infty} \alpha(t)dt = \alpha(0)t_o \qquad \int_{t_o}^{\infty} \alpha(t)dt = e^{-\alpha(0)t_o}$$

Differentiating the setting $c = \alpha(0)$, we conclude that

$$\alpha(t_o) = c\,e^{ct} \qquad P\{t \le t_1\} = 1 - e^{-ct_1}$$

2-14 If A and B are independent, then $P(AB) = P(A)P(B)$. If they are mutually exclusive, then $P(AB) = 0$. Hence, A and B are mutually exclusive and independent iff $P(A)P(B) = 0$.

2-15 Clearly, $A_1 = A_1A_2 + A_1\bar{A}_2$ hence

$$P(A_1) = P(A_1A_2) + P(A_1\bar{A}_2)$$

If the events A_1 and $\bar{A}_2$ are independent, then

$$P(A_1\bar{A}_2) = P(A_1) - P(A_1A_2) = P(A_1) - P(A_1)P(A_2)$$
$$= P(A_1)[1 - P(A_2)] = P(A_1)P(\bar{A}_2)$$

hence, the events A_1 and $\bar{A}_2$ are independent. Furthermore, S is independent with any A because SA = A. This yields

$$P(SA) = P(A) = P(S)P(A)$$

Hence, the theorem is true for $n = 2$. To prove it in general we use induction: Suppose that A_{n+1} is independent of $A_1, \ldots, A_n$. Clearly, A_{n+1} and $\bar{A}_{n+1}$ are independent of $B_1, \ldots, B_n$. Therefore

$$P(B_1 \cdots B_nA_{n+1}) = P(B_1 \cdots B_n)P(A_{n+1})$$
$$P(B_1 \cdots B_n\bar{A}_{n+1}) = P(B_1 \cdots B_n)P(\bar{A}_{n+1})$$

2.16 The desired probabilities are given by (a)

$$\frac{\binom{m-1}{k-1}}{\binom{n}{k}}$$

(b)

$$\frac{\binom{m}{k}}{\binom{n}{k}}$$

2.17 Let A_1, A_2 and A_3 represent the events

$A_1 =$ *"ball numbered less than or equal to m is drawn"*
$A_2 =$ *"ball numbered m is drawn"*
$A_3 =$ *"ball numbered greater than m is drawn"*

$$
\begin{aligned}
P(A_1 \text{ occurs } n_1 = k-1, \quad & A_2 \text{ occurs } n_2 = 1 \text{ and } A_3 \text{ occurs } n_3 = 0) \\
&= \frac{(n_1+n_2+n_3)!}{n_1!\,n_2!\,n_3!}\, p_1^{n_1} p_2^{n_2} p_3^{n_3} \\
&= \frac{k!}{(k-1)!}\left(\frac{m}{n}\right)^{k-1}\left(\frac{1}{n}\right) \\
&= \frac{k}{n}\left(\frac{m}{n}\right)^{k-1}
\end{aligned}
$$

2.18 All cars are equally likely so that the first car is selected with probability $p = 1/3$. This gives the desired probability to be

$$\binom{10}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^7 = 0.26$$

2.19 $P\{$*"drawing a white ball "*$\} = \dfrac{m}{m+n}$
$P($*"atleat one white ball in k trials "*$)$

$$
\begin{aligned}
&= 1 - P(\text{"all black balls in } k \text{ trials"}) \\
&= 1 - \frac{\binom{n}{k}}{\binom{m+n}{k}}
\end{aligned}
$$

2.20 Let $D = 2r$ represent the penny diameter. So long as the center of the penny is at a distance of r away from any side of the square, the penny will be entirely inside the square. This gives the desired probability to be

$$\frac{(1-2r)^2}{1} = \left(1 - \frac{3}{4}\right)^2 = \frac{1}{16}.$$

2.21 Refer to Example 3.14.
(a) Using (3.39), we get

$$P(\text{"}all\ one-digit\ numbers\text{"}) = \frac{\binom{9}{6}\binom{42}{0}}{\binom{51}{6}} = 5 \times 10^{-6}.$$

(b)

$$P(\text{"}two\ one\text{–}digit\ and\ four\ two\text{–}digit\ numbers\text{"}) = \frac{\binom{9}{2}\binom{42}{4}}{\binom{51}{6}} = 0.224.$$

2-22 The number of equations of the form $P(A_iA_k) = P(A_i)P(A_k)$ equals $\binom{n}{2}$. The number of equations involving r sets equals $\binom{n}{r}$. Hence the total number N of such equations equals

$$N = \binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n}$$

And since

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = (1+1)^n = 2^n$$

we conclude that

$$N = 2^n - \binom{n}{0} - \binom{n}{1} = 2^n - 1 - n$$

2-23 We denote by B_1 and B_2 respectively the balls in boxes 1 and 2 and by R the set of red balls. We have (assumption)

$$P(B_1) = P(B_2) = 0.5 \qquad P(R|B_1) = 0.999 \qquad P(R|B_2) = 0.001$$

Hence (Bayes' theorem)

$$P(B_1|R) = \frac{P(R|B_1)P(B_1)}{P(R|B_1)P(B_1) + P(R|B_2)P(B_2)} = \frac{0.999}{0.999 + 0.001} = 0.999$$

2-24 We denote by B_1 and B_2 respectively the ball in boxes 1 and 2 and by D all pairs of defective parts. We have (assumption)

$$P(B_1) = P(B_2) = 0.5$$

To find $P(D|B_1)$ we proceed as in Example 2-10:

First solution. In box B_1 there are 1000×999 pairs. The number of pairs with both elements defective equals 100×99. Hence,

$$P(D|B_1) = \frac{100 \times 99}{1000 \times 999}$$

Second solution. The probability that the first bulb selected from B_1 is defective equals 100/1000. The probability that the second is defective assuming the first was effective equals 99/999. Hence,

$$P(D|B_1) = \frac{100}{1000} \times \frac{99}{999}$$

We similarly find

$$P(D|B_2) = \frac{100}{2000} \times \frac{99}{1999}$$

(a) $P(D) = P(D|B_1)P(B_1) + P(D|B_2)P(B_2) = 0.0062$

(b) $P(B_1|D) = \dfrac{P(D|B_1)P(B_1)}{P(D)} = 0.80$

2-25 Reasoning as in Example 2-13, we conclude that the probability that the bus and the train meet equals

$$(10+x)60 - \frac{10^2}{2} - \frac{x^2}{2}$$

Equating with 0.5, we find $x = 60 - 10\sqrt{11}$.

2-26 We wish to show that the number $N_n(k)$ of the element subsets of S equals

$$N_n(k) = \frac{n(n-1)\cdots(n-k+1)}{1 \cdot 2 \cdots k}$$

This is true for $k=1$ because the number of 1-element subsets equals n. Using induction in k, we shall show that

$$N_n(k+1) = N_n(k)\,\frac{n-k}{k+1} \qquad 1 < k < n \qquad \text{(i)}$$

We attach to each k-element subset of S one of the remaining $n-k$ elements of S. We, then, form $N_n(k)(n-k)$ $k+1$-element subsets. However, these subsets are not all different. They form groups each of which has $k+1$ identical elements. We must, therefore, divide by $k+1$.

2-27 In this experiment we have 8 outcomes. Each outcome is a selection of a particular coin and a specific sequence of heads or tails; for example fhh is the outcome "we selected the fair coin and we observed hh". The event F = {the selected coin is fair} consists of the four outcomes fhh, fht, fth and fhh. Its complement $\overline{F}$ is the selection of the two-headead coin. The event HH = {heads at both tosses} consists of two outcomes. Clearly,

$$P(F) = P(\overline{F}) = \frac{1}{2} \qquad P(HH|F) = \frac{1}{4} \qquad P(HH|\overline{F}) = 1$$

Our problem is to find P(F|HH). From (2-41) and (2-43) it follows that

$$P(HH) = P(HH|F)P(F) + P(HH|\overline{F})P(\overline{F}) = \frac{5}{8}$$

$$P(F|HH) = \frac{P(HH|F)P(F)}{P(HH)} = \frac{1/4 \times 1/2}{5/8} = \frac{1}{5}$$

3.11

$$P_n = pP_{n+\alpha} + qP_{n-\beta}$$

Arguing as in (3.43), we get the corresponding iteration equation

$$P_n = P_{n+\alpha} + qP_{n-\beta}$$

and proceed as in Example 3.15.

3.12 Suppose one best on $k = 1, 2, \cdots, 6$.
Then

$$p_1 = P(k \text{ appears on one dice}) = \binom{3}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$$

$$p_2 = P(k \text{ appear on two dice}) = \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$$

$$p_3 = P(k \text{ appear on all the tree dice}) = \left(\frac{1}{6}\right)^3$$

$$p_0 = P(k \text{ appear none}) = \left(\frac{5}{6}\right)^3$$

Thus, we get

$$\text{Net gain} = 2p_1 + 3p_2 + 4p_3 - p_0 = 0.343.$$

which gives for $i = a$

$$N_a = \begin{cases} \dfrac{a+b}{p-q} \cdot \dfrac{1-(q/p)^a}{1-(q/p)^{a+b}} - \dfrac{a}{p-q}, & p \neq q \\ ab, & p = q \end{cases}$$

$$= \begin{cases} \dfrac{b}{2p-1} - \dfrac{a+b}{2p-1} \cdot \dfrac{1-(p/q)^b}{1-(p/q)^{a+b}}, & p \neq q. \\ ab, & p = q \end{cases}$$

3.10 Using the hint, we obtain

$$p\,(N_{k+1} - N_k) = q\,(N_k - N_{k-1}) - 1$$

Let

$$M_{k+1} = N_{k+1} - N_k$$

so that the above iteration gives

$$\begin{aligned} M_{k+1} &= \frac{q}{p}\,M_k - \frac{1}{p} \\ &= \begin{cases} \left(\frac{q}{p}\right)^k M_1 - \dfrac{1}{p-q}\left\{1 - (\frac{q}{p})^k\right\}, & p \neq q \\ M_1 - \frac{k}{p}, & p = q \end{cases} \end{aligned}$$

This gives

$$\begin{aligned} N_i &= \sum_{k=0}^{i-1} M_{k+1} \\ &= \begin{cases} \left(M_1 + \dfrac{1}{p-q}\right) \displaystyle\sum_{k=0}^{i-1} \left(\frac{q}{p}\right)^k - \dfrac{i}{p-q}, & p \neq q \\ iM_1 - \dfrac{i(i-1)}{2p}, & p = q \end{cases} \end{aligned}$$

where we have used $N_o = 0$. Similarly $N_{a+b} = 0$ gives

$$M_1 + \frac{1}{p-q} = \frac{a+b}{p-q} \cdot \frac{1 - q/p}{1 - (q/p)^{a+b}}.$$

Thus

$$N_i = \begin{cases} \dfrac{a+b}{p-q} \cdot \dfrac{1 - (q/p)^i}{1 - (q/p)^{a+b}} - \dfrac{i}{p-q}, & p \neq q \\ i(a+b-i), & p = q \end{cases}$$

3.8 Define the events

A=" r successes in n Bernoulli trials"

B="success at the i^{th} Bernoulli trial"

C="$r-1$ successes in the remaining $n-1$ Bernoulli trials excluding the i^{th} trial"

$$P(A) = \binom{n}{r} p^r \, q^{n-r}$$

$$P(B) = p$$

$$P(C) = \binom{n-1}{r-1} p^{r-1} \, q^{n-r}$$

We need

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(BC)}{P(A)} = \frac{P(B)\,P(C)}{P(A)} = \frac{r}{n}.$$

3.9 There are $\binom{52}{13}$ ways of selecting 13 cards out of 52 cards. The number of ways to select 13 cards of any suit (out of 13 cards) equals $\binom{13}{13} = 1$. Four such (mutually exclusive) suits give the total number of favorable outcomes to be 4. Thus the desired probability is given by

$$\frac{4}{\binom{52}{13}} = 6.3 \times 10^{-12}$$

3.7 (a) Let n represent the number of wins required in 50 games so that the net gain or loss *does not* exceed \$1. This gives the net gain to be

$$-1 < n - \frac{50-n}{4} < 1$$

$$16 < n < 17.3$$

$$n = 17$$

$$P(\text{net gain } \textit{does not} \text{ exceed } \$1) = \binom{50}{17}\left(\frac{1}{4}\right)^{17}\left(\frac{3}{4}\right)^{33} = 0.432$$

$$P(\text{net gain or loss exceeds } \$1) = 1 - 0.432 = 0.568$$

(b) Let n represent the number of wins required so that the net gain or loss *does not* exceed \$5. This gives

$$-5 < n - \frac{(50-n)}{2} < 5$$

$$13.3 < n < 20$$

$$P(\text{net gain } \textit{does not} \text{ exceed } \$5) = \sum_{n\,=\,14}^{19}\binom{50}{n}\left(\frac{1}{4}\right)^{n}\left(\frac{3}{4}\right)^{50-n} = 0.349$$

$$P(\text{net gain or loss exceeds } \$5) = 1 - 0.349 = 0.651$$

Problem Solutions for Chapter 3

3.1 (a) $P(\text{A } occurs\ atleast\ twice\ in\ n\ trials)$

$$
\begin{aligned}
&= 1 - P(\text{A } never\ occurs\ in\ n\ trials) - P(\text{A } occurs\ once\ in\ n\ trials) \\
&= 1 - (1-p)^n - np(1-p)^{n-1}
\end{aligned}
$$

(b) $P(\text{A } occurs\ atleast\ thrice\ in\ n\ trials)$

$$
\begin{aligned}
&= 1 - P(\text{A never occurs in } n \text{ trials}) - P(\text{A occurs once in } n \text{ trials}) \\
&\quad\ - P(\text{A occurs twice in } n \text{ trials}) \\
&= 1 - (1-p)^n - np(1-p)^{n-1} - \frac{n(n-1)}{2}\,p^2(1-p)^{n-2}
\end{aligned}
$$

3.2

$$
P(doublesix) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}
$$

$$
\begin{aligned}
&P(\text{“double six atleast three times in n trials''}) \\
&= 1 - \binom{50}{0}\left(\tfrac{1}{36}\right)^0\left(\tfrac{35}{36}\right)^{50} - \binom{50}{1}\left(\tfrac{1}{36}\right)\left(\tfrac{35}{36}\right)^{49} - \binom{50}{2}\left(\tfrac{1}{36}\right)^2\left(\tfrac{35}{36}\right)^{48} \\
&= 0.162
\end{aligned}
$$

3.6 (a)

$$
p_1 = 1 - \left(\frac{5}{6}\right)^6 = 0.665
$$

(b)

$$
1 - \left(\frac{5}{6}\right)^{12} - \binom{12}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{11} = 0.619
$$

(c)

$$
1 - \left(\frac{5}{6}\right)^{18} - \binom{18}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{17} - \binom{18}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{16} = 0.597
$$

CHAPTER 4

4-1 From the evenness of $f(x)$: $1 - F(x) = F(-x)$.

From the definition of x_u: $u = F(x_u)$, $1 - u = F(x_{1-u})$. Hence

$$1 - u = 1 - F(x_u) = F(-x_u) = F(x_{1-u}) \qquad -x_u = x_{1-u}$$

4-2 From the symmetry of $f(x)$: $1 - F(\eta+a) = F(\eta-a)$. Hence [see (4-8)]

$$P\{\eta-a < \underset{\sim}{x} < \eta+a\} = F(\eta+a) - F(\eta-a) = 2F(\eta+a) - 1$$

This yields

$$1-\alpha = 2F(\eta+a) - 1 \qquad F(\eta+a) = 1 - \alpha/2 \qquad \eta+a = x_{1-\alpha/2}$$

$$F(a-\eta) = \alpha/2 \qquad a-\eta = x_{\alpha/2}$$

4-3 (a) In a linear interpolation:

$$x_u \simeq x_a + \frac{x_b-x_a}{u_b-u_a}(u-u_a) \qquad \text{for } x_a < x_u < x_b$$

From Table 4-1 page 106

$$z_{0.9} \simeq 1.25 + \frac{0.00565}{0.00885} \times 0.05 = 1.2819$$

Proceeding simiplarly, we obtain

$u =$	0.9	0.925	0.95	0.975	0.99
$z_u =$	1.282	1.440	1.645	1.960	2.327

(b) If $\underset{\sim}{z}$ is such that $\underset{\sim}{x} = \eta + \sigma\underset{\sim}{z}$ then $\underset{\sim}{z}$ is N(0,1) and $G(z) = F_x(\eta+\sigma z)$. Hence,

$$u = G(z_u) = F_x(\eta+\sigma z_u) = F_x(x_u) \qquad x_u = \eta+\sigma z_u$$

4-4 $p_k - 2G(k) = 1 = 2\,\text{erf}k$

(a) From Table 4-1

$k =$	1	2	3
$p_k =$	0.6827	0.9545	0.9973

(b) From Table 3-1 with linear interpolation:

$p_k =$	0.9	0.99	0.999
$k =$	1.282	2.32	3.090

(c) $$P\{\eta - z_u\sigma < \underset{\sim}{x} < \eta + z_u\sigma\} = 2G(z_u) - 1 = \gamma$$

Hence, $G(z_u) = (1+\gamma)/2 \qquad u = (1+\gamma)/2$

4-5 (a) $F(x) = x$ for $0 \le x \le 1$; hence, $u = F(x_u) \doteq x_u$

(b) $F(x) = 1 - e^{-2x}$ for $x \ge 0$; hence, $u = 1 - e^{-2x_u}$

$$x_u = -\frac{1}{2}\ln(1-u)$$

$u =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$x_u =$	0.0527	0.1116	0.1783	0.2554	0.3466	0.4581	0.6020	0.847	1.1513

4-6 Percentage of units between 96 and 104 ohms equals 100p where $p = P\{96 < \underset{\sim}{R} < 104\} = F(104) - F(96)$

(a) $F(R) = 0.1(R-95)$ for $95 \le R \le 105$. Hence,

$p = 0.1(104-95) - 0.1(96-95) = 0.8$

(b) $p = G(2.5) - G(-2.5) = 0.9876$

4-7 From (4-34), with $\alpha = 2$ and $\beta = 1/\lambda$ we get $f(x) = c^2 xe^{-cx}U(x)$

$$F(x) = c^2 \int_0^x ye^{-cy}dy = 1 - e^{-cx} - cxe^{-cx}$$

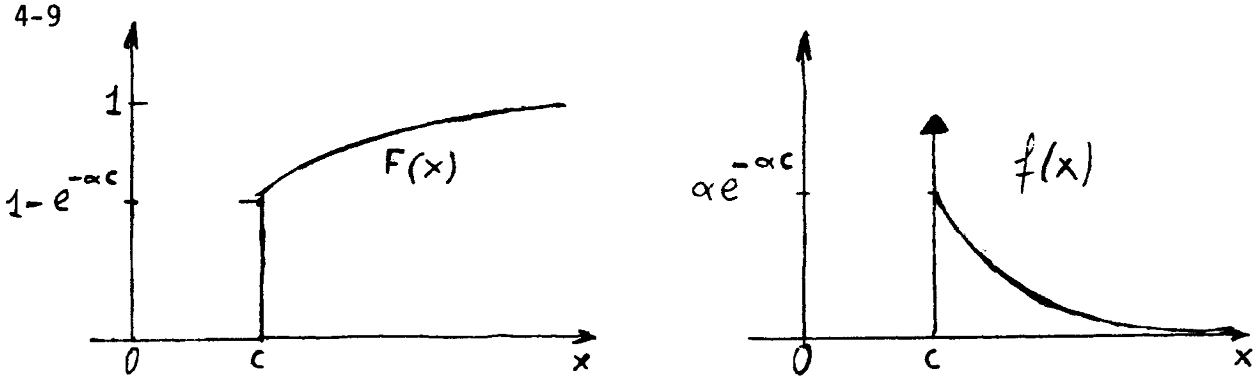
4-8 $\{(\underset{\sim}{x} - 10)^2 < 4\} = \{8 < \underset{\sim}{x} < 12\}$

$$P\{(\underset{\sim}{x} - 10)^2 < 4\} = G(12 - 10) - G(8 - 10) = 0.954$$

$$f(x|(\underset{\sim}{x} - 10)^2 < 4\} = \frac{f(x)}{P\{8 < \underset{\sim}{x} < 12\}} = \frac{1}{0.954\sqrt{2\pi}} e^{-\frac{(x-10)^2}{2}}$$

for $8 < x < 12$ and zero otherwise

4-9



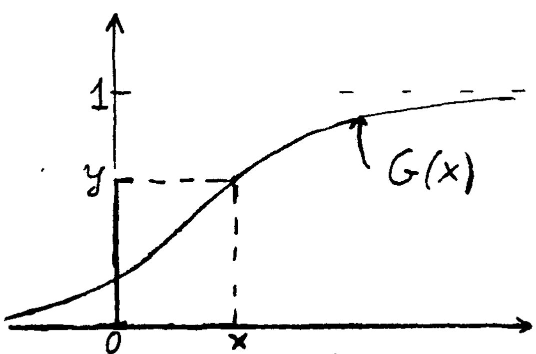
$F(x) = (1 - e^{-\alpha x})U(x-c)$ $\qquad$ $f(x) = (1 - e^{-\alpha c})\delta(x-c) + e^{-\alpha x}U(x-c)$

4-10 (a) $P\{1 \le x \le 2\} = G(\frac{2}{2}) - G(\frac{1}{2}) = 0.1499$

(b) $P\{1 \le \underset{\sim}{x} \le 2 | \underset{\sim}{x} \ge 1\} = \dfrac{G(1) - G(0.5)}{1 - G(0.5)} = \dfrac{0.1499}{0.3085} = 0.4857$

because $\{1 \le \underset{\sim}{x} \le 2, \underset{\sim}{x} \ge 1\} = \{1 \le \underset{\sim}{x} \le 2\}$

4-11



$x = G^{-1}(y) = H(y)$

If $\underset{\sim}{x}(t_1) \le x$

then

$t_i \le y = G(x)$

Hence,

$P\{\underset{\sim}{x} \le x\} = P\{t_i \le y\} = y = G(x)$

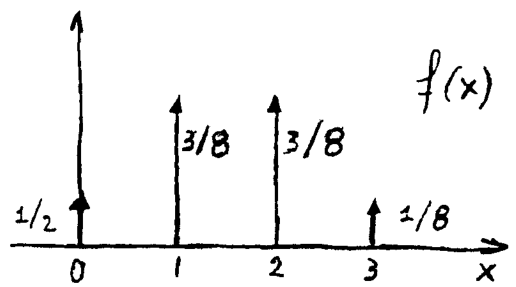
4-12 (a) $P\{x < 1024\} = G(\frac{1024 - 1000}{20}) = G(1.2) = 0.8849$

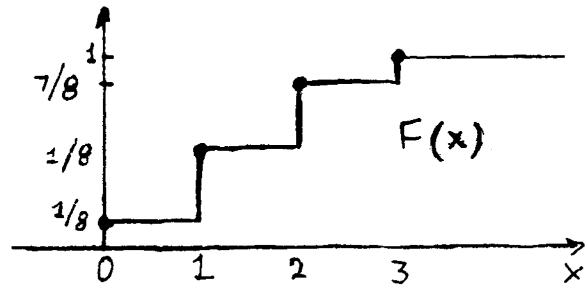
(b) $P\{x < 1024 | x > 961\} = \dfrac{P\{961 < x < 1024\}}{P\{x > 961\}}$

$$= \frac{G(1.2) - G(-1.95)}{1 - G(-1.95)} = 0.8819$$

(c) $P\{31 < \sqrt{x} \leq 32\} = P\{961 < x \leq 1024\} = 0.8593$

4-13 $P\{x = 0\} = \frac{1}{8}$ $\quad P\{x = 1\} = \frac{3}{8}$ $\quad P\{x = 2\} = \frac{3}{8}$ $\quad P\{x = 3\} = \frac{1}{8}$





4-14 (a) 1. $f_x(x) = \dfrac{1}{2^{900}} \sum_{k=0}^{900} \binom{900}{k} \delta(x-k)$

2. $f_x(x) = \dfrac{1}{15\sqrt{2\pi}} \sum_{k=0}^{900} e^{-(k-450)^2/450} \delta(x-k)$

(b) $P\{435 \leq x \leq 460\} = G(\frac{10}{15}) - G(-\frac{15}{15}) = 0.5888$

4-15 If $x > b$ then $\{x \leq x\} = S$ $\quad F(x) = 1$

If $x < a$ then $\{x \leq x\} = \{\emptyset\}$ $\quad F(x) = 0$

4-16 If $y(\zeta_i) \leq w$, then $x(\zeta_i) \leq w$ because $x(\zeta_i) \leq y(\zeta_i)$.

Hence,

$\{y \leq w\} \subset \{x \leq w\}$ $\qquad P\{y \leq w\} \leq P\{x \leq w\}$

Therefore $\qquad F_y(w) \leq F_x(w)$

4-17 From (4-80)

$$f(x) = kx\, e^{-\int_0^x kt\,dt} = kx\, e^{-kx^2/2}$$

4-18 It follows from (2-41) with

$$A_1 = \{\underset{\sim}{x} \le x\} \qquad A_2 = \{\underset{\sim}{x} > x\}$$

4-19 It follows from

$$F_x(x|A) = \frac{P\{\underset{\sim}{x} \le x, A\}}{P(A)} \qquad P\{A|\underset{\sim}{x} \le x\} = \frac{P\{\underset{\sim}{x} \le x, A\}}{P\{\underset{\sim}{x} \le x\}}$$

4-20 We replace in (4-80) all probabilities with conditional probabilities assuming $\{\underset{\sim}{x} \le x_o\}$. This yields

$$\int_{-\infty}^{\infty} P(A|\underset{\sim}{x} = x,\ \underset{\sim}{x} \le x_o) f(x|\underset{\sim}{x} \le x_o)dx = P(A|\underset{\sim}{x} \le x_o)$$

But $f(x|\underset{\sim}{x} \le x_o) = 0$ for $x > x_o$ and

$\{\underset{\sim}{x} = x,\ \underset{\sim}{x} \le x_o\} = \{\underset{\sim}{x} = x\}$ for $x \le x_o$. Hence,

$$\int_{-\infty}^{x_o} P(A|\underset{\sim}{x} = x) f(x|\underset{\sim}{x} \le x_o)dx = P(A|\underset{\sim}{x} \le x_o)$$

Writing a similar equation for $P(B|\underset{\sim}{x} \le x_o)$ we conclude that, if $P(A|\underset{\sim}{x} = x) = P(B|\underset{\sim}{x} = x)$ for $x \le x_o$, then $P(A|\underset{\sim}{x} \le x_o) = P(B|\underset{\sim}{x} \le x_o)$

4-21 (a) Clearly, $f(p) = 1$ for $0 \le p \le 1$ and 0 otherwise; hence

$$P\{0.3 \le \underset{\sim}{p} \le 0.7\} = \int_{0.3}^{0.7} dp = 0.4$$

(b) We wish to find the conditional probability $P\{0.3 \le \underset{\sim}{p} \le 0.7|A\}$ where $A = \{6$ heads in 10 tosses$\}$. Clearly $P\{A|\underset{\sim}{p}=p\} = p^6(1-p)^4$. Hence, [see (4-81)]

$$f(p|A) = \frac{p^6(1-p)^4}{\int_0^1 p^6(1-p)^4 dp} = \frac{p^6(1-p)^4}{4329\times 10^{-7}}$$

This yields

$$P\{0.3 \le \underset{\sim}{p} \le 0.7|A\} = \int_{0.3}^{0.7} f(p|A)dp = \frac{10^7}{4329}\int_{0.3}^{0.7} p^6(1-p)^4 dp = 0.768$$

4-22 (a) In this problem, $f(p) = 5$ for $0.4 \le \underset{\sim}{p} \le 0.6$ and zero otherwise; hence [see(4-82)]

$$P(H) = 5\int_{0.4}^{0.6} pdp = 0.5$$

(b) With $A = \{60$ heads in 100 tosses$\}$ it follows from (4-82) that

$$f(p|A) = p^{60}(1-p)^{40} \Big/ \int_{0.4}^{0.6} p^{60}(1-p)^{40}dp$$

for $0.4 \le p \le 0.6$ and 0 otherwise. Replacing $f(p)$ by $f(p|A)$ in (4-82), we obtain

$$P(H|A) = \int_{0.4}^{0.6} pf(p|A)dp = 0.56$$

4-23 $\quad n = 900 \quad p = q = 0.5 \quad np = 450 \quad \sqrt{npq} = 15$

$$k_1 = 420 \qquad k_2 = 465 \qquad \frac{k_2 - np}{\sqrt{npq}} = 1 \qquad \frac{k_1 - np}{\sqrt{npq}} = -2$$

$$P\{420 \le k \le 465\} = G(1) - [1 - G(-2)] = G(1) + G(2) - 1 = 0.819$$

4-24 For a fair coin $\sqrt{npq} = \sqrt{n}/2$. If

$k_1 = 0.49n$ and $k_2 = 0.52n$ then

$$\frac{k_2 - np}{\sqrt{npq}} = \frac{0.52n - n/2}{\sqrt{n}/2} = 0.04\sqrt{n} \qquad \frac{k_1 - np}{\sqrt{npq}} = -\,0.02\sqrt{n}$$

$$P\{k_1 \le k \le k_2\} = G(0.04\sqrt{n}) + G(0.02\sqrt{n}) - 1 \ge 0.9$$

From Table 4-1 (page 106) it follows that

$$0.02\sqrt{n} > 1.3 \qquad n > 65^2$$

4-25

(a) Assume n = 1,000 (Note correction to the problem)

$P(A) = 0.6 \qquad np = 600 \qquad npq = 240 \qquad k_2 = 650 \qquad k_1 = 550$

$$\frac{k_2 - np}{\sqrt{npq}} = \frac{50}{\sqrt{240}} = 3.23 \qquad\qquad \frac{k_1 - np}{\sqrt{npq}} = -3.23$$

$P\{550 \le k \le 650\} = 2G(3.23) - 1 = 0.999$

(b) $P\{0.59n \le k \le 0.61n\} = 2G\left(\frac{0.01n}{\sqrt{0.24n}}\right) - 1$

$$= 2G\left(\sqrt{\frac{n}{2400}}\right) - 1 = 0.476$$

Hence, (Table 3-1) $\quad n \simeq 9220$

4-26 With $a = 0$, $b = T/4$ it follows that

$p = 1-e^{-1/4} = 0.22 \qquad np = 220 \qquad npq = 171.6 \qquad k_2 = 100$

$$\frac{k_2 - np}{\sqrt{npq}} = -9.16 \quad \text{and (4-100) yields}$$

$P\{0 \le k \le 100\} \simeq G(-9.16) \simeq 0.$

4-27 The event

A = {k heads show at the first n tossings but not earlier}

occurs iff the following two events occur

B = {k-1 heads show at the first n-1 tossing}

C = {heads show at the nth tossing}

And since these two events are independent and

$$P(B) = \binom{n-1}{k-1} p^{k-1} q^{n-1-(k-1)} \qquad\qquad P(C) = p$$

we conclude that

$$P(A) = P(B)P(C) = \binom{n-1}{k-1} p^k q^{n-k}$$

4-28 $$-\frac{d}{dx}\left(\frac{1}{x}e^{-x^2/2}\right) = \left(1+\frac{1}{x^2}\right)e^{-x^2/2} > e^{-x^2/2}$$

Multiplying by $1/\sqrt{2\pi}$ and integrating from x to ∞, we obtain

$$\frac{1}{x\sqrt{2\pi}}e^{-x^2/2} > \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\zeta^2/2}\,d\zeta = 1 - G(x)$$

because

$$\frac{1}{x}e^{-x^2/2} \xrightarrow[x\to\infty]{} 0$$

The first inequality follows similarly because

$$-\frac{d}{dx}\left[\left(\frac{1}{x}-\frac{1}{x^3}\right)e^{-x^2/2}\right] = \left(1-\frac{3}{x^4}\right)e^{-x^2/2} < e^{-x^2/2}$$

4-29 If $P(A) = p$ then $P(\bar{A}) = 1-p$. Clearly $P_1 = 1-Q_1$ where Q_1 equals the probability that A does not occur at all. If $pn \ll 1$, then $Q_1 = (1-p)^n \simeq 1 - np \quad P_1 \simeq p$

4-30 With $p = 0.02$, $n = 100$, $k = 3$, it follows from (4-107) that the unknown probability equals

$$\binom{100}{3}(0.02)^3(0.98)^{97} \simeq \frac{2^3}{3!}e^{-2} = \frac{4}{3}e^{-2}$$

4-31 With $n = 3$, $r = 3$, $k_1 = 2$, $k_2 = 2$, $k_3 = 1$, $p_1 = p_2 = p_3 = 1/6$, it follows from (4-102) that the unknown probability equals

$$\frac{5!}{1!2!2!}\frac{1}{6^6} = 0.00386$$

4-32 With $r = 2$, $k_1 = k$, $k_2 = n-k$, $p_1 = p$, $p_2 = 1-p = q$, we obtain

$$k_1 - np_1 = k - np \qquad k_2 - np_2 = n-k-nq = np - k$$

Hence, the bracket in (4-103) equals

$$\frac{(k_1-np_1)^2}{np_1} + \frac{(k_2-np_2)^2}{np_2} = \frac{(k-np)^2}{n}\left(\frac{1}{p}+\frac{1}{q}\right) = \frac{(k-np)^2}{npq}$$

as in (4-90).

4-33 $P(M) = 2/36$ $P(\overline{M}) = 34/36$. The events M and $\overline{M}$ form a partition, hence, [see (2-41)]

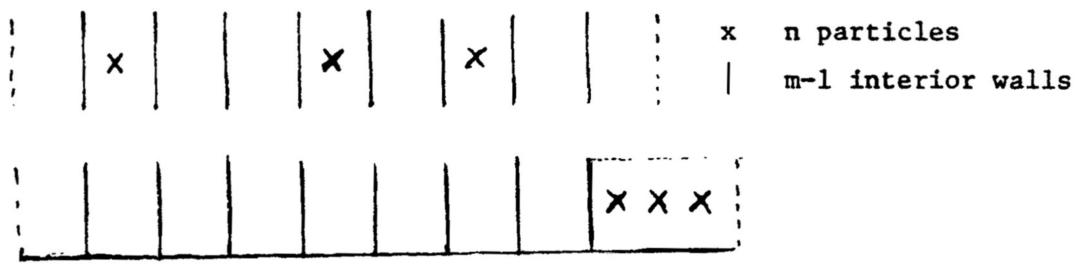
$$P(A) = P(A|M)P(M) + P(A|\overline{M})P(\overline{M}) \qquad \text{(i)}$$

Clearly, $P(A|M) = 1$ because, if M occurs at first try, X wins. The probability that X wins after the first try equals $P(A|\overline{M})$. But in the experiment that starts at the second rolling, the first player is Y and the probability that he wins equals $P(\overline{A}) = 1-p$. Hence, $P(A|\overline{M}) = P(\overline{A}) = 1-p$. And since $P(M) = 1/18$ $P(\overline{M}) = 17/18$ (i) yields

$$p = \frac{1}{18} + (1-p)\,\frac{17}{18} \qquad p = \frac{18}{35}$$

4-34

(a) Each of the n particles can be placed in any one of the m boxes. There are n particles, hence, the number of possibilities equals $N = m^n$. In the m preselected boxes, the particles can be placed in $N_A = n!$ ways (all permutations of n objects). Hence $p = n!/m^n$.

(b) 

All possibilities are obtained by permuting the n+m-1 objects consisting of the m-1 interior walls with and n particles. The (m-1)! permutations of the walls and the n! permutations of the particles must count as one. Hence

$$N = \frac{(n+m-1)!}{n!\,(m-1)!} \qquad N_A = 1$$

(c) Suppose that S is a set consisting of the m boxes. Each placing of the particles specifies a subset of S consisting of n elements (box). The number of such subsets equals $\binom{m}{n}$ (see Prob. 2-26). Hence,

$$N = \binom{m}{n} \qquad N_A = 1$$

4-35 If $k_1 + k_2 << n$, then $k_3, \simeq n$ and

$$k_3(p_1 + p_2) = [n-(k_1 + k_2)](p_1 + p_2) \simeq n(p_1 + p_2)$$

$$p_3 = 1 - (p_1 + p_2) \simeq e^{-(p_1+p_2)} \qquad p_3^{k_3} \simeq e^{-n(p_1+p_2)}$$

$$\frac{n!}{k_1!k_2!k_3!} = \frac{n(n-1)\cdots(n-k_3+1)}{k_1!k_2!} \simeq \frac{n^{k_1+k_2}}{k_1!k_2!}$$

Hence,

$$\frac{n!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3} \simeq e^{-np_1} \frac{(np_1)^{k_1}}{k_1!} e^{-np_2} \frac{(np_2)^{k_2}}{k_2!}$$

4-36 The probability p that a particular point is in the interval (0,2) equals 2/100. (a) From (3-13) it follows that the probability p_1 that only one out of the 200 points is in the interval (0,2) equals

$$p_1 = \binom{200}{1} \times 0.02 \times 0.09^{199}$$

(b) With $np = 200 \times 0.02 = 4$ and $k = 1$, (3-41) yields $p_1 \simeq e^{-4} \times 4 = 0.073$

CHAPTER 5

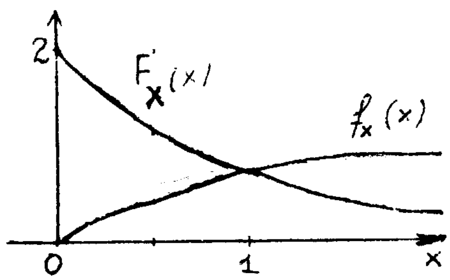
5-1 $\eta = 2\eta_x + 4 = 14 \qquad \sigma_y^2 = 4\sigma_x^2 = 16$

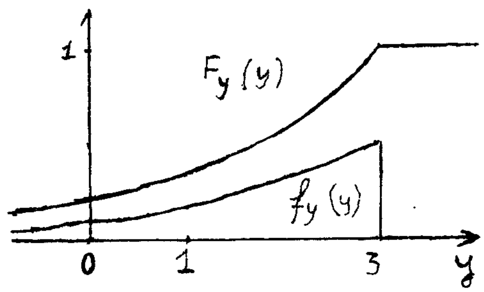
5-2 $\{\underset{\sim}{y} \le y\} = \{-4\underset{\sim}{x} + 3 \le \underset{\sim}{y}\}\ \{\underset{\sim}{x} \le (y-3)/4\}$. Hence

$$F_y(y) = P\left\{\underset{\sim}{x} \ge \frac{3-y}{4}\right\} = 1 - F_x\left(\frac{3-y}{4}\right) \qquad f_y(y) = \frac{1}{4} f_x\left(\frac{3-y}{4}\right)$$

Since $F_x(x) = (1-e^{-2x})U(x)$, this yields

$$F_y(y) = e^{(y-3)/2}U\left(\frac{y-3}{2}\right) \qquad f_y(y) = \frac{1}{2}e^{(y-3)/2}U\left(\frac{y-3}{2}\right)$$

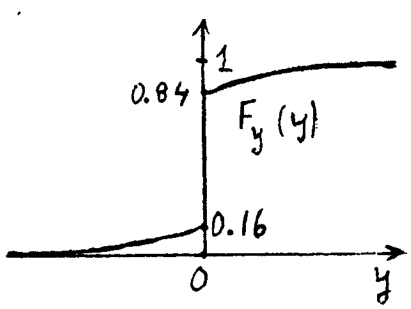


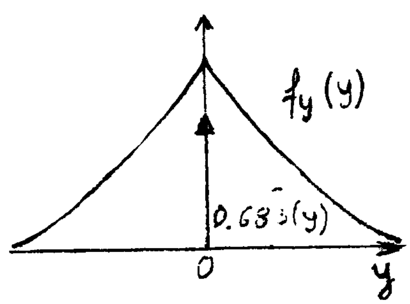


5-3 From Example 5-3 with $F_x = G(x/c)$:

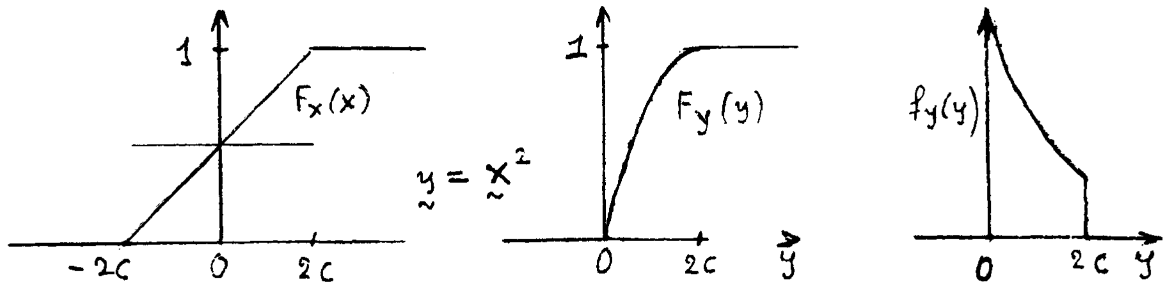
$$f_y(y) = \begin{cases} G(y/c+1) & y \ge 0 \\ G(y/c-1) & y < 0 \end{cases}$$

$$f_y(y) = 0\ 0.68\ \delta(y) + \frac{1}{c\sqrt{2\pi}}\left[e^{-(y+c)^2/2c^2}U(y) + e^{-(y-c)^2/2c^2}U(-y)\right]$$



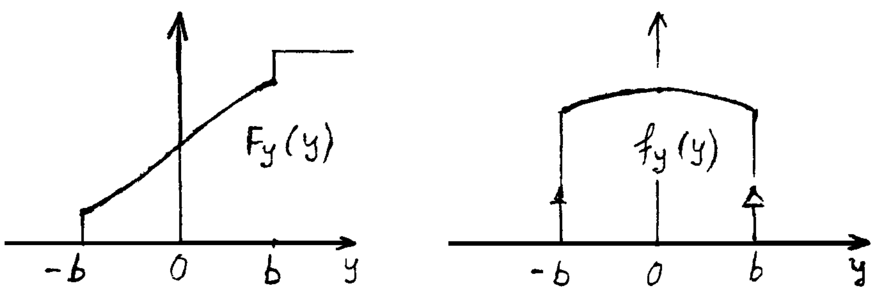


5-4 If $\underset{\sim}{y} = \underset{\sim}{x}^2$ and $F_x(x) = (x+2c)/4c$ for $|x| \le 2c$, then (see Example 5-2) $F_y(y) = \sqrt{y}\,/2c$ and $f_y(y) = 1/4\sqrt{y}$ for $0 < y < 2c$.



5-5 From Example 5-4 with $F_x(x) = G(x/b)$: For $|x| \le b$ $F_u(y) = G(y/b)$ and

$$f_y(y) = 0.16\delta(y+b) + \frac{1}{b\sqrt{2\pi}} e^{-y^2/2b^2} + 0.16\delta(y-b)$$



5-6 The equation $y = -\ell n x$ has a single solution $x = e^{-y}$ for $y > 0$ and no solutions for $y < 0$. Furthermore, $g'(x) = -1/x = -e^{y}$. Hence

$$f_y(y) = \frac{f_x(e^{-y})}{e^{y}}\, U(y) = e^{-y}U(y)$$

5-7 Clearly, $\underset{\sim}{z} \le z$ iff the number $\underset{\sim}{n}(0,z)$ of the points in the interval (0,z) is at least one. Hence,

$$F_z(z) = P\{\underset{\sim}{z} \le z\} = P\{\underset{\sim}{n}(0,z) > 0\} = 1 - P\{\underset{\sim}{n}(0,z) = 0\}$$

The probability p that a particular point is in the integral (0,z) equals z/100. With n = 200, k = 0, and p = z/100, (3-21) yields $P\{\underset{\sim}{n}(0,z) = 0\} = (1-p)^{200}$. Hence,

(a) $F_z(z) = 1 - \left(1 - \dfrac{z}{100}\right)^{100}$

(b) From (4-107) it follows thata $F_z(z) \simeq 1 - e^{-2z}$ for $z << 100$.

5.8

$$Y = \sqrt{X} \quad \Rightarrow \quad x_1 = y^2$$

$$\frac{dy}{dx} = \frac{1}{2\sqrt{x}} = \frac{1}{2y}$$

Thus

$$f_Y(y) = \frac{1}{|\frac{dy}{dx}|} f_X(x_1) = 2y f_X(y^2)$$

$$\frac{2y}{\lambda} e^{-y^2/\lambda} = \begin{cases} \frac{y}{\sigma^2} e^{-y^2/2\sigma^2}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

which represents Rayleigh density function (with $\lambda = 2\sigma^2$).

5-9 For both cases, $f_y(y) = 0$ for $y < 0$.

(a) If $y > 0$ and $|x| = y$, then $x_1 = y$, $x_2 = -y$. Hence

$$f_y(y) = [f_x(y) + f_x(-y)]U(y)$$

(b) If $y > 0$ and $e^{-x}U(x) = y$, then $x = -\ell ny$.

Furthermore, $P\{\underset{\sim}{y} = 0\} = P\{\underset{\sim}{x} \le 0\} = F_x(0)$. Hence

$$f_y(y) = F_x(0)\delta(y) + \frac{1}{y} f_x(-\ell ny)U(y)$$

5-10 (a) If $y \geq 0$ and $(x-1)U(x-1) = y$, then $\{\underset{\sim}{y} \leq y\} = \{\underset{\sim}{x} \leq y+1\}$.
If $y < 0$, then $\{\underset{\sim}{y} < y\} = \{\emptyset\}$

$$F_y(y) = F_x(1+y)U(y) = [1-e^{-2(y+1)}]U(y)$$

$$f_y(y) = (1-e^{-2})\delta(y) + 2e^{-2(y+1)}U(y)$$

(b) If $y > 0$ and $y = x^2$, then $\{\underset{\sim}{y} \leq y\} = \{-\sqrt{y} \leq \underset{\sim}{x} \leq \sqrt{y}\}$

$$F_y(y) = F_x(\sqrt{y}) - F_x(-\sqrt{y}) = (1-e^{-2\sqrt{y}})U(y)$$

$$f_y(y) = \frac{1}{\sqrt{y}} e^{-2\sqrt{y}}U(y)$$

5-11 If $y = \arctan x$, then $\dfrac{dy}{dx} = \dfrac{1}{1+x^2}$

$$f_y(y) = (1+x^2)f_x(\tan y) = \frac{1+x^2}{\pi(1+x^2)} = \frac{1}{\pi} \qquad \frac{\pi}{2} < y < \frac{\pi}{2}$$

5-12 (a) If $y = x^3$ then $x = \sqrt[3]{y}$ for any y

$$f_y(y) = \frac{1}{3\sqrt[3]{y^2}} f_x(\sqrt[3]{y}) = \frac{1}{12\pi\sqrt[3]{y^2}}$$

for $|y| < 8\pi^3$ and zero otherwise

(b) If $y = x^4$ and $y > 0$, then $x_1 = \sqrt[4]{y}$ $x_1 = -\sqrt[4]{y}$

$$f_y(y) = \frac{1}{4\sqrt[4]{y^3}}\left[f_x(\sqrt[4]{y}) + f_x(-\sqrt[4]{y})\right] = \frac{1}{8\pi\sqrt[4]{y^3}}$$

for $0 < y < 16\pi^4$ and zero otherwise

(c) If $y = 2\sin(3x + 40°)$ and $|y| < 2$ then $x = x_i$ as shown.

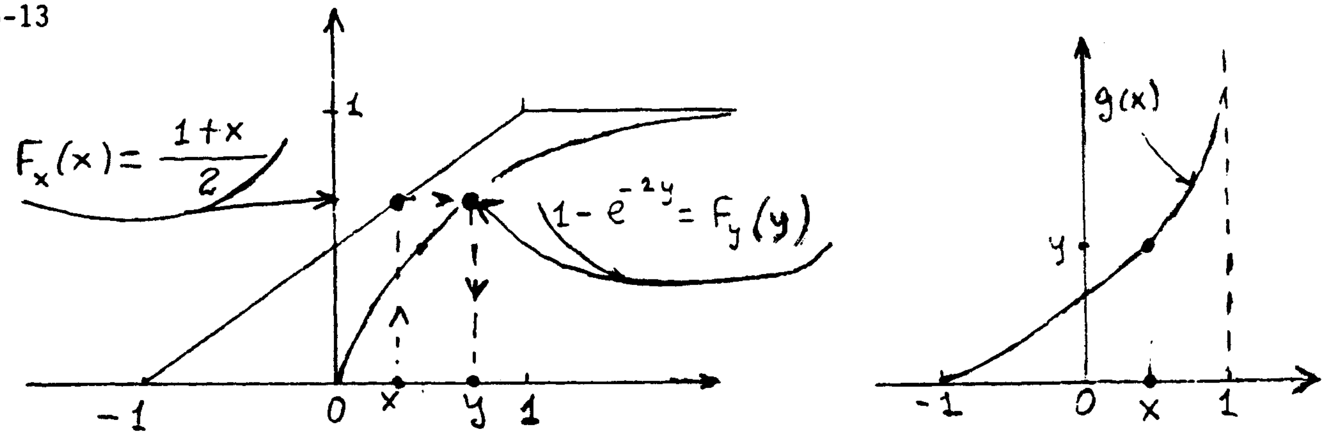
$$\frac{dy}{dx} = \frac{1}{6\sqrt{1-y^2/4}}$$

In the interval $(-2\pi, 2\pi)$ there are 12 x_i's. Hence

$$f_y(y) = \frac{1}{3\sqrt{4-y^2}} \sum_i f_x(x_i) = \frac{12}{12\pi\sqrt{4-y^2}} = \frac{1}{\pi\sqrt{4-y^2}}$$

for $|y| < 2$ and zero otherwise.

5-13



As in (5-43) $\qquad F_y[g(x)] = F_x(x)$

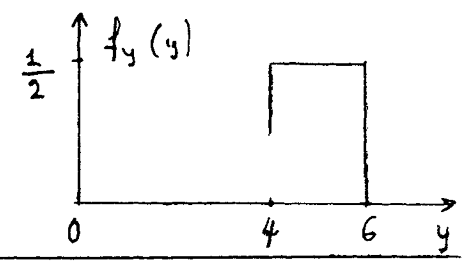
$$\frac{1+x}{2} = 1 - e^{-2y} \qquad y = g(x) = -\frac{1}{2}\ln\frac{1-x}{2}$$

for $|x| < 1$. For $x \le -1$, $g(x) = 0$; for $x \ge 1$, $g(x) = \infty$.

5-14 (a) $\quad g(x) = 2F_x(x) + 4 \qquad g'(x) = 2f_x(x)$

If $4 < y < 6$ then $y = 2F_x(x) + 4$ has a unique solution x_1 and

$$f_y(y) = \frac{f_x(x_1)}{2f_x(x_1)} = \frac{1}{2}$$



(b) Similarly $g(x) = 2F_x(x) + 8$

5-15 (a) The RV $\underset{\sim}{x}$ takes the values $k = 0,1,\dots,10$ and

$$P\{x = k\} = p_k = \binom{10}{k}\frac{1}{2^{10}} \qquad 0 \le k \le 10$$

$F_x(x)$ is a staircase function with discontinuities at the points $x = k$ and jumps equal to p_k.

(b) The RY $\underset{\sim}{y} = (\underset{\sim}{x} - 3)^2$ takes the values $y = k^2$ for $k = 0,1,\dots,7$ and probabilities $P\{y = k^2\} = q_k$.

$k =$	0	1	2	3	4	5	6	7
$q_k =$	p_3	p_2+p_4	p_1+p_5	p_0+p_6	p_7	p_8	p_9	p_{10}

5.16

$$X \sim Beta(\alpha, \beta) \quad \text{gives}$$

$$f_X(x) = \frac{1}{B(\alpha,\beta)}\, x^{\alpha-1}\,(1-x)^{\beta-1}, \quad 0 < x < 1.$$

$$Y = 1 - X \quad \Rightarrow \quad x_1 = 1 - y, \quad |\,\frac{dy}{dx}\,| = 1$$

$$\Rightarrow F_Y(y) = \frac{1}{|\frac{dy}{dx}|}\, f_X(1-y) = \begin{cases} \frac{1}{B(\beta,\alpha)}\, y^{\beta-1}\,(1-y)^{\alpha-1}, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

This gives

$$Y \sim Beta(\beta, \alpha).$$

5.17

$$X \sim \chi^2(n) \Rightarrow$$

$$f_X(x) = \frac{1}{2^{n/2}\Gamma(n/2)}\, x^{n/2-1}\, e^{-x/2}\, U(x)$$

$$y = \sqrt{x} \quad \Rightarrow \quad x_1 = y^2$$

$$\frac{dy}{dx} = \frac{1}{2y}$$

Thus

$$f_Y(y) = 2y f_X(y^2) = \frac{y^{n-1}}{2^{n/2-1}\,\Gamma(n/2)}\, e^{-y^2/2}\, U(y)$$

and it represents the chi-distribution.

5.18

$$X \sim U(0,1)$$

$$Y = -2logX \quad \Rightarrow \quad x_1 = e^{-y/2}$$

$$\frac{dy}{dx} = -\frac{2}{x} = -2e^{y/2}$$

$$\begin{aligned} f_Y(y) \;&= \frac{1}{|\frac{dy}{dx}|}\, f_X(x_1) = \tfrac{1}{2} e^{-y/2}\, U(y) \\ &\sim \text{ Exponential}(2) \equiv \chi^2(2) \end{aligned}$$

5.19

$$f_X(x) = \lambda\, e^{-\lambda x}\, u(x)$$

$$Y = X^{1/\beta} \quad \Rightarrow \quad x_1 = y^\beta$$

$$|\frac{dy}{dx}| = \frac{1}{\beta}\, x^{1/\beta-1} = \frac{1}{\beta}\, y^{1-\beta}$$

$$f_Y(y) = \frac{1}{|\,\frac{dy}{dx}\,|}\, f_X(x_1) = \lambda\,\beta\, y^{\beta-1}\, e^{-\lambda y^\beta}\, U(y)$$

and it represents Weibull distribution

5-20 For $|y| < a$ the equation $y = a \sin\omega t$ has infinitely many solutions τ_i; in each interval of length $2\pi/\omega$ there are two such solutions. Furthermore, $y'(t) = \omega\sqrt{a^2 - y^2}$

$$\tau_i = \frac{1}{\omega}\sin^{-1}\frac{y}{a} \qquad\qquad \tau_{i+2} - \tau_i \quad \frac{2\pi}{\omega} \xrightarrow[\omega\to\infty]{} 0$$

Hence,

$$\frac{1}{\omega\sqrt{a^2-y^2}} \sum_{i=-\infty}^{\infty} f_t(\tau_i) \xrightarrow[\omega\to\infty]{} \frac{1}{\sqrt{a^2-y^2}} \; \frac{2}{2\pi} \int_{-\infty}^{\infty} f_t(\tau)\,d\tau = \frac{1}{\pi\sqrt{a^2-y^2}}$$

5-21 If $y > 0$ then

$$F_y(y|x \geq 0) = F_x(\sqrt{y}|x \geq 0) + F_x(-\sqrt{y}|x \geq 0) = F_x(\sqrt{y}|x \geq 0)$$

$$F_x(\sqrt{y}|x \geq 0) = \frac{P\{0 < x < \sqrt{y}\}}{P\{x \geq 0\}} = \frac{F_x(\sqrt{y}) - F_x(0)}{1 - F_x(0)}$$

$$f_y(y|x \geq 0) = \frac{d}{dy} F_y(\sqrt{y}|x \geq 0) = \frac{f_x(\sqrt{y})}{2\sqrt{y}[1-F_x(0)]}$$

5-22 (a) $\eta_y = a\,\eta_x + b \qquad\qquad \sigma_y^2 = E\{[a\,x + b - (a\,\eta_x + b)]^2\}$

$$\sigma_y^2 = E\{a^2(x - \eta_x)^2\} = a^2\sigma_x^2$$

(b) $y = \dfrac{x-\eta_x}{\sigma_x} \qquad\qquad E\{y\} = 0 \qquad\qquad \sigma_y^2 = \dfrac{\sigma_x^2}{\sigma_x^2} = 1$

5-23 If x has a Rayleigh density, then [see (5-76)]

$$E\{x^2\} = 2\alpha^2 \qquad\qquad E\{x^4\} = 8\alpha^4$$

If $y = b + cx^2$, then

$$E\{y\} = b + 2\alpha^2 c \qquad\qquad E\{y^2\} = b^2 + 4\alpha^2 bc + 8\alpha^4 c^2$$

$$\sigma_y^2 = E\{y^2\} - E^2\{y\} = 4\alpha^4 c^2$$

5-24 $\quad y = 3x^2 \qquad E\{x^2\} = \sigma_x^2 = 4 \qquad E\{x^4\} = 3\sigma_x^4 = 48$

$E\{y\} = 12 \qquad E\{y^2\} = 9 \times 48 = 432 \qquad \sigma_y^2 = 432 - 144 = 288$

If $y > 0$ then $3x^2 = y$ for $x = \pm\sqrt{y/3} \qquad y' = 6x$

$$f_y(y) = \frac{\cancel{2}\,4}{\sqrt{12y}}\, f_x\left(\sqrt{\frac{y}{3}}\right) = \frac{1}{\sqrt{24\pi y}}\, e^{-y/24} U(y)$$

5.25

$$X \sim B(n,p) \Rightarrow P(X=k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\cdots n.$$

a)

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k\, P(X=k) = \sum_{k=1}^{n} k \frac{n!}{k!\,(n-k)!} p^k q^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,(n-k)!} p^{k-1} q^{n-k} \\ &= np\,(p+q)^{n-1} = np. \end{aligned}$$

b)

$$\begin{aligned} E[X(X-1)] &= \sum_{k=2}^{n} k\,(k-1) \frac{n!}{k!\,(n-k)!} p^k q^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!\,(n-k)!} p^{k-2} q^{n-k} \\ &= n(n-1)p^2\,(p+q)^{n-2} \\ &= n(n-1)p^2 \end{aligned}$$

c)

$$\begin{aligned} E[X(X-1)(X-2)] &= \sum_{k=3}^{n} k(k-1)(k-2) \frac{n!}{k!\,(n-k)!} p^k q^{n-k} \\ &= n(n-1)(n-2)\,p^3 \sum_{k=3}^{n} \frac{(n-3)!}{(k-3)!\,(n-k)!} p^{k-3} q^{n-k} \\ &= n(n-1)(n-2)\,p^3\,(p+q)^{n-3} \\ &= n(n-1)(n-2)\,p^3 \end{aligned}$$

$$E(X^2) = E\,(X(X-1)) + E(X) = n^2p^2 + npq$$

$$\begin{aligned} E(X^3) &= E\,(X(X-1)(X-2)) + 3E(X^2) - 2E(X) \\ &= n(n-1)(n-2)\,p^3 + 3(n^2p^2 + npq) - 2np \\ &= n^3p^3 + 3n^2p^2q + npq(q-p). \end{aligned}$$

5.26

$$X \sim P(\lambda) \quad \Rightarrow \quad P(X = k) = e^{\lambda}\,\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \cdots$$

a)

$$E(X) = \lambda, \quad \mathrm{Var}(X) = \sigma_X^2 = \lambda$$

From Chebyshev's inequality (5-88)

$$P\left(|X - \mu| < \lambda\right) > 1 - \frac{\sigma^2}{\lambda^2} = 1 - \frac{1}{\lambda}$$

But

$$|X - \mu| < \lambda = |X - \lambda| < \lambda \quad \Rightarrow \quad 0 < X < 2\lambda$$

which gives

$$P(0 < X < 2\lambda) > 1 - \frac{1}{\lambda} = \frac{\lambda - 1}{\lambda}.$$

b)

$$\begin{aligned} E\left[X(X-1)\right] &= \textstyle\sum_{k=2}^{\infty} k(k-1)\, e^{-\lambda}\,\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\,\lambda^2 \textstyle\sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = e^{-\lambda}\,\lambda^2\,e^{\lambda} = \lambda^2. \end{aligned}$$

$$\begin{aligned} E\left[X(X-1)(X-2)\right] &= \textstyle\sum_{k=3}^{\infty} k(k-1)(k-2)\, e^{-\lambda}\,\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\,\lambda^3 \textstyle\sum_{k=3}^{\infty} \frac{\lambda^{k-3}}{(k-3)!} = \lambda^3. \end{aligned}$$

5-27 **Follows from (4-74)**

$$E\{\underline{x}\} = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{\infty} x \sum_i f(x|A_i)P(A_i)\,dx$$

because $$E\{\underline{x}|A_i\} = \int_{-\infty}^{\infty} x f(x|A_i)\,dx$$

5-28 **From (5-89) with** $\alpha = \sqrt{\eta}$:

$$P\{\underline{x} \geq \sqrt{\eta}\} \leq \eta/\sqrt{\eta} = \sqrt{\eta}$$

5-29 From (5-86) with $g(x) = x^3$ $g''(x) = 6x$:

$$E\{\underset{\sim}{x}^3\} = \eta^3 + 6\eta\frac{\sigma^2}{2} = 1120$$

5-30 (a) If $y = x^3$, then $x = \sqrt[3]{y}$ $g'(x) = 3x^2 = 3\sqrt[3]{y^2}$

But $f_x(x) = 0.5$ for $10 < x < 12$, i.e., for $10^3 < y < 12^3$

and (5-16) yields

$$f_y(y) = \frac{0.5}{3\sqrt[3]{y^2}} \qquad 10^3 < y < 12^3$$

and zero otherwise.

(b) 1.

$$E\{\underset{\sim}{x}^3\} = 0.5\int_{10}^{12} x^3 dx = 1342$$

2. With $g(x) = x^3$ $E\{\underset{\sim}{x}\} = 11$ $\sigma_x^2 = 1/3$, (5-86) yields

$$E\{x^3\} \simeq 11^3 + 6 \times 11 \times \frac{1}{6} \simeq 1342$$

5-31 With $g(x)=1/x$, $g''(x)=2/x^3$, $\eta=100$, and $\sigma=3$, (5-55) yields

$$E\left\{\frac{1}{\underset{\sim}{x}}\right\} \simeq \frac{1}{100} + \frac{9}{2} \times \frac{2}{100^3} = 0.010009$$

5-32

$$\frac{\partial|x-a|}{\partial a} = \begin{cases} 1 & x < a \\ -1 & x > a \end{cases} \qquad \text{If } I(a) = E\{|\underset{\sim}{x}-a|\} \text{ then}$$

$$\frac{dI(a)}{da} = E\,\frac{\partial|\underset{\sim}{x}-a|}{\partial a} = 1\,P\{\underset{\sim}{x} < a\} - 1\,P\{\underset{\sim}{x} > a\}$$
$$= 2\,F(a) - 1$$

(a)
$$I(a) = I(m) + \int_m^a I'(\alpha)d\alpha = I(m) + \int_m^a [2\,F(\alpha) - 1]d\alpha$$

$$= E\{|\underset{\sim}{x} - m|\} - 2\int_m^a x\,f(x)dx$$

because

$$\int_m^a F(\alpha)d\alpha = a\,F(a) - m\,F(m) - \int_m^a x f(x)dx$$

$$F(m) = \frac{1}{2} \qquad\qquad \int_m^a f(x)dx = F(a) - F(m)$$

(b) $I(a) = E\{|\underset{\sim}{x} - a|\}$ is minimum if

$I'(a) = 2F(a) - 1 = 0$ i.e. if $F(a) = \frac{1}{2}$ $\quad a = m$

5-33
$$E\{|\underset{\sim}{x}|\} = \int_0^\infty xf(x)dx - \int_{-\infty}^0 xf(x)dx$$

$$\eta = E\{\underset{\sim}{x}\} = \int_0^\infty xf(x)dx + \int_{-\infty}^0 xf(x)dx$$

$$\frac{E\{|\underset{\sim}{x}|+\eta\}}{2} = \int_0^\infty xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty x\,e^{-(x-\eta)^2/2\sigma^2}dx$$

$$\frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty (x+\eta)e^{-(x-\eta)^2/2\sigma^2}dx = \frac{1}{\sigma\sqrt{2\pi}}\int_\eta^\infty y\, e^{-y^2/2\sigma^2}dy = \frac{\sigma}{\sqrt{2\pi}}\, e^{-\eta^2/2\sigma^2}$$

$$\frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty e^{-(x-\eta)^2/2\sigma^2}dx = G(\frac{\eta}{\sigma})$$

Multiplying the last line by η and subtracting from the fourth line, we obtain

$$\frac{E\{|x|+\eta\}}{2} = \frac{\sigma}{\sqrt{2\pi}}\, e^{-\eta^2/2\sigma^2} + G(\frac{\eta}{\sigma})$$

5-34 The proof is given in sec 14-3: [see (14-100)].

5-35 (a) Follows from (5-89) (b) $e^{sx} \geq e^{sA}$ iff $x \geq A$ for $s > 0$ and $x \leq A$ for $s < 0$.

5.36 See proof for Lyapunov inequality (Ch.5, Eq.(5-92).)

5-37 (a) If $\Phi(\omega) = e^{-\alpha|\omega|}$ then [see (5-102)]

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\alpha|\omega|}\, e^{j\omega x} d\omega = \frac{1}{\pi}\int_{0}^{\infty} \cos\omega x\, e^{-\alpha\omega} d\omega = \frac{\alpha}{\pi(\alpha^2+x^2)}$$

(b) If $f(x) = \frac{\alpha}{2} e^{-\alpha|x|}$, then [see (5-94)]

$$\Phi(\omega) = \frac{\alpha}{2}\int_{-\infty}^{\infty} e^{-\alpha|x|} e^{-j\omega x} dx = \alpha\int_{0}^{\infty} e^{-\alpha x}\cos\omega x\, dx = \frac{\alpha^2}{\alpha^2+\omega^2}$$

5.38 a) On comparing Eq.(4-34) with Eq.(5-106), Example 5-29, we get

$$X \sim G(\alpha,\beta) \quad \Rightarrow \quad \phi_X(\omega) = (1-j\beta\omega)^{-\alpha}$$

$$\phi_X'(\omega) = -\alpha(1-j\beta\omega)^{-(\alpha+1)}\,(-j\beta)$$

so that

$$E(X) = \frac{1}{j}\,\phi_X'(0) = \alpha\,\beta.$$

Similarly

$$\phi_X''(\omega) = j\alpha\beta(\alpha+1)\,(1-j\beta\,\omega)^{-(\alpha+2)}\,(j\beta)$$

and hence

$$E(X^2) = \frac{1}{j^2}\,\phi_X''(0) = \alpha\beta^2\,(\alpha+1).$$

Thus

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \alpha\beta^2.$$

b)

$$X \sim \chi^2(n) \quad \Rightarrow \quad \alpha = \frac{n}{2}, \quad \beta = 2$$

in Gamma(α,β). This gives

$$\begin{aligned} \phi_X(\omega) &= (1-j2\omega)^{-n/2} \\ E(X) &= n \\ \mathrm{Var}(X) &= 2n. \end{aligned}$$

c)

$$X \sim \; B(n,p).$$

From Prob 5-25 (a)-(b)

$$\begin{aligned} E(X) &= np \\ \mathrm{Var}(X) &= E\left(X(X-1)\right) + E(X) = npq. \end{aligned}$$

and

$$\begin{aligned} \phi_X(\omega) &= \sum_{k=0}^{n} e^{jk\omega}\,P(X=k) \\ &= \sum_{k=0}^{n} \binom{n}{k}\,(p\,e^{j\omega})^k\,q^{n-k} = (pe^{j\omega}+q)^n. \end{aligned}$$

d)

$$X \sim N\,Binomial\,(r, p).$$

From (4-64)

$$\begin{aligned}\phi_X(\omega) &= \sum_{k=0}^{\infty} e^{jk\omega}\, P(X=k) \\ &= \sum_{k=0}^{\infty} \binom{r+k-1}{k} p^r\,(qe^{j\omega})^k \\ &= p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-qe^{j\omega})^k \\ &= p^r\,(1-qe^{j\omega})^{-r}.\end{aligned}$$

5-39

$$\Gamma(z) = \sum_{k=0}^{\infty} p\, q^k z^k = \frac{p}{1-qz} \qquad q = 1-p$$

$$\Gamma'(z) = \frac{pq}{(1-qz)^2} \qquad \Gamma'(1) = \frac{pq}{(1-q)^2} = \frac{p}{q} = \eta_x$$

$$\Gamma''(z) = \frac{2pq^2}{(1-qz)^3} \qquad \Gamma''(1) = \frac{2q^2}{p^2} = m_2 - m_1$$

$$\sigma^2 = m_2 - m_1^2 = 2\frac{q^2}{p^2} + m_1 - m_1^2 = \frac{q}{p^2}$$

5-40

$$\Gamma(z) = p^n \sum_{k=0}^{\infty} \binom{-n}{k} (-q)^k\, z^k = p^n (1-qz)^{-n}$$

(binomial expansion with negative exponent)

$$\Gamma'(z) = \frac{n\,p^n q}{(1-qz)^{n+1}} \qquad \Gamma'(1) = \frac{nq}{p} = \eta_x$$

$$\Gamma''(z) = \frac{n(n+1)p^n q^2}{(1-qz)^{n+2}} \qquad \Gamma''(1) = \frac{n(n+1)q^2}{p^2} = m_2 - m_1$$

$$\sigma_x^2 = \Gamma''(1) + m_1 - m_1^2 = \frac{nq}{p^2}$$

5.41 We have

$$P(X=k)=\binom{k-1}{r-1}p^r\,q^{k-r},\quad k=r,r+1,\cdots$$

Let $k=n+r$ so that

$$\begin{aligned}
P(X=n+r) &= \binom{n+r-1}{r-1}p^r\,q^n,\quad n=0,1,2,\cdots\\
&= \frac{(n+r-1)!}{n!\,(r-1)!}\,p^r\,(1-p)^n\\
&= \frac{1}{n!}\,\frac{(n+r-1)(n+r-2)\cdots(r)}{r^n}\,[r(1-p)]^n\,p^r\\
&= \frac{\lambda^n}{n!}\left\{\left(1+\frac{n-1}{r}\right)\left(1+\frac{n-2}{r}\right)\cdots\right\}\left(1-\frac{r(1-p)}{r}\right)^r\\
&= \frac{\lambda^n}{n!}\left\{\prod_{k=1}^{n}\left(1+\frac{n-k}{r}\right)\right\}\left(1-\frac{\lambda}{r}\right)^r,
\end{aligned}$$

where $\lambda=r(1-p)$. Thus

$$\lim_{r\to\infty}P(X=n+r)=\frac{\lambda^n}{n!}\left\{\lim_{r\to\infty}\prod_{k=1}^{n}\left(1+\frac{n-k}{r}\right)\right\}\lim_{r\to\infty}\left(1-\frac{\lambda}{r}\right)^r$$

$$\to\frac{\lambda^n}{n!}e^{-\lambda}\sim P(\lambda).$$

5-42 $E\{e^{s\underline{x}}\} = e^{s\eta} E\{e^{s(\underline{x}-\eta)}\} = e^{s\eta} E\left\{\sum_{n=0}^{\infty} \frac{s^n}{n!} (\underline{x}-\eta)^n\right\}$

$$= e^{s\eta} \sum_{n=0}^{\infty} \frac{s^n}{n!} \mu_n$$

5-43 If $\Phi(\omega_1) = 0$, then [see also (9-176]

$$\int_{-\infty}^{\infty} (1-e^{j\omega_1 x}) f(x)dx = 0, \text{ hence, } f(x) = \sum_{n=\infty}^{\infty} p_n \delta(x - \frac{2\pi n}{\omega_1})$$

5-44 (a) If $\eta = 0$, then $m_n = \mu_n \qquad \lambda_1 = \eta = 0$

$$\Phi(s) = \sum_{n=0}^{\infty} \frac{\mu_n}{n!} s^n \qquad\qquad \Psi(s) = \sum_{n=2}^{\infty} \frac{\lambda_n}{n!} s^n$$

$$1 + \frac{\mu_2}{2!}s^2 + \frac{\mu_3}{3!} s^3 + \frac{\mu_4}{4!} s^4 + \cdots = \exp\left\{\frac{\lambda_2}{2!} s^2 + \frac{\lambda_3}{3!} s^3 + \frac{\lambda_4}{4!} s^4 + \cdots\right\}$$

Expanding the exponential and equating powers of s, we obtain

$$\mu_2 = \lambda_2 \qquad \mu_3 = \lambda_3 \qquad\qquad \frac{\mu_4}{4!} = \frac{\lambda_4}{4!} + \frac{1}{2!} \left(\frac{\lambda_2}{2!}\right)^2$$

(b) If $\underline{y}$ is $N(0;\sigma_y)$ then

$$\Psi_y(s) = \frac{\lambda_2}{2} s^2 , \quad \text{hence,} \quad \lambda_n = 0 \quad \text{for} \quad n \geq 3$$

5-45 $$P\{\underset{\sim}{y} = 0\} = P\{\underset{\sim}{x} \le 1\} = p_o + p_1$$

$$P\{\underset{\sim}{y} = k\} = P\{\underset{\sim}{x} = k + 1\} = p_{k+1} \qquad k \ge 1$$

$$\Gamma_y(z) = p_o + p_1 + \sum_{k=1}^{\infty} p_{k+1}\, z^k = p_o + z^{-1}[\Gamma_x(z) - p_o]$$

$$\eta_y = \sum_{k=1}^{\infty} k\, p_{k+1} = \sum_{r=1}^{\infty} r\, p_r - \sum_{r=1}^{\infty} p_r = \eta_x - 1 + p_o$$

$$E\{\underset{\sim}{y}^2\} = \sum_{k=1}^{\infty} k^2\, p_{k+1} = \sum_{r=1}^{\infty} (r-1)^2\, p_r = E\{\underset{\sim}{x}^2\} - 2\eta_x + 1 - p_o$$

5-46 $$0 \le E\left\{\left|\sum_{i=1}^{n} a_i\, e^{j\omega_i \underset{\sim}{x}}\right|^2\right\} = E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^*\, e^{j(\omega_i - \omega_j)\underset{\sim}{x}}\right\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^*\, \Phi(\omega_i - \omega_j)$$

5-47 From the assumptions it follows that

$$g'(-x) = -g'(x) \qquad g''(x) \ge 0 \qquad f(x-\eta) = f(\eta-x)$$

Hence, if $I(a) = E\{g(\underset{\sim}{x}-a)\}$, then

$$I'(a) = -\int_{-\infty}^{\infty} g'(x-a)f(x)\,dx \qquad I'(\eta) = 0$$

$$I''(a) = \int_{-\infty}^{\infty} g''(x-a)f(x)\,dx \ge 0 \qquad \text{all } a$$

Hence, $I(a)$ is minimum for $a = \eta$.

5-48 $$f(x,v) = \frac{1}{\sqrt{2\pi v}}\, e^{-x^2/2v}$$

$$\sqrt{2\pi}\,\frac{\partial f}{\partial v} = \frac{-1 + x^2/v}{2v\,\sqrt{v}}\, e^{-x^2/2v}$$

$$\sqrt{2\pi}\,\frac{\partial^2 f}{\partial x^2} = \frac{-1 + x/.}{v\,\sqrt{v}}\, e^{-x^2/2v}$$

Hence

(see also (6-198) - (6-199)) $$\boxed{\frac{\partial f}{\partial v} = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}} \qquad (1)$$

(a) Integrating by parts, using (1) and assuming that $g^{(k)}(x)f(x) \to 0$ as $|x| \to \infty$, $k = 0, 1, 2$, we obtain

$$E\{g''(x)\} = \int_{-\infty}^{\infty} \frac{d^2g}{dx^2}\, f dx = \int_{-\infty}^{\infty} g\, \frac{\partial^2 f}{\partial x^2}\, dx = 2\int_{-\infty}^{\infty} g\, \frac{\partial f}{\partial v}\, dx$$

$$= 2\,\frac{d}{dv}\int_{-\infty}^{\infty} g\, f dx = 2\,\frac{d}{dv}\, E\{g(\underset{\sim}{x})\}$$

(b) The moments $\mu_n(u) = E\{\underset{\sim}{x}^n\}$ of $\underset{\sim}{x}$ depend on the variance v of $\underset{\sim}{x}$ and (i) yields

$$\mu_n'(v) = \frac{d}{dv}\, E\{\underset{\sim}{x}^n\} = \frac{1}{2}\, E\{n(n-1)\underset{\sim}{x}^{n-2}\} = \frac{n(n-1)}{2}\, \mu_{n-2}(v)$$

Furthermore, $\mu_n(0) = 0$ because, if $v = 0$, then $\underset{\sim}{x} = 0$.

Hence
$$\mu_n(v) = \frac{n(n-1)}{2}\int_0^v \mu_{n-2}(\beta)\, d\beta$$

5-49 **The function**

$$\Gamma(e^{j\omega}) = E\{e^{j\mathbf{x}\omega}\} = \sum_{k=0}^{\infty} p_k \, e^{jk\omega}$$

is periodic with period 2π and Fourier series coefficients $p_k = E\{\mathbf{x} = k\}$.

5.50 The event $\{X = 1\}$ is given by the disjoint union "$TH \cup HT$". Similarly, the event "$X = k$" is given by the union of the disjoint events (k "T"s followed by "H" or k "H"s followed by "T")

$$\text{"}TT\cdots TTH\text{"} \cup \text{"}HH\cdots HHT\text{"}, \qquad k = 1, 2, \cdots$$

Thus

$$\begin{aligned} P(X = k) &= P(\text{"}TT\cdots TH\text{"} \cup \text{"}HH\cdots HT\text{"}) \\ &= P(TT\cdots TH) + P(HH\cdots HT) = q^k p + p^k q, \quad k = 1, 2, \cdots \end{aligned}$$

Also

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} kP(X = k) \\ &= \sum_{k=1}^{\infty} kq^k p + \sum_{k=1}^{\infty} kp^k q = pq\left\{\sum_{k=1}^{\infty} kq^{k-1} + \sum_{k=1}^{\infty} kp^{k-1}\right\} \\ &= pq\left\{\frac{\partial}{\partial q}\sum_{k=1}^{\infty} q^k + \frac{\partial}{\partial p}\sum_{k=1}^{\infty} p^k\right\} = pq\left\{\frac{\partial}{\partial q}\left(\frac{q}{1-q}\right) + \frac{\partial}{\partial p}\left(\frac{p}{1-p}\right)\right\} \\ &= pq\left\{\frac{1}{p^2} + \frac{1}{q^2}\right\} = \frac{p}{q} + \frac{q}{p}. \end{aligned}$$

5.51 (a) When samples are drawn with replacement, probability of each item being defective is given by

$$p = \frac{M}{N} < 1 \quad \text{(constant)}$$

and

$$q = 1 - p = \frac{N-M}{M} < 1$$

represents the constant probability that the chosen item is not defective. In that case (with replacement), there are $\binom{n}{k}$ possible ways of arranging k defective items among n chosen items, and each such arrangement has probability $p^k q^{n-k}$. This gives

$$P(X = k) = \binom{n}{k} p^k \, q^{n-k}, \quad k = 0, 1, 2, \cdots n$$

which represents the Binomial distribution.
(b) If the samples are drawn without replacement, there are $\binom{M}{k}$ possible ways of choosing k defective item from a total of M defective items, and $\binom{N-M}{n-k}$ possible ways of choosing $n-k$ "good" items from $(N-M)$ "good" items independently. This gives

$$\binom{M}{k}\binom{N-M}{n-k}$$

to be the total number of ways of selecting k defective items and $n-k$ "good" items from a subsample of M and $N-M$ items respectively (favorable ways). But there are a total of $\binom{N}{n}$ ways of selecting n items among N items. This gives

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},$$

since $0 \le k \le M$ and $n - k \le N - M, n - k \ge 0$, i.e. $0 \le k \le M, k \le n, k \ge n + M - N$.
(c) From (b)

$$P(X = k) = \frac{M!}{k!(M-k)!} \, \frac{(N-M)!}{(n-k)!(N-M-n+k)!} \, \frac{n!(N-n)!}{N!}$$

$$= \binom{n}{k} \frac{M(M-1)\cdots(M-k+1)}{N(N-1)\cdots(N-k+1)} \, \frac{(N-M)(N-M-1)\cdots(N-M-n+k+1)}{(N-k)(N-k-1)\cdots(N-n+1)} \, \frac{(I}{(I}$$

$$\simeq \binom{n}{k} \left(\tfrac{M}{N}\right)^k \left(\tfrac{N-M}{N}\right)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \cdots n$$

since $N \to \infty, M \to \infty$ such that $M/N \to p$, and $n \ll N$. Thus

$$P(X = k) \to \text{Binomial}(n, p = M/N)$$

under the above conditions.

5.52 (a) Refer to discussions in problem 5.51 (a) if sampleing is done with replacement, then

$$p = \frac{n}{n+m}$$

represents the probability of selecting a white marble on any trial. The event "$X = k$" is given by "$r-1$ white mables among the first $k-1$ trials" followed by "a white marble at the k^{th} trial". But from problem 5.51 (a), the event $r-1$ white mables among the first $k-1$ trials has a binomial distribution whose probability is given by $\binom{k-1}{r-1}p^{r-1}q^{k-r}$. Thus

$$P(X = k) = \binom{k-1}{r-1}p^{r-1}q^{k-r}p = \binom{k-1}{r-1}p^r q^{k-r}, \quad k = r, r+1, \cdots$$

which represents the Negative-binomial distribution

(b) If sampling is done with replacement, then the favarable ways of choosing the white balls are given by:

(i) $\binom{k-1}{r-1}$ ways of selecting $r-1$ white balls among the first $k-1$ trials/balls.

(ii) One ways of selecting (the r^{th}) white ball at the k^{th} trial

(iii) $\binom{m+n-k}{n-r}$ ways of selecting the remaining $n-r$ white balls among the remaining $m+n-k$ balls.

This gives $\binom{k-1}{r-1}\cdot 1 \cdot \binom{m-n-k}{n-r}$ to be the total number of favorable ways of selecting the white balls. Since there are $n+m$ balls there are a total of $\binom{n+m}{n}$ ways of selecting n white balls. This gives

$$P(X = k) = \binom{k-1}{r-1}\frac{\binom{m+n-k}{n-r}}{\binom{n+m}{n}}, \qquad k = r, r+1, \cdots$$

(c) From (b)

$$P(X = k) = \binom{k-1}{r-1}\frac{(m+n-k)!}{(n-r)!\,(m-k+r)!}\frac{n!m!}{(m+n)!}$$

$$= \binom{k-1}{r-1}\left(\frac{n}{m+n}\right)\left(\frac{n-1}{m+n-1}\right)\cdots\left(\frac{n-r+1}{m+n-r+1}\right)\left(\frac{m!(m+n-k)!}{(m+n-r)!\,(m-k+r)!}\right)$$

$$\simeq \binom{k-1}{r-1}\left(\frac{n}{m+n}\right)^r\left(\frac{m}{m+n-r}\right)\left(\frac{m-1}{m+n-r-1}\right)\cdots\left(\frac{m-k+r+1}{m+n-k+1}\right)$$

$$\simeq \binom{k-1}{r-1}\left(\frac{n}{m+n}\right)^r\left(\frac{m}{m+n}\right)^{k-r} \text{ as } m+n \to \infty$$

$$= \binom{k-1}{r-1}p^r q^{k-r}, \qquad k = r, r+1, \cdots, \quad q = 1-p$$

$$\sim \mathrm{NB}\left(r, p = n/(n+m)\right).$$

CHAPTER 6

6.1 (a) Define

$$Z = X + Y$$

Note that both X and Y positive random variables hence (use Eq. (6-45))

$$\begin{aligned} f_Z(z) &= \int_0^z f_{XY}(z-y, y)dy = \int_0^z e^{-(z-y+y)}dy \\ &= z\, e^{-z} U(z). \end{aligned}$$

(b)

$$Z = X - Y$$

Z ranges over the entire real axis for the random variables X and Y (see Eq. (6-55))

$$F_Z(z) = \begin{cases} \displaystyle\int_0^\infty \int_0^{z+y} f_{XY}(x,y)dx\,dy, & z > 0 \\ \displaystyle\int_{-z}^\infty \int_0^{z+y} f_{XY}(x,y)dx\,dy, & z < 0 \end{cases}$$

Differrentiation gives

$$f_Z(z) = \begin{cases} \displaystyle\int_0^\infty f_{XY}(z+y, y)\,dy, & z > 0 \\ \displaystyle\int_{-z}^\infty f_{XY}(z+y, y)\,dy, & z < 0 \end{cases}$$

$$f_Z(z) = \begin{cases} \displaystyle\int_0^\infty e^{-(z+y+y)}\,dy = e^{-z}\int_0^\infty e^{-2y}\,dy = \tfrac{1}{2}\,e^{-z}, & z > 0 \\ \displaystyle\int_{-z}^\infty e^{-(z+y+y)}\,dy = e^{-z}\int_{-z}^\infty e^{-2y}\,dy = \tfrac{1}{2}\,e^{z}, & z < 0 \end{cases}$$

or

$$f_Z(z) = \frac{1}{2}\,e^{-|z|}, \qquad -\infty \le z \le \infty.$$

(c)

$$Z = XY.$$

$$\begin{aligned} F_Z(z) &= P\{Z \le z\} = P\{XY \le z\} \\ &= \int_0^\infty \int_0^{z/y} f_{XY}(x,y)dx\,dy \end{aligned}$$

or (see Eq. (6-148))

$$f_Z(z) = \int_0^\infty \frac{1}{y} f_{XY}(\frac{z}{y}, y)dy = \int_0^\infty \frac{1}{y} e^{-((z/y)+y)}dy$$

(d)

$$Z = X/Y$$

$$\begin{aligned} F_Z(z) &= P\{Z \le z\} = P\{\tfrac{X}{Y} \le z\} \\ &= \int_0^\infty \int_0^{yz} f_{XY}(x,y)dx\,dy \end{aligned}$$

(use Eq. (6-60))

$$\begin{aligned} f_Z(z) &= \int_0^\infty y f_{XY}(yz,y)dy = \int_0^\infty y e^{y(z+1)}dy = \int_0^\infty y e^{(1+z)y} \\ &= \left[y\frac{e^{-(1+z)y}}{-(1+z)}\right]_0^\infty + \left(\frac{1}{1+z}\right)\int_0^\infty e^{(1+z)y}dy \\ &= \left(\frac{1}{1+z}\right)\left[\frac{e^{-(1+z)y}}{-(1+z)}\right]_0^\infty = \frac{1}{(1+z)^2}\,U(z) \end{aligned}$$

(e)

$$Z = \min\,(X,Y)$$

$$\begin{aligned} F_Z(z) &= P\{\min\,(X,Y) \le z\} \\ &= 1 - P\{Z > z, Y > z\} \\ &= 1 - [1 - F_X(z)]\,[1 - F_Y(z)] \\ &= F_X(z) + F_Y(z) - F_X(z)\,F_Y(z) \end{aligned}$$

(see Eq. (6-81))

$$f_Z(z) = f_X(z) + f_Y(z) - F_X(z)f_Y(z) - f_X(z)F_Y(z).$$

We have

$$f_X(z) = f_Y(z) = e^{-z}\,U(z)$$

so that

$$F_X(z) = \int_0^z e^{-x}dx = (1 - e^{-z})\,U(z) = F_Y(z)$$

$$\begin{aligned} f_Z(z) &= [e^{-z} + e^{-z} - 2(1 - e^{-z})\,e^{-z}]U(z) \\ &= 2e^{-z}\,[1 - 1 + e^{-z}]\,U(z) \\ &= 2e^{-2z}\,U(z) \sim \text{Exponential (2).} \end{aligned}$$

(f)

$$Z = \max(X,Y)$$

$$\begin{aligned} F_Z(z) &= P\{\max(X,Y) \le z\} = P\{X \le z, Y \le z\} \\ &= P\{X \le z\}\,P\{Y \le z\} = F_X(z)\,F_Y(z) \end{aligned}$$

$$\begin{aligned} f_Z(z) &= F_X(z)\,f_Y(z) + f_X(z)\,F_Y(z) \\ &= e^{-z}\,(1 - e^{-z}) + e^{-z}\,(1 - e^{-z}) \\ &= 2e^{-z}\,(1 - e^{-z})\,U(z) \end{aligned}$$

(g)

$$Z = \frac{\min(X,Y)}{\max(X,Y)}, \quad 0 < z < 1$$

$$\begin{aligned}
F_Z(z) &= P\left\{\left(\frac{\min(X,Y)}{\max(X,Y)} \le z\right) \bigcap ((X \le Y) \bigcup (X > Y))\right\} \\
&= P\left\{\left(\frac{\min(X,Y)}{\max(X,Y)} \le z\right) \bigcap (X \le Y)\right\} + P\left\{\left(\frac{\min(X,Y)}{\max(X,Y)} \le z\right) \bigcap (X > Y)\right\} \\
&= P\{\tfrac{X}{Y} \le z, X \le Y\} + P\{\tfrac{Y}{X} \le z, X > Y\} \\
&= P\{X \le Yz, X \le Y\} + P\{Y \le Xz, X > Y\} \\
&= \int_0^\infty \int_0^{yz} f_{XY}(x,y)\, dx\, dy + \int_0^\infty \int_0^{xz} f_{XY}(x,y)\, dy\, dx
\end{aligned}$$

$$\begin{aligned}
f_Z(z) &= \int_0^\infty y\, f_{XY}(yz, y) dy + \int_0^\infty x f_{XY}(x, xz) dx \\
&= \int_0^\infty y\, f_{XY}(yz, y) dy + \int_0^\infty y\, f_{XY}(y, yz) dy \\
&= \int_0^\infty y \left(e^{-(yz+y)} + e^{-(y+yz)}\right) dy \\
&= 2\int_0^\infty y e^{-y(1+z)} dz = \begin{cases} \dfrac{2}{(1+z)^2}, & 0 \le z \le 1 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

6.2

$$f_{XY}(x,y) = f_X(x)\, f_Y(y) = \frac{1}{a^2}, \qquad 0 < x \le a, \quad 0 < y \le a$$

(a)

$$F_Z(z) = P\left\{\frac{X}{Y} \le z\right\} = P\{X \le zY\}$$

(i) $z < 1$

$$\begin{aligned}
F_Z(z) &= P\{X \le zY\} \\
&= \int_0^a \int_0^{zy} \tfrac{1}{a} \cdot \tfrac{1}{a}\, dx\, dy = \tfrac{z}{2}, \quad z \le 1
\end{aligned}$$

(ii) $z \ge 1$

$$\begin{aligned}
F_Z(z) &= P\{X \le zY\} \\
&= 1 - \int_0^a \int_0^{x/z} \tfrac{1}{a} \cdot \tfrac{1}{a}\, dy\, dx \\
&= 1 - \int_0^1 \tfrac{x}{z}\, dx = 1 - \tfrac{1}{2z} \quad z > 1
\end{aligned}$$

$$f_Z(z) = \begin{cases} \tfrac{1}{2}, & z \le 1 \\ \dfrac{1}{2z^2}, & z > 1 \end{cases}$$

(b)

$$\begin{aligned} F_Z(z) &= P(Z \le z) = P\left\{\frac{Y}{X+Y} \le z\right\} \\ &= P\left\{\frac{X}{Y} \ge \frac{1}{z} - 1\right\} = 1 - P\left(\frac{X}{Y} \le \frac{1-z}{z}\right) \\ &= \begin{cases} \frac{1}{2}\left(\frac{z}{1-z}\right), & 0 < z \le 1/2 \\ 1 - \frac{1}{2}\left(\frac{1-z}{z}\right), & 1/2 < z < 1 \end{cases} \end{aligned}$$

$$f_Z(z) = \begin{cases} \dfrac{1}{2(1-z)^2}, & 0 < z \le 1/2 \\ \dfrac{1}{2z^2}, & 1/2 < z < 1 \end{cases}$$

(c)

$$\begin{aligned} F_Z(z) &= P\{Z \le z\} = P\{|X-Y| \le z\} \\ &= P\{\{|X-Y| \le z\} \textstyle\bigcap (X \ge Y)\} + P\{\{|X-Y| \le z\} \textstyle\bigcap (X < Y)\} \\ &= P\{X - Y \le z, X \ge Y\} + P\{Y - X \le z, X < Y\} \\ &= \int_0^\infty \int_y^{y+z} f_{XY}(x,y)\,dx\,dy + \int_0^\infty \int_x^{x+z} f_{XY}(x,y)\,dy\,dx \\ &= \int_0^\infty \int_y^{y+z} f_{XY}(x,y)\,dx\,dy + \int_0^\infty \int_y^{y+z} f_{XY}(y,x)\,dx\,dy \\ &= \int_0^\infty \int_y^{y+z} \{f_{XY}(x,y) + f_{XY}(y,x)\}\,dx\,dy. \end{aligned}$$

In general

$$\begin{aligned} f_Z(z) &= \int_0^\infty \frac{d}{dz} \int_y^{y+z} f_{XY}(x,y) + f_{XY}(y,x)\,dx\,dy \\ &= \int_0^\infty \{f_{XY}(y+z,y) + f_{XY}(y,y+z)\}\,dy. \end{aligned}$$

Here

$$X \sim U(0,a), \qquad Y \sim U(0,a)$$

$$F_Z(z) = 1 - \frac{1}{a^2} \cdot 2 \cdot \frac{(a-z)^2}{2} = 1 - \left(1 - \frac{z}{a}\right)^2$$

and

$$f_Z(z) = \frac{2}{a}\left(1 - \frac{z}{a}\right) \qquad 0 \le z \le a.$$

6.3
$$\begin{aligned} F_Z(z) &= P\{Z \le z\} = P\{X+Y \le z\} \\ &= \tfrac{1}{2} - \tfrac{z^2}{2}, \quad -1 < z < 0, \end{aligned}$$

(which represents the area below the line $X + Y = z$.)

$$\begin{aligned} F_Z(z) &= P\{Z \le z\} = P\{X+Y \le z\} \\ &= \tfrac{1}{2} + \tfrac{z^2}{2}, \quad 0 \le z < 1 \end{aligned}$$

$$f_Z(z) = \begin{cases} -z, & -1 \le z < 0 \\ z, & 0 \le z < 1 \end{cases}$$

6.4
$$Z = X - Y$$

For $z < 0$

$$\begin{aligned} F_Z(z) &= P\{Z \le z\} \\ &= \int_0^{(1+z)/2} \int_{x-z}^{1-x} f_{XY}(x,y)\,dy\,dx = \int_0^{(1+z)/2} \int_{x-z}^{1-x} 6x\,dy\,dx \\ &= \int_0^{(1+z)/2} 6x\,[y]_{x-z}^{1-x}\,dx = \int_0^{(1+z)/2} 6x(1-x-x+z)\,dx \\ &= 6\left[(1+z)\frac{x^2}{2} - \frac{2x^3}{3}\right]_0^{(1+z)/2} = 6\left[\frac{(1+z)^3}{8} - \frac{(1+z)^3}{12}\right] \\ &= \frac{(1+z)^3}{4}, \quad z \le 0. \end{aligned}$$

For $z > 0$

$$\begin{aligned} F_Z(z) &= P\{Z \le z\} = 1 - P\{Z > z\} \\ &= 1 - \int_0^{(1-z)/2} \int_{z+y}^{1-y} f_{XY}(x,y)\,dx\,dy = 1 - \int_0^{(1-z)/2} \int_{z+y}^{1-y} 6x\,dy \\ &= 1 - \int_0^{(1-z)/2} \left[\frac{6x^2}{2}\right]_{z+y}^{1-y} dy = 1 - 3\int_0^{(1-z)/2} \left[(1-y)^2 - (z-y)^2\right] dy \\ &= 1 - 3\,(1+z)\left[\frac{(1-z)^2}{2} - \frac{(1-z)^2}{4}\right] = 1 - \tfrac{3}{4}\,(1+z)(1-z)^2 \quad z \le 0. \end{aligned}$$

$$f_Z(z) = \begin{cases} \tfrac{3}{4}\,(1-z)(1+3z), & 0 \le z \le 1 \\ \tfrac{3}{4}\,(1+z)^2, & -1 < z < 0 \end{cases}$$

6.5 (a) See Example 6-15 for solutions
(b) See Example 6-14 for solutions
(c)
$$U = X - Y \sim N(0, 2\sigma^2)$$

since linear combinations of jointly Gaussian random variables are Gaussian random variables (see Eq. (6-120) Text.). Here $Var(U) = Var(X) + Var(Y) = 2\sigma^2$.

6.6

$$Z = XY$$

$$\begin{aligned} F_Z(z) &= P(XY \le z) = 1 - P(XY > z) \\ &= 1 - \int_z^1 \int_{z/y}^1 f_{XY}(x,y)\,dx\,dy \\ f_Z(z) &= 1 + \int_z^1 \frac{1}{y} f_{XY}(z/y, y)\,dy = 1 + \int_z^1 \left\{ \frac{2}{y} - \frac{2z}{y^2} \right\} dy \\ &= 1 - 2\ln z + 2z, \quad 0 \le z \le 1 \end{aligned}$$

6.7 (a)

$$Z_1 = X + Y$$

$$F_{Z_1}(z) = P(X+Y \le z) = \begin{cases} \displaystyle\int_0^z \int_0^{z-y} f_{XY}(x,y)\,dx\,dy, & 0 < z < 1 \\ \displaystyle 1 - \int_{z-1}^1 \int_{z-y}^1 f_{XY}(x,y)\,dx\,dy, & 1 < z < 2 \end{cases}$$

$$f_{Z_1}(z) = \begin{cases} \displaystyle\int_0^z f_{XY}(z-y, y)\,dy, & 0 < z < 1 \\ \displaystyle\int_{z-1}^1 f_{XY}(z-y, y)\,dy, & 1 < z < 2 \end{cases}$$

$$= \begin{cases} z^2, & 0 < z < 1 \\ z(2-z), & 1 < z < 2 \\ 0, & \text{otherwise} \end{cases}$$

(b)

$$Z_2 = XY$$

$$F_{Z_2}(z) = P(XY \le z) = 1 - \int_z^1 \int_{z/y}^1 f_{XY}(x,y)\,dx\,dy$$

$$\begin{aligned} f_{Z_2}(z) &= \int_z^1 \frac{1}{y} f_{XY}(z/y, y)\,dy = \int_z^1 \frac{1}{y} \left(\frac{z}{y} + y \right) dy \\ &= 2(1-z), \quad 0 < z < 1 \end{aligned}$$

(c)

$$Z_3 = \frac{Y}{X}$$

$$F_{Z_3}(z) = P(Y/X \le z) = \begin{cases} \displaystyle\int_0^1 \int_0^{zx} f_{XY}(x,y)\,dy\,dx, & 0 < z < 1 \\ \displaystyle 1 - \int_0^1 \int_0^{y/z} f_{XY}(x,y)\,dx\,dy, & z > 1 \end{cases}$$

$$f_{Z_3}(z) = \begin{cases} \displaystyle\int_0^1 x\, f_{XY}(x, zx)\, dx, & 0 < z < 1 \\ \displaystyle\int_0^1 \frac{y}{z^2}\, f_{XY}(y/z, y)\, dy, & z > 1 \end{cases}$$

$$= \begin{cases} \dfrac{1+z}{3}, & 0 < z < 1 \\ \dfrac{1+z}{3z^3}, & z > 1 \end{cases}$$

(d)

$$Z_4 = Y - X$$

$$F_{Z_4}(z) = P(Y - X \le z) = \begin{cases} 1 - \displaystyle\int_z^1 \int_0^{y-z} f_{XY}(x,y)\, dx\, dy & 0 < z < 1 \\ \displaystyle\int_0^{z+1} \int_{y-z}^1 f_{XY}(x,y)\, dx\, dy, & -1 < z < 0 \end{cases}$$

$$f_{Z_4}(z) = \begin{cases} \displaystyle\int_z^1 f_{XY}(y-z, y)\, dy, & 0 < z < 1 \\ \displaystyle\int_0^{z+1} f_{XY}(y-z, y)\, dy, & -1 < z < 0 \end{cases}$$

$$= \begin{cases} 1 - z, & 0 < z < 1 \\ 1 + z, & -1 < z < 0 \end{cases} = 1 - |z|, \quad |z| < 1$$

6.8

$$F_Z(z) = P(X + Y \le z)$$

$$= \begin{cases} \displaystyle\int_0^{z/3} \int_{2y}^{z-y} f_{XY}(x,y)\, dx\, dy = \frac{z^2}{6}, & 0 < z < 2 \\ 1 - \displaystyle\int_{2z/3}^{2} \int_{z-x}^{x/2} f_{XY}(x,y)\, dy\, dx = 2z - \frac{z^2}{3} - 2, & 2 < z < 3 \end{cases}$$

Thus

$$f_Z(z) = \begin{cases} \displaystyle\int_0^{z/3} f_{XY}(z-y, y)\, dy & 0 < z < 2 \\ \displaystyle\int_{2z/3}^{2} f_{XY}(x, z-x)\, dx & 2 < z < 3 \end{cases}$$

$$f_Z(z) = \begin{cases} \frac{1}{3}\, z, & 0 < z < 2 \\ 2 - \frac{2z}{3}, & 2 < z < 3 \\ 0, & \text{otherwise} \end{cases}$$

6.9 (a)

$$Z = \frac{X}{Y}, \qquad z \geq 1$$

$$F_Z(z) = P(X \leq Yz) = \int_0^1 \int_{x/z}^{x} f_{XY}(x,y)\, dy\, dx$$

$$f_Z(z) = \int_0^1 \frac{x}{z^2} f_{XY}(x, x/z)\, dx = \frac{1}{z^2}, \quad z \geq 1$$

(b)

$$W = XY$$

$$\begin{aligned} F_W(w) &= P(W \leq w) = P(XY \leq w) = 1 - P(XY > w) \\ &= 1 - \int_{\sqrt{w}}^1 \int_{w/x}^{x} f_{XY}(x,y)\, dy\, dx \end{aligned}$$

Hence

$$\begin{aligned} f_W(w) &= \int_{\sqrt{w}}^1 \frac{1}{x} f_{XY}(x, w/x)\, dx = \int_{\sqrt{w}}^1 \frac{2}{x}\, dx \\ &= \ln(1/w), \quad 0 < w \leq 1 \end{aligned}$$

6.10 (a)

$$Z = X + Y$$

$$F_Z(z) = \int_0^{z/2} \int_x^{2-x} f_{XY}(x,y)\, dx = \frac{z^2}{4}, \quad 0 < z < 2$$

$$f_Z(z) = \frac{z}{2}, \qquad 0 < z < 2$$

(b)

$$W = X - Y$$

$$F_W(w) = \frac{1}{2}(2+w)(1+\frac{w}{2}) = \left(1 + \frac{w}{2}\right)^2$$

$$f_W(w) = \begin{cases} 1 + \frac{w}{2}, & -2 < w < 0 \\ 0, & \text{otherwise} \end{cases}$$

6.11 (a) The charaterristic function of $X+Y$ is given by

$$\begin{aligned}\phi_{X+Y}(\omega) &= \phi_X(\omega)\,\phi_Y(\omega) = \tfrac{1}{(1-j\omega\beta)^\alpha}\cdot\tfrac{1}{(1-j\omega\beta)^\alpha} \\ &= \tfrac{1}{(1-j\omega\beta)^{2\alpha}} \sim \text{Gamma}\,(2\alpha,\beta)\end{aligned}$$

(b)

$$f_{XY}(x,y) = f_X(x)\,f_Y(y) = \frac{(xy)^{(\alpha-1)}}{(\Gamma(\alpha)\beta^\alpha)^2}\;e^{(x+y)/\beta}, \quad x>0,\ y>0$$

Let

$$Z = \frac{X}{Y}$$

Using (Eq. 6-60) we get

$$\begin{aligned}f_Z(z) &= \int_0^\infty y\,\frac{(y^2 z)^{(\alpha-1)}}{(\Gamma(\alpha)\beta^\alpha)^2}\;e^{-(1+z)y/\beta}\,dy \\ &= \frac{z^{(\alpha-1)}}{(\Gamma(\alpha)\beta^\alpha)^2}\int_0^\infty y^{(2\alpha-1)}\,e^{-(1+z)y/\beta}\,dy \\ &= \frac{z^{(\alpha-1)}}{(\Gamma(\alpha))^2\,\beta^{2\alpha}}\,\frac{\beta^{(2\alpha-1)}}{(1+z)^{2\alpha-1}}\,\frac{\beta}{(1+z)}\int_0^\infty u^{2\alpha-1}\,e^{-u}\,du \\ &= \frac{(\Gamma(2\alpha))\;z^{\alpha-1}}{(\Gamma(\alpha))^2\,(1+z)^{2\alpha}}, \qquad z>0\end{aligned}$$

(see also Example 6-27 for the answer).

(c)

$$W = \frac{X}{X+Y} = \frac{X/Y}{X/Y+1} = \frac{Z}{Z+1}$$

$$F_W(w) = P\left(\frac{Z}{Z+1}\le w\right) = P\left(Z \le \frac{w}{1-w}\right) = F_Z\left(\frac{w}{1-w}\right)$$

This gives

$$\begin{aligned}f_W(w) &= \frac{1}{(1-w)^2}\,f_Z\left(\tfrac{w}{1-w}\right) \\ &= \frac{\Gamma(2\alpha)}{(\Gamma(\alpha))^2}\,w^{\alpha-1}\,(1-w)^{\alpha-1} \\ &\sim \ \mathit{Beta}\,(\alpha,\alpha)\end{aligned}$$

where we have used results from (b) above.

6.12

$$X\sim U(0,1), \quad Y\sim U(0,1), \quad X,Y \text{ are independent, and}$$

$$U = X+Y, \qquad V = X-Y \ \Rightarrow\ |v|<u<2.$$

U and V have one pair of solutions given by

$$x_1 = \frac{u+v}{2},\, y_1 = \frac{u-v}{2}.$$

Also the Jacobian is given by

$$J = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

so that

$$f_{UV}(u,v) = \frac{1}{|J|}\,f_{XY}(x_1,y_1) = \frac{1}{2}, \quad 0<|v|<u<2$$

6.13

$$f_{XY}(x,y) = \frac{xy}{\sigma^4}\, e^{-(x^2+y^2)/2\sigma^2}, \quad x, y \geq 0$$

$$Z = \frac{X}{Y}$$

$$F_Z(z) = P(Z \leq z) = P(X/Y \leq z) = \int_0^\infty \int_0^{zy} f_{XY}(x,y)\, dx\, dy.$$

This gives the density function of z to be

$$\begin{aligned} f_Z(z) &= \int_0^\infty y\, f_{XY}(zy, y)\, dy = \int_0^\infty \frac{zy^3}{\sigma^4}\, e^{-(z^2y^2+y^2)/2\sigma^2}\, dy \\ &= \frac{z}{\sigma^4}\int_0^\infty y^3\, e^{-y^2(z^2+1)/2\sigma^2}\, dy \quad \text{Let, } t = y^2(z^2+1)/2\sigma^2 \\ &= \frac{2z}{(z^2+1)^2}\int_0^\infty t\, e^{-t}\, dt = \frac{2z}{(z^2+1)^2}, \quad 0 \leq z \leq \infty. \end{aligned}$$

6-14 $\quad \underset{\sim}{z} = \underset{\sim}{x} + \underset{\sim}{y} \qquad f_z(z) = f_x(z) * f_y(z)$

For $z > 0$

$$c^2 z\, e^{-cz} = \int_0^z c\, e^{-c(z-y)}\, f_y(y)\, dy$$

$$c\, z = \int_0^z e^{cy}\, f_y(y)\, dy \qquad c = e^{cz}\, f_y(z)$$

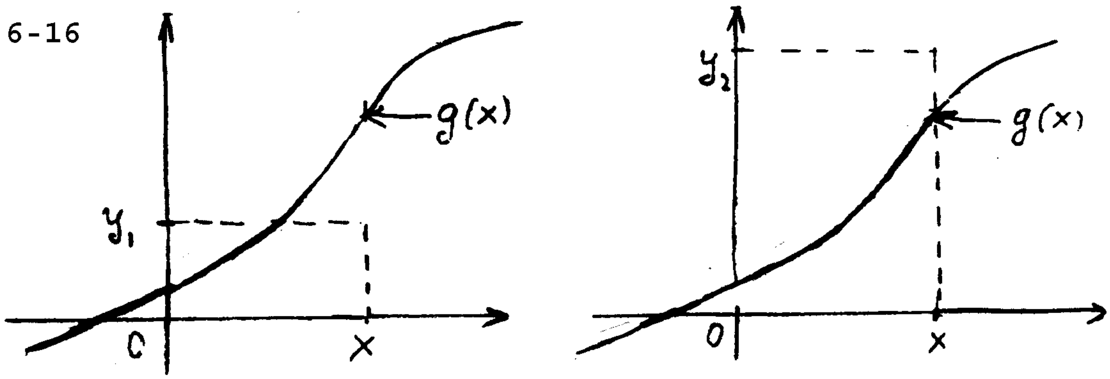
(differentiation). Hence, $f_y(z) = c\, e^{-cz}$; and zero for $z < 0$.

6-15

$$f_z(z) = \int_{-\infty}^{\infty} f_x(x) f_y(z-x)\, dx = \int_{z-1}^{z} f_x(x)\, dx = F_x(z) - F_x(z-1)$$

because $f_y(z-x) = 1$ for $z - 1 < x < z$ and zero otherwise.

6-16



All probability masses are on the line $y = g(x)$.

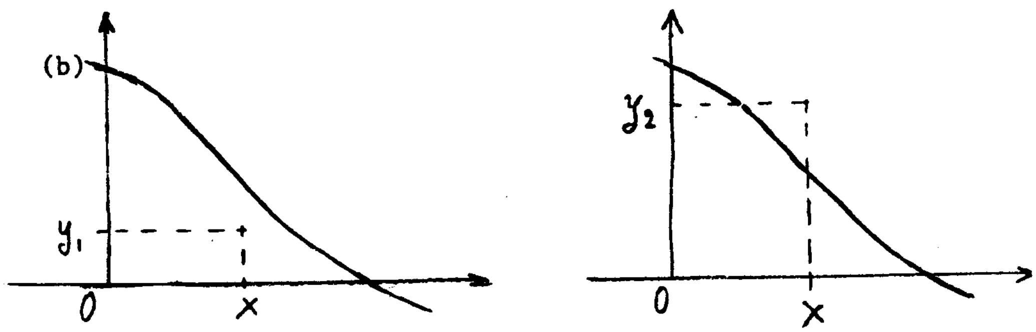
(a) If $y = y_1 < g(x)$ then

$$F(x,y) = P\{\underset{\sim}{x} \le x,\ \underset{\sim}{y} \le y_1\} = P\{\underset{\sim}{y} \le y_1\} = F_y(y_1).$$

If $y = y_2 > g(x)$ then

$$F(x,y) = P\{\underset{\sim}{x} \le x,\ \underset{\sim}{y} \le y_2\} = P\{\underset{\sim}{x} \le x\} = F_x(x)$$

(b)



If $y = y_1 < g(x)$ then

$$F(x,y) = P\{\underset{\sim}{x} \le x,\ \underset{\sim}{y} \le y_1\} = 0$$

If $y = y_2 > g(x)$ then

$$F(x,y) = P\{\underset{\sim}{x} \le x,\ \underset{\sim}{y} \le y_2\} = P\{\underset{\sim}{x} \le x\} - P\{\underset{\sim}{y} > y_2\}$$

$$= F_x(x) - [1 - F_y(y_2)]$$

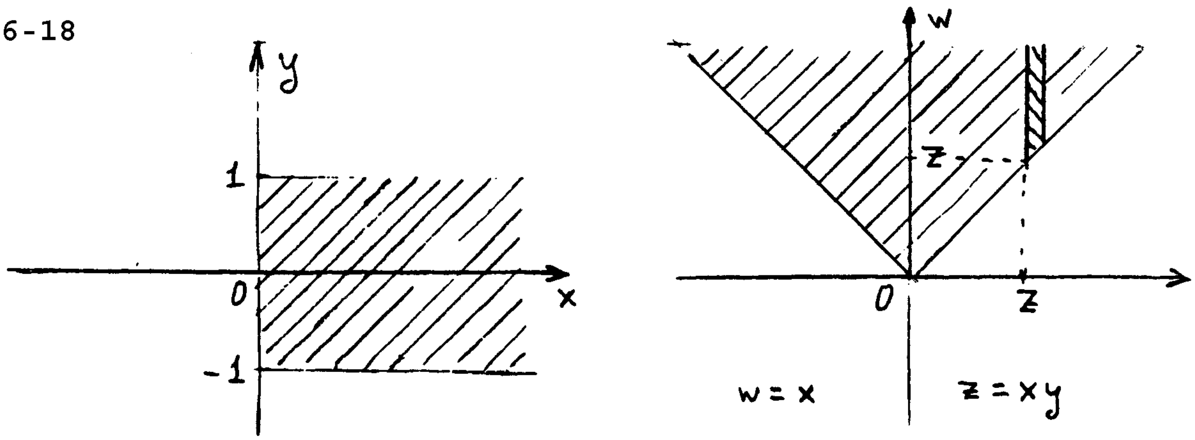
6-17 (a) If $\underset{\sim}{z} = 2\underset{\sim}{x} + 3\underset{\sim}{y}$ then $E\{\underset{\sim}{z}\} = 0$ $\quad \sigma_z^2 = 4\sigma_x^2 + 9\sigma_y^2 = 5^2$

Hence, $\underset{\sim}{z}$ is $N(0;\sqrt{52})$

(b) If $\underset{\sim}{z} = \underset{\sim}{x}/\underset{\sim}{y}$, then from (6-63) with $\sigma_1 = \sigma_2 = 2$, $r = 0$

$$F_z(z) = \frac{1}{2} + \frac{1}{\pi}\,\text{arc tan } z \qquad f_z(z) = \frac{1}{\pi(1+z^2)}$$

6-18



$$f_{zw}(z,w) = \frac{1}{|x|}\, f_{xy}(x,y) \qquad x = w \qquad y = z/w$$

The function $f_{zw}(z,w)$ is different from zero in the shaded areas shown. Hence, with $w^2 - z^2 = s^2$

$$f_z(z) = \frac{1}{\pi\alpha^2}\int_{|z|}^{\infty} e^{-w^2/2\alpha^2}\,\frac{dw}{\sqrt{1 - z^2/w^2}}$$

$$= \frac{1}{\pi\alpha^2}\int_0^{\infty} e^{-(z^2+s^2)/2\alpha^2}\,ds = \frac{1}{\alpha\sqrt{2\pi}}\, e^{-z^2/2\alpha^2}$$

6-19 (a) $\quad z = x/y \qquad w = y \qquad J = 1/y$

$$f_z(z) = \int_{-\infty}^{\infty} |w| f_x(zw) f_y(w)\,dw \qquad z > 0$$

$$= \frac{z}{\alpha^2\beta^2} \int_0^{\infty} w^3\, e^{-cw^2} dw = \frac{z}{2\alpha^2\beta^2 c^2} \qquad c = \frac{z^2}{2\alpha^2} + \frac{1}{2\beta^2}$$

$$= \frac{2\alpha^2}{\beta^2} \frac{z}{(z^2 + \alpha^2/\beta^2)^2} \qquad \text{for } z > 0 \text{ and zero otherwise}$$

(b) $$F_z(z) = \int_0^z \frac{2\alpha^2 z\,dz}{\beta^2(z^2 + \alpha^2/\beta^2)^2} = \frac{\alpha^2}{\beta^2} \int_{\alpha^2/\beta^2}^{z^2+\alpha^2/\beta^2} \frac{dt}{t^2}$$

$$= \frac{z^2}{z^2 + \alpha^2/\beta^2} = P\{z \le z\} = P\{x \le zy\}$$

6-20 1. The density of $2\underset{\sim}{x}$ equals $\frac{1}{2} f_x(\frac{x}{2})$. Hence, if $\underset{\sim}{z} = 2\underset{\sim}{x} + \underset{\sim}{y}$, then

$$f_z(z) = \int_0^z \frac{\alpha}{2} e^{-\alpha x/2} \beta e^{-\beta(z-x)} dx = \frac{\alpha\beta}{\alpha - 2\beta}(e^{-\beta z} - e^{-\alpha z/2})U(z)$$

2. The density of $\underset{\sim}{y}$ equals $f_y(-y)$. Hence, if $\underset{\sim}{z} = \underset{\sim}{x} - \underset{\sim}{y}$, then

$$f_z(z) = f_x(z) * f_y(-z)$$

$$= \alpha\beta \begin{cases} \displaystyle\int_z^\infty e^{-\alpha x} e^{-\beta(x-z)} dx = \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha z} & z > 0 \\ \displaystyle\int_0^\infty e^{-\alpha x} e^{-\beta(x-z)} dx = \frac{\alpha\beta}{\alpha+\beta} e^{\beta z} & z < 0 \end{cases}$$

3. $\underset{\sim}{z} = \underset{\sim}{x}/\underset{\sim}{y} \qquad \underset{\sim}{w} = \underset{\sim}{y} \qquad J = 1/y$

$$f_z(z) = \alpha\beta \int_0^\infty w e^{-\alpha z w} e^{-\beta w} dw = \frac{\alpha\beta}{(\alpha z + \beta)^2} U(z)$$

4. $\underset{\sim}{z} = \max(\underset{\sim}{x},\underset{\sim}{y}) \qquad F_z(z) = F_{xy}(z,z) = F_x(z)F_y(z)$

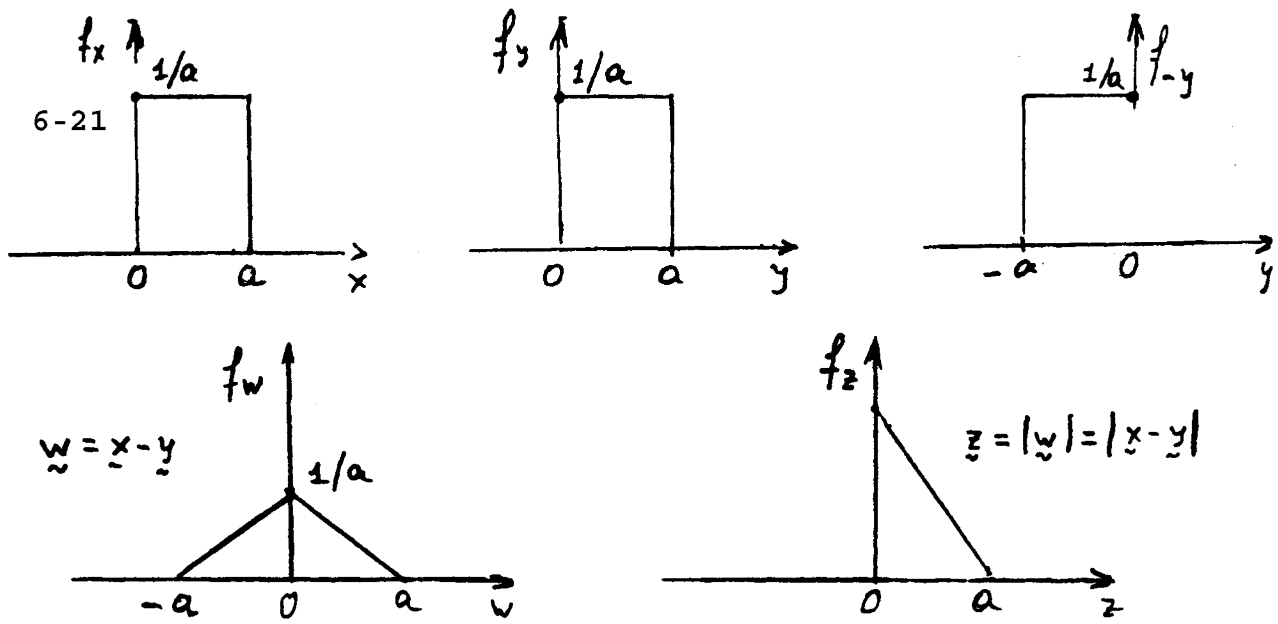
$$f_z(z) = f_x(z)F_y(z) + f_y(z)F_x(z)$$

$$= \left[\alpha e^{-\alpha z}(1 - e^{-\beta z}) + \beta e^{-\beta z}(1 - e^{-\alpha z})\right] U(z)$$

5. $\underset{\sim}{z} = \min(\underset{\sim}{x},\underset{\sim}{y}) \qquad F_z(z) = F_x(z) + F_y(z) - F_x(z)F_y(z)$

$$f_z(z) = f_x(z)[1 - F_y(z)] + f_y(z)[1 - F_x(z)] = (\alpha+\beta)e^{-(\alpha+\beta)z}U(z)$$

6-21



6-22 (a)

$$\alpha y^2 + \beta(x-y)^2 = (\alpha+\beta)\left(y - \frac{\beta x}{\alpha+\beta}\right)^2 + \frac{\alpha\beta}{\alpha+\beta}x^2$$

$$e^{-\alpha x^2} * e^{-\beta x^2} = \int_{-\infty}^{\infty} e^{-\alpha y^2 - \beta(x-y)^2} dy$$

$$= e^{-\alpha\beta x^2/(\alpha+\beta)} \int_{-\infty}^{\infty} e^{-(\alpha+\beta)\left(y - \frac{\beta x}{\alpha+\beta}\right)^2} dy = \sqrt{\frac{\pi}{\alpha+\beta}}\, e^{\frac{-\alpha\beta x^2}{\alpha+\beta}}$$

(b) $$\frac{\alpha/\pi}{x^2+\alpha^2} * \frac{\beta/\pi}{x^2+\beta^2} = \frac{\alpha\beta}{\pi^2}\int_{-\infty}^{\infty} \frac{dy}{(y^2+\alpha^2)[(x-y)^2+\beta^2]} = \frac{(\alpha+\beta)/\pi}{x^2+(\alpha+\beta)^2}$$

Characteristic functions lead to a simpler derivation of the above [see (6-192)]

6-23 We introduce the auxiliary variable w=y. The Jacobian of the transformation z=nx/my, w=y equals n/my. Since x=mzw/n, y=w and the RVs $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are independent, (6-113) yields

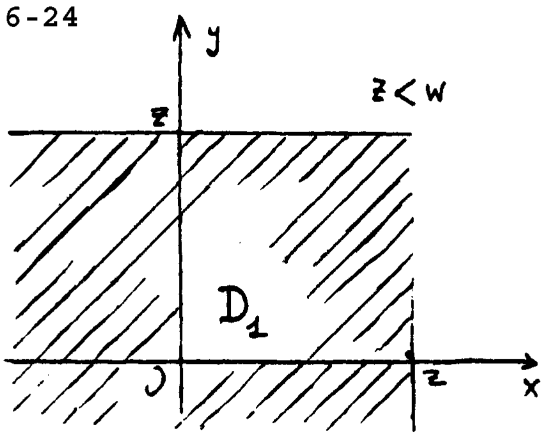
$$f_{zw}(z,w) = \frac{mw}{n} f_x\left(\frac{m}{n}zw\right) f_y(w) \sim w(zw)^{m/2-1}e^{-mzw/2n}w^{n/2-1}e^{-w/2}$$

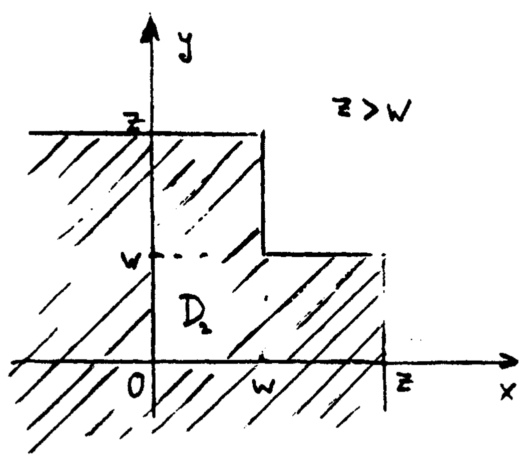
for z>0, w>0 and 0 othrwise. Integrating with respect to w, we obtain

$$f_z(z) \sim z^{m/2-1}\int_0^\infty w^{(m+n)/2-1}\exp\left\{-\frac{w}{2}\left(1+\frac{m}{n}z\right)\right\}dw$$

$$\sim \frac{z^{m/2-1}}{(1+mz/n)^{(m+n/2)}}\int_0^\infty q^{(m+n)/2}e^{-q}dq$$

6-24





If $z \le w$ then

$$P\{\underset{\sim}{z} \le z,\ \underset{\sim}{w} \le w\} = P\{\underset{\sim}{z} \le z\} = P\{(\underset{\sim}{x},\underset{\sim}{y}) \in D_1\} = F_{xy}(z,z)$$

If $z > w$ then

$$P\{\underset{\sim}{z} \le z,\ \underset{\sim}{w} \le w\} = P\{(\underset{\sim}{x},\underset{\sim}{y}) \in D_2\}$$

$$= F_{xy}(z,w) + F_{xy}(w,z) - F_{xy}(w,w)$$

6.25

$$X \sim \text{Exponential}(\lambda), \quad Y \sim \text{Exponential}(\lambda)$$

X and Y are independent so that

$$f_{XY}(x,y) = f_X(x)\, f_Y(y) = \frac{1}{\lambda^2}\, e^{-(x+y)/\lambda}\, U(x)\, U(y)$$

$$Z = X + Y$$

$$\phi_Z(\omega) = \phi_X(\omega)\, \phi_Y(\omega) \frac{1}{(1 - j\omega\lambda)^2}$$

$$Z \sim \text{ Gamma } (2, \lambda)$$

This gives

$$f_Z(z) = \frac{z}{\lambda^2}\, e^{-z/\lambda} U(z)$$

$$P(Z > 2\lambda) = \int_{2\lambda}^{\infty} \frac{z}{\lambda^2}\, e^{-z/\lambda}\, dz = \int_2^{\infty} x e^{-x}\, dx = 3e^{-2} = 0.406$$

Let,

$$W = Y - X$$

Then

$$P(Y - X > \lambda) = P(W > \lambda) = \int_{\lambda}^{\infty} f_W(w) dw$$

Notice that $F_W(w)$ is given by (6-55).
For $w > 0$, this gives

$$\begin{aligned} f_W(w) &= \int_0^{\infty} \frac{1}{\lambda^2}\, e^{-(w+2y)/\lambda}\, dy = \frac{1}{\lambda^2}\, e^{-w/\lambda} \int_0^{\infty} e^{-2y/\lambda}\, dy \\ &= \frac{1}{2\lambda}\, e^{-w/\lambda}, \quad w > 0 \end{aligned}$$

Hence

$$P(Y - X > \lambda) = P(W > \lambda) = \int_{\lambda}^{\infty} \frac{1}{2\lambda}\, e^{-w/\lambda}\, dw = \frac{1}{2e}$$

6.26 (a)

$$\begin{aligned} R &= W - Z \\ &= \max(\mathrm{X},\mathrm{Y}) \; - \; \min(\mathrm{X},\mathrm{Y}) \\ &= \begin{cases} X - Y, & X \geq Y \\ Y - X, & X < Y \end{cases} \end{aligned}$$

$$\begin{aligned} F_R(r) &= P\{R \leq r\} \\ &= P\{R \leq r, X \geq Y\} + P\{R \leq r, X < Y\} \\ &= P\{X - Y \leq r, X \geq Y\} + P\{Y - X \leq r, X < Y\} \\ &= 1 - 2\frac{(1-r)^2}{2} = 1 - (1-r)^2, \quad 0 \leq r \leq 1 \end{aligned}$$

$$f_R(r) = \begin{cases} 2(1-r), & 0 \leq r \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

(b)

$$\begin{aligned} S &= W + Z \\ &= \max(\mathrm{X},\mathrm{Y}) \; + \; \min(\mathrm{X},\mathrm{Y}) = X + Y \end{aligned}$$

Case 1: $0 < s < 1$

$$F_S(s) = P\{S \leq s\} = P\{X + Y \leq s\} = \frac{s^2}{2}, \quad 0 < s < 1$$

Case 2: $1 \leq s \leq 2$

$$F_S(s) = P\{S \leq s\} = P\{X + Y \leq s\} = 1 - \frac{(2-s)^2}{2}, \quad 1 \leq s \leq 2$$

$$F_S(s) = \begin{cases} s, & 0 \leq s \leq 1 \\ (2-s), & 1 \leq s \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

6.27 (a) X, Y are independent, identically distributed exponential random variables.

$$Z = \frac{Y}{\max(X,Y)} = \begin{cases} \frac{Y}{X}, & X \geq Y \\ 1, & X < Y \end{cases} \quad \Rightarrow \; 0 < z \leq 1.$$

$0 < z < 1$

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P\left\{\tfrac{Y}{X} \leq z, \, X > Y\right\} \\ &= P\{Y \leq Xz, \, X > Y\} = \int_0^\infty \int_0^{xz} f_{XY}(x,y)\, dy\, dx \end{aligned}$$

$$f_Z(z) = \int_0^\infty x\, f_{XY}(x, xz)\, dx = \int_0^\infty \frac{x}{\lambda^2}\, e^{-(1+z)x/\lambda}\, dx = \frac{1}{(1+z)^2}, \quad 0 < z < 1.$$

Also

$$P(Z = 1) = P(X < Y) = \int_0^\infty \int_0^y \frac{1}{\lambda^2}\, e^{-(x+y)/\lambda}\, dx\, dy = \frac{1}{2}$$

(b)

$$W = \frac{X}{\min(X, 2Y)} = \begin{cases} \frac{X}{2Y}, & X \geq 2Y \\ 1, & X < 2Y \end{cases} \quad \Rightarrow \; 1 \leq w < \infty$$

$$F_W(w) = P(X \leq 2Yw, \, X > 2Y) = \int_0^\infty \int_{2y}^{2wy} f_{XY}(x,y)\, dx\, dy$$

This gives

$$\begin{aligned} f_W(w) &= \int_0^\infty 2y\, f_{XY}(2wy, y) dy = \int_0^\infty \frac{2y}{\lambda^2}\, e^{-(1+2w)y/\lambda} dy \\ &= \frac{2}{(1+2w)^2}, \quad w > 1 \end{aligned}$$

Also

$$P(W = 1) = P(X < 2Y) = \int_0^\infty \int_0^{2y} \frac{1}{\lambda^2}\, e^{-(x+y)/\lambda}\, dx\, dy = \frac{2}{3}$$

Note that the p.d.f. of Z as well as W has an impulse at $z = 1$ and $w = 1$ respectively.

6.28 X, Y are independent identically distributed exponential random variables.

$$Z = \frac{X}{X+Y}$$

$$\begin{aligned} F_Z(z) &= P\left(\frac{X}{X+Y} \le z\right) = P\left(\frac{X}{Y} \le \frac{z}{1-z}\right) \\ &= P\left\{X \le \frac{zY}{1-z}\right\} = \int_0^\infty \int_0^{(zy)/(1-z)} f_{XY}(x,y)\,dx\,dy \end{aligned}$$

$$\begin{aligned} f_Z(z) &= \int_0^\infty \frac{y}{(1-z)^2} f_{XY}(zy/(1-z), y)\,dy \\ &= \frac{1}{(1-z)^2}\int_0^\infty y\,\frac{1}{\lambda^2}\,e^{-(z/(1-z)+1)(y/\lambda)}\,dy \\ &= \frac{1}{(1-z)^2}\int_0^\infty \frac{y}{\lambda^2}\,e^{-[y/(1-z)\lambda]}\,dy \\ &= \int_0^\infty u\,e^{-u}\,du = 1, \quad 0 < z < 1 \\ &\Rightarrow \frac{X}{X+Y} \sim U(0,1) \end{aligned}$$

6.29 Let

$$f_X(x) = \frac{1}{\lambda}\,e^{-x/\lambda}\,U(x), \quad f_Y(y) = \frac{1}{\lambda}\,e^{-y/\lambda}\,U(y).$$

$$Z = \min(X, Y)$$

$$W = \max(X,Y) - \min(X,Y)$$

$$Z = \begin{cases} Y, & X \ge Y \\ X, & X < Y \end{cases}$$

$$W = \begin{cases} X - Y, & X \ge Y \\ Y - X, & X < Y \end{cases}$$

$Z = \min(X, Y)$. See Example 6-18, Eq.(6-82) for solution. From there (replace λ by $1/\lambda$ in (6-82))

$$f_Z(z) = \frac{2}{\lambda}\,e^{-2z/\lambda}\,U(z).$$

$$\begin{aligned} F_W(w) &= P(X - Y \le w, X \ge Y) + P(Y - X \le w, X < Y) \\ &= \int_0^\infty \int_y^{y+w} f_{XY}(x,y)\,dx\,dy \\ &+ \int_0^\infty \int_x^{x+w} f_{XY}(x,y)\,dy\,dx, \quad w > 0 \end{aligned}$$

This gives

$$\begin{aligned} F_W(w) &= \int_0^\infty f_{XY}(y+w, y)\,dy + \int_0^\infty f_{XY}(x, x+w)\,dx \\ &= 2\int_0^\infty \frac{1}{\lambda^2}\,e^{(2y+w)/\lambda}\,dy \\ &= \frac{2}{\lambda^2}\,e^{-w/\lambda}\,\left.\frac{e^{-2y/\lambda}}{-2/\lambda}\right|_0^\infty = \frac{1}{\lambda}\,e^{-w/\lambda}, \quad w > 0 \end{aligned}$$

Also

$$\begin{aligned}
F_{ZW}(z,w) &= P\{Z \le z, W \le w\} \\
&= P\{Y \le z, X - Y \le w, X \ge Y\} \\
&+ P\{X \le z, Y - X \le w, X < Y\} \\
&= \int_0^z \int_y^{y+w} f_{XY}(x,y)\, dx\, dy + \int_0^z \int_x^{x+w} f_{XY}(x,y)\, dy\, dx
\end{aligned}$$

Repeated use of (6-39)-(6-40) gives

$$\begin{aligned}
f_{ZW}(z,w) &= f_{XY}(z+w, z) + f_{XY}(z, z+w) \\
&= \frac{2}{\lambda^2}\, e^{-(2z+w)/\lambda} = \frac{2}{\lambda}\, e^{-2z/\lambda} \frac{1}{\lambda}\, e^{-w/\lambda} \\
&= f_Z(z)\, f_W(w)
\end{aligned}$$

Thus Z and W are independent exponential random variables.

6.30 (a) Let

$$U = X + Y, \qquad 0 < u < 2\beta.$$

The probability density function of U can be computed as in (6-48)-(6-50). Using Fig. 6-11, for $0 < u \le \beta$, we have

$$F_U(u) = \int_0^u \int_0^{u-x} f_{XY}(x,y)\, dy\, dx$$

which gives

$$\begin{aligned}
f_U(u) &= \int_0^u f_{XY}(x, u-x) dx = \alpha^2 \beta^{-2\alpha} \int_0^u x^{\alpha-1} (u-x)^{\alpha-1}\, dx \\
&= \alpha^2 \beta^{-2\alpha} u^{2\alpha-1} \int_0^1 y^{\alpha-1} (1-y)^{\alpha-1}\, dy \\
&= B(\alpha,\alpha)\, \alpha^2 \beta^{-2\alpha} u^{2\alpha-1} \quad 0 < u \le \beta
\end{aligned}$$

where we have substituted $y = ux$ and made use of the beta function defied in (4-49)-(4-51). Similarly for $\beta < u \le 2\beta$, we get (see (6-49))

$$F_U(u) = 1 - \int_{u-\beta}^{\beta} \int_{u-x}^{\beta} f_{XY}(x,y)\, dy dx$$

and hence

$$\begin{aligned}
f_U(u) &= \int_{u-\beta}^{\beta} f_{XY}(x, u-x) dx = \alpha^2 \beta^{-2\alpha} \int_{u-\beta}^{\beta} x^{\alpha-1} (u-x)^{\alpha-1}\, dx \\
&= \alpha^2 \beta^{-2\alpha} u^{2\alpha-1} \int_{1-\beta/u}^{\beta/u} y^{\alpha-1} (1-y)^{\alpha-1}\, dy, \quad \beta < u \le 2\beta
\end{aligned}$$

(b)

$$Z = \min(X,Y), \qquad W = \max(X,Y)$$

We can proceed as in Example 6-21 to complete this problem. From (6-92) and (6-93), we get

$$F_{ZW}(z,w) = \begin{cases} F_{XY}(z,w) + F_{XY}(w,z) - F_{XY}(z,z), & w \ge z \\ F_{XY}(w,w), & w < z \end{cases}$$

which gives

$$f_{ZW}(z,w) = f_X(z)f_Y(w) + f_X(w)f_Y(z), \quad 0 < z \le w < \beta$$

$$f_{ZW}(z,w) = \begin{cases} 2\alpha^2\beta^{-2\alpha}z^{\alpha-1}w^{\alpha-1}, & 0 < z \le w < \beta \\ 0, & \text{otherwise} \end{cases}$$

check:

$$\begin{aligned}\int_0^\beta \int_0^w f_{ZW}(z,w)dzdw &= 2\alpha^2\beta^{-2\alpha}\int_0^\beta w^{\alpha-1}\left(\left.\frac{z^\alpha}{\alpha}\right|_0^w\right)dw \\ &= 2\alpha\beta^{-2\alpha}\int_0^\beta w^{2\alpha-1}dw = 1\end{aligned}$$

Note: Z and W are not independent random variables, since

$$f_Z(z) = 2\alpha\beta^{-2\alpha}\, z^{\alpha-1}\,(\beta^\alpha - z^\alpha), \quad 0 < z < \beta$$

and

$$f_W(w) = 2\alpha\beta^{-2\alpha}\, w^{2\alpha-1}, \quad 0 < w < \beta$$

(c) Let

$$V = \frac{Z}{W} = \frac{\min(X,Y)}{\max(X,Y)} = \begin{cases} \frac{Y}{X}, & X \ge Y \\ \frac{X}{Y}, & X < Y \end{cases}$$

and

$$W = \max(X,Y) = \begin{cases} X, & X \ge Y \\ Y, & X < Y \end{cases}$$

For $0 < v < 1,\ 0 < w < \beta$

$$\begin{aligned}F_{VW}(v,w) &= P(V \le v,\, W \le w) \\ &= P\{V \le v,\, W \le w,\ (X \ge Y)\bigcup(X < Y)\} \\ &= P\{Y \le Xv,\, X \le w,\, X \ge Y\} \\ &+ P\{X < Yv,\, Y \le w,\, X < Y\} \\ &= \int_0^w \int_0^{xv} f_{XY}(x,y)\,dy\,dx + \int_0^w \int_0^{yv} f_{XY}(x,y)\,dx\,dy\end{aligned}$$

Hence

$$\begin{aligned}
f_{VW}(v,w) &= \frac{\partial^2 F_{VW}(v,w)}{\partial v\,\partial w} \\
&= \frac{\partial}{\partial v}\left\{\int_0^{vw} f_{XY}(w,y)\,dy + \int_0^{vw} f_{XY}(x,w)\,dx\right\} \\
&= w\left\{f_{XY}(w,vw) + f_{XY}(vw,w)\right\} \\
&= 2\alpha^2\,\beta^{-2\alpha}\,w^{2\alpha-1}\,v^{\alpha-1}, \qquad 0<v<1,\ 0<w<\beta
\end{aligned}$$

Hence

$$f_V(v) = \int_0^{\beta} f_{VW}(v,w)dw = \alpha\,v^{\alpha-1}, \quad 0<v<1$$

$$f_W(w) = \int_0^1 f_{VW}(v,w)dv = 2\alpha\,\beta^{-2\alpha}\,w^{2\alpha-1}, \quad 0<w<\beta$$

and

$$f_{VW}(v,w) = f_V(v)\,f_W(w).$$

Thus V and W are independent random variables.

6.31 (a) Solved in Examples 6-27 and 6-12.

(b) Solved in Example 6-27.

(c)

$$Z = X+Y, \qquad W = \frac{X}{X+Y}$$

$$x_1 = zw, \qquad y_1 = z - x_1 = z(1-w)$$

$$J = \begin{vmatrix} 1 & 1 \\ \dfrac{y}{(x+y)^2} & -\dfrac{x}{(x+y)^2} \end{vmatrix} = \frac{1}{x+y} = \frac{1}{z}$$

$$\begin{aligned}
f_{ZW}(z,w) &= \frac{z}{\alpha^{m+n}\Gamma(m)\Gamma(n)}\,(zw)^{m-1}\left\{z(1-w)\right\}^{n-1} \\
&= \left(\frac{z^{m+n-1}}{\alpha^{m+n}\Gamma(\alpha+\beta)}\,e^{-z/\alpha}\right)\left(\frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}\,w^{m-1}\,(1-w)^{n-1}\right) \\
&= f_Z(z)\,f_W(w)
\end{aligned}$$

Thus Z and W are independent random variables.

6.32 (a)

$$Z = \frac{X}{|Y|}, \qquad W = \frac{|X|}{|Y|} = |Z|$$

$$\begin{aligned} F_Z(z) &= P(Z \le z) = P(X \le |Y|z) = \int_{-\infty}^{\infty} \int_{0}^{|y|z} f_{XY}(x,y)\,dx\,dy \\ &= 2\int_0^\infty |y| f_{XY}(|y|z, y) dy = \frac{2}{2\pi\sigma^2}\int_0^\infty y e^{-(z^2+1)y^2/2\sigma^2}\,dy \\ &= \frac{1/\pi}{1+z^2}, \qquad -\infty < z < \infty \end{aligned}$$

Thus Z is a Cauchy random variable. Interestingly, the random variable X/Y is also a Cauchy random variable (see Example 6-11).

$$W = |Z|$$

so that

$$\begin{aligned} F_W(w) &= P(W \le w) = P(|Z| \le w) \\ &= P(-w < Z < w) = F_Z(w) - F_Z(-w) \end{aligned}$$

and hence

$$f_W(w) = f_Z(w) + f_Z(-w) = \frac{2/\pi}{1+w^2}, \quad w > 0.$$

(b)

$$U = X + Y \sim N(0, 2)$$
$$V = X^2 + Y^2 \sim \text{Exponential }(2)$$

(see Example 6-14). Here U, V are *not* independent, since

$$J(x,y) = \begin{vmatrix} 1 & 1 \\ 2x & 2y \end{vmatrix} = -2(x-y) = 2\sqrt{2v - u^2}$$

and

$$\begin{aligned} f_{UV}(u,v) &= \frac{1}{2\sqrt{2v-u^2}}\frac{1}{2\pi\sigma^2}\,e^{-v/2\sigma^2} \\ &\ne f_U(u)\,f_V(v), \qquad -\infty < u < \infty, \;\; v > 0. \end{aligned}$$

6.33

$$Z = X + Y, \qquad W = X - Y$$

are jointly normal random variables. Hence if they are uncorrelated, then they are also independent.

$$\begin{aligned} Cov(Z,W) &= E[(Z-\mu_Z)(W-\mu_W)] \\ &= E\left[\{(X-\mu_X) + (Y-\mu_Y)\}\{(X-\mu_X) - (Y-\mu_Y)\}\right] \\ &= \text{Var}(X) - \text{Var}(Y) = \sigma_X^2 - \sigma_Y^2. \end{aligned}$$

The random variables Z and W are uncorrelated implies that $Cov(Z,W) = 0$. Hence $\sigma_X^2 = \sigma_Y^2$ is the necessary and sufficient condition for the independence of $X+Y$ and $X-Y$.

6.34 (a)-(b) Let

$$R = \sqrt{X^2+Y^2}, \qquad \theta = \tan^{-1}\left(\frac{Y}{X}\right)$$

Form Example 6-22, R and θ are independent random variables with joint p.d.f. as in (6-128). (see (6-131)). In term of R and θ, we have $X = R\cos\theta, Y = R\sin\theta$ and hence we obtain

$$U = \frac{X^2-Y^2}{\sqrt{X^2+Y^2}} = R\cos 2\theta$$

$$V = \frac{2XY}{\sqrt{X^2+Y^2}} = R\sin 2\theta$$

This gives

$$J = \begin{vmatrix} \cos 2\theta & -2r\sin 2\theta \\ \sin 2\theta & 2r\cos 2\theta \end{vmatrix} = 2r = 2\sqrt{u^2+v^2}$$

$$r = \sqrt{u^2+v^2}, \quad \theta_1 = \frac{1}{2}\tan^{-1}\left(\frac{v}{u}\right), \quad 2\theta_2 = \pi + 2\theta_1.$$

There are two sets of solutions (r,θ_1) and (r,θ_2). Substituting into (6-128) we get

$$\begin{aligned} f_{UV}(u,v) &= \frac{1}{|J|}\{f_{r,\theta}(r,\theta_1)+f_{r,\theta}(r,\theta_2)\} = \frac{2}{|J|}f_{r,\theta}(r,\theta_1) \\ &= \frac{2}{2\sqrt{u^2+v^2}}\,\frac{\sqrt{u^2+v^2}}{2\pi\sigma^2}\,e^{-(u^2+v^2)/2\sigma^2} \\ &= \frac{1}{2\pi\sigma^2}\,e^{-(u^2+v^2)/2\sigma^2} = \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-u^2/2\sigma^2}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-v^2/2\sigma^2} \\ &= f_U(u)f_V(v) \end{aligned}$$

Thus U and V are independent normal random variables. Hence it follows that $U = \dfrac{X^2-Y^2}{\sqrt{X^2+Y^2}}$ and $V/2 = \dfrac{XY}{\sqrt{X^2+Y^2}}$ are independent random variables.

(c)

$$\begin{aligned} Z &= \frac{(X-Y)^2-2Y^2}{\sqrt{X^2+Y^2}} = \frac{(X^2-Y^2)-2XY}{\sqrt{X^2+Y^2}} \\ &= \frac{X^2-Y^2}{\sqrt{X^2+Y^2}} - \frac{2XY}{\sqrt{X^2+Y^2}} \\ &= U - V \sim N(0,2\sigma^2). \end{aligned}$$

6.35 (a) $Z \sim F(m,n)$ is given by (6-157) Let

$$Y = \frac{1}{Z}$$

Then

$$\begin{aligned} F_Y(y) &= \frac{1}{|dy/dz|}\, f_Z(1/y) \\ &= \frac{1}{y^2}\, \frac{(m/n)^{m/2}}{\beta(m/2,n/2)}\, \frac{1}{y^{m/2-1}}\, \frac{1}{(1+m/ny)^{m+n/2}} \\ &= \frac{(n/m)^{n/2}}{\beta(n/2,m/2)}\, y^{n/2-1} \left(1+\frac{n}{my}\right)^{-(m+n)/2} \\ &\sim F(n,m). \end{aligned}$$

(b)

$$W = \frac{Zm}{Zm+n}$$

$$\begin{aligned} F_W(w) &= P(W \le w) = P\left(\frac{Zm}{Zm+n} \le w\right) \\ &= P\left(Z \le \frac{nw}{m(1-w)}\right) = F_Z\left(\frac{nw}{m(1-w)}\right) \end{aligned}$$

which gives

$$\begin{aligned} f_W(w) &= \frac{n}{m(1-w)^2}\, f_Z\left(\frac{nw}{m(1-w)}\right) \\ &= \frac{n}{m(1-w)^2}\, \frac{(m/n)^{m/2}}{\beta(m/2,n/2)} \left(\frac{nw}{m(1-w)}\right)^{m/2-1} \left(1+\frac{w}{(1-w)}\right)^{-(m+n)/2} \\ &= \frac{1}{\beta(m/2,n/2)}\, w^{m/2-1}\,(1-w)^{n/2-1}, \quad 0<w<1. \end{aligned}$$

Thus W has Beta distribution.

6.36

$$Z = X+Y > 0, \qquad W = X-Y > 0$$

$$x_1 = \frac{z+w}{2}, \qquad y_1 = \frac{z-w}{2}$$

is the only solution. Moreover

$$J = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

so that

$$f_{ZW}(z,w) = \frac{1}{|J|} f_{XY}(x_1,y_1) = \frac{1}{2}\, e^{-(z+w)/2}, \quad 0<w<z<\infty$$

$$\begin{aligned} F_Z(z) &= \int_0^z f_{ZW}(z,w)\,dw = \tfrac{1}{2}\, e^{-z/2}\, \left.\frac{e^{-w/2}}{(-1/2)}\right|_0^z \\ &= \tfrac{1}{2}\, e^{-z/2}\, \left.\frac{e^{-w/2}}{(-1/2)}\right|_0^z = e^{-z/2}(1-e^{-z/2}), \quad z>0 \end{aligned}$$

6.37

$$Z = X + Y > 0, \qquad W = \frac{Y}{X} > 1$$

$$y = xw, \quad x(1+w) = z, \quad x_1 = \frac{z}{1+w}, \quad y_1 = \frac{zw}{1+w}$$

is the only solution. Also

$$J = \begin{vmatrix} 1 & 1 \\ -\frac{y}{x^2} & \frac{1}{x} \end{vmatrix} = \frac{x+y}{x^2} = \frac{(1+w)^2}{z}$$

This gives

$$\begin{aligned} f_{ZW}(z,w) &= \frac{1}{|J|} f_{XY}(x_1, y_1) \\ &= \frac{z}{(1+w)^2} 2\, e^{-z}, \quad z > 0,\ w > 1 \\ &= z\, e^{-z} \frac{2}{(1+w)^2} = f_Z(z)\, f_W(w) \end{aligned}$$

since

$$\begin{aligned} f_Z(z) &= \int_1^\infty f_{ZW}(z,w)\, dw \\ &= 2ze^{-z} \int_1^\infty \frac{1}{(1+w)^2}\, dw = z\, e^{-z}, \quad z > 0 \end{aligned}$$

and

$$\begin{aligned} f_w(w) &= \int_0^\infty f_{ZW}(z,w)\, dz \\ &= \frac{2}{(1+w)^2} \int_0^\infty ze^{-z}\, dz = \frac{2}{(1+w)^2}, \quad w > 1. \end{aligned}$$

Thus Z and W are independent random variables.

6-38 $\quad \underset{\sim}{z} = \underset{\sim}{x}\underset{\sim}{y} \qquad \underset{\sim}{y} = \cos(\omega t + \underset{\sim}{\theta})$

$\underset{\sim}{w} = \underset{\sim}{y} \qquad J = |y| \qquad f_y(y) = \begin{cases} \dfrac{1}{\pi\sqrt{1-y^2}} & |y| < 1 \\ 0 & |y| > 1 \end{cases}$

The RVs $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are independent. Hence,

$$f_{zw}(z,w) = \frac{1}{|w|} f_x\left(\frac{z}{w}\right) f_y(w)$$

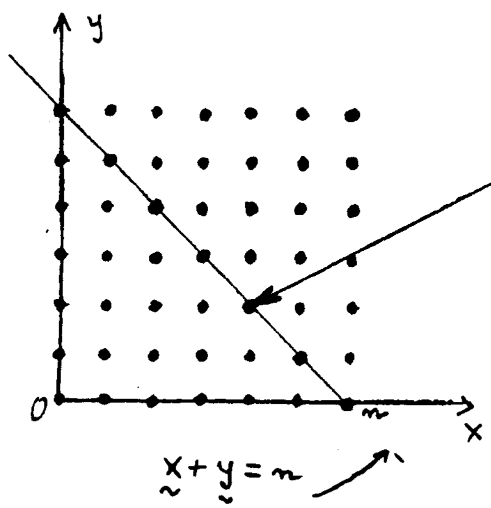
$$f_z(z) = \frac{1}{\pi}\int_{-1}^{1} \frac{f_x(z/w)}{|w|\sqrt{1-w^2}}\, dw = \frac{1}{\pi}\int_{|x|>z} \frac{f_x(x)}{\sqrt{x^2-z^2}}\, dx$$

6-39 $\quad \underset{\sim}{z} = \underset{\sim}{x} + \underset{\sim}{s} \qquad \underset{\sim}{s} = a\cos\underset{\sim}{y}$

$f_z(z) = f_x(z) * f_s(z) \qquad f_s(s) = \begin{cases} \dfrac{1}{\pi\sqrt{a^2-s^2}} & |s| < a \\ 0 & |s| > a \end{cases}$

$$f_z(z) = \frac{1}{\pi\sigma\sqrt{2\pi}}\int_{-a}^{a} \frac{e^{-(z-s)^2/2\sigma^2}}{\sqrt{a^2-s^2}}\, ds = \frac{1}{\pi\sigma\sqrt{2\pi}}\int_{-\pi}^{\pi} e^{-(z-a\cos y)^2/2\sigma^2}\, dy$$

6-40



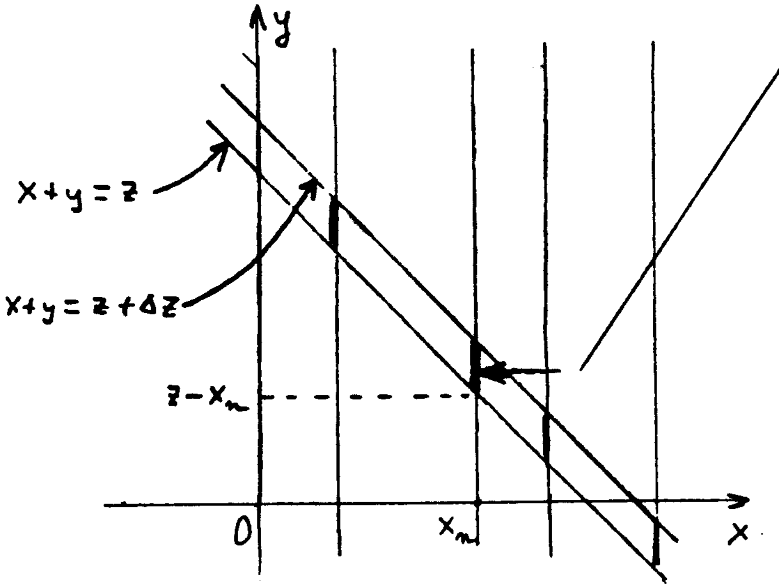
Point masses

$P\{\underset{\sim}{x} = k,\ \underset{\sim}{y} = n-k\} = a_k b_{n-k}$

$$\{\underset{\sim}{z} = n\} = \sum_{k=0}^{n} \{\underset{\sim}{x} = k,\ \underset{\sim}{y} = n-k\}$$

$$P\{\underset{\sim}{z} = n\} = \sum_{k=0}^{n} P\{\underset{\sim}{x} = k,\ \underset{\sim}{y} = n-k\}$$

6-41 (a) **Line masses**

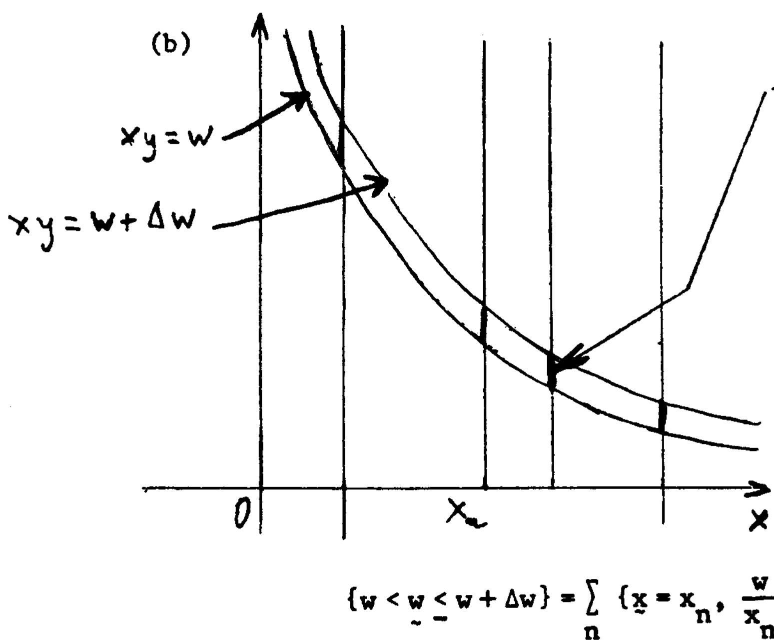


$$P\{\underset{\sim}{x} = x_n,\ z - x_n < \underset{\sim}{y} \le z - x_n + \Delta z\} = p_n f_y(z - x_n)\Delta z$$

$$\{z < \underset{\sim}{z} \le z + \Delta z\} = \sum_n \{\underset{\sim}{x} = x_n,\ z - x_n < \underset{\sim}{y} \le z - x_n + \Delta z\}$$

$$f_z(z)\Delta z = \sum_n p_n f_y(z - x_n)\Delta z$$

(b)



$$P\{\underset{\sim}{x} = x_n,\ \frac{w}{x_n} < \underset{\sim}{y} \le \frac{w+\Delta w}{x_n}\} = p_n f_y\left(\frac{w}{x_n}\right)\frac{\Delta w}{x_n}$$

$$\{w < \underset{\sim}{w} \le w + \Delta w\} = \sum_n \{\underset{\sim}{x} = x_n,\ \frac{w}{x_n} < \underset{\sim}{y} \le \frac{w+\Delta w}{x_n}\}$$

$$f_w(w)\Delta w = \sum_n p_n f_y\left(\frac{w}{x_n}\right)\Delta w$$

6.42 X, Y are independent geometric random variables. Thus

$$\begin{aligned} P\{X=k, Y=m\} &= P\{X=k\}\, P\{Y=m\} \\ &= (pq^k)\,(pq^m) = p^2 q^{k+m}, \qquad k, m = 0, 1, 2, \cdots \end{aligned}$$

(a) Let

$$Z = X + Y$$

$$\begin{aligned} P\{Z=n\} &= P\{X+Y=n\} = \sum_k P\{X=k, Y=n-k\} \\ &= \sum_{k=0}^{n} P\{X=k, Y=n-k\} \\ &= \sum_{k=0}^{n} P\{X=k\}\, P\{Y=n-k\} \\ &= \sum_{k=0}^{n} pq^k\, pq^{n-k} = \sum_{k=0}^{n} p^2\, q^n \\ &= (n+1)\, p^2\, q^n, \quad n = 0, 1, 2, \cdots \end{aligned}$$

(b) Let

$$W = X - Y$$

Case 1: $\quad W \geq 0 \Rightarrow X \geq Y$. Thus for $m \geq 0$

$$\begin{aligned} P\{W=m\} &= P\{X-Y=m\} = \sum_{k=0}^{\infty} P\{X=m+k, Y=k\} \\ &= \sum_{k=0}^{\infty} P\{X=m+k, Y=k\} \\ &= \sum_{k=0}^{\infty} P\{X=m+k\}\, P\{Y=k\} \\ &= \sum_{k=0}^{\infty} (pq^{m+k})\,(pq^k) = p^2 q^m \sum_{k=0}^{\infty} q^{2k} \\ &= p^2 q^m\,(1 + q^2 + q^4 + \cdots) = \frac{p^2 q^m}{(1-q^2)} \\ &= \frac{pq^m}{1+q}, \qquad m = 0, 1, 2, \cdots \qquad (1) \end{aligned}$$

Case 2: $\quad W < 0 \Rightarrow X < Y$. Thus for $m < 0$

$$\begin{aligned} P\{W=m\} &= P\{X-Y=m\} = \sum_k P\{X=k, Y=k-m\} \\ &= \sum_{k=0}^{\infty} P\{X=k, Y=k-m\} \\ &= \sum_{k=0}^{\infty} P\{X=k\}\, P\{Y=k-m\} \\ &= \sum_{k=0}^{\infty} (pq^k)\,(pq^{k-m}) = p^2 q^{-m} \sum_{k=0}^{\infty} q^{2k} \\ &= \frac{p^2 q^{-m}}{(1-q^2)} = \frac{pq^{-m}}{1+q}, \quad m = -1, -2, \cdots \qquad (2) \end{aligned}$$

Thus combining (1) and (2) we can write

$$P\{W=m\} = \frac{pq^{|m|}}{1+q}, \qquad m = 0, \pm 1, \pm 2, \cdots$$

6.43 We have X and Y are independent and $P(X=k)=P(Y=k)=p_k$. Also

$$\begin{aligned} P(X=k|X+Y=k) &= \frac{P(X=k,Y=0)}{P(X+Y=k)} \\ &= \frac{p_k p_0}{\sum\limits_{i=0}^{k} p_i p_{k-i}} = \frac{1}{k+1}. \qquad (1) \end{aligned}$$

Also

$$\begin{aligned} &P(X=k-1|X+Y=k) \\ &= \frac{P(X=k-1,Y=1)}{P(X+Y=k)} \frac{p_{k-1}p_1}{\sum_{i=0}^{k} p_i p_{k-i}} = \frac{1}{k+1}. \qquad (2) \end{aligned}$$

From (1) and (2),

$$\frac{p_k}{p_{k-1}} = \frac{p_1}{p_0} \Rightarrow p_k = \lambda p_{k-1} = \lambda^k p_0$$

where $\lambda \triangleq p_1/p_0$. Since $\sum_{k=0}^{\infty} p_k = 1$, we must have $\lambda < 1$, and this gives

$$\sum_{k=0}^{\infty} p_k = \frac{p_0}{1-\lambda} = 1 \rightarrow p_0 = 1-\lambda.$$

Thus

$$p_k = p_0\,\lambda^k = (1-\lambda)\lambda^k, \quad k=0,1,2,\cdots,\ 0<\lambda<1$$

represents a geometric distribution. Thus X and Y are geometric random variables.

6.44 The moment generating functions of X and Y are given by (see (5-117))

$$\Gamma_X(z) = (pz+q)^n, \qquad \Gamma_Y(z) = (pz+q)^n$$

Also

$$\Gamma_{X+Y}(z) = E[z^{X+Y}] = \Gamma_X(z)\Gamma_Y(z) = (pz+q)^{2n} \sim \text{Binomial}\,(2n,p)$$

6.45 (a) Let

$$Z = \min(X, Y), \qquad W = X - Y$$

$$\begin{aligned}
&P\{Z = k, W = m\} \\
&= P\left\{\min(X, Y) = k, X - Y = m\right\} \\
&= P\left\{(\min(X, Y) = k, X - Y = m) \bigcap (X \geq Y \bigcup X < Y)\right\} \\
&= P\{Y = k, X - Y = m, X \geq Y\} + P\{X = k, X - Y = m, X < Y\} \\
&= P\{X = m + k, Y = k, X \geq Y\} + P\{X = k, Y = k - m, X < Y\}
\end{aligned}$$

Note that $k \geq 0$, and m takes both positive, zero and negative values. Hence

$$\begin{aligned}
P\{Z = k, W = m\} &= \begin{cases} P\{X = k + m, Y = k, X \geq Y\}, & k \geq 0,\ m \geq 0 \\ P\{X = k, Y = k - m, X < Y\}, & k \geq 0,\ m < 0 \end{cases} \\
&= \begin{cases} pq^{k+m}\, pq^k, & k \geq 0,\ m \geq 0 \\ pq^k\, pq^{k-m}, & k \geq 0,\ m < 0 \end{cases}
\end{aligned}$$

$$P\{Z = k, W = m\} = p^2 q^{2k+|m|}, \qquad k = 0, 1, 2, \cdots, \quad m = 0, \pm 1, \pm 2, \cdots$$

Also

$$\begin{aligned}
P\{Z = k\} &= \sum_{m=-\infty}^{\infty} P\{Z = k, W = m\} \\
&= p^2 q^{2k} \sum_{m=-\infty}^{\infty} q^{|m|} = p^2 q^{2k} \left(1 + 2\sum_{m=1}^{\infty} q^m\right) \\
&= p^2 q^{2k} \left(1 + \frac{2q}{p}\right) = p(1+q)q^{2k}, \qquad k = 0, 1, 2, \cdots
\end{aligned}$$

and

$$\begin{aligned}
P\{W = m\} &= \sum_{k=0}^{\infty} P\{Z = k, W = m\} \\
&= p^2 q^{|m|} \sum_{k=0}^{\infty} q^{2k} \\
&= \frac{p}{1+q} q^{|m|}, \qquad m = 0, \pm 1, \pm 2, \cdots
\end{aligned}$$

Note that

$$P\{Z = k, W = m\} = P\{Z = k\}\, P\{W = m\}$$

and hence Z and W are independent random variables.
(b) Let

$$Z = \min(X, Y), \quad W = \max(X, Y) - \min(X, Y)$$

Proceeding as in (a), we obtain

$$\begin{aligned}
&P\{Z = k, W = m\} \\
&= P(Y = k, X - Y = m, X \geq Y) + P(X = k, Y - X = m, X < Y) \\
&= P(X = k + m, Y = k, X \geq Y) + P(X = k, Y = k + m, X < Y) \\
&= \begin{cases} pq^{k+m}\, pq^k + pq^k\, pq^{k+m}, & k = 0, 1, 2, \cdots, \ m = 1, 2, \cdots \\ pq^{k+m}\, pq^k, & k = 0, 1, 2, \cdots, \ m = 0 \end{cases} \\
&= \begin{cases} 2p^2 q^{2k+m}, & k = 0, 1, 2, \cdots, \ m = 1, 2, \cdots \\ p^2 q^{2k}, & k = 0, 1, 2, \cdots, \ m = 0 \end{cases}
\end{aligned}$$

This gives

$$\begin{aligned}
P\{Z = k\} &= \sum_{m=0}^{\infty} P\{Z = k, W = m\} \\
&= p^2 q^{2k} \left(1 + 2\sum_{m=1}^{\infty} q^m\right) = p^2 q^{2k} \left(1 + \frac{2q}{p}\right) \\
&= p(1 + q) q^{2k}, \quad k = 0, 1, 2, \cdots
\end{aligned}$$

Also

$$\begin{aligned}
P\{W = m\} &= \sum_{k=0}^{\infty} P\{Z = k, W = m\} \\
&= \begin{cases} \dfrac{p}{1 + q}, & m = 0 \\ \dfrac{2p}{1 + q}\, q^m, & m = 1, 2, \cdots \end{cases}
\end{aligned}$$

Notice that

$$P\{Z = k, W = m\} = P\{Z = k\}\, P\{W = m\}$$

and hence Z and W are also indepedndent random variables in this case also.

6.46 The moment generating function of X and Y are given by (see (5-119))

$$\Gamma_X(z) = e^{\lambda_1(z-1)}, \qquad \Gamma_Y(z) = e^{\lambda_2(z-1)}$$

Also

$$\Gamma_{X+Y}(z) = \Gamma_X(z)\Gamma_Y(z) = e^{(\lambda_1+\lambda_2)(z-1)}$$

so that

$$Z \sim P(\lambda_1 + \lambda_2)$$

Thus

$$P(X+Y=k) = e^{-(\lambda_1+\lambda_2)}\,\frac{(\lambda_1+\lambda_2)^k}{k!}$$

and

$$\begin{aligned}
&P(X=k|X+Y=n)\\
&\quad= \frac{P(X=k, X+Y=n)}{P(X+Y=n)} = \frac{P(X=k)P(Y=n-k)}{P(X+Y=n)}\\
&\quad= \frac{e^{-\lambda_1}(\lambda_1^k/k!)\,e^{-\lambda_2}(\lambda_2^{n-k}/(n-k)!)}{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n/n!}\\
&\quad= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}, \quad k = 0,1,2,\cdots n\\
&\quad\sim \text{Binomial}(n,p), \text{ where } p = \frac{\lambda_1}{\lambda_1+\lambda_2}.
\end{aligned}$$

See also (6-222). From there the converse is also true (proceed as in Example 6-43).

6-47

$$C = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \qquad \Delta = \sigma_1^2\sigma_2^2(1-r^2)$$

$$C^{-1} = \begin{bmatrix} \dfrac{1}{(1-r^2)\sigma_1^2} & \dfrac{r}{(1-r^2)\sigma_1\sigma_2} \\ \dfrac{r}{(1-r^2)\sigma_1\sigma_2} & \dfrac{1}{(1-r^2)\sigma^2} \end{bmatrix}$$

$$XC^{-1}X^t = \frac{1}{(1-r^2)}\left(\frac{x_1^2}{\sigma_1^2} - 2r\,\frac{x_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)$$

6-48

$$\{\underset{\sim}{x}\,\underset{\sim}{y} < 0\} = \{\underset{\sim}{x} < 0,\ \underset{\sim}{y} > 0\} + \{\underset{\sim}{x} > 0,\ \underset{\sim}{y} < 0\}$$

$$P\{\underset{\sim}{x}\,\underset{\sim}{y} < 0\} = F_x(0)[1 - F_y(0)] + [1 - F_x(0)]F_y(0)$$

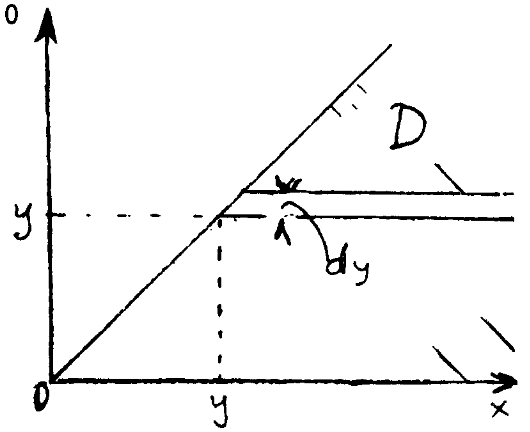
$$F_x(0) = 1 - G\left(\frac{\eta_x}{\sigma_x}\right) \qquad F_y(0) = 1 - G\left(\frac{\eta_y}{\sigma_y}\right)$$

6-49 If $w = x - y$, then $E\{w\} = 0$ $\qquad \sigma_w^2 = \sigma_x^2 + \sigma_y^2 = 2\sigma^2$

Thus, $w = 1$, $N(0;\ \sigma\sqrt{2})$ and [see (5-74)]

$$E\{z\} = E\{|w|\} = \sqrt{2}\,\sigma\sqrt{\frac{2}{\pi}} \qquad E\{z^2\} = E\{w^2\} = 2\sigma^2$$

6-50



$$E\{z\} = \iint_D (x-y)f(x,y)\,dxdy$$

$$= \int_0^\infty\int_y^\infty (x-y)e^{-x}\,e^{-y}\,dxdy = \frac{1}{2}$$

6-51 Since $|E\{x\ y\}| \le E\{|x||y|\}$, we can assume that the RVs x and y are real

(a) $D \le E\{[z\ x - y]^2\} = z^2\ E\{x^2\} - 2z\ E\{x\ y\} + E\{y^2\}$

The above is a non-negative quadratic in z for any z. Hence, its discriminant is non-positive.

(b) Using (a), we obtain

$$E\{x^2\} + E\{y^2\} + 2\sqrt{E\{x^2\}E\{y^2\}}$$

$$\ge E\{x^2\} + E\{y^2\} + 2\ E\{x\ y\} = E\{(x + y)^2\}$$

6-52 If $r_{xy} = 1$ then

$$E^2\{(x - \eta_x)(y - \eta_y)\} = E\{(x - \eta_x)^2\}E\{(y - \eta_y)^2\}$$

i.e., the discriminant of the quadratic

$$E\{[z(x - \eta_x) - (y - \eta_y)]^2\}$$

is zero. This is possible only if the quadratic is zero for some $z = z_o$. This shows that $z(x - \eta_x) - (y - \eta_y) = 0$ in the MS sense.

6-53 If $E\{x\} = E\{y^2\} = E\{x\ y\}$, then

$$E\{(x - y)^2\} = E\{x^2\} + E\{y^2\} - 2\ E\{x\ y\} = 0.$$

Hence, $x = y$ in the MS sense.

6-54 If x has a Cauchy density, then (Prob. 5-31)

$$E\{e^{j\omega x}\} = e^{-\alpha|\omega|} \qquad E\{e^{j\omega kx}\} = e^{-\alpha k|\omega|}$$

Hence, [see (6-240)]

$$\Phi_z(\omega) = E\{e^{j\omega nx}\} = E\{E\{e^{j\omega nx}|n\}\} =$$

$$\sum_{k=0}^{\infty} E\{e^{j\omega kx}\}e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty} e^{-\alpha k|\omega|}\frac{k}{k!} = e^{-\lambda}e^{-\lambda e^{-\alpha|\omega|}}$$

6.55 If $X = k$, then

$$Y = n - k$$

and

$$Z = X - Y = 2X - n,$$

where Z takes the values $-n, -(n-2), \cdots n-2, n$.

$$\begin{aligned} P\{Z = z\} &= P\{2X - n = z\}P\{X = \tfrac{n+z}{2}\} \\ &= \binom{n}{n + z/2} p^{(n+z)/2}\ q^{(n-z)/2}. \end{aligned}$$

Also

$$E(Z) = E[2X - n] = 2np - n = n(2p - 1).$$

$$\text{Var}(Z) = E[(z - \mu_z)^2] = 4E[(X - np)^2] = 4\text{Var}(X) = 4npq$$

6.56 (a)

$$\begin{aligned}\phi_Z(\omega) &= E[e^{j\omega Z}] = E[e^{j\omega(aX+bY+c)}] \\ &= \phi_X(a\omega)\,\phi_Y(b\omega)e^{j\omega c} = e^{j\omega c-(a^2\sigma_1^2+b^2\sigma_2^2)\omega^2/2}\end{aligned}$$

(see (5-100)).
(b) On comparing with (5-100) we obtain

$$Z \sim \mathrm{N}(c, a^2\,\sigma_1^2 + b^2\,\sigma_2^2)$$

(c)

$$E[Z] = c, \qquad \mathrm{Var}(Z) = a^2\,\sigma_1^2 + b^2\,\sigma_2^2$$

6.57

$$P(X = k|Y = n) = \binom{n}{k} p_1^k\, q_1^{n-k}, \quad k = 0, 1, 2, \cdots n$$

$$E[e^{j\omega X}|Y = n] = \sum_{k=0}^{n} e^{j\omega k}\, P(X = k|Y = n) = \left(p_1 e^{j\omega} + q_1\right)^n$$

use (5-117). Also

$$\begin{aligned}\phi_X(\omega) &= E[e^{j\omega X}] = E\left\{E[e^{j\omega X}|Y = n]\right\} \\ &= \sum\nolimits_{n=0}^{M} E[e^{j\omega X}|Y = n]\,P(Y = n) \\ &= \sum\nolimits_{n=0}^{\infty} (p_1 e^{j\omega} + q_1)^n \binom{M}{n} p_2^n\, q_2^{M-n} \\ &= \sum\nolimits_{n=0}^{M} \binom{M}{n} [p_2(p_1\, e^{j\omega} + q_1)]^n\, q_2^{M-n} \\ &= (p_2 p_1 e^{j\omega} + q_1 p_2 + q_2)^M\end{aligned}$$

But

$$1 - p_1 p_2 = 1 - (1 - q_1)(1 - q_2) = q_1 p_2 + q_2$$

Hence

$$\phi_X(\omega) = \left(pe^{j\omega} + q\right)^M$$

where $p = p_1 p_2$. Thus

$$X \sim \mathrm{Binomial}(M, p_1 p_2).$$

6.58

$$\int\int f_{XY}(x,y)\,dx\,dy = \int_0^1 \int_x^1 kx\,dy\,dx = k\int_0^1 x(1-x)\,dx$$

$$\frac{k}{6} = 1 \quad \Rightarrow \quad k = 6.$$

$$f_X(x) = \int_x^1 6x\,dy = 6x(1-x), \quad 0 < x < 1.$$

$$f_Y(y) = \int_0^y 6x\,dy = 3y^2, \quad 0 < y < 1.$$

$$E[X] = \int_0^1 x\,f_X(x)\,dx = 6\left(\frac{x^3}{3} - \frac{x^4}{4}\right)\Big|_0^1 = \frac{1}{2}.$$

$$E[X^2] = \int_0^1 x^2\,f_X(x)\,dx = 6\left(\frac{x^4}{4} - \frac{x^5}{5}\right)\Big|_0^1 = \frac{3}{10}.$$

$$\mathrm{Var}(X) = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}.$$

$$E[Y] = \int_0^1 y\,f_Y(y)\,dy = 3\left.\frac{y^4}{4}\right|_0^1 = \frac{3}{4}.$$

$$E[Y^2] = \int_0^1 y^2\,f_Y(y)\,dy = 3\left.\frac{y^5}{5}\right|_0^1 = \frac{3}{5}.$$

$$\mathrm{Var}(Y) = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}.$$

$$\begin{aligned} E[XY] &= \int\int xy\,f_{XY}(x,y)\,dy\,dx \\ &= \int_0^1\int_x^1 xy\,6x\,dy\,dx = \int_0^1 3x^2\left(1-x^2\right)dx \\ &= 3\left(\tfrac{x^3}{3} - \tfrac{x^5}{5}\right)\Big|_0^1 = 3\left(\tfrac{1}{3} - \tfrac{1}{5}\right) = \tfrac{2}{5} \end{aligned}$$

$$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= \tfrac{2}{5} - \tfrac{1}{2}\tfrac{3}{4} = \tfrac{1}{40} \end{aligned}$$

6.59 (a)

$$\begin{aligned} \phi_{X,Y}(\omega_1,\omega_2) &= E[e^{j(\omega_1 X + \omega_2 Y)}] \\ &= E[e^{j\omega_1 X}]\;E[e^{j\omega_2 Y}] = \phi_X(\omega_1)\,\phi_Y(\omega_2) \\ &= e^{\lambda(e^{j\omega_1}-1)}\;e^{(j\mu\omega_2 - \sigma^2\omega_2^2/2)} \end{aligned}$$

(b)

$$\begin{aligned} \phi_Z(\omega) &= E[e^{j\omega Z}] \\ &= E[e^{j\omega(X+Y)}] = \phi_{X,Y}(\omega,\omega) \\ &= e^{\{\lambda(e^{j\omega}-1) + (j\mu\omega - \sigma^2\omega^2/2)\}} \end{aligned}$$

6.60 (a)

$$Z = \min(X, Y)$$

From Example 6-18, we have

$$f_Z(z) = 2\lambda\, e^{-2\lambda z}, \qquad z \geq 0$$

and hence

$$E[Z] = E[\min(X, Y)] = \frac{1}{2\lambda}$$

(b)

$$\begin{aligned}
E[\max(2X, Y)] &= \int\int \max(2x, y)\, f_{XY}(x, y)\, dx\, dy \\
&= \int\int_{2x \geq y} 2x\, f_{XY}(x, y)\, dx\, dy + \int\int_{2x < y} y\, f_{XY}(x, y)\, dx\, dy \\
&= \int_0^\infty \int_0^{2x} 2x\, \lambda^2\, e^{-\lambda x}\, e^{-\lambda y}\; dy\, dx + \int_0^\infty \int_0^{y/2} y\, \lambda^2\, e^{-\lambda x}\, e^{-\lambda y}\; dx\, dy \\
&= \lambda \int_0^\infty 2x\, e^{-\lambda x}(1 - e^{-2\lambda x})\, dx + \lambda \int_0^\infty y\, e^{-\lambda y}(1 - e^{-\lambda y/2})\, dy \\
&= 2\lambda \int_0^\infty \left(xe^{-\lambda x} + 2xe^{-2\lambda x} - 3xe^{-3\lambda x}\right) dx \\
&= \frac{2}{\lambda} \int_0^\infty (ue^{-u} + 2ue^{-2u} - 3ue^{-3u})\; du \\
&= \frac{2}{\lambda}\left(1 + \frac{2}{4} - \frac{3}{9}\right) = \frac{7}{3\lambda}.
\end{aligned}$$

6.61 (a)

$$Z = X - Y \quad \rightarrow \quad -1 < z < 1.$$

$z > 0$

$$\begin{aligned}
F_Z(z) &= P(X - Y \leq z) = 1 - P(X - Y > z) \\
&= 1 - \int_0^{(1-z)/2} \int_{y+z}^{1-y} f_{XY}(x, y)\, dx\, dy \\
&= 1 - \int_0^{(1-z)/2} \left(\int_{y+z}^{1-y} 6x dx\right)\; dy \\
&= 1 - 3\int_0^{(1-z)/2} \{(1 - z^2) - 2(1 + z)y\}\, dy \\
&= 1 - \frac{3}{4}(1 + z)(1 - z)^2, \quad z \geq 0.
\end{aligned}$$

$z < 0$

$$\begin{aligned}
F_Z(z) &= P(X - Y \leq z) \\
&= \int_0^{(1+z)/2} \int_{x-z}^{1-x} 6x\, dy\, dx = \int_0^{(1+z)/2} 6x\,(1 + z - 2x)\, dx \\
&= \frac{(1 + z)^3}{4}, \quad z < 0.
\end{aligned}$$

This gives

$$f_Z(z) = \begin{cases} \frac{3}{4}(1-z)(1+3z), & 0 < z < 1 \\ \dfrac{3(1+z)^2}{4}, & -1 < z < 0 \end{cases}$$

(b)

$$f_X(x) = \int_0^{1-x} 6x\,dy = 6x(1-x), \quad 0 < x < 1$$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{F_X(x)} = \frac{1}{1-x}, \quad 0 < y \le 1-x$$

(c)

$$W = X + Y$$

we have

$$F_W(w) = P(X+Y \le w) = \int_0^w \left(\int_0^{w-x} 6x\,dy \right) dx = w^3,$$

and

$$f_W(w) = \int_0^w 6x dx = 3w^2, \quad 0 < w < 1$$

$$E[W] = \frac{3}{4}$$

$$E[W^2] = \frac{3}{5}$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(W) = E(W^2) - (E(W))^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}.$$

6.62

$$X = \frac{1}{Z}.$$

where Z represents a Chi-square random variable. Thus (see (4-39))

$$f_Z(z) = \frac{z^{-1/2}}{\sqrt{2}\Gamma(1/2)} e^{-z/2} = \frac{z^{-1/2}}{\sqrt{2\pi}} e^{-z/2}$$

or

$$f_X(x) = \frac{1}{\left|\frac{dx}{dz}\right|} f_Z(1/x) = \frac{1}{x^2} \frac{x^{1/2}}{\sqrt{2\pi}} e^{-1/2x} = \frac{1}{\sqrt{2\pi}x^{3/2}} e^{-1/2x}, \quad x > 0$$

Also it is given that

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi x}} e^{-y^2/2x}$$

so that

$$f_{XY}(x,y) = f_{Y|X}(y|x)\, f_X(x) = \frac{1}{2\pi x^2} e^{-(1+y^2)/2x}$$

and hence

$$\begin{aligned} f_Y(y) &= \int_0^\infty f_{XY}(x,y)\,dx \\ &= \frac{1}{2\pi} \int_0^\infty \frac{1}{x^2} e^{-(1+y^2)/2x}\,dx \\ &= \frac{1}{2\pi} \frac{2}{1+y^2} \int_0^\infty e^{-u}\,du = \frac{1/\pi}{1+y^2}, \quad -\infty < y < \infty. \end{aligned}$$

Thus Y represents a Cauchy random variable.

6.63 (a) For any two random variables X and Y we have

$$\begin{aligned}\sigma_{X+Y}^2 &= \text{Var}(X+Y) = E[\{(X-\mu_X)+(Y-\mu_Y)\}^2] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y\rho_{XY} \\ &\leq (\sigma_X + \sigma_Y)^2\end{aligned}$$

since $|\rho_{XY}| \leq 1$. Thus

$$\sigma_{X+Y} \leq \sigma_X + \sigma_Y,$$

and hence it easily follows that

$$\frac{\sigma_{X+Y}}{\sigma_X + \sigma_Y} \leq 1.$$

(However, (b) is not so easy!)
(b) We shall prove this result in three parts by making use of Holder's inequality.
(i) **Holder's inequality:** The function $\log x$ is concave, for $0 < \alpha < 1$, and hence we have

$$\log[\alpha x_1 + (1-\alpha)x_2] \geq \alpha \log x_1 + (1-\alpha)\log x_2$$

or

$$x_1^\alpha x_2^{1-\alpha} \leq \alpha x_1 + (1-\alpha)x_2, \quad 0 < \alpha < 1. \qquad (6.63-1)$$

Let

$$x_1 = |x|^p, \quad \alpha = \frac{1}{p}, \text{ so that } 1-\alpha = 1 - \frac{1}{p} \triangleq \frac{1}{q}, \quad x_2 = |y|^q \quad (6.63-2)$$

so that (6.63-1) becomes

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}, \quad p > 1, \qquad (6.63-3)$$

the Holder's inequality. From (6.63-2), note that

$$\frac{1}{p} + \frac{1}{q} = 1, \quad p > 1, \quad q > 1 \qquad (6.63-4)$$

(ii) Define

$$x = X\left(E\{|X|^p\}\right)^{-1/p}, \quad y = Y\left(E\{|Y|^q\}\right)^{-1/q}$$

where p and q are as in (6.63-4). Substituting these into the Holder's inequality in (6.63-3), we get

$$\begin{aligned}|XY| &\le p^{-1}\,|X|^p\,\left(E\{|X|^p\}\right)^{1/p-1}\,\left(E\{|Y|\}\right)^{1/q}\\ &+q^{-1}\,|Y|^q\,\left(E\{|Y|^q\}\right)^{1/q-1}\,\left(E\{|X|^p\}\right)^{1/p}.\end{aligned} \qquad (6.63-5)$$

Taking expected values on both sides of (6.63-5), we get

$$E\{|XY|\} \le \left(E\{|X|^p\}\right)^{1/p}\,\left(E\{|Y|^q\}\right)^{1/q} \qquad (6.63-6)$$

which represents the generalization of the Cauchy-Schwarz inequality. (Note $p = q = 2$ corresponds to Cauchy-Schwarz inequality)

(iii) To prove the desired inequality, notice that

$$\begin{aligned}|X+Y|^p &= |X+Y||X+Y|^{p-1}\\ &\le |X||X+Y|^{p-1} + |Y||X+Y|^{p-1}, \quad p > 1\end{aligned}$$

and taking expected values on both sides we get

$$E\{|X+Y|^p\} \le E\{|X||X+Y|^{p-1}\} + E\{|Y||X+Y|^{p-1}\}. \quad (6.63-7)$$

Applying (6.63-6) to each term on the right side of (6.63-7) we get

$$E\{|X||X+Y|^{p-1}\} \le \left(E\{|X|^p\}\right)^{1/p}\left(E\{|X+Y|^{(p-1)q}\}\right)^{1/q} \quad (6.63-8)$$

and

$$E\{|Y||X+Y|^{p-1}\} \le \left(E\{|Y|^p\}\right)^{1/p}\left(E\{|X+Y|^{(p-1)q}\}\right)^{1/q} \quad (6.63-9)$$

Using (6.63-8) and (6.63-9) together with $(p-1)q = p$ in (6.63-7) we get

$$E\{|X+Y|^p\} \le \left[\left(E\{|X|^p\}\right)^{1/p} + \left(E\{|Y|^p\}\right)^{1/p}\right]\cdot\left(E\{|X+Y|^p\}\right)^{1/q}$$

or for $p > 1$

$$\left(E\{|X+Y|^p\}\right)^{1/p} \le \left(E\{|X|^p\}\right)^{1/p} + \left(E\{|Y|^p\}\right)^{1/p}.$$

the desired inequality. Since $p = 1$ follows trivially, we get

$$\frac{\left(E\{|X+Y|^p\}\right)^{1/p}}{\left(E\{|X|^p\}\right)^{1/p} + \left(E\{|Y|^p\}\right)^{1/p}} \le 1, \quad p \ge 1.$$

6.64 (a) See Example 6-41. From there

$$E(Y|X = x) = \mu_Y + \frac{\rho_{XY}\sigma_Y(x - \mu_X)}{\sigma_X}$$

(b) Similarly

$$f_{X|Y}(X|Y = y) \sim N(\mu, \sigma^2)$$

where

$$\mu = \mu_X + \frac{\rho_{XY}\sigma_X(y - \mu_Y)}{\sigma_Y}$$

and

$$\sigma^2 = \sigma_X^2(1 - \rho_{XY}^2).$$

Since

$$E(X^2|Y = y) = \mathrm{Var}(X|Y = y) + (E[X|Y = y])^2$$

we obtain

$$E(X^2|Y = y) = \sigma^2 + \mu^2$$

6.65 (a) See footnote 4, Chapter 8, Page 337. From there (or directly) we have

$$\mathrm{Var}(X|Y) \triangleq E(X^2|Y) - (E\{X|Y\})^2$$

$$\mathrm{Var}\left(E\{X|Y\}\right) \triangleq E\left[E\{X|Y\}\right]^2 - (E[E\{X|Y\}])^2$$

so that

$$\begin{aligned} E[\mathrm{Var}(X|Y)] + \mathrm{Var}\left(E\{X|Y\}\right) &= E\left[E\{X^2|Y\}\right] - (E[E\{X|Y\}])^2 \\ &= E(X^2) - [E(X)]^2 = \mathrm{Var}(X) \qquad (1) \end{aligned}$$

or

$$\mathrm{Var}(X) \geq E[\mathrm{Var}\{X|Y\}]$$

Also

$$\mathrm{Var}(X) \geq \mathrm{Var}[E\{X|Y\}]$$

(b) See (1).

6.66

$$Z = aX + (1-a)Y, \qquad 0 < a < 1$$

$$\sigma_Z^2 = \mathrm{Var}(Z) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$$

$$\frac{\partial \sigma_Z^2}{\partial a} = 2a\sigma_1^2 + 2(1-a)(-1)\sigma_2^2 = 0$$

or

$$a(\sigma_1^2 + \sigma_2^2) = \sigma_2^2$$

$$a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} < 1$$

minimizes $\mathrm{Var}(Z)$.

6-67 **From** (6-240)

$$E\{g(\underset{\sim}{x},\underset{\sim}{y})\} = E\{E\{g(\underset{\sim}{x},\underset{\sim}{y})\,|\,\underset{\sim}{y}\}\} = E\{g(x_n,\underset{\sim}{y})P\{\underset{\sim}{x} = x_n\}\} \ .$$

From (4-74) with $A_n = \{\underset{\sim}{x} = x_n\}$

$$f_z(z) = \sum_n f_z(z\,|\,\underset{\sim}{x} = x_n)P\{\underset{\sim}{x} = x_n\}$$

6-68 (a) The conditional density $f(y|x)$ is $N(rx;\sigma\sqrt{1-r^2})$ [see (7-42)]. Hence

$$E\{f_y(\underset{\sim}{y}|x)\} = \int_{-\infty}^{\infty} f_y(y|x)f_y(y)dy$$

$$= \frac{1}{2\pi\sigma^2\sqrt{1-r^2}}\int_{-\infty}^{\infty} \exp\left\{\frac{-(y-rx)^2}{2\sigma^2(1-r^2)}\right\}\exp\left\{\frac{-y^2}{2\sigma^2}\right\}dy = \frac{1}{\sigma\sqrt{2\pi(2-r^2)}}\exp\left\{\frac{-r^2x^2}{2\sigma^2(2-r^2)}\right\}$$

(b) From (6-241) it follows that

$$E\{f_x(\underset{\sim}{x})f_y(\underset{\sim}{y})\} = E\{f_x(\underset{\sim}{x})E\{f_y(\underset{\sim}{y}|\underset{\sim}{x})\}\} = \int_{-\infty}^{\infty} f_x(x)\ E\{f_y(y|x)\}f_x(x)dx$$

$$= \frac{1}{2\pi\sigma^3\sqrt{2\pi(2-r^2)}}\int_{-\infty}^{\infty}\exp\left\{-\frac{x^2}{\sigma^2}\right\}\exp\left\{\frac{-r^2x^2}{2\sigma^2(2-r^2)}\right\}dx = \frac{1}{2\pi\sigma^2\sqrt{4-r^2}}$$

6-69 We shall use (6-64) and Price's theorem (10-94):

$$\frac{\partial E\{|xy|\}}{\partial \mu} = E\left\{\frac{d|x|}{dx}\frac{d|y|}{\partial y}\right\} = E\{\operatorname{sgn} x \operatorname{sgn} y\}$$

$$= P\{xy > 0\} - P\{xy < 0\} = \frac{2\alpha}{\pi} = \frac{2}{\pi}\arcsin\frac{\mu}{\sigma_1\sigma_2}$$

If $\mu = 0$, then the RVs x and y are independent, hence,

$$E\{|x\,y|\}\Big|_{\mu=0} = E\{|x|\}E\{|y|\} = \frac{2}{\pi}\sigma_1\sigma_2$$

[see (5-74)]. Integrating (i) and using the above, we obtain

$$E\{|x\,y|\} = \frac{2}{\pi}\int_0^{\mu}\arcsin\frac{C}{\sigma_1\sigma_2}\,dC + \frac{2}{\pi}\sigma_1\sigma_2 = \frac{2\sigma_1\sigma_2}{\pi}(\cos\alpha + \alpha\sin\alpha)$$

6-70 From Example 6-41

$$f(y|x) : N\left(\eta_2 + \frac{r\sigma_2}{\sigma_1}x;\ \sigma_2\sqrt{1-r^2}\right) = N(4+x;\ \sqrt{3})$$

$$f(x|y) : N\left(\eta_1 + \frac{r\sigma_1}{\sigma_2}y;\ \sigma_1\sqrt{1-r^2}\right) = N\left(3+\frac{y}{4};\ \sqrt{3}/2\right)$$

6-71 The mass density in the square $|x|\le 1$, $|y|\le 1$ of the xy plane equals 1/4; hence, $P\{r\le 1\}=\pi/4$ and $P\{r\le r\} = \pi r^2/4$ for $r<1$. This yields

$$P\{r\le r, r\le 1\} = \begin{cases} P\{r\le r\} = \pi r^2/4 & r\le 1 \\ P\{r\le 1\} = \pi/4 & r>1 \end{cases}$$

$$F_r(r|M) = \frac{P\{r\le r, M\}}{P(M)} = \begin{cases} r^2 & r\le 1 \\ 1 & r>1 \end{cases} \qquad f_r(r|m) = \begin{cases} 2r, & r<+1 \\ 0 & \text{otherwise} \end{cases}$$

6-72 $\underset{\sim}{z} = \underset{\sim}{x} + \underset{\sim}{y} \qquad \underset{\sim}{w} = \underset{\sim}{x} \qquad f_{xz}(x,z) = f_{xy}(x, z-x)$

If $f_{xy}(x,y) = f_x(x)f_y(y)$, then

$$f_z(z|x) = \frac{f_{xz}(x,z)}{f_x(x)} = f_y(z-x)$$

6-73 The system $z = F_x(x) \qquad w = F_y(y|x)$ has a solution only if $z \le z \le 1$ and $0 \le w \le 1$. Furthermore,

$$\frac{\partial z}{\partial x} = f_x(x) \qquad \frac{\partial z}{\partial y} = 0$$

$$\frac{\partial w}{\partial x} \qquad \frac{\partial w}{\partial y} = f_y(y|x) \qquad J = f_x(x)f_y(y|x)$$

$$f_{zw}(z,w) = \frac{f_{xy}(x,y)}{f_x(x)f_y(y|x)} = 1 \quad \text{for} \quad 0 \le z,w \le 1$$

6-74 We introduce the events C_r = {we selected the rth coin} and A_k = {heads in a specific order}. From the assumptions it follows that

$$P(C_r) = \frac{1}{m} \qquad P(A_k|C_r) = p_r{}^k(1-p_r)^{n-k}$$

We wish to find the probability $P(C_r|A_k)$. The events C_r form a pertition; hence,

$$P(C_r|A_k) - \frac{\frac{1}{m}P(a_k|C_r)}{\frac{1}{m}\sum_{i-1}^{m} P(A_k|C_i)}$$

6-75 We wish to show that

$$E\{\underset{\sim}{x}^2\} = \frac{n}{n-1}$$

From page 207: $\underset{\sim}{x}^2 = n\underset{\sim}{y}^2/\underset{\sim}{z}$ where $\underset{\sim}{y}$ is N(0,1) and $\underset{\sim}{z}$ is $\chi^2(n)$. Hence, $E\{\underset{\sim}{y}^2\} = 1$ and (also (4-35) and (4-39))

$$E\left\{\frac{1}{\underset{\sim}{z}}\right\} = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty z^{n/2-2}e^{-z/2}dz = \frac{2^{m/2-1}\Gamma(n/2-1)}{2^{n/2}\Gamma(n/2)}$$

From this and the independence of $\underset{\sim}{y}$ and $\underset{\sim}{z}$ it follows that

$$E\{\underset{\sim}{x}^2\} = n\, E\{\underset{\sim}{y}^2\}\, E\left\{\frac{1}{\underset{\sim}{z}}\right\} = \frac{n}{n-2}$$

6-76 From (6-222):

$$R_x(x) = \exp\left\{-\int_0^x \beta_x(t)dt\right\} = \exp\left\{-k\int_0^x \beta_y(t)dt\right\} = R_y^k(t)$$

6-77 From (5-89) it follows with $\underset{\sim}{x} = |\underset{\sim}{z}|^2$ and $\alpha = \varepsilon^2$ that

$$E\{|\underset{\sim}{z}|^2 > \varepsilon^2\} \le \frac{E\{|\underset{\sim}{z}|^2\}}{\varepsilon^2}$$

for any $\underset{\sim}{z}$. And the result follows with $\underset{\sim}{z} = \underset{\sim}{x} - \underset{\sim}{y}$.

6-78 $$E\{U(a-\underset{\sim}{x})\} = \int_{-\infty}^{\infty} U(a-x)f(x)dx = \int_{-\infty}^{a} f(x)dx = F_x(a)$$

$$E\{U(b-\underset{\sim}{y})\} = F_y(b)$$

$$E\{U(a-\underset{\sim}{x})U(b-\underset{\sim}{y}) = \int_{-\infty}^{a}\int_{-\infty}^{b} f(x,y)dxdy = F_{xy}(a,b)$$

Hence

$$F_{xy}(a,b) = F_x(a)F_y(b)$$

6-79 From Example 6-38

$$E\{\underset{\sim}{y}|\underset{\sim}{x} \le 0\} = \int_{-\infty}^{\infty} y\ f_y(y|\underset{\sim}{x} \le 0)dy = \frac{1}{F_x(0)} \int_{-\infty}^{\infty} y \frac{\partial F(0,y)}{\partial y}\, dy$$

From (7-41) and (7-57)

$$\int_{-\infty}^{\infty} E\{\underset{\sim}{y}|x\}f_x(x)dx = \int_{-\infty}^{\infty} y \int_{-\infty}^{0} f(x,y)dxdy = \int_{-\infty}^{\infty} y \frac{\partial F(0,y)}{\partial y}\, dy$$

CHAPTER 7

7-1 $$0 \le P\{x_1 < \underset{\sim}{x} \le x_2,\ y_1 < \underset{\sim}{y} \le y_2,\ z_1 < \underset{\sim}{z} \le z_2\} =$$

$$= P\{\underset{\sim}{x} \le x_2,\ y_1 < \underset{\sim}{y} \le y_2,\ z_1 < \underset{\sim}{z} \le z_2\} - P\{\underset{\sim}{x} \le x_1,\ y_1 < \underset{\sim}{y} \le y_2,\ z_1 < \underset{\sim}{z} \le z_2\} =$$

$$= P\{\underset{\sim}{x} \le x_2,\ \underset{\sim}{y} \le y_2,\ z_1 < \underset{\sim}{z} \le z_2\} - P\{\underset{\sim}{x} \le x_2,\ \underset{\sim}{y} \le y_1,\ z_1 < \underset{\sim}{z} \le z_2\}$$

$$- P\{\underset{\sim}{x} \le x_1,\ \underset{\sim}{y} \le y_2,\ z_1 < \underset{\sim}{z} \le z_2\} + P\{\underset{\sim}{x} \le x_1,\ \underset{\sim}{y} \le y_1,\ z_1 < \underset{\sim}{z} \le z_2\} =$$

$$= P\{\underset{\sim}{x} \le x_2,\ \underset{\sim}{y} \le y_2,\ \underset{\sim}{z} \le z_2\} - P\{\underset{\sim}{x} \le x_2,\ \underset{\sim}{y} \le y_2,\ \underset{\sim}{z} \le z_1\}$$

$$- P\{\underset{\sim}{x} \le x_2,\ \underset{\sim}{y} \le y_1,\ \underset{\sim}{z} \le z_2\} + P\{\underset{\sim}{x} \le x_2,\ \underset{\sim}{y} \le y_1,\ \underset{\sim}{z} \le z_1\}$$

$$- P\{\underset{\sim}{x} \le x_1,\ \underset{\sim}{y} \le y_2,\ \underset{\sim}{z} \le z_2\} + P\{\underset{\sim}{x} \le x_1,\ \underset{\sim}{y} \le y_2,\ \underset{\sim}{z} \le z_1\}$$

$$+ P\{\underset{\sim}{x} \le x_1,\ \underset{\sim}{y} \le y_1,\ \underset{\sim}{z} \le z_2\} - P\{\underset{\sim}{x} \le x_1,\ \underset{\sim}{y} \le y_1,\ \underset{\sim}{z} \le z_1\}$$

7-2 $P\{\underset{\sim}{x}_A = 1,\ \underset{\sim}{x}_B = 1,\ \underset{\sim}{x}_C = 1\} = P(ABC) = 1/4$

$P\{\underset{\sim}{x}_A = 1\} = P(A) = 1/2$ $\qquad$ $P\{\underset{\sim}{x}_B = 1\} = P(B) = 1/2$

$P\{\underset{\sim}{x}_C = 1\} = P(C) = 1/2$ hence

$$P\{\underset{\sim}{x}_A = 1,\ \underset{\sim}{x}_B = 1,\ \underset{\sim}{x}_C = 1\} \ne P\{\underset{\sim}{x}_A = 1\}P\{\underset{\sim}{x}_B = 1\}P\{\underset{\sim}{x}_C = 1\}$$

hence $\underset{\sim}{x}_A$, $\underset{\sim}{x}_B$, $\underset{\sim}{x}_C$ are not independent. But

$P\{\underset{\sim}{x}_A = 1,\ \underset{\sim}{x}_B = 1\} = P(AB) = 1/4 = P\{\underset{\sim}{x}_A = 1\}P\{\underset{\sim}{x}_B = 1\}$

Similarly for any other combination, e.g.,
since $P(A) = P(AB) + P(A\bar{B})$, we conclude that
$P(A\bar{B}) = 1/2 - 1/4 = 1/4$ $\qquad$ $P(\bar{B}) = 1 - P(B) = 1/2$
$P\{\underset{\sim}{x}_A = 1,\ \underset{\sim}{x}_B = 0\} = P(A\bar{B}) = 1/4$
$P\{\underset{\sim}{x}_B = 0\} = P(\bar{B}) = 1/2$ hence
$P\{\underset{\sim}{x}_A = 1,\ \underset{\sim}{x}_B = 0\} = P\{\underset{\sim}{x}_A = 1\}P\{\underset{\sim}{x}_B = 0\}$

7-3 If x, y, z are independent in pairs, then

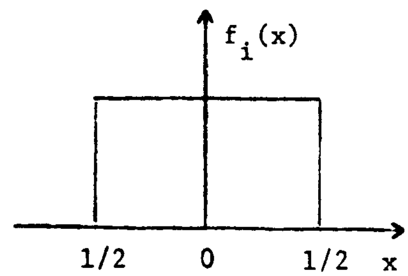
$$r_{xy} = r_{xz} = r_{yz} = 0$$

and (7-60) yields (we assume $\eta_x = \eta_y = \eta_z = 0$)

$$\Phi(\omega_1, \omega_2, \omega_3) = \exp\left\{-\frac{1}{2}\left(\sigma_1^2\omega_1^2 + \sigma_2^2\omega_2^2 + \sigma_3^2\omega_3^2\right)\right\}$$

$$f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3)$$

7-4 $x = x_1 + x_2 + x_3$. To determine $E\{x^4\}$ we shall use char. functions



$$\Phi_i(\omega) = \int_{-1/2}^{1/2} e^{j\omega x} dx = \frac{2\sin(\omega/2)}{\omega}$$

$$\Phi(\omega) = \left[\frac{2\sin(\omega/2)}{\omega}\right]^3 = \left(1 - \frac{\omega^2}{24} + \frac{\omega^4}{1920} - \cdots\right)^3$$

The coefficient of ω^4 in this expansion equals

$$\frac{13}{1920} \text{ hence } \frac{1}{4!}\ \frac{d^4\Phi(0)}{d\omega^4} = \frac{13}{1920}$$

and [see (5-103)]

$$E\{x^4\} = m_4 = \frac{13 \times 4!}{1920} = \frac{13}{80}$$

7-5 (a) The joint density $f(x,y)$ has circular symmetry because

$$f(x,y) = \int_{-\infty}^{\infty} f(\sqrt{x^2+y^2+z^2})\,dz$$

depends only on $x^2 + y^2$. The same holds for $f(x,z)$ and $f(y,z)$. And since the RVs $\underset{\sim}{x}, \underset{\sim}{y}$, and $\underset{\sim}{z}$ are independent, they must be normal [see (6-29)].

(b) From (a) it follows that the RVs $\underset{\sim}{v}_x, \underset{\sim}{v}_y, \underset{\sim}{v}_z$ are $N(0; \sqrt{kT/m})$.

With $\sigma^2 = kT/m$ and $n = 3$ it follows from (7-62)-(7-63) and (5-25) that

$$f_v(v) = \sqrt{\frac{2m^3}{\pi k^3 T^3}}\; v^2\, e^{-mv^2/2kT}\, U(v)$$

$$E\{\underset{\sim}{v}\} = 2\sqrt{\frac{2kT}{\pi m}} \qquad E\{\underset{\sim}{v}^{2n}\} = 1\times 3\cdots(2n+1)\left(\frac{kT}{m}\right)^n$$

7-6 From Prob. 6-52: $\underset{\sim}{y} = a\underset{\sim}{x} + b$, $\underset{\sim}{z} = c\underset{\sim}{y} + d$, hence,

$$\underset{\sim}{z} = A\underset{\sim}{x} + B \qquad \eta_z = A\eta_x + B \qquad \sigma_z = A\sigma_x$$

$$E\{(\underset{\sim}{z} - \eta_z)(\underset{\sim}{x} - \eta_x)\} = E\{A(\underset{\sim}{x} - \eta_x)(\underset{\sim}{x} - \eta_x)\} = A\sigma_x^2 = \sigma_x\sigma_z$$

7-7 It follows from (6-241) with $g_1(x) = x$, $g_2(y) = y$ if we replace all densities with conditional densities assuming $\underset{\sim}{x}_3$.

7-8 Reasoning as in (7-82), we conclude that

$E\{[\mathbf{y} - (a_1\mathbf{x}_1 + a_2\mathbf{x}_1)]^2\}$ is minimum if

$$E\{[\mathbf{y} - (a_1\mathbf{x}_1 + a_2\mathbf{x}_2)]\mathbf{x}_i\} = 0 \qquad i = 1,2$$

With $R_{0i} = E\{\mathbf{y}\,\mathbf{x}_i\}$, $R_{ij} = E\{\mathbf{x}_i\mathbf{x}_j\}$, the above yields

$$R_{01} = a_1R_{11} + a_2R_{12} \qquad\qquad R_{02} = a_1R_{12} + a_2R_{22}$$

But $\hat{E}\{\mathbf{y}|\mathbf{x}_1\} = A\,\mathbf{x}_1 \qquad\qquad A = R_{01}/R_{11} = a_1 + a_2R_{12}/R_{11}$

$$\hat{E}\{\hat{E}\{\mathbf{y}|\mathbf{x}_1,\mathbf{x}_2\}|\mathbf{x}_1\} = \hat{E}\{a_1\mathbf{x}_1 + a_2\mathbf{x}_2|\mathbf{x}_1\}$$

$$= a_1\mathbf{x}_1 + a_2\hat{E}\{\mathbf{x}_2|\mathbf{x}_1\} = \left(a_1 + a_2\frac{R_{12}}{R_{11}}\right)\mathbf{x}_1 = A\,\mathbf{x}_1$$

7-9 As in Probl. 6-51

$$E^2\{\mathbf{x}_i\mathbf{x}_j\} \le E^2\{\mathbf{x}_i\}E^2\{\mathbf{x}_j\} = M^2 \qquad\qquad |E\{\mathbf{x}_i\mathbf{x}_j\}| \le M$$

$$E\{\mathbf{s}^2|\mathbf{n} = n\} = E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{x}_i\mathbf{x}_j\right\} \le Mn^2$$

Hence [see (6-240)]

$$E\{\mathbf{s}^2\} = E\{E\{\mathbf{s}^2|\mathbf{n}\}\} < E\{M\mathbf{n}^2\}$$

7-10 As we know,

$$1 + x + \cdots + x^n + \cdots = \frac{1}{1-x} \qquad |x| < 1$$

Differentiating, we obtain

$$1 + 2x + \cdots + n\,x^{n-1} + \cdots = \sum_{k=1}^{\infty} k\,x^{k-1} = \frac{1}{(1-x)^2} \qquad \text{(i)}$$

The RV $\underset{\sim}{x}_1$ equals the number of tosses until heads shows for the first time, Hence, $\underset{\sim}{x}_1$ takes the values 1,2,... with $P\{\underset{\sim}{x}_1 = k\} = pq^{k-1}$. Hence, [see (3-12) and (i)]

$$E\{\underset{\sim}{x}_1\} = \sum_{k=1}^{\infty} k\,P\{\underset{\sim}{x}_1 = k\} = \sum_{k=1}^{\infty} k\,p\,q^{k-1} = \frac{p}{(1-q)^2} = \frac{1}{p}$$

Starting the count after the first head shows, we conclude that the RV $\underset{\sim}{x}_2 - \underset{\sim}{x}_1$ has the same statistics as the RV $\underset{\sim}{x}_1$. Hence,

$$E\{\underset{\sim}{x}_2 - \underset{\sim}{x}_1\} = E\{\underset{\sim}{x}_1\} \qquad E\{\underset{\sim}{x}_2\} = 2E\{\underset{\sim}{x}_1\} = \frac{2}{p}$$

Reasoning similarly, we conclude that

$E\{\underset{\sim}{x}_n - \underset{\sim}{x}_{n-1}\} = E\{\underset{\sim}{x}_1\}$. Hence (induction)

$$E\{\underset{\sim}{x}_n\} = E\{\underset{\sim}{x}_{n-1}\} + E\{\underset{\sim}{x}_1\} = \frac{n-1}{p} + \frac{1}{p} = \frac{n}{p}$$

7-11 If n accidents occur in a day, the probability that m of them will be fatal equals $\binom{n}{m} p^m q^{n-m}$ for $m \le n$ and zero for $m > n$. Hence,

$$P\{\underset{\sim}{m} = m \mid \underset{\sim}{n} = n\} = \begin{cases} 0 & m > n \\ \binom{n}{m} p^m q^{n-m} & m \le n \end{cases}$$

This yields

$$E\{e^{j\omega \underset{\sim}{m}} \mid \underset{\sim}{n} = n\} = \sum_{m=0}^{n} e^{j\omega m} \binom{n}{m} p^m q^{n-m} = (p\,e^{j\omega} + q)^n$$

But

$$P\{\underset{\sim}{n} = n\} = e^{-a}\frac{a^n}{n!} \qquad n = 0,1,\ldots$$

Hence,

$$E\{e^{j\omega \underset{\sim}{m}}\} = E\{E\{e^{j\omega \underset{\sim}{m}} \mid \underset{\sim}{n}\}\} = E\{(p\, e^{j\omega} + q)^{\underset{\sim}{n}}\}$$

$$\sum_{n=0}^{\infty} (p\, e^{j\omega} + q)^n\, e^{-a} \frac{a^n}{n!} = e^{a(p\, e^{j\omega} + q)} e^{-a} = e^{a\, p\, (e^{j\omega} - 1)}$$

This shows that the RV $\underset{\sim}{m}$ is Poisson distributed with parameter $a\, p$ [see (5-119)].

7-12 We shall determine first the conditional distribution

$$F_s(s \mid \underset{\sim}{n} = n) = \frac{P\{\underset{\sim}{s} \le s,\ \underset{\sim}{n} = n\}}{P\{\underset{\sim}{n} = n\}}$$

The event $\{\underset{\sim}{s} < s,\ \underset{\sim}{n} = n\}$ consists of all outcomes such that $\underset{\sim}{n} = n$ and $\sum_{k=1}^{n} \underset{\sim}{x}_k \le s$. Since the RV $\underset{\sim}{n}$ is independent of the RVs $\underset{\sim}{x}_k$, this yields

$$F_s(s \mid \underset{\sim}{n} = n) = P\{\sum_{k=1}^{n} \underset{\sim}{x}_k \le s\} P\{\underset{\sim}{n} = n\} / P\{\underset{\sim}{n} = n\}$$

From the above and the independence of the RVs $\underset{\sim}{x}_k$ it follows that [see (7-51)]

$$f_s(s \mid \underset{\sim}{n} = n) = f_1(s) * f_2(s) * \cdots * f_n(s)$$

Setting $A_k = \{\underset{\sim}{n} = k\}$ in (4-74), we obtain

$$f_s(s) = \sum_k p_k\, [f_1(s) * \cdots * f_k(s)]$$

7-13 From the independence of the RVs n and x_i it follows that

$$E\{e^{sy} \mid n = k\} = E\{e^{s(x_1 + \cdots + x_k)}\}$$

$$= E\{e^{sx_1}\} \cdots E\{e^{sx_k}\} = \Phi_x^k(s)$$

Hence,

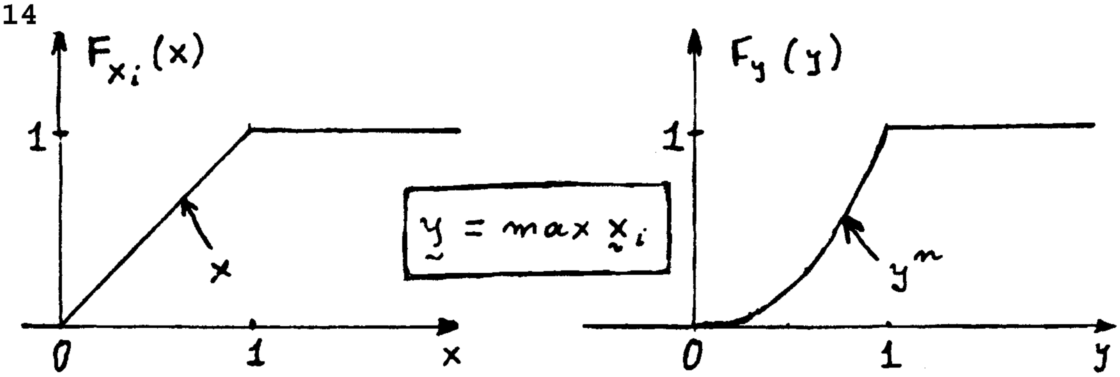
$$\Phi_y(s) = E\{e^{sy}\} = E\{E\{e^{sy} \mid n\}\} = E\{\Phi_x^n(s)\}$$

$$= \Gamma_n[\Phi_x(s)] \quad \text{because} \quad E\{z^n\} = \Gamma_n(z)$$

Special case. If n is Poisson with parameter a, then [see (5-119)]

$$\Gamma_n(z) = e^{az - a} \qquad \Phi_y(s) = e^{a\Phi_x(s) - a}$$

7-14



$$\{y \le y\} = \{x_1 \le y, x_2 \le y, \ldots, x_n \le y\}$$

From the independence of x_i and the abcve it follows that

$$F_y(y) = P\{y \le y\} = P\{x_1 \le y\} \cdots P\{x_n \le y\}$$

$$= F_1(y) \cdots F_n(y)$$

where $F_i(y) = y$ for $0 \le y \le 1$.

7-15 The RV $\underset{\sim}{x}$ is defined in the space S. The set

$$C = \{z < \underset{\sim}{z} \le z + dz,\ w < \underset{\sim}{w} \le w + dz\} \qquad z > w$$

is an event in the space S_n of repeated trials and its probability equals

$$P(C) = f_{zw}(z,w)dzdw$$

We introduce the events

$$D_1 = \{\underset{\sim}{x} \le w\} \qquad D_2 = \{w < \underset{\sim}{x} \le w + dw\} \qquad D_3 = \{w + dw < \underset{\sim}{x} \le z\}$$

$$D_4 = \{z < \underset{\sim}{x} \le z + dz\} \qquad D_5 = \{z + dz < \underset{\sim}{x}\}$$

These events form a partition of S and their probabilities $p_i = P(D_i)$ equal

$$F_x(w) \qquad f_x(w)dw \qquad F_x(z)-F_x(w+dw) \qquad f_z(z)dz \qquad 1-F_x(z+dz)$$

respectively. The event C occurs iff the smallest of the RVs $\underset{\sim}{x}_i$ is in the interval (w, w+dw), the largest is in the interval (z, z+dz), and, consequently, all others are between w+dw and z. This is the case iff D_1 does not occur at all, D_2 occurs once, D_3 occurs n-2 times, D_4 occurs once, and D_5 does not occur at all. With

$$k_1=0 \qquad k_2=1 \qquad k_3=n-2 \qquad k_4=1 \qquad k_5=0$$

it follows from (4-102) that

$$P(C) = \frac{n!}{(n-2)!}p_2\, p_3^{n-2}\, p_4 = n(n-1)f_x(w)dw\,[f_x(z)-F_x(w+dw)]^{n-1}f_x(z)dz$$

for z > w, and 0 otherwise.

7-16 If $\underset{\sim}{z}$ is $N(\eta,1)$ then

$$E\{e^{s\underset{\sim}{z}^2}\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{sz}e^{-(z-\eta)^2/2}dz$$

$$sz^2 - \frac{(z-\eta)^2}{2} = \left(s - \frac{1}{2}\right)\left(z - \frac{\eta}{1-2s}\right)^2 + \frac{\eta^2 s}{1-1s}$$

Since

$$\frac{1}{\sqrt{2\pi}}\int_{-\eta}^{\infty} e^{-a(z-b)^2} dz = \frac{1}{\sqrt{2a}}$$

the above yields

$$E\{e^{sz^2}\} = \frac{1}{\sqrt{2(1/2-S)}} \exp\left\{\frac{\eta^2 S}{1-2S}\right\}$$

$$\Phi_w(s) = \frac{1}{\sqrt{1-2s}} \exp\left\{\frac{\eta_1 s}{1-2s}\right\} \cdots \frac{1}{\sqrt{1-2s}} \exp\left\{\frac{\eta_n s}{1-2s}\right\}$$

7-17 We wish to show that the RVs

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

are independent. Since s^2 is a function of the n RVs $x_i - \bar{x}$, it suffices to show that each of these RVs is independent of $\bar{x}$. We assume for simplicity that $E\{x_i\}=0$. Clearly,

$$E\{x_i \bar{x}\} = \frac{1}{n} E\{x_i^2\} = \frac{\sigma^2}{n} \qquad E\{\bar{x}\bar{x}\} = \frac{1}{n^2}\sum_{i=1}^{n} x_i^2 = \frac{\sigma^2}{n}$$

because $E\{x_i x_j\}=0$ for $i \neq i$. Hence,

$$E\{(x_i - \bar{x})\bar{x}\} = 0$$

Thus, the RVs $x_i - \bar{x}$ and $\bar{x}$ are orthogonal; and since they are jointly normal, they are independent.

7-18 Since $\eta_s = \alpha_0 + \alpha_1\eta_1 + \alpha_2\eta_2$ [see (7-87)], the mean of the error

$$\underset{\sim}{\varepsilon} = \underset{\sim}{s} - (\alpha_0 + \alpha_1 \underset{\sim}{x}_1 + \alpha_2 \underset{\sim}{x}_2) = (\underset{\sim}{s} - \eta_s) - [\alpha_1(x_1 - \eta_1) + \alpha_2(\underset{\sim}{x}_2 - \eta_2)]$$

is zero. Furthermore, $\underset{\sim}{\varepsilon}$ is orthogonal to $\underset{\sim}{x}_i$, hence, it is also orthogonal to $\underset{\sim}{x}_i - \eta_i$.

7-19 From the orthogonality principle:

$$\hat{E}\{\underset{\sim}{y} | \underset{\sim}{x}_1, \underset{\sim}{x}_2\} = a_1 \underset{\sim}{x}_1 + a_2 \underset{\sim}{x}_2 \qquad \underset{\sim}{y} - \{a_1 x_1 + a_2 \underset{\sim}{x}_2\} \perp \underset{\sim}{x}_1, x_2$$

$$\hat{E}\{y | x_1\} = A \underset{\sim}{x}_1 \qquad y - A \underset{\sim}{x}_1 \perp \underset{\sim}{x}_1$$

Hence

$$\underset{\sim}{y} - (a_1 \underset{\sim}{x}_1 + a_2 \underset{\sim}{x}_2) - (\underset{\sim}{y} - A x_1) = a_1 \underset{\sim}{x}_1 + a_2 \underset{\sim}{x}_2 - A \underset{\sim}{x}_1 \perp \underset{\sim}{x}_1$$

From this it follows that

$$\hat{E}\{a_1 x_1 + a_2 \underset{\sim}{x}_2 | \underset{\sim}{x}_1\} = A \underset{\sim}{x}_1$$

$$\hat{E}\{\hat{E}\{\underset{\sim}{y} | \underset{\sim}{x}_1, \underset{\sim}{x}_2\} | \underset{\sim}{x}_1\} = \hat{E}\{\underset{\sim}{y} | \underset{\sim}{x}_1\}$$

7-20 The event $\{\underline{x} \le x\}$ occurs if there is at least one point in the interval $(0,x)$; the event $\{\underline{y} \le y\}$ occurs if all the points are in the interval $(0,y)$:

$$A_x = \{\text{at least one point in } (0,x)\} = \{\underline{x} \le x\}$$

$$B_y = \{\text{no points in } (y,1)\}$$

$$= \{\text{all points in } (0,y)\} = \{\underline{y} \le y\}$$

Hence, for $0 \le x \le 1, \quad 0 \le y \le 1$

$$F_x(x) = P(A_x) = 1 - P(\bar{A}_x) = 1 - (1 - x)^n$$

$$F_y(y) = P(B_y) = y^n$$

Furthermore,

$$\{\underline{x} \le x,\ \underline{y} \le y\} = A_x B_y \qquad\qquad A_x B_y + \bar{A}_x B_y = B_y$$

If $x \le y$ then

$$\bar{A}_x B_y = \{\text{all points in } (x,y)\}$$

$$P(\bar{A}_x B_y) = (y - x)^n$$

If $x > y$, then $\bar{A}_x B_y = \{\emptyset\}$. Hence

$$F_{xy}(x,y) = P(A_x B_y) = \begin{cases} y^n - (y - x)^n & x \le y \\ y^n & x > y \end{cases}$$

7-21 Suppose that $E\{\underset{\sim}{x}_i\} = 0$, $E\{\underset{\sim}{x}_i^2\} = \sigma^2$, $E\{\underset{\sim}{x}_i^4\} = \mu_4$

If $\underset{\sim}{A} = \sum_{i=1}^{n} \underset{\sim}{x}_i^2$, then $E\{\underset{\sim}{A}\} = n\sigma^2$

$$E\{\underset{\sim}{A}^2\} = \sum_{i,j=1}^{n} E\{\underset{\sim}{x}_i^2\underset{\sim}{x}_j^2\} = n\mu_4 + (n^2 - n)\sigma^4$$

because

$$E\{\underset{\sim}{x}_i^2\underset{\sim}{x}_j^2\} = \begin{cases} \mu_4 & i = j \\ \sigma^4 & i \neq j \end{cases}$$

Furthermore

$$E\{\bar{\underset{\sim}{x}}^2\underset{\sim}{x}_j^2\} = \frac{1}{n^2} E\left\{\left(\sum_{i=1}^{n} \underset{\sim}{x}_i\right)^2 \underset{\sim}{x}_j^2\right\} = \frac{1}{n^2} [\mu_4 + (n - 1)\sigma^4]$$

$$E\{\bar{\underset{\sim}{x}}^2\underset{\sim}{A}\} = \frac{1}{n} [\mu_4 + (n - 1)\sigma^4]$$

$$E\{\bar{\underset{\sim}{x}}^4\} = \frac{1}{n^4} E\left\{\left(\sum_{i=1}^{n} \underset{\sim}{x}_i\right)^4\right\} = \frac{1}{n^4} [n\mu_4 + 3n(n - 1)\sigma^4]$$

because

$$E\{\underset{\sim}{x}_i\underset{\sim}{x}_j\underset{\sim}{x}_k\underset{\sim}{x}_r\} = \begin{cases} \mu_4 & i = j = k = r & \text{[n such terms]} \\ \sigma^4 & i = j \neq k = r & \text{[3n(n-1) such terms]} \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $(n - 1)\, \bar{\underset{\sim}{V}} = \sum_{i=1}^{n} (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})^2 = \underset{\sim}{A} - n\bar{\underset{\sim}{x}}^2$, $E\{\bar{\underset{\sim}{V}}\} = \sigma^2$. Hence

$$(n - 1)^2E\{\underset{\sim}{V}^2\} = E\{\underset{\sim}{A}^2\} - 2nE\{\bar{\underset{\sim}{x}}^2\underset{\sim}{A}\} + n^2E\{\bar{\underset{\sim}{x}}^4\}$$

$$= n\mu_4 + (n^2 - n)\sigma^4 - 2[\mu_4 + (n - 1)\sigma^4] + \frac{1}{n}[\mu_4 + 3(n - 1)\sigma^4]$$

This yields

$$E\{\bar{\underset{\sim}{V}}^2\} = \frac{\mu_4}{n} + \frac{n^2 - 2n + 3}{n(n - 1)}\, \sigma^4 = \sigma^4 + \sigma_{\bar{V}}^2$$

<u>Note</u> If the RVs $\underset{\sim}{x}_i$ are $N(0,\sigma^2)$, then $\mu_4 = 3\sigma^4$

$$\sigma_{\bar{V}}^2 = \frac{1}{n} \left(3\sigma^4 - \frac{n - 3}{n - 1}\, \sigma^4\right) = \frac{2}{n - 1}\, \sigma^4$$

7-22 From Prob. 6-49:

$$E\{|\underset{\sim}{x}_{2i} - \underset{\sim}{x}_{2i-1}|\} = \frac{2\sigma}{\sqrt{\pi}} \qquad E\{|\underset{\sim}{x}_{2i} - \underset{\sim}{x}_{2i-1}|^2\} = 2\sigma^2$$

Hence,

$$E\{|\underset{\sim}{x}_{2i} - \underset{\sim}{x}_{2i-1}||\underset{\sim}{x}_{2j} - \underset{\sim}{x}_{2j-1}|\} = \begin{cases} 2\sigma^2 & i = j \\ 4\sigma^2/\pi & i \neq j \end{cases}$$

$$E\{\underset{\sim}{z}\} = \frac{\sqrt{\pi}}{2n}\,\frac{2\sigma n}{\sqrt{\pi}} = \sigma$$

$$E\{\underset{\sim}{z}^2\} = \frac{\pi}{4n^2}\,[2n\sigma^2 + \frac{4\sigma^2}{\pi}(n^2 - n)]$$

$$\sigma_z^2 = \frac{\pi}{2n}\sigma^2 + (1 - \frac{1}{n})\sigma^2 - \sigma^2 = \frac{\pi - 2}{2n}\,\sigma^2$$

7-23 If $R^{-1} = \begin{bmatrix} a_{11} \cdots a_{1n} \\ \\ a_{n1} \cdots a_{nn} \end{bmatrix}$ then $\sum_j a_{ij}R_{ji} = 1$

Hence,

$$E\{\underset{\sim}{X}R^{-1}\underset{\sim}{X}^t\} = E\{\sum_{i=1}^{n}\sum_{j=1}^{n} \underset{\sim}{x}_i a_{ij} \underset{\sim}{x}_j\}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}R_{ji} = \sum_{i=1}^{n} 1 = n$$

7-24 The density $f_z(z)$ of the sum $\underset{\sim}{z} = \underset{\sim}{x}_1 + \cdots + \underset{\sim}{x}_n$ tends to a normal curve with variance $\sigma_1^2 + \cdots + \sigma_n^2 \to \infty$ as $n \to \infty$ (we assume $\sigma_i > c > 0$). Hence, $f_z(z)$ tends to a constant in any interval of length 2π. The result follows as in (5-37) and Prob. 5-20.

7-25 Since $a_n - a \to 0$, we conclude that

$$E\{(\underset{\sim}{x}_n - a)^2\} = E\{[(\underset{\sim}{x}_n - a_n) + (a_n - a)]^2\}$$

$$= E\{(\underset{\sim}{x}_n - a_n)^2\} + 2(a_n - a)E\{\underset{\sim}{x}_n - a_n\} + (a_n - a)^2 \to 0$$

as $n \to \infty$.

7-26 If $E\{\underset{\sim}{x}_n\underset{\sim}{x}_m\} \to a$ as $n,m \to \infty$, then, given $\varepsilon > 0$, we can find a number n_o such that

$$E\{\underset{\sim}{x}_n\underset{\sim}{x}_n\} = a + \theta(n,m) \qquad |\theta| < \varepsilon \qquad \text{if } n,m > 0$$

Hence,

$$E\{(\underset{\sim}{x}_n - \underset{\sim}{x}_m)^2\} = E\{\underset{\sim}{x}_n^2\} + E\{\underset{\sim}{x}_m^2\} - 2E\{\underset{\sim}{x}_n\underset{\sim}{x}_m\}$$

$$= a + \theta_1 + a + \theta_2 - 2(a + \theta_3) = \theta_1 + \theta - 2\theta_3$$

and since $|\theta_1 + \theta_2 - 2\theta_3| < 4\,\varepsilon$ for any ε, it follows that

$E\{(\underset{\sim}{x}_n - \underset{\sim}{x}_m)^2\} \to 0$, hence (Cauchy) $\underset{\sim}{x}_n$ tends to a limit.

<u>Conversely</u> If $\underset{\sim}{x}_n \to \underset{\sim}{x}$ in the MS sense, then

$E\{(\underset{\sim}{x}_n - \underset{\sim}{x})^2\} \to 0$. Furthermore,

$$E\{\underset{\sim}{x}_n^2\} \to E\{\underset{\sim}{x}^2\} \qquad E\{\underset{\sim}{x}\,\underset{\sim}{x}_n\} \to E\{\underset{\sim}{x}^2\}$$

because (see Prob. 6-51)

$$E^2\{\underset{\sim}{x}_n^2 - \underset{\sim}{x}^2\} = E^2\{(\underset{\sim}{x}_n - \underset{\sim}{x})(\underset{\sim}{x}_n + \underset{\sim}{x})\}$$

$$\le E\{(\underset{\sim}{x}_n - \underset{\sim}{x})^2\}E\{(\underset{\sim}{x}_n + \underset{\sim}{x})^2\} \to 0$$

$$E^2\{\underset{\sim}{x}(\underset{\sim}{x}_n - \underset{\sim}{x})\} \le E\{\underset{\sim}{x}^2\}E\{(\underset{\sim}{x}_n - \underset{\sim}{x})^2\} \to 0$$

Similarly, $E\{(\mathbf{x}_n - \mathbf{x})(\mathbf{x}_m - \mathbf{x})\} \to 0$. Hence,

$$E\{\mathbf{x}_n\mathbf{x}_m\} + E\{\mathbf{x}^2\} - E\{\mathbf{x}\,\mathbf{x}_n\} - E\{\mathbf{x}\,\mathbf{x}_m\} \to 0$$

Combining, we conclude that $E\{\mathbf{x}_n\mathbf{x}_m\} \to E\{\mathbf{x}^2\}$.

7-27 $$E\{\mathbf{x}_k\} = 0 \qquad\qquad E\{\mathbf{x}_k^2\} = \sigma_k^2$$

$$E\left\{\left(\sum_{k=n_1}^{n_2} \mathbf{x}_k\right)^2\right\} = \sum_{k=n_1}^{n_2} E\{\mathbf{x}_k^2\}$$

If $\sum_{k=1}^{\infty} \sigma_k^2 < \infty$, then given $\varepsilon > 0$, we can find n_o such that $\sum_{k=n+1}^{n+m} \sigma_k^2 < \varepsilon$ for any m and $n > n_o$. Thus

$$E\{(\mathbf{y}_{n+m} - \mathbf{y}_n)^2\} = E\left\{\left(\sum_{k=n+1}^{n+m} \mathbf{x}_k\right)^2\right\} = \sum_{k=n+1}^{n+m} \sigma_k^2 < \varepsilon$$

This shows that (Cauchy), $\mathbf{y}_k$ converges in the MS sense. The proof of the converse is similar.

7-28 If $f_1(x) = c\,e^{-cx}U(x)$ then $\Phi_1(s) = \dfrac{c}{c-s}$

$$\Phi(s) = \Phi_1(s)\cdots\Phi_n(s) = \frac{c^n}{(c-s)^n}$$

Hence (see Example 5-29) $f(x) = \dfrac{c^n x^{n-1}}{(n-1)!}\,e^{-cx}\,U(x)$

7-29 From Prob. 7-28 it follows that $f(x)$ is the density of the sum $\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_n$. Furthermore,

$$E\{\mathbf{x}\} = \frac{n}{c} \qquad\qquad \sigma_x^2 = \frac{n}{c^2}$$

From the central limit theorem it follows, therefore, that for large n, the Erlang density is nearly equal to a normal curve with mean n/c and variance n/c^2.

7-30 $E\{r_i\} = 500 \qquad \sigma_i^2 = 50^2/3$

$$r = r_1 + r_2 + r_3 + r_4 \qquad E\{r\} = 2{,}000 \qquad \sigma_r^2 = 10^4/3$$

Thus, r is approximately $N(2000;10^2/\sqrt{3})$

$$P\{1900 \le r \le 2100\} = 2\,G\left(\frac{100\sqrt{3}}{100}\right) - 1 = 0.9169.$$

7-31 The RVs x_i are independent with (see Prob. 5-37)

$$f_i(x) = \frac{c_i}{\pi(c_i^2 + x^2)} \qquad \Phi_i(\omega) = e^{-c_i|\omega|}$$

In that case, (7-104) does not hold because

$$\int_{-\infty}^{\infty} x^\alpha\, f(x)\,dx = \frac{c_i}{\pi}\int_{-\infty}^{\infty} \frac{x^\alpha}{c_i^2 + x^2}\,dx = \infty \qquad \alpha > 2$$

In fact, the density of $x = x_1 + \cdots + x_n$ is Cauchy with parameter $c = c_1 + \cdots + c_n$ because

$$\Phi(\omega) = e^{-c_1|\omega|} \cdots e^{-c_n|\omega|} = e^{-(c_1 + \cdots + c_n)|\omega|}$$

7-32 In this problem, $\sigma_z^2 = E\{|z|^2\} = E\{x^2 + y^2\} = 2\sigma^2$

$$f_z(x) = f_x(x)f_y(y) = \frac{1}{2\pi\sigma^2}\,e^{-(x^2+y^2)/2\sigma^2} = \frac{1}{2\pi\sigma_z^2}\,e^{-|z|^2/\sigma_z^2}$$

$$\Phi_z(\Omega) = \Phi_x(u)\Phi_y(v) = \exp\left\{-\frac{1}{2}\sigma^2(u^2+v^2)\right\} = \exp\left\{-\frac{1}{4}\sigma_z^2|\Omega|^2\right\}$$

CHAPTER 8

8-1 (a) From (8-11) with γ=.95, u=.975, $z_{.975} \simeq 2$, σ=0.1, and n=9 we obtain

$$c = \frac{z_u\sigma}{\sqrt{n}} = 0.066$$

(b) From (8-11) with c=91.01-91=0.05mm:

$$z_u = \frac{c\sqrt{n}}{\sigma} = 1.5 \qquad u = .933 \qquad \gamma = .866$$

8-2 (a) From (8-11) with σ=1 and n=4: $\bar{x} \pm \sigma z_u/\sqrt{n} \simeq 203 \pm 1$mm

(b) From (8-12) with δ=.05: $c = \sigma/\sqrt{n\delta} = 2.236$mm

---

8-3 From (8-4)with γ=.9, u=.95: $\bar{x} \pm z_u\sigma/\sqrt{n} = 25{,}000 \pm 1{,}028$ miles

---

8-4 We wish to find n such that $P\{|\bar{x}-a|<0.2\} = 0.95$ where $a=E\{\bar{x}\}$.From (8-4) it follows with u=.975 and σ=0.1mm that

$$\frac{z_u\sigma}{\sqrt{n}} \le 0.2, \text{ hence, } n=1$$

8-5 In this problem, x is uniform with $E\{x\}=\theta$ and $\sigma^2=4/3$.We can use, however, the normal approximation for $\bar{x}$ because n=100. With γ=.95, (8-11) yields the interval

$$\bar{x} \pm z_{.975}\ \sigma\sqrt{n} = 30 \pm 0.227$$

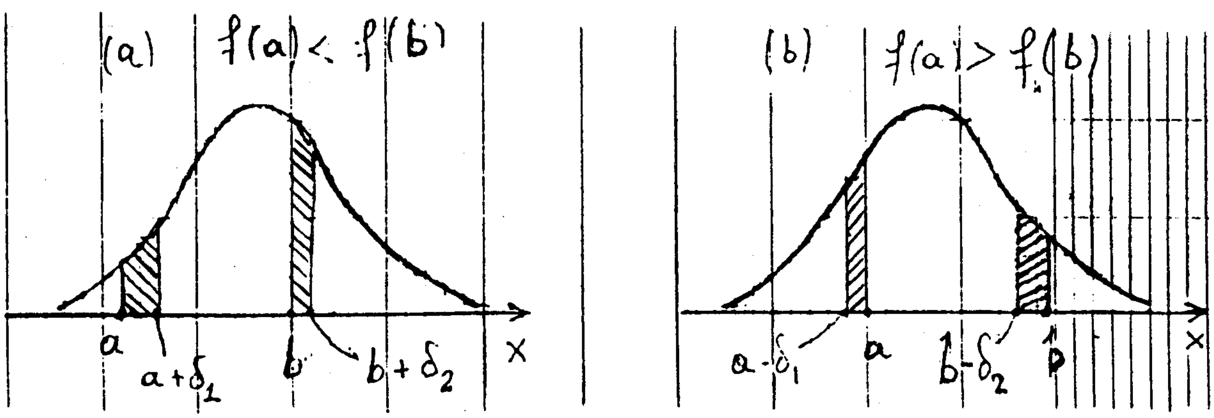
8-6 We shall show that if $f(x)$ is a density with a single maximum and $P\{a<x<b\}=\gamma$, then $b-a$ is minimum if $f(a)=f(b)$. The density $xe^{-x}U(x)$ is a special case. It suffices to show that $b-a$ is not minimum if $f(a)<f(b)$ or $f(a)>f(b)$.

Suppose first that $f(a)<f(b)$ as in figure (a). Clearly. $f'(a)>0$ and $f'(b)<0$, hence, we can find two constants $\delta_1>0$ and$\delta_2>0$ such that $P\{a+\delta_1 < x < b+\delta_2\} = \gamma$ and

$$f(a) < f(a+\delta_1) < f(b+\delta_2) < f(b)$$

From this it follows that $\delta_1 > \delta_2$, hence, the length of thenew interval $(a+\delta_1, b+\delta_2)$ is smaller than b-a.

If $f(a) > f(b)$, we form the interval $(a-\delta_1, b-\delta_2)$ (Fig. 8-6b) and proceed similarly.



Special case. If $f(x)=xe^{-x}$ then (see Problem 4-9) $F(x)=1-e^{-x}-xe^{-x}$, hence,

$$P\{a < x < b\} = e^{-a} + ae^{-a} - e^{-b} - be^{-b} = .95$$

And since $f(a)=f(b)$, the system

$$ae^{-a} = be^{-b} \qquad e^{-a} - e^{-b} = .95$$

results. Solving, we obtain $a \simeq 0.04 \quad b \simeq 5.75$.

A numerically simpler solution results if we set

$$0.025 = P\{x \le a\} = F(a) \qquad 0.025 = P\{x > b\} = 1 - F(b)$$

as in (9-5). This yields the system

$$0.025 = 1 - e^{-a} - ae^{-a} \qquad 0.025 = e^{-b} + be^{-b}$$

Solving, we obtain $a=0.242$, $b=5.572$. However, the length $5.572-0.242=5.33$ of the resulting interval is larger than the length $4.75-0.04=4.71$ of the optimum interval.

8-7 We start with the general problem: We observe the n samples x_i of an $N(\eta,10)$ RV x and we wish to predict the value x of x at a future trial in terms of the average $\bar{x}$ of the observations. If η is known, we have an ordinary prediction problem. If it is unknown, we must first estimate it. To do so, we form the RV $w=x-\bar{x}$. This RV is

$N(0,\sigma_w)$ where $\sigma_w^2=\sigma_x^2+\sigma_{\bar{x}}^2=\sigma^2+\sigma^2/n$. With $c=z_{.975}\sigma_w$$ it follows that $P\{|w| < c\}=.95$. Hence

$$P\{\bar{x} - c < x < \bar{x} + c\} = 0.95$$

For $n=20$ and $\sigma=10$ the above yields $\sigma_w=10.25$ and $c \simeq 20.5$. Thus, we can expect with .95 confidence coefficient that our bulb will last at least $80-20.5=59.5$ and at most $80+20=100.5$ hours.

8-8 The time of arrival of the 40th patient is the sum $x_1 + \cdots + x_n$ of $n=39$ RVs with exponential distribution. We shall estimate the mean $\eta=1/\theta$ of x in terms of its sample mean $\bar{x}=240/39=6.15$ minutes using two methods. The first is approximate (large n) and is based on (8-11).

<u>Normal approximation</u>. With $\lambda=\eta$ and$z_{.975}/\sqrt{39}=0.315$:

$$P\left\{\frac{\bar{x}}{1.315} < \eta < \frac{\bar{x}}{0.685}\right\} = .95 \qquad 4.68 < \eta < 8.98 \text{ minutes}$$

<u>Exact solution</u>. The RVs $\underset{\sim}{x}_i$ are i.i.d. with exponential distribution. From this and (7-52) it follows that their sum $\underset{\sim}{y}=\underset{\sim}{x}_1 + \cdots + \underset{\sim}{x}_n = n\underset{\sim}{\bar{x}}$ has an Erlang distribution:

$$\Phi_y(s) = \frac{\theta^n}{(\theta-s)^n} \qquad f_y(y) = \frac{\theta^n}{(n-1)!} y^{n-1} e^{-\theta y} U(y)$$

and the RV $\underset{\sim}{z}=2\theta\underset{\sim}{y} = 2n\theta\underset{\sim}{\bar{x}}$ has a $\chi^2(2n)$ distribution:

$$f_z(z) = \frac{1}{2\theta} f_y\left(\frac{z}{2\theta}\right) U(z) = \frac{z^{n-1}}{2^n(n-1)!} e^{-z/2} U(z)$$

Hence,

$$P\left\{ \chi^2_{\delta/2}(2n) < \frac{2n\bar{x}}{\eta} < \chi^2_{1-\delta/2}(2n) \right\} = \gamma = 1-\delta$$

Since $\chi^2_{.025}(78) = 54.6$, $\chi^2_{.975}(78)=104.4$, and $2n\bar{x}=480$, this yields the interval

$$4.60 < \eta < 8.79 \text{ minutes}$$

8-9 From (8-19) with $\bar{x}=2,550/200=12.75$ $n=200$ and $z_u \simeq 2$

$$\lambda^2 - 25.52\,\lambda + 12.75^2 = 0 \qquad \lambda_1 = 12.255 < \lambda < 13.265 = \lambda_2$$

8-10 From (8-21) with $\bar{x}=2,080/4000=0.52$, $n=4,000$ and $z_u \simeq 2.326$.

$$p_{1,2} \simeq \bar{x} \pm z_u \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} = .52 \pm .018$$

Hence, $.502 < p < .538$.

8-11 (a) In this problem, $\bar{x}=0.40$, $n=900$ and $z_u \simeq 2$. From (8-21): Margin of error

$$\pm\ 100\ z_u\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} = \pm\ 3.27\%$$

(b) We wish to find z_u. From (9-21) and Table 1a:

$$100 z_u\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} = 2 \qquad z_u = 1.225 \qquad u = .89$$

This yields the confidence coefficient $\gamma = 2u - 1 = .78$

8-12 From (8-21) with $\bar{x}=0.29$ and $z_u=2$:

$$z_u\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} = 0.04 \qquad\qquad n > \frac{\bar{x}(1-\bar{x})}{.04^2}\, z^2_u = 515$$

8-13 The problem is to find n such that [see (8-20)] $z_u\sqrt{\frac{p(1-p)}{n}} \le .02$ *for every p.* Since $z_u \simeq 2$ and $p(1-p) \le 1/4$, this is the case if

$$z_u\sqrt{1/4n} \le .02 \qquad n \ge 2{,}500$$

8-14 From (8-36) with k=1

$$f(p) = \begin{cases} 5 & .4 < p < .6 \\ 0 & \text{otherwise} \end{cases} \qquad P\{k=1\} = 5\int_{.4}^{.6} p\,dp = .5 = \frac{1}{\gamma}$$

$$f_p(p|1) = \begin{cases} 10p & .4 < p < .6 \\ 0 & \text{otherwise} \end{cases} \qquad \hat{p} = 10\int_{.4}^{.6} p^2\,dp = .5067$$

8-15 From Prob. 8-8: $f_{\bar{x}}(\bar{x}|\theta) = \frac{(\theta n)^n}{(n-1)!}\bar{x}^{n-1}e^{-n\theta\bar{x}}$

From (8-32):
$$f_\theta(\theta\,|\bar{x}) = \frac{(c+n\bar{x})^{n+1}}{n!}\theta^n e^{-(c+n\bar{x})\theta}$$

From (8-31):
$$\hat{\theta} = \frac{(c+n\bar{x})^{n+1}}{n!}\int_0^\infty \theta^{n+1}e^{-(c+n\bar{x})\theta}d\theta = \frac{n+1}{c+n\bar{x}}$$

8-16 The sum $n\bar{x}$ is a Poisson RV with mean $n\theta$ (see Prob. 8-8). In the context of Bayesian estimation, this means that

$$f_{\bar{x}}(\bar{x}|\theta) = e^{-n\theta}\frac{(n\theta)^k}{k!} \qquad k = n\bar{x} = 0,1,\ldots$$

Inserting into (8-32), we obtain [see (4-76)]

$$f_\theta(\theta|(\bar{x}) = \frac{(n+c)^{n\bar{x}+b+1}}{\Gamma(n\bar{x}+b+a)}\theta^{n\bar{x}+b}e^{-(n+c)\theta}$$

and (8-31) yields

$$\hat{\theta} = \frac{(n+c)^{n\bar{x}+b+1}}{\Gamma(n\bar{x}+b+1)}\frac{\Gamma(n\bar{x}+b+2)}{(n+c)^{n\bar{x}+b+2}} = \frac{n\bar{x}+b+1}{n+c} \xrightarrow[n\to\infty]{} \bar{x}$$

8-17 From (8-17) with $t_{.95}(9)=2.26$

$$\bar{x} \pm \frac{t_u s}{\sqrt{n}} = 90 \pm 3.57 \qquad 86.43 < \eta \cdot 93.57$$

From (8-24) with $\chi^2_{.975}(9)=19.02$, $\chi^2_{.025}(9)=2.70$.

$$\frac{9 \times 5^2}{19.02} = 11.83 < \sigma^2 < \frac{9 \times 5^2}{2.70} = 83.33 \qquad 3.44 < \sigma < 9.13$$

8-18 The RVs x_i/σ are N(0,1), hence, the sum $z=(x_1^2 + \cdots + x^2_{10})/\sigma^2$ has a $\chi^2(10)$ distribution. This yields

$$P\{\chi^2_{.025}(10) < z < \chi^2_{.975}(10)\} = .95$$

$$\chi^2_{.025}(10) = 3.25 < \frac{4}{\sigma^2} < \chi^2_{.975}(10) = 20.48$$

$$0.442 < \sigma < 1.109$$

8-19 From (8-23) with $n=4, \chi^2_{.025}(4)=0.48,\ \chi^2_{.975}(4)=11.14$

$$n\hat{v} = .1^2 + .15^2 + .05^2 + .04^2 = .0366$$

$$\frac{.0366}{.048} > \sigma^2 > \frac{.0366}{11.14} \qquad 0.276 > \sigma > 0.057$$

8-20 In this problem $n=3,\ x_1+x_2+x_3=9.8$

$$f(x,c) \sim c^4x^3e^{-cx} \qquad f(X,c) = c^{4n}(x_1\ldots x_n)^{3n}e^{-cn\bar{x}}$$

$$\frac{\partial f(X,c)}{\partial c} = \left(\frac{4n}{c} - n\bar{x}\right)f(X,\theta) = 0 \qquad \hat{c} = \frac{4}{\bar{x}} = 1.224$$

8-21 The joint density

$$f(X,c) = c^n e^{-cn(\bar{x}-x_0)} \qquad x_i > x_0$$

has an interior maximum if

$$\frac{\partial f(X,c)}{\partial c} = 0 \qquad \hat{c} = \frac{1}{\bar{x}-x_0}$$

8-22 The probability

$$p = 1 - F_x(200) = e^{-200c}$$

of the event $\{x > 200\}$ is a monoton decreasing function of c. To find the ML estimate $\hat{c}$ of c it suffices to find the ML estimate $\hat{p}$ of p. From Example 8-28 it follows with k=62 and n=80 that

$$\hat{p} = \frac{62}{80} = .775 \text{ hence}$$

$$\hat{c} = -\frac{1}{200} \ln \hat{p} = 0.0013$$

8-23 The samples of x are the integers x_i and the joint density of the RVs x_i equals

$$f(X,\theta) = e^{-n\theta} \prod \frac{\theta^{x_i}}{x_i!} = e^{-n\theta} \frac{\theta^{n\bar{x}}}{\prod x_i!}$$

Hence, $f(X,\theta)$ is maximum if $-n + n\bar{x}/\theta = 0$. This yields $\hat{\theta} = \bar{x}$

8-24 If $L = \ln f(x,\theta)$ then

$$\frac{\partial L}{\partial \theta} = \frac{1}{f}\frac{\partial f}{\partial \theta} \qquad \frac{\partial^2 L}{\partial \theta^2} = \frac{1}{f}\frac{\partial^2 f}{\partial \theta^2} - \frac{1}{f^2}\left(\frac{\partial f}{\partial \theta}\right)^2 \qquad \frac{\partial^2 L}{\partial \theta^2} + \left(\frac{\partial L}{\partial \theta}\right)^2 = \frac{1}{f}\frac{\partial^2 f}{\partial \theta^2}$$

But

$$E\left\{\frac{1}{f}\frac{\partial^2 f}{\partial \theta^2}\right\} = \int_R \frac{1}{f}\frac{\partial^2 f}{\partial \theta^2} f dX = 0 \quad \text{hence} \quad E\left\{\frac{\partial^2 L}{\partial \theta^2} + \left(\frac{\partial L}{\partial \theta}\right)^2\right\} = 0$$

8-25 (a) From (8-307): Critical region

$$\bar{x} > c = \eta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}} = 8+2.326 \times \frac{2}{8} = 8.58$$

If η=8.7, then $\eta_q = \frac{8.7-8}{218} = 2.8$

$$\beta(\eta) = G(2.36 - 2.8) = .32$$

(b) We assume that α=.01, β (8.7)=.05 and wish to find n and c.

$$G(z_{1-\alpha}-\eta_q) = \beta \qquad z_{1-\alpha}-\eta_q = z_\beta$$

$$\eta_q = z_{.99} - z_{.05} = 4.97 = \frac{8.7-8}{2/\sqrt{n}}$$

$$n = 129 \qquad c = 8 + \frac{2}{\sqrt{129}} z_{.99} = 8.41$$

8-26 Our objective is to test the composite null hypothesis $\eta > \eta_0 = 28$ against the hypothesis $\eta < \eta_0$. Consider first the simple null hypothesis $\eta = \eta_0 = 28$. In this case, we can use (8-301) with

$$q = \frac{\bar{x}-\eta_0}{s/\sqrt{n}} \qquad \bar{x} = \frac{1}{17}\sum x_i = 27.67 \qquad s^2 = \frac{1}{16}\sum(x_i-\bar{x}) = 17.6$$

This yields s=4.2 and q=-0.33. Since

$$q_u = t_u\,(n-1) = t_{0.05}(16) = -\ 1.95 < -0.33$$

we conclude that the evidence does not support the rejection of the hypothesis η=28. The resulting OC function $\beta_0(\eta)$ is determined from (9-60c).

If η_0>28, then the corresponding value of q is larger than -0.33. From this it follows that the evidence does not support the hypothesis η_0 for any $\eta_0>28$. We note, however, that the corresponding OC function $\beta(\eta)$ is smaller than the function $\beta_0(\eta)$ obtained from (8-301) with $\eta_0=28$.

8-27 From (8-297) with $q_u=t_u(n-1)$: Critical region $|\bar{x}-\eta_0| > t_{1-\alpha/2}(n-1)s/\sqrt{n}$

1. $\underline{\alpha = .1}$ $\quad t_{.95}(63) = 1.67 \quad |\bar{x}-8| > 1.67 \times 1.5/8 = 0.313$

Since $\bar{x}=7.7$ is in the interval 8 ± 0.317, we accept H_0

2. $\underline{\alpha = .01}$ $\quad t_{.995}(63) = 2.62 \quad |\bar{x}-8| > 2.62 \times 1.5/8 = 0.49$

Since $\bar{x}=7.7$ is outside the interval 8 ± 0.49, we reject H_0.

8-28 We assume that the RVs $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are normal and independent. We form the difference $\underset{\sim}{w}=\underset{\sim}{\bar{x}}-\underset{\sim}{\bar{y}}$ of their sample means

$$\underset{\sim}{\bar{x}} = \frac{1}{16}\sum_{i=1}^{16}\underset{\sim}{x}_1 \qquad \underset{\sim}{\bar{y}} = \frac{1}{26}\sum_{i=1}^{26}\underset{\sim}{y}_i$$

and use as test statistic the ratio

$$\underset{\sim}{q} = \frac{\underset{\sim}{w}}{\sigma_w} \qquad \sigma_w^2 = \frac{\sigma_x^2}{16} + \frac{\sigma_y^2}{26}$$

The RV $\underset{\sim}{q}$ is normal with $\sigma_q=1$ and under hypothesis H_0, $E\{\underset{\sim}{q}\}=0$. We can, therefore, use (8-307) because $q_u = z_u$. To find q, we must determine σ_w.

Since σ_x and σ_y are not specified, we shall use the approximations $\sigma_x \simeq s_x = 1.1$ and $\sigma_y \simeq s_y = 0.9$. This yields

$$\sigma_w^2 \simeq \frac{1.1^2}{16} + \frac{0.9^2}{26} = 0.107 \qquad q = \frac{\bar{x} - \bar{y}}{\sigma_w} = \frac{0.4}{0.327} = 1.223$$

Since $z_{0.95} = 1.645 > 1.223$, we accept H_0.

8-29 (a) In this problem, n=64, k=22, $p_0 = q_0 = 0.5$

$$q = \frac{k - np_0}{\sqrt{np_0 q_0}} = 2.5 \qquad z_{\alpha/2} = -z_{1-\alpha/2} \simeq -2$$

Since 2.5 is outside the interval (2, -2), we reject the fair coin hypothesis [see (8-313)].

(b) From (8-313) with n=16, $p_0 = q_0 = 0.5$:

$$\frac{k_1 - np_0}{\sqrt{np_0 q_0}} = z_{\alpha/2} \qquad \frac{k_2 - np_0}{\sqrt{np_0 q_0}} = -z_{\alpha/2}$$

This yields $k_1 = 8 - 2 \times 2 = 4$, $k_2 = 8 + 2 \times 2 = 12$

8-30 We shall use as test statistic the sum

$$\underset{\sim}{q} = \underset{\sim}{x}_1 + \cdots + \underset{\sim}{x}_m \qquad n = 22$$

The critical region of the test is $q<q_\alpha$ where $q=x_1+\cdots+x_n=90$ [see (8-301)].

The RV $\underset{\sim}{q}$ is Poisson distributed with parameter $n\lambda$. Under hypothesis H_0, $\lambda=\lambda_0=5$; hence, $\eta_q=n\lambda_0=110=\sigma_q^2$. To find q_α we shall use the normal approximation. With $\alpha=0.05$ this yields

$$q_\alpha = n\lambda_0 + z_\alpha \sqrt{n\lambda_0} = 90-17.25 = 72.75$$

Since 90 > 72.75, we accapt the hypothesis that $\lambda=5$.

8-31 From (9-75) with n=102 and $p_{0i}=1/6$

$$q = \sum_{i=1}^{6} \frac{(k_i-17)^2}{17} = 2 \qquad \chi^2_{.95}(5) \simeq 11$$

Since 2<11, we accept the fair die hypothesis.

8-32 Uniformly distributed integers from 0 to 9 means that they have the same probability of appearing. With m=10, $p_{01}=.1$, and n=1,000, it follows from (8-325) that

$$q = \sum_{j=0}^{9} \frac{(n_j-100)^2}{100} = 17.76 \qquad \chi^2_{.95}(9) = 16.92$$

Since 17.76 > 16.92, we reject the uniformity hypothesis.

8-33 In this problem

$$f(x,\theta) = e^{-\theta}\frac{\theta^x}{x!} \qquad f(X,\theta) = \frac{e^{-n\theta}\,\theta^{n\bar{x}}}{x_1!\cdots x_n!}$$

$f(X,\theta)$ is maximum for $\theta=\theta_m=\bar{x}$. And since $\theta_{m0}=\theta_0$ we conclude that

$$\lambda(X) = \frac{e^{-n\theta_0}\theta_0^{n\bar{x}}}{e^{-n\bar{x}}\bar{x}^{n\bar{x}}} \qquad w = -2\,ln\,\lambda = 2n(\theta_0-\bar{x}) + \bar{x}\,ln(\bar{x}/\theta_0)$$

With $n=50$, $\theta_0=20$, $\bar{x}=1{,}058/50=21.16$, this yields $w=3$. Since $m_0=1$, $m=1$, and $\chi^2_{.95}(1)=3.84>3$, we accept H_0.

8-34 We form the RVs

$$\underset{\sim}{z} = \sum_{i=1}^{m}\left(\frac{x_i-\eta_x}{\sigma_x}\right)^2 \qquad \underset{\sim}{w} = \sum_{i=1}^{n}\left(\frac{y_i-\eta_y}{\sigma_y}\right)^2$$

These RVs are $\chi^2(m)$ and $\chi^2(n$ respectively. If $\sigma_x=\sigma_y$, then

$$\underset{\sim}{q} = \frac{z/m}{\underset{\sim}{w}/n}$$

Hence (see Prob. 6-23), $\underset{\sim}{q}$ has a Snedecor distribution. To test the hypothesis $\sigma_x=\sigma_y$, we use (8-297) where $q_u=F_u(m,n)$ is the tabulated u percentile of the Snedecor distribution. This yields the following test:

$$\text{Accept } H_o \text{ iff} \qquad F_{\alpha/2}(m,n) < q < F_{1-\alpha/2}(m,n).$$

8-35 If $\underset{\sim}{x}$ has a student-t distribution, then $f(-x)=f(x)$, hence (see Prob. 6-75)

$$E\{\underset{\sim}{x}\} = 0 \qquad \sigma_x^2 = E\{\underset{\sim}{x}^2\} = \frac{n}{n-2}$$

8-36 (a) Suppose that the probability $P(A)$ that player A wins a set equals $p=1-q$. He wins the match in five sets if he wins two of the first four sets and the fifth set. Hence, the probability $p_5(A)$ that he wins in five equals $6p^3q^2$. Similarly, the probability $p_5(B)$ that player B wins in five equals $6p^2q^3$. Hence,

$$p_5 = p_5(A) + p_5(B) = 6p^3q^2 + 6p^2q^3 = 6p^2q^2$$

is the probability that the match lasts five sets. If $p=q=1/2$, then $p_5=3/8$.

(b) Suppose now that $P(A) = \underset{\sim}{p}$ is an RV with density $f(p)$. In this case,

$$\underset{\sim}{p}_5 = 6\underset{\sim}{p}^2(1-\underset{\sim}{p}^2)$$

is an RV. We wish to find its best bayesian estimate. Using the MS criterion, we obtain

$$\hat{p}_5 = E\{\underset{\sim}{p}_5\} = \int_0^1 6p^2(1-p^2)f(p)dp$$

If $f(p)=1$, then $\hat{p}_5 = 1/5$.

8-37 Given

$$f_v(v) \sim e^{-v^2/2\sigma^2} \qquad f_\theta(\theta) \sim e^{-(\theta-\theta_o)^2/2\sigma_o^2}$$

To show that

$$f_\theta(\theta|X) \sim e^{-(\theta-\theta_1)^2/2\sigma_1^2}$$

where

$$\frac{1}{\sigma_1^2} \equiv \frac{1}{\sigma_o^2} + \frac{n}{\sigma^2} \qquad \theta_1 \equiv \frac{\sigma_1^2}{\sigma_o^2}\theta_o + \frac{n\sigma_1^2}{\sigma^2}\bar{x}$$

Proof

$$f_x(x|\theta) = f_v(x-\theta) \sim \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}$$

$$f(X)|\theta \sim \exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\theta)^2\right\}$$

Since $\sum(x_i-\theta)^2 = \sum(x_i-\bar{x})^2 + n(\bar{x}-\theta)^2$, we conclude from (8-32) omitting factors that do not depend on θ that

$$f(\theta|X) \sim \exp\left\{-\frac{1}{2}\left[\frac{(\theta-\theta_o)^2}{\sigma_o^2} + \frac{n(\bar{x}-\theta)^2}{\sigma^2}\right]\right\}$$

The above bracket equals

$$\left(\frac{1}{\sigma_o^2}+\frac{n}{\sigma^2}\right)\theta^2-2\left(\frac{\theta_o}{\sigma_o^2}+\frac{n\bar{x}}{\sigma^2}\right)\theta+\cdots=\frac{1}{\sigma_1^2}(\theta^2-2\theta\theta_1)+\cdots$$

and (i) follows.

8-38 The likelihood function of X equals

$$f(X,\theta)=\frac{1}{(\sqrt{2\pi\theta})^n}\exp\left\{-\frac{1}{2\theta}\sum(x_i-\eta)^2\right\}$$

where $\theta=\sigma^2$ is the unknown parameter. Hence

$$L(X,\theta)=-\frac{n}{2}\ell n\,(2\pi\theta)-\frac{1}{2\theta}\sum(x_i-\eta)^2$$

$$\frac{\partial L(X,\theta)}{\partial\theta}=-\frac{n}{2\theta}+\frac{1}{2\theta^2}\sum(x_i-\theta)^2=0 \qquad \hat{\theta}=\frac{1}{n}\sum(x_i-\eta)^2$$

8-39 The estimators $\hat{\underset{\sim}{\theta}}_1$ and $\hat{\underset{\sim}{\theta}}_2$ have the same variance because otherwise one or the other would not be best. Thus

$$E\{\hat{\underset{\sim}{\theta}}_1\}=E\{\hat{\underset{\sim}{\theta}}_2\}=\theta \qquad \text{var}\,\hat{\underset{\sim}{\theta}}_1=\text{var}\,\hat{\underset{\sim}{\theta}}_2=\sigma^2$$

If $\hat{\underset{\sim}{\theta}}=\frac{1}{2}(\hat{\underset{\sim}{\theta}}_1+\hat{\underset{\sim}{\theta}}_2)$, then

$$E\{\hat{\underset{\sim}{\theta}}\}=\theta \qquad \sigma_{\hat{\theta}}^2=\frac{1}{2}(\sigma^2+\sigma^2+2r\sigma^2)=\frac{1}{2}(1+r)\sigma^2$$

where σ is the correlation coefficient of $\hat{\underset{\sim}{\theta}}_1$ and $\hat{\underset{\sim}{\theta}}_2$. If r<1 then $\sigma_{\hat{\theta}}<\sigma$ which is impossible. Hence, r=1 and $\hat{\underset{\sim}{\theta}}_1=\hat{\underset{\sim}{\theta}}_2$ (see Prob. 6-53).

8-40 $k_1+k_2-np_1-np_2=n-n(p_1+p_2)=0$; Hence, $|k_1-np_1|=|k_2-np_2|$

$$\frac{(k_1-np_1)^2}{np_1}+\frac{(k_2-np_2)^2}{np_2}=(k_1-np_1)^2\left(\frac{1}{np_1}+\frac{1}{np_2}\right)=\frac{(k_1-np_1)^2}{np_1p_2}$$

8.41 It is given that

$$E\{T(X)\} = \int_{-\infty}^{\infty} T(X)\, f(X;\theta)\, dx = \psi(\theta),$$

so that after differentiating and making use of (8-81) we get

$$\int_{-\infty}^{\infty} T(X)\, \frac{\partial f(X;\theta)}{\partial \theta}\, dx = \psi^{'}(\theta) \qquad (8.41-1)$$

Also using (8-80)

$$\int_{-\infty}^{\infty} \psi(\theta)\, \frac{\partial f(X;\theta)}{\partial \theta}\, dx = 0, \qquad (8.41-2)$$

and the above two expressions give

$$\int_{-\infty}^{\infty} [T(X) - \psi(\theta)]\, \frac{\partial f(X;\theta)}{\partial \theta}\, dx = \psi^{'}(\theta) \qquad (8.41-3)$$

But

$$\frac{\partial f(X;\theta)}{\partial \theta} = \frac{1}{f(X;\theta)}\, \frac{\partial \log f(X;\theta)}{\partial \theta}$$

so that (8.41-3) simplifies to

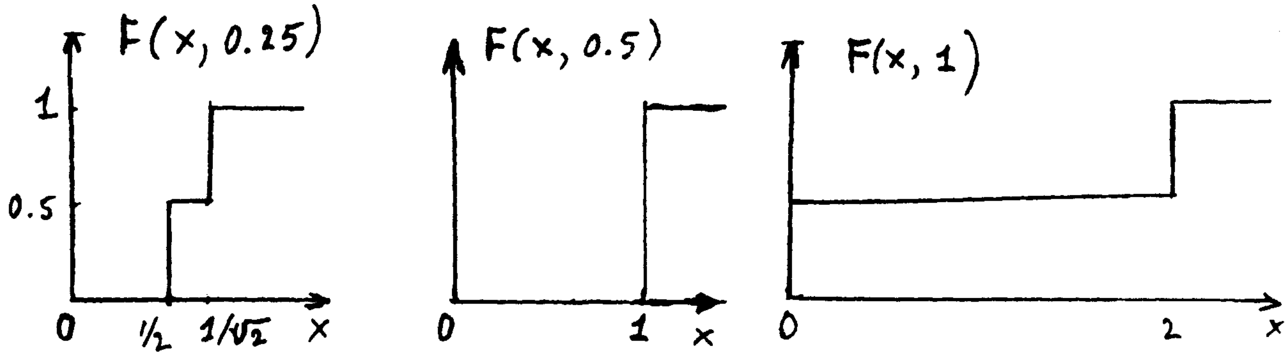
$$\int_{-\infty}^{\infty} \left[\{T(X) - \psi(\theta)\} \sqrt{f(X;\theta)}\right] \left[\sqrt{f(X;\theta)}\ \frac{\partial \log f(X;\theta)}{\partial \theta}\right] dx = \psi^{'}(\theta)$$

and application of Cauchy-Schwarz inequality as in (8-89)-(8-92), Text gives

$$E\left[\{T(X) - \psi(\theta)\}^2\right] \geq \frac{[\psi^{'}(\theta)]^2}{E\left\{\left(\frac{\partial \log f(X;\theta)}{\partial \theta}\right)^2\right\}}$$

9-1 (a) $E\{x(t)\} = t + 0.5 \sin \pi t$

$$x(t,\ \text{heads}) = \sin \pi t = \begin{cases} 1/\sqrt{2} & t = 0.25 \\ 1 & t = 0.5 \\ 0 & t = 1 \end{cases} \qquad x(t,\ \text{tails}) = 2t = \begin{cases} 0.5 \\ 1 \\ 2 \end{cases}$$



9-2

$$x(t) = e^{at}$$

$$\eta(t) = \int_{-\infty}^{\infty} e^{at} f_a(a)\,da \qquad R(t_1,t_2) = \int_{-\infty}^{\infty} e^{at_1} e^{at_2} f_a(a)\,da$$

From (5-16) with $x = g(a) = e^{ta}$ $\quad g'(a) = t\, e^{ta} = tx$

$$f(x,t) = \frac{1}{x|t|} f_a\left(\frac{1}{t} \ln x\right) U(x)$$

9-3 As we know, $E\{\underset{\sim}{x}(t)\} = \lambda t$ and var $\underset{\sim}{x}(t) = \lambda^2 t^2$ [see (9-18)]. But $E\{\underset{\sim}{x}(9) = 6\}$ by assumption, hence, $\lambda = 2/3$

(a) $E\{\underset{\sim}{x}(8)\} = 24$ var $\underset{\sim}{x}^2(t) = 24^2$

(b) The RV $\underset{\sim}{x}(2)$ is Poisson distributed with parameter $2\lambda = 6$. Hence,

$$P\{\underset{\sim}{x}(2) \le 3\} = e^{-2\lambda} \sum_{k=0}^{3} \frac{(2\lambda)^k}{k!}$$

(c) The RVs $\underset{\sim}{z} = \underset{\sim}{x}(2)$ and $\underset{\sim}{w} = \underset{\sim}{x}(4) - \underset{\sim}{x}(2)$ are independent and Poisson distributed with parameter 2λ. Hence,

$$P\{\underset{\sim}{z}=k\} = e^{-2\lambda} \frac{(2\lambda)^k}{k!} \qquad P\{\underset{\sim}{z} = k, \underset{\sim}{w} = m\} = e^{-4\lambda} \frac{(2\lambda)^k}{k!} \frac{(2\lambda)^m}{m!}$$

$$P\{\underset{\sim}{x}(4) \le 5 \mid \underset{\sim}{x}(2) \le 3\} = \frac{P\{z\le 3,\ w \le 5-z\}}{P\{\underset{\sim}{z} \le 3\}} \qquad P\{\underset{\sim}{z} \le 3\} = \sum_{k=0}^{3} p\{\underset{\sim}{z}=k\}$$

$$P\{\underset{\sim}{z} \le 3, \underset{\sim}{w} \le 5 - \underset{\sim}{z}\} = \sum_{k=0}^{3} \sum_{m=0}^{5-k} P\{\underset{\sim}{z} = k, \underset{\sim}{w} = m\}$$

9-4 $$\underset{\sim}{x}(t) = U(t - \underset{\sim}{c}) \qquad \underset{\sim}{y}(t) = \delta(t - \underset{\sim}{c}) = \underset{\sim}{x}'(t)$$

For t_1 or $t_2 < 0$, $R(t_1,t_2) = 0$; for t_1 and $t_2 > T$, $R(t_1,t_2) = 1$. Otherwise,

$$R_x(t_1,t_2) = \frac{1}{T}\min(t_1,t_2) \qquad \frac{\partial R_x}{\partial t_1} = \frac{1}{T}U(t_1 - t_2) \qquad -\frac{\partial^2 R_x}{\partial t_1 \partial t_2} = \frac{1}{T}\delta(t_1 - t_2)$$

From this and (9-105) it follows that $T R_y(t_1 - t_2) = \delta(t_1 - t_2)$ for $0 < t_1, t_2 < T$ and 0 otherwise.

9-5 $\underset{\sim}{a} - \underset{\sim}{b}\, t = 0$ iff $t = \underset{\sim}{t}_1 = \underset{\sim}{a}/\underset{\sim}{b}$. Setting $\sigma_1 = \sigma_2 = \sigma$ and $r = 0$ in (6-63), we obtain

$$P\{0 < \underset{\sim}{t}_1 < T\} = \frac{1}{2} + \frac{1}{\pi}\text{arc tan} T - \left(\frac{1}{2} + \frac{1}{\pi}\text{arc tan } 0\right)$$

9-6 The equations

$$w''(t) = \nu(t)U(t) \qquad w(0) = w'(0) = 0$$

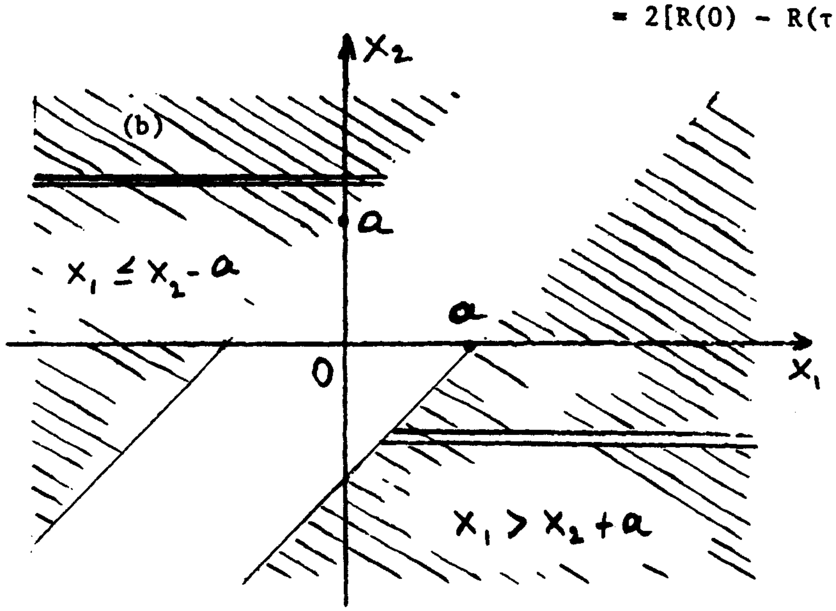
specify a system with input $\nu(t)U(t)$ and impulse response $h(t) = t\,U(t)$. Hence [see (9-100)]

$$E\{w^2(t)\} = q(t)U(t) * t^2U(t) = \int_0^t (t-\tau)^2 q(\tau)\,d\tau$$

9-7 (a) From (5-88) with $x = x(t+\tau) - x(t)$, and (8-101):

$$P\{|x(t+\tau) - x(t)| \geq a\} \leq \frac{E\{[x(t+\tau) - x(t)]^2\}}{a^2}$$

$$= 2[R(0) - R(\tau)]/a^2$$



The above probability equals the mass in the regions (shaded) $x_2 - x_1 > a$ and $x_2 - x_1 < -a$. Hence,

$$P\{|x(t+\tau) - x(t)| \geq a\}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{x_2 - a} f(x_1,x_2;\tau)\,dx_1 dx_2 + \int_{-\infty}^{\infty}\int_{x_2+a}^{\infty} f(x_1,x_2;\tau)\,dx_1 dx_2$$

9-8 (a) The RV $\underset{\sim}{x}(t)$ is normal with zero mean and variance $E\{\underset{\sim}{x}^2(t)\} = R(0)=4$, hence it is N(0,2) and $P\{\underset{\sim}{x}(t)\le 3\} = F(3) = G(1.5) = 0.933$

(b) $E\{[\underset{\sim}{x}(t+1) - \underset{\sim}{x}(t-1)]\} = 2[R(0)-R(2)] = 8(1-e^{-4})$

9-9 If $\underset{\sim}{x}(t) = \underset{\sim}{c}e^{j(\omega t+\theta)}$ and $\eta_c = 0$ then

$$\eta_x(t) = \eta_c e^{j(\omega t+\theta)} = 0 \qquad R_{xx}(t+\tau,t) = \sigma_c^2 e^{j\omega\tau}$$

hence, $\underset{\sim}{x}(t)$ is WSS. We shall prove the converse:

If the process $\underset{\sim}{x}(t) = \underset{\sim}{c}\,w(t)$ is WSS, then $\eta_c=0$ and $w(t) = e^{j(\omega t+\theta)}$ within a constant factor.

<u>Proof</u> $\eta_x(t) = \eta_c w(t)$ is independent of t; hence, $\eta_c=0$. The function $R_{xx}(t_1,t_2) = \sigma_c^2 w(t_1)w^*(t_2)$ depends only on $\tau=t_1-t_2$; hence, $w(t+\tau)w^*(t)=g(\tau)$. With $\tau=0$ this yields

$$|w(t)|^2 = g(0) = \text{constant} \qquad w(t) = ae^{j\phi(t)}$$

$$w(t+\tau)w^*(t) = a^2 e^{j[\phi(t+\tau)-\phi(t)]}$$

Hence the difference $\phi(t+\tau)-\phi(t)$ depends only on τ:

$$\phi(t+\tau)-\phi(t) = f(\tau) \tag{i}$$

From this it follows that, if $\phi(t)$ is continuous then, $\phi(t)$ is a linear function of t. To simplify the proof, we shall assume that $\phi(t)$ is differentiable. Differentiating with respect to t, we obtain $\phi'(t+\tau) = \phi'(t)$ for every τ. With t=0 this yields

$$\phi'(\tau) = \phi'(0) = \text{constant} \qquad \phi(\tau) = a\tau+b$$

9-10 We shall show that if $\underset{\sim}{x}(t)$ is a normal process with zero mean and $\underset{\sim}{z}(t) = \underset{\sim}{x}^2(t)$, then $C_{zz}(\tau) = 2C_{xx}^2(\tau)$.

From (7-61): If the RVs $\underset{\sim}{x}_k$ are normal and $E\{\underset{\sim}{x}_k\}=0$, then

$$E\{\underset{\sim}{x}_1\underset{\sim}{x}_2\underset{\sim}{x}_3\underset{\sim}{x}_4\} = E\{\underset{\sim}{x}_1\underset{\sim}{x}_2\}\ E\{\underset{\sim}{x}_3\underset{\sim}{x}_4\} + E\{\underset{\sim}{x}_1\underset{\sim}{x}_3\}\ E\{\underset{\sim}{x}_2\underset{\sim}{x}_4\} + E\{\underset{\sim}{x}_1\underset{\sim}{x}_4\}\ E\{\underset{\sim}{x}_2\underset{\sim}{x}_3\}$$

With $\underset{\sim}{x}_1=\underset{\sim}{x}_2=\underset{\sim}{x}(t+\tau)$ and $\underset{\sim}{x}_3 = \underset{\sim}{x}_4 = \underset{\sim}{x}(t)$, we conclude that the autocorrelation of $\underset{\sim}{z}(t)$ equals

$$E\{\underset{\sim}{x}^2(t+\tau)\underset{\sim}{x}^2(t)\} = E^2\{\underset{\sim}{x}^2(t+\tau)\} + 2E^2\ \{\underset{\sim}{x}(t+\tau)\underset{\sim}{x}(t)\} = R_{xx}{}^2(0) + 2R_{xx}{}^2(\tau)$$

And since $R_{xx}(\tau)=C_{xx}(\tau)$, and $E\{\underset{\sim}{z}(t)\} = R_{xx}(0)$, the above yields

$$C_{zz}(\tau) = R_{zz}(\tau) - E^2\{\underset{\sim}{z}(t)\} = 2C_{xx}{}^2(\tau)$$

9-11 $$\underset{\sim}{y}''(t) + 4\underset{\sim}{y}'(t) + 13\underset{\sim}{y}(t) = \underset{\sim}{x}(t) \text{ all } t$$

The process $\underset{\sim}{y}(t)$ is the response of a system with input $\underset{\sim}{x}(t) = 26 + \underset{\sim}{\nu}(t)$ and

$$H(s) = \frac{1}{s^2+45+13} \qquad h(t) = \frac{1}{3}\, e^{-2t}\sin 3tU(t)$$

Since $\eta_x = 26$, this yields $\eta_y = \eta_x H(0) = 2$. The centered process $\underset{\sim}{\bar{y}}(t) = \underset{\sim}{y}(t)-\eta_y$ is the response due to $\underset{\sim}{\nu}(t)$. Hence [see (9-100)]

$$E\{\underset{\sim}{\bar{y}}^2(t)\} = q \int_0^{\infty} h^2(t)dt = \frac{10}{104}$$

With b=4 and c=13 it follows that (see Example 9-276)

$$R_{\underset{\sim}{yy}}(\tau) = \frac{10}{104} e^{-2|\tau|} \left(\cos 3\tau - \frac{2}{3} \sin 3|\tau| \right) + 4$$

If $\underset{\sim}{\nu}$ is normal, then $\underset{\sim}{y}(t)$ is normal with mean 2 and variance $R_{yy}(0) - 4 = 10/104$; hence,

$$P\{\underset{\sim}{y}(t) \le 3\} = G\left(\frac{3-2}{0.31}\right) = G(3.24)$$

9-12 $$E\{\underset{\sim}{y}(t)\} = 0 \qquad R_{yy}(t_1,t_2) = \frac{R_{xx}(t_1,t_2)}{f(t_1)f(t_2)} = w(t_1-t_2)$$

$$E\{\underset{\sim}{z}(t)\} = 0 \qquad R_{zz}(t_1,t_2) = \frac{R_{xx}(t_1,t_2)}{\sqrt{q(t_1)}\ \sqrt{q(t_2)}} = \delta(t_1-t_2)$$

because $q(t_1)\delta(t_1-t_2) = \sqrt{q(t_1)}\ \sqrt{q(t_2)}\ \delta(t_1-t_2)$.

9-13 From (9-181) and the identity $4ab \le (a+b)^2$ it follows that

$$|R_{xy}(\tau)|^2 \le R_{xx}(0)R_{yy}(0) \le \frac{1}{4}[R_{xx}(0)+R_{yy}(0)]^2$$

9-14 Clearly (stationarity assumption)

$$E\{|\underset{\sim}{x}^*(t) - \underset{\sim}{y}^*(t)|^2\} = E\{|\underset{\sim}{x}(0) - \underset{\sim}{y}(0)|^2\} = 0$$

Furthermore,

$$E\{\underset{\sim}{x}(t+\tau)[\underset{\sim}{x}^*(t) - \underset{\sim}{y}^*(t)]\} = R_{xx}(\tau) - R_{xy}(\tau)$$

and [see (9-177)]

$$|E\{\underset{\sim}{x}(t+\tau)[\underset{\sim}{x}^*(t) - \underset{\sim}{y}^*(t)]\}|^2 \le E\{|\underset{\sim}{x}(t+\tau)|^2\}E\{|\underset{\sim}{x}^*(t) - \underset{\sim}{y}^*(t)|^2\} = 0$$

Hence, $R_{xx}(\tau) - R_{xy}(\tau) = 0$; similarly, $R_{yy}(\tau) = R_{xy}(\tau)$

9-15 $E\{|\underset{\sim}{x}(t+\tau) - \underset{\sim}{x}(t)|^2\} = E\{[\underset{\sim}{x}(t+\tau) - \underset{\sim}{x}(t)][\underset{\sim}{x}^*(t+\tau) - \underset{\sim}{x}^*(t)]\}$

$= R(0) - R(\tau) - R^*(\tau) + R(0) = 2R(0) - 2\,\underline{\mathrm{Re}}\,R(\tau)$

9-16 From $\Phi(1) = \Phi(2) = 0$ it follows that

$$E\{\cos \underset{\sim}{\phi}\} = E\{\sin \underset{\sim}{\phi}\} = E\{\cos 2\underset{\sim}{\phi}\} = E\{\sin 2\underset{\sim}{\phi}\} = 0$$

Hence, $E\{\underset{\sim}{x}(t)\} = \cos\omega t\, E\{\cos\underset{\sim}{\phi}\} - \sin\omega t\, E\{\sin\underset{\sim}{\phi}\} = 0$

and as in Example 9-14

$$2\cos[\omega(t+\tau)+\underset{\sim}{\phi}]\cos(\omega t+\underset{\sim}{\phi}) = \cos\omega\tau + \cos(2\omega t + \omega\tau + 2\underset{\sim}{\phi})$$

$$2R_x(\tau) = \cos\omega\tau$$

If $\underset{\sim}{\phi}$ is uniform in $(-\pi,\pi)$, then

$$\Phi(\lambda) = \frac{\sin\pi\omega}{\pi\omega} \qquad \Phi(1) = \Phi(2) = 0$$

9-17 (a) $\underset{\sim}{x}(t_1)\underset{\sim}{x}(t_2) = [\underset{\sim}{x}(t_1) - \underset{\sim}{x}(0)][\underset{\sim}{x}(t_2) - \underset{\sim}{x}(t_1) + \underset{\sim}{x}(t_1) - x(0)]$

$$R(t_1,t_2) = E\{[\underset{\sim}{x}(t_1) - \underset{\sim}{x}(0)]^2\} = E\{\underset{\sim}{x}^2(t_1)\} = R(t_1,t_1)$$

(b) If $t_1 + \varepsilon < t_2$, then $R_y(t_1,t_2) = 0$; if

$t_1 < t_2 < t_1 + \varepsilon$ then

$$E\{[\underset{\sim}{x}(t_1+\varepsilon) - \underset{\sim}{x}(t_1)][\underset{\sim}{x}(t_2+\varepsilon) - \underset{\sim}{x}(t_2)]\} = q(t_1 + \varepsilon - t_2)$$

Hence, $\varepsilon^2 R_y(\tau) = q(\varepsilon - |\tau|)$ for $|\tau| = |t_2 - t_1| \le \varepsilon$

9-18

$$E\{\underset{\sim}{x}(t)\underset{\sim}{y}(t)\} = \int_{-\infty}^{\infty} E\{\underset{\sim}{x}(t)\underset{\sim}{x}(t-\tau)\}h(\tau)d\tau$$

$$= \int_{-\infty}^{\infty} R_{xx}(t,t-\tau)h(\tau)d\tau = \int_{-\infty}^{\infty} q(t)\delta(\tau)h(\tau)d\tau = h(0)q(t)$$

9-19 As in Prob. 5-14, $g(x) = 6 + 3\,F_x(x)$. In this case,

$E\{\underset{\sim}{x}^2(t)\} = 4$, hence, $\underset{\sim}{x}(t)$ is $N(0,2)$ and $F_x(x) = G(x/2)$

9-20 $\underset{\sim}{x}(t)$ is SSS, hence, $P\{x(t) \le y\} = F_x(y)$ does not depend on t. The RVs $\underset{\sim}{\varepsilon}$ and $\underset{\sim}{x}(t)$ are independent, hence, [see (6-238)]

$$F_y(y) = P\{\underset{\sim}{x}(t-\underset{\sim}{\varepsilon}) \le y \mid \underset{\sim}{\varepsilon} = \varepsilon\} = P\{\underset{\sim}{x}(t-\varepsilon) \le y \mid \underset{\sim}{\varepsilon} = \varepsilon\}$$

$$= P\{\underset{\sim}{x}(t-\varepsilon) < y\} = F_x(y)$$

is independent of t. Similarly for higher order distributions.

9-21 $E\{x(t)\} = \eta =$ constant, hence, [see (9-102)] $E\{x'(t)\} = 0$
Furthermore, $R_{xx}(-\tau) = R_{xx}(\tau)$. hence, $R'_{xx}(0) = 0$ and (10-97) yields

$$E\{x(t)x'(t)\} = R_{xx'}(0) = 0$$

9-22 (a) $E\{z\,w\} = R_x(2) = 4e^{-4}$ $\qquad E\{z^2\} = E\{w^2\} = R_x(0) = 4$

$$E\{(z+w)^2\} = R_x(0) + R_x(0) + 2R_x(2) = 8(1+e^{-4})$$

(b) z is $N(0,2)$ $\qquad P\{z<1\} = F_z(1) = G(1/2)$

$r_{zw} = e^{-4}$, $\quad f_{zw}(z,w)$: $N(0,0;2,2;e^{-4})$

9-23 The RV $x'(t)$ is normal with zero mean and variance

$$E\{|x'(t)|^2\} = R_{x'x'}(0) = -R''(0)$$

Hence, $P\{x'(t) \le a\} = F_{x'}(a) = G\left[a/\sqrt{|R''(0)|}\right]$

9-24 The function arc sin x is odd, hence, it can be expanded into a sine series in the interval (-1,1):

$$\alpha(x) \equiv \arc\sin x = \sum_{n=1}^{\infty} b_n \sin n\pi x \qquad |x| \le 1$$

$$b_n = \int_{-1}^{1} \alpha(x)\sin n\pi x\,dx = -\frac{1}{n\pi}\int_{-1}^{1}\alpha(x)\,d\cos n\pi x$$

$$= -\frac{\alpha(x)\cos n\pi x}{n\pi}\Bigg|_{-1}^{1} + \frac{1}{n\pi}\int_{-1}^{1}\cos n\pi x\,d\alpha(x)$$

$$= -\frac{\cos n\pi}{n} + \frac{1}{n\pi}\int_{-\pi/2}^{\pi/2}\cos(n\pi\sin x)\,dx$$

and the result follows because [see (9-81)]

$$R_y(\tau) = \frac{2}{\pi}\arc\sin\frac{R_x(\tau)}{R_x(0)} \qquad J_o(z) = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\cos(z\sin x)\,dx$$

9-25 As we know [see (5-100) and (6-193)]

$$E\{e^{j\omega \underset{\sim}{x}(t)}\} = \exp\{-\frac{R(0)}{2}\omega^2\}$$

$$E\{e^{j[\omega_1 \underset{\sim}{x}(t+\tau)+\omega_2 \underset{\sim}{x}(t)]}\} = \exp\{-\frac{1}{2}[R(0)\omega_1^2 + 2R(\tau)\omega_1\omega_2 + R(0)\omega_2^2]\}$$

Hence, with $j\omega = a$

$$E\{I e^{a\underset{\sim}{x}(t)}\} = \exp\{\frac{a^2}{2} R_x(0)\} I$$

$$E\{I e^{a\underset{\sim}{x}(t+\tau)} I e^{a\underset{\sim}{x}(t)}\} = I^2 \exp\{a [R_x(0) + R_x(\tau)]\}$$

9-26 (a) $R_y(\tau) = a^2 E\{\underset{\sim}{x}[c(t+\tau)]\underset{\sim}{x}(ct)\} = a^2 R(c\tau)$

(b) If $\underset{\sim}{z}_\varepsilon(t) = \sqrt{\varepsilon}\,\underset{\sim}{x}(\varepsilon t)$ then $R_{z_\varepsilon}(\tau) = \varepsilon R_x(\varepsilon\tau)$ [as in (a)].

If $\delta > 0$ is sufficiently small and $\phi(t)$ is continuous at the origin, then

$$\int_{-\delta}^{\delta} R_{z_\varepsilon}(\tau)\phi(\tau)d\tau \simeq \phi(0)\int_{-\delta}^{\delta} \varepsilon R_x(\varepsilon\tau)\,d\tau$$

$$= \phi(0)\int_{-\varepsilon\delta}^{\varepsilon\delta} R(\tau)d\tau \xrightarrow[\varepsilon\to\infty]{} \phi(0)\int_{-\infty}^{\infty} R(\tau)d\tau = q\,\phi(0)$$

Hence, $R_{z_\varepsilon}(\tau) \to q\,\delta(\tau)$ as $\varepsilon \to \infty$.

9-27

$$\underset{\sim}{y}(t) = \int_{t-T}^{t} \underset{\sim}{x}(\tau)h(t-\tau)d\tau$$

Hence, $\underset{\sim}{y}(t_1)$ and $\underset{\sim}{y}(t_2)$ depend linearly on the values of $\underset{\sim}{x}(t)$ in the intervals $(t_1 - T,\ t_1)$ and $(t_2 - T,\ t_2)$ respectively. If $|t_1 - t_2| > T$ then these intervals do not overlap and since $E\{\underset{\sim}{x}(\tau_1)\underset{\sim}{x}(\tau_2)\} = 0$ for $\tau_1 \neq \tau_2$, it follows that $E\{\underset{\sim}{y}(t_1)\underset{\sim}{y}(t_2)\} = 0$.

9-28 (a)

$$I(t) = E\left\{\int_0^t \int_0^t h(t,\alpha)\underset{\sim}{x}(\alpha)h(t,\beta)\underset{\sim}{x}(\beta)\ d\alpha d\beta\right\}$$

$$= \int_0^t \int_0^t h(t,\alpha)h(t,\alpha)q(\alpha)\delta(\alpha-\beta)d\alpha d\beta = \int_0^t h^2(t,\alpha)q(\alpha)d\alpha$$

(b) If $\underset{\sim}{y}'(t) + c(t)\underset{\sim}{y}(t) = \underset{\sim}{x}(t)$, then $\underset{\sim}{y}(t)$ is the output of a linear time-varying system as in (a) with impulse response $h(t,\alpha)$ such that

$$\frac{\partial h(t,\alpha)}{\partial t} + c(t)h(t,\alpha) = \delta(t-\alpha) \qquad h(\alpha^-,\alpha) = 0$$

or equivalently

$$\frac{\partial h(t,\alpha)}{\partial t} + c(t)h(t,\alpha) = 0 \qquad t > 0 \quad h(\alpha^+,\alpha) = 1$$

This yields

$$h(t,\alpha) = e^{-\int_\alpha^t c(\tau)d\tau}$$

Hence, if

$$I(t) = \int_0^t h^2(t,\alpha)q(\alpha)d\alpha \quad \text{then} \quad I'(t) + 2c(t)I(t) = q(t)$$

because the impulse response of this equation equals

$$e^{-2\int_\alpha^t c(\tau)d\tau} = h^2(t,\alpha)$$

9-29 (a) If $\underset{\sim}{y}'(t) + 2\underset{\sim}{y}(t) = \underset{\sim}{x}(t)$, then $\underset{\sim}{y}(t) = \underset{\sim}{x}(t)*h(t)$

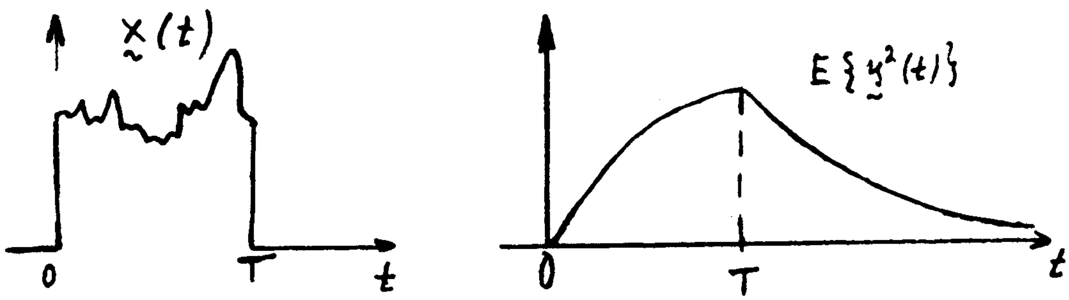
where $h(t) = e^{-2t}U(t)$ and with $q(t) = 5$, (10-90) yields

$$E\{\underset{\sim}{y}^2(t)\} = 5*e^{-4t}U(t) = 5\int_0^{\infty} e^{-4\tau}d\tau = \frac{5}{4}$$

(b) As in (a) with $q(t) = 5U(t)$. Hence, for $t > 0$

$$E\{\underset{\sim}{y}^2(t)\} = 5U(t)*e^{-4t}U(t) = 5\int_0^{t} e^{-4\tau}d\tau = \frac{5}{4}(1-e^{-4t})$$

9-30



From (9-90) with $q(t) = N[U(t) - U(t-T)]$

$$E\{\underset{\sim}{y}^2(t)\} = \begin{cases} AN\int_0^{t} e^{-2\alpha(t-\tau)}d\tau = \frac{AN}{2\alpha}(1-e^{-2\alpha t}) & 0 \le t < T \\ \\ AN\int_0^{T} e^{-2\alpha(t-\tau)}d\tau = \frac{AN}{2\alpha}(e^{2\alpha T}-1)e^{-2\alpha t} & t > T \end{cases}$$

9-31 Since $\underset{\sim}{x}(t)$ is WSS, the moments of S equal the moments of

$$\underset{\sim}{z} = \int_{-5}^{5} \underset{\sim}{x}(t)dt$$

Hence, (see Fig. 9-5)

$$E\{\underset{\sim}{s}^2\} = \int_{-5}^{5}\int_{-5}^{5} R_x(t_1-t_2)dt_1dt_2 = \int_{-10}^{10}(10-|\tau|)R_x(\tau)d\tau$$

$$E\{\underset{\sim}{s}\} = 80 \qquad \sigma_s^2 = 2\int_0^{10}(10-\tau)10e^{-2\tau}d\tau$$

9-32 $\underset{\sim}{y}(t) = \underset{\sim}{x}(t)*h(t) \qquad h(t) = e^{-2t}U(t)$

(a) $E\{\underset{\sim}{y}^2(t)\} = 5*e^{-4t}\,U(t) = 5/4$

$$R_{xy}(t_1,t_2) = 5\,\delta(t_1-t_2)*e^{-2t_2}\,U(t_2) = 5e^{-2(t_2-t_1)}U(t_2-t_1)$$

$$R_{yy}(t_1,t_2) = 5e^{-2(t_2-t_1)}U(t_2-t_1)*e^{-2t_1}\,U(t_1)$$

$$= \frac{5}{4}\,e^{-2|t_1-t_2|}$$

The first equation follows from (9-100) with $q(t) = 5$; the second from (9-94) with $R_{xx}(t_1,t_2) = 5\delta(t_1-t_2)$, and the third from (9-96).

(b) With $R_{xx}(t_1,t_2) = 5\delta(t_1-t_2)U(t_1)U(t_2)$, (9-94) and (9-96) yield the following: For t_1 or $t_2 < 0$, $R_{xy}(t_1,t_2) = R_{yy}(t_1,t_2) = 0$. For $0 < t_1 < t_2$

$$R_{xy}(t_1,t_2) = 5\delta(t_1-t_2)*e^{-2t_2} = 5\,e^{-2t_2}$$

$$R_{yy}(t_1,t_2) = \int_0^{t_1} 5\,e^{-2(t_1-\tau)}\,e^{-2(t_1-\tau)}\,d\tau = \frac{5}{4}\,e^{-2(t_2-t_1)}(1-e^{-4t_1})$$

9-33 $$\int_{-\infty}^{\infty} e^{-\alpha\tau^2} e^{-s\tau} d\tau = e^{s^2/4\alpha} \int_{-\infty}^{\infty} e^{-\alpha(\tau+s/2\alpha)^2} d\tau = \sqrt{\frac{\pi}{\alpha}}\, e^{s^2/4\alpha}$$

This yields

$$e^{-\alpha\tau^2} \longleftrightarrow \sqrt{\frac{\pi}{\alpha}}\, e^{-\omega^2/4\alpha}$$

$$e^{-\alpha\tau^2} \cos\omega_o\tau \longleftrightarrow \frac{1}{2}\sqrt{\frac{\pi}{\alpha}} \left[e^{-(\omega-\omega_o)^2/4\alpha} + e^{-(\omega+\omega_o)^2/4\alpha} \right]$$

9-34 $$G(x_1,x_2;\omega) = \int_{-\infty}^{\infty} f(x_1,x_2;\tau) e^{-j\omega\tau} d\tau$$

$$R(\tau) = E\{\mathbf{x}(t+\tau)\mathbf{x}(t)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2\, f(x_1,x_2;\tau)dx_1dx_2$$

$$S(\omega) = \int_{-\infty}^{\infty} R(\tau)e^{-j\omega\tau}d\tau = \int_{-\infty}^{\infty} e^{-j\omega\tau}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2\, f(x_1,x_2;\tau)dx_1\, dx_2 d\tau$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 \int_{-\infty}^{\infty} e^{-j\omega\tau}\, f(x_1,x_2;\tau)d\tau dx_1 dx_2$$

9-35 The process $\mathbf{y}(t) = \mathbf{x}(t+a) - \mathbf{x}(t-a)$ is the output of a system with input $\mathbf{x}(t)$ and system function

$$H(\omega) = e^{ja\omega} - e^{-ja\omega} = 2j\sin a\omega$$

Hence [see (9-150)]

$$S_y(\omega) = 4\sin^2 a\omega\, S_x(\omega) = (2 - e^{j2a\omega} - e^{-j2a\omega})S_x(\omega)$$

$$R_y(\tau) = 2R_x(\tau) - R_x(\tau+2a) - R_x(\tau-2a)$$

9-36 Since $S(\omega) \geq 0$, we conclude with (9-136) that

$$R(0) - R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)(1 - \cos \omega\tau)\, d\omega$$

$$\geq \frac{1}{8\pi} \int_{-\infty}^{\infty} S(\omega)(1 - \cos 2\omega\tau)\, d\omega = \frac{1}{4}[R(0) - R(2\tau)]$$

and the result follows for $n = 1$. Repeating the above, we obtain the general result.

9-37 **From** (6-197)

$$E\{\underset{\sim}{x}^2(t+\tau)\underset{\sim}{x}^2(t)\} = E\{\underset{\sim}{x}^2(t+\tau)\}E\{\underset{\sim}{x}^2(t)\} + 2E^2\{\underset{\sim}{x}^2(t+\tau)\underset{\sim}{x}^2(t)\}$$

Hence,

$$R_y(\tau) = R_x^2(0) + 2\, R_x^2(\tau) = I^2(1 + e^{-2\alpha|\tau|} + e^{-2\alpha|\tau|}\cos 2\beta\tau)$$

$$S_y(\omega) = \left[2\pi\delta(\omega) + \frac{4\alpha}{4\alpha^2 + \omega^2} + \frac{2\alpha}{4\alpha^2 + (\omega - 2\beta)^2} + \frac{2\alpha}{4\alpha^2 + (\omega + 2\beta)^2}\right]$$

Furthermore,

$$\eta_y = E\{\underset{\sim}{x}^2(t)\} = R_x(0) \qquad C_y(\tau) = 2R_x^2(\tau)$$

9-38

$$\int_{-\infty}^{\infty} S(\omega)\left|\sum_i a_i\, e^{j\omega\tau_i}\right|^2 d\omega = \int_{-\infty}^{\infty} S(\omega) \sum_{i,k} a_i a_k^*\, e^{j\omega(\tau_i - \tau_k)}\, d\omega$$

$$= \sum_{i,k} a_i a_k^* R(\tau_i - \tau_k) \geq 0$$

9-39 (a) $S(s) = \dfrac{1}{1+s^4} = \dfrac{1}{(s^2+\sqrt{2}s+1)(s^2-\sqrt{2}s+1)}$

A special case of example 9-27b with $b = \sqrt{2}$, $c = 1$. Hence,

$$R(\tau) = \frac{1}{2\sqrt{2}}\, e^{-|\tau|/\sqrt{2}} \left(\cos\frac{\tau}{\sqrt{2}} + \sin\frac{|\tau|}{\sqrt{2}}\right)$$

(b) From the pair $e^{-2|\tau|} \leftrightarrow 4/(4+\omega^2)$ and the convolution theorem it follows that

$$e^{-2|\tau|} * e^{-2|\tau|} \leftrightarrow \frac{16}{(4+\omega^2)^2}$$

Hence, for $\tau > 0$

$$16\, R(\tau) = \int_{\infty}^{\infty} e^{-2|x|}\, e^{-2|\tau-x|} dx = \int_{-\infty}^{0} e^{2x}\, e^{-2(\tau-x)} dx$$

$$+ \int_{0}^{\tau} e^{-2x}\, e^{-2(\tau-x)}\, dx + \int_{\tau}^{\infty} e^{-2x}\, e^{2(\tau-x)}\, dx = \frac{1}{2} e^{-2\tau}(1+2\tau)$$

And since $R(-\tau) = R(\tau)$, the above yields

$$e^{-2|\tau|}\, \frac{1+2|\tau|}{32} \longleftrightarrow \frac{1}{(4+\omega^2)^2}$$

9-40 $H^*(-s^*)\Big|_{s=j\omega} = H^*(j\omega)$ $\qquad H^*(1/z^*)\Big|_{z=e^{j\omega T}} = H^*(e^{j\omega T})$

Hence

$$H(s)H^*(-s^*)\Big|_{s=j\omega} = |H(j\omega)|^2 \qquad H(z)H^*(1/z^*)\Big|_{z=j\omega T} = |H(e^{j\omega T})|^2$$

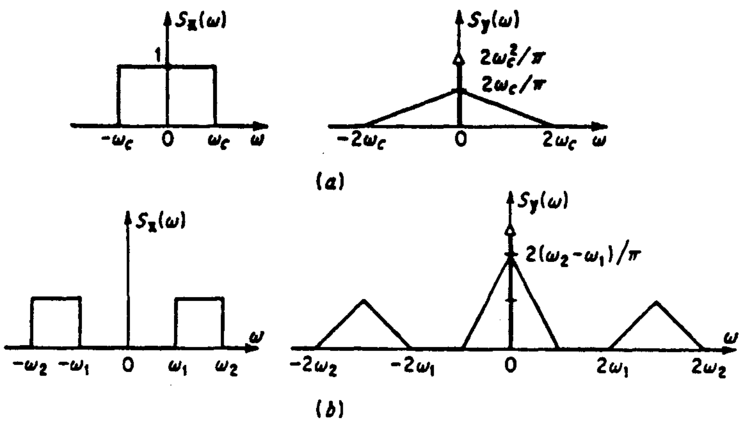
9-41 From (6-197)

$$R_y(\tau) = E\{\underset{\sim}{x}^2(t+\tau)\underset{\sim}{x}^2(t)\}$$

$$= E\{\underset{\sim}{x}^2(t+\tau)\}E\{\underset{\sim}{x}^2(t)\} + 2\,E^2\{\underset{\sim}{x}(t+\tau)\underset{\sim}{x}(t)\} = R_x^2(0) + 2\,R_x^2(\tau)$$

From the above and the frequency convolution theorem it follows that

$$S_y(\omega) = 2\pi R_x^2(0)\delta(\omega) + \frac{1}{\pi}\,S_x(\omega) * S_x(\omega)$$



9-42 $\quad \underset{\sim}{y}(t) = 2\underset{\sim}{x}(t) + 3\underset{\sim}{x}'(t) \qquad \eta_x = 5 \qquad C_{xx}(\tau) = 4e^{-2|\sigma|}$

The process $\underset{\sim}{y}(t)$ is the output of the system $H(s) = 2+3s$ with input $\underset{\sim}{x}(t)$. Hence, $\eta_y=5H(0)=10$

$$S_{yy}{}^c(\omega) = S_{xx}{}^c(\omega)|2+3j\omega|^2 = \frac{16}{4+\omega^2}(4+9\omega^2) = 144 - \frac{512}{4+\omega^2} = S_{yy}(\omega) - 2\pi\eta_y{}^2\delta(\omega)$$

9-43 (a) $y'(t) + 3y(t) = x(t)$, $R_{xx}(\tau) = 5\delta(\tau)$. The process $y(t)$ is the output of the system

$$H(s) = \frac{1}{s+3} \qquad h(t) = e^{-3t}U(t)$$

Hence, [see (9-100) and (9-150)]

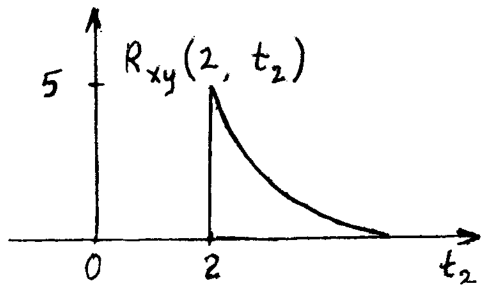
$$E\{y^2(t)\} = 5\int_0^{\infty} e^{-6t}dt = \frac{5}{6}$$

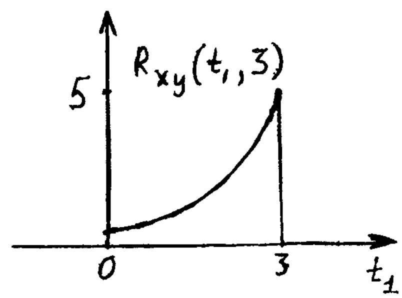
$$S_{yy}(\omega) = \frac{5}{\omega^2+9} \qquad R_{yy}(\tau) = \frac{5}{6}e^{-3|\tau|}$$

(b) As in Example 9-18:

$$E\{y^2(t)\} = 5\int_0^{t} e^{-6\alpha}d\alpha = \frac{5}{6}(1-e^{-6t}) \qquad t > 0$$

$$R_{xy}(t_1,t_2) = 5e^{-2|t_2-t_1|}U(t_1)U(t_2)U(t_2-t_1)$$





9-44 We shall show that: If $x(t)$ is a complex process with autocorrelation $R(\tau)$ and $|R(\tau_1)|=R(0)$ for some τ_1, then $R(\tau)=e^{j\omega_o\tau}w(\tau)$ where $w(\tau)$ is a periodic function with period τ_1. Furthermore, the process $y(t) = e^{-j\omega_o t}x(t)$ is MS periodic.

Proof Clearly, $R(\tau_1) = R(0)e^{j\phi}$. With $\omega_o = \phi/\tau_1$,

$$R_{yy}(\tau) = E\{x(t+\tau)e^{-j\omega_o(t+\tau)}x^*(t)e^{j\omega_o t}\} = R(\tau)e^{-j\omega\tau}$$

Hence, $R_{yy}(\tau_1)=e^{-j\omega_o\tau_1}R(\tau_1) = R(0) = R_{yy}(0)$. From this and (10-168) it follows that the function $w(\tau) = R_{yy}(\tau)$ is periodic.

9-45 (a) The cross spectrum $S_{\check{x}x}(\omega) = -j\,\mathrm{sgn}\,\omega\; S_{xx}(\omega)$ is an odd function. Hence,

$$E\{\underset{\sim}{x}(t)\underset{\sim}{\check{x}}'(t)\} = \frac{-j}{2\pi}\int_{\infty}^{\infty} \mathrm{sgn}\,\omega S_{xx}(\omega)\,d\omega = 0$$

(b) The process $\underset{\sim}{\check{x}}(t)$ is the output of the system

$$(-j\,\mathrm{sgn}\,\omega)(-j\,\mathrm{sgn}\,\omega) = -1$$

with input $\underset{\sim}{x}(t)$. Hence, $\underset{\sim}{\check{\check{x}}}(t) = -\underset{\sim}{x}(t)$.

9-46 In general

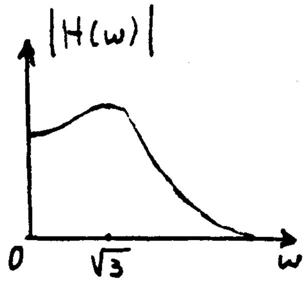
$$E\{\underset{\sim}{y}^2(t)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)\,|H(\omega)|^2 d\omega$$

$$\leq |H(\omega_m)|^2 \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)\,d\omega = E\{\underset{\sim}{x}^2(t)\}\,|H(\omega_m)|^2$$

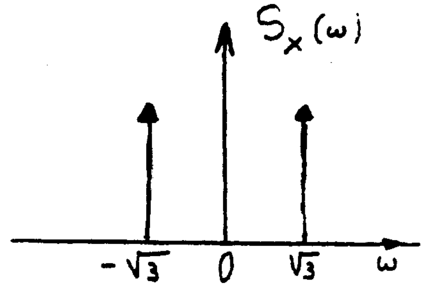
where $|H(\omega_m)|$ is the maximum of $|H(\omega)|$. In our case,

$$|H(\omega)|^2 = \frac{1}{(5-\omega)^2 + 4\omega^2} \quad \text{is maximum for } \omega = \sqrt{3}$$

and $|H(\omega_m)|^2 = 1/16$. Hence $E\{\underset{\sim}{y}^2(t)\} \leq 10/16$ with equality if $R_x(10) = 10\cos\sqrt{3}\,\tau$ (Fig. b).



(a)



(b)

9-47 If $R_x(\tau) = e^{j\omega_o\tau}$, then $S_x(\omega) = 2\pi\delta(\omega-\omega_o)$, hence, the integral of $S_x(\omega)$ equals zero in any interval not including the point $\omega = \omega_o$. From (9-182) it follows that the same is true for the integral of $S_{xy}(\omega)$. This shows that $S_{xy}(\omega)$ is a line at $\omega = \omega_o$ for any $y(t)$.

9-48 (a) As in (9-147) and (9-149)

$$R_{yx}(\tau) = R_{xx}(\tau) * h(\tau) = \int_{-\infty}^{\infty} e^{j\alpha(\tau-\gamma)} h(\gamma)\, d\gamma = e^{j\alpha\tau} H(\alpha)$$

$$R_{yy}(\tau) = R_{xx}(\tau) * \rho(\tau) = \int_{-\infty}^{\infty} e^{j\alpha(\tau-\gamma)} \rho(\gamma)\, d\gamma = e^{j\alpha\tau} |H(\alpha)|^2$$

(b) As in (9-94) and (9-95)

$$R_{yx}(t_1,t_2) = e^{-j\beta t_2} \int_{-\infty}^{\infty} e^{j\alpha(t_1-\gamma)} h(\gamma)\, d\gamma = e^{j(\alpha t_1 - \beta t_2)} H(\alpha)$$

$$R_{yy}(t_1,t_2) = e^{-j\alpha t_1} H(\alpha) \int_{-\infty}^{\infty} e^{-j\beta(t_2-\gamma)} h(\gamma)\, d\gamma = e^{j(\alpha t_1 - \beta t_2)} H(\alpha) H^*(\beta)$$

because $h(t)$ is real and $H(-\beta) = H^*(\beta)$.

9-49 If $S_{xx}(\omega) S_{yy}(\omega) \equiv 0$ then $S_{xx}(\omega) = 0$ or $S_{yy}(\omega) = 0$ in any interval (a,b). From this and (10-168) it follows that the integral of $S_{xy}(\omega)$ in any interval equals zero, hence, $S_{xy}(\omega) \equiv 0$.

9-50 This is the discrete-time version of theorem (9-162). From (9-163)

$$E^2\{(\underline{x}[n+m+1] - \underline{x}[n+m])\underline{x}[n]\} \le E\{|\underline{x}[n+m+1] - \underline{x}[n+m]|^2\}E\{|\underline{x}[n]|^2\}$$

$$(R[m+1] - R[m])^2 \le 2(R[0] - R[1])R[0] = 0$$

Hence, $R[m+1] = R[m]$ for any m.

9-51 We shall show that

$$2\frac{R^2[1]}{R[0]} - R[0] \le R[2] \le R[0] \qquad (i)$$

The covariance matrix of the RVs $\underline{x}[n]$, $\underline{x}[n+1]$, and $\underline{x}[n+2]$ is nonnegative [see (7-29)]:

$$\begin{vmatrix} R[0] & R[1] & R[2] \\ R[1] & R[0] & R[1] \\ R[2] & R[1] & R[0] \end{vmatrix} \ge 0$$

This yields

$$R[0]R^2[2] - 2R^2[1]R[2] - R^3[0] + 2R[0]R^2[1] \le 0$$

The above is a quadratic in R[2] with roots

$$R[0] \text{ and } -R[0] + 2R^2[1]/R[0]$$

Since it is nonpositive, R[2] must be between the roots as in (i)

9-52 If $\underline{x}[n] = Ae^{jn\underline{\omega}T}$ then

$$R_x[m] = A^2E\{e^{j(m+n)\underline{\omega}T}e^{-jn\underline{\omega}T}\} = A^2\int_{-\sigma}^{\sigma} e^{jm\omega T} f(\omega)d\omega$$

But [see (9-194)]

$$R[m] = \frac{1}{2\sigma}\int_{\sigma}^{\sigma} S_x(\omega)e^{jm\omega T}d\omega$$

hence, $A^2f(\omega) = S_x(\omega)/2\sigma$

9-53 (a) If $\underset{\sim}{y}(0) = \underset{\sim}{y}'(0) = 0$, then $\underset{\sim}{y}(t)$ is the output of a system with input $\underset{\sim}{x}(t)U(t)$ and impulse response $h(t)$ such that

$$h''(t) + 7h'(t) + 10h(t) = \delta(t) \qquad h(0^-) = h'(0^-) = 0$$

$$h(t) = \frac{1}{3}(e^{-2t} - e^{-5t})U(t)$$

and with $q(t) = 5\,U(t)$, (9-100) yields

$$E\{\underset{\sim}{y}^2(t)\} = \frac{5}{9}\int_0^t (e^{-2\tau} - e^{-5\tau})^2 d\tau$$

(b) If $\underset{\sim}{y}[-1] = \underset{\sim}{y}[-2] = 0$, then $\underset{\sim}{y}[n]$ is the output of a system with input $\underset{\sim}{x}[n]U[n]$ and delta response $h[n]$ such that

$$8h[n] - 6h[n-1] + h[n-2] = \delta[n] \qquad h[-1] = h[-2] = 0$$

$$h[n] = \left(\frac{1}{2^{n+2}} - \frac{1}{2^{2n+3}}\right) U[n]$$

and with $q[n] = 5\,U[n]$, (10-176) yields

$$E\{\underset{\sim}{y}^2[n]\} = 5\sum_{k=0}^{n}\left(\frac{1}{2^{k+2}} - \frac{1}{2^{2k+3}}\right)^2$$

9-54 $$\underset{\sim}{y}[n] = \underset{\sim}{x}[n]*h[n] \qquad h[n] = 2^{-n}U[n]$$

(a) $$E\{\underset{\sim}{y}^2[n]\} = 5*2^{-2n}U[n] = 0$$

$$R_{xy}[m_1,m_2] = 5\delta[m_1-m_2]*2^{-m_2}U[m_2] = 5\;2^{-(m_2-m_1)}U[m_2-m_1]$$

$$R_{yy}[m_1,m_2] = 5\times 2^{-(m_2-m_1)}U[m_2-m_1]*2^{-m_1}U[m_1]$$

$$= \frac{20}{3}\times 2^{-|m_1-m_2|}$$

The first equation follows from (9-190) with $q[n] = 5$; the second and third from (9-191) with $R_{xx}[m_1,m_2] = 5\,\delta[m_1-m_2]$.

(b) With $R_{xx}[m_1,m_2] = 5\,\delta[m_1-m_2]U[m_1]U[m_2]$, Prob. 9-25a yields the following: For m_1 or $m_2 < 0$, $R_{xy}[m_1,m_2] = R_{yy}[m_1,m_2] = 0$. For $0 < m_1 < m_2$

$$R_{xy}[m_1,m_2] = 5\,\delta[m_1-m_2]*2^{-m_2} = 5\times 2^{-m_2}$$

$$R_{yy}[m_1,m_2] = \sum_{k=0}^{m_1} 5\times 2^{-(m_2-k)}\,2^{-(m_1-k)} = \frac{5}{3}\,2^{-(m_2-m_1)}(4 - 2^{-2m_1})$$

9-55

(a) $R_x[m_1,m_2] = q[m_1]\delta[m_1-m_2]$

$$E\{\underset{\sim}{s}^2\} = \sum_{n=0}^{N} \sum_{k=0}^{N} a_n a_k \; E\{\underset{\sim}{x}[n]\underset{\sim}{x}[k]\}$$

$$= \sum_{n=0}^{N} \sum_{k=0}^{N} a_n a_k \; q[n]\delta[n-k] = \sum_{n=0}^{N} a_n^2 \; q[n]$$

(b) $R_x(t_1,t_2) = q(t_1)\delta(t_1-t_2)$

$$E\{s^2\} = \int_0^T\int_0^T a(t)a(\tau) \; E\{x(t)x(\tau)\}d\tau dt$$

$$= \int_0^T\int_0^T a(t)a(\tau)q(t)\delta(t-\tau)d\tau dt = \int_0^T a^2(t)q(t)dt$$

CHAPTER 10

10-1

(a) If $x(t)$ is a Poisson process as in Fig. 9-3a, then for a fixed t, $x(t)$ is a Poisson RV with parameter λt. Hence [see (5-119)] its characteristic function equals $\exp\{\lambda t(e^{j\omega} - 1)\}$.

(b) If $x(t)$ is a Wiener process then $f(x,t)$ is $N(0, \sqrt{\alpha t})$. Hence [see (5-100)] its first order characteristic function equals $\exp\{-\alpha t \omega^2/2\}$.

10-2 For large t, $x(t)$ and $y(t)$ can be approximated by two independent Wiener processes as in (10-52):

$$f_x(x,t) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-x^2/2\alpha t} \qquad f_y(y,t) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-y^2/2\alpha t}$$

Hence, $z(t)$ has a Rayleigh density [see (6-70)]. [Note. Exactly, $z(t)$ is a discrete-type RV taking the values $s\sqrt{m^2+n^2}$ where m and n are integers]. The product $f_z(z,t)dz$ equals approximately the probability that $z(t)$ is between z and $z+dz$ provided that $dz \gg T$.

10-3 The voltage $\underset{\sim}{v}(t)$ is the output of a system with input $\underset{\sim}{n}_e(t)$ and system function

$$H_1(s) = \frac{1}{LCs^2 + RCs + 1}$$

Hence,

$$S_v(\omega) = S_{n_e}(\omega)\,|H_1(j\omega)|^2 = \frac{2kTR}{(1-\omega^2 LC)^2 + R^2C^2\omega^2}$$

Furthermore,

$$Z_{ab}(s) = \frac{R + Ls}{LCs^2 + RCs + 1} \qquad \underline{Re}\; Z_{ab}(j\omega) = \frac{R}{(1-\omega^2 LC)^2 + R^2C^2\omega^2}$$

in agreement with (10-75).

The current $\underset{\sim}{i}(t)$ is the output of a system with input $\underset{\sim}{n}_e(t)$ and system function

$$H_2(s) = \frac{1}{R + Ls}$$

Hence,

$$S_i(\omega) = S_{n_e}(\omega)\,|H_2\;(j\omega)|^2 = \frac{2kTR}{R^2 + \omega^2 L^2}$$

Furthermore (short circuit admittance)

$$Y_{ab}(s) = \frac{1}{R + LS} \qquad \underline{Re}Y_{ab}(j\omega) = \frac{2kTR}{R^2 + L^2\omega^2}$$

in agreement with (10-78).

10-4 The equation $m\underset{\sim}{x}''(t) + f\underset{\sim}{x}'(t) = \underset{\sim}{F}(t)$ specifies a system with

$$H(s) = \frac{1}{ms^2 + fs} \qquad h(t) = \frac{1}{f}(1 - e^{-ft/m})U(t)$$

and (9-100) yields

$$E\{\underset{\sim}{x}^2(t)\} = \frac{2kTf}{f^2}\int_0^t (1 - e^{-2\alpha\tau})^2\, d\tau \qquad \alpha = \frac{f}{2m}$$

10-5 As in Example 12-2, a and b are such that

$$\underset{\sim}{x}(\tau) - a\,\underset{\sim}{x}(0) - b\underset{\sim}{v}(0) \perp \underset{\sim}{x}(0),\ \underset{\sim}{v}(0)$$

This yields

$$R_{xx}(\tau) = aR_{xx}(0) + b\,R_{xv}(0) \qquad (i)$$

$$R_{xv}(\tau) = aR_{xv}(0) + b\,R_{vv}(0)$$

where [see (10-163)]

$$R_{xx}(\tau) = A\,e^{-\alpha\tau}(\cos\beta\tau + \frac{\alpha}{\beta}\sin\beta\tau) \qquad \tau > 0$$

$$R_{xv}(\tau) = -R'_{xx}(\tau) = A\,e^{-\alpha\tau}(\sin\beta\tau)\,\frac{\alpha^2+\beta^2}{\beta}$$

$$R_{vv}(\tau) = R'_{xv}(\tau) = A\,e^{-\alpha\tau}(\cos\beta\tau - \frac{\alpha}{\beta}\sin\beta\tau)\,\frac{\alpha^2+\beta^2}{\beta^2}$$

Inserting into (i) and solving, we obtain

$$a = e^{-\alpha\tau}(\cos\beta\tau + \frac{\alpha}{\beta}\sin\beta\tau)$$

$$b = \frac{1}{\beta}e^{-\alpha\tau}\sin\beta\tau$$

Finally,

$$P = E\{[\underset{\sim}{x}(t) - a\,\underset{\sim}{x}(0) - b\,\underset{\sim}{v}(0)]\underset{\sim}{x}(t)\} = R_{xx}(0) - a\,R_{xx}(t) - b\,R_{xv}(t)$$

$$= \frac{2kTf}{m^2}\left[1 - e^{-2\alpha t}(1 + \frac{2\alpha^2}{\beta}\sin^2\beta t + \frac{\alpha}{\beta}\sin 2\beta t)\right]$$

10-6 If $\underset{\sim}{x}(t) = \underset{\sim}{w}(t^2)$ then [see (10-70)]

$$R_x(t_1,t_2) = E\{\underset{\sim}{w}(t_1^2)\underset{\sim}{w}(t_2^2)\} = \alpha t_1^2$$

If $\underset{\sim}{y}(t) = \underset{\sim}{w}^2(t)$ then [see (6-197)]

$$R_y(t_1,t_2) = E\{\underset{\sim}{w}^2(t_1)\underset{\sim}{w}^2(t_2)\}$$

$$= E\underset{\sim}{w}^2(t_1)E\{\underset{\sim}{w}^2(t_2)\} + 2\,E^2\{\underset{\sim}{w}(t_1)\underset{\sim}{w}(t_2)\} = \alpha^2 t_1 t_2 + 2\alpha^2 t_1^2$$

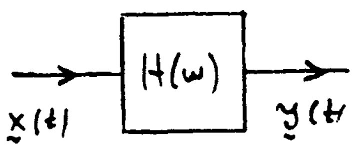
10-7 From (10-112):

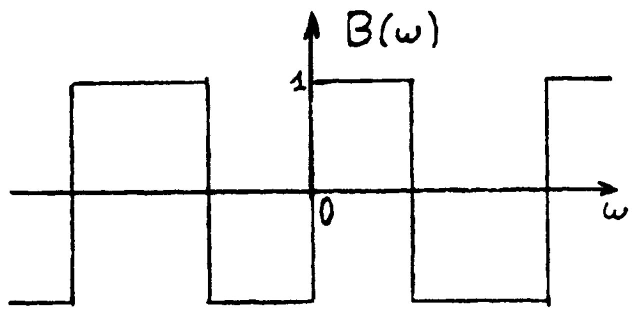
$$\eta_s = 3\int_0^{10} 2\,dt = 60 \qquad \sigma_s^2 = 3\int_0^{10} 4dt = 120 \qquad E\{\underset{\sim}{s}^2\} = 3720$$

$\underset{\sim}{s}(7) = 0$ if there are no points in the interval (7-10, 7). The number of points in this interval is a Poission RV with parameter $10\lambda = 30$. Hence, $P\{\underset{\sim}{s}(7) = 0\} = e^{-30}$.

10-8

$H(\omega) = j\,B(\omega)$





From the assumption: $S_{xx}(\omega) = S_{yy}(\omega) \qquad S_{xy}(-\omega) = -S_{xy}(\omega)$

From (9-148): $S_{yy}(\omega) = S_{xx}(\omega)\,|H(\omega)|^2 \qquad S_{xy}(\omega) = S_{xx}(\omega)H^*(\omega)$

Combining, we obtain

$$|H(\omega)|^2 = 1 \qquad H(-\omega) = -H(\omega)$$

Since h(t) is real, the second equation yields $H(\omega) = jB(\omega)$ and from the first it follows that

$$|B(\omega)| = 1$$

as in the figure.

10-9 With $\underset{\sim}{i}(t) = \underset{\sim}{a}(t)$, $\underset{\sim}{q}(t) = \underset{\sim}{b}(t)$, (11-63) yields

$$S_i(\omega) = S_q(\omega) \qquad S_{iq}(\omega) = -S_{qi}(\omega) = S_{qi}(-\omega)$$

Hence [see (11-75) and (11-82)],

$$S_w(\omega) = 2\,S_i(\omega) + 2j\,S_{qi}(\omega)$$

$$S_w(-\omega) = 2\,S_i(\omega) - 2j\,S_{qi}(\omega)$$

Adding and subtracting, we obtain

$$4\,S_i(\omega) = S_w(\omega) + S_w(-\omega) \qquad 4j\,S_{iq}(\omega) = S_w(-\omega) - S_w(\omega)$$

10-10 From (10-133)

$$\underset{\sim}{x}(t) = \underset{\sim}{\mathrm{Re}}\,[\underset{\sim}{w}(t)e^{j\omega_0 t}]$$

$$\underset{\sim}{x}(t-\tau) = \underset{\sim}{\mathrm{Re}}[\underset{\sim}{w}_\tau(t)e^{j\omega_0 t}] = \underset{\sim}{\mathrm{Re}}[\underset{\sim}{w}(t-\tau)e^{j\omega_0(t-\tau)}]$$

$$\underset{\sim}{w}_\tau(t) = \underset{\sim}{w}(t-\tau)e^{-j\omega_0\tau}$$

10-11 $$R_x''(\tau) \longleftrightarrow -\omega^2 S_x(\omega)$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 S_x(\omega)\,d\omega = -\,R_x''(0)$$

and with ω_0 the optimum carrier frequency,(10-150) yields

$$E\{|\underset{\sim}{w}'(t)|^2\} = \frac{M}{2\pi} = -\,2R_x''(0) - 2\omega_0^2 R_x(0)$$

10-12 From the stationarity of the process $\underset{\sim}{x}(t)\cos\omega t + \underset{\sim}{y}(t)\sin\omega t$ it follows that [see (10-130)]

$$C_{xx}(\tau) = C_{yy}(\tau) \qquad C_{xy} = -C_{yx}(\tau) \tag{i}$$

Using these identities, we shall express the joint density f(X,Y) of the 2n RVs

$$\underset{\sim}{X} = [\underset{\sim}{x}(t_1), \ldots, \underset{\sim}{x}(t_n)] \qquad \underset{\sim}{Y} = [\underset{\sim}{y}(t_1), \ldots, \underset{\sim}{y}(t_n)]$$

in terms of the covariance matrix C_{ZZ} of the complex vector $\underset{\sim}{Z} = \underset{\sim}{X} + j\underset{\sim}{Y}$. From (i) it follows that

$$E\{\underset{\sim}{x}(t_i)\underset{\sim}{x}(t_j)\} = E\{\underset{\sim}{y}(t_i)\underset{\sim}{y}(t_j)\} \qquad E\{\underset{\sim}{x}(t_i)\underset{\sim}{y}(t_j)\} = -E\{\underset{\sim}{y}(t_i)\underset{\sim}{x}(t_j)\}$$

This yields

$C_{XX} = C_{YY}$, and $C_{XY} = -C_{YX}$; hence, f(X,Y) is given by (8-62).

10-13 The signal $\underset{\sim}{c}(t) = f(t)$ is an extreme case of a cyclostationary process as in (10-178) with

$$h(t) = \begin{cases} f(t) & 0 \le t < T \\ 0 & \text{otherwise} \end{cases} \longleftrightarrow H(\omega) = \int_0^T f(t)e^{-j\omega t}dt$$

and $\underset{\sim}{c}_m = 1$, $R[m] = 1$. Hence [see (10A-2)]

$$\sum_{m=-\infty}^{\infty} R_m e^{-jm\omega T} = \sum_{m=-\infty}^{\infty} e^{-jm\omega T} = T\sum_{m=-\infty}^{\infty} \delta(\omega - \frac{2\pi}{T}m)$$

From the above and (10-180) it follows that the process $\underset{\sim}{x}(t) = f(t-\underset{\sim}{\theta})$ is stationary with power spectrum

$$S(\omega) = \left|\int_0^T f(t)e^{-j\omega t}dt\right|^2 \sum_{m=-\infty}^{\infty} \delta(\omega - \frac{2\pi}{T}m)$$

10-14 The process

$$\underline{y}_N(t) = \underline{x}(t+\tau) - \sum_{n=-N}^{N} \underline{x}(t+nT)\,\frac{\sin\sigma(\tau-nT)}{\sigma(\tau-nT)}$$

is the output of a system with input $\underline{x}(t)$ and system function

$$H_N(\omega) = e^{j\omega\tau} - \sum_{n=-N}^{N} \frac{\sin\sigma(\tau-nT)}{\sigma(\tau-nT)}\,e^{jnT\omega}$$

Furthermore, $\underline{\varepsilon}_N(\tau) = \underline{y}_N(0)$, hence [see (9-153)]

$$E\{\underline{\varepsilon}_N^2(\tau)\} = E\{\underline{y}^2(0)\} = \frac{1}{2\pi}\int_{\infty}^{\infty} S(\omega)\,|H_N(\omega)|^2 d\omega \qquad \text{(i)}$$

The function $H_N(\omega)$ is the truncation error in the Fourier series expansion of $e^{j\omega\tau}$ in the interval $(-\sigma,\sigma)$. Hence, for $N > N_0$

$$|H_N(\omega)| < \varepsilon \qquad\qquad |\omega| < \sigma$$

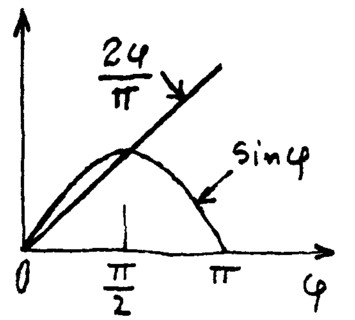
From this and (i) it follows that, if $S(\omega) = 0$ for $|\omega| < \sigma$, then

$$E\{\underline{\varepsilon}_N^2(\tau)\} = \frac{1}{2\pi}\int_{-\sigma}^{\sigma} S(\omega)\,|H_N(\omega)|^2 d\omega < \varepsilon\, R(0) \qquad N > N_0$$

10-15 [see after (10-195)]

$$R(0) - R(\tau) = \frac{1}{2\pi}\int_{-\sigma}^{\sigma} S(\omega)(1-\cos\omega\tau)\,d\omega$$

$$\leq \frac{\tau^2}{4\pi}\int_{-\sigma}^{\sigma}\omega^2 S(\omega)\,d\omega = \frac{-\tau^2}{2}R''(0)$$



Furthermore, since

$$\sin\phi \geq \frac{2\phi}{\pi} \qquad 0 \leq \phi \leq \frac{\pi}{2}$$

we obtain

$$R(0) - R(\tau) = \frac{1}{2\pi}\int_{-\sigma}^{\sigma} S(\omega)\, 2\sin^2\frac{\omega\tau}{2}\, d\omega$$

$$\geq \frac{2\tau^2}{\pi^2}\frac{1}{2\pi}\int_{-\sigma}^{\sigma}\omega^2 S(\omega)\,d\omega = \frac{-2\tau^2}{\pi^2}R''(0)$$

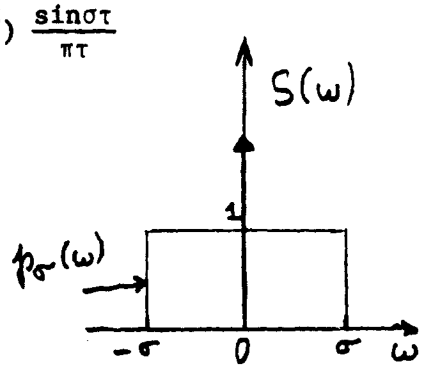
10-16 With $T = \pi/\sigma$

$$R(mT) = E\{\underset{\sim}{x}(nT+mT)\underset{\sim}{x}(nT)\} = \begin{cases} I & m = 0 \\ \eta^2 & m \neq 0 \end{cases}$$

Hence [see (10-196)]

$$R(\tau) = \sum_{m=-\infty}^{\infty} R(mT)\frac{\sin\sigma(\tau-mT)}{\sigma(\tau-mT)} = \eta^2 + (I-\eta^2)\frac{\sin\sigma\tau}{\pi\tau}$$

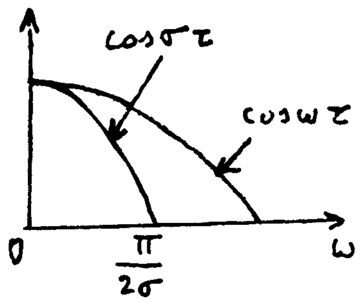
$$S(\omega) = 2\pi\eta^2\delta(\omega) + 2\pi(I-\eta^2)p_\sigma(\omega)$$



10-17 Given $E\{\underset{\sim}{x}(n+m)\underset{\sim}{x}(n)\} = N\delta[m]$

This is a special case of Prob. 10-16 with $\eta = 0$, $I = N$.

10-18 If $|\tau| < \pi/2\sigma$, then



$$\cos \omega\tau \geq \cos \sigma\tau \qquad |\omega| \leq \sigma$$

$$R(\tau) = \frac{1}{2\pi}\int_{-\sigma}^{\sigma} S(\omega)\cos\omega\tau\, d\omega$$

$$\geq \frac{\cos\sigma\tau}{2\pi}\int_{-\sigma}^{\sigma} S(\omega)\, d\omega = R(0)\cos\sigma\tau$$

10-19 From (10-133) with $c = \sigma$

$$P_1(\omega,\tau) + j\omega\, P_2(\omega,\tau) = 1$$

$$P_1(\omega,\tau) + j(\omega+\tau)P_2(\omega,\tau) = e^{j\sigma\tau}$$

Hence,

$$P_1(\omega,\tau) = 1 - \frac{\omega}{\sigma}(e^{j\sigma\tau} - 1) \qquad P_2(\omega,\tau) = \frac{1}{j\sigma}(e^{j\sigma\tau} - 1)$$

Inserting into (11-141), we obtain

$$p_1(\tau) = \frac{4\sin^2(\sigma\tau/2)}{\sigma^2\tau^2} \qquad p_2(\tau) = \frac{4\sin^2(\sigma\tau/2)}{\sigma^2\tau}$$

and with $t = 0$, the desired result follows from (10-206) because $\bar{T} = 2T$ and

$$\sin^2 \frac{\sigma(\tau-2nT)}{2} = \sin^2\left(\frac{\sigma\tau}{2} - n\pi\right) = \sin^2 \frac{\sigma\tau}{2}$$

10-20 As in (10-213)

$$\underset{\sim}{P}(\omega) = \frac{1}{\lambda}\int_{-a}^{a} \cos\omega t \; \underset{\sim}{z}(t) \cos\omega_c t \, dt$$

$$E\{\underset{\sim}{P}(\omega)\} = \int_{-a}^{a} \cos\omega t \cos\omega_c t \, dt$$

$$\sigma^2_{P(\omega)} = \frac{1}{\lambda}\int_{-a}^{a} \cos^2\omega_c t_2 \cos^2 \omega t_2 \, dt_2$$

10-21 We shall show that if

$$\underset{\sim}{X}_c(\omega) = \frac{1}{\lambda} \sum_{|t_i|<c} \underset{\sim}{x}(t_i)e^{-j\omega t_i} = \frac{1}{\lambda}\int_{-a}^{a} \underset{\sim}{x}(t)\underset{\sim}{z}(t)e^{-j\omega t}dt$$

where $\underset{\sim}{z}(t) = \sum \delta(t-\underset{\sim}{t}_i)$ is a Poisson impulse train, then

$$E\{|\underset{\sim}{X}_c(\omega)|^2\} \simeq 2cS_x(\omega) + \frac{2c}{\lambda} R_x(0)$$

Proof

Since $R_z(\tau) = \lambda^2 + \lambda\delta(\tau)$, it follows that

$$E\left\{|\underset{\sim}{X}_c(\omega)|^2\right\} = \frac{1}{\lambda^2}\int_{-c}^{c}\int_{-c}^{c} R_x(t_1-t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2$$

$$= \int_{-c}^{c} e^{j\omega t_2}\int_{-c}^{c} R_x(t_1-t_2)e^{-j\omega t_1}dt_1dt_2 + \frac{1}{\lambda}\int_{-c}^{c} R_x(0)dt_2$$

If $\int_{-\infty}^{\infty} |R_x(\tau)| < \infty$ then for sufficient large c, the inner integral on the right is nearly equal to $S_x(\omega)^{-j\omega t_2}$ and (i) follows.

10-22 $E\{\underset{\sim}{z}(t)\} = g(t) \qquad E\{\underset{\sim}{w}(t)\} = g(t) - g(T)t/T = g(t)$

$$\underset{\sim}{w}(t) = \left(1 - \frac{t}{T}\right)\int_0^t \underset{\sim}{x}(\alpha)d\alpha - \frac{t}{T}\int_t^T \underset{\sim}{x}(\alpha)d\alpha$$

The above two integrals are uncorrelated because $\underset{\sim}{n}(t)$ is white noise. Hence, as in Example 9-5

$$\sigma_w^2 = \left(1 - \frac{t}{T}\right)^2 Nt + \frac{t^2}{T^2} N(T - t) = Nt\left(1 - \frac{t}{T}\right)$$

<u>Note</u> The above shows that the information that $g(T) = 0$ can be used to improve the estimate of $g(t)$. Indeed, if we use $\underset{\sim}{w}(t)$ instead of $\underset{\sim}{z}(t)$ for the estimate of $g(t)$ in terms of the data $\underset{\sim}{x}(t)$, the variance is reduced from Nt to $Nt(1-t/T)$.

10-23 (a) Since $|\sum_i a_i b_i| \le \sum |a_i||b_i|$, it suffices to assume that the numbers a_i and b_i are real. The quadratic

$$I(z) = \sum_i (a_i - z\,b_i)^2 = z^2 \sum_i b_i^2 - 2z \sum_i a_i b_i + \sum_i a_i^2$$

is nonnegative for every real z, hence, its discriminant cannot be positive. This yields (i).

(b) With $f[n]$ and $R_\nu[m] = S_0\delta[m]$ as in Prob. 10-24a (white noise)

$$y_f[n_0] = \sum h[n] f[n_0 - n] \qquad \underset{\sim}{y}_\nu[n] = \sum h[n] \underset{\sim}{\nu}[n]$$

$$E\{\underset{\sim}{y}_\nu^2[n]\} = S_0\,\rho[0] = S_0 \sum |h[n]|^2$$

[see (9-213)] And (i) yields

$$\frac{y_f^2[n_0]}{E\{\underset{\sim}{y}_\nu^2[n]\}} = \frac{|\sum h[n] f[n_0 - n]|^2}{S_0 \sum h^2[n]} \le \frac{1}{S_0} \sum |h[n]|^2$$

with equality iff $h[n] = kf^*[n_0 - n]$.

10-24 (a) Given $F(z)$ and $S_\nu(\omega) = S_0 \equiv$ constant . The z transform of $y_f[n]$ equals $F(z)H(z)$. Hence, [see (9-109)]

$$y_f[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(e^{j\omega T})H(e^{j\omega T})e^{jn\omega T}d\omega$$

$$\frac{y_f^2[n]}{E\{\underset{\sim}{y}_\nu^2[n]\}} = \frac{\left|\int_{-\pi}^{\pi} F(e^{j\omega T})H(e^{j\omega T})d\omega\right|^2}{S_0\int_{-\pi}^{\pi}\left|H(e^{j\omega T})\right|^2 d\omega}$$

$$\le \frac{1}{S_0}\int_{\pi}^{\pi}\left|F(e^{j\omega T})\right|^2 d\omega$$

The last inequality follows from Schwarz's inequality with equality iff

$$H(e^{j\omega T}) = kF^*(e^{j\omega T}) = kF(e^{-j\omega T}),\text{ i.e., iff } H(z) = kF(z^{-1})$$

(b) Given arbitrary $R_\nu[m]$, $F(z)$, and the form of $H(z)$ (FIR); to find the coefficients a_m of $H(z)$. In this case

$$y_f[n] = a_0f[n] + a_1f[n-1] + \cdots + a_Nf[n-N]$$

$$\underset{\sim}{y}_\nu[n] = a_0\underset{\sim}{\nu}[n] + a_1\underset{\sim}{\nu}[n-1] + \cdots + a_N\underset{\sim}{\nu}[n-N]$$

To maximize the signal-to-noise ratio it suffices to minimize

$$E\{\underset{\sim}{y}_\nu^2[n]\} = \sum_{k,r=0}^{N} a_ka_rR_\nu[k-r]$$

subject to the constraint that the sum

$$y_f[0] = a_0f[0] + a_1f[-1] + \cdots + a_Nf[-N]$$

is constant. With λ a constant (Lagrange multiplier), we minimize the sum

$$I = \sum_{k,r=0}^{N} a_k a_r R[k-r] - \lambda\left[\sum_{k=0}^{N} a_k f[-k] - y_f[0]\right]$$

this yields the system

$$\frac{\partial I}{\partial a_k} = 0 = \sum_{r=0}^{N}\left[a_r R_\nu[k-r] - \lambda f[-k]\right] \qquad k = 0,\ldots,N$$

whose solution yields a_k.

10-25 $B = A|H(\omega_0)| = \dfrac{A}{\sqrt{\alpha^2+\omega_0^2}}$ $\qquad S_{y_n}(\omega) = \dfrac{N}{\alpha^2+\omega^2}$

$$R_{y_n}(\tau) = \frac{N}{2\alpha}\, e^{-\alpha|\tau|} \qquad E\{\underset{\sim}{y}_n^2(t)\} = R_{y_n}(0) = \frac{N}{2\alpha}$$

$$\frac{B^2}{E\{\underset{\sim}{y}_n^2(t)\}} = \frac{2A^2}{N}\,\frac{\alpha}{\alpha^2+\omega_0^2} \qquad \text{Max. if } \alpha = \omega_0$$

10-26 Since $H(\omega)$ is determined within a constant factor, we can sssume that the response $y_f(t_o)$ of the optimum $H(\omega)$ due to $f(t)$ is constant:

$$y_f(t_o) = \sum_{i=0}^{m} a_i\, f(t_o-iT) = c \qquad \text{(i)}$$

Our problem is to minimize the variance

$$V = E\{\underset{\sim}{y}_\nu^2(t)\} = \sum_{n=0}^{m} a_n \sum_{i=0}^{m} a_i\, R(nT-iT) \qquad \text{(ii)}$$

of $\underset{\sim}{y}_\nu(t)$ subject to the constraint (i). This yields the system

$$\frac{\partial V}{\partial a_n} = \sum_{i=0}^{m} a_i\, R(nT-iT) - kf(t_o-nT) = 0$$

where k is a constant (lagrange multiplier). With a_n so determined, we conclude from (ii) that

$$V = \sum_{n=0}^{m} ka_n\, f(t_o-nT) = ky_f(t_o) \qquad r^2 = \frac{y_f^2(t_o)}{ky_f(t_o)}$$

10-27 $R_{yyy}(\mu,\nu) = E\{\underset{\sim}{x}(t+\mu)+c[\,[\underset{\sim}{x}(t+\nu)+c]\,[\underset{\sim}{x}(t)+c]\} = R(\mu,\nu) + cR(\mu) + cR(\nu) + cR(\mu-\nu) + c^3$

because $E\{\underset{\sim}{x}(t)\} = 0$. Furthermore,

$$R(\mu) \Leftrightarrow 2\pi S(u)\delta(v) \qquad R(\nu) = 2\pi\delta(u)S(v) \qquad c^3 \Leftrightarrow 4\pi^2\delta(u)\delta(v)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(\mu-\nu)e^{-j(u\mu+v\nu)}d\mu d\nu = \int_{-\infty}^{\infty} R(\tau)e^{-ju\tau}d\tau \int_{-\infty}^{\infty} e^{-j(u+v)\nu}d\nu = 2\pi S(u)\delta(u+v)$$

10-28 We shall use the equations $E\{\underset{\sim}{\bar{x}}(t)\} = 0$, $E\{\underset{\sim}{\bar{x}}^2(t)\} = \lambda t$. Suppose that $t_1 < t_2 < t_3$. Clearly,

$$\underset{\sim}{\bar{x}}(t_2) = \underset{\sim}{\bar{x}}(t_1) + [\underset{\sim}{\bar{x}}(t_2)-\underset{\sim}{\bar{x}}(t_1)]$$

$$\underset{\sim}{\bar{x}}(t_3) = \underset{\sim}{\bar{x}}(t_1) + [\underset{\sim}{\bar{x}}(t_2) - \underset{\sim}{\bar{x}}(t_1)] + [\underset{\sim}{\bar{x}}(t_3)-\underset{\sim}{\bar{x}}(t_2)] \qquad \text{(i)}$$

Inserting into the product $\underset{\sim}{\bar{x}}(t_1)\underset{\sim}{\bar{x}}(t_2)\underset{\sim}{\bar{x}}(t_3)$ and using the identity $E\{\underset{\sim}{\bar{x}}(t_i)-\underset{\sim}{\bar{x}}(t_j)\} = 0$ and the independence of the three terms on the right of (i), we obtain

$$E\{\underset{\sim}{x}(t_1)\underset{\sim}{x}(t_2)\underset{\sim}{x}(t_3)\} = E\{\underset{\sim}{x}^3(t_1)\} = \lambda t_1 = \lambda \min(t_1, t_2, t_3)$$

Since $\underset{\sim}{\bar{z}}(t) = \underset{\sim}{\bar{x}}'(t)$, we conclude from (9-120)-(9-122) that

$$R_{\bar{z}\bar{z}\bar{z}}(t_1,t_2,t_3) = \frac{\partial^3 R_{xxx}(t_1,t_2,t_3)}{\partial t_1 \partial t_2 \partial t_3} = \lambda \frac{\partial^3 \min(t_1,t_2,t_3)}{\partial t_1 \partial t_2 \partial t_3}$$

It suffices therefore to show that the right side equals $\lambda\delta(t_1-t_2)\delta(t_1-t_2)$. This is a consequence of the following:

$$\frac{\partial \min(t_1,t_2,t_3)}{\partial t_3} = t_1 U(t_2-t_1)\delta(t_3-t_1) + t_2\, U(t_1-t_2)\,\delta\,(t_3-t_2)$$

$$+ U(t_1-t_3)U(t_2-t_3) - t_3\delta(t_1-t_3)U(t_2-t_3) - t_3 U(t_1-t_3)\delta(t_2-t_3)$$

$$= U(t_1-t_3)U(t_2-t_3)$$

because $t_i\delta(t_i-t_j) = t_j\delta(t_j-t_i)$. Hence,

$$\frac{\partial^2 \min(t_1,t_2,t_3)}{\partial t_2 \partial t_3} = U(t_1-t_3)\delta(t_2-t_3) \qquad \frac{\partial^2 \min(t_1,t_2,t_3)}{\partial t_1 \delta t_2 \partial t_3} = \delta(t_1-t_2)\delta(t_1-t_3)$$

10-29 See outline given in text.

CHAPTER 12

12-1 $\underset{\sim}{x}(t) = 10 + \underset{\sim}{\nu}(t) \qquad R_\nu(\tau) = 2\,\delta(\tau) \qquad E\{\underset{\sim}{\nu}(t)\} = 0$

$$E\{\underset{\sim}{n}_T\} = E\{\underset{\sim}{x}(t)\} = 10 \qquad C_x(\tau) = 2\delta(\tau)$$

From (12-5)

$$\sigma^2_{n_T} = \frac{1}{2T}\int_{-T}^{T} C_x(\tau)(1 - \frac{|\tau|}{2T})\,d\tau = \frac{1}{2T}\int_{-T}^{T} 2\delta(\tau)\ (1 - \frac{|\tau|}{2T})d\tau = \frac{1}{T}$$

12-2 The process $\underset{\sim}{x}(t)$ is normal (note correction) and such that

$$F(x,x;\tau) \longrightarrow F^2(x) \quad \text{as} \quad \tau \to \infty \qquad (i)$$

We shall show that it is mean-ergodic. It suffices to show that [see (12-10)]

$$C(\tau) \longrightarrow 0 \qquad \text{as} \qquad \tau \to \infty$$

Proof. We can assume (scaling and centering) that $\eta = 0$ $C(0) = 1$. With this assumption, the RVs $\underset{\sim}{x}(t+\tau)$ and $\underset{\sim}{x}(t)$ are $N(0,0;1,1;r)$ where $r = r(\tau) = C(\tau)$ is the autocovariance of $\underset{\sim}{x}(t)$. Hence,

$$f(x_1,x_2;\tau) = \frac{1}{2\pi\sqrt{1-r^2}}\ \exp\left\{-\frac{1}{2(1-r^2)}\ (x_1^2 - 2rx_1x_2 + x_2^2)\right\}$$

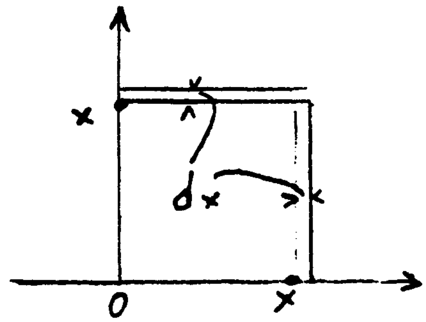
$$= \frac{1}{2\pi\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}(x_1 - rx_2)^2\right\} e^{-x_2^2/2}$$

Clearly, $f(x,y) = f(y,x)$, hence, (see figure)

$$F(x+dx,\ x+dx;\tau) - F(x,x,\tau) = 2\int_{-\infty}^{x} f(\xi,x)\,d\xi\,dx$$

$$= \frac{1}{\pi\sqrt{1-r^2}}\int_{-\infty}^{x}\exp\left\{-\frac{1}{2(1-r^2)}(\xi - xr)^2\right\} d\xi\ e^{-x^2/2}\,dx$$

Furthermore,



$$F^2(x+dx) - F^2(x) = 2\,F(x)f(x)dx$$

From the above and (i) it follows that

$$G\left(\frac{x-rx}{\sqrt{1-r^2}}\right) \xrightarrow[\tau\to\infty]{} G(x)$$

Hence, $r(\tau) \longrightarrow 0$ as $\tau \to \infty$

12-3 If $\underline{x}(t)$ is normal, then [see (12-27)]

$$C_{zz}(\tau) = R_x(\lambda+\tau)R_x(\lambda-\tau) + R_x^2(\tau) \qquad \underline{z}(t) = \underline{x}(t+\lambda)\underline{x}(t)$$

If, therefore, $R_x(\tau) = 0$ for $|\tau| > a$, then $C_{zz}(\tau) = 0$ for $|\tau| > \lambda + a$.

12-4 If $\underline{x}(t) = \underline{a}\, e^{j(\omega t + \underline{\phi})}$ then the time-average

$$\frac{1}{2T}\int_{-T}^{T} \underline{x}(t+\tau)\underline{x}^*(t)dt = e^{j\omega\tau}\,|\underline{a}|^2$$

12-5 If $\underline{z}(t) = \underline{x}(t+\lambda)\underline{y}(t)$, then

$$C_{zz}(\tau) = E\{\underline{x}(t+\lambda+\tau)\underline{y}(t+\tau)\underline{x}(t+\lambda)\underline{y}(t)\} - R_{xy}^2(\lambda)$$

and the result follows from (12-5).

12-6 The process $\bar{\underset{\sim}{x}}(t) = \underset{\sim}{x}(t - \underset{\sim}{\theta})$ is stationary with mean $\bar{\eta}$ and covariance $\bar{C}(\tau)$ given by [see (10-176) and (10-177)]

$$\bar{\eta} = \frac{1}{T}\int_0^T \eta(t)\,dt \qquad \bar{C}(\tau) = \frac{1}{T}\int_0^T C(t+\tau,t)\,dt$$

If $R(t+\tau,t) \to \eta^2(t)$ as $\tau \to \infty$ (note correction), then

$$C(t+\tau,t) \xrightarrow[\tau\to\infty]{} 0 \qquad \text{hence} \qquad \bar{C}(\tau) \xrightarrow[\tau\to\infty]{} 0$$

This shows that [see (12-10)], $\bar{\underset{\sim}{x}}(t)$ is ergodic, therefore,

$$\frac{1}{2c}\int_{-c}^{c} \bar{\underset{\sim}{x}}(t)\,dt = \frac{1}{2c}\int_{-c+\underset{\sim}{\theta}}^{c+\underset{\sim}{\theta}} \underset{\sim}{x}(t)\,dt \longrightarrow \bar{\eta}$$

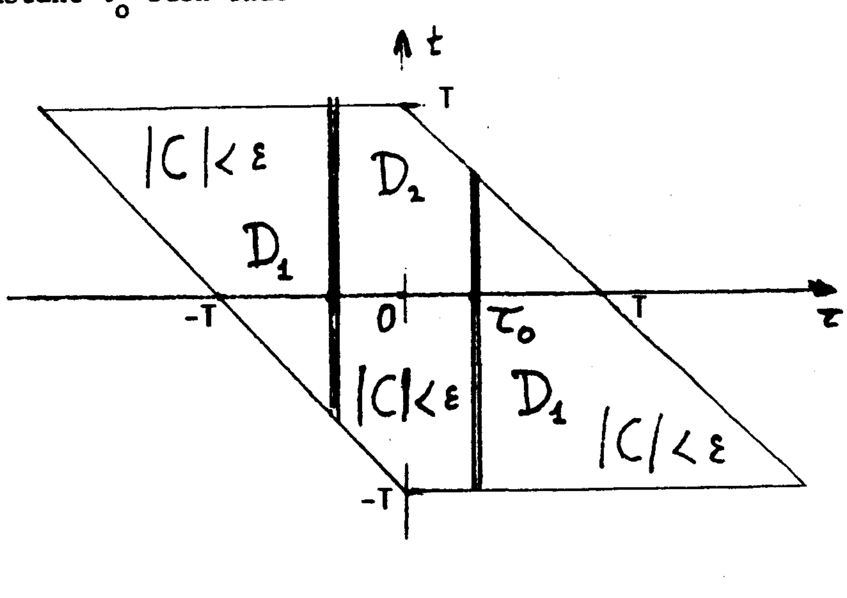
This yields the desired result because for a specific outcome, $\underset{\sim}{\theta}(\zeta) = \theta$ is a constant and

$$\lim_{c\to\infty} \frac{1}{2c}\int_{-c+\theta}^{c+\theta} \underset{\sim}{x}(t)\,dt = \lim_{c\to\infty} \frac{1}{2c}\int_{-c}^{c} \underset{\sim}{x}(t)\,dt$$

12-7 From (9-38) it follows that

$$4T^2\sigma_T^2 = \int_{-T}^{T}\int_{-T}^{T} C(t_1,t_2)\,dt_1,dt_2 = \iint_D C(t+\tau,t)\,d\tau\,dt$$

where D is the parallelogram in the figure. Given $\varepsilon > 0$, we can find a constant τ_o such that



$|C(t+\tau,t)| < \varepsilon$ for $|\tau| > \tau_o$ (uniform continuity). Furthermore, if $C(t,t) < P$ then

$$C^2(t_1,t_2) \le C(t_1,t_1)C(t_2,t_2) < P^2$$

Thus,

$|C| < \varepsilon$ in D_1 and $|C| < P$ in D_2

The area of D_1 is less than $4T^2$; the area of D_2 is less than $4\tau_o T$. Hence

$$\sigma_T^2 < \varepsilon + \frac{\tau_o}{T} \xrightarrow[T\to\infty]{} \varepsilon$$

And since ε is arbitrary, we conclude that $\sigma_T \to 0$.

12-8 It follows from (6-234) with $\underset{\sim}{x}(t) = \underset{\sim}{x}$, $\underset{\sim}{x}(t+\lambda) = \underset{\sim}{y}$

$$\eta_1 = \eta_2 = 0 \qquad \sigma_1^2 = \sigma_2^2 = R(0) \qquad r\sigma_1,\sigma_2 = R(\lambda)$$

(b) The proof is based on the identity

$$E\{\underset{\sim}{y}|M\} = E\,\{E\{\underset{\sim}{y}|\underset{\sim}{x}\}|M\} \qquad M = \{\underset{\sim}{x}(t)\in D\} \tag{i}$$

<u>Proof</u> Suppose first that D consists of the union of open intervals. In this case, if $x\in D$, then for small δ the interval $(x,x+\delta)$ is a subset of D, hence

$$\{x \le \underset{\sim}{x} < x + dx,\ M\} = \{x \le \underset{\sim}{x} < x + dx\}$$

for $x\in D$ and $\{\emptyset\}$ otherwise. This yields

$$f(x|M)\,dx = \frac{P\{x\le \underset{\sim}{x}<x+dx\}}{P(M)} = \frac{1}{p} f(x)dx \qquad p = P(M)$$

for $x\in D$ and 0 otherwise. Similarly, $f(x,y|M) = f(x,y)/p$ for $x\in D$ and 0 otherwise. From the above it follows that

$$E\{E\{\underset{\sim}{y}|\underset{\sim}{x}|M\}\} = \int_D \left(\int_{-\infty}^{\infty} yf(y|x)dy\right) f_x(x|M)dx$$

$$= \int_D \int_{-\infty}^{\infty} \frac{yf(x,y)f(x)}{f(x)\,p}\,dydx = \int_D \int_{-\infty}^{\infty} yf(x,y|M)dydx = E\{\underset{\sim}{y}|M\}$$

If D has isolated points, we replace each $x \in D$ by an open interval $(x-\varepsilon,\ x+\varepsilon)$ forming an open set D_ε. Clearly, $D_\varepsilon \to D$ as $\varepsilon \to 0$ and (i) follows if at the isolated points x_i of D, $E\{\underset{\sim}{y}|x_i\}$ is interpreted as a limit.

Since $E\{\underset{\sim}{x}(t+\lambda)|\underset{\sim}{x}(t)\} = R(\lambda)\underset{\sim}{x}(t)/R(0)$, (i) yields

$$E\{\underset{\sim}{x}(t+\lambda)|M\} = E\{E\{\underset{\sim}{x}(t+\lambda)|\underset{\sim}{x}(t)\}|M\} = E\left\{\frac{R(\lambda)}{R(0)}\underset{\sim}{x}(t)|M\right\} = \frac{R(\lambda)}{R(0)}\ \overline{x}$$

(c) We select for D the interval (a,b) and we form the samples $\underset{\sim}{x}(nT)$, $\underset{\sim}{x}(nT+\lambda)$ of a single realization of $\underset{\sim}{x}(t)$ retaining only the pairs $\underset{\sim}{x}(t_i)$, $\underset{\sim}{x}(t_i+\lambda)$ such that $a < \underset{\sim}{x}(t_i) < b$. Using (5-51), we obtain

$$E\{\underset{\sim}{x}(t+\lambda)|\ a < \underset{\sim}{x}(t) < b\} = \frac{R(\lambda)}{R(0)}\overline{x} \simeq \frac{1}{N}\sum_{i=1}^{N} \underset{\sim}{x}(t_i+\lambda)$$

where $\overline{x} = E\{\underset{\sim}{x}(t)|a<\underset{\sim}{x}(t)<b\}$. This approximation is satisfactory if N is large and $R(\tau)\simeq 0$ for $\tau > T$.

12-9 (a) From (7-61) with $E\{w(t)\} = C_{xy}(\lambda)$:

$$R_{ww}(\tau) = C_{xy}(\lambda+\tau)C_{xy}(\lambda-\tau) + C_{xx}(\tau)C_{yy}(\tau) + C_{xy}^2(\lambda) = C_{ww}(\tau) + C_{xy}^2(\lambda)$$

(b) It follows from (a) that if

$$C_{xx}(\tau) \to 0 \qquad C_{yy}(\tau) \to 0 \qquad C_{xy}(\sigma) \to 0$$

then $C_{ww}(\tau) \to 0$ as $|\tau| \to \infty$; hence [see (12-10)] the process $\underset{\sim}{x}(t)$ and $\underset{\sim}{y}(t)$ are covariance ergodic.

12-10 From (10B-1) with $g(x) = 1$:

$$\left| \int_a^b f(x)dx \right|^2 \le \int_a^b |f(x)|^2 dt \int_a^b 1^2 \times dx = (b-a)\int_a^b |f(x)|^2 dx$$

12-11 We use as estimate of η the time average $\hat{\eta}_T$ in (12-1): As we know (see Example 12-4)

$$E\{\eta_T\} = \eta \qquad \sigma_T^2 = \frac{5}{2T}$$

We wish to find ε such that

$$P\{\eta - \varepsilon < \eta_T < \eta + \varepsilon\} = 0.95$$

(a) From (5-88):

$$0.95 = P\{|\eta_T - \eta| \le \varepsilon\} \le 1 - \frac{\sigma_T^2}{\varepsilon^2} \qquad \varepsilon = \varepsilon_a \le \frac{\sigma_T}{\sqrt{0.05}} = \frac{50}{T}$$

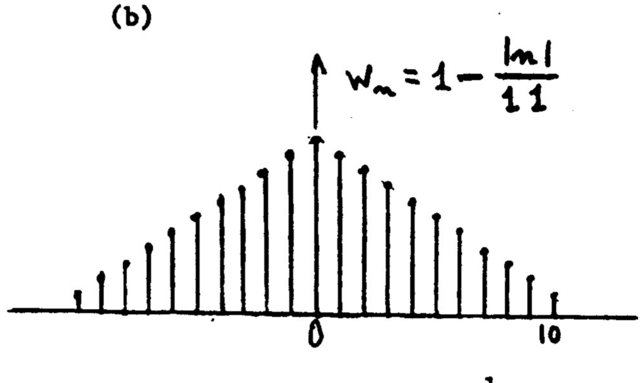
(b) If $\nu(t)$ is normal, then η_T is normal; hence,

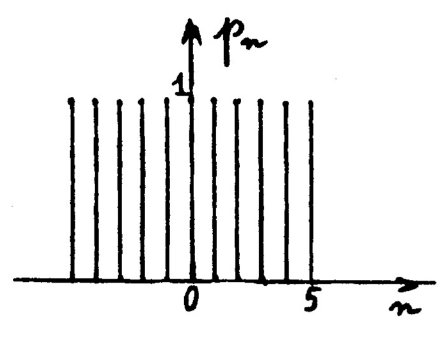
$$0.95 = 2G\left(\frac{\varepsilon}{\sigma_T}\right) - 1 \qquad G\left(\frac{\varepsilon}{\sigma_T}\right) = 0.975 \qquad \frac{\varepsilon}{\sigma_T} = z_{0.975}$$

This yields $\varepsilon = \varepsilon_b \simeq \sqrt{10/5} = \varepsilon_a\sqrt{5}$

12-12 (a) It follows from the convolution theorem for Fourier series

(b)





With p_n as above, $w_n = \frac{1}{11} p_n * p_n$

$$P(\omega) = \sum_{n=-5}^{5} e^{-jnT\omega} = \frac{\sin 5.5\omega T}{\sin 0.5\omega T} \qquad W(\omega) = \frac{1}{11} P^2(\omega)$$

12-13
$$\underset{\sim}{X}_T(\omega) = \frac{1}{\sqrt{2T}} \int_{-T}^{T} x(t)e^{-j\omega t}dt \qquad \underset{\sim}{S}_T(\omega) = |X_T(\omega)|^2$$

and
$$\Gamma(u,v) = \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} R(t_1 - t_2)e^{-j(u\,t_1 + v\,t_2)}\, dt_1 dt_2 \tag{i}$$

as in (9-173) and (9-174) yield

$$E\{\underset{\sim}{S}_T(\omega)\} = \Gamma(\omega, -\omega)$$

$$\text{Var } \underset{\sim}{S}_T(\omega) = |\Gamma(\omega, -\omega)|^2 + |\Gamma(\omega,\omega)|^2 \geq E^2\{\underset{\sim}{S}_T(\omega)\}$$

$$\text{Var } \underset{\sim}{S}_T(0) = 2|\Gamma(0,0)|^2 = 2\{E^2\{\underset{\sim}{S}_T(0)\}$$

The remaining part of the problem is more difficult. We outline the proof (For details see Papoulis, Signal Analysis). From (i) and the convolution theorem it follows that

$$\Gamma(u,v) = \int_{-\infty}^{\infty} \frac{\sin T(u + v - \alpha)}{\pi\, T(u + v - \alpha)} \frac{\sin T\alpha}{\alpha} S(v - \alpha)d\alpha$$

If $S(\omega)$ is nearly constant in an interval of length $1/T$, then it can be taken outside the integral sign. Hence,

$$\Gamma(u,v) \simeq S(v)\int_{-\infty}^{\infty} \frac{\sin T(u + v - \alpha)}{\pi\, T(u + v - \alpha)} \frac{\sin T\alpha}{\alpha} d\alpha = S(v)\, \frac{\sin T(u + v)}{T(u + v)}$$

This yields

$$\Gamma(\omega,-\omega) \simeq S(\omega) < \Gamma(\omega,\omega) \simeq S(\omega)\, \frac{\sin 2\,T\omega}{T\omega} \xrightarrow[\omega T\to\infty]{} 0$$

$$\text{Var } \underset{\sim}{S}_T(\omega) \leq 2\, E^2\{\underset{\sim}{S}_T(\omega)\} \xrightarrow[\omega T\to\infty]{} E^2\{\underset{\sim}{S}_T(\omega)\}$$

12-14 The function

$$\underset{\sim}{X}_c(\omega) = \int_{-T}^{T} c(t)\underset{\sim}{x}(t)e^{-j\omega t}dt$$

is the Fourier transform of the product

$$c(t)\underset{\sim}{x}_T(t) \qquad\qquad \underset{\sim}{x}_T(t) = \begin{cases} 1 & |t| < T \\ 0 & |t| > T \end{cases}$$

Hence, the function

$$2\,T\,\underset{\sim}{S}_T(\omega) = |\underset{\sim}{X}_c(\omega)|^2$$

is the Fourier transform of

$$\underset{\sim}{c}(t)\underset{\sim}{x}_T(t) * c(-t)\underset{\sim}{x}_T(-t)$$

$$= \int_{-T+|\tau|/2}^{T-|\tau|/2} c(t + \tfrac{\tau}{2})\underset{\sim}{x}_T(t + \tfrac{\tau}{2})c(t - \tfrac{\tau}{2})\underset{\sim}{x}(t - \tfrac{\tau}{2})dt$$

12-15 Since $C(-\tau) = C(\tau)$, it follows from (12-28) that for large T,

$$\text{Var}\ \underset{\sim}{R}_T(\lambda) \simeq \frac{1}{2T}\int_{-\infty}^{\infty} [C(\lambda+\tau)C(\lambda-\tau) + C^2(\tau)]\ d\tau$$

Since $S(\omega)$ is real, it follows from Parseval's formula and the pairs

$$C(\lambda+\tau) \leftrightarrow e^{j\lambda\omega}S(\omega) \qquad\qquad C(\lambda-\tau) \leftrightarrow e^{-j\lambda\omega}S(\omega)$$

that the above integral equals

$$\int_{-\infty}^{\infty} \left[e^{j\lambda\omega}\ S(\omega)e^{j\lambda\omega}S(\omega) + S^2(\omega)\right]\ d\omega$$

12-16 With $c = T - |\tau|/2$

$$\underset{\sim}{z}(t) = \underset{\sim}{x}(t + \frac{\tau}{2})\underset{\sim}{x}(t - \frac{\tau}{2}) \qquad E\{\underset{\sim}{R}_T(\tau)\} = R(\tau)(1 - \frac{|\tau|}{T})$$

(7-37) yields

$$E\{\underset{\sim}{z}(t_1)\underset{\sim}{z}(t_2)\} - E\{\underset{\sim}{z}(t_1)\}E\{\underset{\sim}{z}(t_2)\}$$

$$= R^2(t_1 - t_2) + R(t_1 - t_2 + \tau)R(t_1 - t_2 - \tau)$$

$$4\,T^2 \operatorname{Var} \underset{\sim}{R}_T(\tau) = \int_{-c}^{c}\int_{-c}^{c} [R^2(t_1 - t_2) + R(t_1 - t_2 + \tau)R(t_1 - t_2 - \tau)]dt_1 dt_2$$

$$= \int_{-2c}^{2c} [R^2(\alpha) + R(\alpha + \tau)R(\alpha - \tau)](2T - |\tau| - |\alpha|)d\alpha$$

12-17 Equating coefficients of z^k in (12-98), we obtain

$$(1-K_N^2)\,\alpha_k^{N-1} = \alpha_k^N + K_N\,\alpha_{N-k}^N$$

12-18 $R[0] = 8 \qquad R[1] = 4$

From (13-67)

$$P_0 = 8 \qquad a_1^1 = K_1 = 0.5 \qquad P_1 = (1 - K_1^2)P_0 = 6$$

$$E_1(z) = 1 - 0.5z^{-1} \qquad S(\omega) = \frac{6}{|E_1(e^{j\omega})|^2}$$

12-19 From page (13-67)

$P_0 = 13 \qquad a_1^1 = K_1 = \frac{5}{13} \qquad P_1 = \frac{144}{13}$

$P_1 K_2 = R[2] - a_1^1 R[1] \qquad K_2 = \frac{1}{144}$

$a_1^2 = \frac{55}{144} \qquad a_2^2 = \frac{1}{144} \qquad P_2 = \frac{1595}{144}$

$$S_{MEM}(\omega) = \frac{1595 \times 144}{\left|144 - 55e^{-j\omega T} - e^{-j2\omega T}\right|^2}$$

From (12-119)

$$\begin{vmatrix} 13-q & 5 & 2 \\ 5 & 13-q & 5 \\ 2 & 5 & 13-q \end{vmatrix} = 0 \qquad q_o = 14-\sqrt{51} \approx 6.86$$

Inserting the modified data 6.14, 5, 2 into the Yule-Walker equations (12-82), we obtain

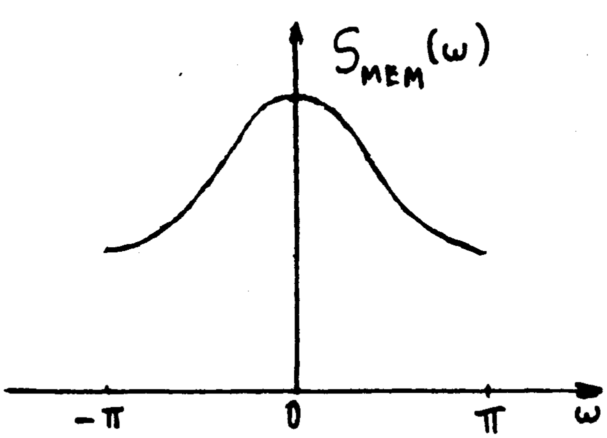
$a_1^2 \approx 4.07 \qquad a_2^2 = -1 \qquad E_2(z) = 1 - 4.07z^{-1} + z^{-2}$

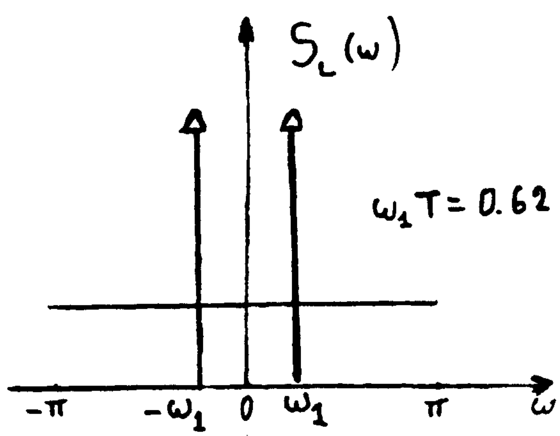
$E_2(z) = 1 - 4.07z^{-1} + z^{-2} \qquad z_{1,2} \approx e^{\pm j0.62}$

Solving (12-91) we obtain

$R_L[m] = 6.86\,\delta[m] + 3.07 \cos 0.62m$

$S_L(\omega) = 6.86 + \frac{2\pi}{T} \times 3.07\,[\delta(\omega - 0.62) + \delta(\omega + 0.62)]$





12.20 (a) Let $z = e^{j\theta_1}$ represent one of the roots of the Levinsion Polynomial $P_n(z)$ that lie on the unit circle. In that case

$$P_n\left(e^{j\theta_1}\right) = 0$$

and substituting this into the recursion equation (12-177) we get

$$|s_n| = \left|\frac{P_{n-1}\left(e^{j\theta_1}\right)}{\tilde{P}_{n-1}\left(e^{j\theta_1}\right)}\right| = 1$$

so that

$$s_n = e^{j\alpha}.$$

Let

$$P_{n-1}\left(e^{j\theta}\right) = R(\theta)\, e^{j\psi(\theta)}$$

and since $P_{n-1}(z)$ is free of zeros in $|z| \leq 1$, we have $R(\theta) > 0, 0 < \theta < 2\pi$, and once again substituting these into (12-177) we obtain

$$\begin{aligned}
\sqrt{1 - s_n^2}\, P_n\left(e^{j\theta}\right) &= R(\theta)\, e^{j\psi(\theta)} - e^{j(\theta+\alpha)}\, e^{j(n-1)\theta}\, R(\theta)\, e^{-j\psi(\theta)} \\
&= R(\theta)\left[e^{j\psi(\theta)} - e^{j(n\theta+\alpha)}\, e^{-j\psi(\theta)}\right] \\
&= 2j\, R(\theta)\, e^{j(n\theta+\alpha)/2}\, sin\left(\psi(\theta) - \frac{n\theta}{2} - \frac{\alpha}{2}\right).
\end{aligned}$$

Due to the strict Hurwitz nature of $P_{n-1}(z)$, as θ varies from 0 to 2π, there is no net increment in the phase term $\psi(\theta)$, and the entire argument of the sine term above increases by $n\pi$. Consequently $P_n(e^{j\theta})$ equals zero atleast at n distinct points $\theta_1, \theta_2, \cdots \theta_n, \quad 0 < \theta_i < 2\pi$. However $P_n(z)$ is a polynomial od degree n in z and can have atmost n zeros. Thus all the above zeros are simple and they all lie on the unit circle.

(b) Suppose $P_n(z)$ and $P_{n-1}(z)$ has a common zero at $z = z_0$. Then $|z_0| > 1$ and from (12-137), we get

$$z_0\, s_n\, \tilde{P}_{n-1}(z_0) = 0$$

which gives $s_n = 0$, since $\tilde{P}_{n-1}(z_0) \neq 0$, ($\tilde{P}_{n-1}(z)$ has all its zeros in $|z| < 1$). Hence $s_n \neq 0$ implies $P_n(z)$ and $P_{n-1}(z)$ do not have a common zero.

12.21 Substituting $s_n = \rho^n, |\rho| < 1$ in (12-177) we get

$$\sqrt{1-\rho^{2n}}\, P_n(z) = P_{n-1}(z) - (z\rho)^n\, P^*_{n-1}(1/z^*)$$

Let $x = z\rho$ and

$$P_n(z) = P_n(x/\rho) \triangleq A_n(x)$$

so that the above iteration reduces to

$$\begin{aligned} \sqrt{1-\rho^{2n}}\, A_n(x) &= A_{n-1}(x) - x^n\, A^*_{n-1}(1/x^*) \\ &= A_{n-1}(x) - x\, \tilde{A}_{n-1}(x) \end{aligned}$$

From problem (12-20), the polynomial $A_n(x)$ has all its zeros on the unit circle (since $s_n = 1$). i.e.,

$$x_k = e^{j\theta_k} = z_k\, \rho.$$

Hence the zeros $z_k = (1/\rho)e^{j\theta_k}$ or $|z_k| = 1/\rho$. (The zeros of $P_n(z)$ lie on a circle of radius $1/\rho$).

12.22 The Levinson Polynomials $P_n(z)$ satisfy the recusion in (12-177). Define $s'_n = \lambda^n s_n$, $|\lambda| = 1$, and replacing s_n by s'_n and $P_n(z)$ by $P'_n(z)$ in (12-177) we get

$$\begin{aligned} P'_n(z) &= P'_{n-1}(z) - z s'_n\, \tilde{P'}_{n-1}(z) \\ &= P'_{n-1}(z) - (z\lambda)^n\, s_n\, P'^*_{n-1}(z)\,(1/z^*) \end{aligned}$$

Let $y = z\lambda$ and define $P'_n(y/\lambda) = A_n(y)$ so that the above recursion simplifies to

$$\begin{aligned} A_n(y) &= A_{n-1}(y) - y^n\, s_n\, A^*_{n-1}(1/y^*) \\ &= A_{n-1}(y) - y s_n\, A^*_{n-1}(y) \end{aligned}$$

and on comparing with (12-177), we notice that $A_n(y) = P_n(y) = P_n(\lambda z)$. Thus $P_n(\lambda z)$ represents the new set of Levinson Polynomials.

12.23 (a) In this case

$$S(\theta) = \left|H(e^{j\theta})\right|^2 = |1 - e^{j\theta}|^2 = 2 - e^{j\theta} - e^{-j\theta} = 2(1 - cos\theta)$$

so that $r_0 = 2, r_1 = -1, r_k = 0, |k| \geq 2$. Substituing these values into (9-196) and taking the determinant of the tridiagonal matrix $\mathbf{T}_n$ we obtain the recursion

$$|\mathbf{T}_n| = \Delta_n = 2\Delta_{n-1} - \Delta_{n-2}$$

where $\Delta_0 = 2, \Delta_1 = 3$. Let $D(z) = \sum_{n=0}^{\infty} \Delta_n \, z^n$ so that the above recursion gives

$$D(z) = \frac{2 - z}{(1 - z)^2} = \frac{1}{1 - z} + \frac{1}{(1 - z)^2} = \sum_{n=0}^{\infty}(n + 2)\, z^n$$

and hence we get

$$\Delta_n = n + 2, \qquad n \geq 0.$$

Using (12-192) and (9-196) we get

$$s_n = (-1)^{n-1} \, \frac{\Delta_n^{(1)}}{\Delta_{n-1}} = \frac{(-1)^{n-1}(-1)^n}{n + 1} = -\frac{1}{n + 1}, \quad k \geq 1.$$

(b) The new set of reflection coefficient $s'_k = -s_k$ switches around the Levinson Polynomials $P_n(z)$ and $Q(z)$, and hence it follows that they correspond to the positive-real function

$$Z'(z) = \frac{2}{1 - z}$$

which gives $r'_0 = 2, r'_k = 1, \, k \geq 1$.

CHAPTER 13

13-1
$$\hat{\mathbf{s}}\left(t - \frac{T}{2}\right) = a\,\mathbf{s}(t) + b\,\mathbf{s}(t-T)$$

$$\mathbf{s}\left(t - \frac{T}{2}\right) - [a\,\mathbf{s}(t) + b\,\mathbf{s}(t-T)] \perp \mathbf{s}(t),\ \mathbf{s}(t-T)$$

$$R(T/2) = a\,R(0) + b\,R(T)$$

$$a = b = \frac{R(T/2)}{R(0)+R(T)} = \frac{e^{-1/2}}{1+e^{-1}}$$

$$R(T/2) = a\,R(T) + b\,R(0)$$

$$P = E\{[\mathbf{s}(t - \tfrac{T}{2}) - \hat{\mathbf{s}}(t - \tfrac{T}{2})]\mathbf{s}(t - \tfrac{T}{2})\}$$

$$= R(0) - aR(T/2) - bR(T/2) = R(0) - \frac{R^2(T/2)}{R(0)+R(T)} = \frac{I}{1+e^{-1}}$$

13-2
$$\int_0^T \mathbf{s}(t)dt - [a\,\mathbf{s}(0) + b\,\mathbf{s}(T)] \perp \mathbf{s}(0),\ \mathbf{s}(T)$$

$$\int_0^T R(t)dt = aR(0) + bR(T) \qquad \int_0^T R(T-t)dt = aR(T) + bR(0)$$

The above two integrals are equal. Hence,

$$a = b = \frac{\int_0^T R(t)dt}{R(0)+R(T)}$$

13-3
$$\hat{\mathbf{s}}'(t) = a\,\mathbf{x}(t) + b\,\mathbf{x}(t-\tau)$$

$$\mathbf{s}'(t) - [a\,\mathbf{x}(t) + b\,\mathbf{x}(t-\tau)] \perp \mathbf{x}(t),\ \mathbf{x}(t-\tau)$$

$$R_{s'x}(0) = a\,R_{xx}(0) + b\,R_{xx}(\tau) \qquad R_{s'x}(0) = R_{s's}(0) = 0$$

$$R_{s'x}(\tau) = a\,R_{xx}(\tau) + b\,R_{xx}(0) \qquad R_{xx}(\tau) = R_{ss}(\tau) + R_{\nu\nu}(\tau)$$

For small τ

$$R_{s'x}(\tau) = R_{s's}(\tau) = R'_{ss}(\tau) \simeq \tau R''_{ss}(0) \qquad R_{xx}(\tau) \simeq R_{xx}(0) + \tau^2 R''_{xx}(0)/2$$

Hence,

$$a = -b + O(\tau^2) \qquad \tau R''_{ss}(0) = a\,\tau^2\,R''_{xx}(0)/2 + O(\tau^3)$$

13-4 It suffices to show that, for any m,

$$E\left\{\left[\mathbf{x}(t) - \sum_{n=-\infty}^{\infty} \frac{\sin(\sigma t - n\pi)}{\sigma t - n\pi}\,\mathbf{x}(nT)\right]\mathbf{x}(mT)\right\} = 0$$

The left side equals

$$R(t-mT) - \sum_{n=-\infty}^{\infty} \frac{\sin(\sigma t - n\pi)}{\sigma t - n\pi} R(nT - mT)$$

From the sampling theorem (10-140) it follows that this is zero because the Fourier transform

$$e^{-jmT}S(\omega)$$

of $R(t-mT)$ is zero for $|\omega| > \sigma$.

13-5 Since

$$\hat{E}\{\mathbf{x}(t+\lambda)|\mathbf{s}(t)\} = a\mathbf{s}(t) \qquad a = R(\lambda)/R(0)$$

it follows from the assumption that

$$\mathbf{s}(t+\lambda) - a\,\mathbf{s}(t) \perp \mathbf{s}(t-\tau)$$

Hence

$$R(\lambda+\tau) = \frac{R(\lambda)}{R(0)} R(\tau) \tag{i}$$

The only continuous function satisfying the above is an exponential. This is easily shown if we assume that $R(\lambda)$ is differentiable for $\lambda > 0$. Differentiating (i) with respect to λ and setting $\lambda = 0^+$, we obtain

$$R'(\tau) + \alpha R(\tau) = 0 \qquad \alpha = -R'(0^+)/R(0) \qquad \tau > 0$$

This yields $R(\tau) = Ie^{-\alpha\tau}$ for $\tau > 0$.

13-6 Given:

$$E\{\mathbf{y}_n\} = 0 \qquad \mathbf{x}_n = \mathbf{y}_1 + \cdots + \mathbf{y}_n = \mathbf{x}_{n-1} + \mathbf{y}_n$$

Furthermore the RVs $\mathbf{y}_n$ are independent. Hence, $\mathbf{y}_n$ is independent of $\mathbf{x}_{n-1}, \ldots, \mathbf{x}_1$. This yields

$$E\{\mathbf{x}_n|\mathbf{x}_{n-1},\ldots,\mathbf{x}_1\} = E\{\mathbf{x}_{n-1} + \mathbf{y}_n|\mathbf{x}_{n-1},\ldots,\mathbf{x}_1\}$$

$$= E\{\mathbf{x}_{n-1}|\mathbf{x}_{n-1},\ldots,\mathbf{x}_1\} + E\{\mathbf{y}_n\} = \mathbf{x}_{n-1}$$

13-7 (a) If $\hat{E}\{\mathbf{x}_n|\mathbf{x}_{n-1},\ldots,\mathbf{x}_1\} = \mathbf{x}_{n-1}$, then

$$= \mathbf{x}_n - \mathbf{x}_{n-1} \perp \mathbf{x}_{n-1},\ldots,\mathbf{x}_1$$

From this it follows that the RVs $\mathbf{y}_n = \mathbf{x}_n - \mathbf{x}_{n-1}$ are orthogonal and

$$\mathbf{x}_n = \mathbf{y}_n + \mathbf{x}_{n-1} = \mathbf{y}_n + \mathbf{y}_{n-1} + \cdots + \mathbf{y}_1 \qquad \text{(i)}$$

Conversely, if (i) is true and the RVs $\mathbf{y}_n$ are orthogonal, then

$$\mathbf{x}_n - \mathbf{x}_{n-1} = \mathbf{y}_n \perp \mathbf{x}_{n-1},\ldots,\mathbf{x}_1$$

(b) $E\{\mathbf{x}_n^2\} = E\{[(\mathbf{x}_n - \mathbf{x}_{n-1}) + \mathbf{x}_{n-1}]^2\}$

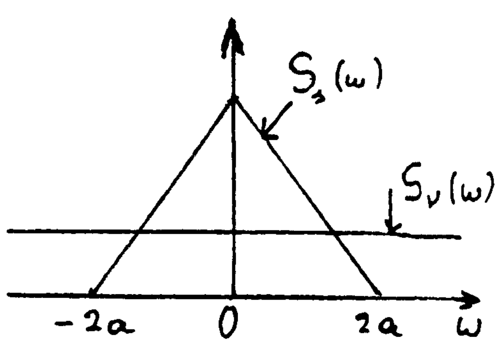
$$= E\{(\mathbf{x}_n - \mathbf{x}_{n-1})^2\} + E\{\mathbf{x}_{n-1}^2\} \geq E\{\mathbf{x}_{n-1}^2\}$$

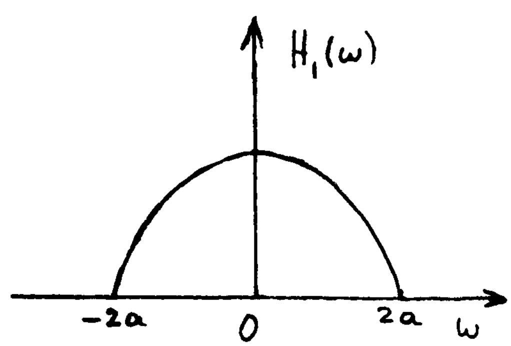
for any n.

13-8 The Fourier transform $S_s(\omega)$ of the function

$$R_s(\tau) = A\frac{\sin^2 a\tau}{\tau^2}$$

is a triangle as shown





And since $S_\nu(\omega) = N$, (13-16) yields

$$H_1(\omega) = \frac{S_s(\omega)}{S_s(\omega) + S_\nu(\omega)} = \frac{Aa\pi(1 - |\omega|/2a)}{Aa\pi(1 - |\omega|/2a) + N}$$

We show next that

$$H_2(\omega) = j\omega H_1(\omega)$$

<u>Proof</u>

$$\varepsilon(t) = s(t) \quad \hat{s}'(t) \perp x(\xi) \quad \text{all } t,\xi. \quad \text{Hence } R_{\varepsilon x}(\tau) = 0.$$

This yields [see (9-131)]

$$R_{\varepsilon' x}(\tau) = R'_{\varepsilon x}(\tau) = 0, \text{ hence } s'(t) - \hat{s}'(t) \perp x(\xi)$$

$$\xrightarrow[s(t)]{} \boxed{j\omega} \xrightarrow[s'(t)]{} \qquad\qquad \xrightarrow[s(t)]{} \boxed{H_1(\omega)} \xrightarrow[\hat{s}(t)]{} \boxed{j\omega} \xrightarrow[\widehat{s'}(t)]{}$$

In other words, the estimate of $s'(t)$ equals the derivative of the estimate $\hat{s}(t)$ of $s(t)$. This follows from Prob. 13-9 with $T(\omega) = j\omega$.

13-9 We wish to show that the estimator of

$$y(t) = \int_{-\infty}^{\infty} p(t-\alpha)s(\alpha)\,d\alpha \qquad\qquad p(t) \leftrightarrow T(\omega)$$

equals

$$\hat{y}(t) = \int_{-\infty}^{\infty} p(t-\alpha)\hat{s}(\alpha)\,d\alpha$$

where $\hat{s}(t)$ is the estimator of $s(t)$.

<u>Proof</u>. Clearly

$$E\{[s(t) - \hat{s}(t)]x(\xi)\} = 0 \qquad\qquad \text{all } t,\xi$$

Hence

$$E\{[y(t) - \hat{y}(t)]x(\xi)\}$$

$$= \int_{-\infty}^{\infty} p(t-\alpha)E\{[s(\alpha) - \hat{s}(\alpha)]x(\xi)\}\,d\alpha = 0$$

13-10 [See (13-46) and beyond]

(a) $S(s) = \dfrac{1}{s^4+1} = \dfrac{1}{(s^2+\sqrt{2}\,s+1)(s^2-\sqrt{2}\,s+1)}$

(b) $L(s) = \dfrac{1}{(s+\alpha)^2+\beta^2} \qquad \ell(t) = \dfrac{1}{\beta}\, e^{-\alpha t} \sin\beta t U(t) \qquad \alpha = \beta = \dfrac{1}{\sqrt{2}}$

(c) $h_1(t) = \dfrac{1}{\beta}\, e^{-\alpha\lambda}\, e^{-\alpha t} \sin\beta(t+\lambda)U(t)$

$$H_1(s) = \frac{1}{\beta}\, e^{-\alpha\lambda}\, \frac{(s+\alpha)\sin\beta\lambda + \beta\cos\beta\lambda}{(s+\alpha)^2+\beta^2}$$

$$b_0 = e^{-\alpha\lambda}(\cos\beta\lambda + \frac{\alpha}{\ }\sin\beta\lambda)$$

$$H(s) = \frac{H_1(s)}{L(s)} = b_0 + b_1 s$$

$$b_1 = \frac{\sin\beta\lambda}{\beta}\, e^{-\alpha\lambda}$$

13-11 (a) The given equation is the Wiener-Hopf equation (13-40) for the prediction problem with $\lambda = \ln 2$. We can, therefore, use the method described after (13-46):

$$S(s) = \frac{3}{1-s^2} + \frac{22}{9-s^2} = \frac{49-25s^2}{(1-s^2)(9-s^2)}$$

$$L(s) = \frac{7+5s}{(1+s)(3+s)} \qquad \ell(t) = (e^{-t}+4e^{-3t})U(t)$$

$$h_i(t) = (e^{-\ln 2}e^{-t}+4e^{-3\ln 2}e^{-3t})U(t)$$

$$H_i(s) = \frac{1/2}{1+s} + \frac{4/8}{3+s} = \frac{2+s}{(1+s)(3+s)} \qquad H(s) = \frac{2+s}{7+5s}$$

(b)

$$H(s) = \frac{N(s)}{D(s)} \qquad \frac{N(s)-2^sD(s)}{D(s)}\, L(s)L(-s) = Y(s)$$

Since $Y(s)$ is analytic for $\underline{\mathrm{Re}}\, s \le 0$, all roots of $D(s)$ must be cancelled by the zeros of $L(s)$, hence, $D(s) = 7+5s$. Similarly the poles $s = -1$ and $s = -3$ of $L(s)$ must be cancelled by the zeros of the term $N(s) - 2^sD(s)$. With $N(s) = As + B$, this yields

$$N(-1) - 2^{-1}D(-1) = -A + B - 2^{-1}(7-5) = 0 \qquad A = 1$$

$$N(-3) - 2^{-3}D(-3) = -3A + B - 2^{-3}(7-15) = 0 \qquad B = 2$$

$$H(s) = \frac{2+s}{7+5s}$$

(c) The Laplace transform of the function $R(\tau)$ in (a) equals

$$\frac{49-25s^2}{9-10s^2+s^4}$$

Hence (convolution theroem), the inverse transform $y(t)$ of $Y(s)$ equals

$$y(t) = \int_0^\infty h(\alpha)R(t-\alpha)\,d\alpha - R(t+\ln 2)$$

From the analyticity of $Y(s)$ for $\underline{\mathrm{Re}}\, s < 0$ it follows that $y(t) = 0$ for $t > 0$. Therefore, (b) gives a direct method for solving the Wiener-Hopf equation (13-40).

13-12 (a) The given equation is identical with equation (13-22) for the prediction problem with $r=1$. We can, therefore, use the method in (13-31)-(13-33):

$$S(z) = \frac{3}{5-2w} + \frac{8}{10-3w} = \frac{70-25w}{(5-2w)(10-3w)} \qquad w = z+\frac{1}{z}$$

$$L(z) = \frac{a - bz^{-1}}{(2-z^{-1})(3-z^{-1})} \qquad a = \sqrt{30}+\sqrt{5} \simeq 7.75 \qquad b = \sqrt{30}-\sqrt{5} \simeq 3.25$$

$$\ell[0] = \frac{a}{6} \simeq 1.3 \qquad H(z) = 1 - \frac{\ell[0]}{L(z)} \simeq \frac{0.41\,z^{-1} - 0.167\,z^{-2}}{1-0.42z^{-1}}$$

(b)

$$H(z) = \frac{N(z)}{D(z)} \qquad \frac{N(z) - zD(z)}{D(z)} L(z)L(z^{-1}) = Y(z)$$

Since $Y(z)$ is analytic for $|z|<1$, all roots of $D(z)$ must be cancelled by the zeros of $L(z)$, hence, $D(z) = 1-0.42\,z^{-1}$. Similarly, the poles $z=1/2$ and $z=1/3$ of $L(z)$ must be cancelled by the zeros of the term $N(z)-zD(z)$. With $N(z) = A+Bz^{-1}$, this yields

$$N(\tfrac{1}{2}) - \tfrac{1}{2}D(\tfrac{1}{2}) \simeq A+2B - 0.08 = 0 \qquad A \simeq 0.42$$

$$N(\tfrac{1}{3}) - \tfrac{1}{3}D(\tfrac{1}{3}) \simeq A+3B + 0.09 = 0 \qquad B = -0.17$$

$$H(z) \simeq \frac{0.42-0.17z^{-1}}{1-0.42z^{-1}}$$

The z transform of the sequence R_m in (a) equals

$$\frac{70-25w}{6w^2 - 35w + 50} \qquad w = z+z^{-1}$$

Hence, the inverse transform y_n of $Y(z)$ equals

$$y_n = \sum_{k=0}^{n} h_k R_{n-k} - R_{n+1}$$

From the analyticity of $Y(z)$ for $|z|<1$ it follows that $y_n = 0$ for $n \geq 0$. Therefore, (b) gives a direct method for solving (13-22).

13-13 A predictor is a stable function H(z) vanishing at ∞. Since $H(z) \to 0$ 1 as $z \to \infty$, we conclude that $E_N(z) = 1-H(z) \to 1$ and $E_N(z)H_a(z) \to 1$ as $z \to \infty$. From this and (13-25) it follows that the difference $1 - E_N(z)H_a(z)$ is a predictor and the MS error equals

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} | E_N(e^{j\omega}) |^2 S(\omega)d\omega = P$$

because $|A_a(e^{j\omega})| = 1$.

13-14 As we know, if

$$\underline{s}[n] = a_1\underline{s}[n-1] + \cdots + a_m\underline{s}[n-m] + \underline{\varepsilon}[n]$$

where $\underline{\varepsilon}[n]$ is white noise, then the one-step predictor of $\underline{s}[n]$ equals

$$\hat{\underline{s}}_1[n] = a_1\,\underline{s}[n-1] + \cdots + a_m\,\underline{s}[n-m]$$

We wish to show that the sum

$$\hat{\underline{s}}_2[n] = a_1\hat{\underline{s}}_1[n-1] + a_2\underline{s}[n-2] + \cdots + a_m\underline{s}[n-m]$$

is its two-step predictor. It suffices to show that

$$\underline{s}[n] - \hat{\underline{s}}_2[n] \perp \underline{s}[n-k] \qquad k \ge 2$$

<u>Proof</u>

$$\underline{s}[n] - \hat{\underline{s}}_2[n] = a_1\left(\underline{s}_1[n-1] - \hat{\underline{s}}_1[n-1]\right) + \underline{\varepsilon}[n]$$

This completes the proof because

$$\underline{s}_1[n-1] - \hat{\underline{s}}_1[n-1] \perp \underline{s}(n-k], \quad k \ge 2 \text{ and } \underline{\varepsilon}[n] \perp \underline{s}[n-k] \quad k \ge 1.$$

13-15 The Nth order MS estimation error P_N equals [see (13-66)]

$$P_N = \frac{\Delta_{N+1}}{\Delta_N}$$

This tends to the MS estimation error in (13-34). Hence,

$$\lim_{N\to\infty} \ln P_N = \frac{1}{2\sigma}\int_{-\sigma}^{\sigma} \ln S(\omega)d\omega = \lim_{N\to\infty} \ln \frac{\Delta_{N+1}}{\Delta_N}$$

To complete the proofs, we use (14-129)

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \ln \frac{\Delta_{n+1}}{\Delta_n} = \lim_{N\to\infty} \ln \frac{\Delta_{N+1}}{\Delta_N}$$

and the result follows because

$$\frac{1}{N}\sum_{n=1}^{N} (\ln \Delta_{n+1} - \ln \Delta_n) = \frac{\ln\Delta_{N+1}}{N} - \frac{\ln\Delta_1}{N}$$

and the last term tends to zero as $N\to\infty$.

13-16 $P_0 = R[0] = 15 \qquad R[1] = 10 \qquad R[2] = 5 \qquad R[3] = 0$

We use Levinson's algorithm [see (13-67)]

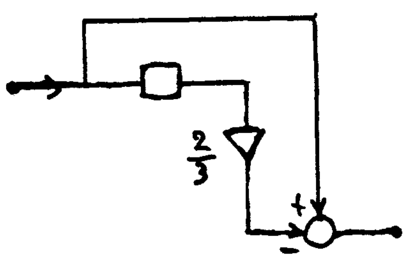
$P_0K_1 = R[1] \qquad K_1 = a_1^1 = \frac{2}{3} \qquad P_1 = (1-k_1^2)P_0 = \frac{25}{3}$

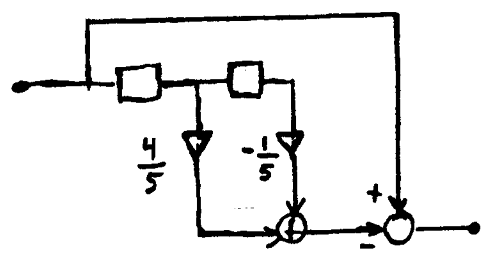
$P_1K_2 = R[2] - R[1]a_1^1 = -\frac{5}{3} \qquad K_2 = -\frac{1}{5}$

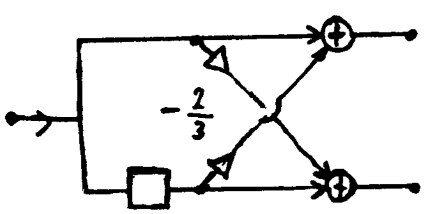
$a_1^2 = a_1^1 - K_2a_1^1 = \frac{4}{5} \qquad a_2^2 = -\frac{1}{5} \qquad P_2 = 8$

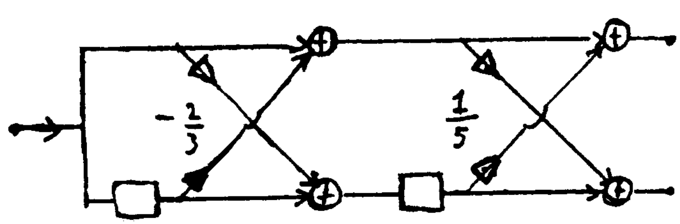
$P_2K_3 = R[3] - R[2]a_1^2 - R[1]a_2^2 \qquad K_3 = -\frac{1}{4}$

$a_1^3 = \frac{3}{4} \qquad a_2^3 = 0 \qquad a_3^3 = -\frac{1}{4} \qquad P_3 = 7.5$









13-17 $\quad P_0 = R[0] = 5 \qquad K_1 = 0.4 \qquad K_2 = 0.6 \qquad K_3 = 0.8$

$$R[1] = P_0K_1 = 2 \qquad a_1^1 = 0.4 \qquad P_1 = 4.2$$

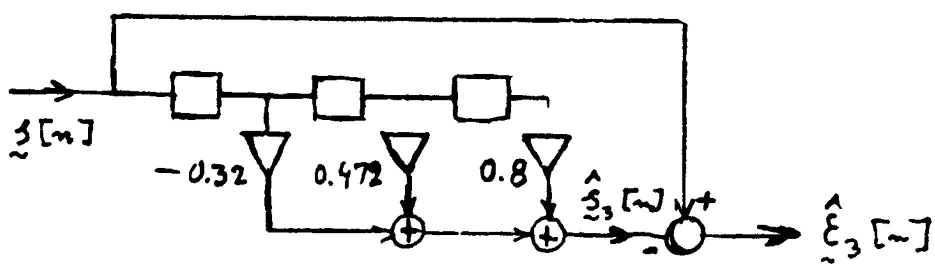
$$R[2] = R[1]a_1^1 + P_1K_2 = 3.32$$

$$a_1^2 = 0.16 \qquad a_2^2 = 0.6 \qquad P_2 = 2.688$$

$$R[3] = R[2]a_1^2 + R[1]a_2^2 + P_2K_3 = 3.8816$$

$$a_1^3 = a_1^2 - K_3a_2^2 = -0.32 \qquad a_2^3 = a_2^2 - K_3a_1^2 = 0.472 \qquad a_3^3 = 0.8$$

$$a_3^3 = 0.8 \qquad P_3 = 0.968$$



13-18

$$S_x(s) = \frac{4\lambda}{-s^2 + 4\lambda^2} + N = \frac{N(-s^2 + c^2)}{-s^2 + 4\lambda^2}$$

$$\Gamma_x(s) = \frac{s + 2\lambda}{\sqrt{N}\,(s + c)} \qquad c = 2\lambda\sqrt{1 + \frac{1}{\lambda N}}$$

and (13-104) yields

$$H_x(s) = \frac{c - 2\lambda}{s + c} \qquad h_x(t) = (c - 2\lambda)e^{-ct}U(t)$$

13-19 (a) $\hat{\varepsilon}_{N+m}[n+m] \perp \underset{\sim}{s}[n-k] \qquad k = -m+1,\dots,0,\dots,N$

$$\hat{\varepsilon}_N[n] = \underset{\sim}{s}[n] - a_1\underset{\sim}{s}[n-1] - \cdots - a_N\underset{\sim}{s}[n-N]$$

(b) $\check{\underset{\sim}{\varepsilon}}_{N+m}[n-m] \perp \underset{\sim}{s}[n+k] \qquad k = -m+1,\dots,0,\dots,N$

$$\check{\varepsilon}_N[n] = \underset{\sim}{s}[n] - a_1\underset{\sim}{s}[n+1] - \cdots - a_N\underset{\sim}{s}[n+N]$$

(c) $\check{\underset{\sim}{\varepsilon}}_{N+m}[n-N-m] \perp \underset{\sim}{s}[n+k] \qquad k = -N-m+1,\dots,-N,\dots,0$

$$\hat{\underset{\sim}{\varepsilon}}_N[n] = \underset{\sim}{s}[n] - a_1\underset{\sim}{s}[n-1] - \cdots - a_N\underset{\sim}{s}[n-N]$$

13-20

$$S_s(\omega) = \frac{2}{\omega^2+0.04} \qquad S_x(\omega) = \frac{5\omega^2+2.2}{\omega^2+0.04} \qquad L_x(s) = \sqrt{5}\,\frac{s+0.66}{s+0.2}$$

(a) From (13-16)

$$H(\omega) = \frac{2}{5\omega^2+2.2}$$

(b) From (13-104)

$$H_x(s) = 1 - \sqrt{5}\,\Gamma_x(s) = \frac{0.46}{s+0.66}$$

(c) Using (13-48)

$$L_s(s) = \frac{\sqrt{2}}{s+0.2} \qquad \ell(t) = \sqrt{2}\,e^{-0.2t}\,U(t)$$

$$h_i(t) = \sqrt{2}\,e^{-0.2(t+2)}U(t) \qquad H_i(s) = \frac{\sqrt{2}\,e^{-0.4}}{s+0.2}$$

$$H(s) = e^{-0.4} \qquad \hat{\underset{\sim}{s}}(t+2) = e^{-0.4}\underset{\sim}{s}(t)$$

(d) [see (13-99) and beyond]

$$S_{sx}(s) = \frac{1}{0.04 - s^2} \qquad \Gamma_x(s) = \frac{s+0.2}{\sqrt{5}(s+0.66)} \qquad S_{si_x}(s) = \sqrt{20}\,\frac{0.66-s}{s+0.2}$$

$$R_{si_x}(\tau) = \sqrt{20}\left[\delta(\tau) + \frac{0.86}{s+0.2}\right] \qquad h_{i_x}(\tau) = 0.86\sqrt{20}\; e^{-0.2(t+2)}U(t)$$

$$H_{i_x}(s) = \frac{0.86\ 20\ e^{-0.4}}{s+0.2} \qquad H_x(s) = \frac{1.72\ e^{-0.4}}{s+0.66}$$

13-21 As in Example 13-2 with $N_0 = 1.8$, $N = 5$, $a = 0.8$

$$S_s(z) = \frac{1.8}{(1-0.8\ z^{-1})(1-0.8z)} \qquad L_s(z) = \frac{\sqrt{1.8}}{1-0.8\ z^{-1}}$$

$$S_x(z) = \frac{8(1-0.5\ z^{-1})(1-0.5z)}{(1-0.8\ z^{-1})(1-0.8z)} \qquad L_x(z) = \frac{\sqrt{8}(1-0.5\ z^{-1}}{1-0.8\ z^{-1}}$$

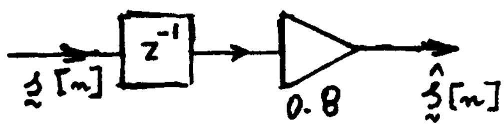
(a)
$$H(z) = \frac{9}{40(1-0.52\ z^{-1})(1-0.5z)} \qquad h[n] = 2\times 2^{-|n|}$$

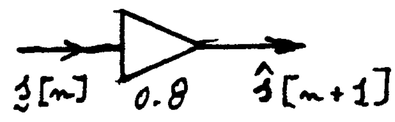
(b) From (13-114) with $\ell_x[0] = \sqrt{8}$

$$H_x(z) = \frac{3/8}{1-0.5\ z^{-1}} \qquad h_x[n] = \frac{3}{8}\times 0.5\ U[n]$$

(c) From (13-33) with $\ell[0] = \sqrt{1.8}$

$$H\ (z) = 1 - \frac{\ell[0]}{L_s(z)} = 0.8\ z^{-1}$$





(d) The power spectrum of the estimate $\hat{\underline{s}}_0[n]$ of $\underline{s}[n]$ obtained in (b) equals

$$S_{\hat{s}_0}(z) = S_x(z)H_x(z)H_x(z^{-1}) = \frac{9/8}{(1-0.8\ z^{-1})(1-0.8\ z)}$$

Hence, $$L_{\hat{s}_0}(z) = \frac{\sqrt{9/8}}{1-0.8\ z^{-1}}$$

Therefore, the pure predictor of $\hat{s}_0[n]$ equals [see (13-33)]

$$\hat{H}^1(z) = 1 - \frac{\ell[0]}{L(z)} = 0.8\ z^{-1}$$

And (13-117) yields

$$H_x^1(z) = H_x^0(z)\hat{H}_1(z) = \frac{0.3\,z^{-1}}{1-0.5\ z^{-1}}$$

13-22

$$R_s[m] = 5\times 0.8^{|m|} \longleftrightarrow \frac{1.8}{(1-0.8\,z^{-1})(1-0.8\,z)}$$

Hence, as in (13-135) with $V_n = 1.8$, $N_n = 5$. And (13-143) yields

$$F_n = 0.64\,F_{n-1} + V_n\,G_{n-1} \qquad F_0 = V_0 N_0 = 9$$

$$5\,G_n = 0.64\,F_{n-1} + 6.5\,G_{n-1} \qquad G_0 = V_0 + N_0 = 6.8$$

Solving, we obtain

$$F_n = 12(1.6)^n - 3(0.4)^n \qquad G_n = 6.4(1.6)^n + 0.4(0.4)^n$$

$$P_n = \frac{F_n}{G_n} \xrightarrow[n\to\infty]{} \frac{12}{6.4} = 1.875$$

This agrees with Prob. 13-21c because the MS error of the Wiener filter equals

$$P = R_s(0) - \sum_{k=0}^{\infty} R_s[k]h_x[k] = 5 - \sum_{k=0}^{\infty} 5\times 0.8^m \times \frac{3}{8}\times 0.5^m = 1.875$$

13-23 $R_s(\tau) = 5\, e^{-0.2|\tau|}$ $\qquad R(\tau) = \frac{10}{3}\,\delta(\tau)$

$$S_s(\omega) = \frac{2}{\omega^2 + 0.2^2} \qquad A(t) = 0.2 \qquad V(t) = 2 \qquad N(t) = \frac{10}{3}$$

From (13-159)

$$F'(t) = 0.2\,F(t) + 2G(t) \qquad G'(t) = 0.3\,F(t) + 0.2G(t)$$

Case 1. If $\mathbf{s}(0) = 0$, then $P(0) = F(0) = 0$, $G(0) = 1$

Solving, we obtain

$$P(t) = \frac{F(t)}{G(t)} = \frac{1.25\, e^{0.8t} - 1.25\, e^{-0.8t}}{0.625\, e^{0.8t} + 0.375\, e^{-0.8t}}$$

Case 2. If $\mathbf{s}(t)$ is stationary, then $F(0) = P(0) = R_s(0) = 5$

$$P(t) = \frac{F(t)}{G(t)} = \frac{5\, e^{0.8t} + 3\, e^{-0.8t}}{2.5\, e^{0.8t} - 0.9\, e^{-0.8t}}$$

13-24 The sequences $\hat{q}_N[n]$ and $\check{q}_N[n]$ are the responses of the filters

$$\hat{E}_N(z) = 1 - \sum_{k=1}^{N} a_k^N z^{-k} \qquad \check{E}_N(z) = z^{-N}\hat{E}_N(1/z)$$

respectively, with input R[m] (see Fig. 13-11a). Hence,

$$\hat{q}_N[m] = R[m] - \sum_{k=1}^{N} R[m-k]a_k^N$$

$$\check{q}_N[m] = \hat{q}_N[N-m] = R[m-N] - \sum_{k=1}^{N} R[m-N+k]a_k^N$$

From this and the Yule-Walker equation (13-65) it follows that

$$\hat{q}_N[m] = \check{q}_N[N-m] = 0 \text{ for } 1 \le m \le N-1$$

$$\hat{q}_N[0] = \check{q}_N[N] = P_N$$

This completes the proof.

CHAPTER 14

14-1 It suffices to show that [see (14-41)]

$$H(A\cdot B|B_j) = H(A|B_j)$$

Since

$$A_iB_kB_j = \begin{cases} A_iB_j & k=j \\ \{\emptyset\} & k\neq j \end{cases} \qquad \text{and } P(A_iB_j|B_j) = P(A_i|B_j)$$

(14-40) yields

$$H(A\cdot B|B_j) = -\sum_{i,k} P(A_iB_k|B_j)\log P(A_iB_k|B_j)$$

$$= -\sum_i P(A_i|B_j)\log P(A_i|B_j) = H(A|B_j)$$

14-2 If $\alpha<\beta$, then $\phi'(\alpha)>\phi'(\beta)$ because

$$\phi'(\alpha)-\phi'(\beta) = \log(\beta/\alpha) > 0. \quad \text{Hence,}$$

$$\int_a^b \phi'(\alpha)d\alpha > \int_{a+c}^{b+c} \phi'(\alpha)d\alpha \qquad c>0$$

This yields

$$\phi(p_1+p_2)-\phi(p_1) = \int_{p_1}^{p_1+p_2}\phi'(\alpha)d\alpha < \int_0^{p_2}\phi'(\alpha)d\alpha = \phi(p_2)$$

Similarly

$$\phi(p_1+\varepsilon)-\phi(p_1)-\phi(p_2)+\phi(p_2-\varepsilon)$$

$$= \int_{p_1}^{p_1+\varepsilon}\phi'(\alpha)d\alpha - \int_{p_2-\varepsilon}^{p_2}\phi'(\alpha)d\alpha > 0$$

14-3 Applying the identity

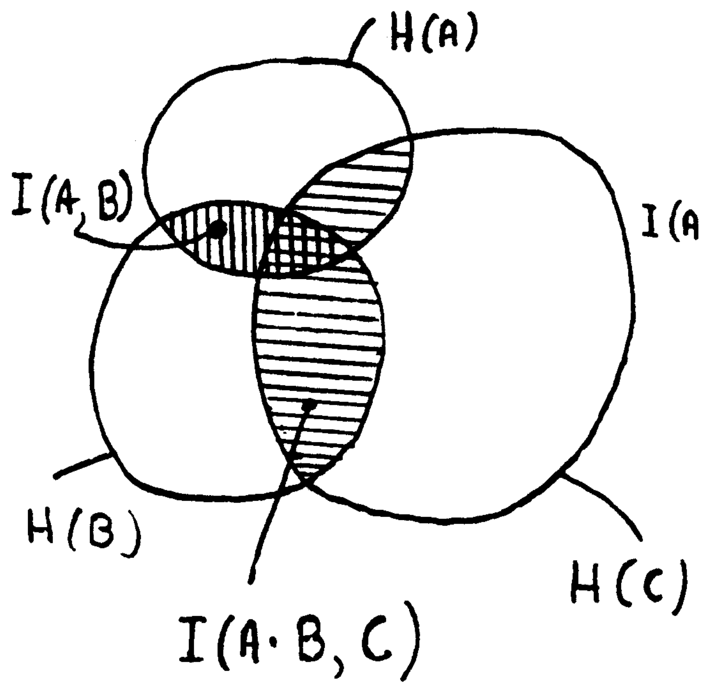
$$H(A_1 \cdot A_2) = H(A_1) + H(A_2|A_1) \tag{i}$$

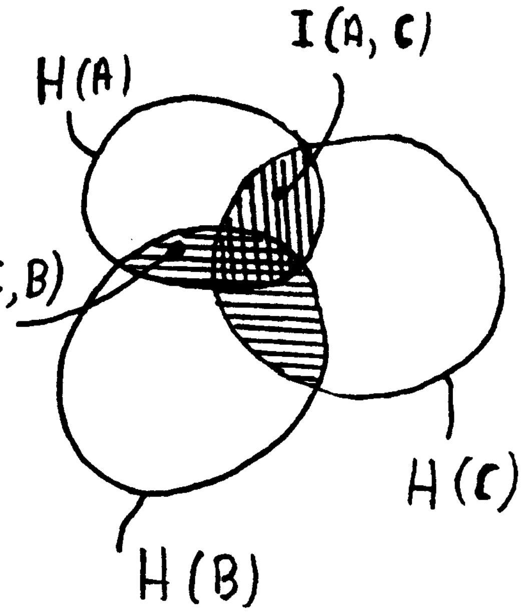
to the partitions $A_1 = A$, $A_2 = B \cdot C$ and $A_1 = A \cdot B$, $A_2 = C$, we obtain the first line. The second line follows from the first [see (i)]. The third line is a consequence of the first two.

14-4 It follows if we apply the identity

$$I(A_1, A_2) = H(A_1) + H(A_2) - H(A_1 \cdot A_2)$$

to the partitions $A_1 = A \cdot B$, $A_2 = C$.





14-5 **(a)** From (14-53)

$$I(A,B \cdot C) = H(A) + H(B \cdot C) - H(A \cdot B \cdot C)$$

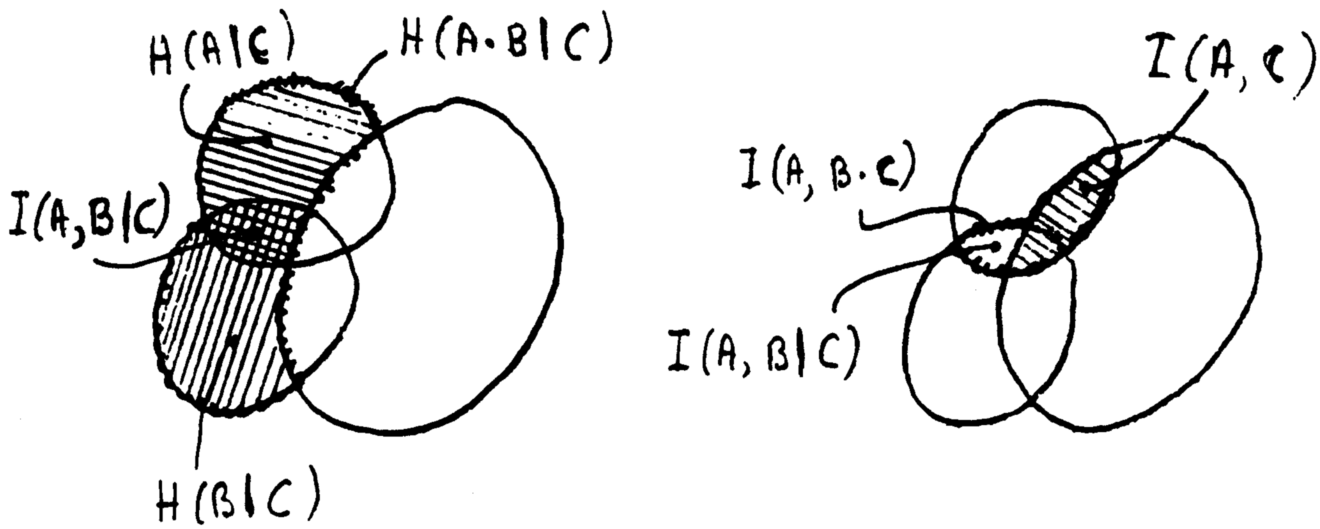
$$I(A,C) = H(A) + H(C) - H(A \cdot C)$$

and since (see Prob. 14-4)

$$H(A \cdot B \cdot C) - H(A \cdot C) = H(A \cdot B|C) - H(A|C)$$

we conclude with (14-49) that

$$I(A,B \cdot C) - I(A \cdot C) = H(B|C) + H(A|C) - H(A \cdot B|C)$$



(b) If $B \cdot C$ is observed, then the resulting prediction in the uncertainty of A equals $I(A, B \cdot C)$. But, if $B \cdot C$ is observed, then C is observed, hence, the reduction in the uncertainty of A is at least $I(A,C)$. Hence

$$I(A,B \cdot C) \geq I(A,C)$$

with equality only if $I(A,B|C) = 0$, i.e., if in the subsequence of trials in which C occurred, knowledge of the occurrence of B gives no information about A.

14-6 The partition $H(A^3)$ has eight elements with respective probabilities

$$p^3, p^2q, p^2q, p^2q, pq^2, pq^2, pq^2, q^3$$

Hence

$$H(A^3) = -p^3\log p^3 - 3p^2q\log p^2q - 3pq^2\log pq^2 - q^3\log q^3$$

$$= -3p(p^2 + 2pq + q^2)\log p - 3q(p^2 + 2pq + q^2)\log q$$

$$= -3p\log p - 3q\log q = 3H(A)$$

14-7 The density of the RV $\underset{\sim}{w} = \underset{\sim}{x} + a$ equals $f_x(w-a)$. Hence,

$$H(\underset{\sim}{x} + a) = -\int_{-\infty}^{\infty} f_x(w-a)\log f_x(w-\alpha)\,dw$$

$$= -\int_{-\infty}^{\infty} f_x(x)\log f_x(x)\,dx = H(\underset{\sim}{x})$$

The joint density of the RVs $\underset{\sim}{x}$ and $\underset{\sim}{z} = \underset{\sim}{x} + \underset{\sim}{y}$ equals $f_{xy}(x, z-x)$. Hence [see (14-90)]

$$H(\underset{\sim}{z}|\underset{\sim}{x}) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x, z-x)\log f_{xy}(x, z-x)/f_x(x)\,dx\,dz$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{xy}(x,y)\log f_{xy}(x,y)/f_x(x)\,dx\,dy = H(\underset{\sim}{y}|\underset{\sim}{x})$$

14-8 The RVs $\underset{\sim}{x}$ and $\underset{\sim}{y}$ take the values x_i and y_j respectively then $\underset{\sim}{z} = x_i + y_j$ iff $\underset{\sim}{x} = x_i$ and $\underset{\sim}{y} = y_j$ (assumption). Hence,

$$\{\underset{\sim}{z} = x_i + y_j\} = \{\underset{\sim}{x} = x_i\} \cap \{\underset{\sim}{y} = y_j\}$$

This shows that $A_z = A_x \cdot B_y$. Furthermore, since the RVs $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are independent, the events $\{\underset{\sim}{x} = x_i\}$ and $\{\underset{\sim}{y} = y_j\}$ are also independent. This shows that the partitions A_x and B_y are independent and [see (14-44) and Prob. 14-1]

$$H(A_z|A_x) = H(A_x \cdot A_y|A_x) = H(A_y|A_x) = H(A_y)$$

From this it follows that $H(\underset{\sim}{z}|\underset{\sim}{x}) = H(\underset{\sim}{y})$ because [see (14-88) and (14-41)]

$$H(\underset{\sim}{z}|\underset{\sim}{x}) = H(A_z|A_x)$$

14-9 As we see from (14-80)

$H(\underset{\sim}{x}) = \ln a$ where we assume that $a = N\delta$. The RV $\underset{\sim}{y}$ takes the values $0, \delta, \ldots, (N-1)\delta$ with probability $1/N$. The conditional density of $\underset{\sim}{x}$ assuming $\underset{\sim}{y} = k\delta$ is uniform in the interval $(k\delta, k\delta + \delta)$. Hence,

$$H(\underset{\sim}{x}|\underset{\sim}{y} = k\delta) = -\int_{k\delta}^{k\delta+\delta} f(x|\underset{\sim}{y} = k\delta) \ln f(x|\underset{\sim}{y} = k\delta)\,dx = \ln \delta$$

And as in (14-41)

$$H(\underset{\sim}{x}|\underset{\sim}{y}) = \sum_{k=0}^{N} H(\underset{\sim}{x}|\underset{\sim}{y} = k\delta) P\{\underset{\sim}{y} = k\delta\} = \ln \delta$$

Finally [see (14-95)]

$$I(\underset{\sim}{x},\underset{\sim}{y}) = H(\underset{\sim}{x}) - H(\underset{\sim}{x}|\underset{\sim}{y}) = \ln a - \ln \delta$$

14-10 If $y_i = g(x_i)$, $y_j = g(x_j)$ and $y_i = y_j$ then $x_i = x_j$. Hence,

$$p_{ij} = \begin{cases} p_i & i = j \\ 0 & i \neq j \end{cases} \qquad p_i = P\{\underset{\sim}{x} = x_i\}$$

and

$$H(\underset{\sim}{x},\underset{\sim}{y}) = -\sum_{i,j} p_{ij} \log p_{ij} = -\sum_i p_i \log p_i = H(\underset{\sim}{x})$$

14-11 From Prob. 10-10 it follows with $g(x) = x$ that $H(\underset{\sim}{x},\underset{\sim}{x}) = H(\underset{\sim}{x})$. And since [see (14-103)] $H(\underset{\sim}{x},\underset{\sim}{x}) = H(\underset{\sim}{x}|\underset{\sim}{x}) + H(\underset{\sim}{x})$ we conclude that $H(\underset{\sim}{x}|\underset{\sim}{x}) = 0$. From Prob. 14-3 it follows that

$$H(\underset{\sim}{y},\underset{\sim}{x}|\underset{\sim}{x}) = H(A_y \cdot A_x|A_x) = H(A_x \cdot A_x) + H(A_y|A_x \cdot A_x)$$

$$= H(A_y|A_x) = H(\underset{\sim}{y}|\underset{\sim}{x})$$

because $A_x \cdot A_x = A_x$ and $H(A_x \cdot A_x) = H(\underset{\sim}{x},\underset{\sim}{x}) = 0$.

14-12 $$E\{\underset{\sim}{x}_n\} = 0 \qquad E\{\underset{\sim}{x}_n^2\} = 5 \qquad E\{\underset{\sim}{y}_n\} = 0$$

$$E\{\underset{\sim}{y}_n^2\} = \sum_{k=0}^{\infty} 2^{-2k} E\{\underset{\sim}{x}_{n-k}^2\} = \frac{20}{3} \qquad E\{\underset{\sim}{x}_n\underset{\sim}{y}_n\} = E\{\underset{\sim}{x}_n^2\} = 5$$

(a) From (14-95), (14-84), and (15-86) with $\mu_{11} = 5$, $\mu_{22} = 20/3$, and $\mu_{12} = 5$

$$H(\underset{\sim}{x}) = \ln\sqrt{10\pi e} \qquad H(\underset{\sim}{y}) = \ln\sqrt{40\pi e/3} \qquad H(\underset{\sim}{x},\underset{\sim}{y}) = \ln 10\pi e/\sqrt{3}$$

$$I(\underset{\sim}{x},\underset{\sim}{y}) = \ln 2$$

(b) The process $\underset{\sim}{y}(t)$ is the output of the system

$$L(z) = \frac{1}{1 - 0.5\, z^{-1}} \qquad \ell_o = 1$$

with input $\underset{\sim}{x}_n$. Since $\bar{H}(\underset{\sim}{x}) = H(\underset{\sim}{x})$ and [see (12A-1)

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|L(e^{j\phi})|\,d\phi = \ln\ell_o = 0$$

(14-133) yields $\bar{H}(\underset{\sim}{y}) = \bar{H}(\underset{\sim}{x}) = H(\underset{\sim}{x}) = \ln\sqrt{10\pi e}$.

14-13
$$\bar{H}(\underset{\sim}{x}) = H(\underset{\sim}{x}) = -\frac{1}{2}\int_4^6 \ln\frac{1}{2}\,dx = \ln 2$$

And as in Prob. 14-12 with $\ell_o = 5$,

$$\bar{H}(\underset{\sim}{y}) = \bar{H}(x) + \ln 5 = \ln 10$$

14-14 Given that $f(x) = 0$ for $|x|>1$ and $E\{\underset{\sim}{x}\} = 0.3$, find $f(x)$. With $g(x) = x$, (14-143) yields $f(x) = Ae^{-\lambda x}$ where

$$A\int_{-1}^{1} e^{-\lambda x}\,dx = \frac{A}{\lambda}(e^{\lambda} - e^{-\lambda}) = 1$$

$$A\int_{-1}^{1} xe^{-\lambda x}\,dx = \frac{A}{\lambda^2}(e^{\lambda} - e^{-\lambda}) - \frac{A}{\lambda}(e^{\lambda} - e^{-\lambda}) = 0.31$$

Solving, we obtain $A \simeq 0.425$, $\lambda \simeq -1$

14-15 $f(x) = Ae^{-\lambda x}$ for $1<x<5$ and 0 otherwise,

$$A\int_{1}^{5} e^{-\lambda x}\,dx = 0.31 \qquad A\int_{1}^{5} xe^{-\lambda x}dx = 3\,\frac{37}{60}$$

Hence, $A \simeq 1.06$, $\lambda \simeq 0.5$

14-16 From (14-151) with $x_k=k$, $g_1(x_k) = g_1(k) = k$, $k=1, \ldots, 6$

$$g_2(x_k)=\begin{cases} 0 & k=1,3,5 \\ 1 & k=2,4,6 \end{cases} \quad p_k=\begin{cases} Ae^{-\lambda_1 k} & k=1,3,5 \\ Ae^{-\lambda_1 x-\lambda_2} & k=2,4,6 \end{cases}$$

Since $p_1 + p_3 + p_5 = 0.5$ and $E\{\underset{\sim}{x}\} = 4.44$, we conclude with $z = e^{-\lambda_2}$ and $w = e^{-\lambda_2}$ that

$$A(z+z^3+z^5) = Aw(z^2+z^4z^6)$$

$$A(\underset{\sim}{z}+3z^3+5z^5) + Aw(2z^2+4z^4+6z^6) = 4.44$$

This yields $A \simeq 0.0437$, $\underset{\sim}{z} = 1/w \simeq 1.468$

14-17 (a) The transformation $\underset{\sim}{y} = 3\underset{\sim}{x}$ is one-to-one, hence, $H(\underset{\sim}{y}) = H(\underset{\sim}{x})$

(b) From (14-113) with $g(x) = 3x$: $H(\underset{\sim}{y}) = H(\underset{\sim}{x}) + \ln 3$

14-18 (a) For fair dice, $P\{7\} = \frac{1}{6}$, $P(11) = \frac{1}{18}$, $P\{\text{neither 7 nor 11}\} = \frac{14}{18}$

$$H(A) = -\left(\frac{1}{6}\ln\frac{1}{6} + \frac{1}{18}\ln\frac{1}{18} + \frac{14}{18}\ln\frac{14}{18}\right) = 0.655$$

(b) From (14-10) with $n=100$ and $N=3$:

$$n_T \simeq e^{nH(A)} \simeq 2.79 \times 10^{28} \qquad n_a \simeq N^n \simeq 5.16 \times 10^{47}$$

14-19 The process $\underset{\sim}{x}_n$ is WSS with entropy rate $\overline{H}(x)$. Show that, if

$$\underset{\sim}{w}_n = \sum_{k=0}^{n} \underset{\sim}{x}_{n-k}\, \ell_k$$

then

$$\lim_{n\to\infty} \frac{1}{n+1} H(\underset{\sim}{w}_o,\ldots,\underset{\sim}{w}_n) = \overline{H}(x) + \ln|\ell_o| \tag{i}$$

Proof. The RVs $\underset{\sim}{w}_o,\ldots,\underset{\sim}{w}_n$ are linear transformations of the RVs $\underset{\sim}{x}_o,\ldots,\underset{\sim}{x}_n$ and the transformation matrix equals

$$\begin{bmatrix} \ell_0 & 0 & \cdots & 0 \\ \ell_1 & \ell_0 & \cdots & 0 \\ \hline \ell_n & \ell_{n-1} & \cdots & 0 \end{bmatrix}$$

Since the determinant of this transformation equals $|\ell_o|^{n+1}$, (14-115) yields

$$H(\underset{\sim}{w}_o,\ldots,\underset{\sim}{w}_n) = H(\underset{\sim}{x}_o,\ldots\underset{\sim}{x}_n) + (n+1)\ln|\ell_o|$$

Dividing by (n+1) we obtain (i) as $n\to\infty$.

14-20 As in Example 14-19, $f(p) = A\, e^{-\lambda p}$. To find λ, we use the $\lambda-\eta$ curve of Fig. 14-16. This yields

$$\lambda = -1.23 \qquad f(p) = 0.51\, e^{1.23p}$$

14-21 As in Example 14-22, $p_k = A\,e^{-\lambda k}$. To find λ, we use the w-η curve of Fig. 14-17. This yields (see also Jaynes)

$$w \simeq 1.449 \qquad \lambda \simeq -0.371$$

p_1	p_2	p_3	p_4	p_5	p_6
0.054	0.079	0.114	0.165	0.240	0.348

14-22 The unknown density is normal as in (14-157) where

$$\Delta = \begin{vmatrix} 4 & 1 & 1 \\ 1 & 4 & m_{23} \\ 1 & m_{23} & 4 \end{vmatrix} = -4m_{23}^2 + 2m_{23} + 56$$

The moment $m_{23} = E\{\mathbf{x}_2\mathbf{x}_3\}$ must be such as to maximize Δ. This yields $m_{23} = 0.25$.

14-23

Shannon

L = 2.7

p_i	0.3	0.2	0.15	0.15	0.1	0.06	0.04	
	$\frac{1}{4} \le p_i < \frac{1}{2}$	$\frac{1}{8} \le p_i < \frac{1}{4}$			$\frac{1}{16} \le p_i < \frac{1}{8}$	$\frac{1}{32} \le p_i < \frac{1}{6}$		$\sum_{i=1}^{7} \frac{1}{2^{n_i}}$
n_i	2	3	3	3	4	5	5	0.75
	2	3	3	3	4	4	4	0.8125
	2	3	3	3	3	4	4	0.875
	2	3	3	3	3	3	4	0.9375
	2	3	3	3	3	3	3	1
x_i	00	010	011	100	101	110	111	

Fano

$L = 2.6$

p_i	0.3	0.2	0.15	0.15	0.1	0.06	0.04
	A_0 0.5		A_1 0.5				
	A_{00} 0.3	A_{01} 0.2	A_{10} 0.3		A_{11} 0.2		
			A_{100} 0.15	A_{101} 0.15	A_{110} 0.1	A_{111} 0.1	
						A_{1110}	A_{1111}
x_i	00	01	100	101	110	1110	1111

Huffman

$L = 2.6$

	1	2	3	4	5	6	7
	1	2	3	4	5	6 0	7 1
	1	2	5 0	6 10	7 11	3	4
	1	3 0	4 1	2	5 0	6 10	7 11
	2 0	5 10	6 110	7 111	1	3 0	4 1
	1 0	3 10	4 11	2 0	5 10	6 110	7 111
	1 00	3 010	4 011	2 10	5 110	6 1110	7 1111
x_i	00	10	010	011	110	1110	1111

14-24 If $\underline{x}_n = 0$, then $\bar{x}_n = 000$ and $\underline{y}_n = 1$ iff $\bar{y}_n$ consists of one 0 or no zeros. The probability of one and only one zero equals $3\beta^2(1-\beta)$ [see (3-13)]; the probability of no zeros equals β^3. Hence,

$$P\{\underline{y}_n = 1 | \underline{x}_n = 0\} = 3\beta^2(1-\beta) + \beta^3$$

Thus, the redundantly coded channel of Example 14-29 is symmetrical as in (14-191) with probability of error $\beta_1 = \beta^2$.

14-25 If the received information is always wrong, then

$$P\{\underline{y}_n = 1 | \underline{x}_n = 0\} = \beta = 1, \text{ hence } C = 1 - r(\beta) = 1$$

Chapter 15

15.1 The chain represented by

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

is irreducible and aperiodic.

The second chain is also irreducible and aperiodic.

The third chain has two aperiodic closed sets $\{e_1, e_2\}$ and $\{e_3, e_4\}$ and a transient state e_5.

15.2 Note that both the row sums and column sums are unity in this case. Hence P represents a doubly stochastic matrix here, and

$$P^n = \frac{1}{m+1} \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}$$

$$\lim_{n\to\infty} P\{\mathbf{x}_n = e_k\} = \frac{1}{m+1}, \qquad k = 0, 1, 2, \cdots m.$$

15.3 This is the "success runs" problem discussed in Example 15-11 and 15-23. From Example 15-23, we get

$$u_{i+1} = p_{i,i+1} u_i = \frac{1}{i+1} u_i = \frac{u_o}{(i+1)!}$$

so that from (15-206)

$$\sum_{k=1}^{\infty} u_k = u_0 \sum_{k=1}^{\infty} \frac{1}{k!} = e \cdot u_0 = 1$$

gives $u_0 = 1/e$ and the steady state probabilities are given by

$$u_k = \frac{1/e}{k!}, \qquad k = 1, 2, \cdots$$

15.4 If the zeroth generation has size m, then the overall process may be considered as the sum of m independent and identically distributed branching processes $\mathbf{x}_n^{(k)}$, $k = 1, 2, \cdots m$, each corresponding to unity size at the zeroth generation. Hence if π_0 represents the probability of extinction for any one of these individual processes, then the overall probability of extinction is given by

$$\begin{aligned}
&\lim_{n\to\infty} P[\mathbf{x}_n = 0|\mathbf{x}_0 = m] = \\
= \;& P[\{\mathbf{x}_n^{(1)} = 0|\mathbf{x}_0^{(1)} = 1\} \bigcap \{\mathbf{x}_n^{(2)} = 0|\mathbf{x}_0^{(2)} = 1\} \bigcap \cdots \{\mathbf{x}_n^{(m)} = 0|\mathbf{x}_0^{(m)} = 1\}] \\
= \;& \textstyle\prod_{k=1}^m P[\mathbf{x}_n^{(k)} = 0|\mathbf{x}_0^{(k)} = 1] \\
= \;& \pi_0^m
\end{aligned}$$

15.5 From (15-288)-(15-289),

$$P(z) = p_0 + p_1 z + p_2 z^2, \quad \text{since } p_k = 0, \quad k \geq 3.$$

Also $p_0 + p_1 + p_2 = 1$, and from (15-307) the extinction probability is given by sloving the equation

$$P(z) = z.$$

Notice that

$$\begin{aligned}
P(z) - z &= p_0 - (1 - p_1)z + p_2 z^2 \\
&= p_0 - (p_0 + p_2)z + p_2 z^2 \\
&= (z - 1)(p_2 z - p_0)
\end{aligned}$$

and hence the two roots of the equation $P(z) = z$ are given by

$$z_1 = 1, \qquad z_2 = \frac{p_0}{p_2}.$$

Thus if $p_2 < p_0$, then $z_2 > 1$ and hence the smallest positive root of $P(z) = z$ is 1, and it represents the probability of extinction. It follows that such a tribe which does not produce offspring in abundence is bound to extinct.

15.6 Define the branching process $\{\mathbf{x}_n\}$

$$\mathbf{x}_{n+1} = \sum_{k=1}^{\mathbf{x}_n} \mathbf{y}_k$$

where $\mathbf{y}_k$ are i.i.d random variables with common moment generating function $P(z)$ so that (see (15-287)-(15-289))

$$P'(1) = E\{\mathbf{y}_k\} = \mu.$$

Thus

$$\begin{aligned} E\{\mathbf{x}_{n+1}|\mathbf{x}_n\} &= E\{\textstyle\sum_{k=1}^{\mathbf{x}_n} \mathbf{y}_k|\mathbf{x}_n = m\} \\ &= E\{\textstyle\sum_{k=1}^{m} \mathbf{y}_k|\mathbf{x}_n = m\} \\ &= E\{\textstyle\sum_{k=1}^{m} \mathbf{y}_k\} = mE\{\mathbf{y}_k\} = \mathbf{x}_n\,\mu \end{aligned}$$

Similarly

$$\begin{aligned} E\{\mathbf{x}_{n+2}|\mathbf{x}_n\} &= E\{E\{\mathbf{x}_{n+2}|\mathbf{x}_{n+1}, \mathbf{x}_n\}\} \\ &= E\{E\{\mathbf{x}_{n+2}|\mathbf{x}_{n+1}\}|\mathbf{x}_n\} \\ &= E\{\mu\mathbf{x}_{n+1}|\mathbf{x}_n\} = \mu^2\,\mathbf{x}_n \end{aligned}$$

and in general we obtain

$$E\{\mathbf{x}_{n+r}|\mathbf{x}_n\} = \mu^r\,\mathbf{x}_n. \tag{i}$$

Also from (15-310)-(15-311)

$$E\{\mathbf{x}_n\} = \mu^n. \tag{ii}$$

Define

$$\mathbf{w}_n = \frac{\mathbf{x}_n}{\mu^n}. \tag{iii}$$

This gives

$$E\{\mathbf{w}_n\} = 1.$$

Dividing both sider of (i) with μ^{n+r} we get

$$E\{\frac{\mathbf{x}_{n+r}}{\mu^{n+r}}|\mathbf{x}_n = x\} = \mu^r \cdot \frac{\mathbf{x}_n}{\mu^{n+r}} = \frac{\mathbf{x}_n}{\mu^n} = \mathbf{w}_n$$

or

$$E\{\mathbf{w}_{n+r}|\mathbf{w}_n = \frac{x}{\mu^n} \triangleq w\} = \mathbf{w}_n$$

which gives

$$E\{\mathbf{w}_{n+r}|\mathbf{w}_n\} = \mathbf{w}_n,$$

the desired result.

15.7

$$\mathbf{s}_n = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n$$

where $\mathbf{x}_n$ are i.i.d. random variables. We have

$$\mathbf{s}_{n+1} = \mathbf{s}_n + \mathbf{x}_{n+1}$$

so that

$$E\{\mathbf{s}_{n+1}|\mathbf{s}_n\} = E\{\mathbf{s}_n + \mathbf{x}_{n+1}|\mathbf{s}_n\} = \mathbf{s}_n + E\{\mathbf{x}_{n+1}\} = \mathbf{s}_n.$$

Hence $\{\mathbf{s}_n\}$ represents a Martingale.

15.8 (a) From Bayes' theorem

$$\begin{aligned} P\{\mathbf{x_n} = j|\mathbf{x_{n+1}} = i\} &= \frac{P\{\mathbf{x}_{n+1} = i|\mathbf{x}_n = j\}\, P\{\mathbf{x}_n = j\}}{P\{\mathbf{x}_{n+1} = i\}} \\ &= \frac{q_j\, p_{ji}}{q_i} = p^*_{ij}, \end{aligned} \tag{i}$$

where we have assumed the chain to be in steady state.

(b) Notice that time-reversibility is equivalent to

$$p^*_{ij} = p_{ij}$$

and using (i) this gives

$$p^*_{ij} = \frac{q_j\, p_{ji}}{q_i} = p_{ij} \tag{ii}$$

or, for a time-reversible chain we get

$$q_j\, p_{ji} = q_i\, p_{ij}. \tag{iii}$$

Thus using (ii) we obtain by direct substitution

$$\begin{aligned} p_{ij}\, p_{jk}\, p_{ki} &= \left(\frac{q_j}{q_i}\, p_{ji}\right) \left(\frac{q_k}{q_j}\, p_{kj}\right) \left(\frac{q_i}{q_k}\, p_{ik}\right) \\ &= p_{ik}\, p_{kj}\, p_{ji}, \end{aligned}$$

the desired result.

15.9 (a) It is given that $A = A^T$, $(a_{ij} = a_{ji})$ and $a_{ij} > 0$. Define the i^{th} row sum

$$r_i = \sum_k a_{ik} > 0, \qquad i = 1, 2, \cdots$$

and let

$$p_{ij} = \frac{a_{ij}}{\sum_k a_{ik}} = \frac{a_{ij}}{r_i}.$$

Then

$$\begin{aligned} p_{ji} &= \frac{a_{ji}}{\displaystyle\sum_m a_{jm}} = \frac{a_{ji}}{r_j} = \frac{a_{ij}}{r_j} \\ &= \frac{r_i}{r_j}\, \frac{a_{ij}}{r_i} = \frac{r_i}{r_j}\, p_{ij} \end{aligned} \tag{i}$$

or

$$r_i\, p_{ij} = r_j\, p_{ji}.$$

Hence

$$\sum_i r_i\, p_{ij} = \sum_i r_j\, p_{ji} = r_j \sum_i p_{ji} = r_j, \tag{ii}$$

since

$$\sum_i p_{ji} = \frac{\sum_i a_{ji}}{r_j} = \frac{r_j}{r_j} = 1.$$

Notice that (ii) satisfies the steady state probability distribution equation (15-167) with

$$q_i = c\, r_i, \qquad i = 1, 2, \cdots$$

where c is given by

$$c \sum_i r_i = \sum_i q_i = 1 \Longrightarrow c = \frac{1}{\sum_i r_i} = \frac{1}{\sum_i \sum_j a_{ij}}.$$

Thus

$$q_i = \frac{r_i}{\sum_i r_i} = \frac{\sum_j a_{ij}}{\sum_i \sum_j a_{ij}} > 0 \tag{iii}$$

represents the stationary probability distribution of the chain.

With (iii) in (i) we get

$$p_{ji} = \frac{q_i}{q_j}\, p_{ij}$$

or

$$p_{ij} = \frac{q_j\, p_{ji}}{q_i} = p_{ij}^*$$

and hence the chain is time-reversible.

15.10 (a) $M = (m_{ij})$ is given by

$$M = (I - W)^{-1}$$

or

$$(I - W)M = I$$
$$M = I + WM$$

which gives

$$\begin{aligned} m_{ij} &= \delta_{ij} + \textstyle\sum_k w_{ik}\, m_{kj}, \quad e_i, e_j \in T \\ &= \delta_{ij} + \textstyle\sum_k p_{ik}\, m_{kj}, \quad e_i, e_j \in T \end{aligned}$$

(b) The general case is solved in pages 743-744. From page 744, with $N = 6$ (2 absorbing states; 5 transcient states), and with $r = p/q$ we obtain

$$m_{ij} = \begin{cases} \dfrac{(r^j - 1)(r^{6-i} - 1)}{(p-q)(r^6 - 1)}, & j \le i \\[2ex] \dfrac{(r^i - 1)(r^{6-i} - r^{j-i})}{(p-q)(r^6 - 1)}, & j \ge i. \end{cases}$$

15.11 If a stochastic matrix $A = (a_{ij}), a_{ij} > 0$ corresponds to the two-step transition matrix of a Markov chain, then there must exist another stochastic matrix P such that

$$A = P^2, \qquad P = (p_{ij})$$

where

$$p_{ij} > 0, \quad \sum_j p_{ij} = 1,$$

and this may not be always possible. For example in a two state chain, let

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}$$

so that

$$A = P^2 = \begin{pmatrix} \alpha^2 + (1-\alpha)(1-\beta) & (\alpha+\beta)(1-\alpha) \\ (\alpha+\beta)(1-\beta) & \beta^2 + (1-\alpha)(1-\beta) \end{pmatrix}.$$

This gives the sum of this its diagonal entries to be

$$\begin{aligned} a_{11} + a_{22} &= \alpha^2 + 2(1-\alpha)(1-\beta) + \beta^2 \\ &= (\alpha+\beta)^2 - 2(\alpha+\beta) + 2 \\ &= 1 + (\alpha+\beta-1)^2 \ \geq 1. \end{aligned} \qquad \text{(i)}$$

Hence condition (i) necessary. Since $0 < \alpha < 1$, $0 < \beta < 1$, we also get $1 < a_{11} + a_{22} \leq 2$. Futher, the condition (i) is also sufficient in the 2×2 case, since $a_{11} + a_{22} > 1$, gives

$$(\alpha+\beta-1)^2 = a_{11} + a_{22} - 1 > 0$$

and hence

$$\alpha + \beta = 1 \pm \sqrt{a_{11} + a_{22} - 1}$$

and this equation may be solved for all admissible set of values $0 < \alpha < 1$ and $0 < \beta < 1$.

15.12 In this case the chain is irreducible and aperiodic and there are no absorption states. The steady state distribution $\{u_k\}$ satisfies (15-167),and hence we get

$$u_k = \sum_j u_j \, p_{jk} = \sum_{j=0}^{N} u_j \binom{N}{k} p_j^k \, q_j^{N-k}.$$

Then if $\alpha > 0$ and $\beta > 0$ then "fixation to pure genes" does not occur.

15.13 The transition probabilities in all these cases are given by (page 765) (15A-7) for specific values of $A(z) = B(z)$ as shown in Examples 15A-1, 15A-2 and 15A-3. The eigenvalues in general satisfy the equation

$$\sum_j p_{ij}\, x_j^{(k)} = \lambda_k\, x_i^{(k)}, \quad k = 0, 1, 2, \cdots N$$

and trivially $\sum_j p_{ij} = 1$ for all i implies $\lambda_0 = 1$ is an eigenvalue in all cases.

However to determine the remaining eigenvalues we can exploit the relation in (15A-7). From there the corresponding conditional moment generating function in (15-291) is given by

$$G(s) = \sum_{j=0}^{N} p_{ij}\, s^j \tag{i}$$

where from (15A-7)

$$\begin{aligned} p_{ij} &= \frac{\{A^i(z)\}_j\, \{B^{N-i}(z)\}_{N-j}}{\{A^i(z)\, B^{N-i}(z)\}_N} \\ &= \frac{\text{coefficient of } s^j\, z^N \text{ in } \{A^i(sz)\, B^{N-i}(z)\}}{\{A^i(z)\, B^{N-i}(z)\}_N} \end{aligned} \tag{ii}$$

Substituting (ii) in (i) we get the compact expression

$$G(s) = \frac{\{A^i(sz)\, B^{N-i}(z)\}_N}{\{A^i(z)\, B^{N-i}(z)\}_N}. \tag{iii}$$

Differentiating $G(s)$ with respect to s we obtain

$$\begin{aligned} G'(s) &= \sum_{j=0}^{N} P_{ij}\, j\, s^{j-1} \\ &= \frac{\{iA^{i-1}(sz)\, A'(sz) z\, B^{N-i}(z)\}_N}{\{A^i(z)\, B^{N-i}(z)\}_N} \\ &= i \cdot \frac{\{A^{i-1}(sz)\, A'(sz)\, B^{N-i}(z)\}_{N-1}}{\{A^i(z)\, B^{N-i}(z)\}_N}. \end{aligned} \tag{iv}$$

Letting $s = 1$ in the above expression we get

$$G'(1) = \sum_{j=0}^{N} p_{ij}\, j = i\, \frac{\{A^{i-1}(z)\, A'(z)\, B^{N-i}(z)\}_{N-1}}{\{A^{i}(z)\, B^{N-i}(z)\}_{N}}. \tag{v}$$

In the special case when $A(z) = B(z)$, Eq.(v) reduces to

$$\sum_{j=0}^{N} p_{ij}\, j = \lambda_1\, i \tag{vi}$$

where

$$\lambda_1 = \frac{\{A^{N-1}(z)\, A'(z)\}_{N-1}}{\{A^{N}(z)\}_{N}}. \tag{vii}$$

Notice that (vi) can be written as

$$P x_1 = \lambda_1\, x_1, \quad x_1 = [0, 1, 2, \cdots N]^T$$

and by direct computation with $A(z) = B(z) = (q + pz)^2$ (Example 15A-1) we obtain

$$\begin{aligned}
\lambda_1 &= \frac{\{(q+pz)^{2(N-1)}\, 2p(q+pz)\}_N}{\{(q+pz)^{2N}\}_N} \\
&= \frac{2p\{(q+pz)^{2N-1}\}_{N-1}}{\{(q+pz)^{2N}\}_N} = \frac{2p\binom{2N}{N-1} q^N\, p^{N-1}}{\binom{2N}{N} q^N\, p^N} = 1.
\end{aligned}$$

Thus $\sum_{j=0}^{N} p_{ij}\, j = i$ and from (15-224) these chains represent Martingales. (Similarly for Examples 15A-2 and 15A-3 as well).

To determine the remaining eigenvalues we differentiate $G'(s)$ once more. This gives

$$\begin{aligned}
G''(s) &= \sum_{j=0}^{N} p_{ij}\, j(j-1)\, s^{j-2} \\
&= \frac{\{i(i-1)A^{i-2}\,(sz)[A'(sz)]^2\, z\, B^{N-i}(z) + iA^{i-1}(sz)\, A''(sz) z\, B^{N-i}(z)\}_{N-1}}{\{A^{i}(z)\, B^{N-i}(z)\}_N}\ . \\
&= \frac{\{i\, A^{i-2}(sz)\, B^{N-i}(z)[(i-1)\,(A'(sz))^2 + A(sz)\, A''(sz)]\}_{N-2}}{\{A^{i}(z)\, B^{N-i}(z)\}_N}.
\end{aligned}$$

With $s = 1$, and $A(z) = B(z)$, the above expression simplifies to

$$\sum_{j=0}^{N} p_{ij}\, j(j-1) = \lambda_2\, i(i-1) + i\mu_2 \tag{viii}$$

where

$$\lambda_2 = \frac{\{A^{N-2}(z)\,[A'(z)]^2\}_{N-2}}{\{A^N(z)\}_N}$$

and

$$\mu_2 = \frac{\{A^{N-1}(z)\,A''(z)\}_{N-2}}{\{A^N(z)\}_N}.$$

Eq. $(viii)$ can be rewritten as

$$\sum_{j=0}^{N} p_{ij}\, j^2 = \lambda_2\, i^2 + (\text{polynomial in } i \text{ of degree } \le 1)$$

and in general repeating this procedure it follows that (show this)

$$\sum_{j=0}^{N} p_{ij}\, j^k = \lambda_k\, i^k + (\text{polynomial in } i \text{ of degree } \le k-1) \tag{ix}$$

where

$$\lambda_k = \frac{\{A^{N-k}(z)\,[A'(z)]^k\}_{N-k}}{\{A^N(z)\}_N}, \quad k = 1, 2, \cdots N. \tag{x}$$

Equations $(viii)$–(x) motivate to consider the identities

$$P\, q_k = \lambda_k\, q_k \tag{xi}$$

where q_k are polynomials in i of degree $\le k$, and by proper choice of constants they can be chosen in that form. It follows that λ_k, $k = 1, 2, \cdots N$ given by (ix) represent the desired eigenvalues.

(a) The transition probabilities in this case follow from Example 15A-1 (page 765-766) with $A(z) = B(z) = (q + pz)^2$. Thus using (ix) we obtain the desired eigenvalues to be

$$\begin{aligned}\lambda_k &= \frac{\{(q+pz)^{2(N-k)}[2p(q+pz)]^k\}_{N-k}}{\{(q+pz)^{2N}\}_N} \\ &= 2^k\, p^k\, \frac{\{(q+pz)^{2N-k}\}_{N-k}}{\{(q+pz)^{2N}\}_N\}} \\ &= 2^k\, \frac{\binom{2N-k}{N-k}}{\binom{2N}{N}}, \quad k = 1, 2, \cdots N.\end{aligned}$$

(b) The transition probabilities in this case follows from Example 15A-2 (page 766) with

$$A(z) = B(z) = e^{\lambda(z-1)}$$

and hence

$$\begin{aligned}\lambda_k &= \frac{\{e^{\lambda(N-k)(z-1)}\,\lambda^k\, e^{\lambda k(z-1)}\}_{N-k}}{\{e^{\lambda N(z-1)}\}_N} \\ &= \frac{\lambda^k\,\{e^{\lambda N z}\}_{N-k}}{\{e^{\lambda N z}\}_N} = \frac{\lambda^k\,(\lambda N)^{N-k}/(N-k)!}{(\lambda N)^N/N!} \\ &= \frac{N!}{(N-k)!\,N^k} = \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{k-1}{N}\right), \quad k = 1, 2, \cdots N\end{aligned}$$

(c) The transition probabilities in this case follow from Example 15A-3 (page 766-767) with

$$A(z) = B(z) = \frac{q}{1-pz}.$$

Thus

$$\begin{aligned}\lambda_k &= p^k\,\frac{\{1/(1-pz)^{N+k}\}_{N-k}}{\{1/(1-pz)^N\}_N} \\ &= (-1)^k\,\frac{\binom{-(N+k)}{N-k}}{\binom{-N}{N}} = \frac{\binom{2N-1}{N-k}}{\binom{2N-1}{N}}, \quad r = 2, 3, \cdots N\end{aligned}$$

15.14 From (15-240), the mean time to absorption vector is given by

$$m = (I - W)^{-1}\, E, \quad E = [1, 1, \cdots 1]^T,$$

where

$$W_{ik} = p_{jk}, \quad j, k = 1, 2, \cdots N - 1,$$

with p_{jk} as given in (15-30) and (15-31) respectively.

15.15 The mean time to absorption satisfies (15-240). From there

$$\begin{aligned} m_i &= 1 + \sum_{k \in T} p_{ik}\, m_k = 1 + p_{i,i+1}\, m_{i+1} + p_{i,i-1}\, m_{i-1} \\ &= 1 + p\, m_{i+1} + q\, m_{i-1}, \end{aligned}$$

or

$$m_k = 1 + p\, m_{k+1} + q\, m_{k-1}.$$

This gives

$$p\,(m_{k+1} - m_k) = q\,(m_k - m_{k-1}) - 1$$

Let

$$M_{k+1} = m_{k+1} - m_k$$

so that the above iteration gives

$$\begin{aligned} M_{k+1} &= \frac{q}{p}\, M_k - \frac{1}{p} \\ &= \left(\frac{q}{p}\right)^k M_1 - \frac{1}{p}\left[1 + \frac{q}{p} + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{k-1}\right] \\ &= \begin{cases} \left(\frac{q}{p}\right)^k M_1 - \dfrac{1}{p - q}\left\{1 - (\frac{q}{p})^k\right\}, & p \neq q \\ M_1 - \dfrac{k}{p}, & p = q \end{cases} \end{aligned}$$

This gives

$$
\begin{aligned}
m_i &= \sum_{k=0}^{i-1} M_{k+1} \\
&= \begin{cases} \left(M_1 + \frac{1}{p-q}\right) \sum_{k=0}^{i-1} \left(\frac{q}{p}\right)^k - \frac{i}{p-q}, & p \neq q \\ iM_1 - \frac{i(i-1)}{2p}, & p = q \end{cases} \\
&= \begin{cases} \left(M_1 + \frac{1}{p-q}\right) \frac{1-(q/p)^i}{1-q/p} - \frac{i}{p-q}, & p \neq q \\ iM_1 - \frac{i(i-1)}{2p}, & p = q \end{cases}
\end{aligned}
$$

where we have used $m_o = 0$. Similarly $m_{a+b} = 0$ gives

$$
M_1 + \frac{1}{p-q} = \frac{a+b}{p-q} \cdot \frac{1-q/p}{1-(q/p)^{a+b}}.
$$

Thus

$$
m_i = \begin{cases} \frac{a+b}{p-q} \cdot \frac{1-(q/p)^i}{1-(q/p)^{a+b}} - \frac{i}{p-q}, & p \neq q \\ i(a+b-i), & p = q \end{cases}
$$

which gives for $i = a$

$$
\begin{aligned}
m_a &= \begin{cases} \frac{a+b}{p-q} \cdot \frac{1-(q/p)^a}{1-(q/p)^{a+b}} - \frac{a}{p-q}, & p \neq q \\ ab, & p = q \end{cases} \\
&= \begin{cases} \frac{b}{2p-1} - \frac{a+b}{2p-1} \cdot \frac{1-(p/q)^b}{1-(p/q)^{a+b}}, & p \neq q \\ ab, & p = q \end{cases}
\end{aligned}
$$

by writing

$$
\frac{1-(q/p)^a}{1-(q/p)^{a+b}} = 1 - \frac{(q/p)^a - (q/p)^{a+b}}{1-(q/p)^{a+b}} = 1 - \frac{1-(p/q)^b}{1-(p/q)^{a+b}}
$$

(see also problem 3-10).

Chapter 16

16.1 Use (16-132) with $r = 1$. This gives

$$p_n = \begin{cases} \dfrac{\rho^n}{n!}\, p_0, & n \le 1 \\ \rho^n\, p_0, & 1 < n \le m \end{cases}$$

$$= \rho^n\, p_0, \qquad 0 \le n \le m$$

Thus

$$\sum_{n=0}^{m} p_n = p_0 \sum_{n=0}^{m} \rho^n = p_0\, \frac{(1-\rho^{m+1})}{1-\rho} = 1$$

$$\Longrightarrow \quad p_0 = \frac{1-\rho}{1-\rho^{m+1}}$$

and hence

$$p_n = \frac{1-\rho}{1-\rho^{n+1}}\, \rho^n, \qquad 0 \le n \le m, \quad \rho \ne 1$$

and $lim\, \rho \to 1$, we get

$$p_n = \frac{1}{m+1}, \qquad \rho = 1.$$

16.2 (a) Let $n_1(t) = X + Y$, where X and Y represent the two queues. Then

$$\begin{aligned} p_n &= P\{n_1(t) = n\} = P\{X+Y = n\} \\ &= \sum_{k=0}^{n} P\{X = k\}\, P\{Y = n-k\} \\ &= \sum_{k=0}^{n} (1-\rho)\rho^k\, (1-\rho)\rho^{n-k} \\ &= (n+1)(1-\rho)^2\rho^n, \qquad n = 0, 1, 2, \cdots \end{aligned} \tag{i}$$

where $\rho = \lambda/\mu$.

(b) When the two queues are merged, the new input rate $\lambda^{'} = \lambda + \lambda = 2\lambda$. Thus from (16-102)

$$p_n = \begin{cases} \dfrac{(\lambda^{'}/\mu)^n}{n!} p_0 = \dfrac{(2\rho)^n}{n!}\, p_0, & n < 2 \\ \dfrac{2^2}{2!}\,(\dfrac{\lambda^{'}}{2\mu})^n\, p_0 = 2\rho^n p_0, & n \geq 2. \end{cases}$$

Hence

$$\begin{aligned} \sum_{k-0}^{\infty} p_k &= p_0(1 + 2\rho + 2\sum_{k=2}^{\infty}\rho^k) \\ &= p_0(1 + 2\rho + \frac{2\rho^2}{1-\rho}) \\ &= \frac{p_0}{1-\rho}((1+2\rho)\,(1-\rho) + 2\rho^2) \\ &= \frac{p_0}{1-\rho}\,(1+\rho) = 1 \end{aligned}$$

$$\Longrightarrow p_0 = \frac{1-\rho}{1+\rho}, \qquad (\rho = \lambda/\mu). \tag{ii}$$

Thus

$$p_n = \begin{cases} 2\,(1-\rho)\,\rho^n/(1+\rho), & n \leq 1 \\ (1-\rho)/(1+\rho), & n = 0 \end{cases} \tag{iii}$$

(c) For an $M/M/1$ queue the average number of items waiting is given by (use (16-106) with $r = 1$)

$$E\{X\} = L_1^{'} = \sum_{n=2}^{\infty}(n-1)\,p_n$$

where p_n is an in (16-88). Thus

$$\begin{aligned} L_1' &= \sum_{n=2}^{\infty} (n-1)(1-\rho)\,\rho^n \\ &= (1-\rho)\,\rho^2 \sum_{n=2}^{\infty} (n-1)\,\rho^{n-2} \\ &= (1-\rho)\,\rho^2 \sum_{k=1}^{\infty} k\,\rho^{k-1} \\ &= (1-\rho)\,\rho^2 \frac{1}{(1-\rho)^2} = \frac{\rho^2}{(1-\rho)}. \end{aligned} \tag{iv}$$

Since $n_1(t) = X + Y$ we have

$$\begin{aligned} L_1 &= E\{n_1(t)\} = E\{X\} + E\{Y\} \\ &= 2L_1' = \frac{2\rho^2}{1-\rho} \end{aligned} \tag{v}$$

For L_2 we can use (16-106)-(16-107) with $r = 2$. Using (iii), this gives

$$\begin{aligned} L_2 &= p_r \frac{\rho}{(1-\rho)^2} \\ &= 2\,\frac{(1-\rho)\,\rho^2}{1+\rho}\,\frac{\rho}{(1-\rho)^2} = \frac{2\,\rho^3}{1-\rho^2} \\ &= \frac{2\,\rho^2}{1-\rho}\left(\frac{\rho}{1+\rho}\right) < L_1 \end{aligned} \tag{vi}$$

From (vi), a single queue configuration is more efficient then two separate queues.

16.3 The only non-zero probabilities of this process are

$$\lambda_{0,0} = -\lambda_0 = -m\lambda, \quad \lambda_{0,1} = \mu$$

$$\lambda_{i,i+1} = (m-i)\,\lambda, \quad \lambda_{i,i-1} = i\mu$$

$$\lambda_{i,i} = [(m-i)\,\lambda + i\mu], \quad i = 1, 2, \cdots, m-1$$

$$\lambda_{m,m} = -\lambda_{m,m-1} = -m\mu.$$

Substituting these into (16-63) text, we get

$$m\,\lambda\,p_0 = \mu\,p_1 \tag{i}$$

$$[(m-i)\lambda + i\mu]\,p_i = (m-i+1)\,p_{i-1} + (i+1)\,\mu\,p_{i+1}, \quad i = 1, 2, \cdots, m-1 \tag{ii}$$

and

$$m\,\mu\,p_m = \lambda\,p_{m-1}. \tag{iii}$$

Solving (i)-(iii) we get

$$p_i = \binom{m}{i} \left(\frac{\lambda}{\lambda+\mu}\right)^i \left(\frac{\mu}{\lambda+\mu}\right)^{m-i}, \quad i = 0, 1, 2, \cdots, m$$

16.4 (a) In this case

$$p_n = \begin{cases} \dfrac{\lambda}{\mu_1}\,\dfrac{\lambda}{\mu_1}\cdots\dfrac{\lambda}{\mu_1} = \left(\dfrac{\lambda}{\mu_1}\right)^n p_0, & n < m \\ \dfrac{\lambda}{\mu_1}\,\dfrac{\lambda}{\mu_1}\cdots\dfrac{\lambda}{\mu_1}\,\dfrac{\lambda}{\mu_2}\cdots\dfrac{\lambda}{\mu_2}\,p_0, & n \geq m \end{cases}$$

$$= \begin{cases} \rho_1^n\,p_0, & n < m \\ \rho_1^{m-1}\,\rho_2^{n-m+1}\,p_0, & n \geq m, \end{cases}$$

where

$$\begin{aligned} \sum_{n=0}^{\infty} p_n &= p_0 \left[\sum_{k=0}^{m-1} \rho_1^k + \rho_1^{m-1}\rho_2 \sum_{n=0}^{\infty} \rho_2^n\right] \\ &= p_0 \left[\frac{1-\rho_1^m}{1-\rho_1} + \frac{\rho_2\rho_1^{m-1}}{1-\rho_2}\right] = 1 \end{aligned}$$

gives

$$p_0 = \left(\frac{1-\rho_1^m}{1-\rho_1} + \frac{\rho_2 \rho_1^{m-1}}{1-\rho_2} \right)^{-1}.$$

(b)

$$\begin{aligned}
L &= \sum_{n=0}^{\infty} n\, p_n \\
&= p_0 \left[\sum_{n=0}^{m-1} n\, \rho_1^n + \sum_{n=m}^{\infty} n\, \rho_1^{m-1}\, \rho_2^{n-m+1} \right] \\
&= p_0 \left[\rho_1 \sum_{n=0}^{m-1} n\, \rho_1^{n-1} + \rho_1 \, \left(\frac{\rho_1}{\rho_2} \right)^{m-2} \, \sum_{n=m}^{\infty} n\, \rho_2^{n-1} \right] \\
&= p_0 \left[\rho_1 \frac{d}{d\rho_1} \, \left(\sum_{n=0}^{m-1} \rho_1^n \right) + \rho_1 \, \left(\frac{\rho_1}{\rho_2} \right)^{m-2} \, \frac{d}{d\rho_2} \sum_{n=m}^{\infty} \rho_2^n \right] \\
&= p_0 \left[\rho_1 \, \frac{d}{d\rho_1} \, \left(\frac{1-\rho_1^m}{1-\rho} \right) + \rho_1 \, \left(\frac{\rho_1}{\rho_2} \right)^{m-2} \, \frac{d}{d\rho} \, \left(\frac{\rho^m}{1-\rho} \right) \right] \\
&= p_0 \left[\frac{\rho_1 [1 + (m-1)\rho_1^m - m\rho_1^{m-1}]}{(1-\rho_1)^2} + \frac{\rho_2 \, \rho_1^{m-1} + [m - (m-1)\rho_2]}{(1-\rho_2)^2} \right].
\end{aligned}$$

16.5 In this case

$$\lambda_i = \begin{cases} \lambda, & j < r \\ p\lambda, & j \geq r \end{cases} \qquad \mu_i = \begin{cases} j\mu, & j < r \\ r\mu, & j \geq r. \end{cases}$$

Using (16-73)-(16-74), this gives

$$p_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!}\, p_0, & n < r \\ \dfrac{(\lambda/\mu)^r}{r!}\, (p\lambda/r\mu)^{n-r}, & n \geq r. \end{cases}$$

16.6

$$\begin{aligned} P\{w>t\} &= \sum_{n=r}^{m-1} p_n \, P(w>t|n) \\ &= \sum_{n=r}^{m-1} p_n \,(1-F_w(t|n)) = \sum p_r \left(\frac{\lambda}{r\mu}\right)^{n-r} \;(1-F_w(t|n)) \end{aligned}$$

$$f_w(t|n) = e^{-\gamma\mu t}\,\frac{(\gamma\mu)^{n-r+1}\,t^{n-r}}{(n-r)!)} \quad (\textit{see } 16.116)$$

and

$$F_w(t|n) = 1-\sum_{k=0}^{n-r}\frac{(\gamma\mu t)^k}{k!}\,e^{-\gamma\mu t} \quad (\textit{see } 4.)$$

so that

$$1-F_w(t|n) = \sum_{k=0}^{n-r}\frac{(\gamma\mu t)^k}{k!}\,e^{-\gamma\mu t}$$

$$\begin{aligned} P\{w>t\} &= \sum_{n=r}^{m-1} p_r \left(\frac{\lambda}{\gamma\mu}\right)^{n-r} \sum_{k=0}^{n-r}\frac{(\gamma\mu t)^k}{k!}\,e^{-\gamma\mu t} \\ &= \sum_{i=0}^{m-r-1} p_r\,\rho^i \sum_{k=0}^{i}\frac{(\gamma\mu t)^k}{k!}\,e^{-\gamma\mu t}, \quad n-r=i \\ &= p_r\,e^{-\gamma\mu t}\sum_{k=0}^{m-r-1}\rho^k\sum_{i=0}^{k}\frac{(\gamma\mu t)^i}{i!} \\ &= \sum_{k=0}^{m-r-1}\sum_{i=0}^{k} = \sum_{i=0}^{m-r-1}\sum_{k=i}^{m-r-1} \end{aligned}$$

$$\begin{aligned} P\{w>t\} &= p_r\,e^{-\gamma\mu t}\sum_{i=0}^{m-r-1}\frac{(\gamma\mu t)^i}{i!}\sum_{k=i}^{m-r-1}\rho^k \\ &= \frac{p_r}{1-\rho}\,e^{-\gamma\mu t}\sum_{i=0}^{m-r-1}\frac{(\gamma\mu t)^i}{i!}\,(\rho^i-\rho^{m-r}), \quad \rho=\lambda/\gamma\mu. \end{aligned}$$

Note that $m \to \infty \Longrightarrow M/M/r/m \Longrightarrow M/M/r$ and

$$\begin{aligned} P(w > t) &= \frac{p_r}{1-\rho}\, e^{-\gamma\mu t} \sum_{i=0}^{\infty} \frac{(\gamma\mu\rho t)^i}{i!} \\ &= \frac{p_r}{1-\rho}\, e^{-\gamma\mu(1-\rho)t} \quad t > 0. \end{aligned}$$

and it agrees with (16.119)

16.7 (a) Use the hints

(b)

$$-\sum_{n=1}^{\infty} (\lambda+\mu)\, p_n\, z^n + \frac{\mu}{z} \sum_{n=1}^{\infty} p_{n+1}\, z^{n+1} + \lambda \sum_{n=1}^{\infty} \sum_{k=1}^{n} p_{n-k}\, c_k\, z^n = 0$$

$$-(\rho+1)\,(P(z)-p_0) + \frac{\mu}{z}\,(P(z)-p_0-p_1 z) + \lambda \sum_{k=1}^{\infty} c_k z^k \sum_{m=0} p_m\, z^m = 0$$

which gives

$$P(z)[1-z-\rho z\,(1-C(z))] = p_0(1-z)$$

or

$$P(z) = \frac{p_0(1-z)}{1-z-\rho z\,(1-C(z))}.$$

$$\begin{aligned} 1 = P(1) = \frac{-p_0}{-1-\rho+\rho_z\, C'(z)+\rho C(z)} &= \frac{-p_0}{-1+\rho C'(1)} \\ \Longrightarrow p_0 = 1-\rho_0, \quad \rho_0 &= \rho C'(1). \end{aligned}$$

Let

$$D(z) = \frac{1-C(z)}{1-z}.$$

Then

$$P(z) = \frac{1-\rho_L}{1-\rho z D(z)}.$$

(c) This gives

$$P'(z) = \frac{(1-\rho_c)}{(1-\rho z D(z))^2}\ (\rho D(z) + \rho z D'(z))$$

$$\begin{aligned} L = P'(1) &= \frac{(1-\rho_c)}{(1-\rho_c)^2}\,\rho\,(D(1)+D'(1)) \\ &= \frac{1}{(1-\rho_c)}\;(C'(1)+D'(1)) \end{aligned}$$

$$C'(1) = E(x)$$

$$D(z) = \frac{1-C(z)}{1-z}$$

$$\begin{aligned} D'(z) &= \frac{(1-z)\,(-C'(z)) - (1-C(z))\,(-1)}{(1-z)^2} \\ &= \frac{1-C(z)-(1-z)C'(z)}{(1-z)^2} \end{aligned}$$

By L-Hopital's Rule

$$\begin{aligned} D'(1) &= \lim_{z\to 1} \frac{-C'(z)-(-1)C'(z)-(1-z)C''(z)}{-2(1-z)} \\ &= \lim_{z\to 1} = 1/2C''(z) = \frac{C''(z)}{2} \\ &= 1/2\sum k(k-1)\,C_k = \frac{E(X^2)-E(X)}{2} \end{aligned}$$

$$L = \frac{\rho\,(E(X)+E(X^2))}{2\,(1-\rho E(X))}.$$

(d)

$$C(z)z^m \quad E(X) = m$$

$$P(z) = \frac{1-\rho}{1-\rho\sum_{k=1}^{m} z^k}$$

$$D(z) = \frac{1-z^m}{1-z} = \sum_{k=0}^{m-1} z^k$$

$$E(X) = m, \quad E(X^2) = m^2$$

$$L = \frac{\rho(m+m^2)}{2(1-\rho m)}$$

(e)

$$C(z) = \frac{qz}{1 - Pz}$$

$$P(z) = \frac{1-\rho_0}{1-\rho z D(z)}, \quad C(z) = \frac{qz}{1-pz}$$

$$D(z) = \frac{1-C(z)}{1-z} = \frac{1-\frac{qz}{1-P(z)}}{1-z} = \frac{1-Pz-(1-P)_z}{(1-z)(1-Pz)} = \frac{1-z}{(1-z)(1-Pz)} = \frac{1}{1-Pz}$$

$$P(z) = \frac{(1-\rho_0)(1-pz)}{1-pz-\rho z} = \frac{(1-\rho_0)(1-pz)}{1-(p+\rho)z}$$

$$C'(1) = \frac{(1-pz)q - qz(-p)}{(1-Pz)^2} = \frac{q}{q^2} = \frac{1}{q}$$

$$D(z) = \frac{1-C(z)}{1-z}$$

$$D(1) = C'(1)$$

$$L = P'(1) = \frac{1-\rho_c}{(1-\rho_c)^2}\ (\rho \cdot C'(1) + \rho \cdot D'(1))$$

$$D'(z) = \frac{-(1-z)C'(z) - (1-C(z))\,(\rho-1)}{(1-z)^2} = \frac{1-C(z)-(1-z)C'(1)}{(1-z)^2}$$

$$\begin{aligned} \lim\nolimits_{z\to 1} D'(z) \quad &= \lim\nolimits_{z\to 1} \frac{-C'(z)-(-1)C'(z)-(1-z)C''(z)}{2(1-z)} \\ &= \frac{-(1-z)C''(z)}{-2(1-z)} = \frac{\rho''(z)}{2} \end{aligned}$$

$$D'(1) = \frac{C''(1)}{2}$$

$$L = \frac{1}{(1-\rho_c)}\left(\rho E(X) + \frac{\rho\,(E(X^2)-E(X))}{2}\right) = \frac{\rho E(X) + \rho E(X^2)}{2(1-\rho_c)}.$$

16.8 (a) Use the hints.
(b)

$$-\sum_{n=1}^{\infty} (\lambda+\mu)\, p_n\, z^n + \frac{\mu}{z^n}\sum_{n=1}^{\infty} p_{n+m}\, z^{n+m} + \lambda z \sum_{n=1}^{\infty} p_{n-1}\, z^{n-1} = 0$$

or

$$-(1+\rho)\,(P(z)-p_0)+\frac{1}{z^m}\left(P(z)-\sum_{k=0}^{m}p_k\,z^k\right)+\rho zP(z)=0$$

which gives

$$P(z)\left[\rho\,z^{m+1}-(\rho+1)\,z^m+1\right]=\sum_{k=0}^{m}p_k\,z^k-p_0\,(1+\rho)\,z^m$$

or

$$P(z)=\frac{\displaystyle\sum_{k=0}^{m}p_k\,z^k-p_0\,(1+\rho)\,z^m}{\rho\,z^{m+1}-(\rho+1)\,z^m+1}=\frac{N(z)}{M(z)}. \tag{i}$$

(c) Consider the denominator polynomial $M(z)$ in (i) given by

$$M(z)=\rho\,z^{m+1}-(1+\rho)\,z^m+1=f(z)+g(z)$$

where

$$f(z)=-(1+\rho)\,z^m,$$
$$g(z)=1+\rho\,z^{m+1}.$$

Notice that $|f(z)|>|g(z)|$ in a circle defined by $|z|=1+\varepsilon$, $\varepsilon>0$. Hence by Rouche's Theorem $f(z)$ and $f(z)+g(z)$ have the same number of zeros inside the unit circle ($|z|=1+\varepsilon$). But $f(z)$ has m zeros inside the unit circle. Hence $f(z)+g(z)=M(z)$ also has m zeros inside the unit circle. Hence

$$M(z)=M_1(z)\,(z-z_0) \tag{ii}$$

where $|z_0|>1$ and $M_1(z)$ is a polynomial of degree m whose zeros are all *inside* or on the unit circle. But the moment generating function $P(z)$ is analytic inside and on the unit circle. Hence all the m zeros of $M(z)$ that are inside or on the unit circle must cancel out with the zeros of the numerator polynomial of $P(z)$. Hence

$$N(z)=M_1(z)\,a. \tag{iii}$$

Using (ii) and (iii) in (i) we get

$$P(z) = \frac{N(z)}{M(z)} = \frac{a}{z - z_0}.$$

But $P(1) = 1$ gives $a = 1 - z_0$
or

$$\begin{aligned} P(z) &= \frac{z_0 - 1}{z_0 - z} \\ &= \left(1 - \frac{1}{z_0}\right) \sum_{n=0}^{\infty} (z/z_0)^n \end{aligned}$$

$$\Longrightarrow p_n = \left(1 - \frac{1}{z_0}\right) \left(\frac{1}{z_0}\right)^n = (1 - r)\, r^n, \quad n \geq 0 \tag{iv}$$

where $r = 1/z_0$.

(d) Average system size

$$L = \sum_{n=0}^{\infty} n\, p_n = \frac{r}{1 - r}.$$

16.9 (a) Use the hints in the previous problem.
(b)

$$\begin{aligned} -\sum_{n=m}^{\infty} (\lambda + \mu)\, p_n\, z^n + \mu \sum_{n=m}^{\infty} p_{n+m}\, z^n + \lambda \sum_{n=m}^{\infty} p_{n-1}\, z^n \\ -(1+\rho)\left(P(z) - \sum_{k=0}^{m-1} p_k\, z^k\right) + \frac{1}{z^m}\left(P(z) - \sum_{k=0}^{2m-1} p_k\, z^k\right) \\ +\rho\, z \left(P(z) - \sum_{k=0}^{m-2} p_k\, z^k\right) = 0. \end{aligned}$$

After some simplifications we get

$$P(z)\left[\rho\, z^{m+1} - (\rho+1)\, z^m + 1\right] = (1 - z^m) \sum_{k=0}^{m-1} p_k\, z^k$$

or

$$P(z) = \frac{(1 - z^m) \sum_{k=0}^{m-1} p_k\, z^k}{\rho\, z^{m+1} - (\rho+1)\, z^m + 1} = \frac{(z_0 - 1) \sum_{k=0}^{m-1} z^k}{m\,(z_0 - z)}$$

where we have made use of Rouche's theerem and $P(z) \equiv 1$ as in problem 16-8.
(c)

$$P(z) = \sum_{n=0}^{\infty} p_n\, z^n = \frac{1-r}{m}\, \frac{\sum_{k=0}^{m-1} z^k}{1 - rz}$$

gives

$$p_n = \begin{cases} (1 + r + \cdots + r^k)\, p_0, & k \le m-1 \\ r^{n-m+1}\, (1 + r + \cdots + r^{m-1})\, p_0, & k \ge m \end{cases}$$

where

$$p_0 = \frac{1-r}{m}, \qquad r = \frac{1}{z_0}.$$

Finally

$$L = \sum_{n=0}^{\infty} n\, p_n = P_n^{'}(1).$$

But

$$P^{'}(z) = \left(\frac{1-r}{m}\right) \frac{\sum_{k=1}^{m-1} k\, z^{k-1}\, (1 - rz) - \sum_{k=0}^{m-1} z^k\, (-r)}{(1 - rz)^2}$$

so that

$$\begin{aligned} L &= P^{'}(1) = \frac{1-r}{m}\, \frac{m - 1 + r}{(1-r)^2} = \frac{m - (1-r)}{m\,(1-r)} \\ &= \frac{1}{1-r} - \frac{1}{m}. \end{aligned}$$

16.10 Proceeding as in (16-212),

$$\begin{aligned} \psi_A(u) &= \int_0^{\infty} e^{-u\tau} dA(\tau) \\ &= \left(\frac{\lambda\, m}{u + \lambda\, m\, z}\right)^m. \end{aligned}$$

This gives

$$\begin{aligned} B(z) &= \psi_A(\psi(1-z)) \\ &= \left(\frac{\lambda\, m}{\mu\,(1-z)+\lambda\, m}\right)^m \\ &= \left(\frac{1}{1+\frac{1}{\rho}(1-z)}\right)^m \\ &= \left(\frac{\rho}{(1+\rho)-z}\right)^m, \qquad \rho = \frac{\lambda}{m\,\mu}. \end{aligned} \tag{i}$$

Thus the equation $B(z) = z$ for π_0 reduce to

$$\left(\frac{\rho}{(1+\rho)-z}\right)^m = z$$

or

$$\frac{\rho}{(1+\rho)-z} = z^{1/m},$$

which is the same as

$$\rho\, z^{-1/m} = (1+\rho) - z \tag{ii}$$

Let $x = z^{-1/m}$. Sustituting this into (ii) we get

$$\rho\, x = (1+\rho) - x^{-m}$$

or

$$\rho\, x^{m+1} - (1+\rho)\, x^m + 1 = 0 \tag{iii}$$

16.11 From Example 16.7, Eq.(16-214), the characteristic equation for $Q(z)$ is given by ($\rho = \lambda/m\,\mu$)

$$1 - z[1+\rho\,(1-z)]^m = 0$$

which is equivalent to

$$1 + \rho\,(1 - z) = z^{-1/m}. \tag{i}$$

Let $x = z^{1/m}$ in this case, so that (i) reduces to

$$[(1 + \rho) - \rho\, x^m]\, x = 1$$

or the characteristic equation satisfies

$$\rho\, x^{m+1} - (1 + \rho)\, x + 1 = 0. \tag{ii}$$

16.12 Here the service time distribution is given by

$$\frac{dB(t)}{dt} = \sum_{i=1}^{k} d_i\, \delta(t - T_i)$$

and this Laplace transform equals

$$\Phi_s(s) = \sum_{i=1}^{k} d_i\, e^{-s\,T_i} \tag{i}$$

substituting (i) into (15.219), we get

$$\begin{aligned} A(z) &= \Phi_s\,(\lambda\,(1 - z)) \\ &= \sum_{i=1}^{k} d_i\, e^{-\lambda\, T_i\,(1-z)} \\ &= \sum_{i=1}^{k} d_i\, e^{-\lambda\, T_i}\, e^{\lambda\, T_i\, z} \\ &= \sum_{i=1}^{k} d_i\, e^{-\lambda\, T_i} \sum_{j=0}^{\infty} \tfrac{(\lambda\, T_i)^j\, z^j}{j!} = \sum_{j=0}^{\infty} a_j\, z^j. \end{aligned}$$

Hence

$$a_j = \sum_{i=1}^{k} d_i\, e^{-\lambda\, T_i}\, \frac{(\lambda\, T_i)^j}{j!}, \qquad j = 0, 1, 2, \cdots. \tag{i}$$

To get an explicit formula for the steady state probabilities $\{q_n\}$, we can make use of the analysis in (16.194)-(16.204) for an $M/G/1$ queue. From (16.203)-(16.204), let

$$c_0 = 1 - a_0, \qquad c_n = 1 - \sum_{k=0}^{n} a_k, \qquad n \geq 1$$

and let $\{c_k^{(m)}\}$ represent the $m-$fold convolution of the sequence $\{c_k\}$ with itself. Then the steady-state probabilities are given by (16.203) as

$$q_n = (1-\rho) \sum_{m=0}^{\infty} \sum_{k=0}^{n} a_k \, c_{n-k}^{(m)}.$$

(b) *State-Dependent Service Distribution*
Let $B_i(t)$ represent the service-time distribution for those customers entering the system, where the most recent departure left i customers in the queue. In that case, (15.218) modifies to

$$a_{k,i} = P\{A_k | B_i\}$$

where

$$A_k = \text{"}k \text{ customers arrive during a service time"}$$

and

$$B_i = \text{"}i \text{ customers in the system at the most recent departure."}$$

This gives

$$\begin{aligned} a_{k,i} &= \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \, dB_i(t) \\ &= \begin{cases} \displaystyle\int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \mu_1 \, e^{-\mu_1 t} \, dt = \frac{\mu_1 \lambda^k}{(\lambda+\mu_1)^{k+1}}, & i = 0 \\ \displaystyle\int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \mu_2 \, e^{-\mu_2 t} \, dt = \frac{\mu_2 \lambda^k}{(\lambda+\mu_2)^{k+1}}, & i \geq 1 \end{cases} \end{aligned} \qquad \text{(i)}$$

This gives

$$A_i(z) = \sum_{k=0}^{\infty} a_{k,i}\, z^k = \begin{cases} \dfrac{1}{1+\rho_1(1-z)}, & i = 0 \\ \dfrac{1}{1+\rho_2(1-z)}, & i \geq 1 \end{cases} \tag{ii}$$

where $\rho_1 = \lambda/\mu_1$, $\rho_2 = \lambda/\mu_2$. Proceeding as in Example 15.24, the steady state probabilities satisfy [(15.210) gets modified]

$$q_j = q_0\, a_{j,0} + \sum_{i=1}^{j+1} q_i\, a_{j-i+1,i} \tag{iii}$$

and (see(15.212))

$$\begin{aligned} Q(z) &= \sum_{j=0}^{\infty} q_j\, z^j \\ &= q_0 \sum_{j=0}^{\infty} a_{j,0}\, z^j + \sum_{j=0}^{\infty} q_i\, a_{j-i+1,i} \\ &= q_0\, A_0(z) + \sum_{i=1}^{\infty} q_i\, z^i \sum_{m=0}^{\infty} a_{m,i}\, z^m\, z^{-1} \\ &= q_0\, A_0(z) + (Q(z) - q_0)\, A_1(z)/z \end{aligned} \tag{iv}$$

where (see (ii))

$$A_0(z) = \frac{1}{1+\rho_1(1-z)} \tag{v}$$

and

$$A_1(z) = \frac{1}{1+\rho_2(1-z)}. \tag{vi}$$

From (iv)

$$Q(z) = \frac{q_0\,(z\, A_0(z) - A_1(z))}{z - A_1(z)}. \tag{vii}$$

Since

$$
\begin{aligned}
Q(1) &= 1 = \frac{q_0\,\left[A_0'(1) + A_0(1) - A_1'(1)\right]}{1 - A_1'(1)} \\
&= \frac{q_0\,(1 + \rho_1 - \rho_2)}{1 - \rho_2}
\end{aligned}
$$

we obtain

$$
q_0 = \frac{1 - \rho_2}{1 + \rho_1 - \rho_2}. \tag{viii}
$$

Substituting (viii) into (vii) we can rewrite $Q(z)$ as

$$
\begin{aligned}
Q(z) &= (1 - \rho_2)\,\frac{(1 - z)\,A_1(z)}{A_1(z) - z} \cdot \frac{1}{1 + \rho_1 - \rho_2}\,\frac{1 - z\,A_0(z)/A_1(z)}{1 - z} \\
&= \left(\frac{1 - \rho_2}{1 - \rho_2\,z}\right) \frac{1}{1 + \rho_1 - \rho_2}\,\frac{1 - \frac{\rho_2}{1+\rho_1}\,z}{1 - \frac{\rho_1}{1+\rho_1}\,z} \\
&= Q_1(z)\,Q_2(z)
\end{aligned} \tag{ix}
$$

where

$$
Q_1(z) = \frac{1 - \rho_2}{1 - \rho_2\,z} = (1 - \rho_2)\sum_{k=0}^{\infty} \rho_2^k\,z^k
$$

and

$$
Q_2(z) = \frac{1}{1 + \rho_1 - \rho_2}\,\left(1 - \frac{\rho_2}{1 + \rho_1}\,z\right)\sum_{i=0}^{\infty}\left(\frac{\rho_1}{1 + \rho_1}\right)^i\,z^i.
$$

Finally substituting. $Q_1(z)$ and $Q_2(z)$ into (ix) we obtain

$$
q_n = q_0\left[\sum_{i=0}^{n}\left(\frac{\rho_1}{1 + \rho_1}\right)^{n-i}\,\rho_2^i - \sum_{i=0}^{n-1}\rho_2^{i+1}\,\frac{\rho_1^{n-i-1}}{(1 + \rho_1)^{n-i}}\right]. \quad n = 1, 2, \cdots
$$

with q_0 as in (viii).

16.13 From (16-209), the Laplace transform of the waiting time distribution is given by

$$\begin{aligned}\Psi_w(s) &= \frac{1-\rho}{1-\lambda\left(\frac{1-\Phi_s(s)}{s}\right)} \\ &= \frac{1-\rho}{1-\rho\,\mu\left(\frac{1-\Phi_s(s)}{s}\right)}.\end{aligned} \tag{i}$$

Let

$$\begin{aligned}F_r(t) &= \mu\int_0^t [1-B(\tau)]d\tau \\ &= \mu\left[t-\int_0^t B(\tau)d\tau\right].\end{aligned} \tag{ii}$$

represent the residual service time distribution. Then its Laplace transform is given by

$$\begin{aligned}\Phi_F(s) &= L\left\{F_r(t)\right\} = \mu\left(\frac{1}{s}-\frac{\Phi_s(s)}{s}\right) \\ &= \mu\left(\frac{1-\Phi_s(s)}{s}\right).\end{aligned} \tag{iii}$$

Substituting (iii) into (i) we get

$$\Psi_w(s) = \frac{1-\rho}{1-\rho\,\Phi_F(s)} = (1-\rho)\sum_{n=0}^{\infty}[\rho\,\Phi_F(s)]^n, \quad |\Phi_F(s)|<1. \tag{iv}$$

Taking inverse transform of (iv) we get

$$F_w(t) = (1-\rho)\sum_{n=0}^{\infty}\rho^n\,F_r^{(n)}(t),$$

where $F_r^{(n)}(t)$ is the n^{th} convolution of $F_r(t)$ with itself.

16.14 Let ρ in (16.198) that represents the average number of customers that arrive during any service period be greater than one. Notice that

$$\rho = A^{'}(1) > 1$$

where

$$A(z) = \sum_{k=0}^{\infty} a_k \, z^k$$

From Theorem 15.9 on Extinction probability (pages 759-760) it follows that if $\rho = A^{'}(1) > 1$, the eqution

$$A(z) = z \tag{i}$$

has a unique positive root $\pi_0 < 1$. On the other hand, the transient state probabilities $\{\sigma_i\}$ satisfy the equation (15.236). By direct substitution with $x_i = \pi_0^i$ we get

$$\sum_{j=1}^{\infty} p_{ij} \, x_j = \sum_{j=1}^{\infty} a_{j-i+1} \, \pi_0^j \tag{ii}$$

where we have made use of $p_{ij} = a_{j-i+1}$, $i \geq 1$ in (15.33) for an $M/G/1$ queue. Using $k = j - i + 1$ in (ii), it reduces to

$$\sum_{k=2-i}^{\infty} a_k \, \pi_0^{k+i-1} = \pi_0^{i-1} \sum_{k=0}^{\infty} a_k \, \pi_0^k$$

$$= \pi_0^{i-1} \, \pi_0 = \pi_0^i = x_i \tag{iii}$$

since π_0 satisfies (i). Thus if $\rho > 1$, the $M/G/1$ system is transient with probabilities $\sigma_i = \pi_0^i$.

16.15 (a) The transition probability matrix here is the truncated version of (15.34) given by

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & \cdot & \cdot & \cdot & \cdot & a_{m-2} & 1 - \sum_{k=0}^{m-2} a_k \\ a_0 & a_1 & a_2 & \cdot & \cdot & \cdot & \cdot & a_{m-2} & 1 - \sum_{k=0}^{m-2} a_k \\ 0 & a_0 & a_1 & \cdot & \cdot & \cdot & \cdot & a_{m-3} & 1 - \sum_{k=0}^{m-3} a_k \\ \vdots & \vdots & \vdots & & & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & a_0 & a_1 & 1-(a_0+a_1) \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & a_0 & 1 - a_0 \end{pmatrix} \tag{i}$$

and it corresponds to the upper left hand block matrix in (15.34) followed by an m^{th} column that makes each row sum equal to unity.

(b) By direct sybstitution of (i) into (15-167), the steady state probabilities $\{q_j^*\}_{j=0}^{m-1}$ satisfy

$$q_j^* = q_0^* \, a_j + \sum_{i=1}^{j+1} q_i^* \, a_{j-i+1}, \qquad j = 0, 1, 2, \cdots, m-2 \tag{ii}$$

and the normalization condition gives

$$q_{m-1}^* = 1 - \sum_{i=0}^{m-2} q_i^*. \tag{iii}$$

Notice that (ii) in the same as the first $m-1$ equations in (15-210) for an $M/G/1$ queue. Hence the desired solution $\{q_j^*\}_{j=0}^{m-1}$ must satisfy the first $m-1$ equations in (15-210) as well. Since the unique solution set to (15.210) is given by $\{q_j\}_{j=0}^{\infty}$ in (16.203), it follows that the desired probabilities satisfy

$$q_j^* = c \, q_j, \qquad j = 0, 1, 2, \cdots, m-1 \tag{iv}$$

where $\{q_j\}_{j=0}^{m-1}$ are as in (16.203) for an $M/G/1$ queue. From (iii) we also get the normalization constant c to be

$$c = \frac{1}{\sum_{i=0}^{m-1} q_i}. \tag{v}$$

16.16 (a) The event $\{X(t) = k\}$ can occur in several mutually exclusive ways, *viz.*, in the interval $(0, t)$, n customers arrive and k of them continue their service beyond t. Let A_n = "n arrivals in $(0, t)$", and $B_{k,n}$ ="exactly k services among the n arrivals continue beyond t", then by the theorem of total probability

$$P\{X(t) = k\} = \sum_{n=k}^{\infty} P\{A_n \cap B_{k,n}\} = \sum_{n=k}^{\infty} P\{B_{k,n}|A_n\}P(A_n).$$

But $P(A_n) = e^{-\lambda t}(\lambda t)^n/n!$, and to evaluate $P\{B_{k,n}|A_n\}$, we argue as follows: From (9.28), under the condition that there are n arrivals in $(0, t)$, the joint distribution of the arrival instants agrees with the joint distribution of n independent random variables arranged in increasing order and distributed uniformly in $(0, t)$. Hence the probability that a service time S does not terminate by t, given that its starting time $\mathbf{x}$ has a uniform distribution in $(0, t)$ is given by

$$p_t = \int_0^t P(S > t - x | \mathbf{x} = x) f_{\mathbf{x}}(x) dx$$

$$= \int_0^t [1 - B(t - x)] \frac{1}{t} dx = \frac{1}{t} \int_0^t (1 - B(\tau))\, d\tau = \frac{\alpha(t)}{t}$$

Thus $B_{k,n}$ given A_n has a Binomial distribution, so that

$$P\{B_{k,n}|A_n\} = \binom{n}{k} p_t^k (1 - p_t)^{n-k}, \quad k = 0, 1, 2, \cdots n,$$

and

$$\begin{aligned}
P\{X(t)=k\} &= \sum_{n=k}^{\infty} e^{-\lambda t}\,\frac{(\lambda t)^n}{n!}\binom{n}{k}\left(\frac{\alpha(t)}{t}\right)^k\,\left(\frac{1}{t}\int_0^t B(\tau)d\tau\right)^{n-k} \\
&= e^{-\lambda t}\,\frac{[\lambda\alpha(t)]^k}{k!}\sum_{n=k}^{\infty}\frac{\left(\lambda t\frac{1}{t}\int_0^t B(\tau)d\tau\right)^{n-k}}{(n-k)!} \\
&= \frac{[\lambda\alpha(t)]^k}{k!}\,e^{-\lambda\left[t-\int_0^t B(\tau)d\tau\right]} \\
&= \frac{[\lambda\alpha(t)]^k}{k!}\,e^{-\lambda\int_0^t[1-B(\tau)]d\tau} \\
&= \frac{[\lambda\alpha(t)]^k}{k!}\,e^{-\lambda\,\alpha(t)}, \qquad k=0,1,2,\cdots
\end{aligned} \tag{i}$$

(b)

$$\begin{aligned}
\lim_{t\to\infty}\alpha(t) &= \int_0^{\infty}[1-B(\tau)]d\tau \\
&= E\{\mathbf{s}\}
\end{aligned} \tag{ii}$$

where we have made use of (5-52)-(5-53). Using (ii) in (i), we obtain

$$\lim_{t\to\infty} P\{x(t)=k\} = e^{-\rho}\,\frac{\rho^k}{k!} \tag{iii}$$

where $\rho = \lambda\,E\{\mathbf{s}\}$.

CHAPTER 11

11-1 $$S_x(z) = \frac{5-2(z+1/z)}{10-3(z+1/z)} = \frac{2}{3} + \frac{5/9}{10/3 - (z+1/z)}$$

$$R[m] = \frac{2}{3} + \frac{5}{18}\, 3^{-|m|} \qquad \Gamma(z) = \frac{3z-1}{2z-1}$$

11-2 $$S_x(s) = \frac{s^4+64}{s^4-10s^2+9} = \frac{s^2+4s+8}{s^2+4s+3}\,\frac{s^2-4s+8}{s^2-4s+3}$$

$$L(s) = \frac{s^2+4s+8}{s^2+4s+3}$$

11-3 First proof

$$\underline{s}[n] = \sum_{k=0}^{\infty} \ell[n]\underline{i}[n-k] \qquad E\{\underline{x}^2[n]\} = \sum_{k=0}^{\infty} \ell^2[k]$$

Second proof

$$S(z) = L(z)L(1/z) \qquad R[m] = \ell[m] * \ell[-m] = \sum_{k=0}^{\infty} \ell[k]\ell[k-m]$$

$$R[0] = \sum_{k=0}^{\infty} \ell^2[k]$$

11-4 (a) This is a special case of (11-22) and (11-23).

(b) From (a) it follows that

$$R''_{yx}(\tau) + 3\,R'_{yx}(\tau) + 2\,R_{yx}(\tau) = q\delta(\tau)$$

Since $R_{xx}(\tau) = 0$ for $\tau < 0$, the above shows that

$$R_{yx}(\tau) = 0 \text{ for } \tau \le 0^- \qquad R'_{yx}(0^-) = 0$$

Furthermore,

$$S_{yx}(s) = \frac{q}{s^2 + 3s + 2}$$

hence (initial value theorem)

$$R_{yx}(0^+) = \lim_{s\to\infty} s\; S_{yx}(s) = 0 \qquad R'_{yx}(0^+) = \lim_{s\to\infty} s^2\; S_{yx}(s) = q$$

Similarly,

$$R''_{yy}(\tau) + 3\,R'_{yy}(\tau) + 2\,R_{yy}(\tau) = R_{xy}(\tau) = R_{yx}(-\tau) = 0 \quad \text{for } \tau > 0$$

$$S_{yy}(s) = \frac{q}{(s^2 + 3s + 2)(s^2 - 3s + 2)} = \frac{qs/12 + q/4}{s^2 + 3s + 2} + \frac{-qs/12 + q/4}{s^2 - 3s + 2}$$

$$S^+_{yy}(s) = \frac{qs/12 + q/4}{s^2 + 3s + 2}$$

$$R^+_{yy}(0^+) = R_{yy}(0) = \lim_{s\to\infty} s^2\; S^+_{yy}(s) = \frac{q}{12}$$

$$R'_{yy}(0) = \lim_{s\to\infty} s\,[s\,S^+_{yy}(s) - \frac{q}{12}] = 0$$

11-5 $$S_x(z) = S_s(z) + S_\nu(z) = \frac{1}{D(z)} + q = \frac{1 + qD(z)}{D(z)}$$

If $R_s[m] = 2^{-|m|}$ and $S_\nu(z) = 5$, then (see Example 9-31)

$$S_s(z) = \frac{1.5}{2.5 - (z^{-1} + z)} \qquad S_x(z) = \frac{5 - 14z^{-1} + 5z^{-2}}{1 - 2.5z^{-1} + z^{-2}}$$

11-6 The process

$$\underset{\sim}{y}[n] = \frac{1}{n}\sum_{k=1}^{n} \underset{\sim}{x}(nT + kT)$$

is the output of a system with input $\underset{\sim}{x}[n]$ and system function

$$H(z) = \frac{1}{n}\sum_{k=1}^{n} z^k$$

Furthermore, $\underset{\sim}{s} = \underset{\sim}{y}[0]$ and

$$n^2\left|H(e^{j\omega T})\right|^2 = \left|\sum_{k=1}^{n} e^{jk\omega T}\right|^2$$

$$= \left|\frac{e^{j\omega T} - e^{j(n+1)\omega T}}{1 - e^{j\omega T}}\right|^2 = \frac{\sin^2 n\omega T/2}{\sin^2 \omega T/2}$$

Hence [see (9-51)]

$$E\{\underset{\sim}{s}^2\} = R_y[0] = \frac{1}{2\pi n^2}\int_{-\infty}^{\infty} S_x(\omega)\,\frac{\sin^2 n\omega T/2}{\sin^2 \omega T/2}\,d\omega$$

11-7 Since $R(t_1,t_2) = e^{-c|t_1-t_2|}$, (12-58) yields

$$\int_{-a}^{t_1} e^{-c(t_1-t_2)}\phi(t_2)dt_2 + \int_{t_1}^{a} e^{c(t_1-t_2)}\phi(t_2)dt_2 = \lambda\phi(t_1) \qquad \text{(i)}$$

Differentiating twice and using (i) we obtain (omitting details)

$$\lambda\phi''(t) + (2c - \lambda c^2)\phi(t) = 0$$

Hence,

$$\phi(t) = \beta\cos\omega t \quad \text{and} \quad \phi(t) = \beta'\cos\omega' t$$

To determine ω, we insert into (i). This yields

$$\frac{2c}{c^2+\omega^2} + \frac{\omega\sin a\omega - c\cos a\omega}{c^2+\omega^2}\, e^{-ac}\,(e^{ct} + e^{-ct}) = 2c\,\lambda\cos\omega t$$

This yields

$$\omega_n \sin a\omega_n - c\cos a\omega_n = 0 \qquad\qquad \lambda_n = \frac{2c}{c^2+\omega_n^2}$$

The constants β_n are determined from (normalization)

$$1 = \int_{-a}^{a} \beta_n^2 \cos^2\omega_n t\,dt \qquad\qquad \beta_n^2 = \frac{1}{a+c\lambda_n}$$

Similarly for $\beta'_n \sin\omega'_n t$.

11-8 As in (9-60)

$$E\{|\underline{X}_T(\omega)|^2\} = \int_{-T/2}^{T/2} R(t_1 - t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2$$

$$= \int_{-T}^{T} (T - |\tau|)R(\tau)\, e^{-j\omega\tau}d\tau$$

Differentiating with respect to T and using the fact that if

$$\phi(t) = \int_{-t}^{t} f(x;t)dx$$

then

$$\frac{d\phi(t)}{dt} = f(t;t) - f(-t,t) + \int_{-t}^{t} \frac{\partial f}{\partial t}(x,t)dx$$

we obtain

$$\frac{\partial E\{|\underline{X}_T(\omega)|^2\}}{\partial T} = \int_{-T}^{T} R(\tau)e^{-j\omega\tau}d\tau = E\{\frac{\partial}{\partial T}|\underline{X}_T(\omega)|^2\}$$

The above approaches $S(\omega)$ as $T \to \infty$.

11-9 $$E\{\underset{\sim}{X}(\omega)\} = \int_{-a}^{a} 5\cos 3t\, e^{-j\omega t}dt = \frac{5\sin a(\omega-3)}{\omega-3} + \frac{5\sin a(\omega+3)}{\omega+3}$$

Var. $\underset{\sim}{X}(\omega) = 2qa = 4a.$

11-10 $$E\{\underset{\sim}{X}(u)\underset{\sim}{X}(v)\} = \sum_{n=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} \sigma_n^2\delta[n-k]\, e^{-j(nu-kv)T}$$

$$= \sum_{n=-\infty}^{\infty} \sigma_n^2\, e^{-jn(u-v)T}$$

11-11 Shifting the origin, we set

$$\underset{\sim}{c}_n = \frac{1}{T}\int_{-T/2}^{T/2} \underset{\sim}{x}(t)e^{-jn\omega_o t}dt \qquad \beta_n(\alpha) = \frac{1}{T}\int_{-T/2}^{T/2} R(\tau-\alpha)e^{-jn\omega_o\tau}d\tau$$

(a) We shall show that if

$$\hat{\underset{\sim}{x}}(t) = \sum_{n=-\infty}^{\infty} \underset{\sim}{c}_n e^{jn\omega_o t} \text{ then } E\{|\underset{\sim}{x}(t)-\hat{\underset{\sim}{x}}(t)|^2\} = 0 \text{ for } |t| < T/2 \qquad \text{(i)}$$

<u>Proof</u> $$E\{\underset{\sim}{c}_n\underset{\sim}{x}^*(\alpha)\} = \frac{1}{T}\int_{-T/2}^{T/2} E\{\underset{\sim}{x}(t)\underset{\sim}{x}^*(\alpha)\}e^{-jn\omega_o t}dt = \beta_n(\alpha)$$

The functions $\beta_n(\alpha)$ are the coefficients of the Fourier expansion of $R(\tau-\alpha)$:

$$R(\tau-\alpha) = \sum_{n=-\infty}^{\infty} \beta_n(\alpha)e^{jn\omega_o\tau} \qquad |\tau| < T/2 \qquad \text{(ii)}$$

Hence

$$E\{\hat{\underset{\sim}{x}}(t)\underset{\sim}{x}^*(t)\} = \sum_{n=-\infty}^{\infty} E\{\underset{\sim}{c}_n\underset{\sim}{x}^*(t)\}^{jn\omega_o t} = \sum_{n=-\infty}^{\infty} \beta_n(t)e^{jn\omega_o t}$$

From (ii) it follows with $\tau = \alpha = t$ that the last sum equals R(0). Similarly, $E\{\hat{\underset{\sim}{x}}^*(t)\underset{\sim}{x}(t)\} = R(0)$ and (i) results.

(b) $E\{\underset{\sim}{c}_n\underset{\sim}{c}_m^*\} = \frac{1}{T}\int_{-T/2}^{T/2} E\{c_n\underset{\sim}{x}^*(t)\}e^{jn\omega_o t}dt = \frac{1}{T}\int_{-T/2}^{T/2} \beta_n(t)e^{jn\omega_o t}dt$

(c) If T is sufficiently large, then

$$T\beta_n(\alpha) = \int_{-T/2}^{T/2} R(\tau-\alpha)e^{-jn\omega_o\tau}d\tau \simeq S(n\omega_o)e^{-jn\omega_o\alpha}$$

$$E\{c_n c_m^*\} = \frac{S(n\omega_o)}{T^2}\int_{-T/2}^{T/2} e^{j(m-n)\omega_o\alpha}d\alpha = \begin{cases} S(n\omega_o)/T & m=n \\ 0 & m\neq n \end{cases}$$

Thus, for large T, the coefficients $\underset{\sim}{c}_n$ of an arbitrary WSS process are nearly orthogonal.

11-12 $$E\{\underset{\sim}{x}(t_1)\underset{\sim}{x}^*(t_2)\} = \frac{1}{4\pi^2} E\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\{\underset{\sim}{X}(u)\underset{\sim}{X}^*(v)\}e^{j(ut_1-vt_2)}dudv\right.$$

$$= \frac{1}{4\pi^2} E\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Q(u)\delta(u-v)e^{j(ut_1-vt_2)}dudv\right. = \frac{1}{4\pi^2}\int_{-\infty}^{\infty} Q(u)e^{ju(t_1-t_2)}du$$

This depends only on $\tau = t_1-t_2$:

$$R_{xx}(\tau) = \frac{1}{x_2}\int_{-\infty}^{\infty} Q(u)e^{ju\tau}du \qquad S_{xx}(\omega) = \frac{Q(\omega)}{2\pi}$$

11-13 Equations (11-79) can be written in the following form:

$$E\{\underset{\sim}{A}(u)\underset{\sim}{A}(v)\} = Q(u)\delta(u-v) = E\{\underset{\sim}{B}(u)\underset{\sim}{B}(v)\} \qquad E\{\underset{\sim}{A}(u)\underset{\sim}{B}(v)\} = 0$$

for $u \geq 0, v \geq 0$. We shall show that if the above is true and $E\{\underset{\sim}{A}(\omega)\} = E\{\underset{\sim}{B}(\omega)\} = 0$, then the process

$$\underset{\sim}{x}(t) = \frac{1}{\pi}\int_0^{\infty}\left[\underset{\sim}{A}(\omega)\cos\omega t - \underset{\sim}{B}(\omega)\sin\omega t\right] d\omega$$

is WSS.

<u>Proof</u> Clearly, $E\{\underset{\sim}{x}(t)\} = 0$ and

$$E\{\underset{\sim}{x}(t+\tau)\underset{\sim}{x}(t)\}$$

$$= \frac{1}{\pi^2}\int_0^\infty\int_0^\infty E\{\underset{\sim}{A}(u)\cos u(t+\tau) - \underset{\sim}{B}(u)\sin u(t+\tau)\}\,[\underset{\sim}{A}(v)\cos vt - B(v)\sin vt]\,du dv$$

$$= \frac{1}{\pi}\int_0^\infty\int_0^\infty Q(u)\delta(u-v)\,[\cos u(t+\tau)\cos vt + \sin u(t+\tau)\sin v(t)\,du dv\,]\,dv du$$

$$= \frac{1}{\pi^2}\int_0^\infty Q(u)\,[\cos u(t+\tau)\cos u\,t + \sin u(t+\tau)\sin ut]\,du$$

$$= \frac{1}{\pi^2}\int_0^\infty Q(u)\cos u\tau du$$

From this and (9-136) it follows that $\underset{\sim}{x}(t)$ is WSS with $S_{xx}(\omega) = Q(\omega)/\pi$.

11-14 $\quad E\{\underset{\sim}{v}(t)\} = 0 \qquad E\{\underset{\sim}{X}_T(\omega)\} = \int_{-T}^{T} f(t)e^{-j\omega t}dt$

The above integral is the transform of the product $f(t)p_T(t)$, hence (frequency convolution theorem), it equals $F(\omega)*\sin T\omega/\pi\omega$.

$$\text{Var}\ \underset{\sim}{X}_T(\omega) = E\left\{\ \left|\int_{-T}^{T}\underset{\sim}{\nu}(t)e^{-j\omega t}dt\right|^2\right\}$$

The integral is the transform of the nonstationary white noise $\underset{\sim}{\nu}(t)p_T(t)$. The autocorrelation of this process equals $q(t_1)\delta(t_1-t_2)$ where $q(t) = qp_T(t)$. Hence, [see (11-69)]

$$\text{Var}\ \underset{\sim}{X}_T(\omega) = Q(0) = \int_{-T}^{T} q dt = 2qT$$

TABLE OF CONTENTS

PROBABILITY THEORY

STOCHASTIC PROCESSES

Note: These lecture notes are revised periodically with new materials and examples added from time to time. Lectures 1 $\rightarrow$ 11 are used at Polytechnic for a first level graduate course on “Probability theory and Random Variables”. Parts of lectures 14 $\rightarrow$ 19 are used at Polytechnic for a “Stochastic Processes” course. These notes are intended for unlimited worldwide use. Any feedback may be addressed to pillai@hora.poly.edu

S. UNNIKRISHNA PILLAI

PROBABILITY THEORY

1. Basics

Probability theory deals with the study of random phenomena, which under repeated experiments yield different outcomes that have certain underlying patterns about them. The notion of an experiment assumes a set of repeatable conditions that allow any number of identical repetitions. When an experiment is performed under these conditions, certain elementary events ξ_i occur in different but *completely uncertain* ways. We can assign nonnegative number $P(\xi_i)$, as the probability of the event ξ_i in various ways:

Laplace's Classical Definition: The Probability of an event A is defined a-priori without actual experimentation as

$$P(A) = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of possible outcomes}}, \qquad (1\text{-}1)$$

provided all these outcomes are *equally likely*.

Consider a box with n white and m red balls. In this case, there are two elementary outcomes: white ball or red ball. Probability of "selecting a white ball" $= \dfrac{n}{n+m}$.

Relative Frequency Definition: The probability of an event A is defined as

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n} \tag{1-2}$$

where n_A is the number of occurrences of A and n is the total number of trials.

The axiomatic approach to probability, due to Kolmogorov, developed through a set of axioms (below) is generally recognized as superior to the above definitions, as it provides a solid foundation for complicated applications.

The totality of all ξ_i, *known a priori*, constitutes a set Ω, the set of all experimental outcomes.

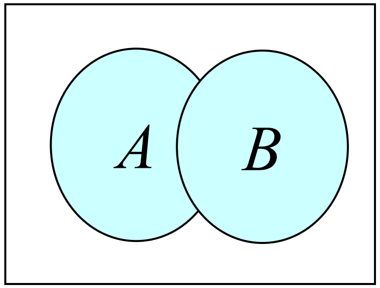
$$\Omega = \{\xi_1, \xi_2, \cdots, \xi_k, \cdots\} \tag{1-3}$$

Ω has subsets $A, B, C, \cdots$. Recall that if A is a subset of Ω, then $\xi \in A$ implies $\xi \in \Omega$. From A and B, we can generate other related subsets $A \cup B$, $A \cap B$, $\overline{A}$, $\overline{B}$, etc.

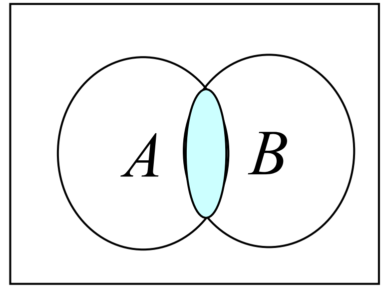
$$A \cup B = \{\xi \mid \xi \in A \text{ or } \xi \in B\}$$
$$A \cap B = \{\xi \mid \xi \in A \text{ and } \xi \in B\}$$

and

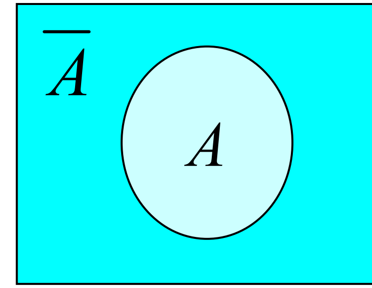
$$\overline{A} = \{\xi \mid \xi \notin A\} \tag{1-4}$$



$A \cup B$



$A \cap B$

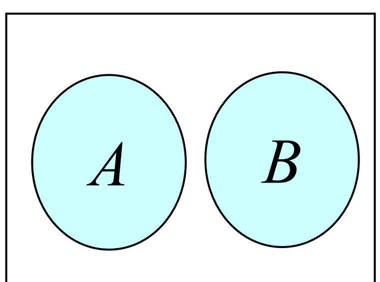


$\overline{A}$

Fig.1.1

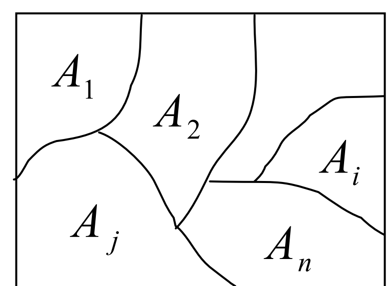
- If $A \cap B = \phi$, the empty set, then A and B are said to be mutually exclusive (M.E).
- A partition of Ω is a collection of mutually exclusive subsets of Ω such that their union is Ω.

$$A_i \cap A_j = \phi, \quad \text{and} \quad \bigcup_{i=1} A_i = \Omega. \tag{1-5}$$



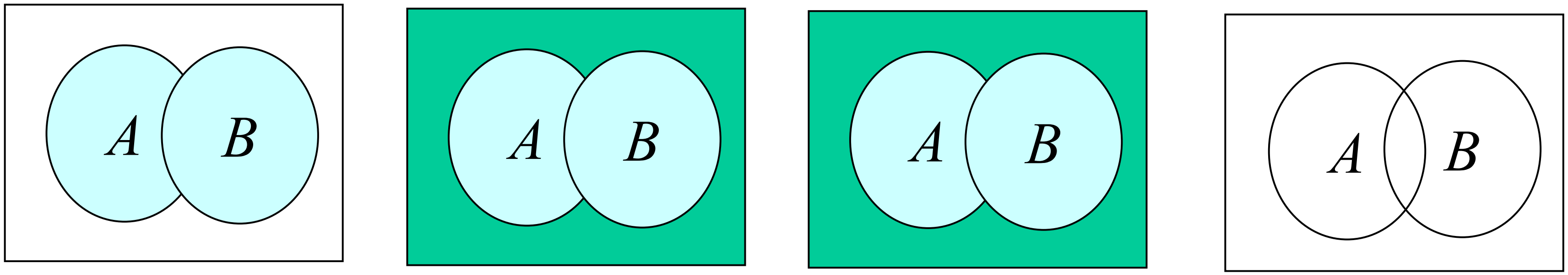
$A \cap B = \phi$

Fig. 1.2



De-Morgan's Laws:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}; \qquad \overline{A \cap B} = \overline{A} \cup \overline{B} \tag{1-6}$$



$A \cup B$ $\quad$ $\overline{A \cup B}$ $\quad$ $\overline{A} \cap \overline{B}$

Fig.1.3

- Often it is meaningful to talk about at least some of the subsets of Ω as events, for which we must have mechanism to compute their probabilities.

Example 1.1: Consider the experiment where two coins are simultaneously tossed. The various elementary events are

$$\xi_1 = (H,H),\ \ \xi_2 = (H,T),\ \ \xi_3 = (T,H),\ \ \xi_4 = (T,T)$$

and

$$\Omega = \{\xi_1, \xi_2, \xi_3, \xi_4\}.$$

The subset $A = \{\xi_1, \xi_2, \xi_3\}$ is the same as "Head has occurred at least once" and qualifies as an event.

Suppose two subsets A and B are both events, then consider

"Does an outcome belong to A or $B = A \cup B$"

"Does an outcome belong to A and $B = A \cap B$"

"Does an outcome fall outside A"?

Thus the sets $A \cup B$, $A \cap B$, $\overline{A}$, $\overline{B}$, etc., also qualify as events. We shall formalize this using the notion of a Field.

•**Field**: A collection of subsets of a nonempty set Ω forms a field F if

$$
\begin{aligned}
&\text{(i)} \quad \Omega \in F \\
&\text{(ii)} \quad \text{If } A \in F, \quad \text{then } \overline{A} \in F \\
&\text{(iii)} \quad \text{If } A \in F \quad \text{and} \quad B \in F, \quad \text{then } A \cup B \in F.
\end{aligned}
\tag{1-7}
$$

Using (i) - (iii), it is easy to show that $A \cap B$, $\overline{A} \cap B$, etc., also belong to F. For example, from (ii) we have $\overline{A} \in F$, $\overline{B} \in F$, and using (iii) this gives $\overline{A} \cup \overline{B} \in F$; applying (ii) again we get $\overline{\overline{A} \cup \overline{B}} = A \cap B \in F$, where we have used De Morgan's theorem in (1-6).

Thus if $A \in F$, $B \in F$, then

$$F = \left\{ \Omega, A, B, \overline{A}, \overline{B}, A \cup B, A \cap B, \overline{A} \cup B, \cdots \right\}. \quad (1\text{-}8)$$

From here on wards, we shall reserve the term 'event' only to members of F.

Assuming that the probability $p_i = P(\xi_i)$ of elementary outcomes ξ_i of Ω are apriori defined, how does one assign probabilities to more 'complicated' events such as A, B, AB, etc., above?

The three axioms of probability defined below can be used to achieve that goal.

Axioms of Probability

For any event A, we assign a number $P(A)$, called the probability of the event A. This number satisfies the following three conditions that act the axioms of probability.

$$
\begin{array}{lll}
\text{(i)} & P(A) \geq 0 & \text{(Probability is a nonnegative number)} \\
\text{(ii)} & P(\Omega) = 1 & \text{(Probability of the whole set is unity)} \\
\text{(iii)} & \text{If } A \cap B = \phi, \quad \text{then} & P(A \cup B) = P(A) + P(B).
\end{array}
\tag{1-9}
$$

(Note that (iii) states that if A and B are mutually exclusive (M.E.) events, the probability of their union is the sum of their probabilities.)

The following conclusions follow from these axioms:

a. Since $A \cup \overline{A} = \Omega$, we have using (ii)

$$\mathrm{P}(A \cup \overline{A}) = P(\Omega) = 1.$$

But $A \cap \overline{A} \in \phi$, and using (iii),

$$\mathrm{P}(A \cup \overline{A}) = P(A) + \mathrm{P}(\overline{A}) = 1 \quad \text{or} \quad \mathrm{P}(\overline{A}) = 1 - P(A). \qquad \text{(1-10)}$$

b. Similarly, for any *A*, $A \cap \{\phi\} = \{\phi\}$.

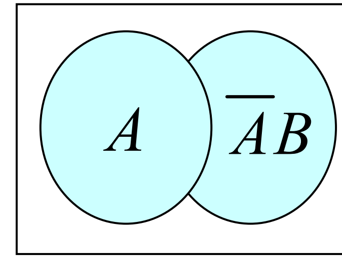
Hence it follows that $P\left(A \cup \{\phi\}\right) = P(A) + P(\phi)$.

But $A \cup \{\phi\} = A$, and thus $P\{\phi\} = 0$. (1-11)

c. Suppose *A* and *B* are *not* mutually exclusive (M.E.)?

How does one compute $P(A \cup B) = ?$

To compute the above probability, we should re-express $A \cup B$ in terms of M.E. sets so that we can make use of the probability axioms. From Fig.1.4 we have



$A \cup B$

Fig.1.4

$$A \cup B = A \cup \overline{A}B, \qquad (1\text{-}12)$$

where A and $\overline{A}B$ are clearly M.E. events. Thus using axiom (1-9-iii)

$$P(A \cup B) = P(A \cup \overline{A}B) = P(A) + P(\overline{A}B). \quad (1\text{-}13)$$

To compute $P(\overline{A}B)$, we can express B as

$$B = B \cap \Omega = B \cap (A \cup \overline{A})$$

$$= (B \cap A) \cup (B \cap \overline{A}) = BA \cup B\overline{A} \qquad (1\text{-}14)$$

Thus

$$P(B) = P(BA) + P(B\overline{A}), \qquad (1\text{-}15)$$

since $BA = AB$ and $B\overline{A} = \overline{A}B$ are M.E. events.

From (1-15),

$$P(\overline{A}B) = P(B) - P(AB) \tag{1-16}$$

and using (1-16) in (1-13)

$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{1-17}$$

- Question: Suppose every member of a denumerably infinite collection A_i of pair wise disjoint sets is an event, then what can we say about their union

$$A = \bigcup_{i=1}^{\infty} A_i \,? \tag{1-18}$$

i.e., suppose all $A_i \in F$, what about A? Does it belong to F? (1-19)

Further, if A also belongs to F, what about $P(A)$? (1-20)

The above questions involving infinite sets can only be settled using our intuitive experience from plausible experiments. For example, in a coin tossing experiment, where the same coin is tossed indefinitely, define

$$A = \text{“head eventually appears”}. \tag{1-21}$$

Is A an event? Our intuitive experience surely tells us that A is an event. Let

$$\begin{aligned} A_n &= \{\text{head appears for the 1st time on the } n\text{th toss}\} \\ &= \{\underbrace{t, t, t, \cdots, t}_{n-1}, h\} \end{aligned} \tag{1-22}$$

Clearly $A_i \cap A_j = \phi$. Moreover the above A is

$$A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_i \cup \cdots. \tag{1-23}$$

We cannot use probability axiom (1-9-iii) to compute $P(A)$, since the axiom only deals with two (or a finite number) of M.E. events.

To settle both questions above (1-19)-(1-20), extension of these notions must be done based on our intuition as new axioms.

σ-Field (Definition):

A field F is a σ-field if in addition to the three conditions in (1-7), we have the following:

For every sequence $A_i, i = 1 \to \infty,$ of pair wise disjoint events belonging to F, their union also belongs to F, i.e.,

$$A = \bigcup_{i=1}^{\infty} A_i \in F. \tag{1-24}$$

In view of (1-24), we can add yet another axiom to the set of probability axioms in (1-9).

(iv) If A_i are pair wise mutually exclusive, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{1-25}$$

Returning back to the coin tossing experiment, from experience we know that if we keep tossing a coin, eventually, a head must show up, i.e.,

$$P(A) = 1. \tag{1-26}$$

But $A = \bigcup_{n=1}^{\infty} A_n$, and using the fourth probability axiom in (1-25),

$$P(A) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{1-27}$$

From (1-22), for a fair coin since only one in 2^n outcomes is in favor of A_n, we have

$$P(A_n) = \frac{1}{2^n} \quad \text{and} \quad \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1, \tag{1-28}$$

which agrees with (1-26), thus justifying the

'reasonableness' of the fourth axiom in (1-25).

In summary, the triplet (Ω, F, P) composed of a nonempty set Ω of elementary events, a σ-field F of subsets of Ω, and a probability measure P on the sets in F subject the four axioms ((1-9) and (1-25)) form a probability model.

The probability of more complicated events must follow from this framework by deduction.

Conditional Probability and Independence

In N independent trials, suppose N_A, N_B, N_{AB} denote the number of times events A, B and AB occur respectively. According to the frequency interpretation of probability, for large N

$$P(A) \approx \frac{N_A}{N}, \quad P(B) \approx \frac{N_B}{N}, \quad P(AB) \approx \frac{N_{AB}}{N}. \qquad \text{(1-29)}$$

Among the N_A occurrences of A, only N_{AB} of them are also found among the N_B occurrences of B. Thus the ratio

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)} \qquad \text{(1-30)}$$

is a measure of "the event A given that B has already occurred". We denote this conditional probability by

$P(A|B) =$ Probability of "the event A given that B has occurred".

We define

$$P(A \mid B) = \frac{P(AB)}{P(B)}, \tag{1-31}$$

provided $P(B) \neq 0.$ As we show below, the above definition satisfies all probability axioms discussed earlier.

We have

(i) $$P(A \mid B) = \frac{P(AB) \geq 0}{P(B) > 0} \geq 0, \qquad (1\text{-}32)$$

(ii) $$P(\Omega \mid B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1, \text{ since } \Omega B = B. \qquad (1\text{-}33)$$

(iii) Suppose $A \cap C = 0$. Then

$$P(A \cup C \mid B) = \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P(AB \cup CB)}{P(B)}. \qquad (1\text{-}34)$$

But $AB \cap AC = \phi$, hence $P(AB \cup CB) = P(AB) + P(CB)$.

$$P(A \cup C \mid B) = \frac{P(AB)}{P(B)} + \frac{P(CB)}{P(B)} = P(A \mid B) + P(C \mid B), \qquad (1\text{-}35)$$

satisfying all probability axioms in (1-9). Thus (1-31) defines a legitimate probability measure.

Properties of Conditional Probability:

a. If $B \subset A$, $AB = B$, and

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1 \qquad (1\text{-}36)$$

since if $B \subset A$, then occurrence of B implies automatic occurrence of the event A. As an example, but

$$A = \{\text{outcome is even}\}, \;\; B = \{\text{outcome is 2}\},$$

in a dice tossing experiment. Then $B \subset A$, and $P(A \mid B) = 1$.

b. If $A \subset B$, $AB = A$, and

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} > P(A). \qquad (1\text{-}37)$$

(In a dice experiment, $A=\{\text{outcome is 2}\}$, $B=\{\text{outcome is even}\}$, so that $A \subset B$. The statement that B has occurred (outcome is even) makes the odds for "outcome is 2" greater than without that information).

c. We can use the conditional probability to express the probability of a complicated event in terms of "simpler" related events.

Let $A_1, A_2, \cdots, A_n$ are pair wise disjoint and their union is Ω.
Thus $A_i A_j = \phi$, and

$$\bigcup_{i=1}^{n} A_i = \Omega . \tag{1-38}$$

Thus

$$B = B(A_1 \cup A_2 \cup \cdots \cup A_n) = BA_1 \cup BA_2 \cup \cdots \cup BA_n . \tag{1-39}$$

But $A_i \cap A_j = \phi \Rightarrow BA_i \cap BA_j = \phi$, so that from (1-39)

$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i). \qquad (1\text{-}40)$$

With the notion of conditional probability, next we introduce the notion of "independence" of events.

Independence: *A* and *B* are said to be independent events, if

$$P(AB) = P(A) \cdot P(B). \qquad (1\text{-}41)$$

Notice that the above definition is a probabilistic statement, *not* a set theoretic notion such as mutually exclusiveness.

Suppose A and B are independent, then

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A). \qquad (1\text{-}42)$$

Thus if A and B are independent, the event that B has occurred does not shed any more light into the event A. It makes no difference to A whether B has occurred or not. An example will clarify the situation:

Example 1.2: A box contains 6 white and 4 black balls. Remove two balls at random without replacement. What is the probability that the first one is white and the second one is black?

Let $W_1 =$ "first ball removed is white"

$B_2 =$ "second ball removed is black"

We need $P(W_1 \cap B_2) = ?$ We have $W_1 \cap B_2 = W_1 B_2 = B_2 W_1.$ Using the conditional probability rule,

$$P(W_1 B_2) = P(B_2 W_1) = P(B_2 \mid W_1) P(W_1). \qquad (1\text{-}43)$$

But

$$P(W_1) = \frac{6}{6+4} = \frac{6}{10} = \frac{3}{5},$$

and

$$P(B_2 \mid W_1) = \frac{4}{5+4} = \frac{4}{9},$$

and hence

$$P(W_1 B_2) = \frac{5}{9} \cdot \frac{4}{9} = \frac{20}{81} \approx 0.25\,.$$

Are the events W_1 and B_2 independent? Our common sense says No. To verify this we need to compute $P(B_2)$. Of course the fate of the second ball very much depends on that of the first ball. The first ball has two options: W_1 = "first ball is white" or B_1= "first ball is black". Note that $W_1 \cap B_1 = \phi$, and $W_1 \cup B_1 = \Omega$. Hence W_1 together with B_1 form a partition. Thus (see (1-38)-(1-40))

$$P(B_2) = P(B_2 \mid W_1)P(W_1) + P(B_2 \mid R_1)P(B_1)$$
$$= \frac{4}{5+4} \cdot \frac{3}{5} + \frac{3}{6+3} \cdot \frac{4}{10} = \frac{4}{9} \cdot \frac{3}{5} + \frac{1}{3} \cdot \frac{2}{5} = \frac{4+2}{15} = \frac{2}{5},$$

and

$$P(B_2)P(W_1) = \frac{2}{5} \cdot \frac{3}{5} \neq P(B_2 W_1) = \frac{20}{81}.$$

As expected, the events W_1 and B_2 are dependent.

From (1-31),

$$P(AB) = P(A \mid B)P(B). \qquad (1\text{-}44)$$

Similarly, from (1-31)

$$P(B \mid A) = \frac{P(BA)}{P(A)} = \frac{P(AB)}{P(A)},$$

or

$$P(AB) = P(B \mid A)P(A). \qquad (1\text{-}45)$$

From (1-44)-(1-45), we get

$$P(A \mid B)P(B) = P(B \mid A)P(A).$$

or

$$P(A \mid B) = \frac{P(B \mid A)}{P(B)} \cdot P(A) \qquad (1\text{-}46)$$

Equation (1-46) is known as Bayes' theorem.

Although simple enough, Bayes' theorem has an interesting interpretation: $P(A)$ represents the a-priori probability of the event A. Suppose B has occurred, and assume that A and B are not independent. How can this new information be used to update our knowledge about A? Bayes' rule in (1-46) take into account the new information ("B has occurred") and gives out the a-posteriori probability of A given B.

We can also view the event B as new knowledge obtained from a fresh experiment. We know something about A as $P(A)$. The new information is available in terms of B. The new information should be used to improve our knowledge/understanding of A. Bayes' theorem gives the exact mechanism for incorporating such new information.

A more general version of Bayes' theorem involves partition of Ω. From (1-46)

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}, \qquad (1\text{-}47)$$

where we have made use of (1-40). In (1-47), $A_i,\ i = 1 \to n$, represent a set of mutually exclusive events with associated a-priori probabilities $P(A_i),\ i = 1 \to n.$ With the new information "*B* has occurred", the information about A_i can be updated by the n conditional probabilities $P(B \mid A_i),\ i = 1 \to n,$ using (1 - 47).

Example 1.3: Two boxes B_1 and B_2 contain 100 and 200 light bulbs respectively. The first box (B_1) has 15 defective bulbs and the second 5. Suppose a box is selected at random and one bulb is picked out.

(a) What is the probability that it is defective?

Solution: Note that box B_1 has 85 good and 15 defective bulbs. Similarly box B_2 has 195 good and 5 defective bulbs. Let D = "Defective bulb is picked out".

Then

$$P(D \mid B_1) = \frac{15}{100} = 0.15, \quad P(D \mid B_2) = \frac{5}{200} = 0.025 .$$

Since a box is selected at random, they are equally likely.

$$P(B_1) = P(B_2) = \frac{1}{2}.$$

Thus B_1 and B_2 form a partition as in (1-39), and using

(1-40) we obtain

$$P(D) = P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2)$$
$$= 0.15 \times \frac{1}{2} + 0.025 \times \frac{1}{2} = 0.0875 \; .$$

Thus, there is about 9% probability that a bulb picked at random is defective.

(b) Suppose we test the bulb and it is found to be defective. What is the probability that it came from box 1? $P(B_1 \mid D) = ?$

$$P(B_1 \mid D) = \frac{P(D \mid B_1)P(B_1)}{P(D)} = \frac{0.15 \times 1/2}{0.0875} = 0.8571 \,. \quad \text{(1-48)}$$

Notice that initially $P(B_1) = 0.5;$ then we picked out a box at random and tested a bulb that turned out to be defective. Can this information shed some light about the fact that we might have picked up box 1?

From (1-48), $P(B_1 \mid D) = 0.857 > 0.5,$ and indeed it is more likely at this point that we must have chosen box 1 in favor of box 2. (Recall box1 has six times more defective bulbs compared to box2).

2. Independence and Bernoulli Trials

Independence: Events A and B are independent if

$$P(AB) = P(A)P(B). \tag{2-1}$$

- It is easy to show that A, B independent implies $\overline{A}, B$; $A, \overline{B}$; $\overline{A}, \overline{B}$ are all independent pairs. For example, $B = (A \cup \overline{A})B = AB \cup \overline{A}B$ and $AB \cap \overline{A}B = \phi$, so that $P(B) = P(AB \cup \overline{A}B) = P(AB) + P(\overline{A}B) = P(A)P(B) + P(\overline{A}B)$ or

$$P(\overline{A}B) = P(B) - P(A)P(B) = (1 - P(A))P(B) = P(\overline{A})P(B),$$

i.e., $\overline{A}$ and B are independent events.

- If $P(A) = 0$, then since the event $AB \subset A$ always, we have

$$P(AB) \leq P(A) = 0 \Rightarrow P(AB) = 0,$$

and (2-1) is always satisfied. Thus the event of zero probability is independent of every other event!

- Independent events obviously cannot be mutually exclusive, since $P(A) > 0,\ P(B) > 0$ and A, B independent implies $P(AB) > 0.$ Thus if A and B are independent, the event AB cannot be the null set.

- More generally, a family of events $\{A_i\}$ are said to be independent, if for every finite sub collection $A_{i_1}, A_{i_2}, \cdots, A_{i_n},$ we have

$$P\left(\bigcap_{k=1}^{n} A_{i_k}\right) = \prod_{k=1}^{n} P(A_{i_k}). \tag{2-2}$$

• Let

$$A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n, \tag{2-3}$$

a union of n independent events. Then by De-Morgan's law

$$\overline{A} = \overline{A}_1 \overline{A}_2 \cdots \overline{A}_n \tag{2-4}$$

and using their independence

$$P(\overline{A}) = P(\overline{A}_1 \overline{A}_2 \cdots \overline{A}_n) = \prod_{i=1}^{n} P(\overline{A}_i) = \prod_{i=1}^{n} (1 - P(A_i)). \tag{2-5}$$

Thus for any A as in (2-3)

$$P(A) = 1 - P(\overline{A}) = 1 - \prod_{i=1}^{n} (1 - P(A_i)), \tag{2-6}$$

a useful result.

Example 2.1: Three switches connected in parallel operate independently. Each switch remains closed with probability p. (a) Find the probability of receiving an input signal at the output. (b) Find the probability that switch S_1 is open given that an input signal is received at the output.

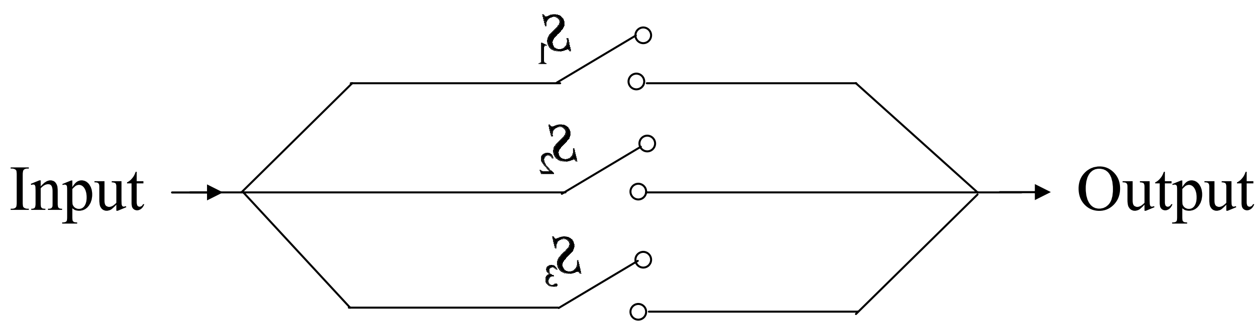


Fig.2.1

Solution: a. Let A_i = "Switch S_i is closed". Then $P(A_i) = p$, $i = 1 \rightarrow 3$. Since switches operate independently, we have

$$P(A_i A_j) = P(A_i)P(A_j); \;\; P(A_1 A_2 A_3) = P(A_1)P(A_2)P(A_3).$$

Let R = "input signal is received at the output". For the event R to occur either switch 1 or switch 2 or switch 3 must remain closed, i.e.,

$$R = A_1 \cup A_2 \cup A_3. \tag{2-7}$$

Using (2-3) - (2-6),

$$P(R) = P(A_1 \cup A_2 \cup A_3) = 1 - (1-p)^3 = 3p - 3p^2 + p^3. \tag{2-8}$$

We can also derive (2-8) in a different manner. Since any event and its compliment form a trivial partition, we can always write

$$P(R) = P(R \mid A_1)P(A_1) + P(R \mid \overline{A_1})P(\overline{A_1}). \tag{2-9}$$

But $P(R \mid A_1) = 1,$ and $P(R \mid \overline{A_1}) = P(A_2 \cup A_3) = 2p - p^2$ and using these in (2-9) we obtain

$$P(R) = p + (2p - p^2)(1-p) = 3p - 3p^2 + p^3, \tag{2-10}$$

which agrees with (2-8).

Note that the events A_1, A_2, A_3 do not form a partition, since they are not mutually exclusive. Obviously any two or all three switches can be closed (or open) simultaneously. Moreover, $P(A_1)+P(A_2)+P(A_3) \neq 1.$

b. We need $P(\overline{A}_1 \mid R).$ From Bayes' theorem

$$P(\overline{A}_1 \mid R) = \frac{P(R \mid \overline{A}_1)P(\overline{A}_1)}{P(R)} = \frac{(2p-p^2)(1-p)}{3p-3p^2+p^3} = \frac{2-2p+p^2}{3p-3p^2+p^3}. \quad (2\text{-}11)$$

Because of the symmetry of the switches, we also have

$$P(\overline{A}_1 \mid R) = P(\overline{A}_2 \mid R) = P(\overline{A}_3 \mid R).$$

Repeated Trials

Consider two independent experiments with associated probability models (Ω_1, F_1, P_1) and (Ω_2, F_2, P_2). Let $\xi \in \Omega_1$, $\eta \in \Omega_2$ represent elementary events. A joint performance of the two experiments produces an elementary events $\omega = (\xi, \eta)$. How to characterize an appropriate probability to this "combined event" ? Towards this, consider the Cartesian product space $\Omega = \Omega_1 \times \Omega_2$ generated from Ω_1 and Ω_2 such that if $\xi \in \Omega_1$ and $\eta \in \Omega_2$, then every ω in Ω is an ordered pair of the form $\omega = (\xi, \eta)$. To arrive at a probability model we need to define the combined trio (Ω, F, P).

Suppose $A \in F_1$ and $B \in F_2$. Then $A \times B$ is the set of all pairs (ξ, η), where $\xi \in A$ and $\eta \in B$. Any such subset of Ω appears to be a legitimate event for the combined experiment. Let F denote the field composed of all such subsets $A \times B$ together with their unions and compliments. In this combined experiment, the probabilities of the events $A \times \Omega_2$ and $\Omega_1 \times B$ are such that

$$P(A \times \Omega_2) = P_1(A), \quad P(\Omega_1 \times B) = P_2(B). \qquad (2\text{-}12)$$

Moreover, the events $A \times \Omega_2$ and $\Omega_1 \times B$ are independent for any $A \in F_1$ and $B \in F_2$. Since

$$(A \times \Omega_2) \cap (\Omega_1 \times B) = A \times B, \qquad (2\text{-}13)$$

we conclude using (2-12) that

$$P(A \times B) = P(A \times \Omega_2) \cdot P(\Omega_1 \times B) = P_1(A) P_2(B) \tag{2-14}$$

for all $A \in F_1$ and $B \in F_2$. The assignment in (2-14) extends to a unique probability measure $P(\equiv P_1 \times P_2)$ on the sets in F and defines the combined trio (Ω, F, P).

Generalization: Given n experiments $\Omega_1, \Omega_2, \cdots, \Omega_n,$ and their associated F_i and $P_i,$ $i = 1 \to n$, let

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n \tag{2-15}$$

represent their Cartesian product whose elementary events are the ordered n-tuples $\xi_1, \xi_2, \cdots, \xi_n,$ where $\xi_i \in \Omega_i.$ Events in this combined space are of the form

$$A_1 \times A_2 \times \cdots \times A_n \tag{2-16}$$

where $A_i \in F_i,$ and their unions an intersections.

If all these n experiments are independent, and $P_i(A_i)$ is the probability of the event A_i in F_i then as before

$$P(A_1 \times A_2 \times \cdots \times A_n) = P_1(A_1)P_2(A_2)\cdots P(A_n). \tag{2-17}$$

Example 2.2: An event A has probability p of occurring in a single trial. Find the probability that A occurs exactly k times, $k \leq n$ in n trials.

Solution: Let (Ω, F, P) be the probability model for a single trial. The outcome of n experiments is an n-tuple

$$\omega = \left\{ \xi_1, \xi_2, \cdots, \xi_n \right\} \in \Omega_0, \tag{2-18}$$

where every $\xi_i \in \Omega$ and $\Omega_0 = \Omega \times \Omega \times \cdots \times \Omega$ as in (2-15). The event A occurs at trial # i, if $\xi_i \in A$. Suppose A occurs exactly k times in ω.

Then k of the ξ_i belong to A, say $\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}$, and the remaining $n-k$ are contained in its compliment in $\overline{A}$. Using (2-17), the probability of occurrence of such an ω is given by

$$P_0(\omega) = P(\{\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}, \cdots, \xi_{i_n}\}) = P(\{\xi_{i_1}\})P(\{\xi_{i_2}\})\cdots P(\{\xi_{i_k}\})\cdots P(\{\xi_{i_n}\})$$

$$= \underbrace{P(A)P(A)\cdots P(A)}_{k}\ \underbrace{P(\overline{A})P(\overline{A})\cdots P(\overline{A})}_{n-k} = p^k q^{n-k}. \qquad (2\text{-}19)$$

However the k occurrences of A can occur in any particular location inside ω. Let $\omega_1, \omega_2, \cdots, \omega_N$ represent all such events in which A occurs exactly k times. Then

$$\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"} = \omega_1 \cup \omega_2 \cup \cdots \cup \omega_N. \qquad (2\text{-}20)$$

But, all these ω_is are mutually exclusive, and equiprobable.

Thus

$$P(\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"})$$

$$= \sum_{i=1}^{N} P_0(\omega_i) = N P_0(\omega) = N p^k q^{n-k}, \tag{2-21}$$

where we have used (2-19). Recall that, starting with n possible choices, the first object can be chosen n different ways, and for every such choice the second one in $(n-1)$ ways, … and the kth one $(n-k+1)$ ways, and this gives the total choices for k objects out of n to be $n(n-1)\cdots(n-k+1)$. But, this includes the $k!$ choices among the k objects that are indistinguishable for identical objects. As a result

$$N = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!\,k!} \stackrel{\Delta}{=} \binom{n}{k} \tag{2-22}$$

represents the number of combinations, or choices of n identical objects taken k at a time. Using (2-22) in (2-21), we get

$$P_n(k) = P(\text{"} A \text{ occurs exactly } k \text{ times in } n \text{ trials"})$$

$$= \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\cdots,n, \tag{2-23}$$

a formula, due to Bernoulli.

Independent repeated experiments of this nature, where the outcome is either a "success" $(= A)$ or a "failure" $(= \overline{A})$ are characterized as Bernoulli trials, and the probability of k successes in n trials is given by (2-23), where p represents the probability of "success" in any one trial.

Example 2.3: Toss a coin n times. Obtain the probability of getting k heads in n trials ?

Solution: We may identify "head" with "success" (A) and let $p = P(H)$. In that case (2-23) gives the desired probability.

Example 2.4: Consider rolling a fair die eight times. Find the probability that either 3 or 4 shows up five times ?

Solution: In this case we can identify

$$\text{"success"} = A = \{ \text{ either } 3 \text{ or } 4 \ \} = \{f_3\} \cup \{f_4\}.$$

Thus

$$P(A) = P(f_3) + P(f_4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3},$$

and the desired probability is given by (2-23) with $n = 8,\ k = 5$ and $p = 1/3$. Notice that this is similar to a "biased coin" problem.

Bernoulli trial: consists of repeated independent and identical experiments each of which has only two outcomes A or $\overline{A}$ with $P(A) = p,$ and $P(\overline{A}) = q$. The probability of exactly k occurrences of A in n such trials is given by (2-23).

Let

$$X_k = \text{"exactly } k \text{ occurrences in } n \text{ trials"}. \tag{2-24}$$

Since the number of occurrences of A in n trials must be an integer $k = 0,1,2,\cdots,n,$ either X_0 or X_1 or X_2 or $\cdots$ or X_n must occur in such an experiment. Thus

$$P(X_0 \cup X_1 \cup \cdots \cup X_n) = 1. \tag{2-25}$$

But $X_i,\ X_j$ are mutually exclusive. Thus

$$P(X_0 \cup X_1 \cup \cdots \cup X_n) = \sum_{k=0}^{n} P(X_k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}. \quad (2\text{-}26)$$

From the relation

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \quad (2\text{-}27)$$

(2-26) equals $(p+q)^n = 1,$ and it agrees with (2-25).

For a given *n* and *p* what is the most likely value of *k* ? From Fig.2.2, the most probable value of *k* is that number which maximizes $P_n(k)$ in (2-23). To obtain this value, consider the ratio

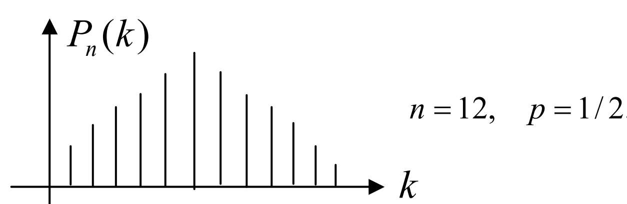


Fig. 2.2

$$\frac{P_n(k-1)}{P_n(k)} = \frac{n!\,p^{k-1}q^{n-k+1}}{(n-k+1)!\,(k-1)!}\,\frac{(n-k)!\,k!}{n!\,p^k q^{n-k}} = \frac{k}{n-k+1}\,\frac{q}{p}. \tag{2-28}$$

Thus $P_n(k) \geq P_n(k-1)$, if $k(1-p) \leq (n-k+1)p$ or $k \leq (n+1)p$. Thus $P_n(k)$ as a function of *k* increases until

$$k = (n+1)p \tag{2-29}$$

if it is an integer, or the largest integer $k_{\max}$ less than $(n+1)p$, and (2-29) represents the most likely number of successes (or heads) in *n* trials.

Example 2. 5: In a Bernoulli experiment with *n* trials, find the probability that the number of occurrences of *A* is between k_1 and k_2.

Solution: With $X_i, i = 0,1,2,\cdots,n,$ as defined in (2-24), clearly they are mutually exclusive events. Thus

$$P(\text{"Occurrences of } A \text{ is between } k_1 \text{ and } k_2\text{"})$$

$$= P(X_{k_1} \cup X_{k_1+1} \cup \cdots \cup X_{k_2}) = \sum_{k=k_1}^{k_2} P(X_k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}. \quad \text{(2-30)}$$

Example 2. 6: Suppose 5,000 components are ordered. The probability that a part is defective equals 0.1. What is the probability that the total number of defective parts does not exceed 400 ?

Solution: Let

$$Y_k = \text{"}k \text{ parts are defective among 5,000 components"}.$$

Using (2-30), the desired probability is given by

$$P(Y_0 \cup Y_1 \cup \cdots \cup Y_{400}) = \sum_{k=0}^{400} P(Y_k)$$

$$= \sum_{k=0}^{400} \binom{5000}{k} (0.1)^k (0.9)^{5000-k}. \tag{2-31}$$

Equation (2-31) has too many terms to compute. Clearly, we need a technique to compute the above term in a more efficient manner.

From (2-29), $k_{\max}$ the most likely number of successes in n trials, satisfy

$$(n+1)p - 1 \le k_{\max} \le (n+1)p \tag{2-32}$$

or

$$p - \frac{q}{n} \le \frac{k_{\max}}{n} \le p + \frac{p}{n}, \tag{2-33}$$

so that

$$\lim_{n \to \infty} \frac{k_m}{n} = p. \tag{2-34}$$

From (2-34), as $n \to \infty$, the ratio of the most probable number of successes (A) to the total number of trials in a Bernoulli experiment tends to p, the probability of occurrence of A in a single trial. Notice that (2-34) connects the results of an actual experiment (k_m/n) to the axiomatic definition of p. In this context, it is possible to obtain a more general result as follows:

Bernoulli's theorem: Let A denote an event whose probability of occurrence in a single trial is p. If k denotes the number of occurrences of A in n independent trials, then

$$P\left(\left\{ \left| \frac{k}{n} - p \right| > \varepsilon \right\} \right) < \frac{pq}{n\varepsilon^2}. \tag{2-35}$$

Equation (2-35) states that the frequency definition of probability of an event $\frac{k}{n}$ and its axiomatic definition (p) can be made compatible to any degree of accuracy.

Proof: To prove Bernoulli's theorem, we need two identities. Note that with $P_n(k)$ as in (2-23), direct computation gives

$$\sum_{k=0}^{n} k\, P_n(k) = \sum_{k=1}^{n-1} k \frac{n!}{(n-k)!\,k!} p^k q^{n-k} = \sum_{k=1}^{n} \frac{n!}{(n-k)!\,(k-1)!} p^k q^{n-k}$$

$$= \sum_{i=0}^{n-1} \frac{n!}{(n-i-1)!\,i!} p^{i+1} q^{n-i-1} = np \sum_{i=0}^{n-1} \frac{(n-1)!}{(n-1-i)!\,i!} p^i q^{n-1-i}$$

$$= np(p+q)^{n-1} = np. \qquad (2\text{-}36)$$

Proceeding in a similar manner, it can be shown that

$$\sum_{k=0}^{n} k^2 P_n(k) = \sum_{k=1}^{n} k \frac{n!}{(n-k)!\,(k-1)!} p^k q^{n-k} = \sum_{k=2}^{n} \frac{n!}{(n-k)!\,(k-2)!} p^k q^{n-k}$$

$$+ \sum_{k=1}^{n} \frac{n!}{(n-k)!\,(k-1)!} p^k q^{n-k} = n^2 p^2 + npq. \qquad (2\text{-}37)$$

Returning to (2-35), note that

$$\left|\frac{k}{n}-p\right|>\varepsilon \quad \text{is equivalent to} \quad (k-np)^2>n^2\varepsilon^2, \tag{2-38}$$

which in turn is equivalent to

$$\sum_{k=0}^{n}(k-np)^2 P_n(k)>\sum_{k=0}^{n} n^2\varepsilon^2 P_n(k)=n^2\varepsilon^2. \tag{2-39}$$

Using (2-36)-(2-37), the left side of (2-39) can be expanded to give

$$\sum_{k=0}^{n}(k-np)^2 P_n(k)=\sum_{k=0}^{n}k^2 P_n(k)-2np\sum_{k=0}^{n}kP_n(k)+n^2p^2$$

$$=n^2p^2+npq-2np\cdot np+n^2p^2=npq. \tag{2-40}$$

Alternatively, the left side of (2-39) can be expressed as

$$\sum_{k=0}^{n}(k-np)^2 P_n(k)=\sum_{|k-np|\le n\varepsilon}(k-np)^2 P_n(k)+\sum_{|k-np|>n\varepsilon}(k-np)^2 P_n(k)$$

$$\ge\sum_{|k-np|>n\varepsilon}(k-np)^2 P_n(k)>n^2\varepsilon^2\sum_{|k-np|>n\varepsilon}P_n(k)$$

$$=n^2\varepsilon^2 P\{|k-np|>n\varepsilon\}. \tag{2-41}$$

Using (2-40) in (2-41), we get the desired result

$$P\left(\left\{\left|\frac{k}{n}-p\right|>\varepsilon\right\}\right)<\frac{pq}{n\varepsilon^2}. \qquad (2\text{-}42)$$

Note that for a given $\varepsilon > 0$, $pq/n\varepsilon^2$ can be made arbitrarily small by letting *n* become large. Thus for very large *n*, we can make the fractional occurrence (relative frequency) $\frac{k}{n}$ of the event *A* as close to the actual probability *p* of the event *A* in a single trial. Thus the theorem states that the probability of event *A* from the axiomatic framework can be computed from the relative frequency definition quite accurately, provided the number of experiments are large enough. Since $k_{\max}$ is the most likely value of k in n trials, from the above discussion, as $n \to \infty$, the plots of $P_n(k)$ tends to concentrate more and more around $k_{\max}$ in (2-32).

Next we present an example that illustrates the usefulness of "simple textbook examples" to practical problems of interest:

Example 2.7 : **Day-trading strategy :** A box contains n *randomly* numbered balls (not 1 through n but arbitrary numbers including numbers greater than n). Suppose a fraction of those balls – say $m = np$; $p < 1$ – are initially drawn one by one with replacement while noting the numbers on those balls. The drawing is allowed to continue *until* a ball is drawn with a number larger than the first m numbers. Determine the fraction p to be initially drawn, so as to maximize the probability of drawing the largest among the n numbers using this strategy.

Solution: Let "$X_k = (k+1)^{st}$ drawn ball has the largest number among all n balls, and the largest among the first k balls is in the group of first m balls, $k > m$." (2.43)

Note that X_k is of the form $A \cap B$,

where

A = "largest among the first k balls is in the group of first m balls drawn"

and

B = "$(k+1)^{st}$ ball has the largest number among all n balls".

Notice that A and B are independent events, and hence

$$P(X_k) = P(A)P(B) = \frac{1}{n}\frac{m}{k} = \frac{1}{n}\frac{np}{k} = \frac{p}{k}. \tag{2-44}$$

Where $m = np$ represents the fraction of balls to be initially drawn. This gives

P ("selected ball has the largest number among all balls")

$$= \sum_{k=m}^{n-1} P(X_k) = p\sum_{k=m}^{n-1}\frac{1}{k} \approx p\int_{np}^{n}\frac{1}{k} = p\ln k \Big|_{np}^{n}$$

$$= -p\ln p. \tag{2-45}$$

Maximization of the desired probability in (2-45) with respect to p gives

$$\frac{d}{dp}(-p\ln p) = -(1+\ln p) = 0$$

or

$$p = e^{-1} \simeq 0.3679. \tag{2-46}$$

From (2-45), the maximum value for the desired probability of drawing the largest number equals 0.3679 also. Interestingly the above strategy can be used to "play the stock market".

Suppose one gets into the market and decides to stay up to 100 days. The stock values fluctuate day by day, and the important question is when to get out?

According to the above strategy, one should get out at the first opportunity after 37 days, when the stock value exceeds the maximum among the first 37 days. In that case the probability of hitting the top value over 100 days for the stock is also about 37%. Of course, the above argument assumes that the stock values over the period of interest are randomly fluctuating without exhibiting any other trend. Interestingly, such is the case if we consider shorter time frames such as inter-day trading.

In summary if one must day-trade, then a possible strategy might be to get in at 9.30 AM, and get out any time after 12 noon (9.30 AM + 0.3679×6.5 hrs = 11.54 AM to be precise) at the first peak that exceeds the peak value between 9.30 AM and 12 noon. In that case chances are about 37% that one hits the absolute top value for that day! (disclaimer : Trade at your own risk)

We conclude this lecture with a variation of the *Game of craps* discussed in Example 3-16, Text.

Example 2.8: **Game of craps using biased dice:**

From Example 3.16, Text, the probability of winning the game of craps is $0.492929\ \cdots$ for the player. Thus the game is slightly advantageous to the house. This conclusion of course assumes that the two dice in question are perfect cubes. Suppose that is not the case.

Let us assume that the two dice are slightly loaded in such a manner so that the faces 1, 2 and 3 appear with probability $\frac{1}{6}-\varepsilon$ and faces 4, 5 and 6 appear with probability $\frac{1}{6}+\varepsilon,\ \ \varepsilon > 0$ for each dice. If T represents the combined total for the two dice (following Text notation), we get

$$p_4 = P\{T=4\} = P\{(1,3),(2,2),(1,3)\} = 3(\tfrac{1}{6}-\varepsilon)^2$$

$$p_5 = P\{T=5\} = P\{(1,4),(2,3),(3,2),(4,1)\} = 2(\tfrac{1}{36}-\varepsilon^2) + 2(\tfrac{1}{6}-\varepsilon)^2$$

$$p_6 = P\{T=6\} = P\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = 4(\tfrac{1}{36}-\varepsilon^2) + (\tfrac{1}{6}-\varepsilon)^2$$

$$p_7 = P\{T=7\} = P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = 6(\tfrac{1}{36}-\varepsilon^2)$$

$$p_8 = P\{T=8\} = P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = 4(\tfrac{1}{36}-\varepsilon^2) + (\tfrac{1}{6}+\varepsilon)^2$$

$$p_9 = P\{T=9\} = P\{(3,6),(4,5),(5,4),(6,3)\} = 2(\tfrac{1}{36}-\varepsilon^2) + 2(\tfrac{1}{6}+\varepsilon)^2$$

$$p_{10} = P\{T=10\} = P\{(4,6),(5,5),(6,4)\} = 3(\tfrac{1}{6}+\varepsilon)^2$$

$$p_{11} = P\{T=11\} = P\{(5,6),(6,5)\} = 2(\tfrac{1}{6}+\varepsilon)^2.$$

(Note that "(1,3)" above represents the event "the first dice shows face 1, and the second dice shows face 3" etc.)
For $\varepsilon = 0.01,$ we get the following Table:

$T = k$	4	5	6	7	8	9	10	11
$p_k = P\{T = k\}$	0.0706	0.1044	0.1353	0.1661	0.1419	0.1178	0.0936	0.0624

This gives the probability of win on the first throw to be (use (3-56), Text)

$$P_1 = P(T = 7) + P(T = 11) = 0.2285 \qquad (2\text{-}47)$$

and the probability of win by throwing a carry-over to be (use (3-58)-(3-59), Text)

$$P_2 = \sum_{\substack{k=4 \\ k \neq 7}}^{10} \frac{p_k^2}{p_k + p_7} = 0.2717 \qquad (2\text{-}48)$$

Thus

$$P\{\text{winning the game}\} = P_1 + P_2 = 0.5002 \qquad (2\text{-}49)$$

Although perfect dice gives rise to an unfavorable game, a slight loading of the dice turns the fortunes around in favor of the player! (Not an exciting conclusion as far as the casinos are concerned).

Even if we let the two dice to have different loading factors ε_1 and ε_2 (for the situation described above), similar conclusions do follow. For example, $\varepsilon_1 = 0.01$ and $\varepsilon_2 = 0.005$ gives (show this)

$$P\{\text{winning the game}\} = 0.5015. \tag{2-50}$$

Once again the game is in favor of the player!

Although the advantage is very modest in each play, from Bernoulli's theorem the cumulative effect can be quite significant when a large number of game are played. All the more reason for the casinos to keep the dice in perfect shape.

In summary, small chance variations in each game of craps can lead to significant counter-intuitive changes when a large number of games are played. What appears to be a favorable game for the house may indeed become an unfavorable game, and when played repeatedly can lead to unpleasant outcomes.

3. Random Variables

Let (Ω, F, P) be a probability model for an experiment, and X a function that maps every $\xi \in \Omega,$ to a unique point $x \in R,$ the set of real numbers. Since the outcome ξ is not certain, so is the value $X(\xi) = x.$ Thus if B is some subset of R, we may want to determine the probability of "$X(\xi) \in B$". To determine this probability, we can look at the set $A = X^{-1}(B) \in \Omega$ that contains all $\xi \in \Omega$ that maps into B under the function X.

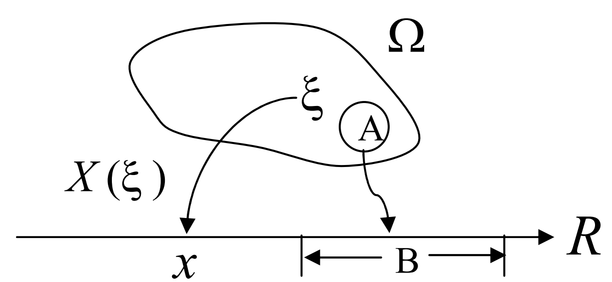


Fig. 3.1

Obviously, if the set $A = X^{-1}(B)$ also belongs to the associated field F, then it is an event and the probability of A is well defined; in that case we can say

$$\text{Probability of the event } "X(\xi) \in B" = P(X^{-1}(B)). \qquad (3\text{-}1)$$

However, $X^{-1}(B)$ may not always belong to F for all B, thus creating difficulties. The notion of random variable (r.v) makes sure that the inverse mapping always results in an event so that we are able to determine the probability for any $B \in R.$

Random Variable (r.v): A finite single valued function $X(\cdot)$ that maps the set of all experimental outcomes Ω into the set of real numbers R is said to be a r.v, if the set $\{\xi \mid X(\xi) \leq x\}$ is an event $(\in F)$ for every x in R.

Alternatively X is said to be a r.v, if $X^{-1}(B) \in F$ where B represents semi-definite intervals of the form $\{-\infty < x \le a\}$ and all other sets that can be constructed from these sets by performing the set operations of union, intersection and negation any number of times. The Borel collection B of such subsets of R is the smallest σ-field of subsets of R that includes all semi-infinite intervals of the above form. Thus if X is a r.v, then

$$\{\xi \mid X(\xi) \le x\} = \{X \le x\} \tag{3-2}$$

is an event for every x. What about $\{a < X \le b\}$, $\{X = a\}$? Are they also events ? In fact with $b > a$ since $\{X \le a\}$ and $\{X \le b\}$ are events, $\{X \le a\}^c = \{X > a\}$ is an event and hence $\{X > a\} \cap \{X \le b\} = \{a < X \le b\}$ is also an event.

Thus, $\left\{ a - \frac{1}{n} < X \le a \right\}$ is an event for every n. Consequently

$$\bigcap_{n=1}^{\infty} \left\{ a - \frac{1}{n} < X \le a \right\} = \{ X = a \} \tag{3-3}$$

is also an event. All events have well defined probability. Thus the probability of the event $\{ \xi \mid X(\xi) \le x \}$ must depend on x. Denote

$$P\{ \xi \mid X(\xi) \le x \} = F_X(x) \ge 0. \tag{3-4}$$

The role of the subscript X in (3-4) is only to identify the actual r.v. $F_X(x)$ is said to the Probability Distribution Function (PDF) associated with the r.v X.

Distribution Function: Note that a distribution function $g(x)$ is nondecreasing, right-continuous and satisfies

$$g(+\infty) = 1, \quad g(-\infty) = 0, \tag{3-5}$$

i.e., if $g(x)$ is a distribution function, then

(i) $g(+\infty) = 1, \quad g(-\infty) = 0,$

(ii) if $x_1 < x_2$, then $g(x_1) \le g(x_2)$, (3-6)

and

(iii) $g(x^+) = g(x)$, for all x.

We need to show that $F_X(x)$ defined in (3-4) satisfies all properties in (3-6). In fact, for any r.v X,

(i) $F_X(+\infty) = P\{\xi \mid X(\xi) \le +\infty\} = P(\Omega) = 1$ (3-7)

and $F_X(-\infty) = P\{\xi \mid X(\xi) \le -\infty\} = P(\phi) = 0.$ (3-8)

(ii) If $x_1 < x_2$, then the subset $(-\infty, x_1) \subset (-\infty, x_2)$. Consequently the event $\{\xi \mid X(\xi) \le x_1\} \subset \{\xi \mid X(\xi) \le x_2\}$, since $X(\xi) \le x_1$ implies $X(\xi) \le x_2$. As a result

$$F_X(x_1) \triangleq P(X(\xi) \le x_1) \le P(X(\xi) \le x_2) \triangleq F_X(x_2), \tag{3-9}$$

implying that the probability distribution function is nonnegative and monotone nondecreasing.

(iii) Let $x < x_n < x_{n-1} < \cdots < x_2 < x_1$, and consider the event

$$A_k = \{\xi \mid x < X(\xi) \le x_k\}. \tag{3-10}$$

since

$$\{x < X(\xi) \le x_k\} \cup \{X(\xi) \le x\} = \{X(\xi) \le x_k\}, \tag{3-11}$$

using mutually exclusive property of events we get

$$P(A_k) = P\left(x < X(\xi) \le x_k\right) = F_X(x_k) - F_X(x). \tag{3-12}$$

But $\cdots A_{k+1} \subset A_k \subset A_{k-1} \cdots,$ and hence

$$\lim_{k\to\infty} A_k = \bigcap_{k=1}^{\infty} A_k = \phi \quad \text{and hence} \quad \lim_{k\to\infty} P(A_k) = 0. \tag{3-13}$$

Thus

$$\lim_{k\to\infty} P(A_k) = \lim_{k\to\infty} F_X(x_k) - F_X(x) = 0.$$

But $\lim_{k\to\infty} x_k = x^+,$ the right limit of x, and hence

$$F_X(x^+) = F_X(x), \tag{3-14}$$

i.e., $F_X(x)$ is right-continuous, justifying all properties of a distribution function.

Additional Properties of a PDF

(iv) If $F_X(x_0) = 0$ for some x_0, then $F_X(x) = 0, \quad x \le x_0.$ (3-15)

This follows, since $F_X(x_0) = P\big(X(\xi) \le x_0\big) = 0$ implies $\{X(\xi) \le x_0\}$ is the null set, and for any $x \le x_0$, $\{\, X(\xi) \le x \,\}$ will be a subset of the null set.

(v) $P\{\, X(\xi) > x \,\} = 1 - F_X(x).$ (3-16)

We have $\{\, X(\xi) \le x \,\} \cup \{\, X(\xi) > x \,\} = \Omega$, and since the two events are mutually exclusive, (16) follows.

(vi) $P\{\, x_1 < X(\xi) \le x_2 \,\} = F_X(x_2) - F_X(x_1), \quad x_2 > x_1.$ (3-17)

The events $\{\, X(\xi) \le x_1 \,\}$ and $\{x_1 < X(\xi) \le x_2\}$ are mutually exclusive and their union represents the event $\{\, X(\xi) \le x_2 \,\}$.

(vii) $P(X(\xi)=x)=F_X(x)-F_X(x^-).$ (3-18)

Let $x_1 = x-\varepsilon,\ \varepsilon > 0,$ and $x_2 = x.$ From (3-17)

$$\lim_{\varepsilon\to 0} P\{\, x-\varepsilon < X(\xi) \le x \,\} = F_X(x) - \lim_{\varepsilon\to 0} F_X(x-\varepsilon), \tag{3-19}$$

or

$$P\{\, X(\xi) = x \,\} = F_X(x) - F_X(x^-). \tag{3-20}$$

According to (3-14), $F_X(x_0^+)$, the limit of $F_X(x)$ as $x \to x_0$ from the right always exists and equals $F_X(x_0)$. However the left limit value $F_X(x_0^-)$ need not equal $F_X(x_0)$. Thus $F_X(x)$ need not be continuous from the left. At a discontinuity point of the distribution, the left and right limits are different, and from (3-20)

$$P\{\, X(\xi) = x_0 \,\} = F_X(x_0) - F_X(x_0^-) > 0. \tag{3-21}$$

Thus the only discontinuities of a distribution function $F_X(x)$ are of the jump type, and occur at points x_0 where (3-21) is satisfied. These points can always be enumerated as a sequence, and moreover they are at most countable in number.

Example 3.1: X is a r.v such that $X(\xi) = c,\ \xi \in \Omega$. Find $F_X(x)$.

Solution: For $x < c,\ \{X(\xi) \le x\} = \{\phi\}$, so that $F_X(x) = 0$, and for $x > c,\ \{X(\xi) \le x\} = \Omega$, so that $F_X(x) = 1$. (Fig.3.2)

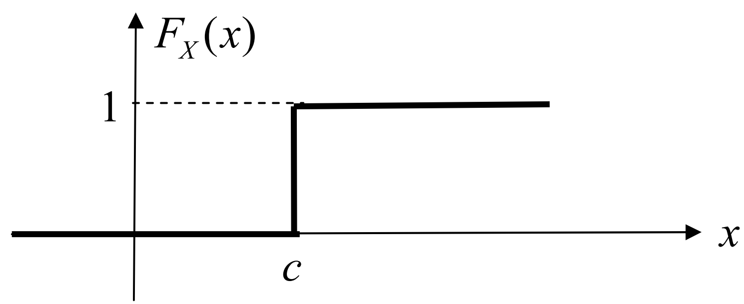


Fig. 3.2

Example 3.2: Toss a coin. $\Omega = \{H, T\}$. Suppose the r.v X is such that $X(T) = 0,\ X(H) = 1$. Find $F_X(x)$.

Solution: For $x < 0, \{ X(\xi) \le x \} = \{\phi\},$ so that $F_X(x) = 0.$

$0 \le x < 1, \; \{ X(\xi) \le x \} = \{ T \},$ so that $F_X(x) = P\{ T \} = 1 - p,$

$x \ge 1, \quad \{ X(\xi) \le x \} = \{ H, T \} = \Omega,$ so that $F_X(x) = 1.$ (Fig. 3.3)

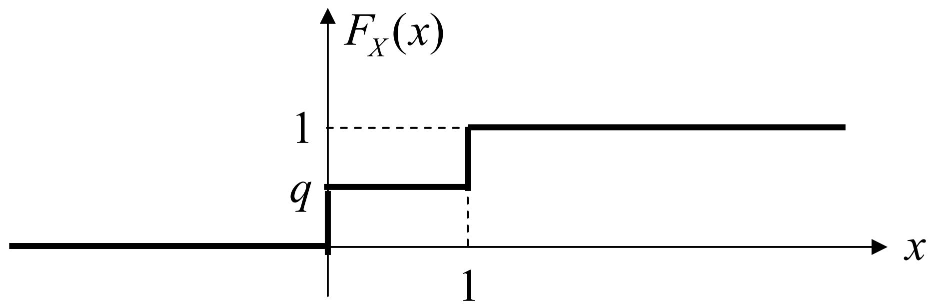


Fig.3.3

•X is said to be a continuous-type r.v if its distribution function $F_X(x)$ is continuous. In that case $F_X(x^-) = F_X(x)$ for all x, and from (3-21) we get $P\{X = x\} = 0.$

•If $F_X(x)$ is constant except for a finite number of jump discontinuities(piece-wise constant; step-type), then X is said to be a discrete-type r.v. If x_i is such a discontinuity point, then from (3-21)

$$p_i = P\{X = x_i\} = F_X(x_i) - F_X(x_i^-). \qquad (3\text{-}22)$$

From Fig.3.2, at a point of discontinuity we get

$$P\{X = c\} = F_X(c) - F_X(c^-) = 1 - 0 = 1.$$

and from Fig.3.3,

$$P\{X = 0\} = F_X(0) - F_X(0^-) = q - 0 = q.$$

Example:3.3 A fair coin is tossed twice, and let the r.v X represent the number of heads. Find $F_X(x)$.
Solution: In this case $\Omega = \{HH, HT, TH, TT\}$, and

$$X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0.$$

$$x < 0, \quad \{X(\xi) \le x\} = \phi \Rightarrow F_X(x) = 0,$$

$$0 \le x < 1, \{X(\xi) \le x\} = \{TT\} \Rightarrow F_X(x) = P\{TT\} = P(T)P(T) = \frac{1}{4},$$

$$1 \le x < 2, \{X(\xi) \le x\} = \{TT, HT, TH\} \Rightarrow F_X(x) = P\{TT, HT, TH\} = \frac{3}{4},$$

$$x \ge 2, \quad \{X(\xi) \le x\} = \Omega \Rightarrow F_X(x) = 1. \text{ (Fig. 3.4)}$$

From Fig.3.4, $P\{X=1\}=F_X(1)-F_X(1^-)=3/4-1/4=1/2.$

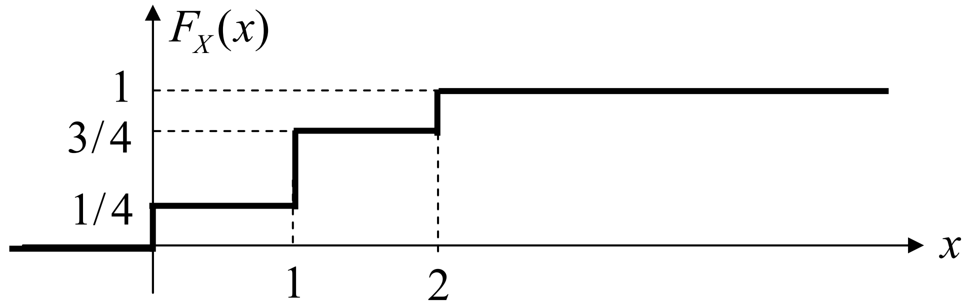


Fig. 3.4

Probability density function (p.d.f)

The derivative of the distribution function $F_X(x)$ is called the probability density function $f_X(x)$ of the r.v X. Thus

$$f_X(x) \stackrel{\Delta}{=} \frac{dF_X(x)}{dx}. \tag{3-23}$$

Since

$$\frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} \geq 0, \tag{3-24}$$

from the monotone-nondecreasing nature of $F_X(x)$, it follows that $f_X(x) \geq 0$ for all x. $f_X(x)$ will be a continuous function, if X is a continuous type r.v. However, if X is a discrete type r.v as in (3-22), then its p.d.f has the general form (Fig. 3.5)

$$f_X(x) = \sum_i p_i \delta(x - x_i), \qquad (3\text{-}25)$$

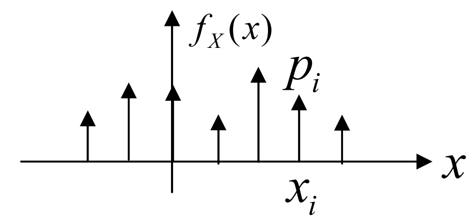


Fig. 3.5

where x_i represent the jump-discontinuity points in $F_X(x)$. As Fig. 3.5 shows $f_X(x)$ represents a collection of positive discrete masses, and it is known as the probability mass function (p.m.f) in the discrete case. From (3-23), we also obtain by integration

$$F_X(x) = \int_{-\infty}^{x} f_x(u)\,du\,. \qquad (3\text{-}26)$$

Since $F_X(+\infty) = 1$, (3-26) yields

$$\int_{-\infty}^{+\infty} f_x(x)\,dx = 1, \qquad (3\text{-}27)$$

which justifies its name as the density function. Further, from (3-26), we also get (Fig. 3.6b)

$$P\{x_1 < X(\xi) \le x_2\} = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx\,. \tag{3-28}$$

Thus the area under $f_X(x)$ in the interval (x_1, x_2) represents the probability in (3-28).

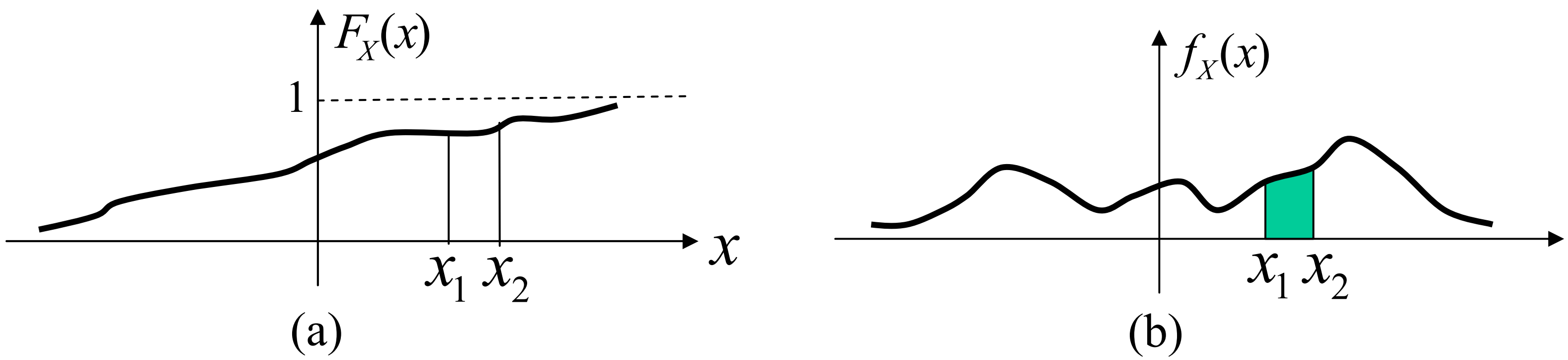


Fig. 3.6

Often, r.vs are referred by their specific density functions - both in the continuous and discrete cases - and in what follows we shall list a number of them in each category.

Continuous-type random variables

1. Normal (Gaussian): X is said to be normal or Gaussian r.v, if

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}. \tag{3-29}$$

This is a bell shaped curve, symmetric around the parameter μ, and its distribution function is given by

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy \stackrel{\Delta}{=} G\left(\frac{x-\mu}{\sigma}\right), \tag{3-30}$$

where $G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ is often tabulated. Since $f_X(x)$ depends on two parameters μ and σ^2, the notation $X \sim N(\mu, \sigma^2)$ will be used to represent (3-29).

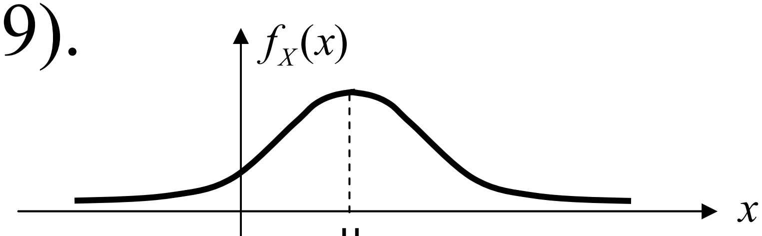


Fig. 3.7

2. Uniform: $X \sim U(a,b),\ a<b,$ if (Fig. 3.8)

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases} \tag{3.31}$$

3. Exponential: $X \sim \varepsilon(\lambda)$ if (Fig. 3.9)

$$f_X(x) = \begin{cases} \dfrac{1}{\lambda} e^{-x/\lambda}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3-32}$$

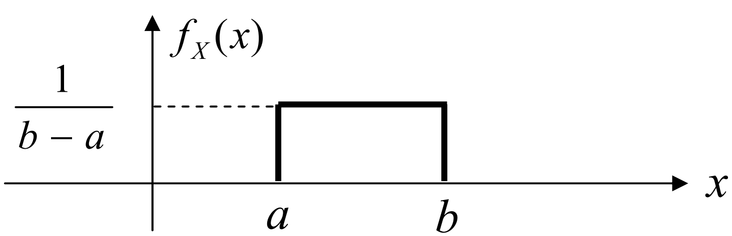


Fig. 3.8

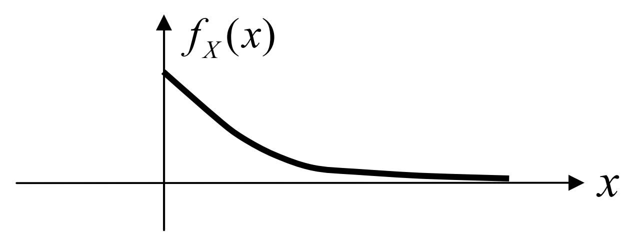


Fig. 3.9

4. Gamma: $X \sim G(\alpha, \beta)$ if $(\alpha > 0,\ \beta > 0)$ (Fig. 3.10)

$$f_X(x) = \begin{cases} \dfrac{x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} e^{-x/\beta}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \qquad (3\text{-}33)$$

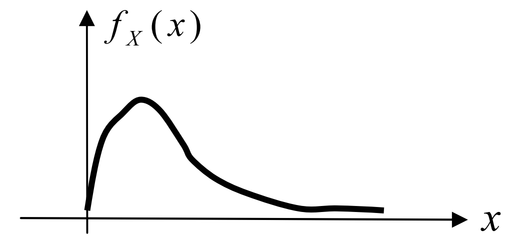


Fig. 3.10

If $\alpha = n$ an integer $\Gamma(n) = (n-1)!$.

5. Beta: $X \sim \beta(a,b)$ if $(a > 0,\ b > 0)$ (Fig. 3.11)

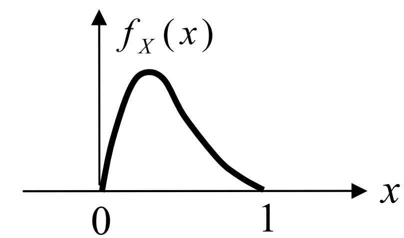


Fig. 3.11

$$f_X(x) = \begin{cases} \dfrac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (3\text{-}34)$$

where the Beta function $\beta(a,b)$ is defined as

$$\beta(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}\, du. \qquad (3\text{-}35)$$

6. Chi-Square: $X \sim \chi^2(n)$, if (Fig. 3.12)

$$f_X(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3-36}$$

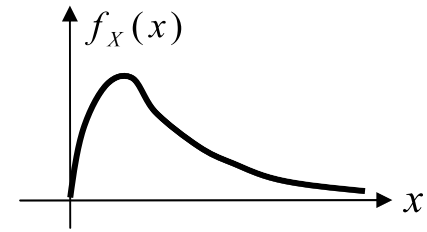


Fig. 3.12

Note that $\chi^2(n)$ is the same as Gamma $(n/2, 2)$.

7. Rayleigh: $X \sim R(\sigma^2)$, if (Fig. 3.13)

$$f_X(x) = \begin{cases} \dfrac{x}{\sigma^2} e^{-x^2/2\sigma^2}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3-37}$$

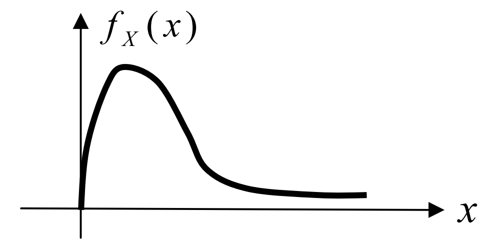


Fig. 3.13

8. Nakagami – m distribution:

$$f_X(x) = \begin{cases} \dfrac{2}{\Gamma(m)} \left(\dfrac{m}{\Omega}\right)^m x^{2m-1} e^{-mx^2/\Omega}, & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3-38}$$

9. Cauchy: $X \sim C(\alpha, \mu)$, if (Fig. 3.14)

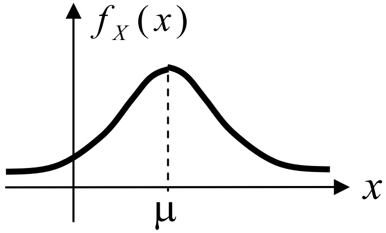


Fig. 3.14

$$f_X(x) = \frac{\alpha/\pi}{\alpha^2 + (x-\mu)^2}, \quad -\infty < x < +\infty. \qquad (3\text{-}39)$$

10. Laplace: (Fig. 3.15)

$$f_X(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}, \quad -\infty < x < +\infty. \qquad (3\text{-}40)$$

11. Student's *t*-distribution with n degrees of freedom (Fig 3.16)

$$f_T(t) = \frac{\Gamma\big((n+1)/2\big)}{\sqrt{\pi n}\,\Gamma(n/2)}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad -\infty < t < +\infty. \qquad (3\text{-}41)$$

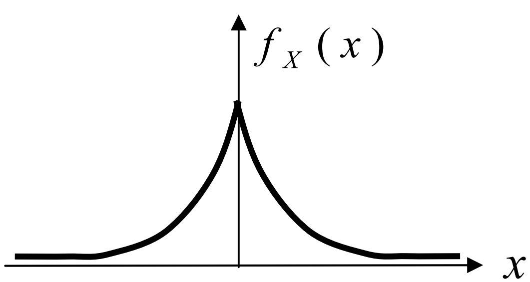


Fig. 3.15

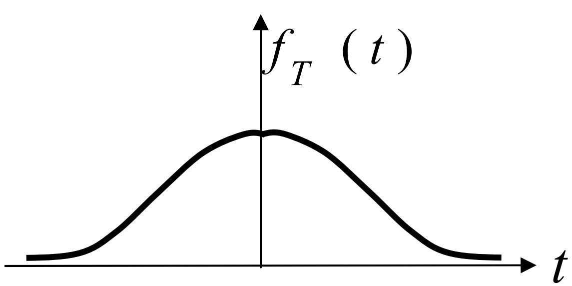


Fig. 3.16

12. Fisher's F-distribution

$$f_z(z) = \begin{cases} \dfrac{\Gamma\{(m+n)/2\}\, m^{m/2} n^{n/2}}{\Gamma(m/2)\,\Gamma(n/2)} \dfrac{z^{m/2-1}}{(n+mz)^{(m+n)/2}}, & z \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad (3\text{-}42)$$

Discrete-type random variables

1. Bernoulli: X takes the values (0,1), and

$$P(X = 0) = q, \quad P(X = 1) = p. \tag{3-43}$$

2. Binomial: $X \sim B(n, p)$, if (Fig. 3.17)

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \cdots, n. \tag{3-44}$$

3. Poisson: $X \sim P(\lambda)$, if (Fig. 3.18)

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \cdots, \infty. \tag{3-45}$$

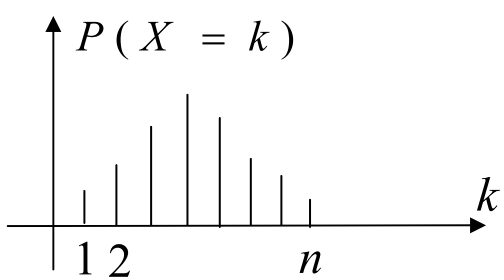


Fig. 3.17

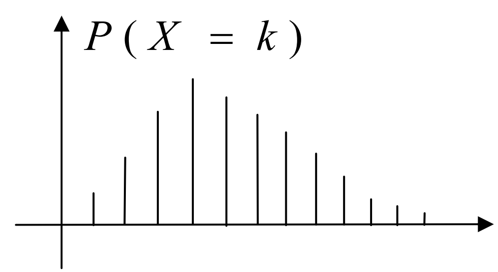


Fig. 3.18

4. Hypergeometric:

$$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}, \qquad \max(0, m+n-N) \le k \le \min(m,n) \tag{3-46}$$

5. Geometric: $X \sim g(p)$ if

$$P(X = k) = pq^{k}, \quad k = 0,1,2,\cdots,\infty, \qquad q = 1 - p. \tag{3-47}$$

6. Negative Binomial: $X \sim NB(r,p)$, if

$$P(X = k) = \binom{k-1}{r-1} p^{r} q^{k-r}, \qquad k = r, r+1, \cdots. \tag{3-48}$$

7. Discrete-Uniform:

$$P(X = k) = \frac{1}{N}, \quad k = 1,2,\cdots,N. \tag{3-49}$$

We conclude this lecture with a general distribution due to Polya that includes both binomial and hypergeometric as special cases.

Polya's distribution: A box contains a white balls and b black balls. A ball is drawn at random, and it is replaced along with c balls of the same color. If X represents the number of white balls drawn in n such draws, $X = 0, 1, 2, \cdots, n$, find the probability mass function of X.

Solution: Consider the specific sequence of draws where k white balls are first drawn, followed by $n - k$ black balls. The probability of drawing k successive white balls is given by

$$p_W = \frac{a}{a+b} \; \frac{a+c}{a+b+c} \; \frac{a+2c}{a+b+2c} \cdots \frac{a+(k-1)c}{a+b+(k-1)c} \qquad \text{(3-50)}$$

Similarly the probability of drawing k white balls followed by $n-k$ black balls is given by

$$p_k = p_w \frac{b}{a+b+kc} \frac{b+c}{a+b+(k+1)c} \cdots \frac{b+(n-k-1)c}{a+b+(n-1)c}$$

$$= \prod_{i=0}^{k-1} \frac{a+ic}{a+b+ic} \prod_{j=0}^{n-k-1} \frac{b+jc}{a+b+(j+k)c}. \qquad (3\text{-}51)$$

Interestingly, p_k in (3-51) also represents the probability of drawing k white balls and $(n-k)$ black balls in *any other specific order* (*i.e.*, The same set of numerator and denominator terms in (3-51) contribute to *all other* sequences as well.) But there are $\binom{n}{k}$ such distinct mutually exclusive sequences and summing over all of them, we obtain the Polya distribution (probability of getting k white balls in n draws) to be

$$P(X=k) = \binom{n}{k} p_k = \binom{n}{k} \prod_{i=0}^{k-1} \frac{a+ic}{a+b+ic} \prod_{j=0}^{n-k-1} \frac{b+jc}{a+b+(j+k)c}, \qquad k = 0,1,2,\cdots,n. \qquad (3\text{-}52)$$

Both binomial distribution as well as the hypergeometric distribution are special cases of (3-52).

For example if draws are done with replacement, then $c = 0$ and (3-52) simplifies to the binomial distribution

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, 2, \cdots, n \tag{3-53}$$

where

$$p = \frac{a}{a+b}, \qquad q = \frac{b}{a+b} = 1 - p.$$

Similarly if the draws are conducted without replacement, Then $c = -1$ in (3-52), and it gives

$$P(X = k) = \binom{n}{k} \frac{a(a-1)(a-2)\cdots(a-k+1)}{(a+b)(a+b-1)\cdots(a+b-k+1)} \frac{b(b-1)\cdots(b-n+k+1)}{(a+b-k)\cdots(a+b-n+1)}$$

$$P(X=k)=\frac{n!}{k!(n-k)!}\;\frac{a!(a+b-k)!}{(a-k)!(a+b)!}\;\frac{b!(a+b-n)!}{(b-n+k)!(a+b-k)!}=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}$$

$$(3\text{-}54)$$

which represents the hypergeometric distribution. Finally $c=+1$ gives (replacements are doubled)

$$P(X=k)=\binom{n}{k}\frac{(a+k-1)!\;(a+b+1)!}{(a-1)!\;(a+b+k-1)!}\;\frac{(b+n-k-1)!\;(a+b+k-1)!}{(b-1)!\;(a+b+n-1)!}$$

$$=\frac{\binom{a+k-1}{k}\binom{b+n-k-1}{n-k}}{\binom{a+b+n-1}{n}}. \qquad (3\text{-}55)$$

we shall refer to (3-55) as Polya's +1 distribution. the general Polya distribution in (3-52) has been used to study the spread of contagious diseases (epidemic modeling).

4. Binomial Random Variable Approximations and Conditional Probability Density Functions

Let X represent a Binomial r.v as in (3-42). Then from (2-30)

$$P(k_1 \le X \le k_2) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k}. \tag{4-1}$$

Since the binomial coefficient $\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$ grows quite rapidly with n, it is difficult to compute (4-1) for large n. In this context, two approximations are extremely useful.

4.1 **The Normal Approximation** (Demoivre-Laplace Theorem) Suppose $n \to \infty$ with p held fixed. Then for k in the $\sqrt{npq}$ neighborhood of np, we can approximate

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq}. \tag{4-2}$$

Thus if k_1 and k_2 in (4-1) are within or around the neighborhood of the interval $\left(np - \sqrt{npq},\ \ np + \sqrt{npq}\right)$, we can approximate the summation in (4-1) by an integration. In that case (4-1) reduces to

$$P\left(k_1 \le X \le k_2\right) = \int_{k_1}^{k_2} \frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq} dx = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \tag{4-3}$$

where

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

We can express (4-3) in terms of the normalized integral

$$erf\,(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy = erf\,(-x) \tag{4-4}$$

that has been tabulated extensively (See Table 4.1).

For example, if x_1 and x_2 are both positive ,we obtain

$$P(k_1 \le X \le k_2) = erf\,(x_2) - erf\,(x_1). \qquad (4\text{-}5)$$

Example 4.1: A fair coin is tossed 5,000 times. Find the probability that the number of heads is between 2,475 to 2,525.

Solution: We need $P(2{,}475 \le X \le 2{,}525)$.Here *n* is large so that we can use the normal approximation. In this case $p = \frac{1}{2}$, so that $np = 2{,}500$ and $\sqrt{npq} \approx 35$. Since $np - \sqrt{npq} = 2{,}465$, and $np + \sqrt{npq} = 2{,}535$, the approximation is valid for $k_1 = 2{,}475$ and $k_2 = 2{,}525$. Thus

$$P(k_1 \le X \le k_2) = \int_{x_1}^{x_2} \frac{1}{2\pi} e^{-y^2/2} dy.$$

Here $x_1 = \dfrac{k_1 - np}{\sqrt{npq}} = -\dfrac{5}{7}$, $x_2 = \dfrac{k_2 - np}{\sqrt{npq}} = \dfrac{5}{7}$.

$$\text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy = G(x) - \frac{1}{2}$$

x	erf(x)	x	erf(x)	x	erf(x)	x	erf(x)
0.05	0.01994	0.80	0.28814	1.55	0.43943	2.30	0.48928
0.10	0.03983	0.85	0.30234	1.60	0.44520	2.35	0.49061
0.15	0.05962	0.90	0.31594	1.65	0.45053	2.40	0.49180
0.20	0.07926	0.95	0.32894	1.70	0.45543	2.45	0.49286
0.25	0.09871	1.00	0.34134	1.75	0.45994	2.50	0.49379
0.30	0.11791	1.05	0.35314	1.80	0.46407	2.55	0.49461
0.35	0.13683	1.10	0.36433	1.85	0.46784	2.60	0.49534
0.40	0.15542	1.15	0.37493	1.90	0.47128	2.65	0.49597
0.45	0.17364	1.20	0.38493	1.95	0.47441	2.70	0.49653
0.50	0.19146	1.25	0.39435	2.00	0.47726	2.75	0.49702
0.55	0.20884	1.30	0.40320	2.05	0.47982	2.80	0.49744
0.60	0.22575	1.35	0.41149	2.10	0.48214	2.85	0.49781
0.65	0.24215	1.40	0.41924	2.15	0.48422	2.90	0.49813
0.70	0.25804	1.45	0.42647	2.20	0.48610	2.95	0.49841
0.75	0.27337	1.50	0.43319	2.25	0.48778	3.00	0.49865

Table 4.1

Since $x_1 < 0$, from Fig. 4.1(b), the above probability is given by $P(2{,}475 \leq X \leq 2{,}525) = \text{erf}(x_2) - \text{erf}(x_1) = \text{erf}(x_2) + \text{erf}(|x_1|)$

$$= 2\,\text{erf}\left(\frac{5}{7}\right) = 0.516,$$

where we have used Table 4.1 $(\text{erf}(0.7) = 0.258)$.

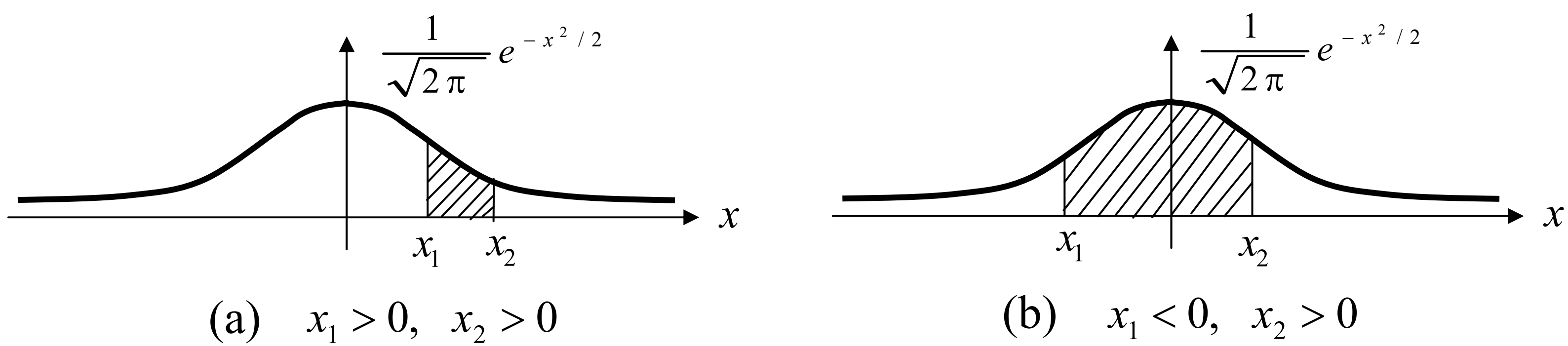


Fig. 4.1

4.2. **The Poisson Approximation**

As we have mentioned earlier, for large *n*, the Gaussian approximation of a binomial r.v is valid only if *p* is fixed, i.e., only if $np \gg 1$ and $npq \gg 1$. what if *np* is small, or if it does not increase with *n*?

Obviously that is the case if, for example, $p \to 0$ as $n \to \infty$, such that $np = \lambda$ is a fixed number.

Many random phenomena in nature in fact follow this pattern. Total number of calls on a telephone line, claims in an insurance company etc. tend to follow this type of behavior. Consider random arrivals such as telephone calls over a line. Let n represent the total number of calls in the interval $(0, T)$. From our experience, as $T \to \infty$ we have $n \to \infty$ so that we may assume $n = \mu T$. Consider a small interval of duration Δ as in Fig. 4.2. If there is only a single call coming in, the probability p of that single call occurring in that interval must depend on its relative size with respect to T.

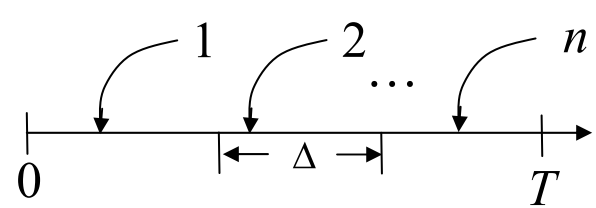


Fig. 4.2

Hence we may assume $p = \frac{\Delta}{T}$. Note that $p \to 0$ as $T \to \infty$. However in this case $np = \mu T \cdot \frac{\Delta}{T} = \mu \Delta = \lambda$ is a constant, and the normal approximation is invalid here.

Suppose the interval Δ in Fig. 4.2 is of interest to us. A call inside that interval is a "success" (H), whereas one outside is a "failure" (T). This is equivalent to the coin tossing situation, and hence the probability $P_n(k)$ of obtaining k calls (in any order) in an interval of duration Δ is given by the binomial p.m.f. Thus

$$P_n(k) = \frac{n!}{(n-k)!\,k!} p^k (1-p)^{n-k}, \tag{4-6}$$

and here as $n \to \infty$, $p \to 0$ such that $np = \lambda$. It is easy to obtain an excellent approximation to (4-6) in that situation. To see this, rewrite (4-6) as

$$
\begin{aligned}
P_n(k) &= \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{(np)^k}{k!}(1-np/n)^{n-k} \\
&= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{\lambda^k}{k!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}.
\end{aligned}
\tag{4-7}
$$

Thus

$$
\lim_{n\to\infty,\, p\to 0,\, np=\lambda} P_n(k) = \frac{\lambda^k}{k!}e^{-\lambda}, \tag{4-8}
$$

since the finite products $\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$ as well as $\left(1-\frac{\lambda}{n}\right)^k$ tend to unity as $n \to \infty$, and

$$
\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}.
$$

The right side of (4-8) represents the Poisson p.m.f and the Poisson approximation to the binomial r.v is valid in situations where the binomial r.v parameters n and p diverge to two extremes $(n \to \infty, p \to 0)$ such that their product np is a constant.

Example 4.2: Winning a Lottery: Suppose two million lottery tickets are issued with 100 winning tickets among them. (a) If a person purchases 100 tickets, what is the probability of winning? (b) How many tickets should one buy to be 95% confident of having a winning ticket?

Solution: The probability of buying a winning ticket

$$p = \frac{\text{No. of winning tickets}}{\text{Total no. of tickets}} = \frac{100}{2\times 10^6} = 5\times 10^{-5}.$$

Here $n = 100$, and the number of winning tickets X in the n purchased tickets has an approximate Poisson distribution with parameter $\lambda = np = 100 \times 5 \times 10^{-5} = 0.005$. Thus

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!},$$

and (a) Probability of winning $= P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda} \approx 0.005.$

(b) In this case we need $P(X \geq 1) \geq 0.95$.

$$P(X \geq 1) = 1 - e^{-\lambda} \geq 0.95 \quad \text{implies} \quad \lambda \geq \ln 20 = 3.$$

But $\lambda = np = n \times 5 \times 10^{-5} \geq 3$ or $n \geq 60{,}000$. Thus one needs to buy about 60,000 tickets to be 95% confident of having a winning ticket!

Example 4.3: A space craft has 100,000 components $(n \to \infty)$ The probability of any one component being defective is 2×10^{-5} $(p \to 0)$. The mission will be in danger if five or more components become defective. Find the probability of such an event.

Solution: Here n is large and p is small, and hence Poisson approximation is valid. Thus $np = \lambda = 100{,}000 \times 2 \times 10^{-5} = 2$, and the desired probability is given by

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{k=0}^{4} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - e^{-2} \sum_{k=0}^{4} \frac{\lambda^k}{k!}$$

$$= 1 - e^{-2}\left(1 + 2 + 2 + \frac{4}{3} + \frac{2}{3}\right) = 0.052.$$

Conditional Probability Density Function

For any two events A and B, we have defined the conditional probability of A given B as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0. \tag{4-9}$$

Noting that the probability distribution function $F_X(x)$ is given by

$$F_X(x) = P\{ X(\xi) \leq x \}, \tag{4-10}$$

we may define the conditional distribution of the r.v X given the event B as

$$F_X(x \mid B) = P\{X(\xi) \le x \mid B\} = \frac{P\{(X(\xi) \le x) \cap B\}}{P(B)}. \qquad (4\text{-}11)$$

Thus the definition of the conditional distribution depends on conditional probability, and since it obeys all probability axioms, it follows that the conditional distribution has the same properties as any distribution function. In particular

$$F_X(+\infty \mid B) = \frac{P\{(X(\xi) \le +\infty) \cap B\}}{P(B)} = \frac{P(B)}{P(B)} = 1,$$
$$F_X(-\infty \mid B) = \frac{P\{(X(\xi) \le -\infty) \cap B\}}{P(B)} = \frac{P(\phi)}{P(B)} = 0. \qquad (4\text{-}12)$$

Further

$$P(x_1 < X(\xi) \le x_2 \mid B) = \frac{P\{(x_1 < X(\xi) \le x_2) \cap B\}}{P(B)}$$
$$= F_X(x_2 \mid B) - F_X(x_1 \mid B), \qquad (4\text{-}13)$$

Since for $x_2 \geq x_1$,

$$\big(X(\xi) \leq x_2\big) = \big(X(\xi) \leq x_1\big) \cup \big(x_1 < X(\xi) \leq x_2\big). \tag{4-14}$$

The conditional density function is the derivative of the conditional distribution function. Thus

$$f_X(x \mid B) = \frac{dF_X(x \mid B)}{dx}, \tag{4-15}$$

and proceeding as in (3-26) we obtain

$$F_X(x \mid B) = \int_{-\infty}^{x} f_X(u \mid B)\,du. \tag{4-16}$$

Using (4-16), we can also rewrite (4-13) as

$$P\big(x_1 < X(\xi) \leq x_2 \mid B\big) = \int_{x_1}^{x_2} f_X(x \mid B)\,dx. \tag{4-17}$$

Example 4.4: Refer to example 3.2. Toss a coin and $X(T)=0$, $X(H)=1$. Suppose $B=\{H\}$. Determine $F_X(x \mid B)$.

Solution: From Example 3.2, $F_X(x)$ has the following form. We need $F_X(x \mid B)$ for all x.

For $x<0$, $\{X(\xi) \le x\} = \phi$, so that $\{(X(\xi) \le x) \cap B\} = \phi$, and $F_X(x \mid B) = 0$.

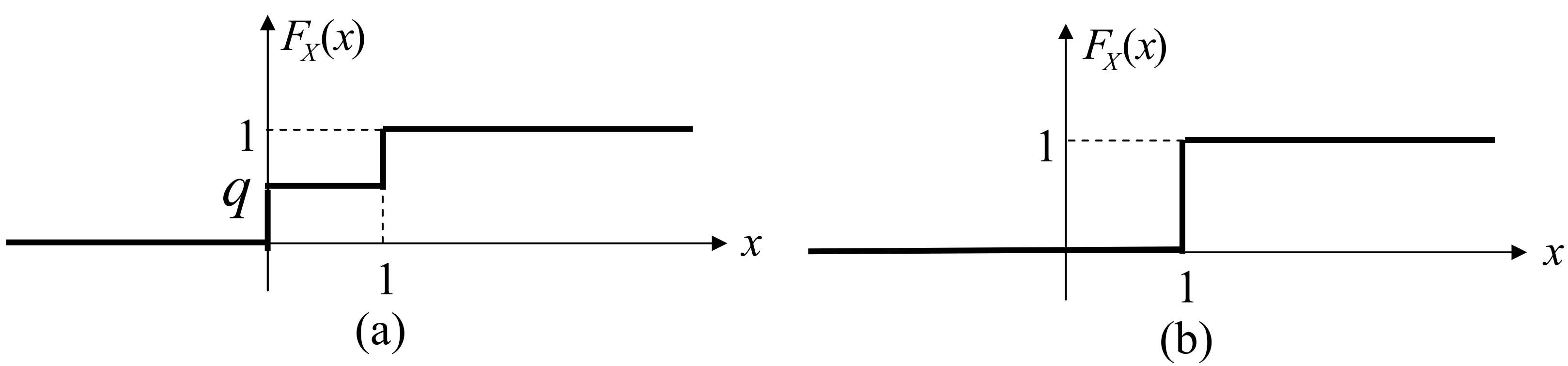


Fig. 4.3

For $0 \le x < 1,\ \{X(\xi) \le x\} = \{T\}$, so that

$\{(X(\xi) \le x) \cap B\} = \{T\} \cap \{H\} = \phi$ and $F_X(x \mid B) = 0.$

For $x \ge 1,\ \{X(\xi) \le x\} = \Omega,$ and

$\{(X(\xi) \le x) \cap B\} = \Omega \cap \{B\} = \{B\}$ and $F_X(x \mid B) = \dfrac{P(B)}{P(B)} = 1$

(see Fig. 4.3(b)).

Example 4.5: Given $F_X(x)$, suppose $B = \{X(\xi) \le a\}$. Find $f_X(x \mid B)$.
Solution: We will first determine $F_X(x \mid B)$. From (4-11) and B as given above, we have

$$F_X(x \mid B) = \frac{P\{(X \le x) \cap (X \le a)\}}{P(X \le a)}. \tag{4-18}$$

For $x < a,\ (X \le x) \cap (X \le a) = (X \le x)$ so that

$$F_X(x \mid B) = \frac{P(X \le x)}{P(X \le a)} = \frac{F_X(x)}{F_X(a)}. \tag{4-19}$$

For $x \ge a,\ (X \le x) \cap (X \le a) = (X \le a)$ so that $F_X(x \mid B) = 1$.
Thus

$$F_X(x \mid B) = \begin{cases} \dfrac{F_X(x)}{F_X(a)}, & x < a, \\ 1, & x \ge a, \end{cases} \tag{4-20}$$

and hence

$$f_X(x \mid B) = \frac{d}{dx} F_X(x \mid B) = \begin{cases} \dfrac{f_X(x)}{F_X(a)}, & x < a, \\ 0, & \text{otherwise.} \end{cases} \tag{4-21}$$

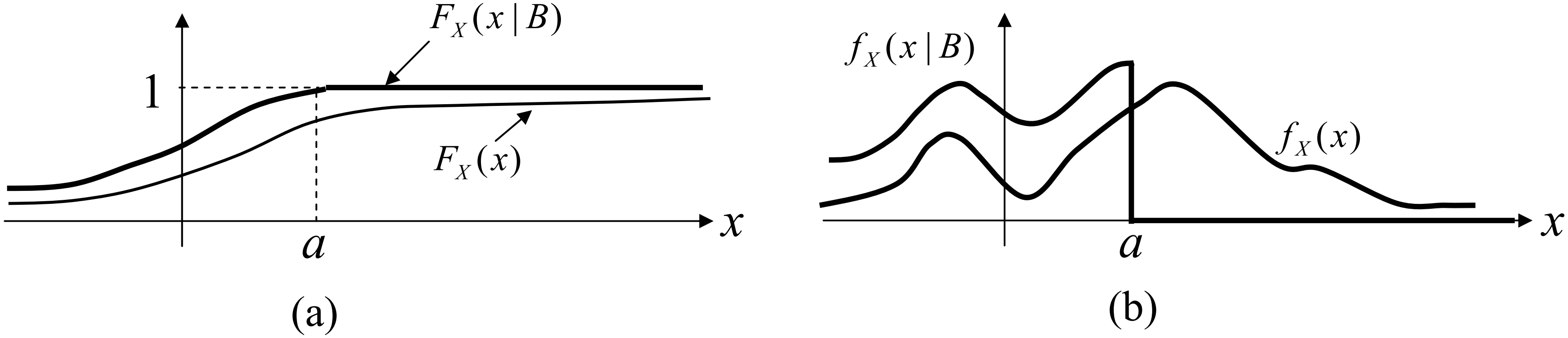


Fig. 4.4

Example 4.6: Let B represent the event $\{a < X(\xi) \le b\}$ with $b > a$. For a given $F_X(x)$, determine $F_X(x \mid B)$ and $f_X(x \mid B)$.

Solution:

$$F_X(x \mid B) = P\{X(\xi) \le x \mid B\} = \frac{P\{(X(\xi) \le x) \cap (a < X(\xi) \le b)\}}{P(a < X(\xi) \le b)}$$

$$= \frac{P\{(X(\xi) \le x) \cap (a < X(\xi) \le b)\}}{F_X(b) - F_X(a)}. \tag{4-22}$$

For $x < a$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \phi$, and hence $F_X(x \mid B) = 0$. (4-23)

For $a \le x < b$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \{a < X(\xi) \le x\}$ and hence

$$F_X(x \mid B) = \frac{P\big(a < X(\xi) \le x\big)}{F_X(b) - F_X(a)} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}. \qquad (4\text{-}24)$$

For $x \ge b$, we have $\{X(\xi) \le x\} \cap \{a < X(\xi) \le b\} = \{a < X(\xi) \le b\}$ so that $F_X(x \mid B) = 1.$ (4-25)

Using (4-23)-(4-25), we get (see Fig. 4.5)

$$f_X(x \mid B) = \begin{cases} \dfrac{f_X(x)}{F_X(b) - F_X(a)}, & a < x \le b, \\ 0, & \text{otherwise.} \end{cases} \qquad (4\text{-}26)$$

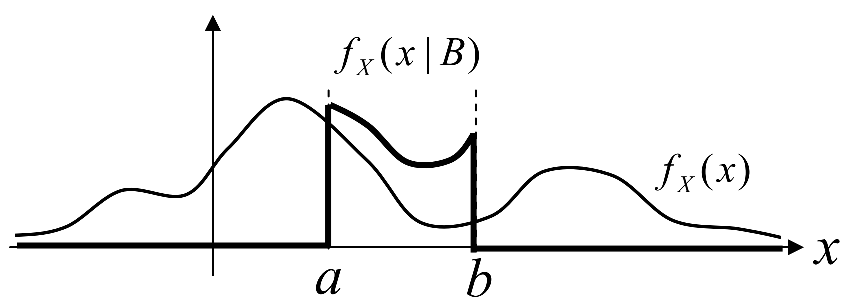


Fig. 4.5

We can use the conditional p.d.f together with the Bayes' theorem to update our a-priori knowledge about the probability of events in presence of new observations. Ideally, any new information should be used to update our knowledge. As we see in the next example, conditional p.d.f together with Bayes' theorem allow systematic updating. For any two events A and B, Bayes' theorem gives

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}. \tag{4-27}$$

Let $B = \{x_1 < X(\xi) \le x_2\}$ so that (4-27) becomes (see (4-13) and (4-17))

$$P\{A \mid (x_1 < X(\xi) \le x_2)\} = \frac{P\big((x_1 < X(\xi) \le x_2) \mid A\big)P(A)}{P\big(x_1 < X(\xi) \le x_2\big)}$$

$$= \frac{F_X(x_2 \mid A) - F_X(x_1 \mid A)}{F_X(x_2) - F_X(x_1)} P(A) = \frac{\int_{x_1}^{x_2} f_X(x \mid A)\,dx}{\int_{x_1}^{x_2} f_X(x)\,dx} P(A). \tag{4-28}$$

Further, let $x_1 = x,\ x_2 = x + \varepsilon,\ \varepsilon > 0,$ so that in the limit as $\varepsilon \to 0,$

$$\lim_{\varepsilon \to 0} P\{A \mid (x < X(\xi) \le x + \varepsilon)\} = P\big(A \mid X(\xi) = x\big) = \frac{f_X(x \mid A)}{f_X(x)} P(A). \tag{4-29}$$

or

$$f_{X|A}(x \mid A) = \frac{P(A \mid X = x) f_X(x)}{P(A)}. \tag{4-30}$$

From (4-30), we also get

$$P(A) \underbrace{\int_{-\infty}^{+\infty} f_X(x \mid A)\, dx}_{1} = \int_{-\infty}^{+\infty} P(A \mid X = x) f_X(x)\, dx, \tag{4-31}$$

or

$$P(A) = \int_{-\infty}^{+\infty} P(A \mid X = x) f_X(x)\, dx \tag{4-32}$$

and using this in (4-30), we get the desired result

$$f_{X|A}(x \mid A) = \frac{P(A \mid X = x) f_X(x)}{\int_{-\infty}^{+\infty} P(A \mid X = x) f_X(x)\, dx}. \tag{4-33}$$

To illustrate the usefulness of this formulation, let us reexamine the coin tossing problem.

Example 4.7: Let $p = P(H)$ represent the probability of obtaining a head in a toss. For a given coin, a-priori p can possess any value in the interval (0,1). In the absence of any additional information, we may assume the a-priori p.d.f $f_P(p)$ to be a uniform distribution in that interval. Now suppose we actually perform an experiment of tossing the coin n times, and k heads are observed. This is new information. How can we update $f_P(p)$?

Solution: Let A= "k heads in n specific tosses". Since these tosses result in a specific sequence,

$$P(A \mid P = p) = p^k q^{n-k}, \qquad \text{(4-34)}$$

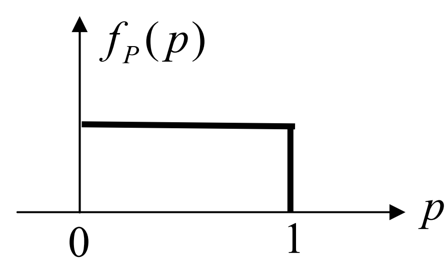


Fig.4.6

and using (4-32) we get

$$P(A) = \int_0^1 P(A \mid P = p) f_P(p) dp = \int_0^1 p^k (1-p)^{n-k} dp = \frac{(n-k)!\, k!}{(n+1)!}. \qquad \text{(4-35)}$$

The a-posteriori p.d.f $f_P(p \mid A)$ represents the updated information given the event *A*, and from (4-30)

$$f_{P|A}(p \mid A) = \frac{P(A \mid P = p) f_P(p)}{P(A)}$$

$$= \frac{(n+1)!}{(n-k)!\, k!} p^k q^{n-k}, \quad 0 < p < 1 \sim \beta(n, k). \qquad \text{(4-36)}$$

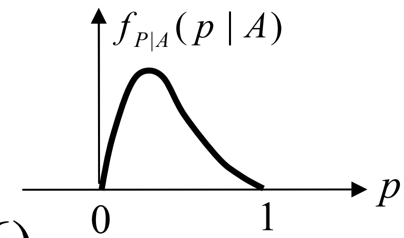


Fig. 4.7

Notice that the a-posteriori p.d.f of *p* in (4-36) is not a uniform distribution, but a beta distribution. We can use this a-posteriori p.d.f to make further predictions, For example, in the light of the above experiment, what can we say about the probability of a head occurring in the next $(n+1)$th toss?

Let B= "head occurring in the (n+1)th toss, given that k heads have occurred in n previous tosses".

Clearly $P(B \mid P = p) = p,$ and from (4-32)

$$P(B) = \int_0^1 P(B \mid P = p) f_P(p \mid A) dp. \tag{4-37}$$

Notice that unlike (4-32), we have used the a-posteriori p.d.f in (4-37) to reflect our knowledge about the experiment already performed. Using (4-36) in (4-37), we get

$$P(B) = \int_0^1 p \cdot \frac{(n+1)!}{(n-k)!\, k!} p^k q^{n-k} dp = \frac{k+1}{n+2}. \tag{4-38}$$

Thus, if n =10, and k = 6, then

$$P(B) = \frac{7}{12} = 0.58,$$

which is better than p = 0.5.

To summarize, if the probability of an event X is unknown, one should make noncommittal judgement about its a-priori probability density function $f_X(x)$. Usually the uniform distribution is a reasonable assumption in the absence of any other information. Then experimental results (A) are obtained, and out knowledge about X must be updated reflecting this new information. Bayes' rule helps to obtain the a-posteriori p.d.f of X given A. From that point on, this a-posteriori p.d.f $f_{X|A}(x \mid A)$ should be used to make further predictions and calculations.

5. Functions of a Random Variable

Let X be a r.v defined on the model (Ω, F, P), and suppose $g(x)$ is a function of the variable x. Define

$$Y = g(X). \tag{5-1}$$

Is Y necessarily a r.v? If so what is its PDF $F_Y(y)$, pdf $f_Y(y)$?

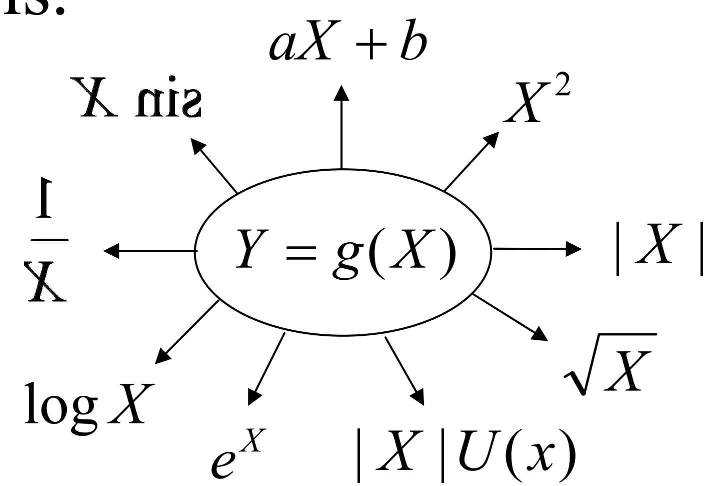
Clearly if Y is a r.v, then for every Borel set B, the set of ξ for which $Y(\xi) \in B$ must belong to F. Given that X is a r.v, this is assured if $g^{-1}(B)$ is also a Borel set, i.e., if $g(x)$ is a Borel function. In that case if X is a r.v, so is Y, and for every Borel set B

$$P(Y \in B) = P(X \in g^{-1}(B)). \tag{5-2}$$

In particular

$$F_Y(y) = P(Y(\xi) \le y) = P\big(g(X(\xi)) \le y\big) = P\big(X(\xi) \le g^{-1}(-\infty, y]\big). \quad (5\text{-}3)$$

Thus the distribution function as well of the density function of Y can be determined in terms of that of X. To obtain the distribution function of Y, we must determine the Borel set on the x-axis such that $X(\xi) \le g^{-1}(y)$ for every given y, and the probability of that set. At this point, we shall consider some of the following functions to illustrate the technical details.



Example 5.1: $Y = aX + b$ (5-4)

Solution: Suppose $a > 0$.

$$F_Y(y) = P(Y(\xi) \le y) = P(aX(\xi) + b \le y) = P\left(X(\xi) \le \frac{y-b}{a} \right) = F_X\left(\frac{y-b}{a} \right). \quad (5\text{-}5)$$

and

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a} \right). \quad (5\text{-}6)$$

On the other hand if $a < 0$, then

$$F_Y(y) = P(Y(\xi) \le y) = P(aX(\xi) + b \le y) = P\left(X(\xi) > \frac{y-b}{a} \right)$$

$$= 1 - F_X\left(\frac{y-b}{a} \right), \quad (5\text{-}7)$$

and hence

$$f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a} \right). \quad (5\text{-}8)$$

From (5-6) and (5-8), we obtain (for all a)

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right). \tag{5-9}$$

Example 5.2: $Y = X^2.$ (5-10)

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P\big(X^2(\xi) \le y\big). \tag{5-11}$$

If $y < 0,$ then the event $\{X^2(\xi) \le y\} = \phi$, and hence

$$F_Y(y) = 0, \quad y < 0. \tag{5-12}$$

For $y > 0,$ from Fig. 5.1, the event $\{Y(\xi) \le y\} = \{X^2(\xi) \le y\}$ is equivalent to $\{x_1 < X(\xi) \le x_2\}.$

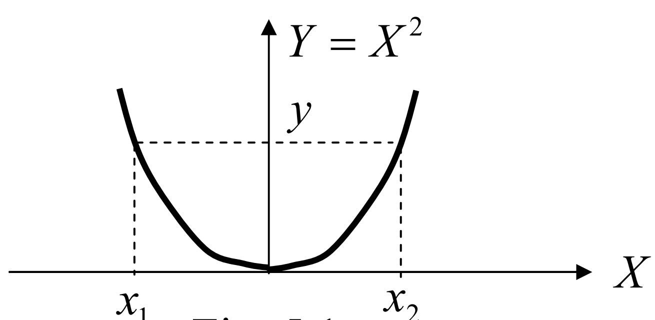


Fig. 5.1

Hence

$$F_Y(y) = P\left(x_1 < X(\xi) \le x_2\right) = F_X(x_2) - F_X(x_1)$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \quad y > 0. \tag{5-13}$$

By direct differentiation, we get

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right), & y > 0, \\ 0, & \text{otherwise} \end{cases}. \tag{5-14}$$

If $f_X(x)$ represents an even function, then (5-14) reduces to

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X\left(\sqrt{y}\right) U(y). \tag{5-15}$$

In particular if $X \sim N(0,1)$, so that

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{5-16}$$

and substituting this into (5-14) or (5-15), we obtain the p.d.f of $Y = X^2$ to be

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} U(y). \tag{5-17}$$

On comparing this with (3-36), we notice that (5-17) represents a Chi-square r.v with $n = 1$, since $\Gamma(1/2) = \sqrt{\pi}$. Thus, if X is a Gaussian r.v with $\mu = 0$, then $Y = X^2$ represents a Chi-square r.v with one degree of freedom ($n = 1$).

Example 5.3: Let

$$Y = g(X) = \begin{cases} X - c, & X > c, \\ 0, & -c < X \le c, \\ X + c, & X \le -c. \end{cases}$$

In this case

$$P(Y=0)=P(-c<X(\xi)\le c)=F_X(c)-F_X(-c). \tag{5-18}$$

For $y>0$, we have $x>c$, and $Y(\xi)=X(\xi)-c$ so that

$$\begin{aligned} F_Y(y) &= P\big(Y(\xi)\le y\big)=P(X(\xi)-c\le y) \\ &= P\big(X(\xi)\le y+c\big)=F_X(y+c), \quad y>0. \end{aligned} \tag{5-19}$$

Similarly $y<0$, if $x<-c$, and $Y(\xi)=X(\xi)+c$ so that

$$\begin{aligned} F_Y(y) &= P\big(Y(\xi)\le y\big)=P(X(\xi)+c\le y) \\ &= P\big(X(\xi)\le y-c\big)=F_X(y-c), \quad y<0. \end{aligned} \tag{5-20}$$

Thus

$$f_Y(y)=\begin{cases} f_X(y+c), & y\ge 0, \\ f_X(y-c), & y<0. \end{cases} \tag{5-21}$$

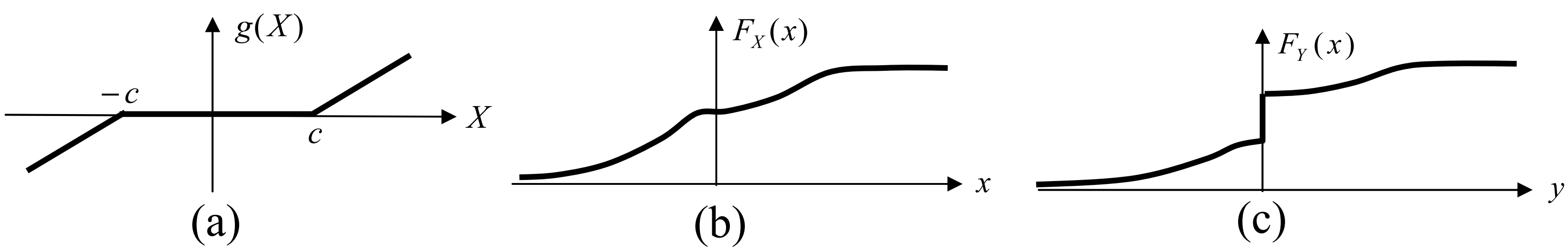


Fig. 5.2

Example 5.4: Half-wave rectifier

$$Y = g(X); \quad g(x) = \begin{cases} x, & x > 0, \\ 0, & x \le 0. \end{cases} \tag{5-22}$$

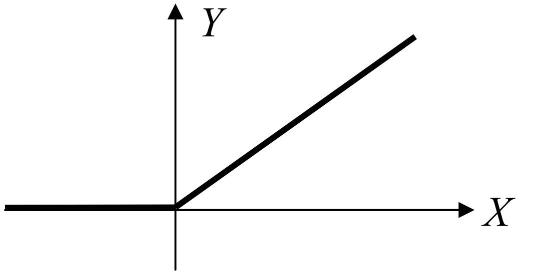


Fig. 5.3

In this case

$$P(Y = 0) = P(X(\xi) \le 0) = F_X(0). \tag{5-23}$$

and for $y > 0$, since $Y = X$,

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P\big(X(\xi) \le y\big) = F_X(y). \tag{5-24}$$

Thus

$$f_Y(y) = \begin{cases} f_X(y), & y > 0, \\ 0, & y \le 0, \end{cases} = f_X(y)U(y). \tag{5-25}$$

Note: As a general approach, given $Y = g(X)$, first sketch the graph $y = g(x)$, and determine the range space of y. Suppose $a < y < b$ is the range space of $y = g(x)$. Then clearly for $y < a$, $F_Y(y) = 0$, and for $y > b$, $F_Y(y) = 1$, so that $F_Y(y)$ can be nonzero only in $a < y < b$. Next, determine whether there are discontinuities in the range space of y. If so evaluate $P\left(Y(\xi) = y_i\right)$ at these discontinuities. In the continuous region of y, use the basic approach

$$F_Y(y) = P\left(g(X(\xi)) \le y\right)$$

and determine appropriate events in terms of the r.v X for every y. Finally, we must have $F_Y(y)$ for $-\infty \le y \le +\infty$, and obtain

$$f_Y(y) = \frac{dF_Y(y)}{dy} \quad \text{in} \quad -a < y < b.$$

However, if $Y = g(X)$ is a continuous function, it is easy to establish a direct procedure to obtain $f_Y(y)$. A continuos function $g(x)$ with $g'(x)$ nonzero at all but a finite number of points, has only a finite number of maxima and minima, and it eventually becomes monotonic as $|x| \to \infty$. Consider a specific y on the y-axis, and a positive increment Δy as shown in Fig. 5.4

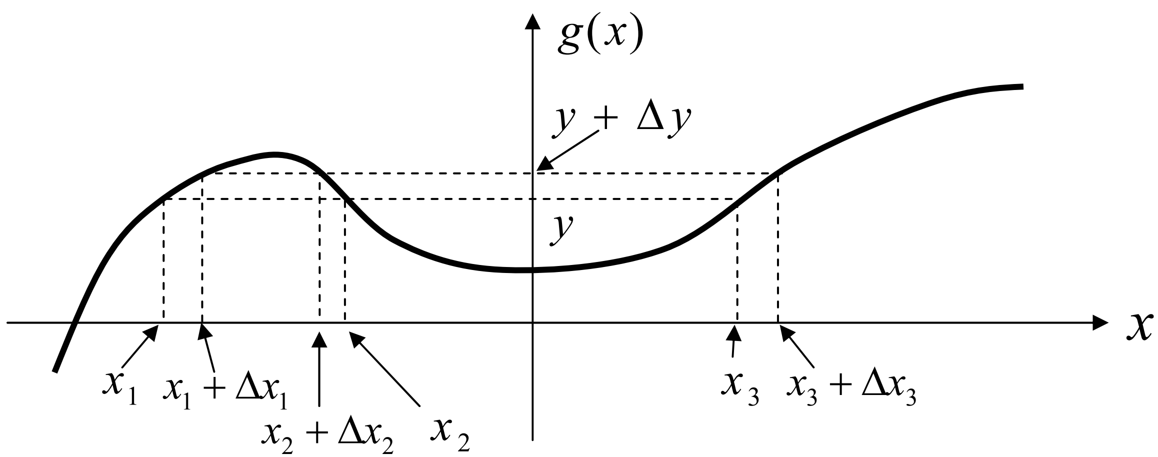


Fig. 5.4

$f_Y(y)$ for $Y = g(X)$, where $g(\cdot)$ is of continuous type.

Using (3-28) we can write

$$P\{y < Y(\xi) \le y + \Delta y\} = \int_{y}^{y+\Delta y} f_Y(u)\,du \approx f_Y(y)\cdot\Delta y. \tag{5-26}$$

But the event $\{y < Y(\xi) \le y + \Delta y\}$ can be expressed in terms of $X(\xi)$ as well. To see this, referring back to Fig. 5.4, we notice that the equation $y = g(x)$ has three solutions x_1, x_2, x_3 (for the specific y chosen there). As a result when $\{y < Y(\xi) \le y + \Delta y\}$, the r.v X could be in any one of the three mutually exclusive intervals

$$\{x_1 < X(\xi) \le x_1 + \Delta x_1\},\ \{x_2 + \Delta x_2 < X(\xi) \le x_2\}\ \text{or}\ \{x_3 < X(\xi) \le x_3 + \Delta x_3\}.$$

Hence the probability of the event in (5-26) is the sum of the probability of the above three events, i.e.,

$$P\{y < Y(\xi) \le y + \Delta y\} = P\{x_1 < X(\xi) \le x_1 + \Delta x_1\}$$
$$+ P\{x_2 + \Delta x_2 < X(\xi) \le x_2\} + P\{x_3 < X(\xi) \le x_3 + \Delta x_3\}. \tag{5-27}$$

For small $\Delta y, \Delta x_i$, making use of the approximation in (5-26), we get

$$f_Y(y)\Delta y = f_X(x_1)\Delta x_1 + f_X(x_2)(-\Delta x_2) + f_X(x_3)\Delta x_3. \qquad (5\text{-}28)$$

In this case, $\Delta x_1 > 0,\ \ \Delta x_2 < 0$ and $\Delta x_3 > 0,$ so that (5-28) can be rewritten as

$$f_Y(y) = \sum_i f_X(x_i)\frac{|\,\Delta x_i\,|}{\Delta y} = \sum_i \frac{1}{\left|\Delta y / \Delta x_i\right|} f_X(x_i) \qquad (5\text{-}29)$$

and as $\Delta y \to 0,$ (5-29) can be expressed as

$$f_Y(y) = \sum_i \frac{1}{\left|dy/dx\right|_{x_i}} f_X(x_i) = \sum_i \frac{1}{\left|g'(x_i)\right|} f_X(x_i). \qquad (5\text{-}30)$$

The summation index *i* in (5-30) depends on *y*, and for every *y* the equation $y = g(x_i)$ must be solved to obtain the total number of solutions at every *y*, and the actual solutions $x_1, x_2, \cdots$ all in terms of *y*.

For example, if $Y = X^2$, then for all $y > 0$, $x_1 = -\sqrt{y}$ and $x_2 = +\sqrt{y}$ represent the two solutions for each y. Notice that the solutions x_i are all in terms of y so that the right side of (5-30) is only a function of y. Referring back to the example $Y = X^2$ (Example 5.2) here for each $y > 0$, there are two solutions given by $x_1 = -\sqrt{y}$ and $x_2 = +\sqrt{y}$. ($f_Y(y) = 0$ for $y < 0$).

Moreover

$$\frac{dy}{dx} = 2x \quad \text{so that} \quad \left|\frac{dy}{dx}\right|_{x=x_i} = 2\sqrt{y}$$

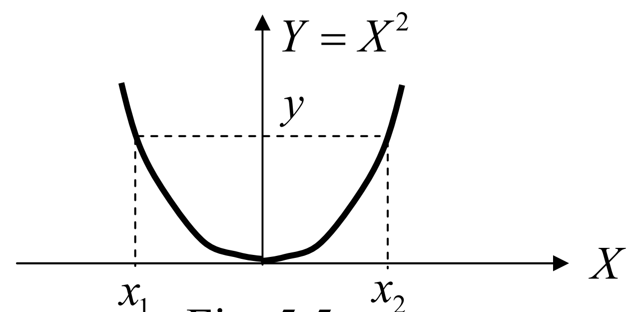


Fig. 5.5

and using (5-30) we get

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right), & y > 0, \\ 0, & \text{otherwise} \end{cases} \tag{5-31}$$

which agrees with (5-14).

Example 5.5: $Y = \dfrac{1}{X}$. Find $f_Y(y)$. (5-32)

Solution: Here for every y, $x_1 = 1/y$ is the only solution, and

$$\frac{dy}{dx} = -\frac{1}{x^2} \quad \text{so that} \quad \left|\frac{dy}{dx}\right|_{x=x_1} = \frac{1}{1/y^2} = y^2,$$

and substituting this into (5-30), we obtain

$$f_Y(y) = \frac{1}{y^2} f_X\left(\frac{1}{y}\right). \tag{5-33}$$

In particular, suppose X is a Cauchy r.v as in (3-38) with parameter α so that

$$f_X(x) = \frac{\alpha/\pi}{\alpha^2 + x^2}, \quad -\infty < x < +\infty. \tag{5-34}$$

In that case from (5-33), $Y = 1/X$ has the p.d.f

$$f_Y(y) = \frac{1}{y^2}\frac{\alpha/\pi}{\alpha^2 + (1/y)^2} = \frac{(1/\alpha)/\pi}{(1/\alpha)^2 + y^2}, \quad -\infty < y < +\infty. \tag{5-35}$$

But (5-35) represents the p.d.f of a Cauchy r.v with parameter $1/\alpha$. Thus if $X \sim C(\alpha)$, then $1/X \sim C(1/\alpha)$.

Example 5.6: Suppose $f_X(x) = 2x/\pi^2, \; 0 < x < \pi$, and $Y = \sin X$. Determine $f_Y(y)$.

Solution: Since X has zero probability of falling outside the interval $(0, \pi)$, $y = \sin x$ has zero probability of falling outside the interval $(0,1)$. Clearly $f_Y(y) = 0$ outside this interval. For any $0 < y < 1$, from Fig.5.6(b), the equation $y = \sin x$ has an infinite number of solutions $\cdots, x_1, x_2, x_3, \cdots$, where $x_1 = \sin^{-1} y$ is the principal solution. Moreover, using the symmetry we also get $x_2 = \pi - x_1$ etc. Further,

$$\frac{dy}{dx} = \cos x = \sqrt{1 - \sin^2 x} = \sqrt{1 - y^2}$$

so that

$$\left|\frac{dy}{dx}\right|_{x=x_i} = \sqrt{1 - y^2}.$$

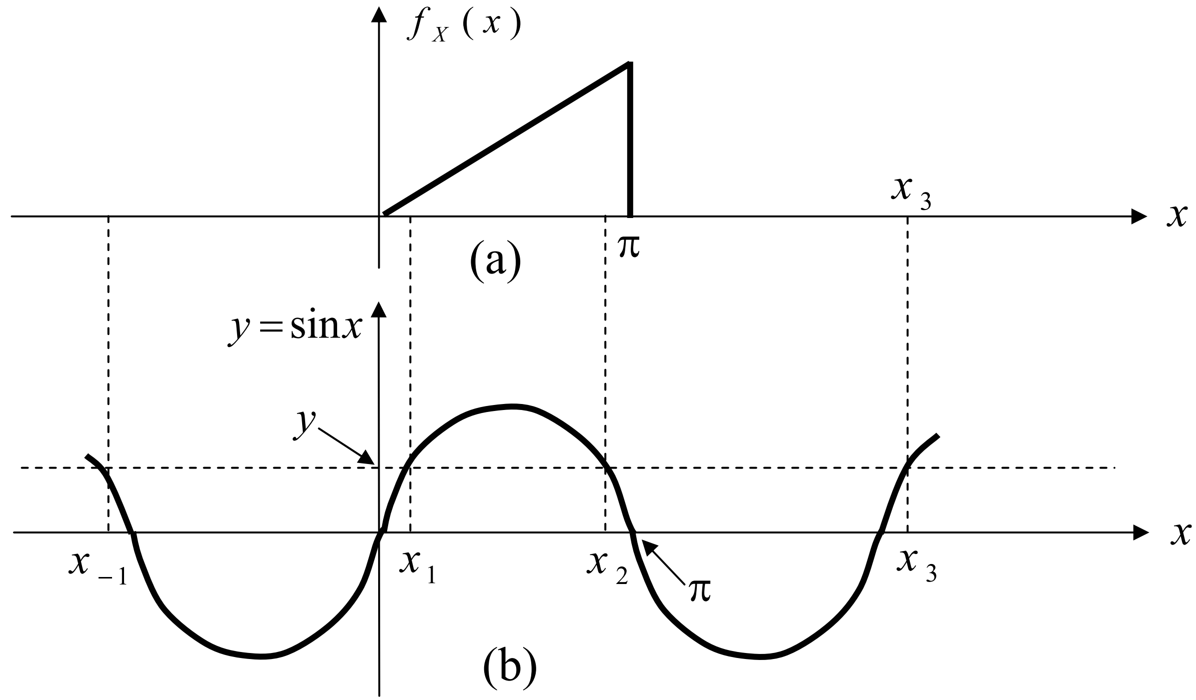


Fig. 5.6

Using this in (5-30), we obtain for $0 < y < 1,$

$$f_Y(y) = \sum_{i=-\infty}^{+\infty} \frac{1}{\sqrt{1-y^2}} f_X(x_i). \tag{5-36}$$

But from Fig. 5.6(a), in this case $f_X(x_{-1}) = f_X(x_3) = f_X(x_4) = \cdots = 0$ (Except for $f_X(x_1)$ and $f_X(x_2)$ the rest are all zeros).

Thus (Fig. 5.7)

$$f_Y(y) = \frac{1}{\sqrt{1-y^2}}\left(f_X(x_1) + f_X(x_2)\right) = \frac{1}{\sqrt{1-y^2}}\left(\frac{2x_1}{\pi^2} + \frac{2x_2}{\pi^2}\right)$$

$$= \frac{2(x_1 + \pi - x_1)}{\pi^2\sqrt{1-y^2}} = \begin{cases} \dfrac{2}{\pi\sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (5\text{-}37)$$

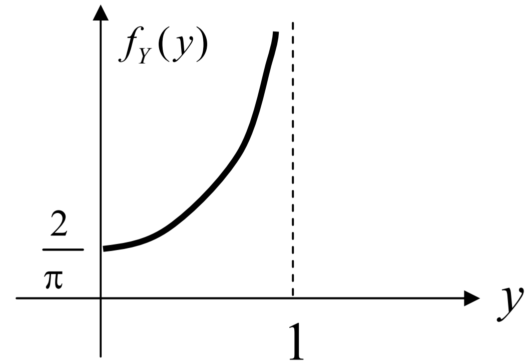


Fig. 5.7

Example 5.7: Let $Y = \tan X$ where $X \sim U(-\pi/2, \pi/2)$. Determine $f_Y(y)$.

Solution: As x moves from $(-\pi/2, \pi/2)$, y moves from $(-\infty, +\infty)$. From Fig.5.8(b), the function $Y = \tan X$ is one-to-one for $-\pi/2 < x < \pi/2$. For any y, $x_1 = \tan^{-1} y$ is the principal solution. Further

$$\frac{dy}{dx} = \frac{d\tan x}{dx} = \sec^2 x = 1 + \tan^2 x = 1 + y^2$$

so that using (5-30)

$$f_Y(y) = \frac{1}{|dy/dx|_{x=x_1}} f_X(x_1) = \frac{1/\pi}{1+y^2}, \quad -\infty < y < +\infty, \tag{5-38}$$

which represents a Cauchy density function with parameter equal to unity (Fig. 5.9).

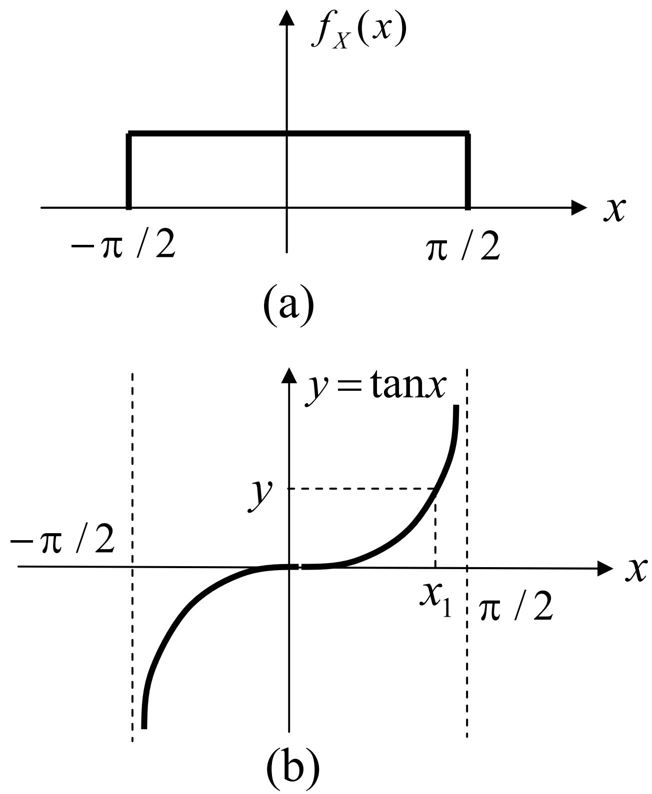


Fig. 5.8

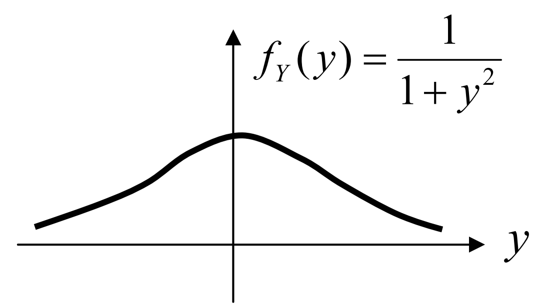


Fig. 5.9

Functions of a discrete-type r.v

Suppose X is a discrete-type r.v with

$$P(X = x_i) = p_i, \quad x = x_1, x_2, \cdots, x_i, \cdots \tag{5-39}$$

and $Y = g(X)$. Clearly Y is also of discrete-type, and when $x = x_i,\ \ y_i = g(x_i)$, and for those y_i

$$P(Y = y_i) = P(X = x_i) = p_i, \quad y = y_1, y_2, \cdots, y_i, \cdots \tag{5-40}$$

Example 5.8: Suppose $X \sim P(\lambda)$, so that

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0,1,2,\cdots \tag{5-41}$$

Define $Y = X^2 + 1$. Find the p.m.f of Y.
Solution: X takes the values $0,1,2,\cdots,k,\cdots$ so that Y only takes the value $1,3,\cdots,k^2+1,\cdots$ and

$$P(Y = k^2 + 1) = P(X = k)$$

so that for $j = k^2 + 1$

$$P(Y = j) = P\left(X = \sqrt{j-1}\right) = e^{-\lambda} \frac{\lambda^{\sqrt{j-1}}}{(\sqrt{j-1})!}, \quad j = 1, 3, \cdots, k^2 + 1, \cdots. \quad (5\text{-}42)$$

6. Mean, Variance, Moments and Characteristic Functions

For a r.v X, its p.d.f $f_X(x)$ represents complete information about it, and for any Borel set B on the x-axis

$$P\big(X(\xi) \in B\big) = \int_B f_X(x)\,dx \,. \tag{6-1}$$

Note that $f_X(x)$ represents very detailed information, and quite often it is desirable to characterize the r.v in terms of its average behavior. In this context, we will introduce two parameters - mean and variance - that are universally used to represent the overall properties of the r.v and its p.d.f.

Mean or the **Expected Value** of a r.v X is defined as

$$\eta_X = \overline{X} = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)dx. \tag{6-2}$$

If X is a discrete-type r.v, then using (3-25) we get

$$\eta_X = \overline{X} = E(X) = \int x \sum_i p_i \delta(x - x_i)dx = \sum_i x_i p_i \underbrace{\int \delta(x - x_i)dx}_{1}$$

$$= \sum_i x_i p_i = \sum_i x_i\, P(X = x_i). \tag{6-3}$$

Mean represents the average (mean) value of the r.v in a very large number of trials. For example if $X \sim U(a,b)$, then using (3-31) ,

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left. \frac{x^2}{2} \right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2} \tag{6-4}$$

is the midpoint of the interval (a,b).

On the other hand if X is exponential with parameter λ as in (3-32), then

$$E(X)=\int_0^\infty \frac{x}{\lambda}e^{-x/\lambda}dx=\lambda\int_0^\infty ye^{-y}dy=\lambda, \tag{6-5}$$

implying that the parameter λ in (3-32) represents the mean value of the exponential r.v.

Similarly if X is Poisson with parameter λ as in (3-43), using (6-3), we get

$$E(X)=\sum_{k=0}^{\infty}kP(X=k)=\sum_{k=0}^{\infty}ke^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=1}^{\infty}k\frac{\lambda^k}{k!}$$

$$=e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}=\lambda e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}=\lambda e^{-\lambda}e^{\lambda}=\lambda. \tag{6-6}$$

Thus the parameter λ in (3-43) also represents the mean of the Poisson r.v.

In a similar manner, if X is binomial as in (3-42), then its mean is given by

$$E(X)=\sum_{k=0}^{n} kP(X=k)=\sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k}=\sum_{k=1}^{n} k\frac{n!}{(n-k)!\,k!}p^k q^{n-k}$$

$$=\sum_{k=1}^{n}\frac{n!}{(n-k)!\,(k-1)!}p^k q^{n-k}=np\sum_{i=0}^{n-1}\frac{(n-1)!}{(n-i-1)!\,i!}p^i q^{n-i-1}=np(p+q)^{n-1}=np. \tag{6-7}$$

Thus np represents the mean of the binomial r.v in (3-42).

For the normal r.v in (3-29),

$$E(X)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} xe^{-(x-\mu)^2/2\sigma^2}dx=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty}(y+\mu)e^{-y^2/2\sigma^2}dy$$

$$=\frac{1}{\sqrt{2\pi\sigma^2}}\underbrace{\int_{-\infty}^{+\infty} ye^{-y^2/2\sigma^2}dy}_{0}+\mu\cdot\underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} e^{-y^2/2\sigma^2}dy}_{1}=\mu. \tag{6-8}$$

Thus the first parameter in $X \sim N(\mu, \sigma^2)$ is infact the mean of the Gaussian r.v X. Given $X \sim f_X(x)$, suppose $Y = g(X)$ defines a new r.v with p.d.f $f_Y(y)$. Then from the previous discussion, the new r.v Y has a mean μ_Y given by (see (6-2))

$$\mu_Y = E(Y) = \int_{-\infty}^{+\infty} y \, f_Y(y) dy. \tag{6-9}$$

From (6-9), it appears that to determine $E(Y)$, we need to determine $f_Y(y)$. However this is not the case if only $E(Y)$ is the quantity of interest. Recall that for any y, $\Delta y > 0$

$$P\left(y < Y \le y + \Delta y\right) = \sum_i P\left(x_i < X \le x_i + \Delta x_i\right), \tag{6-10}$$

where x_i represent the multiple solutions of the equation $y = g(x_i)$. But(6-10) can be rewritten as

$$f_Y(y)\Delta y = \sum_i f_X(x_i)\Delta x_i, \tag{6-11}$$

where the $(x_i,\ x_i+\Delta x_i)$ terms form nonoverlapping intervals. Hence

$$y\ f_Y(y)\Delta y = \sum_i y\ f_X(x_i)\Delta x_i = \sum_i g(x_i) f_X(x_i)\Delta x_i, \qquad \text{(6-12)}$$

and hence as Δy covers the entire y-axis, the corresponding Δx's are nonoverlapping, and they cover the entire x-axis. Hence, in the limit as $\Delta y \to 0$, integrating both sides of (6-12), we get the useful formula

$$E(Y) = E\big(g(X)\big) = \int_{-\infty}^{+\infty} y\ f_Y(y)dy = \int_{-\infty}^{+\infty} g(x) f_X(x)dx. \qquad \text{(6-13)}$$

In the discrete case, (6-13) reduces to

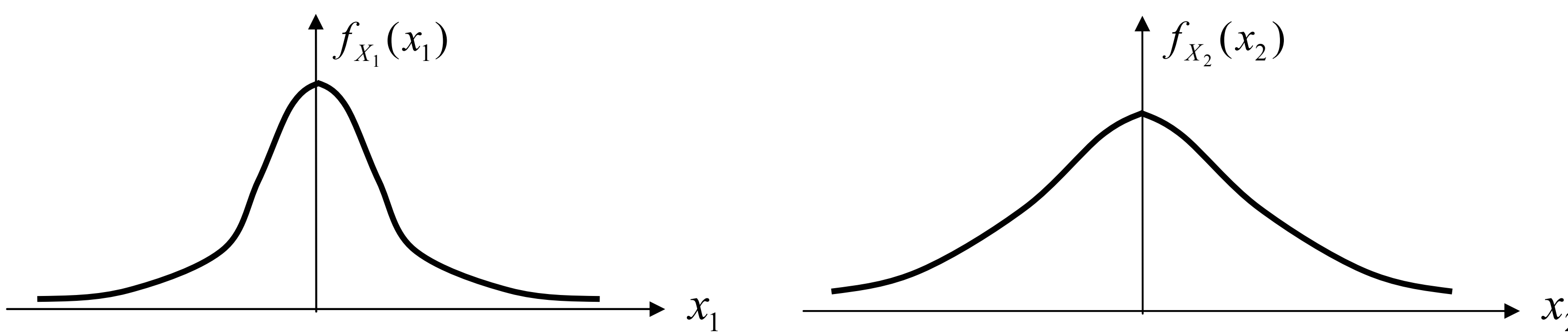
$$E(Y) = \sum_i g(x_i)P(X = x_i). \qquad \text{(6-14)}$$

From (6-13)-(6-14), $f_Y(y)$ is not required to evaluate $E(Y)$ for $Y = g(X)$. We can use (6-14) to determine the mean of $Y = X^2$, where X is a Poisson r.v. Using (3-43)

$$
\begin{aligned}
E\left(X^2\right) &= \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{i=0}^{\infty} (i+1) \frac{\lambda^{i+1}}{i!} \\
&= \lambda e^{-\lambda} \left(\sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} + \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \right) = \lambda e^{-\lambda} \left(\sum_{i=1}^{\infty} i \frac{\lambda^i}{i!} + e^{\lambda} \right) \\
&= \lambda e^{-\lambda} \left(\sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} + e^{\lambda} \right) = \lambda e^{-\lambda} \left(\sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} + e^{\lambda} \right) \\
&= \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e^{\lambda} \right) = \lambda^2 + \lambda. \qquad (6\text{-}15)
\end{aligned}
$$

In general, $E\left(X^k\right)$ is known as the *k*th moment of r.v X. Thus if $X \sim P(\lambda)$, its second moment is given by (6-15).

Mean alone will not be able to truly represent the p.d.f of any r.v. To illustrate this, consider the following scenario: Consider two Gaussian r.vs $X_1 \sim N(0,1)$ and $X_2 \sim N(0,10)$. Both of them have the same mean $\mu = 0$. However, as Fig. 6.1 shows, their p.d.fs are quite different. One is more concentrated around the mean, whereas the other one (X_2) has a wider spread. Clearly, we need atleast an additional parameter to measure this spread around the mean!



(a) $\sigma^2 = 1$ (b) $\sigma^2 = 10$

Fig.6.1

For a r.v X with mean μ, $X - \mu$ represents the deviation of the r.v from its mean. Since this deviation can be either positive or negative, consider the quantity $(X - \mu)^2$, and its average value $E[(X - \mu)^2]$ represents the average mean square deviation of X around its mean. Define

$$\sigma_X^2 \stackrel{\Delta}{=} E[(X - \mu)^2] > 0. \tag{6-16}$$

With $g(X) = (X - \mu)^2$ and using (6-13) we get

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x) dx > 0. \tag{6-17}$$

σ_X^2 is known as the variance of the r.v X, and its square root $\sigma_X = \sqrt{E(X - \mu)^2}$ is known as the standard deviation of X. Note that the standard deviation represents the root mean square spread of the r.v X around its mean μ.

Expanding (6-17) and using the linearity of the integrals, we get

$$Var(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} \left(x^2 - 2x\mu + \mu^2\right) f_X(x)dx$$

$$= \int_{-\infty}^{+\infty} x^2 f_X(x)dx - 2\mu \int_{-\infty}^{+\infty} x\, f_X(x)dx + \mu^2$$

$$= E\left(X^2\right) - \mu^2 = E\left(X^2\right) - \left[E(X)\right]^2 = \overline{X^2} - \overline{X}^2. \quad (6\text{-}18)$$

Alternatively, we can use (6-18) to compute σ_X^2.

Thus , for example, returning back to the Poisson r.v in (3-43), using (6-6) and (6-15), we get

$$\sigma_X^2 = \overline{X^2} - \overline{X}^2 = \left(\lambda^2 + \lambda\right) - \lambda^2 = \lambda. \quad (6\text{-}19)$$

Thus for a Poisson r.v, mean and variance are both equal to its parameter λ.

To determine the variance of the normal r.v $N(\mu, \sigma^2)$, we can use (6-16). Thus from (3-29)

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx. \qquad (6\text{-}20)$$

To simplify (6-20), we can make use of the identity

$$\int_{-\infty}^{+\infty} f_X(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = 1$$

for a normal p.d.f. This gives

$$\int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}\,\sigma. \qquad (6\text{-}21)$$

Differentiating both sides of (6-21) with respect to σ, we get

$$\int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma^3} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}$$

or

$$\int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = \sigma^2, \qquad (6\text{-}22)$$

which represents the $Var(X)$ in (6-20). Thus for a normal r.v as in (3-29)

$$Var(X) = \sigma^2 \tag{6-23}$$

and the second parameter in $N(\mu, \sigma^2)$ infact represents the variance of the Gaussian r.v. As Fig. 6.1 shows the larger the σ, the larger the spread of the p.d.f around its mean. Thus as the variance of a r.v tends to zero, it will begin to concentrate more and more around the mean ultimately behaving like a constant.

Moments: As remarked earlier, in general

$$m_n = \overline{X^n} = E(X^n), \quad n \geq 1 \tag{6-24}$$

are known as the moments of the r.v X, and

$$\mu_n = E[(X-\mu)^n] \tag{6-25}$$

are known as the central moments of X. Clearly, the mean $\mu = m_1$, and the variance $\sigma^2 = \mu_2$. It is easy to relate m_n and μ_n. Infact

$$\mu_n = E[(X-\mu)^n] = E\left(\sum_{k=0}^{n}\binom{n}{k} X^k (-\mu)^{n-k}\right)$$

$$= \sum_{k=0}^{n}\binom{n}{k} E\left(X^k\right)(-\mu)^{n-k} = \sum_{k=0}^{n}\binom{n}{k} m_k (-\mu)^{n-k}. \tag{6-26}$$

In general, the quantities

$$E[(X-a)^n] \tag{6-27}$$

are known as the generalized moments of X about a, and

$$E[|X|^n] \tag{6-28}$$

are known as the absolute moments of X.

For example, if $X \sim N(0, \sigma^2)$, then it can be shown that

$$E(X^n) = \begin{cases} 0, & n \text{ odd}, \\ 1 \cdot 3 \cdots (n-1)\sigma^n, & n \text{ even}. \end{cases} \tag{6-29}$$

$$E(|X|^n) = \begin{cases} 1 \cdot 3 \cdots (n-1)\sigma^n, & n \text{ even}, \\ 2^k k! \sigma^{2k+1} \sqrt{2/\pi}, & n = (2k+1), \text{ odd}. \end{cases} \tag{6-30}$$

Direct use of (6-2), (6-13) or (6-14) is often a tedious procedure to compute the mean and variance, and in this context, the notion of the characteristic function can be quite helpful.

Characteristic Function

The characteristic function of a r.v X is defined as

$$\Phi_X(\omega) \stackrel{\Delta}{=} E\left(e^{jX\omega}\right) = \int_{-\infty}^{+\infty} e^{jx\omega} f_X(x)dx. \qquad (6\text{-}31)$$

Thus $\Phi_X(0) = 1,$ and $\left|\Phi_X(\omega)\right| \le 1$ for all $\omega.$

For discrete r.vs the characteristic function reduces to

$$\Phi_X(\omega) = \sum_k e^{jk\omega} P(X = k). \qquad (6\text{-}32)$$

Thus for example, if $X \sim P(\lambda)$ as in (3-43), then its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} e^{jk\omega} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{j\omega})^k}{k!} = e^{-\lambda} e^{\lambda e^{j\omega}} = e^{\lambda(e^{j\omega}-1)}. \qquad (6\text{-}33)$$

Similarly, if X is a binomial r.v as in (3-42), its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{n} e^{jk\omega} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^{j\omega})^k q^{n-k} = (pe^{j\omega} + q)^n. \qquad (6\text{-}34)$$

To illustrate the usefulness of the characteristic function of a r.v in computing its moments, first it is necessary to derive the relationship between them. Towards this, from (6-31)

$$\Phi_X(\omega) = E\left(e^{jX\omega}\right) = E\left[\sum_{k=0}^{\infty} \frac{(j\omega X)^k}{k!}\right] = \sum_{k=0}^{\infty} j^k \frac{E(X^k)}{k!}\omega^k$$

$$= 1 + jE(X)\omega + j^2 \frac{E(X^2)}{2!}\omega^2 + \cdots + j^k \frac{E(X^k)}{k!}\omega^k + \cdots. \qquad (6\text{-}35)$$

Taking the first derivative of (6-35) with respect to ω, and letting it to be equal to zero, we get

$$\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0} = jE(X) \quad \text{or} \quad E(X) = \frac{1}{j}\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0}. \qquad (6\text{-}36)$$

Similarly, the second derivative of (6-35) gives

$$E(X^2) = \frac{1}{j^2}\left.\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2}\right|_{\omega=0}, \qquad (6\text{-}37)$$

and repeating this procedure k times, we obtain the kth moment of X to be

$$E(X^k) = \frac{1}{j^k} \frac{\partial^k \Phi_X(\omega)}{\partial \omega^k}\bigg|_{\omega=0}, \quad k \geq 1. \tag{6-38}$$

We can use (6-36)-(6-38) to compute the mean, variance and other higher order moments of any r.v X. For example, if $X \sim P(\lambda)$, then from (6-33)

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = e^{-\lambda} e^{\lambda e^{j\omega}} \lambda j e^{j\omega}, \tag{6-39}$$

so that from (6-36)

$$E(X) = \lambda, \tag{6-40}$$

which agrees with (6-6). Differentiating (6-39) one more time, we get

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = e^{-\lambda}\left(e^{\lambda e^{j\omega}}(\lambda j e^{j\omega})^2 + e^{\lambda e^{j\omega}} \lambda j^2 e^{j\omega}\right), \tag{6-41}$$

so that from (6-37)

$$E(X^2) = \lambda^2 + \lambda, \tag{6-42}$$

which again agrees with (6-15). Notice that compared to the tedious calculations in (6-6) and (6-15), the efforts involved in (6-39) and (6-41) are very minimal.

We can use the characteristic function of the binomial r.v $B(n, p)$ in (6-34) to obtain its variance. Direct differentiation of (6-34) gives

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = jnpe^{j\omega}(pe^{j\omega} + q)^{n-1} \tag{6-43}$$

so that from (6-36), $E(X) = np$ as in (6-7).

One more differentiation of (6-43) yields

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = j^2 np\left(e^{j\omega}(pe^{j\omega}+q)^{n-1} + (n-1)pe^{j2\omega}(pe^{j\omega}+q)^{n-2}\right) \qquad (6\text{-}44)$$

and using (6-37), we obtain the second moment of the binomial r.v to be

$$E(X^2) = np\left(1+(n-1)p\right) = n^2p^2 + npq. \qquad (6\text{-}45)$$

Together with (6-7), (6-18) and (6-45), we obtain the variance of the binomial r.v to be

$$\sigma_X^2 = E(X^2) - \left[E(X)\right]^2 = n^2p^2 + npq - n^2p^2 = npq. \qquad (6\text{-}46)$$

To obtain the characteristic function of the Gaussian r.v, we can make use of (6-31). Thus if $X \sim N(\mu, \sigma^2)$, then

$$\Phi_X(\omega) = \int_{-\infty}^{+\infty} e^{j\omega x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \quad (\text{Let } x - \mu = y)$$

$$= e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{j\omega y} e^{-y^2/2\sigma^2} dy = e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-y/2\sigma^2 (y - j2\sigma^2\omega)} dy$$

$$(\text{Let } y - j\sigma^2\omega = u \text{ so that } y = u + j\sigma^2\omega)$$

$$= e^{j\mu\omega} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-(u+j\sigma^2\omega)(u-j\sigma^2\omega)/2\sigma^2} du$$

$$= e^{j\mu\omega} e^{-\sigma^2\omega^2/2} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-u^2/2\sigma^2} du = e^{(j\mu\omega - \sigma^2\omega^2/2)}. \tag{6-47}$$

Notice that the characteristic function of a Gaussian r.v itself has the "Gaussian" bell shape. Thus if $X \sim N(0, \sigma^2)$, then

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}, \tag{6-48}$$

and

$$\Phi_X(\omega) = e^{-\sigma^2\omega^2/2}. \tag{6-49}$$

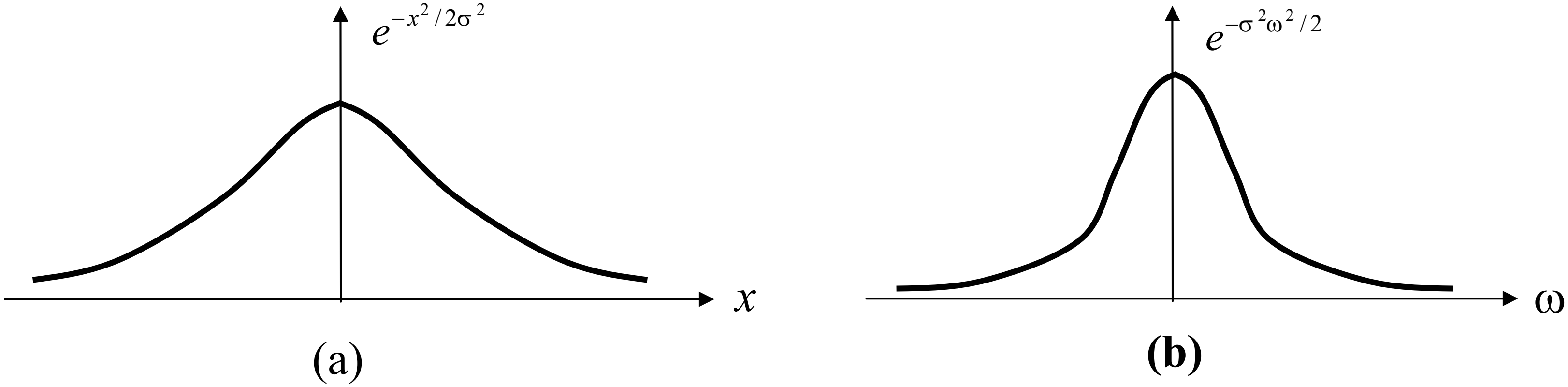


Fig. 6.2

From Fig. 6.2, the reverse roles of σ^2 in $f_X(x)$ and $\Phi_X(\omega)$ are noteworthy $(\sigma^2 \text{ vs } \frac{1}{\sigma^2})$.

In some cases, mean and variance may not exist. For example, consider the Cauchy r.v defined in (3-38). With

$$f_X(x) = \frac{(\alpha/\pi)}{\alpha^2 + x^2},$$

$$E(X^2) = \frac{\alpha}{\pi}\int_{-\infty}^{+\infty} \frac{x^2}{\alpha^2 + x^2}\,dx = \frac{\alpha}{\pi}\int_{-\infty}^{+\infty}\left(1 - \frac{\alpha^2}{\alpha^2 + x^2}\right)dx = \infty, \quad (6\text{-}50)$$

clearly diverges to infinity. Similarly

$$E(X) = \frac{\alpha}{\pi} \int_{-\infty}^{+\infty} \frac{x}{\alpha^2 + x^2} dx. \tag{6-51}$$

To compute (6-51), let us examine its one sided factor

$\int_0^{+\infty} \frac{x}{\alpha^2 + x^2} dx.$ With $x = \alpha \tan \theta$

$$\int_0^{+\infty} \frac{x}{\alpha^2 + x^2} dx = \int_0^{\pi/2} \frac{\alpha \tan \theta}{\alpha^2 \sec^2 \theta} \alpha \sec^2 \theta d\theta = \int_0^{\pi/2} \frac{\sin \theta}{\cos \theta} d\theta$$

$$= -\int_0^{\pi/2} \frac{d(\cos \theta)}{\cos \theta} = -\log \cos \theta \Big|_0^{\pi/2} = -\log \cos \frac{\pi}{2} = -\infty, \tag{6-52}$$

indicating that the double sided integral in (6-51) does not converge and is undefined. From (6-50)-(6-52), the mean and variance of a Cauchy r.v are undefined.

We conclude this section with a bound that estimates the dispersion of the r.v beyond a certain interval centered around its mean. Since σ^2 measures the dispersion of the r.v X around its mean μ , we expect this bound to depend on σ^2 as well.

Chebychev Inequality

Consider an interval of width 2ε symmetrically centered around its mean μ as in Fig. 6.3. What is the probability that X falls outside this interval? We need

$$P(|X-\mu| \geq \varepsilon) \; ? \tag{6-53}$$

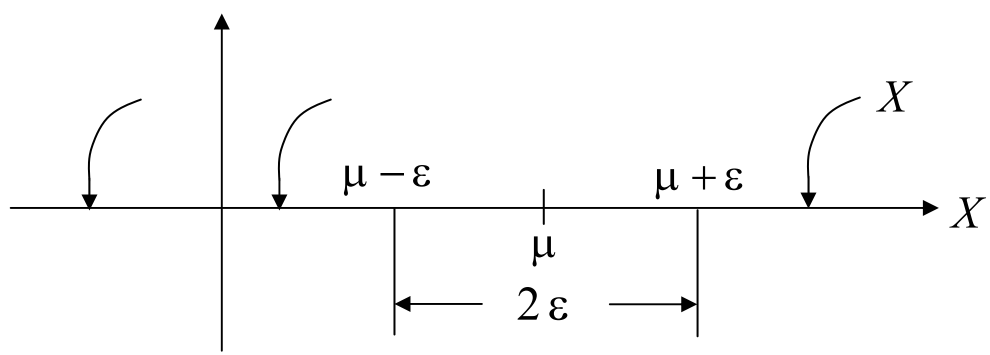


Fig. 6.3

To compute this probability, we can start with the definition of σ^2.

$$\sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)dx \geq \int_{|x-\mu|\geq\varepsilon} (x-\mu)^2 f_X(x)dx$$

$$\geq \int_{|x-\mu|\geq\varepsilon} \varepsilon^2 f_X(x)dx \geq \varepsilon^2 \int_{|x-\mu|\geq\varepsilon} f_X(x)dx \geq \varepsilon^2 P\left(|X-\mu| \geq \varepsilon\right). \quad (6\text{-}54)$$

From (6-54), we obtain the desired probability to be

$$P\left(|X-\mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}, \quad (6\text{-}55)$$

and (6-55) is known as the chebychev inequality. Interestingly, to compute the above probability bound the knowledge of $f_X(x)$ is not necessary. We only need σ^2, the variance of the r.v. In particular with $\varepsilon = k\sigma$ in (6-55) we obtain

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}. \quad (6\text{-}56)$$

Thus with $k = 3$, we get the probability of X being outside the 3σ interval around its mean to be 0.111 for any r.v. Obviously this cannot be a tight bound as it includes all r.vs. For example, in the case of a Gaussian r.v, from Table 4.1 $(\mu = 0, \sigma = 1)$

$$P\left(| X | \geq 3\sigma \right) = 0.0027 . \tag{6-57}$$

which is much tighter than that given by (6-56). Chebychev inequality always underestimates the exact probability.

Moment Identities :

Suppose X is a discrete random variable that takes only nonnegative integer values. i.e.,

$$P(X = k) = p_k \geq 0, \quad k = 0, 1, 2, \cdots$$

Then

$$\sum_{k=0}^{\infty} P(X > k) = \sum_{k=0}^{\infty} \sum_{i=k+1}^{\infty} P(X = i) = \sum_{i=1}^{\infty} P(X = i) \sum_{k=0}^{i-1} 1$$

$$= \sum_{i=0}^{\infty} i\, P(X = i) = E(X) \qquad \text{(6-58)}$$

similarly

$$\sum_{k=0}^{\infty} k\, P(X > k) = \sum_{i=1}^{\infty} P(X = i) \sum_{k=0}^{i-1} k = \sum_{i=1}^{\infty} \frac{i(i-1)}{2} P(X = i) = \frac{E\{X(X-1)\}}{2}$$

which gives

$$E(X^2) = \sum_{i=1}^{\infty} i^2 P(X=i) = \sum_{k=0}^{\infty} (2k+1) P(X>k). \qquad (6\text{-}59)$$

Equations (6-58) – (6-59) are at times quite useful in simplifying calculations. For example, referring to the Birthday Pairing Problem [Example 2-20., Text], let X represent the minimum number of people in a group for a birthday pair to occur. The probability that "the first n people selected from that group have different birthdays" is given by [$P(B)$ in page 39, Text]

$$p_n = \prod_{k=1}^{n-1} \left(1 - \frac{k}{N}\right) \approx e^{-n(n-1)/2N}.$$

But the event the "the first n people selected have

different birthdays" is the same as the event " $X > n$."
Hence

$$P(X > n) \approx e^{-n(n-1)/2N}.$$

Using (6-58), this gives the mean value of X to be

$$E(X) = \sum_{n=0}^{\infty} P(X > n) \approx \sum_{n=0}^{\infty} e^{-n(n-1)/2N} \approx \int_{-1/2}^{\infty} e^{-(x^2-1/4)/2N} dx$$

$$= e^{(1/8N)} \int_{-1/2}^{\infty} e^{-x^2/2N} dx = e^{(1/8N)} \left\{ \frac{1}{2}\sqrt{2\pi N} + \int_0^{1/2} e^{-x^2/2N} dx \right\}$$

$$\approx \sqrt{\pi N/2} + \frac{1}{2} = 24.44. \tag{6-60}$$

Similarly using (6-59) we get

$$E(X^2) = \sum_{n=0}^{\infty} (2n+1) P(X > n)$$

$$= \sum_{n=0}^{\infty} (2n+1) e^{-n(n-1)/2N} = \int_{-1/2}^{\infty} 2\,(x+1) e^{-(x^2-1/4)/2N} dx$$

$$= 2e^{(1/8N)} \left\{ \int_0^{\infty} x e^{-x^2/2N} dx + \int_0^{1/2} x e^{-x^2/2N} dx \right\} + 2 \int_{-1/2}^{\infty} e^{-(x^2-1/4)/2N} dx$$

$$= 2 \left\{ \frac{\sqrt{2\pi\, N}}{2} \sqrt{\frac{2}{\pi}} \sqrt{N} + \frac{1}{8} \right\} + 2E(X)$$

$$= 2N + \frac{1}{4} + \sqrt{2\pi\, N} + 1 = 2N + \sqrt{2\pi\, N} + \frac{5}{4}$$

$$= 779.139.$$

Thus

$$Var(X) = E(X^2) - (E(X))^2 = 181.82$$

which gives

$$\sigma_X \approx 13.48.$$

Since the standard deviation is quite high compared to the mean value, the actual number of people required for a birthday coincidence could be anywhere from 25 to 40.

Identities similar to (6-58)-(6-59) can be derived in the case of continuous random variables as well. For example, if X is a nonnegative random variable with density function $f_X(x)$ and distribution function $F_X(X)$, then

$$\begin{aligned} E\{X\} &= \int_0^\infty x f_X(x)dx = \int_0^\infty \left(\int_0^x dy \right) f_X(x)dx \\ &= \int_0^\infty \left(\int_y^\infty f_X(x)dx \right) dy = \int_0^\infty P(X > y)dy = \int_0^\infty P(X > x)dx \\ &= \int_0^\infty \{1 - F_X(x)\}dx = \int_0^\infty R(x)dx, \end{aligned} \qquad (6\text{-}61)$$

where

$$R(x) = 1 - F_X(x) \geq 0, \qquad x > 0.$$

Similarly

$$\begin{aligned} E\{X^2\} &= \int_0^\infty x^2 f_X(x)dx = \int_0^\infty \left(\int_0^x 2ydy \right) f_X(x)dx \\ &= 2\int_0^\infty \left(\int_y^\infty f_X(x)dx \right) ydy \\ &= 2\int_0^\infty x\,R(x)dx. \end{aligned}$$

A Baseball Trivia (Pete Rose and Dimaggio):

In 1978 Pete Rose set a national league record by hitting a string of 44 games during a 162 game baseball season. How unusual was that event?
As we shall see, that indeed was a rare event. In that context, we will answer the following question: What is the probability that someone in major league baseball will repeat that performance and possibly set a new record in the next 50 year period? The answer will put Pete Rose's accomplishment in the proper perspective.

Solution: As example 5-32 (Text) shows consecutive successes in n trials correspond to a run of length r in n

trials. From (5-133)-(5-134) text, we get the probability of r successive hits in n games to be

$$p_n = 1 - \alpha_{n,r} + p^r \alpha_{n-r,r} \tag{6-62}$$

where

$$\alpha_{n,r} = \sum_{k=0}^{\lfloor n/(r+1) \rfloor} \binom{n-kr}{k} (-1)^k (qp^r)^k \tag{6-63}$$

and p represents the probability of a hit in a game. Pete Rose's batting average is 0.303, and on the average since a batter shows up about four times/game, we get

$$\begin{aligned} p &= P(\text{at least one hit / game}) \\ &= 1 - \mathrm{P}(\text{no hit / game}) \\ &= 1 - (1 - 0.303)^4 = 0.76399 \end{aligned} \tag{6-64}$$

Substituting this value for p into the expressions (6-62)-(6-63) with $r = 44$ and $n = 162$, we can compute the desired probability p_n. However since n is quite large compared to r, the above formula is hopelessly time consuming in its implementation, and it is preferable to obtain a good approximation for p_n.

Towards this, notice that the corresponding moment generating function $\phi(z)$ for $q_n = 1 - p_n$ in Eq. (5-130) Text, is rational and hence it can be expanded in partial fraction as

$$\phi(z) = \frac{1 - p^r z^r}{1 - z + qp^r z^{r+1}} = \sum_{k=1}^{r} \frac{a_k}{z - z_k}, \tag{6-65}$$

where only r roots (out of $r+1$) are accounted for, since the root $z = 1/p$ is common to both the numerator and the denominator of $\phi(z)$. Here

$$a_k = \lim_{z \to z_k} \frac{(1 - p^r z^r)(z - z_k)}{1 - z + qp^r z^{r+1}}$$

$$= \lim_{z \to z_k} \frac{(1 - p^r z^r) - rp^r z^{r-1}(z - z_k)}{-1 + (r+1)qp^r z^r}$$

or

$$a_k = \frac{p^r z_k{}^r - 1}{1 - (r+1)qp^r z_k^r}, \qquad k = 1, 2, \cdots, r \tag{6-66}$$

From (6-65) – (6-66)

$$\phi(z) = \sum_{k=1}^{r} \frac{a_k}{(-z_k)} \; \frac{1}{1 - z / z_k} = \sum_{n=0}^{\infty} \underbrace{(\sum_{k=1}^{r} A_k z_k^{-(n+1)})}_{q_n} \, z^n \;\stackrel{\Delta}{=}\; \sum_{n=0}^{\infty} q_n z^n \tag{6-67}$$

where

$$A_k = -a_k = \frac{1 - p^r z_k^r}{1 - (r+1)qp^r z_k^r}$$

and

$$q_n = 1 - p_n = \sum_{k=1}^{r} A_k \ z_k^{-(n+1)}. \qquad (6\text{-}68)$$

However (fortunately), the roots $z_k, \ \ k = 1,2,\cdots,r$ in (6-65)-(6-67) are all not of the same importance (in terms of their relative magnitude with respect to unity). Notice that since for large *n*, $z_k^{-(n+1)} \to 0$ for $|z_k| > 1,$ only the roots nearest to unity contribute to (6-68) as n becomes larger.

To examine the nature of the roots of the denominator

$$A(z) = z - 1 - qp^r z^{r+1}$$

in (6-65), note that (refer to Fig 6.1) $A(0) = -1 < 0,$ $A(1) = -qp^r > A(0), \ A(1/p) = 0, \ A(\infty) < 0$ implying that for $z \geq 0, \ \ A(z)$ increases from –1 and reaches a positive maximum at z_0 given by

$$\left.\frac{dA(z)}{dz}\right|_{z-z_0} = 1 - qp^r(r+1)z_0^r = 0,$$

which gives

$$z_0^r = \frac{1}{qp^r(r+1)}. \tag{6-69}$$

There onwards $A(z)$ decreases to $-\infty$. Thus there are two positive roots for the equation $A(z) = 0$ given by $z_1 < z_0$ and $z_2 = 1/p > 1$. Since $A(1) = -qp^r \approx 0$ but negative, by continuity z_1 has the form $z_1 = 1 + \varepsilon$, $\varepsilon > 0$. (see Fig 6.1)

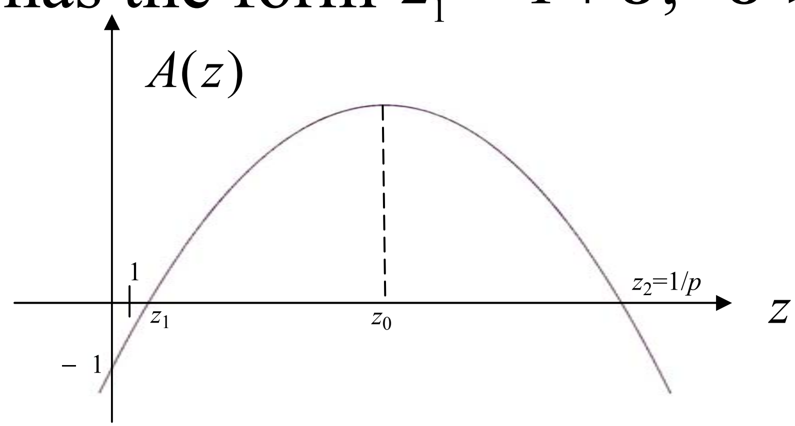


Fig 6.1 $A(z)$ for r odd

It is possible to obtain a bound for z_0 in (6-69). When P varies from 0 to 1, the maximum of $qp^r = (1-p)p^r$ is attained for $p = r/(r+1)$ and it equals $r^r/(r+1)^{r+1}$. Thus

$$qp^r \leq \frac{r^r}{(r+1)^{r+1}} \tag{6-70}$$

and hence substituting this into (6-69), we get

$$z_0 \geq \frac{r+1}{r} = 1 + \frac{1}{r}. \tag{6-71}$$

Hence it follows that the two positive roots of $A(z)$ satisfy

$$1 < z_1 < 1 + \frac{1}{r} < z_2 = \frac{1}{p} > 1. \tag{6-72}$$

Clearly, the remaining roots of $A(z)$ are complex if r is odd , and there is one negative root $-\alpha$ if r is even (see Fig 6.2). It is easy to show that the absolute value of *every* such complex or negative root is greater than $1/p > 1$.

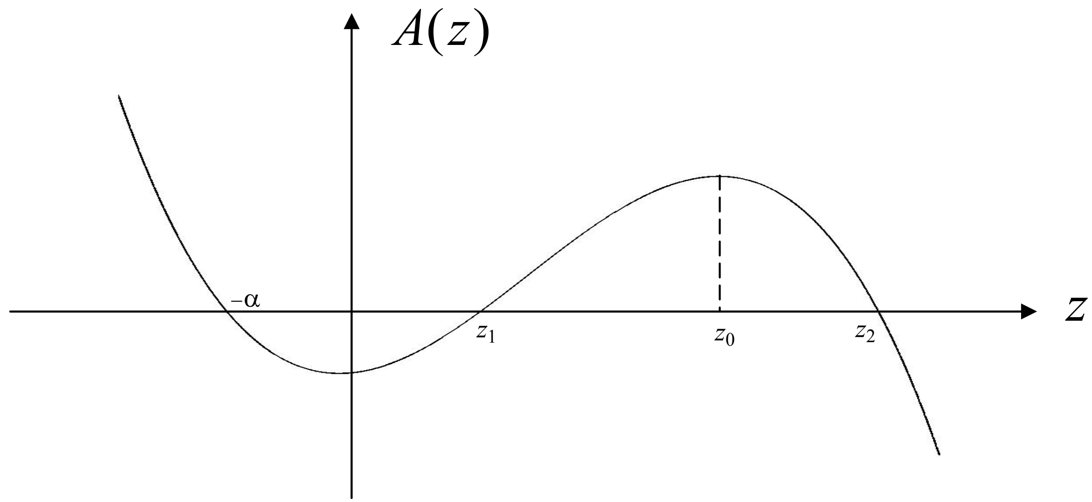


Fig 6.2 $A(z)$ for r even

To show this when r is even, suppose $-\alpha$ represents the negative root. Then

$$A(-\alpha) = -(\alpha + 1 - qp^r \alpha^{r+1}) = 0$$

so that the function

$$B(x) = x + 1 - qp^r x^{r+1} = A(x) + 2 \qquad (6\text{-}73)$$

starts positive, for $x > 0$ and increases till it reaches once again maximum at $z_0 \geq 1 + 1/r$ and then decreases to $-\infty$ through the root $x = \alpha > z_0 > 1$. Since $B(1/p) = 2$, we get $\alpha > 1/p > 1$, which proves our claim.

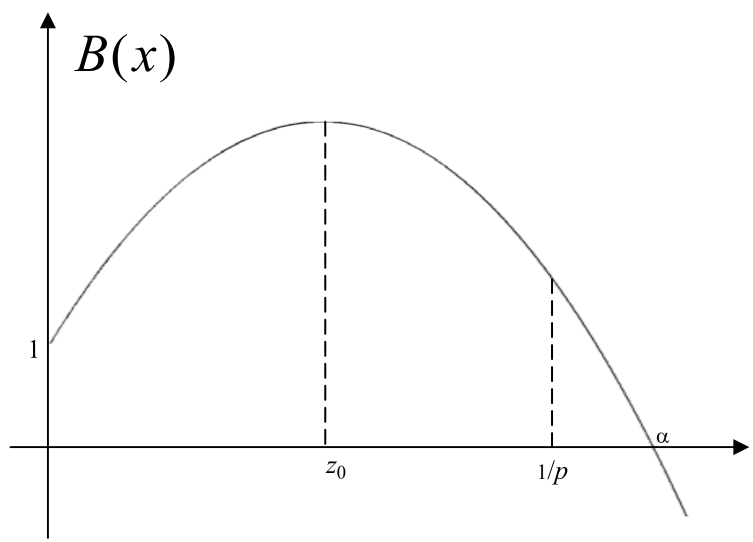


Fig 6.3 Negative root $B(\alpha) = 0$

PILLAI

Finally if $z = \rho\, e^{j\theta}$ is a complex root of $A(z)$, then

$$A(\rho\, e^{j\theta}) = \rho\, e^{j\theta} - 1 - qp^r \rho^{r+1} e^{j(r+1)\theta} = 0 \tag{6-74}$$

so that

$$\rho = |\, 1 + qp^r \rho^{r+1} e^{j(r+1)\theta} \,| \leq 1 + qp^r \rho^{r+1}$$

or

$$A(\rho) = \rho - 1 - qp^r \rho^{r+1} < 0.$$

Thus from (6-72), ρ belongs to either the interval $(0,\ z_1)$ or the interval $(\frac{1}{p}, \infty)$ in Fig 6.1. Moreover , by equating the imaginary parts in (6-74) we get

$$qp^r \rho^r \frac{\sin(r+1)\theta}{\sin\theta} = 1. \tag{6-75}$$

But

$$\left| \frac{\sin(r+1)\theta}{\sin\theta} \right| \le r+1, \tag{6-76}$$

equality being excluded if $\theta \neq 0$. Hence from (6-75)-(6-76) and (6-70)

$$(r+1)qp^r \rho^r > 1 \;\; \Rightarrow \;\; \rho^r > \frac{1}{(r+1)qp^r} = z_0^r > \left(\frac{r+1}{r} \right)^r$$

or

$$\rho > z_0 \ge 1 + \frac{1}{r}.$$

But $z_1 < z_0$. As a result ρ lies in the interval $(\frac{1}{p}, \infty)$ only. Thus

$$\rho > \frac{1}{p} > 1. \tag{6-77}$$

To summarize the two real roots of the polynomial $A(z)$ are given by

$$z_1 = 1 + \varepsilon, \quad \varepsilon > 0; \quad z_2 = \frac{1}{p} > 1, \tag{6-78}$$

and all other roots are (negative or complex) of the form

$$z_k = \rho\, e^{j\theta} \quad \text{where } \rho > \frac{1}{p} > 1. \tag{6-79}$$

Hence except for the first root z_1 (which is very close to unity), for all other roots

$$z_k^{-(n+1)} \;\to\; 0 \;\; \text{rapidly for all } k.$$

As a result, the most dominant term in (6-68) is the first term, and the contributions from all other terms to q_n in (6-68) can be bounded by

$$\left|\sum_{k=2}^{r} A_k z_k^{-(n+1)}\right| \le \sum_{k=2}^{r} |A_k||z_k|^{-(n+1)}$$

$$\le \sum_{k=2}^{r} \frac{1-(p|z_k|)^r}{1-(r+1)q(p|z_k|)^r}\, p^{n+1}$$

$$\le \sum_{k=2}^{r} \frac{(p|z_k|)^r}{(r+1)q(p|z_k|)^r}\, p^{n+1}$$

$$= \frac{r-1}{r+1}\frac{p^{n+1}}{q} \le \frac{p^{n+1}}{q} \;\rightarrow\; 0. \tag{6-80}$$

Thus from (6-68), to an excellent approximation

$$q_n = A_1 z_1^{-(n+1)}. \tag{6-81}$$

This gives the desired probability to be

$$p_n = 1 - q_n = 1 - \left(\frac{1-(pz_1)^r}{1-(r+1)q(pz_1)^r} \right) z_1^{-(n+1)}. \qquad (6\text{-}82)$$

Notice that since the dominant root z_1 is very close to unity, an excellent closed form approximation for z_1 can be obtained by considering the first order Taylor series expansion for $A(z)$. In the immediate neighborhood of $z=1$ we get

$$A(1+\varepsilon) = A(1) + A'(1)\varepsilon = -qp^r + (1-(r+1)qp^r)\varepsilon$$

so that $A(z_1) = A(1+\varepsilon) = 0$ gives

$$\varepsilon = \frac{qp^r}{1-(r+1)qp^r},$$

or

$$z_1 \approx 1 + \frac{qp^r}{1-(r+1)qp^r}. \tag{6-83}$$

Returning back to Pete Rose's case, $p = 0.763989$, $r = 44$ gives the smallest positive root of the denominator polynomial

$$A(z) = z - 1 - qp^{44} z^{45}$$

to be

$$z_1 = 1.00000169360549.$$

(The approximation (6-83) gives $z_1 = 1.00000169360548$).
Thus with n = 162 in (6-82) we get

$$p_{162} = 0.0002069970 \tag{6-84}$$

to be the probability for scoring 44 or more consecutive hits in 162 games for a player of Pete Rose's caliber – a very small probability indeed! In that sense it is a very rare event.

Assuming that during any baseball season there are on the average about $2 \times 25 = 50$ (?) such players over all major league baseball teams, we obtain [use Lecture #2, Eqs.(2-3)-(2-6) for the independence of 50 players]

$$P_1 = 1 - (1 - p_{162})^{50} = 0.0102975349$$

to be the probability that one of those players will hit the desired event. If we consider a period of 50 years, then the probability of *some* player hitting 44 or more consecutive games during one of these game seasons turns out to be

$$1 - (1 - P_1)^{50} = 0.40401874. \qquad (6\text{-}85)$$

(We have once again used the independence of the 50 seasons.)

Thus Pete Rose's 44 hit performance has a 60-40 chance of survival for about 50 years.From (6-85), rare events do indeed occur. In other words, *some* unlikely event is likely to happen.
However, as (6-84) shows a *particular* unlikely event – such as Pete Rose hitting 44 games in a sequence – is indeed rare.

Table 6.1 lists p_{162} for various values of r. From there, every reasonable batter should be able to hit at least 10 to 12 consecutive games during every season!

r	p_n ; $n = 162$
44	0.000207
25	0.03928
20	0.14937
15	0.48933
10	0.95257

Table 6.1 Probability of r runs in n trials for p=0.76399.

As baseball fans well know, Dimaggio holds the record of consecutive game hitting streak at 56 games (1941). With a lifetime batting average of 0.325 for Dimaggio, the above calculations yield [use (6-64), (6-82)-(6-83)] the probability for that event to be

$$p_n = 0.0000504532. \qquad (6\text{-}86)$$

Even over a 100 year period, with an average of 50 excellent hitters / season, the probability is only

$$1-(1-P_0)^{100} = 0.2229669 \qquad (6\text{-}87)$$

(where $P_0 = 1-(1-p_n)^{50} = 0.00251954$) that someone will repeat or outdo Dimaggio's performance.Remember, 60 years have already passed by, and no one has done it yet!

7. Two Random Variables

In many experiments, the observations are expressible not as a single quantity, but as a family of quantities. For example to record the height and weight of each person in a community or the number of people and the total income in a family, we need two numbers.

Let X and Y denote two random variables (r.v) based on a probability model (Ω, F, P). Then

$$P\left(x_1 < X(\xi) \leq x_2\right) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx,$$

and

$$P\left(y_1 < Y(\xi) \leq y_2\right) = F_Y(y_2) - F_Y(y_1) = \int_{y_1}^{y_2} f_Y(y)\,dy.$$

What about the probability that the pair of r.vs (X,Y) belongs to an arbitrary region D? In other words, how does one estimate, for example, $P[(x_1 < X(\xi) \le x_2) \cap (y_1 < Y(\xi) \le y_2)] = ?$ Towards this, we define the joint probability distribution function of X and Y to be

$$\begin{aligned} F_{XY}(x, y) &= P[(X(\xi) \le x) \cap (Y(\xi) \le y)] \\ &= P(X \le x, Y \le y) \ge 0, \qquad (7\text{-}1) \end{aligned}$$

where x and y are arbitrary real numbers.

Properties

(i) $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0, \quad F_{XY}(+\infty, +\infty) = 1.$ (7-2)

since $(X(\xi) \le -\infty, Y(\xi) \le y) \subset (X(\xi) \le -\infty)$, we get

$F_{XY}(-\infty, y) \le P\big(X(\xi) \le -\infty\big) = 0.$ Similarly $\big(X(\xi) \le +\infty, Y(\xi) \le +\infty\big) = \Omega,$

we get $F_{XY}(\infty, \infty) = P(\Omega) = 1.$

(ii) $P\big(x_1 < X(\xi) \le x_2, Y(\xi) \le y\big) = F_{XY}(x_2, y) - F_{XY}(x_1, y).$ (7-3)

$P\big(X(\xi) \le x, y_1 < Y(\xi) \le y_2\big) = F_{XY}(x, y_2) - F_{XY}(x, y_1).$ (7-4)

To prove (7-3), we note that for $x_2 > x_1,$

$$\big(X(\xi) \le x_2, Y(\xi) \le y\big) = \big(X(\xi) \le x_1, Y(\xi) \le y\big) \cup \big(x_1 < X(\xi) \le x_2, Y(\xi) \le y\big)$$

and the mutually exclusive property of the events on the right side gives

$$P\big(X(\xi) \le x_2, Y(\xi) \le y\big) = P\big(X(\xi) \le x_1, Y(\xi) \le y\big) + P\big(x_1 < X(\xi) \le x_2, Y(\xi) \le y\big)$$

which proves (7-3). Similarly (7-4) follows.

(iii) $$P\left(x_1 < X(\xi) \le x_2, y_1 < Y(\xi) \le y_2\right) = F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1). \qquad (7\text{-}5)$$

This is the probability that (X,Y) belongs to the rectangle R_0 in Fig. 7.1. To prove (7-5), we can make use of the following identity involving mutually exclusive events on the right side.

$$\left(x_1 < X(\xi) \le x_2, Y(\xi) \le y_2\right) = \left(x_1 < X(\xi) \le x_2, Y(\xi) \le y_1\right) \cup \left(x_1 < X(\xi) \le x_2, y_1 < Y(\xi) \le y_2\right).$$

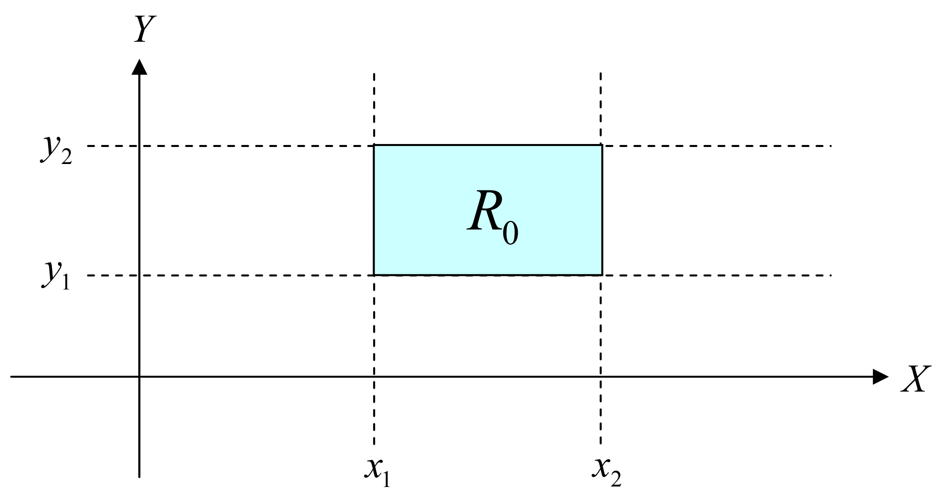


Fig. 7.1

This gives

$$P\big(x_1 < X(\xi) \le x_2, Y(\xi) \le y_2\big) = P\big(x_1 < X(\xi) \le x_2, Y(\xi) \le y_1\big) + P\big(x_1 < X(\xi) \le x_2, y_1 < Y(\xi) \le y_2\big)$$

and the desired result in (7-5) follows by making use of (7-3) with $y = y_2$ and y_1 respectively.

Joint probability density function (Joint p.d.f)

By definition, the joint p.d.f of X and Y is given by

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \, \partial y}. \tag{7-6}$$

and hence we obtain the useful formula

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v) \, du dv \, . \tag{7-7}$$

Using (7-2), we also get

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) \, dx dy = 1. \tag{7-8}$$

To find the probability that (X,Y) belongs to an arbitrary region D, we can make use of (7-5) and (7-7). From (7-5) and (7-7)

$$P\big(x < X(\xi) \le x + \Delta x, y < Y(\xi) \le y + \Delta y\big) = F_{XY}(x + \Delta x, y + \Delta y)$$
$$- F_{XY}(x, y + \Delta y) - F_{XY}(x + \Delta x, y) + F_{XY}(x, y)$$
$$= \int_x^{x+\Delta x} \int_y^{y+\Delta y} f_{XY}(u, v)\,du\,dv = f_{XY}(x, y)\Delta x \Delta y. \tag{7-9}$$

Thus the probability that (X,Y) belongs to a differential rectangle $\Delta x\ \Delta y$ equals $f_{XY}(x, y) \cdot \Delta x \Delta y,$ and repeating this procedure over the union of no overlapping differential rectangles in D, we get the useful result

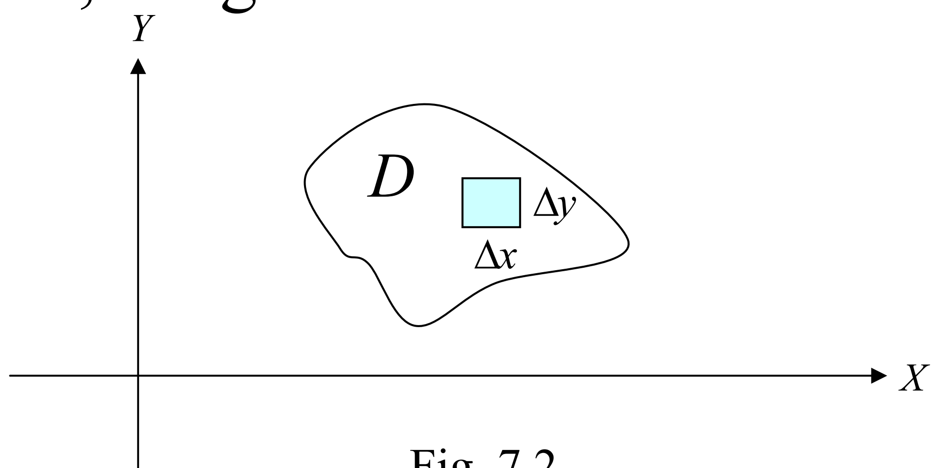


Fig. 7.2

$$P\big((X,Y)\in D\big)=\int\int_{(x,y)\in D} f_{XY}(x,y)\,dxdy\ . \tag{7-10}$$

(iv) Marginal Statistics

In the context of several r.vs, the statistics of each individual ones are called marginal statistics. Thus $F_X(x)$ is the marginal probability distribution function of X, and $f_X(x)$ is the marginal p.d.f of X. It is interesting to note that all marginals can be obtained from the joint p.d.f. In fact

$$F_X(x)=F_{XY}(x,+\infty),\quad F_Y(y)=F_{XY}(+\infty,y). \tag{7-11}$$

Also

$$f_X(x)=\int_{-\infty}^{+\infty} f_{XY}(x,y)dy,\quad f_Y(y)=\int_{-\infty}^{+\infty} f_{XY}(x,y)dx. \tag{7-12}$$

To prove (7-11), we can make use of the identity

$$(X\le x)=(X\le x)\cap(Y\le +\infty)$$

so that $F_X(x) = P(X \le x) = P(X \le x, Y \le \infty) = F_{XY}(x, +\infty).$
To prove (7-12), we can make use of (7-7) and (7-11), which gives

$$F_X(x) = F_{XY}(x, +\infty) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f_{XY}(u, y)\, du dy \tag{7-13}$$

and taking derivative with respect to *x* in (7-13), we get

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy. \tag{7-14}$$

At this point, it is useful to know the formula for differentiation under integrals. Let

$$H(x) = \int_{a(x)}^{b(x)} h(x, y) dy. \tag{7-15}$$

Then its derivative with respect to *x* is given by

$$\frac{dH(x)}{dx} = \frac{db(x)}{dx} h(x, b) - \frac{da(x)}{dx} h(x, a) + \int_{a(x)}^{b(x)} \frac{dh(x, y)}{dx} dy. \tag{7-16}$$

Obvious use of (7-16) in (7-13) gives (7-14).

If X and Y are discrete r.vs, then $p_{ij} \stackrel{\Delta}{=} P(X = x_i, Y = y_j)$ represents their joint p.d.f, and their respective marginal p.d.fs are given by

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij} \tag{7-17}$$

and

$$P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{ij} \tag{7-18}$$

Assuming that $P(X = x_i, Y = y_j)$ is written out in the form of a rectangular array, to obtain $P(X = x_i)$, from (7-17), one need to add up all entries in the i-th row.

It used to be a practice for insurance companies routinely to scribble out these sum values in the left and top margins, thus suggesting the name marginal densities! (Fig 7.3).

$$
\begin{array}{c|cccccc}
 & & & & \sum_i p_{ij} & & \\
\hline
 & p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1n} \\
 & p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2n} \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_j p_{ij} & p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{in} \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & p_{m1} & p_{m2} & \cdots & p_{mj} & \cdots & p_{mn}
\end{array}
$$

Fig. 7.3

From (7-11) and (7-12), the joint P.D.F and/or the joint p.d.f represent complete information about the r.vs, and their marginal p.d.fs can be evaluated from the joint p.d.f. However, given marginals, (most often) it will not be possible to compute the joint p.d.f. Consider the following example:

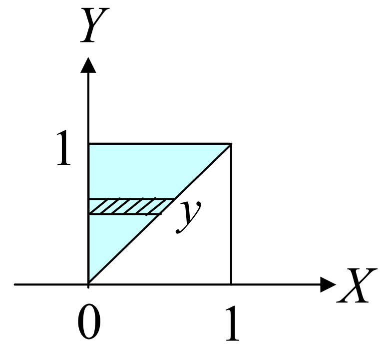


Fig. 7.4

Example 7.1: Given

$$f_{XY}(x,y) = \begin{cases} \text{constant}, & 0 < x < y < 1, \\ 0, & \text{otherwise} \end{cases}. \qquad (7\text{-}19)$$

Obtain the marginal p.d.fs $f_X(x)$ and $f_Y(y)$.

Solution: It is given that the joint p.d.f $f_{XY}(x,y)$ is a constant in the shaded region in Fig. 7.4. We can use (7-8) to determine that constant c. From (7-8)

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx\,dy = \int_{y=0}^{1}\left(\int_{x=0}^{y} c \cdot dx\right)dy = \int_{y=0}^{1} cy\,dy = \left.\frac{cy^2}{2}\right|_0^1 = \frac{c}{2} = 1. \quad (7\text{-}20)$$

Thus $c = 2$. Moreover from (7-14)

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dy = \int_{y=x}^{1} 2dy = 2(1-x), \quad 0 < x < 1, \tag{7-21}$$

and similarly

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dx = \int_{x=0}^{y} 2dx = 2y, \quad 0 < y < 1. \tag{7-22}$$

Clearly, in this case given $f_X(x)$ and $f_Y(y)$ as in (7-21)-(7-22), it will not be possible to obtain the original joint p.d.f in (7-19).

Example 7.2: *X* and *Y* are said to be jointly normal (Gaussian) distributed, if their joint p.d.f has the following form:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}, \tag{7-23}$$

$$-\infty < x < +\infty, \; -\infty < y < +\infty, \; |\rho| < 1.$$

By direct integration, using (7-14) and completing the square in (7-23), it can be shown that

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(x-\mu_X)^2/2\sigma_X^2} \sim N(\mu_X, \sigma_X^2), \qquad (7\text{-}24)$$

and similarly

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-(y-\mu_Y)^2/2\sigma_Y^2} \sim N(\mu_Y, \sigma_Y^2), \qquad (7\text{-}25)$$

Following the above notation, we will denote (7-23) as $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Once again, knowing the marginals in (7-24) and (7-25) alone doesn’t tell us everything about the joint p.d.f in (7-23).

As we show below, the only situation where the marginal p.d.fs can be used to recover the joint p.d.f is when the random variables are statistically independent.

Independence of r.vs

Definition: The random variables X and Y are said to be statistically independent if the events $\{X(\xi) \in A\}$ and $\{Y(\xi) \in B\}$ are independent events for any two Borel sets A and B in x and y axes respectively. Applying the above definition to the events $\{X(\xi) \leq x\}$ and $\{Y(\xi) \leq y\}$, we conclude that, if the r.vs X and Y are independent, then

$$P\big((X(\xi) \leq x) \cap (Y(\xi) \leq y)\big) = P(X(\xi) \leq x)P(Y(\xi) \leq y) \qquad (7\text{-}26)$$

i.e.,

$$F_{XY}(x, y) = F_X(x)F_Y(y) \qquad (7\text{-}27)$$

or equivalently, if X and Y are independent, then we must have

$$f_{XY}(x, y) = f_X(x) f_Y(y). \qquad (7\text{-}28)$$

If X and Y are discrete-type r.vs then their independence implies

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \text{for all } \ i, j. \qquad (7\text{-}29)$$

Equations (7-26)-(7-29) give us the procedure to test for independence. Given $f_{XY}(x, y)$, obtain the marginal p.d.fs $f_X(x)$ and $f_Y(y)$ and examine whether (7-28) or (7-29) is valid. If so, the r.vs are independent, otherwise they are dependent. Returning back to Example 7.1, from (7-19)-(7-22), we observe by direct verification that $f_{XY}(x, y) \neq f_X(x) f_Y(y)$. Hence X and Y are dependent r.vs in that case. It is easy to see that such is the case in the case of Example 7.2 also, unless $\rho = 0$. In other words, two jointly Gaussian r.vs as in (7-23) are independent if and only if the fifth parameter $\rho = 0$.

Example 7.3: Given

$$f_{XY}(x,y) = \begin{cases} xy^2 e^{-y}, & 0 < y < \infty, \quad 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{7-30}$$

Determine whether X and Y are independent.

Solution:

$$f_X(x) = \int_0^{+\infty} f_{XY}(x,y)dy = x\int_0^{\infty} y^2 e^{-y} dy$$

$$= x\left(-2ye^{-y}\Big|_0^{\infty} + 2\int_0^{\infty} ye^{-y}dy \right) = 2x, \quad 0 < x < 1. \tag{7-31}$$

Similarly

$$f_Y(y) = \int_0^1 f_{XY}(x,y)dx = \frac{y^2}{2}e^{-y}, \quad 0 < y < \infty. \tag{7-32}$$

In this case

$$f_{XY}(x,y) = f_X(x) f_Y(y),$$

and hence X and Y are independent random variables.

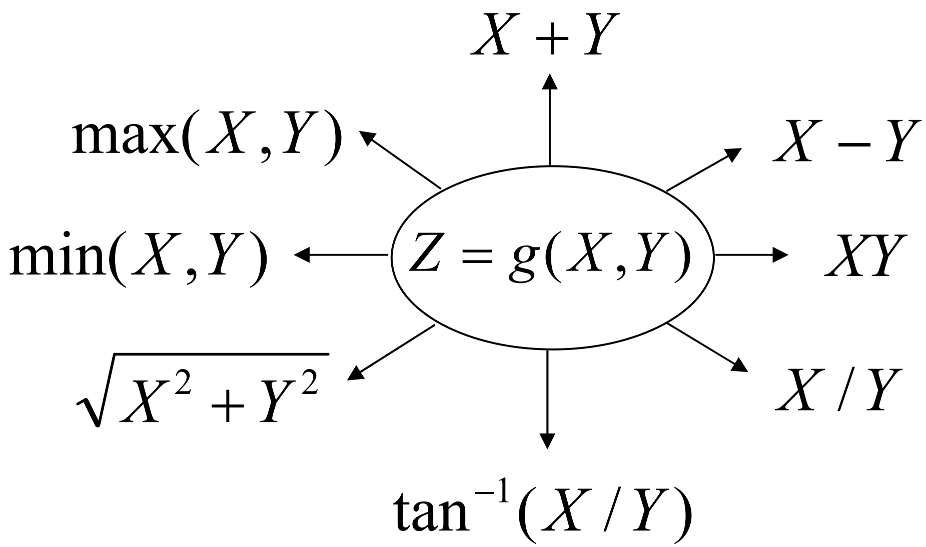
8. One Function of Two Random Variables

Given two random variables X and Y and a function $g(x,y)$, we form a new random variable Z as

$$Z = g(X, Y). \tag{8-1}$$

Given the joint p.d.f $f_{XY}(x, y)$, how does one obtain $f_Z(z)$, the p.d.f of Z? Problems of this type are of interest from a practical standpoint. For example, a receiver output signal usually consists of the desired signal buried in noise, and the above formulation in that case reduces to $Z = X + Y$.

It is important to know the statistics of the incoming signal for proper receiver design. In this context, we shall analyze problems of the following type:



(8-2)

Referring back to (8-1), to start with

$$F_Z(z) = P\big(Z(\xi) \le z\big) = P\big(g(X,Y) \le z\big) = P\big[(X,Y) \in D_z\big]$$

$$= \int\int_{x,y \in D_z} f_{XY}(x,y)\,dxdy\,, \qquad (8\text{-}3)$$

where D_z in the XY plane represents the region such that $g(x,y) \le z$ is satisfied. Note that D_z need not be simply connected (Fig. 8.1). From (8-3), to determine $F_Z(z)$ it is enough to find the region D_z for every z, and then evaluate the integral there.

We shall illustrate this method through various examples.

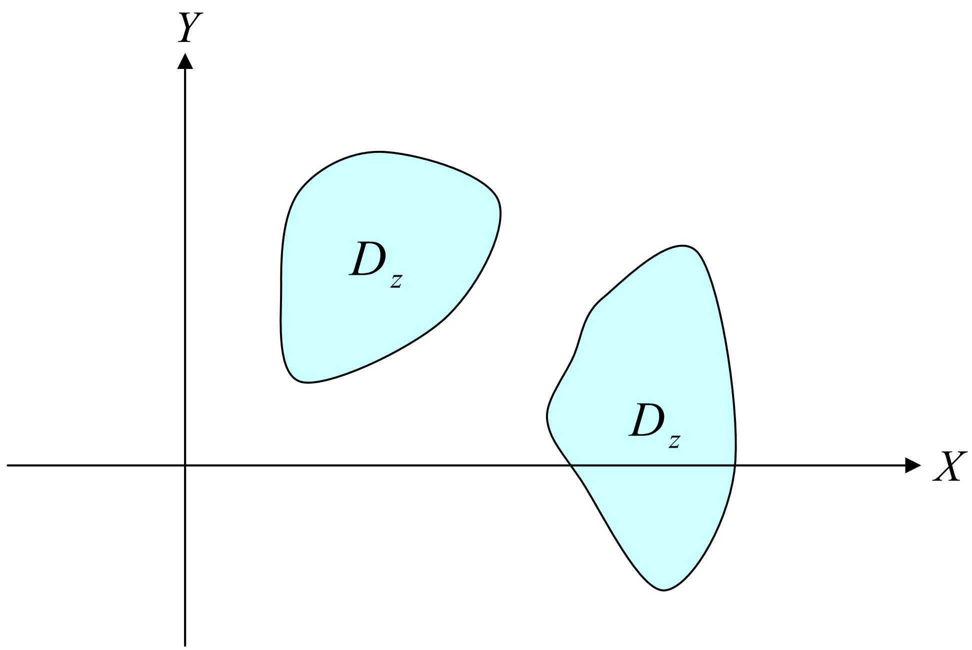


Fig. 8.1

Example 8.1: $Z = X + Y$. Find $f_Z(z)$.

Solution:

$$F_Z(z) = P\big(X + Y \le z\big) = \int_{y=-\infty}^{+\infty} \int_{x=-\infty}^{z-y} f_{XY}(x,y)\,dx\,dy, \tag{8-4}$$

since the region D_z of the *xy* plane where $x + y \le z$ is the shaded area in Fig. 8.2 to the left of the line $x + y = z$. Integrating over the horizontal strip along the *x*-axis first (inner integral) followed by sliding that strip along the *y*-axis from $-\infty$ to $+\infty$ (outer integral) we cover the entire shaded area.

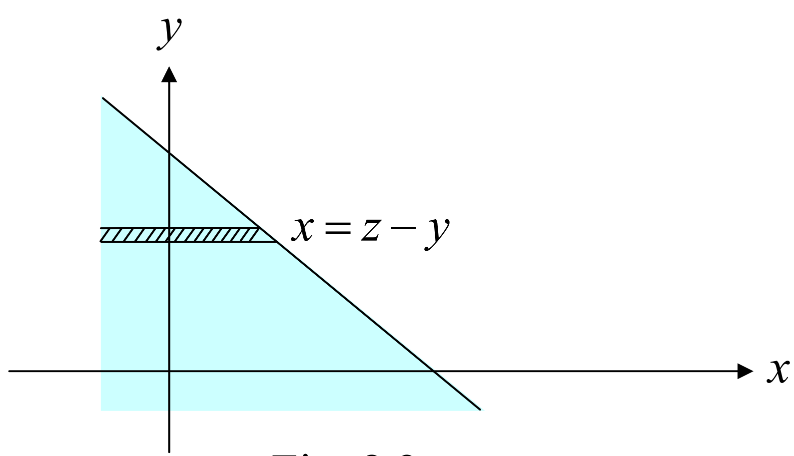


Fig. 8.2

We can find $f_Z(z)$ by differentiating $F_Z(z)$ directly. In this context, it is useful to recall the differentiation rule in (7-15) - (7-16) due to Leibnitz. Suppose

$$H(z) = \int_{a(z)}^{b(z)} h(x, z)\,dx. \tag{8-5}$$

Then

$$\frac{dH(z)}{dz} = \frac{db(z)}{dz} h\big(b(z), z\big) - \frac{da(z)}{dz} h\big(a(z), z\big) + \int_{a(z)}^{b(z)} \frac{\partial h(x, z)}{\partial z}\,dx. \tag{8-6}$$

Using (8-6) in (8-4) we get

$$f_Z(z) = \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial z} \int_{-\infty}^{z-y} f_{XY}(x, y)\,dx \right) dy = \int_{-\infty}^{+\infty} \left(1 \cdot f_{XY}(z-y, y) - 0 + \frac{\partial f_{XY}(x, y)}{\partial z} \right) dy$$

$$= \int_{-\infty}^{+\infty} f_{XY}(z-y, y)\,dy. \tag{8-7}$$

Alternatively, the integration in (8-4) can be carried out first along the y-axis followed by the x-axis as in Fig. 8.3.

In that case

$$F_Z(z) = \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{z-x} f_{XY}(x,y)\,dx\,dy, \qquad (8\text{-}8)$$

and differentiation of (8-8) gives

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{x=-\infty}^{+\infty} \left(\frac{\partial}{\partial z} \int_{y=-\infty}^{z-x} f_{XY}(x,y)\,dy \right) dx$$

$$= \int_{x=-\infty}^{+\infty} f_{XY}(x, z-x)\,dx. \qquad (8\text{-}9)$$

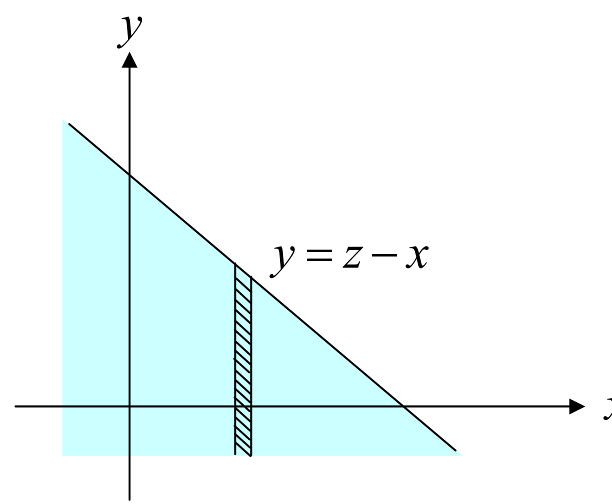


Fig. 8.3

If X and Y are independent, then

$$f_{XY}(x,y) = f_X(x) f_Y(y) \qquad (8\text{-}10)$$

and inserting (8-10) into (8-8) and (8-9), we get

$$f_Z(z) = \int_{y=-\infty}^{+\infty} f_X(z-y) f_Y(y)\,dy = \int_{x=-\infty}^{+\infty} f_X(x) f_Y(z-x)\,dx. \qquad (8\text{-}11)$$

The above integral is the standard convolution of the functions $f_X(z)$ and $f_Y(z)$ expressed two different ways. We thus reach the following conclusion: If two r.vs are independent, then the density of their sum equals the convolution of their density functions.

As a special case, suppose that $f_X(x) = 0$ for $x < 0$ and $f_Y(y) = 0$ for $y < 0,$ then we can make use of Fig. 8.4 to determine the new limits for $D_z.$

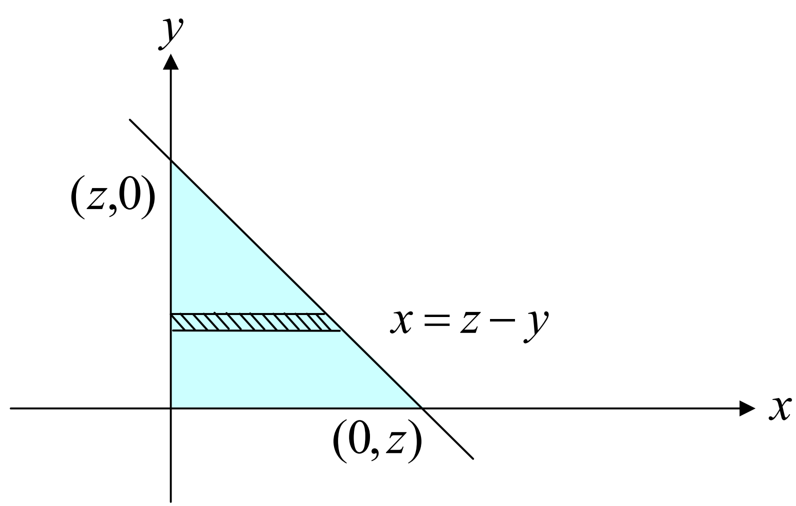


Fig. 8.4

In that case

$$F_Z(z) = \int_{y=0}^{z} \int_{x=0}^{z-y} f_{XY}(x,y)\,dx\,dy$$

or

$$f_Z(z) = \int_{y=0}^{z} \left(\frac{\partial}{\partial z} \int_{x=0}^{z-y} f_{XY}(x,y)\,dx \right) dy = \begin{cases} \int_0^z f_{XY}(z-y, y)\,dy, & z > 0, \\ 0, & z \le 0. \end{cases} \qquad (8\text{-}12)$$

On the other hand, by considering vertical strips first in Fig. 8.4, we get

$$F_Z(z) = \int_{x=0}^{z} \int_{y=0}^{z-x} f_{XY}(x,y)\,dy\,dx$$

or

$$f_Z(z) = \int_{x=0}^{z} f_{XY}(x, z-x)\,dx = \begin{cases} \int_{y=0}^{z} f_X(x) f_Y(z-x)\,dx, & z > 0, \\ 0, & z \le 0, \end{cases} \qquad (8\text{-}13)$$

if X and Y are independent random variables.

Example 8.2: Suppose X and Y are independent exponential r.vs with common parameter λ, and let $Z = X + Y$. Determine $f_Z(z)$.

Solution: We have $f_X(x) = \lambda e^{-\lambda x} U(x), \quad f_Y(y) = \lambda e^{-\lambda y} U(y), \qquad$ (8-14)

and we can make use of (13) to obtain the p.d.f of $Z = X + Y$.

$$f_Z(z) = \int_0^z \lambda^2 e^{-\lambda x} e^{-\lambda(z-x)} dx = \lambda^2 e^{-\lambda z} \int_0^z dx = z\lambda^2 e^{-\lambda z} U(z). \qquad (8\text{-}15)$$

As the next example shows, care should be taken in using the convolution formula for r.vs with finite range.

Example 8.3: X and Y are independent uniform r.vs in the common interval (0,1). Determine $f_Z(z)$, where $Z = X + Y$.

Solution: Clearly, $Z = X + Y \Rightarrow 0 < z < 2$ here, and as Fig. 8.5 shows there are two cases of z for which the shaded areas are quite different in shape and they should be considered separately.

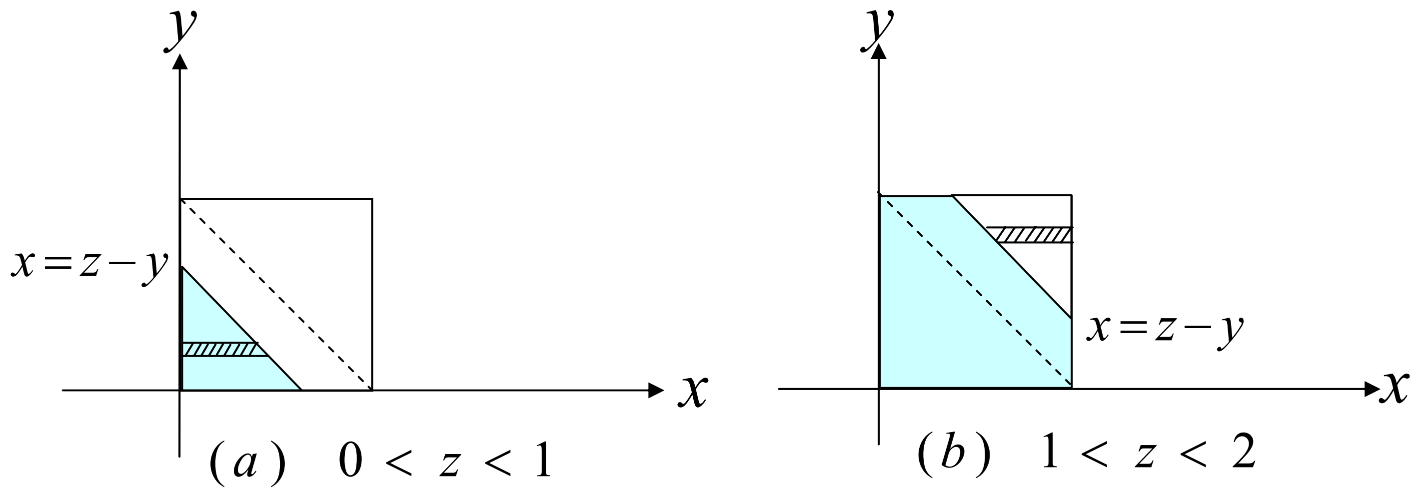


Fig. 8.5

For $0 \le z < 1$,

$$F_Z(z) = \int_{y=0}^{z} \int_{x=0}^{z-y} 1 \; dxdy = \int_{y=0}^{z} (z-y)dy = \frac{z^2}{2}, \quad 0 \le z < 1. \tag{8-16}$$

For $1 \le z < 2$, notice that it is easy to deal with the unshaded region. In that case

$$F_Z(z) = 1 - P\left(Z > z\right) = 1 - \int_{y=z-1}^{1} \int_{x=z-y}^{1} 1 \; dxdy$$

$$= 1 - \int_{y=z-1}^{1} (1 - z + y)dy = 1 - \frac{(2-z)^2}{2}, \quad 1 \le z < 2. \tag{8-17}$$

Using (8-16) - (8-17), we obtain

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \begin{cases} z & 0 \le z < 1, \\ 2 - z, & 1 \le z < 2. \end{cases} \tag{8-18}$$

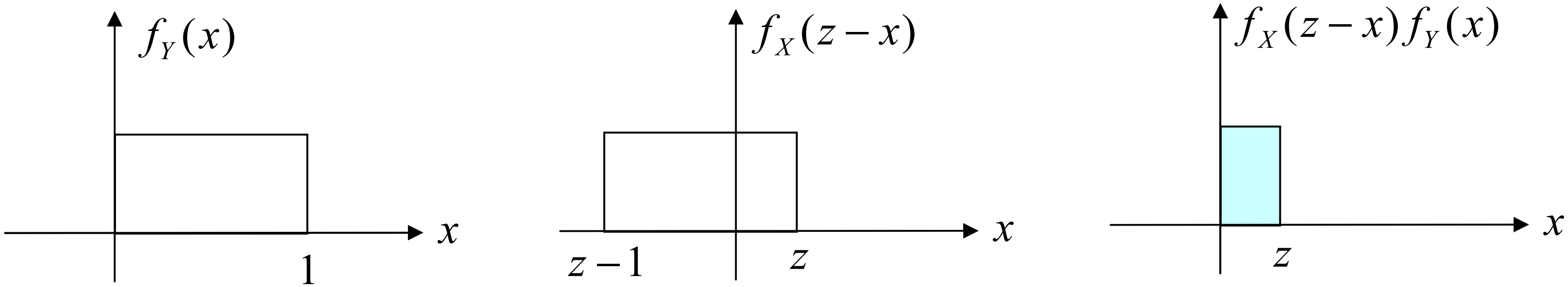
By direct convolution of $f_X(x)$ and $f_Y(y)$, we obtain the same result as above. In fact, for $0 \le z < 1$ (Fig. 8.6(a))

$$f_Z(z) = \int f_X(z-x) f_Y(x) dx = \int_0^z 1 \; dx = z. \tag{8-19}$$

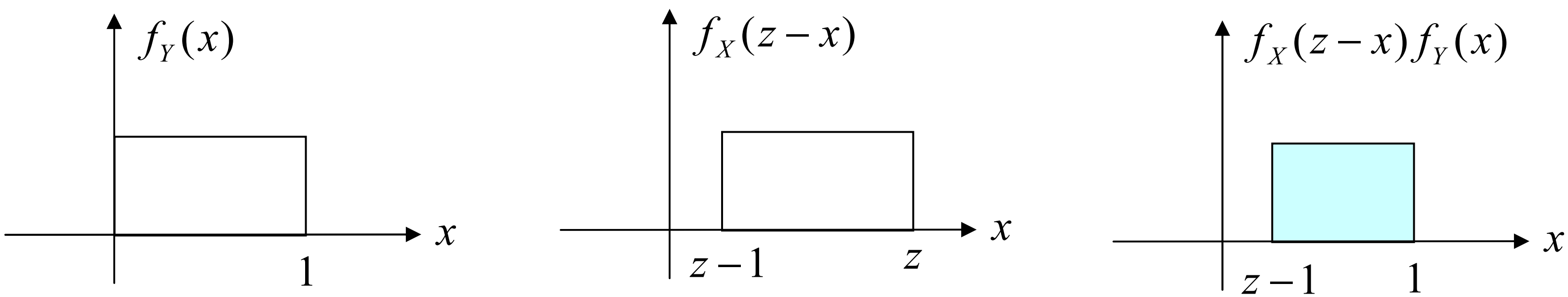
and for $1 \le z < 2$ (Fig. 8.6(b))

$$f_Z(z) = \int_{z-1}^{1} 1 \; dx = 2 - z. \tag{8-20}$$

Fig 8.6 (c) shows $f_Z(z)$ which agrees with the convolution of two rectangular waveforms as well.



$(a)\quad 0 \le z < 1$



$(b)\quad 1 \le z < 2$

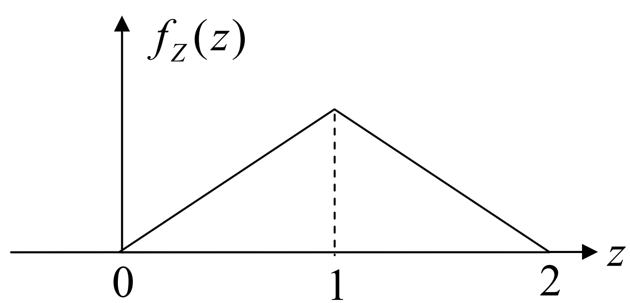


Fig. 8.6 (c)

Example 8.3: Let $Z = X - Y$. Determine its p.d.f $f_Z(z)$.

Solution: From (8-3) and Fig. 8.7

$$F_Z(z) = P(X - Y \le z) = \int_{y=-\infty}^{+\infty} \int_{x=-\infty}^{z+y} f_{XY}(x,y)dxdy$$

and hence

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{y=-\infty}^{+\infty} \left(\frac{\partial}{\partial z} \int_{x=-\infty}^{z+x} f_{XY}(x,y)dx \right) dy = \int_{-\infty}^{+\infty} f_{XY}(y+z,y)dy. \quad (8\text{-}21)$$

If X and Y are independent, then the above formula reduces to

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z+y) f_Y(y)dx = f_X(-z) \otimes f_Y(y), \quad (8\text{-}22)$$

which represents the convolution of $f_X(-z)$ with $f_Y(z)$.

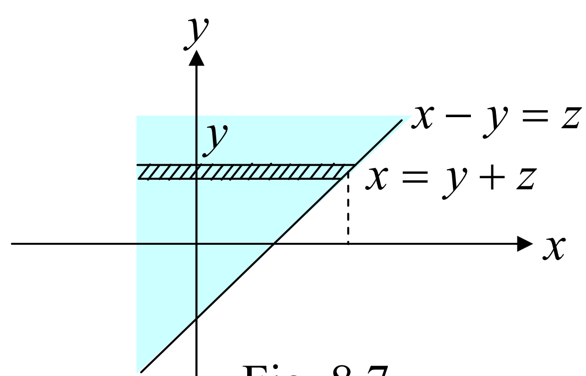


Fig. 8.7

As a special case, suppose

$$f_X(x)=0,\ x<0,\quad \text{and}\quad f_Y(y)=0,\ y<0.$$

In this case, Z can be negative as well as positive, and that gives rise to two situations that should be analyzed separately, since the region of integration for $z \ge 0$ and $z<0$ are quite different. For $z \ge 0$, from Fig. 8.8 (a)

$$F_Z(z)=\int_{y=0}^{+\infty}\int_{x=0}^{z+y} f_{XY}(x,y)dxdy$$

and for $z<0$, from Fig 8.8 (b)

$$F_Z(z)=\int_{y=-z}^{+\infty}\int_{x=0}^{z+y} f_{XY}(x,y)dxdy$$

After differentiation, this gives

$$f_Z(z)=\begin{cases}\int_0^{+\infty} f_{XY}(z+y,y)dy, & z\ge 0,\\ \int_{-z}^{+\infty} f_{XY}(z+y,y)dy, & z<0.\end{cases} \qquad (8\text{-}23)$$

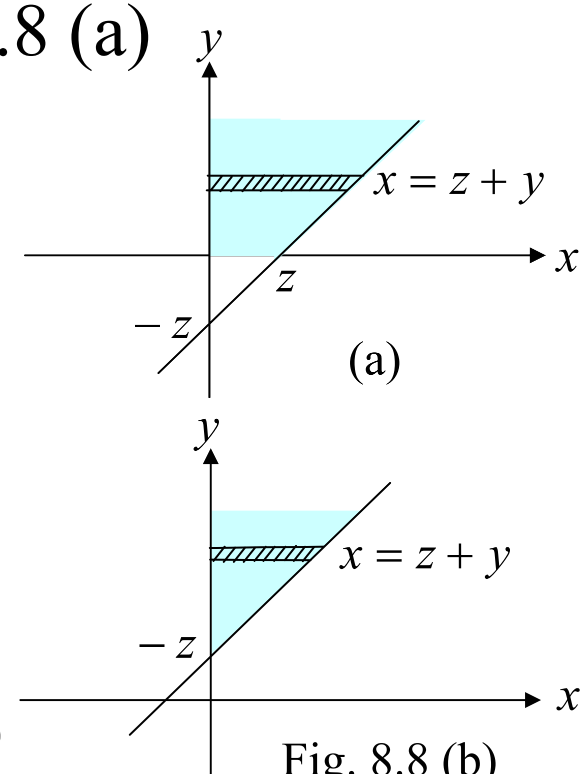


Fig. 8.8 (b)

Example 8.4: Given $Z = X / Y$, obtain its density function.
Solution: We have $F_Z(z) = P(X / Y \le z).$ (8-24)
The inequality $X / Y \le z$ can be rewritten as $X \le Yz$ if $Y > 0$, and $X \ge Yz$ if $Y < 0$. Hence the event $(X / Y \le z)$ in (24) need to be conditioned by the event $A = (Y > 0)$ and its compliment $\overline{A}$. Since $A \cup \overline{A} = S$, by the partition theorem, we have

$$\{X / Y \le z\} = \{(X / Y \le z) \cap (A \cup \overline{A})\} = \{(X / Y \le z) \cap A\} \cup \{(X / Y \le z) \cap \overline{A}\}$$

and hence by the mutually exclusive property of the later two events

$$\begin{aligned} P(X / Y \le z) &= P(X / Y \le z, Y > 0) + P(X / Y \le z, Y < 0) \\ &= P(X \le Yz, Y > 0) + P(X \ge Yz, Y < 0). \end{aligned} \qquad (8\text{-}25)$$

Fig. 8.9(a) shows the area corresponding to the first term, and Fig. 8.9(b) shows that corresponding to the second term in (8-25).

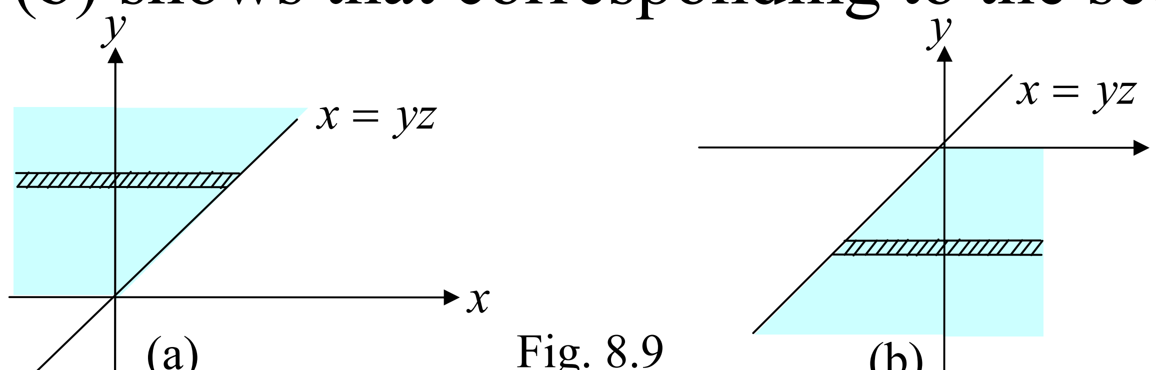


Fig. 8.9

Integrating over these two regions, we get

$$F_Z(z) = \int_{y=0}^{+\infty} \int_{x=-\infty}^{yz} f_{XY}(x, y)\,dx\,dy + \int_{y=-\infty}^{0} \int_{x=yz}^{\infty} f_{XY}(x, y)\,dx\,dy. \qquad (8\text{-}26)$$

Differentiation with respect to z gives

$$f_Z(z) = \int_0^{+\infty} y f_{XY}(yz, y)\,dy + \int_{-\infty}^{0} (-y) f_{XY}(yz, y)\,dy$$

$$= \int_{-\infty}^{+\infty} |y| f_{XY}(yz, y)\,dy, \qquad -\infty < z < +\infty. \qquad (8\text{-}27)$$

Note that if X and Y are nonnegative random variables, then the area of integration reduces to that shown in Fig. 8.10.

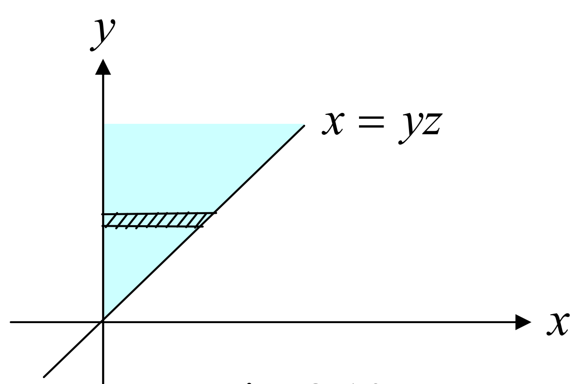


Fig. 8.10

This gives

$$F_Z(z) = \int_{y=0}^{\infty} \int_{x=0}^{yz} f_{XY}(x, y) dx dy$$

or

$$f_Z(z) = \begin{cases} \int_0^{+\infty} y \, f_{XY}(yz, y) dy, & z > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{8-28}$$

Example 8.5: X and Y are jointly normal random variables with zero mean so that

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)}. \tag{8-29}$$

Show that the ratio $Z = X / Y$ has a Cauchy density function centered at $r\sigma_1 / \sigma_2$.

Solution: Inserting (8-29) into (8-27) and using the fact that $f_{XY}(-x, -y) = f_{XY}(x, y)$, we obtain

$$f_Z(z) = \frac{2}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}} \int_0^{\infty} y e^{-y^2/2\sigma_0^2} dy = \frac{\sigma_0^2(z)}{\pi\sigma_1\sigma_2\sqrt{1 - r^2}},$$

where

$$\sigma_0^2(z) = \frac{1 - r^2}{\frac{z^2}{\sigma_1^2} - \frac{2rz}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}}.$$

Thus

$$f_Z(z) = \frac{\sigma_1\sigma_2\sqrt{1-r^2}/\pi}{\sigma_2^2(z - r\sigma_1/\sigma_2)^2 + \sigma_1^2(1-r^2)}, \tag{8-30}$$

which represents a Cauchy r.v centered at $r\sigma_1/\sigma_2$. Integrating (8-30) from $-\infty$ to z, we obtain the corresponding distribution function to be

$$F_Z(z) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{\sigma_2 z - r\sigma_1}{\sigma_1\sqrt{1-r^2}}. \tag{8-31}$$

Example 8.6: $Z = X^2 + Y^2$. Obtain $f_Z(z)$.

Solution: We have

$$F_Z(z) = P\left(X^2 + Y^2 \le z\right) = \int\int_{X^2+Y^2\le z} f_{XY}(x,y)\,dxdy. \tag{8-32}$$

But, $X^2 + Y^2 \le z$ represents the area of a circle with radius $\sqrt{z}$, and hence from Fig. 8.11,

$$F_Z(z) = \int_{y=-\sqrt{z}}^{\sqrt{z}} \int_{x=-\sqrt{z-y^2}}^{\sqrt{z-y^2}} f_{XY}(x,y)\,dx\,dy. \tag{8-33}$$

This gives after repeated differentiation

$$f_Z(z) = \int_{y=-\sqrt{z}}^{\sqrt{z}} \frac{1}{2\sqrt{z-y^2}} \left(f_{XY}(\sqrt{z-y^2}, y) + f_{XY}(-\sqrt{z-y^2}, y) \right) dy. \tag{8-34}$$

As an illustration, consider the next example.

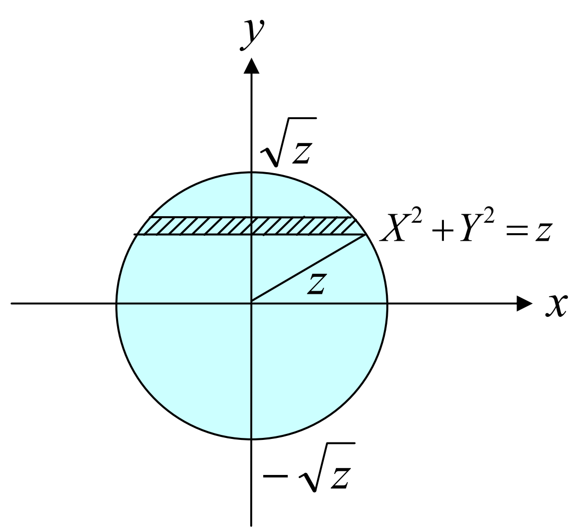


Fig. 8.11

Example 8.7 : X and Y are independent normal r.vs with zero Mean and common variance σ^2. Determine $f_Z(z)$ for $Z = X^2 + Y^2$.
Solution: Direct substitution of (8-29) with $r = 0, \ \ \sigma_1 = \sigma_2 = \sigma$ Into (8-34) gives

$$f_Z(z) = \int_{y=-\sqrt{z}}^{\sqrt{z}} \frac{1}{2\sqrt{z-y^2}} \left(2 \cdot \frac{1}{2\pi\sigma^2} e^{(z-y^2+y^2)/2\sigma^2} \right) dy = \frac{e^{-z/2\sigma^2}}{\pi\sigma^2} \int_0^{\sqrt{z}} \frac{1}{\sqrt{z-y^2}} dy$$

$$= \frac{e^{-z/2\sigma^2}}{\pi\sigma^2} \int_0^{\pi/2} \frac{\sqrt{z}\cos\theta}{\sqrt{z}\cos\theta} d\theta = \frac{1}{2\sigma^2} e^{-z/2\sigma^2} U(z), \tag{8-35}$$

where we have used the substitution $y = \sqrt{z}\sin\theta$. From (8-35) we have the following result: If X and Y are independent zero mean Gaussian r.vs with common variance σ^2, then $X^2 + Y^2$ is an exponential r.vs with parameter $2\sigma^2$.

Example 8.8 : Let $Z = \sqrt{X^2 + Y^2}$. Find $f_Z(z)$.
Solution: From Fig. 8.11, the present case corresponds to a circle with radius z^2. Thus

$$F_Z(z) = \int_{y=-z}^{z} \int_{x=-\sqrt{z^2-y^2}}^{\sqrt{z^2-y^2}} f_{XY}(x, y)\,dxdy \ .$$

And by repeated differentiation, we obtain

$$f_Z(z) = \int_{-z}^{z} \frac{z}{\sqrt{z^2 - y^2}} \Big(f_{XY}(\sqrt{z^2 - y^2}, y) + f_{XY}(-\sqrt{z^2 - y^2}, y) \Big) dy \ . \qquad (8\text{-}36)$$

Now suppose X and Y are independent Gaussian as in Example 8.7. In that case, (8-36) simplifies to

$$f_Z(z) = 2\int_0^z \frac{z}{\sqrt{z^2 - y^2}} \frac{1}{2\pi\sigma^2} e^{(z^2 - y^2 + y^2)/2\sigma^2} dy = \frac{2z}{\pi\sigma^2} e^{-z^2/2\sigma^2} \int_0^z \frac{1}{\sqrt{z^2 - y^2}} dy$$

$$= \frac{2z}{\pi\sigma^2} e^{-z^2/2\sigma^2} \int_0^{\pi/2} \frac{z\cos\theta}{z\cos\theta} d\theta = \frac{z}{\sigma^2} e^{-z^2/2\sigma^2} U(z), \qquad (8\text{-}37)$$

which represents a Rayleigh distribution. Thus, if $W = X + iY$, where X and Y are real, independent normal r.vs with zero mean and equal variance, then the r.v $|W| = \sqrt{X^2 + Y^2}$ has a Rayleigh density. W is said to be a complex Gaussian r.v with zero mean, whose real and imaginary parts are independent r.vs. From (8-37), we have seen that its magnitude has Rayleigh distribution.

What about its phase

$$\theta = \tan^{-1}\left(\frac{X}{Y}\right)? \tag{8-38}$$

Clearly, the principal value of θ lies in the interval $(-\pi/2, \pi/2)$. If we let $U = \tan\theta = X/Y$, then from example 8.5, U has a Cauchy distribution with (see (8-30) with $\sigma_1 = \sigma_2,\ r = 0$)

$$f_U(u) = \frac{1/\pi}{u^2+1}, \quad -\infty < u < \infty. \tag{8-39}$$

As a result

$$f_\theta(\theta) = \frac{1}{|d\theta/du|} f_U(\tan\theta) = \frac{1}{(1/\sec^2\theta)} \frac{1/\pi}{\tan^2\theta + 1} = \begin{cases} 1/\pi, & -\pi/2 < \theta < \pi/2, \\ 0, & \text{otherwise}. \end{cases} \tag{8-40}$$

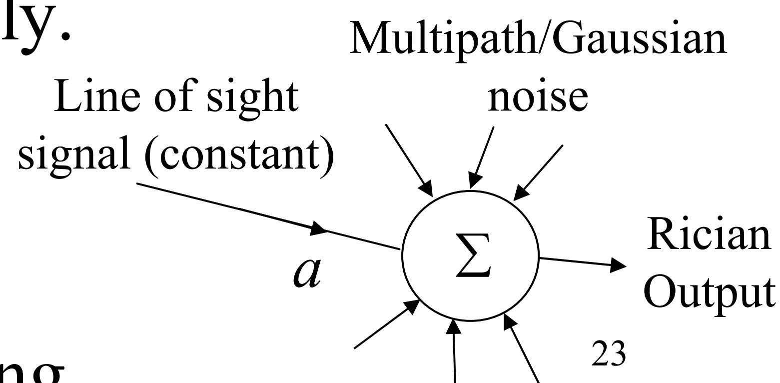
To summarize, the magnitude and phase of a zero mean complex Gaussian r.v has Rayleigh and uniform distributions respectively. Interestingly, as we will show later, these two derived r.vs are also independent of each other!

Let us reconsider example 8.8 where X and Y have nonzero means μ_X and μ_Y respectively. Then $Z=\sqrt{X^2+Y^2}$ is said to be a Rician r.v. Such a scene arises in fading multipath situation where there is a dominant constant component (mean) in addition to a zero mean Gaussian r.v. The constant component may be the line of sight signal and the zero mean Gaussian r.v part could be due to random multipath components adding up incoherently (see diagram below). The envelope of such a signal is said to have a Rician p.d.f.

Example 8.9: Redo example 8.8, where X and Y have nonzero means μ_X and μ_Y respectively.

Solution: Since

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma^2}e^{-[(x-\mu_X)^2+(y-\mu_Y)^2]/2\sigma^2},$$



substituting this into (8-36) and letting

$$y = z\sin\theta,\ \mu = \sqrt{\mu_X^2 + \mu_Y^2},\ \mu_X = \mu\cos\phi,\ \mu_Y = \mu\sin\phi,$$

we get the Rician probability density function to be

$$\begin{aligned}
f_Z(z) &= \frac{z e^{-(z^2+\mu^2)/2\sigma^2}}{2\pi\sigma^2}\int_{-\pi/2}^{\pi/2}\left(e^{z\mu\cos(\theta-\phi)/\sigma^2} + e^{-z\mu\cos(\theta+\phi)/\sigma^2}\right)d\theta \\
&= \frac{z e^{-(z^2+\mu^2)/2\sigma^2}}{2\pi\sigma^2}\left(\int_{-\pi/2}^{\pi/2} e^{z\mu\cos(\theta-\phi)/\sigma^2}\,d\theta + \int_{\pi/2}^{3\pi/2} e^{z\mu\cos(\theta-\phi)/\sigma^2}\,d\theta\right) \\
&= \frac{z e^{-(z^2+\mu^2)/2\sigma^2}}{2\pi\sigma^2} I_0\left(\frac{z\mu}{\sigma^2}\right), \qquad (8\text{-}41)
\end{aligned}$$

where

$$I_0(\eta) \triangleq \frac{1}{2\pi}\int_0^{2\pi} e^{\eta\cos(\theta-\phi)}d\theta = \frac{1}{\pi}\int_0^{\pi} e^{\eta\cos\theta}d\theta \qquad (8\text{-}42)$$

is the modified Bessel function of the first kind and zeroth order.

Example 8.10: $Z = \max(X, Y),\quad W = \min(X, Y).$ Determine $f_Z(z)$.

Solution: The functions *max* and *min* are nonlinear operators and represent special cases of the more general order statistics. In general, given any n-tuple $X_1, X_2, \cdots, X_n$, we can arrange them in an increasing order of magnitude such that

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}, \tag{8-43}$$

where $X_{(1)} = \min(X_1, X_2, \cdots, X_n)$, and $X_{(2)}$ is the second smallest value among $X_1, X_2, \cdots, X_n$, and finally $X_{(n)} = \max(X_1, X_2, \cdots, X_n)$. If $X_1, X_2, \cdots, X_n$ represent r.vs, the function $X_{(k)}$ that takes on the value $x_{(k)}$ in each possible sequence $(x_1, x_2, \cdots, x_n)$ is known as the k-th order statistic. $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ represent the set of order statistics among n random variables. In this context

$$R = X_{(n)} - X_{(1)} \tag{8-44}$$

represents the range, and when $n = 2$, we have the *max* and *min* statistics.

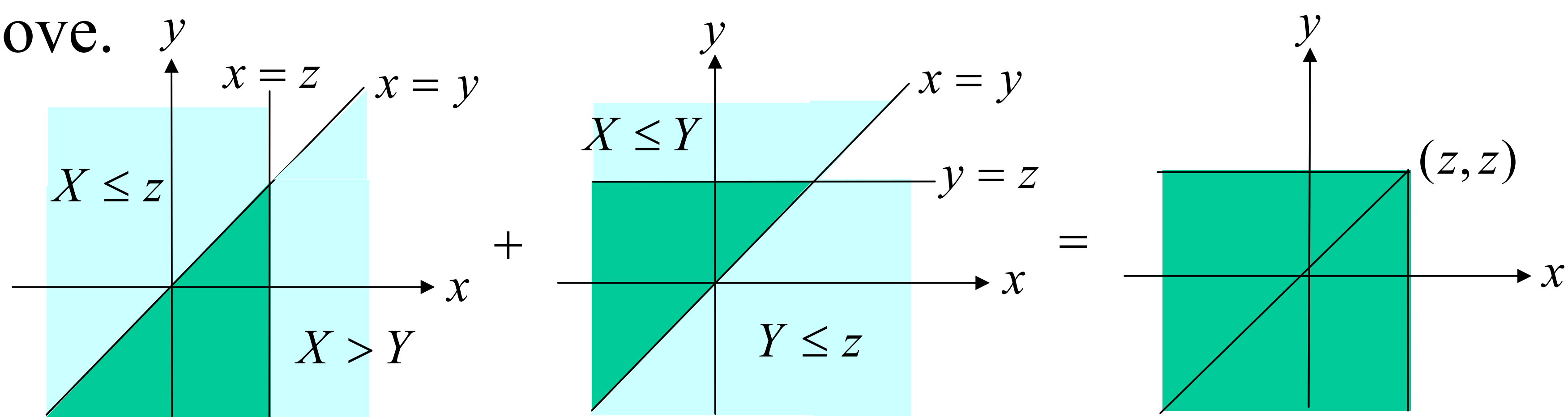
Returning back to that problem, since

$$Z = \max(X, Y) = \begin{cases} X, & X > Y, \\ Y, & X \le Y, \end{cases} \tag{8-45}$$

we have (see also (8-25))

$$\begin{aligned} F_Z(z) &= P(\max(X,Y) \le z) = P[(X \le z, X > Y) \cup (Y \le z, X \le Y)] \\ &= P(X \le z, X > Y) + P(Y \le z, X \le Y), \end{aligned}$$

since $(X > Y)$ and $(X \le Y)$ are mutually exclusive sets that form a partition. Figs 8.12 (a)-(b) show the regions satisfying the corresponding inequalities in each term above.



$(a)\ P(X \le z, X > Y)$ $(b)\ P(Y \le z, X \le Y)$ (c)

Fig. 8.12

Fig. 8.12 (c) represents the total region, and from there

$$F_Z(z) = P\big(X \le z, Y \le z\big) = F_{XY}(z, z). \tag{8-46}$$

If X and Y are independent, then

$$F_Z(z) = F_X(x) F_Y(y)$$

and hence

$$f_Z(z) = F_X(z) f_Y(z) + f_X(z) F_Y(z). \tag{8-47}$$

Similarly

$$W = \min(X, Y) = \begin{cases} Y, & X > Y, \\ X, & X \le Y. \end{cases} \tag{8-48}$$

Thus

$$F_W(w) = P\big(\min(X, Y) \le w\big) = P\big[\big(Y \le w, X > Y\big) \cup \big(X \le w, X \le Y\big)\big].$$

Once again, the shaded areas in Fig. 8.13 (a)-(b) show the regions satisfying the above inequalities and Fig 8.13 (c) shows the overall region.

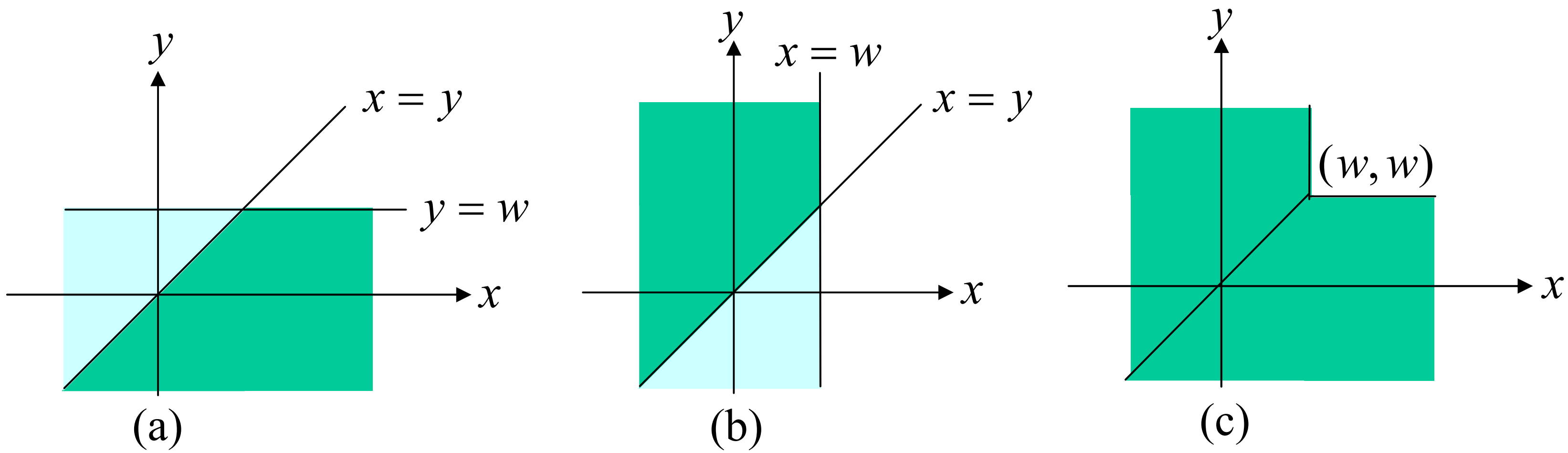


Fig. 8.13

From Fig. 8.13 (c),

$$F_W(w) = 1 - P(W > w) = 1 - P(X > w, Y > w)$$
$$= F_X(w) + F_Y(w) - F_{XY}(w, w), \tag{8-49}$$

where we have made use of (7-5) and (7-12) with $x_2 = y_2 = +\infty$, and $x_1 = y_1 = w$.

Example 8.11: Let X and Y be independent exponential r.vs with common parameter λ. Define $W = \min(X, Y)$. Find $f_W(w)$?
Solution: From (8-49)

$$F_W(w) = F_X(w) + F_Y(w) - F_X(w)F_Y(w)$$

and hence

$$f_W(w) = f_X(w) + f_Y(w) - f_X(w)F_Y(w) - F_X(w)f_Y(w).$$

But $f_X(w) = f_Y(w) = \lambda e^{-\lambda w}$, and $F_X(w) = F_Y(w) = 1 - e^{-\lambda w}$, so that

$$f_W(w) = 2\lambda e^{\lambda w} - 2(1 - e^{-\lambda w})\lambda e^{-\lambda w} = 2\lambda e^{-2\lambda w}U(w). \qquad (8\text{-}50)$$

Thus min (X, Y) is also exponential with parameter 2λ.

Example 8.12: Suppose X and Y are as give in the above example. Define $Z = \left[\min(X, Y) / \max(X, Y)\right]$. Determine $f_Z(z)$.

Solution: Although $\min(\cdot)/\max(\cdot)$ represents a complicated function, by partitioning the whole space as before, it is possible to simplify this function. In fact

$$Z = \begin{cases} X/Y, & X \le Y, \\ Y/X, & X > Y. \end{cases} \tag{8-51}$$

As before, this gives

$$F_Z(z) = P(Z \le z) = P(X/Y \le z, X \le Y) + P(Y/X \le z, X > Y)$$
$$= P(X \le Yz, X \le Y) + P(Y \le Xz, X > Y) \ . \tag{8-52}$$

Since X and Y are both positive random variables in this case, we have $0 < z < 1$. The shaded regions in Figs 8.14 (a)-(b) represent the two terms in the above sum.

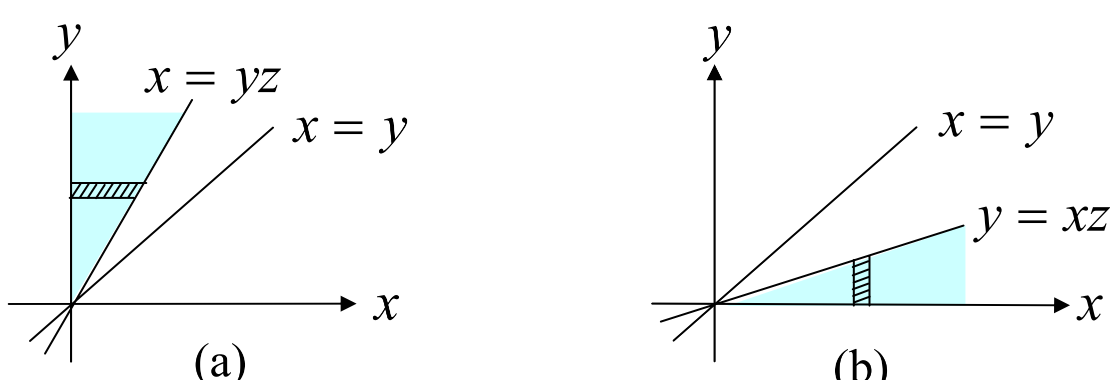


Fig. 8.14

From Fig. 8.14

$$F_Z(z) = \int_0^\infty \int_{x=0}^{yz} f_{XY}(x,y)dxdy + \int_0^\infty \int_{y=0}^{xz} f_{XY}(x,y)dydx. \qquad (8\text{-}53)$$

Hence

$$f_Z(z) = \int_0^\infty y\, f_{XY}(yz,y)dy + \int_0^\infty x\, f_{XY}(x,xz)dx = \int_0^\infty y\{f_{XY}(yz,y) + f_{XY}(y,yz)\}dy$$

$$= \int_0^\infty y\lambda^2 \left\{ e^{-\lambda(yz+y)} + e^{-\lambda(y+yz)} \right\}dy = 2\lambda^2 \int_0^\infty y e^{-\lambda(1+z)y} dy = \frac{2}{(1+z)^2}\int_0^\infty u e^{-u} dy$$

$$= \begin{cases} \dfrac{2}{(1+z)^2}, & 0 < z < 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (8\text{-}54)$$

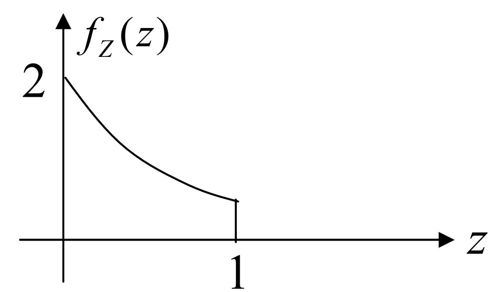


Fig. 8.15

Example 8.13 (Discrete Case): Let X and Y be independent Poisson random variables with parameters λ_1 and λ_2 respectively. Let $Z = X + Y$. Determine the p.m.f of Z.

Solution: Since X and Y both take integer values $\{0, 1, 2, \cdots\}$, the same is true for Z. For any $n = 0, 1, 2, \cdots,\quad X + Y = n$ gives only a finite number of options for X and Y. In fact, if $X = 0$, then Y must be n; if $X = 1$, then Y must be n-1, etc. Thus the event $\{X + Y = n\}$ is the union of $(n + 1)$ mutually exclusive events A_k given by

$$A_k = \{X = k,\ Y = n - k\}, \qquad k = 0,1,2,\cdots,n. \tag{8-55}$$

As a result

$$P(Z = n) = P(X + Y = n) = P\left(\bigcup_{k=0}^{n}\left(X = k,\ Y = n - k\right)\right)$$

$$= \sum_{k=0}^{n} P(X = k,\ Y = n - k). \tag{8-56}$$

If X and Y are also independent, then

$$P\left(X = k,\ Y = n - k\right) = P(X = k)P(Y = n - k)$$

and hence

$$P(Z=n)=\sum_{k=0}^{n} P(X=k,\ Y=n-k)$$

$$=\sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}$$

$$=e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}, \qquad n=0,1,2,\cdots,\infty. \tag{8-57}$$

Thus Z represents a Poisson random variable with parameter $\lambda_1+\lambda_2$, indicating that sum of independent Poisson random variables is also a Poisson random variable whose parameter is the sum of the parameters of the original random variables.

As the last example illustrates, the above procedure for determining the p.m.f of functions of discrete random variables is somewhat tedious. As we shall see in Lecture 10, the joint characteristic function can be used in this context to solve problems of this type in an easier fashion.

9. Two Functions of Two Random Variables

In the spirit of the previous section, let us look at an immediate generalization: Suppose X and Y are two random variables with joint p.d.f $f_{XY}(x,y)$. Given two functions $g(x,y)$ and $h(x,y)$, define the new random variables

$$Z = g(X,Y) \tag{9-1}$$

$$W = h(X,Y). \tag{9-2}$$

How does one determine their joint p.d.f $f_{ZW}(z,w)$? Obviously with $f_{ZW}(z,w)$ in hand, the marginal p.d.fs $f_Z(z)$ and $f_W(w)$ can be easily determined.

The procedure is the same as that in (8-3). In fact for given z and w,

$$F_{ZW}(z,w) = P\big(Z(\xi) \le z, W(\xi) \le w\big) = P\big(g(X,Y) \le z, h(X,Y) \le w\big)$$
$$= P\big((X,Y) \in D_{z,w}\big) = \int\int_{(x,y)\in D_{z,w}} f_{XY}(x,y)\,dxdy, \qquad (9\text{-}3)$$

where $D_{z,w}$ is the region in the xy plane such that the inequalities $g(x,y) \le z$ and $h(x,y) \le w$ are simultaneously satisfied.

We illustrate this technique in the next example.

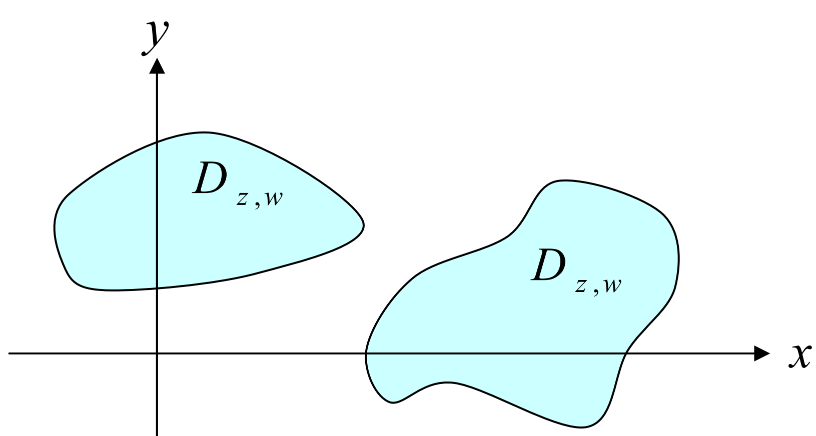


Fig. 9.1

Example 9.1: Suppose X and Y are independent uniformly distributed random variables in the interval $(0,\theta)$. Define $Z = \min(X,Y)$, $W = \max(X,Y)$. Determine $f_{ZW}(z,w)$.

Solution: Obviously both w and z vary in the interval $(0,\theta)$. Thus $F_{ZW}(z,w) = 0$, if $z < 0$ or $w < 0$. (9-4)

$$F_{ZW}(z,w) = P\big(Z \le z, W \le w\big) = P\big(\min(X,Y) \le z,\ \max(X,Y) \le w\big). \quad (9\text{-}5)$$

We must consider two cases: $w \ge z$ and $w < z$, since they give rise to different regions for $D_{z,w}$ (see Figs. 9.2 (a)-(b)).

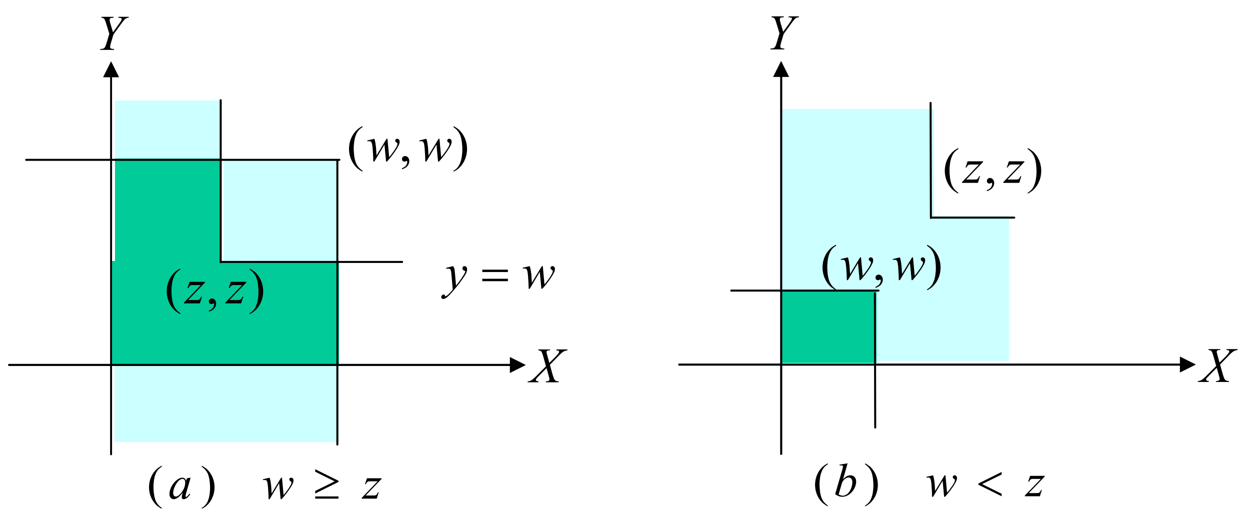


Fig. 9.2

For $w \geq z$, from Fig. 9.2 (a), the region $D_{z,w}$ is represented by the doubly shaded area. Thus

$$F_{ZW}(z,w) = F_{XY}(z,w) + F_{XY}(w,z) - F_{XY}(z,z), \quad w \geq z, \tag{9-6}$$

and for $w < z$, from Fig. 9.2 (b), we obtain

$$F_{ZW}(z,w) = F_{XY}(w,w), \quad w < z. \tag{9-7}$$

With

$$F_{XY}(x,y) = F_X(x)\, F_Y(y) = \frac{x}{\theta} \cdot \frac{y}{\theta} = \frac{xy}{\theta^2}, \tag{9-8}$$

we obtain

$$F_{ZW}(z,w) = \begin{cases} 2(w-z)z/\theta^2, & 0 < z < w < \theta, \\ w^2/\theta^2, & 0 < w < z < \theta. \end{cases} \tag{9-9}$$

Thus

$$f_{ZW}(z,w) = \begin{cases} 2/\theta^2, & 0 < z < w < \theta, \\ 0, & \text{otherwise}. \end{cases} \tag{9-10}$$

From (9-10), we also obtain

$$f_Z(z) = \int_z^{\theta} f_{ZW}(z,w)dw = \frac{2}{\theta}\left(1 - \frac{z}{\theta}\right), \quad 0 < z < \theta, \tag{9-11}$$

and

$$f_W(w) = \int_0^{w} f_{ZW}(z,w)dz = \frac{2w}{\theta^2}, \quad 0 < w < \theta. \tag{9-12}$$

If $g(x,y)$ and $h(x,y)$ are continuous and differentiable functions, then as in the case of one random variable (see (5-30)) it is possible to develop a formula to obtain the joint p.d.f $f_{ZW}(z,w)$ directly. Towards this, consider the equations

$$g(x,y) = z, \qquad h(x,y) = w. \tag{9-13}$$

For a given point (z,w), equation (9-13) can have many solutions. Let us say

$$(x_1, y_1),\ (x_2, y_2),\ \cdots,\ (x_n, y_n),$$

represent these multiple solutions such that (see Fig. 9.3)

$$g(x_i, y_i) = z, \qquad h(x_i, y_i) = w. \tag{9-14}$$

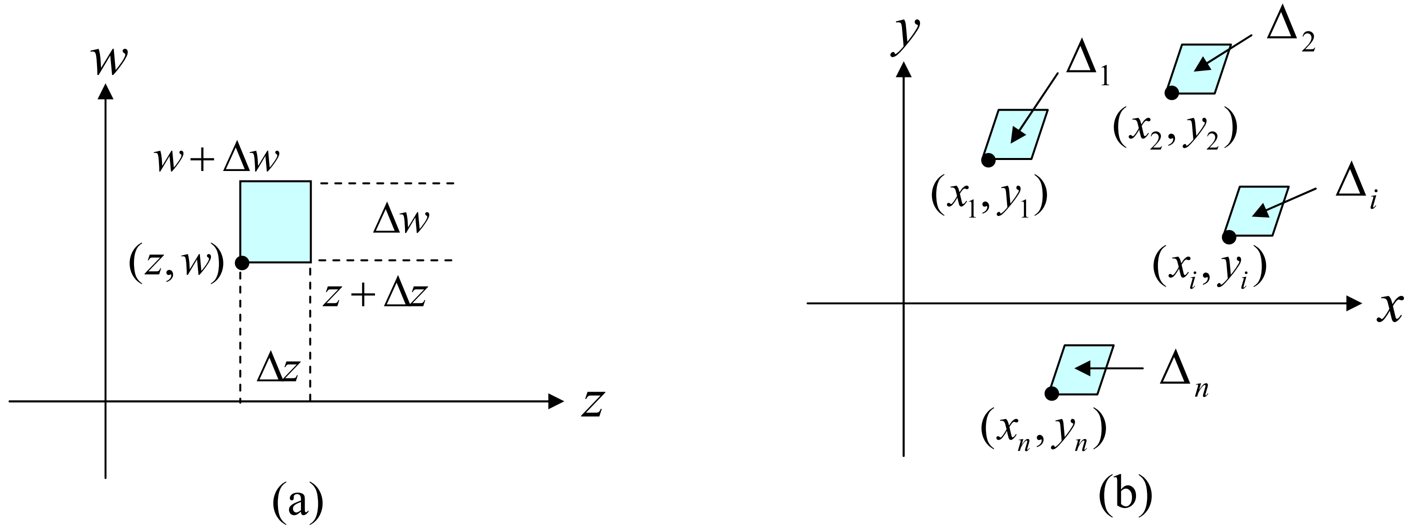


Fig. 9.3

Consider the problem of evaluating the probability

$$P\big(z < Z \le z + \Delta z, w < W \le w + \Delta w\big)$$
$$= P\big(z < g(X,Y) \le z + \Delta z, w < h(X,Y) \le w + \Delta w\big). \tag{9-15}$$

Using (7-9) we can rewrite (9-15) as

$$P\left(z < Z \le z + \Delta z, w < W \le w + \Delta w\right) = f_{ZW}(z,w)\Delta z \Delta w. \qquad (9\text{-}16)$$

But to translate this probability in terms of $f_{XY}(x,y)$, we need to evaluate the equivalent region for $\Delta z\, \Delta w$ in the *xy* plane. Towards this referring to Fig. 9.4, we observe that the point *A* with coordinates (*z*,*w*) gets mapped onto the point A' with coordinates (x_i, y_i) (as well as to other points as in Fig. 9.3(b)). As *z* changes to $z + \Delta z$ to point *B* in Fig. 9.4 (a), let B' represent its image in the *xy* plane. Similarly as *w* changes to $w + \Delta w$ to *C*, let C' represent its image in the *xy* plane.

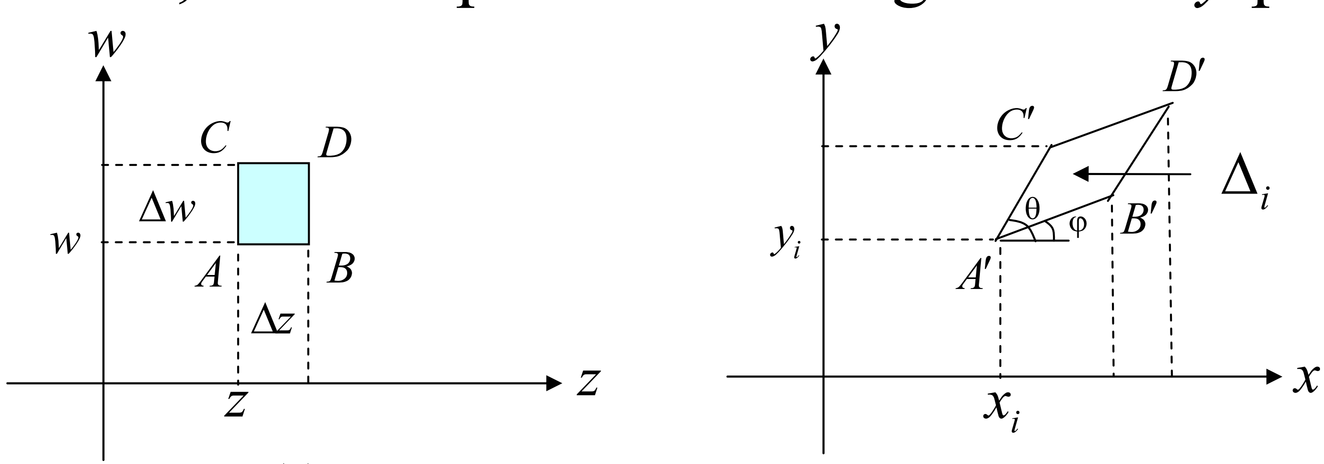


Fig. 9.4

Finally D goes to D', and $A'B'C'D'$ represents the equivalent parallelogram in the XY plane with area Δ_i. Referring back to Fig. 9.3, the probability in (9-16) can be alternatively expressed as

$$\sum_i P\big((X,Y) \in \Delta_i\big) = \sum_i f_{XY}(x_i, y_i)\Delta_i. \tag{9-17}$$

Equating (9-16) and (9-17) we obtain

$$f_{ZW}(z,w) = \sum_i f_{XY}(x_i, y_i)\frac{\Delta_i}{\Delta z \Delta w}. \tag{9-18}$$

To simplify (9-18), we need to evaluate the area Δ_i of the parallelograms in Fig. 9.3 (b) in terms of $\Delta z \Delta w$. Towards this, let g_1 and h_1 denote the inverse transformation in (9-14), so that

$$x_i = g_1(z,w), \qquad y_i = h_1(z,w). \tag{9-19}$$

As the point (z,w) goes to $(x_i, y_i) \equiv A'$, the point $(z+\Delta z, w) \to B'$, the point $(z, w+\Delta w) \to C'$, and the point $(z+\Delta z, w+\Delta w) \to D'$. Hence the respective x and y coordinates of B' are given by

$$g_1(z+\Delta z, w) = g_1(z,w) + \frac{\partial g_1}{\partial z}\Delta z = x_i + \frac{\partial g_1}{\partial z}\Delta z, \tag{9-20}$$

and

$$h_1(z+\Delta z, w) = h_1(z,w) + \frac{\partial h_1}{\partial z}\Delta z = y_i + \frac{\partial h_1}{\partial z}\Delta z. \tag{9-21}$$

Similarly those of C' are given by

$$x_i + \frac{\partial g_1}{\partial w}\Delta w, \quad y_i + \frac{\partial h_1}{\partial w}\Delta w. \tag{9-22}$$

The area of the parallelogram $A'B'C'D'$ in Fig. 9.4 (b) is given by

$$\begin{aligned}\Delta_i &= \left(A'B'\right)\left(A'C'\right)\sin(\theta - \varphi) \\ &= \left(A'B'\cos\varphi\right)\left(A'C'\sin\theta\right) - \left(A'B'\sin\varphi\right)\left(A'C'\cos\theta\right).\end{aligned} \tag{9-23}$$

But from Fig. 9.4 (b), and (9-20) - (9-22)

$$A'B'\cos\varphi = \frac{\partial g_1}{\partial z}\Delta z, \quad A'C'\sin\theta = \frac{\partial h_1}{\partial w}\Delta w, \tag{9-24}$$

$$A'B'\sin\varphi = \frac{\partial h_1}{\partial z}\Delta z, \quad A'C'\cos\theta = \frac{\partial g_1}{\partial w}\Delta w. \tag{9-25}$$

so that

$$\Delta_i = \left(\frac{\partial g_1}{\partial z}\frac{\partial h_1}{\partial w} - \frac{\partial g_1}{\partial w}\frac{\partial h_1}{\partial z} \right)\Delta z \Delta w \tag{9-26}$$

and

$$\frac{\Delta_i}{\Delta z \Delta w} = \left(\frac{\partial g_1}{\partial z}\frac{\partial h_1}{\partial w} - \frac{\partial g_1}{\partial w}\frac{\partial h_1}{\partial z} \right) = \det \begin{pmatrix} \frac{\partial g_1}{\partial z} & \frac{\partial g_1}{\partial w} \\ \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \end{pmatrix} \tag{9-27}$$

The right side of (9-27) represents the Jacobian $|J(z,w)|$ of the transformation in (9-19). Thus

$$|J(z,w)| = \det \begin{pmatrix} \dfrac{\partial g_1}{\partial z} & \dfrac{\partial g_1}{\partial w} \\ \dfrac{\partial h_1}{\partial z} & \dfrac{\partial h_1}{\partial w} \end{pmatrix}. \tag{9-28}$$

Substituting (9-27) - (9-28) into (9-18), we get

$$f_{ZW}(z,w) = \sum_i |J(z,w)|\, f_{XY}(x_i, y_i) = \sum_i \frac{1}{|J(x_i, y_i)|} f_{XY}(x_i, y_i), \tag{9-29}$$

since

$$|J(z,w)| = \frac{1}{|J(x_i, y_i)|} \tag{9-30}$$

where $|J(x_i, y_i)|$ represents the Jacobian of the original transformation in (9-13) given by

$$J(x_i, y_i) = \det \begin{pmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \\ \dfrac{\partial h}{\partial x} & \dfrac{\partial h}{\partial y} \end{pmatrix}_{x = x_i,\, y = y_i}. \tag{9-31}$$

Next we shall illustrate the usefulness of the formula in (9-29) through various examples:

Example 9.2: Suppose X and Y are zero mean independent Gaussian r.vs with common variance σ^2.
Define $Z = \sqrt{X^2 + Y^2}$, $W = \tan^{-1}(Y/X)$, where $|w| \le \pi/2$.
Obtain $f_{ZW}(z,w)$.
Solution: Here

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{9-32}$$

Since

$$z = g(x,y) = \sqrt{x^2 + y^2}; w = h(x,y) = \tan^{-1}(y/x), \quad |w| \le \pi/2, \tag{9-33}$$

if (x_1, y_1) is a solution pair so is $(-x_1, -y_1)$. From (9-33)

$$\frac{y}{x} = \tan w, \quad \text{or} \quad y = x \tan w. \tag{9-34}$$

Substituting this into z, we get

$$z = \sqrt{x^2 + y^2} = x\sqrt{1 + \tan^2 w} = x \ \sec w, \quad \text{or} \quad x = z \ \cos w. \tag{9-35}$$

and

$$y = x \tan w = z \sin w. \tag{9-36}$$

Thus there are two solution sets

$$x_1 = z \cos w, \ \ y_1 = z \sin w, \ \ x_2 = -z \cos w, \ \ y_2 = -z \sin w. \tag{9-37}$$

We can use (9-35) - (9-37) to obtain $J(z, w)$. From (9-28)

$$J(z, w) = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} \cos w & -z \sin w \\ \sin w & z \cos w \end{vmatrix} = z, \tag{9-38}$$

so that

$$|J(z, w)| = z. \tag{9-39}$$

We can also compute $J(x,y)$ using (9-31). From (9-33),

$$J(x,y) = \begin{vmatrix} \dfrac{x}{\sqrt{x^2+y^2}} & \dfrac{y}{\sqrt{x^2+y^2}} \\ \dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{vmatrix} = \frac{1}{\sqrt{x^2+y^2}} = \frac{1}{z}. \qquad (9\text{-}40)$$

Notice that $|J(z,w)| = 1/|J(x_i,y_i)|$, agreeing with (9-30). Substituting (9-37) and (9-39) or (9-40) into (9-29), we get

$$f_{ZW}(z,w) = z\big(f_{XY}(x_1,y_1) + f_{XY}(x_2,y_2)\big)$$

$$= \frac{z}{\pi\sigma^2} e^{-z^2/2\sigma^2}, \qquad 0 < z < \infty, \quad |w| < \frac{\pi}{2}. \qquad (9\text{-}41)$$

Thus

$$f_Z(z) = \int_{-\pi/2}^{\pi/2} f_{ZW}(z,w)\,dw = \frac{z}{\sigma^2} e^{-z^2/2\sigma^2}, \qquad 0 < z < \infty, \qquad (9\text{-}42)$$

which represents a Rayleigh r.v with parameter σ^2, and

$$f_W(w) = \int_0^{\infty} f_{ZW}(z,w)\,dz = \frac{1}{\pi}, \quad |w| < \frac{\pi}{2}, \qquad (9\text{-}43)$$

which represents a uniform r.v in the interval $(-\pi/2, \pi/2)$. Moreover by direct computation

$$f_{ZW}(z,w) = f_Z(z) \cdot f_W(w) \tag{9-44}$$

implying that Z and W are independent. We summarize these results in the following statement: If X and Y are zero mean independent Gaussian random variables with common variance, then $\sqrt{X^2+Y^2}$ has a Rayleigh distribution and $\tan^{-1}(Y/X)$ has a uniform distribution. Moreover these two derived r.vs are statistically independent. Alternatively, with X and Y as independent zero mean r.vs as in (9-32), $X + jY$ represents a complex Gaussian r.v. But

$$X + jY = Ze^{jW}, \tag{9-45}$$

where Z and W are as in (9-33), except that for (9-45) to hold good on the entire complex plane we must have $-\pi < W < \pi$, and hence it follows that the magnitude and phase of a complex Gaussian r.v are independent with Rayleigh and uniform distributions $(U \sim (-\pi, \pi))$ respectively. The statistical independence of these derived r.vs is an interesting observation.

Example 9.3: Let X and Y be independent exponential random variables with common parameter λ.
Define $U = X + Y$, $V = X - Y$. Find the joint and marginal p.d.f of U and V.
Solution: It is given that

$$f_{XY}(x, y) = \frac{1}{\lambda^2} e^{-(x+y)/\lambda}, \quad x > 0, \quad y > 0. \tag{9-46}$$

Now since $u = x + y$, $v = x - y$, always $|v| < u$, and there is only one solution given by

$$x = \frac{u+v}{2}, \quad y = \frac{u-v}{2}. \tag{9-47}$$

Moreover the Jacobian of the transformation is given by

$$J(x,y) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

and hence

$$f_{UV}(u,v) = \frac{1}{2\lambda^2} e^{-u/\lambda}, \qquad 0 < |\,v\,| < u < \infty, \tag{9-48}$$

represents the joint p.d.f of U and V. This gives

$$f_U(u) = \int_{-u}^{u} f_{UV}(u,v)dv = \frac{1}{2\lambda^2}\int_{-u}^{u} e^{-u/\lambda} dv = \frac{u}{\lambda^2} e^{-u/\lambda}, \quad 0 < u < \infty, \tag{9-49}$$

and

$$f_V(v) = \int_{|v|}^{\infty} f_{UV}(u,v)du = \frac{1}{2\lambda^2}\int_{|v|}^{\infty} e^{-u/\lambda} du = \frac{1}{2\lambda} e^{-|u|/\lambda}, \quad -\infty < v < \infty. \tag{9-50}$$

Notice that in this case the r.vs U and V are not independent.

As we show below, the general transformation formula in (9-29) making use of two functions can be made useful even when only one function is specified.

Auxiliary Variables:

Suppose

$$Z = g(X, Y), \tag{9-51}$$

where X and Y are two random variables. To determine $f_Z(z)$ by making use of the above formulation in (9-29), we can define an auxiliary variable

$$W = X \quad \text{or} \quad W = Y \tag{9-52}$$

and the p.d.f of Z can be obtained from $f_{ZW}(z, w)$ by proper integration.

Example 9.4: Suppose $Z = X + Y$ and let $W = Y$ so that the transformation is one-to-one and the solution is given by $y_1 = w, \quad x_1 = z - w.$

The Jacobian of the transformation is given by

$$J(x,y) = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1$$

and hence

$$f_{ZW}(x,y) = f_{XY}(x_1,y_1) = f_{XY}(z-w,w)$$

or

$$f_Z(z) = \int f_{ZW}(z,w)dw = \int_{-\infty}^{+\infty} f_{XY}(z-w,w)dw, \qquad (9\text{-}53)$$

which agrees with (8.7). Note that (9-53) reduces to the convolution of $f_X(z)$ and $f_Y(z)$ if X and Y are independent random variables. Next, we consider a less trivial example.

Example 9.5: Let $X \sim U(0,1)$ and $Y \sim U(0,1)$ be independent. Define $Z = (-2\ln X)^{1/2}\cos(2\pi Y)$. (9-54)

Find the density function of Z.

Solution: We can make use of the auxiliary variable $W = Y$ in this case. This gives the only solution to be

$$x_1 = e^{-\left(z\sec(2\pi w)\right)^2/2}, \tag{9-55}$$

$$y_1 = w, \tag{9-56}$$

and using (9-28)

$$J(z,w) = \begin{vmatrix} \dfrac{\partial x_1}{\partial z} & \dfrac{\partial x_1}{\partial w} \\ \dfrac{\partial y_1}{\partial z} & \dfrac{\partial y_1}{\partial w} \end{vmatrix} = \begin{vmatrix} -z\sec^2(2\pi w)\ e^{-\left(z\sec(2\pi w)\right)^2/2} & \dfrac{\partial x_1}{\partial w} \\ 0 & 1 \end{vmatrix} \tag{9-57}$$

$$= -z\sec^2(2\pi w)e^{-\left(z\sec(2\pi w)\right)^2/2}.$$

Substituting (9-55) - (9-57) into (9-29), we obtain

$$f_{ZW}(z,w) = z\sec^2(2\pi w)\ e^{-\left(z\sec(2\pi w)\right)^2/2}, \qquad -\infty < z < +\infty, \quad 0 < w < 1, \tag{9-58}$$

and

$$f_Z(z) = \int_0^1 f_{ZW}(z,w)dw = e^{-z^2/2} \int_0^1 z \sec^2(2\pi w) \;\; e^{-\left(z \tan(2\pi w)\right)^2/2} dw. \quad (9\text{-}59)$$

Let $u = z\tan(2\pi w)$ so that $du = 2\pi z \sec^2(2\pi w)dw.$ Notice that as w varies from 0 to 1, u varies from $-\infty$ to $+\infty.$ Using this in (9-59), we get

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \underbrace{\int_{-\infty}^{+\infty} e^{-u^2/2} \frac{du}{\sqrt{2\pi}}}_{1} = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty, \quad (9\text{-}60)$$

which represents a zero mean Gaussian r.v with unit variance. Thus $Z \sim N(0,1).$ Equation (9-54) can be used as a practical procedure to generate Gaussian random variables from those of two independent uniformly distributed random sequences.

Example 9.6 : Let X and Y be independent identically distributed Geometric random variables with

$$P(X=k)=P(Y=K)=pq^k, \qquad k=0,1,2,\cdots.$$

(a) Show that $min\ (X, Y)$ and $X-Y$ are independent random variables.
(b) Show that $min\ (X, Y)$ and $max\ (X, Y)-min\ (X, Y)$ are also independent random variables.

Solution: (a) Let

$$Z=\min\,(X, Y)\ , \text{ and } \ W=X-Y. \tag{9-61}$$

Note that Z takes only nonnegative values $\{0,1,2,\cdots\}$, while W takes both positive, zero and negative values $\{0,\pm 1,\pm 2,\cdots\}$. We have $P(Z=m,\ W=n)\ =P\{\min\,(X, Y)=m,\ X-Y\ =n\}$. But

$$Z=\min(X,Y)=\begin{cases} Y & X\geq Y \Rightarrow W=X-Y \text{ is nonnegative} \\ X & X<Y \Rightarrow W=X-Y \text{ is negative.} \end{cases}$$

Thus

$$\begin{aligned} P(Z=m,W=n) &= P\{\min(X,Y)=m, X-Y=n, (X\geq Y \cup X<Y)\} \\ &= P(\min(X,Y)=m, X-Y=n, X\geq Y) \\ &\quad + P(\min(X,Y)=m, X-Y=n, X<Y) \end{aligned} \tag{9-62}$$

$$
\begin{aligned}
P(Z=m,W=n) &= P(Y=m, X=m+n, X\ge Y) \\
&\quad + P(X=m, Y=m-n, X<Y) \\
&= \begin{cases} P(X=m+n)P(Y=m) = pq^{m+n}pq^{m}, m\ge 0, n\ge 0 \\ P(X=m)P(Y=m-n) = pq^{m}pq^{m-n}, m\ge 0, n<0 \end{cases} \\
&= p^2 q^{2m+|n|}, \quad m=0,1,2,\cdots \quad n=0,\pm 1,\pm 2,\cdots \qquad \text{(9-63)}
\end{aligned}
$$

represents the joint probability mass function of the random variables Z and W. Also

$$
\begin{aligned}
P(Z=m) &= \sum_n P(Z=m,W=n) = \sum_n p^2 q^{2m} q^{|n|} \\
&= p^2 q^{2m}(1+2q+2q^2+\cdots) \\
&= p^2 q^{2m}(1+\tfrac{2q}{1-q}) = pq^{2m}(1+q) \\
&= p(1+q)q^{2m}, \quad m=0,1,2,\cdots. \qquad \text{(9-64)}
\end{aligned}
$$

Thus Z represents a Geometric random variable since $1-q^2 = p(1+q),$ and

$$P(W=m)=\sum_{m=0}^{\infty}P(Z=m,W=n)=\sum_{m=0}^{\infty}p^2q^{2m}q^{|n|}$$
$$=p^2q^{|n|}(1+q^2+q^4+\cdots)=p^2q^{|n|}\frac{1}{1-q^2}$$
$$=\frac{p}{1+q}q^{|n|},\qquad n=0,\pm 1,\pm 2,\cdots. \tag{9-65}$$

Note that

$$P(Z=m,W=n)=P(Z=m)P(W=n), \tag{9-66}$$

establishing the independence of the random variables Z and W. The independence of $X-Y$ and $min\,(X\,,\,Y\,)$ when X and Y are independent Geometric random variables is an interesting observation.
(b) Let

$$Z=min\,(X\,,\,Y\,)\,,\quad R=max\,(X\,,\,Y\,)-min\,(X\,,\,Y\,). \tag{9-67}$$

In this case both Z and R take nonnegative integer values $0,1,2,\cdots.$ Proceeding as in (9-62)-(9-63) we get

$$
\begin{aligned}
P\{Z=m, R=n\} &= P\{\min(X,Y)=m, \max(X,Y)-\min(X,Y)=n, X \ge Y\} \\
&\quad + P\{\min(X,Y)=m, \max(X,Y)-\min(X,Y)=n, X<Y\} \\
&= P\{Y=m, X=m+n, X \ge Y) + P(X=m, Y=m+n, X<Y\} \\
&= P\{X=m+n, Y=m, X \ge Y) + P(X=m, Y=m+n, X<Y\} \\
&= \begin{cases} pq^{m+n}pq^{m} + pq^{m}pq^{m+n}, & m=0,1,2,\cdots, \quad n=1,2,\cdots \\ pq^{m+n}pq^{m}, & m=0,1,2,\cdots, \quad n=0 \end{cases} \\
&= \begin{cases} 2p^2q^{2m+n}, & m=0,1,2,\cdots, \quad n=1,2,\cdots \\ p^2q^{2m}, & m=0,1,2,\cdots, \quad n=0. \end{cases} \qquad (9\text{-}68)
\end{aligned}
$$

Eq. (9-68) represents the joint probability mass function of Z and R in (9-67). From (9-68),

$$
\begin{aligned}
P(Z=m) &= \sum_{n=0}^{\infty} P\{Z=m, R=n\} = p^2q^{2m}(1+2\sum_{n=1}^{\infty} q^n) = p^2q^{2m}\left(1+\frac{2q}{p}\right) \\
&= p(1+q)q^{2m}, \qquad m=0,1,2,\cdots \qquad (9\text{-}69)
\end{aligned}
$$

and

$$P(R=n)=\sum_{m=0}^{\infty} P\{Z=m, R=n\}=\begin{cases} \dfrac{p}{1+q}, & n=0 \\ \dfrac{2p}{1+q}q^n, & n=1,2,\cdots. \end{cases} \tag{9-70}$$

From (9-68)-(9-70), we get

$$P(Z=m, R=n)=P(Z=m)P(R=n) \tag{9-71}$$

which proves the independence of the random variables Z and R defined in (9-67) as well.

10. Joint Moments and Joint Characteristic Functions

Following section 6, in this section we shall introduce various parameters to compactly represent the information contained in the joint p.d.f of two r.vs. Given two r.vs X and Y and a function $g(x, y)$, define the r.v

$$Z = g(X, Y) \tag{10-1}$$

Using (6-2), we can define the mean of Z to be

$$\mu_Z = E(Z) = \int_{-\infty}^{+\infty} z \, f_Z(z) dz. \tag{10-2}$$

However, the situation here is similar to that in (6-13), and it is possible to express the mean of $Z = g(X,Y)$ in terms of $f_{XY}(x,y)$ *without* computing $f_Z(z)$. To see this, recall from (5-26) and (7-10) that

$$P\big(z < Z \le z + \Delta z\big) = f_Z(z)\Delta z = P\big(z < g(X,Y) \le z + \Delta z\big)$$
$$= \sum \sum_{(x,y)\in D_{\Delta z}} f_{XY}(x,y)\Delta x \Delta y \qquad (10\text{-}3)$$

where $D_{\Delta z}$ is the region in *xy* plane satisfying the above inequality. From (10-3), we get

$$z\ f_Z(z)\Delta z = g(x,y) \sum \sum_{(x,y)\in D_{\Delta z}} f(x,y)\Delta x \Delta y. \qquad (10\text{-}4)$$

As Δz covers the entire *z* axis, the corresponding regions $D_{\Delta z}$ are nonoverlapping, and they cover the entire *xy* plane.

By integrating (10-4), we obtain the useful formula

$$E(z) = \int_{-\infty}^{+\infty} z\, f_Z(z)dz = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{XY}(x,y)dxdy. \tag{10-5}$$

or

$$E[g(x,y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{XY}(x,y)dxdy. \tag{10-6}$$

If X and Y are discrete-type r.vs, then

$$E[g(x,y)] = \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j). \tag{10-7}$$

Since expectation is a linear operator, we also get

$$E\left(\sum_k a_k g_k(x,y)\right) = \sum_k a_k E[g_k(x_i, y_j)]. \tag{10-8}$$

If X and Y are independent r.vs, it is easy to see that $Z = g(X)$ and $W = h(Y)$ are always independent of each other. In that case using (10-7), we get the interesting result

$$E[g(X)h(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y)f_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{+\infty} g(x)f_X(x)dx \int_{-\infty}^{+\infty} h(y)f_Y(y)dy = E[g(X)]E[h(Y)]. \quad (10\text{-}9)$$

However (10-9) is in general not true (if X and Y are not independent).

In the case of one random variable (see (10- 6)), we defined the parameters mean and variance to represent its average behavior. How does one parametrically represent similar cross-behavior between two random variables? Towards this, we can generalize the variance definition given in (6-16) as shown below:

Covariance: Given any two r.vs X and Y, define

$$Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]. \tag{10-10}$$

By expanding and simplifying the right side of (10-10), we also get

$$\begin{aligned} Cov(X,Y) &= E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y) \\ &= \overline{XY} - \overline{X}\,\overline{Y}. \end{aligned} \tag{10-11}$$

It is easy to see that

$$|Cov(X,Y)| \le \sqrt{Var(X)Var(Y)}. \tag{10-12}$$

To see (10-12), let $U = aX + Y$, so that

$$\begin{aligned} Var(U) &= E\left[\{a(X-\mu_X) + (Y-\mu_Y)\}^2\right] \\ &= a^2 Var(X) + 2a\, Cov(X,Y) + Var(Y) \ge 0. \end{aligned} \tag{10-13}$$

The right side of (10-13) represents a quadratic in the variable *a* that has no distinct real roots (Fig. 10.1). Thus the roots are imaginary (or double) and hence the discriminant

$$[Cov(X,Y)]^2 - Var(X)\,Var(Y)$$

must be non-positive, and that gives (10-12). Using (10-12), we may define the normalized parameter

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \le \rho_{XY} \le 1, \qquad (10\text{-}14)$$

or

$$Cov(X,Y) = \rho_{XY}\sigma_X\sigma_Y \qquad (10\text{-}15)$$

and it represents the correlation coefficient between *X* and *Y*.

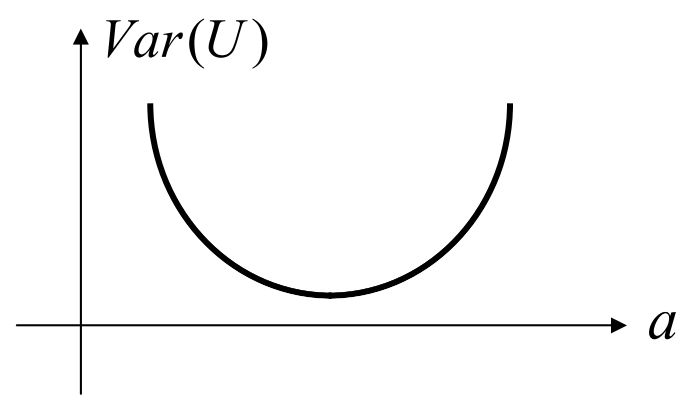


Fig. 10.1

Uncorrelated r.vs: If $\rho_{XY} = 0$, then X and Y are said to be uncorrelated r.vs. From (11), if X and Y are uncorrelated, then

$$E(XY) = E(X)E(Y). \qquad (10\text{-}16)$$

Orthogonality: X and Y are said to be orthogonal if

$$E(XY) = 0. \qquad (10\text{-}17)$$

From (10-16) - (10-17), if either X or Y has zero mean, then orthogonality implies uncorrelatedness also and vice-versa. Suppose X and Y are independent r.vs. Then from (10-9) with $g(X) = X,\ h(Y) = Y,$ we get

$$E(XY) = E(X)E(Y),$$

and together with (10-16), we conclude that the random variables are uncorrelated, thus justifying the original definition in (10-10). Thus independence implies uncorrelatedness.

Naturally, if two random variables are statistically independent, then there cannot be any correlation between them $(\rho_{XY} = 0)$. However, the converse is in general not true. As the next example shows, random variables can be uncorrelated without being independent.

Example 10.1: Let $X \sim U(0,1),\ \ Y \sim U(0,1).$ Suppose X and Y are independent. Define $Z = X + Y,\ \ W = X - Y$. Show that Z and W are dependent, but uncorrelated r.vs.

Solution: $z = x + y, \quad w = x - y$ gives the only solution set to be

$$x = \frac{z + w}{2}, \quad y = \frac{z - w}{2}.$$

Moreover $0 < z < 2,\ -1 < w < 1,\ \ z + w \le 2,\ \ z - w \le 2,\ \ z > |w|$ and $|J(z, w)| = 1/2.$

Thus (see the shaded region in Fig. 10.2)

$$f_{ZW}(z,w) = \begin{cases} 1/2, & 0 < z < 2,\ -1 < w < 1,\ z+w \le 2,\ z-w \le 2,\ |w| < z, \\ 0, & \text{otherwise} \ , \end{cases} \tag{10-18}$$

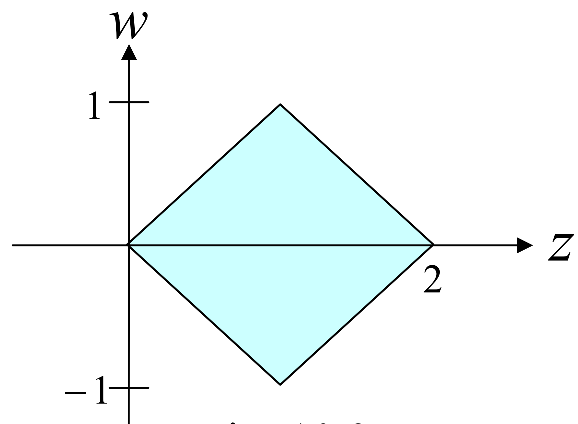


Fig. 10.2

and hence

$$f_Z(z) = \int f_{ZW}(z,w)dw = \begin{cases} \displaystyle\int_{-z}^{z} \frac{1}{2}\,dw = z, & 0 < z < 1, \\ \displaystyle\int_{z\text{-}2}^{2\text{-}z} \frac{1}{2}\,dw = 2 - z, & 1 < z < 2, \end{cases}$$

or by direct computation ($Z = X + Y$)

$$f_Z(z) = f_X(z) \otimes f_Y(z) = \begin{cases} z, & 0 < z < 1, \\ 2 - z, & 1 < z < 2, \\ 0, & \text{otherwise}, \end{cases} \tag{10-19}$$

and

$$f_W(w) = \int f_{ZW}(z,w)dz = \int_{|w|}^{2-|w|} \frac{1}{2} dz = \begin{cases} 1 - |w|, & -1 < w < 1, \\ 0, & \text{otherwise}. \end{cases} \tag{10-20}$$

Clearly $f_{ZW}(z,w) \neq f_Z(z) f_W(w)$. Thus Z and W are not independent. However

$$E(ZW) = E[(X+Y)(X-Y)] = E(X^2) - E(Y^2) = 0, \tag{10-21}$$

and

$$E(W) = E(X - Y) = 0,$$

and hence

$$Cov(Z,W) = E(ZW) - E(Z)E(W) = 0 \tag{10-22}$$

implying that Z and W are uncorrelated random variables.

Example 10.2: Let $Z = aX + bY$. Determine the variance of Z in terms of σ_X, σ_Y and ρ_{XY}.

Solution:

$$\mu_Z = E(z) = E(aX + bY) = a\mu_X + b\mu_Y$$

and using (10-15)

$$\begin{aligned}\sigma_Z^2 = Var(z) &= E\left[(Z-\mu_Z)^2\right] = E\left[\left(a(X-\mu_X) + b(Y-\mu_Y)\right)^2\right] \\ &= a^2 E(X-\mu_X)^2 + 2abE\left((X-\mu_X)(Y-\mu_Y)\right) + b^2 E(Y-\mu_Y)^2 \\ &= a^2\sigma_X^2 + 2ab\rho_{XY}\sigma_X\sigma_Y + b^2\sigma_Y^2. \qquad (10\text{-}23)\end{aligned}$$

In particular if X and Y are independent, then $\rho_{XY} = 0$, and (10-23) reduces to

$$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2. \qquad (10\text{-}24)$$

Thus the variance of the sum of independent r.vs is the sum of their variances $(a = b = 1)$.

Moments:

$$E[X^k Y^m] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^k y^m f_{XY}(x,y)dx\ dy, \qquad (10\text{-}25)$$

represents the joint moment of order (k,m) for X and Y.

Following the one random variable case, we can define the joint characteristic function between two random variables which will turn out to be useful for moment calculations.

Joint characteristic functions:

The joint characteristic function between X and Y is defined as

$$\Phi_{XY}(u,v) = E\left(e^{j(Xu+Yv)}\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{j(Xu+Yv)} f_{XY}(x,y)dxdy. \qquad (10\text{-}26)$$

Note that $\left|\Phi_{XY}(u,v)\right| \le \Phi_{XY}(0,0) = 1.$

It is easy to show that

$$E(XY) = \frac{1}{j^2} \frac{\partial^2 \Phi_{XY}(u,v)}{\partial u \partial v}\Bigg|_{u=0,v=0}. \tag{10-27}$$

If X and Y are independent r.vs, then from (10-26), we obtain

$$\Phi_{XY}(u,v) = E(e^{juX})E(e^{jvY}) = \Phi_X(u)\Phi_Y(v). \tag{10-28}$$

Also

$$\Phi_X(u) = \Phi_{XY}(u,0), \quad \Phi_Y(v) = \Phi_{XY}(0,v). \tag{10-29}$$

More on Gaussian r.vs :

From Lecture 7, X and Y are said to be jointly Gaussian as $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, if their joint p.d.f has the form in (7-23). In that case, by direct substitution and simplification, we obtain the joint characteristic function of two jointly Gaussian r.vs to be

$$\Phi_{XY}(u,v) = E(e^{j(Xu+Yv)}) = e^{j(\mu_X u + \mu_Y v) - \frac{1}{2}(\sigma_X^2 u^2 + 2\rho\sigma_X\sigma_Y uv + \sigma_Y^2 v^2)}. \tag{10-30}$$

Equation (10-14) can be used to make various conclusions. Letting $v = 0$ in (10-30), we get

$$\Phi_X(u) = \Phi_{XY}(u,0) = e^{j\mu_X u - \frac{1}{2}\sigma_X^2 u^2}, \tag{10-31}$$

and it agrees with (6-47).

From (7-23) by direct computation using (10-11), it is easy to show that for two jointly Gaussian random variables

$$Cov(X,Y) = \rho\,\sigma_X\sigma_Y.$$

Hence from (10-14), ρ in $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ represents the actual correlation coefficient of the two jointly Gaussian r.vs in (7-23). Notice that $\rho = 0$ implies

$$f_{XY}(X,Y) = f_X(x) f_Y(y).$$

Thus if X and Y are jointly Gaussian, uncorrelatedness does imply independence between the two random variables. Gaussian case is the only exception where the two concepts imply each other.

Example 10.3: Let X and Y be jointly Gaussian r.vs with parameters $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Define $Z = aX + bY$. Determine $f_Z(z)$.

Solution: In this case we can make use of characteristic function to solve this problem.

$$\begin{aligned}\Phi_Z(u) &= E(e^{jZu}) = E(e^{j(aX+bY)u}) = E(e^{jauX + jbuY}) \\ &= \Phi_{XY}(au, bu). \end{aligned} \qquad (10\text{-}32)$$

From (10-30) with u and v replaced by au and bu respectively we get

$$\Phi_Z(u) = e^{j(a\mu_X + b\mu_Y)u - \frac{1}{2}(a^2\sigma_X^2 + 2\rho ab\sigma_X\sigma_Y + b^2\sigma_Y^2)u^2} = e^{j\mu_Z u - \frac{1}{2}\sigma_Z^2 u^2}, \tag{10-33}$$

where

$$\mu_Z \triangleq a\mu_X + b\mu_Y, \tag{10-34}$$

$$\sigma_Z^2 \triangleq a^2\sigma_X^2 + 2\rho ab\sigma_X\sigma_Y + b^2\sigma_Y^2. \tag{10-35}$$

Notice that (10-33) has the same form as (10-31), and hence we conclude that $Z = aX + bY$ is also Gaussian with mean and variance as in (10-34) - (10-35), which also agrees with (10-23).

From the previous example, we conclude that any linear combination of jointly Gaussian r.vs generate a Gaussian r.v.

In other words, linearity preserves Gaussianity. We can use the characteristic function relation to conclude an even more general result.

Example 10.4: Suppose X and Y are jointly Gaussian r.vs as in the previous example. Define two linear combinations

$$Z = aX + bY, \quad W = cX + dY. \tag{10-36}$$

what can we say about their joint distribution?
Solution: The characteristic function of Z and W is given by

$$\begin{aligned}\Phi_{ZW}(u,v) &= E(e^{j(Zu+Wv)}) = E(e^{j(aX+bY)u+j(cX+dY)v}) \\ &= E(e^{jX(au+cv)+jY(bu+dv)}) = \Phi_{XY}(au+cv, bu+dv).\end{aligned} \tag{10-37}$$

As before substituting (10-30) into (10-37) with u and v replaced by $au + cv$ and $bu + dv$ respectively, we get

$$\Phi_{ZW}(u,v) = e^{j(\mu_Z u + \mu_W v) - \frac{1}{2}(\sigma_Z^2 u^2 + 2\rho_{ZW}\sigma_X \sigma_Y uv + \sigma_W^2 v^2)}, \tag{10-38}$$

where

$$\mu_Z = a\mu_X + b\mu_Y, \tag{10-39}$$

$$\mu_W = c\mu_X + d\mu_Y, \tag{10-40}$$

$$\sigma_Z^2 = a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2, \tag{10-41}$$

$$\sigma_W^2 = c^2\sigma_X^2 + 2cd\rho\sigma_X\sigma_Y + d^2\sigma_Y^2, \tag{10-42}$$

and

$$\rho_{ZW} = \frac{ac\sigma_X^2 + (ad+bc)\rho\sigma_X\sigma_Y + bd\sigma_Y^2}{\sigma_Z\sigma_W}. \tag{10-43}$$

From (10-38), we conclude that Z and W are also jointly distributed Gaussian r.vs with means, variances and correlation coefficient as in (10-39) - (10-43).

To summarize, any two linear combinations of jointly Gaussian random variables (independent or dependent) are also jointly Gaussian r.vs.

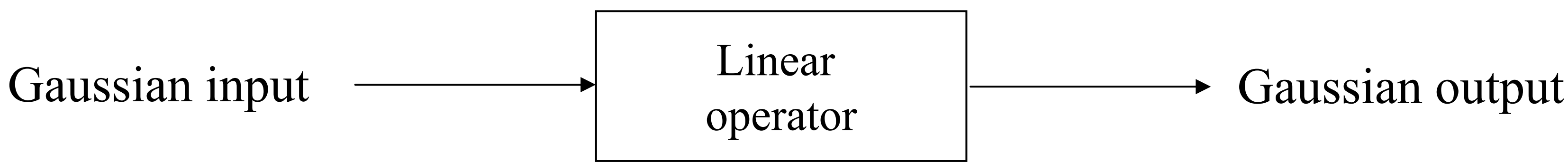


Fig. 10.3

Of course, we could have reached the same conclusion by deriving the joint p.d.f $f_{ZW}(z,w)$ using the technique developed in section 9 (refer (7-29)).

Gaussian random variables are also interesting because of the following result:

Central Limit Theorem: Suppose $X_1, X_2, \cdots, X_n$ are a set of zero mean independent, identically distributed (i.i.d) random variables with some common distribution. Consider their scaled sum

$$Y = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}. \tag{10-44}$$

Then asymptotically (as $n \to \infty$)

$$Y \to N(0, \sigma^2). \tag{10-45}$$

Proof: Although the theorem is true under even more general conditions, we shall prove it here under the independence assumption. Let σ^2 represent their common variance. Since

$$E(X_i) = 0, \tag{10-46}$$

we have

$$Var(X_i) = E(X_i^2) = \sigma^2. \tag{10-47}$$

Consider

$$\Phi_Y(u) = E(e^{jYu}) = E\left(e^{j(X_1+X_2+\cdots+X_n)u/\sqrt{n}}\right) = \prod_{i=1}^{n} E(e^{jX_i u/\sqrt{n}})$$

$$= \prod_{i=1}^{n} \Phi_{X_i}(u/\sqrt{n}) \qquad (10\text{-}48)$$

where we have made use of the independence of the r.vs $X_1, X_2, \cdots, X_n$. But

$$E(e^{jX_i u/\sqrt{n}}) = E\left(1 - \frac{jX_i u}{\sqrt{n}} + \frac{j^2 X_i^2 u^2}{2!\,n} + \frac{j^3 X_i^3 u^3}{3!\,n^{3/2}} + \cdots\right) = 1 - \frac{\sigma^2 u^2}{2n} + o\left(\frac{1}{n^{3/2}}\right), \qquad (10\text{-}49)$$

where we have made use of (10-46) - (10-47). Substituting (10-49) into (10-48), we obtain

$$\Phi_Y(u) = \left[1 - \frac{\sigma^2 u^2}{2n} + o\left(\frac{1}{n^{3/2}}\right)\right]^n, \qquad (10\text{-}50)$$

and as

$$\lim_{n\to\infty} \Phi_Y(n) \to e^{-\sigma^2 u^2/2}, \qquad (51)$$

since

$$\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n \to e^{-x}. \tag{10-52}$$

[Note that $o(1/n^{3/2})$ terms in (10-50) decay faster than $1/n^{3/2}$]. But (10-51) represents the characteristic function of a zero mean normal r.v with variance σ^2 and (10-45) follows.

The central limit theorem states that a large sum of independent random variables each with finite variance tends to behave like a normal random variable. Thus the individual p.d.fs become unimportant to analyze the collective sum behavior. If we model the noise phenomenon as the sum of a large number of independent random variables (eg: electron motion in resistor components), then this theorem allows us to conclude that noise behaves like a Gaussian r.v.

It may be remarked that the finite variance assumption is necessary for the theorem to hold good. To prove its importance, consider the r.vs to be Cauchy distributed, and let

$$Y = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}. \tag{10-53}$$

where each $X_i \sim C(\alpha)$. Then since

$$\Phi_{X_i}(u) = e^{-\alpha |u|}, \tag{10-54}$$

substituting this into (10-48), we get

$$\Phi_Y(u) = \prod_{i=1}^{n} \Phi_X(u/\sqrt{n}) = e^{-\alpha |u|/\sqrt{n}} \sim C(\alpha/\sqrt{n}), \tag{10-55}$$

which shows that Y is still Cauchy with parameter $\alpha/\sqrt{n}$. In other words, central limit theorem doesn't hold good for a set of Cauchy r.vs as their variances are undefined.

Joint characteristic functions are useful in determining the p.d.f of linear combinations of r.vs. For example, with X and Y as independent Poisson r.vs with parameters λ_1 and λ_2 respectively, let

$$Z = X + Y. \tag{10-56}$$

Then

$$\Phi_Z(u) = \Phi_X(u)\Phi_Y(u). \tag{10-57}$$

But from (6-33)

$$\Phi_X(u) = e^{\lambda_1(e^{ju}-1)}, \quad \Phi_Y(u) = e^{\lambda_2(e^{ju}-1)} \tag{10-58}$$

so that

$$\Phi_Z(u) = e^{(\lambda_1+\lambda_2)(e^{ju}-1)} \quad \sim \quad P(\lambda_1 + \lambda_2) \tag{10-59}$$

i.e., sum of independent Poisson r.vs is also a Poisson random variable.

11. Conditional Density Functions and Conditional Expected Values

As we have seen in section 4 conditional probability density functions are useful to update the information about an event based on the knowledge about some other related event (refer to example 4.7). In this section, we shall analyze the situation where the related event happens to be a random variable that is dependent on the one of interest.

From (4-11), recall that the distribution function of X given an event B is

$$F_X(x \mid B) = P\big(X(\xi) \le x \mid B\big) = \frac{P\big((X(\xi) \le x) \cap B\big)}{P(B)}. \qquad (11\text{-}1)$$

Suppose, we let

$$B = \{ y_1 < Y(\xi) \le y_2 \}. \tag{11-2}$$

Substituting (11-2) into (11-1), we get

$$F_X(x \mid y_1 < Y \le y_2) = \frac{P\big(X(\xi) \le x, y_1 < Y(\xi) \le y_2\big)}{P(y_1 < Y(\xi) \le y_2)}$$

$$= \frac{F_{XY}(x, y_2) - F_{XY}(x, y_1)}{F_Y(y_2) - F_Y(y_1)}, \tag{11-3}$$

where we have made use of (7-4). But using (3-28) and (7-7) we can rewrite (11-3) as

$$F_X(x \mid y_1 < Y \le y_2) = \frac{\int_{-\infty}^{x} \int_{y_1}^{y_2} f_{XY}(u, v)\, du dv}{\int_{y_1}^{y_2} f_Y(v)\, dv}. \tag{11-4}$$

To determine, the limiting case $F_X(x \mid Y = y)$, we can let $y_1 = y$ and $y_2 = y + \Delta y$ in (11-4).

This gives

$$F_X(x \mid y < Y \le y+\Delta y) = \frac{\int_{-\infty}^{x}\int_{y}^{y+\Delta y} f_{XY}(u,v)\,du\,dv}{\int_{y}^{y+\Delta y} f_Y(v)\,dv} \approx \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du\,\Delta y}{f_Y(y)\,\Delta y} \tag{11-5}$$

and hence in the limit

$$F_X(x \mid Y = y) = \lim_{\Delta y \to 0} F_X(x \mid y < Y \le y+\Delta y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du}{f_Y(y)}. \tag{11-6}$$

(To remind about the conditional nature on the left hand side, we shall use the subscript $X \mid Y$ (instead of X) there). Thus

$$F_{X|Y}(x \mid Y = y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)\,du}{f_Y(y)}. \tag{11-7}$$

Differentiating (11-7) with respect to x using (8-7), we get

$$f_{X|Y}(x \mid Y = y) = \frac{f_{XY}(x,y)}{f_Y(y)}. \tag{11-8}$$

It is easy to see that the left side of (11-8) represents a valid probability density function. In fact

$$f_X(x \mid Y = y) = \frac{f_{XY}(x, y)}{f_Y(y)} \geq 0 \tag{11-9}$$

and

$$\int_{-\infty}^{+\infty} f_{X|Y}(x \mid Y = y)dx = \frac{\int_{-\infty}^{+\infty} f_{XY}(x, y)dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1, \tag{11-10}$$

where we have made use of (7-14). From (11-9) - (11-10), (11-8) indeed represents a valid p.d.f, and we shall refer to it as the conditional p.d.f of the r.v X given $Y = y$. We may also write

$$f_{X|Y}(x \mid Y = y) = f_{X|Y}(x \mid y). \tag{11-11}$$

From (11-8) and (11-11), we have

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}, \tag{11-12}$$

and similarly

$$f_{Y|X}(y\,|\,x) = \frac{f_{XY}(x,y)}{f_X(x)}. \tag{11-13}$$

If the r.vs X and Y are independent, then $f_{XY}(x,y) = f_X(x) f_Y(y)$ and (11-12) - (11-13) reduces to

$$f_{X|Y}(x\,|\,y) = f_X(x), \qquad f_{Y|X}(y\,|\,x) = f_Y(y), \tag{11-14}$$

implying that the conditional p.d.fs coincide with their unconditional p.d.fs. This makes sense, since if X and Y are independent r.vs, information about Y shouldn't be of any help in updating our knowledge about X.

In the case of discrete-type r.vs, (11-12) reduces to

$$P\left(X = x_i \,|\, Y = y_j\right) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}. \tag{11-15}$$

Next we shall illustrate the method of obtaining conditional p.d.fs through an example.

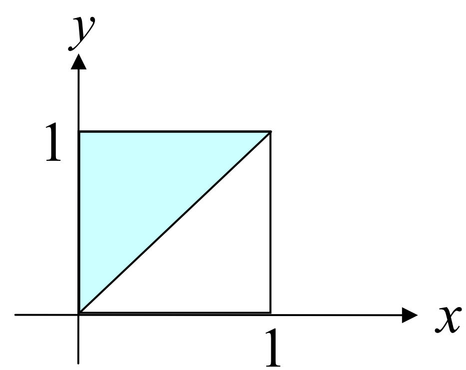


Fig. 11.1

Example 11.1: Given

$$f_{XY}(x,y) = \begin{cases} k, & 0 < x < y < 1, \\ 0, & \text{otherwise}, \end{cases} \qquad (11\text{-}16)$$

determine $f_{X|Y}(x \mid y)$ and $f_{Y|X}(y \mid x)$.

Solution: The joint p.d.f is given to be a constant in the shaded region. This gives

$$\iint f_{XY}(x,y)dxdy = \int_0^1 \int_0^y k\ dx\ dy = \int_0^1 k\ y\ dy = \frac{k}{2} = 1 \Rightarrow k = 2.$$

Similarly

$$f_X(x) = \int f_{XY}(x,y)dy = \int_x^1 k\ dy = k\,(1-x), \quad 0 < x < 1, \qquad (11\text{-}17)$$

and

$$f_Y(y) = \int f_{XY}(x,y)dx = \int_0^y k\ dx = k\ y, \quad 0 < y < 1. \qquad (11\text{-}18)$$

From (11-16) - (11-18), we get

$$f_{X|Y}(x\,|\,y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\frac{1}{y},\qquad 0<x<y<1, \tag{11-19}$$

and

$$f_{Y|X}(y\,|\,x)=\frac{f_{XY}(x,y)}{f_X(x)}=\frac{1}{1-x},\qquad 0<x<y<1. \tag{11-20}$$

We can use (11-12) - (11-13) to derive an important result. From there, we also have

$$f_{XY}(x,y)=f_{X|Y}(x\,|\,y)f_Y(y)=f_{Y|X}(y\,|\,x)f_X(x) \tag{11-21}$$

or

$$f_{Y|X}(y\,|\,x)=\frac{f_{X|Y}(x\,|\,y)f_Y(y)}{f_X(x)}. \tag{11-22}$$

But

$$f_X(x)=\int_{-\infty}^{+\infty}f_{XY}(x,y)dy=\int_{-\infty}^{+\infty}f_{X|Y}(x\,|\,y)f_Y(y)dy \tag{11-23}$$

and using (11-23) in (11-22), we get

$$f_{Y|X}(y \mid x) = \frac{f_{X|Y}(x \mid y) f_Y(y)}{\int_{-\infty}^{+\infty} f_{X|Y}(x \mid y) f_Y(y) dy}. \tag{24}$$

Equation (11-24) represents the p.d.f version of Bayes' theorem. To appreciate the full significance of (11-24), one need to look at communication problems where observations can be used to update our knowledge about unknown parameters. We shall illustrate this using a simple example.

Example 11.2: An unknown random phase θ is uniformly distributed in the interval $(0, 2\pi)$, and $r = \theta + n$, where $n \sim N(0, \sigma^2)$. Determine $f(\theta \mid r)$.

Solution: Initially almost nothing about the r.v θ is known, so that we assume its a-priori p.d.f to be uniform in the interval $(0, 2\pi)$.

In the equation $r = \theta + n,$ we can think of n as the noise contribution and r as the observation. It is reasonable to assume that θ and n are independent. In that case

$$f(r \,|\, \boldsymbol{\theta} = \theta) \sim N(\theta, \sigma^2) \tag{11-25}$$

since it is given that $\boldsymbol{\theta} = \theta$ is a constant, $r = \theta + n$ behaves like n. Using (11-24), this gives the a-posteriori p.d.f of θ given r to be (see Fig. 11.2 (b))

$$f(\theta \,|\, r) = \frac{f(r \,|\, \theta) f_\theta(\theta)}{\int_0^{2\pi} f(r \,|\, \theta) f_\theta(\theta) d\theta} = \frac{e^{-(r-\theta)^2/2\sigma^2}}{\frac{1}{2\pi}\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta}$$

$$= \varphi(r) e^{-(\theta - r)^2/2\sigma^2}, \qquad 0 < \theta < 2\pi, \tag{11-26}$$

where

$$\varphi(r) = \frac{2\pi}{\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta}.$$

Notice that the knowledge about the observation r is reflected in the a-posteriori p.d.f of θ in Fig. 11.2 (b). It is no longer flat as the a-priori p.d.f in Fig. 11.2 (a), and it shows higher probabilities in the neighborhood of $\theta = r$.

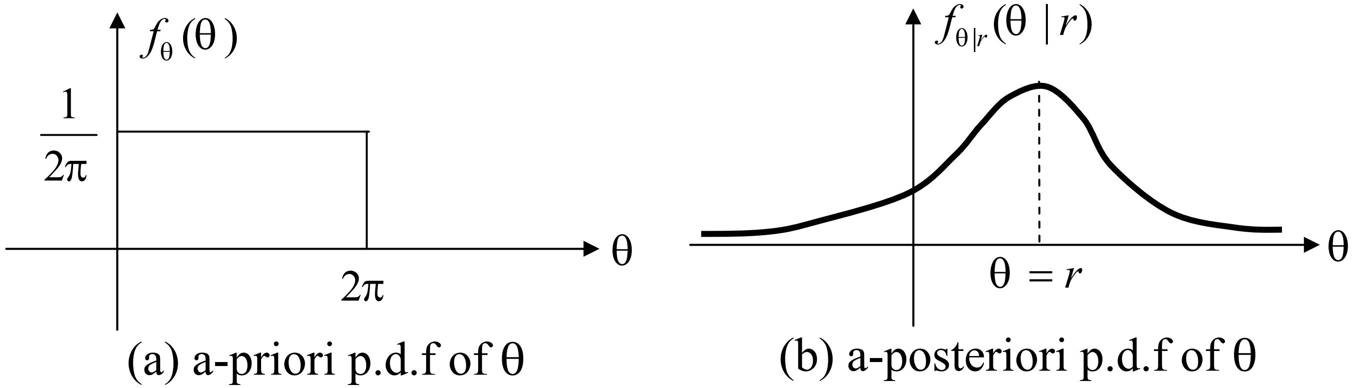


Fig. 11.2

Conditional Mean:

We can use the conditional p.d.fs to define the conditional mean. More generally, applying (6-13) to conditional p.d.fs we get

$$E\big(g(x)\,|\,B\big)=\int_{-\infty}^{+\infty} g(x)\,f_X(x\,|\,B)\,dx. \tag{11-27}$$

and using a limiting argument as in (11-2) - (11-8), we get

$$\mu_{X|Y}=E\big(X\,|\,Y=y\big)=\int_{-\infty}^{+\infty} x\, f_{X|Y}(x\,|\,y)\,dx \tag{11-28}$$

to be the conditional mean of X given $Y=y$. Notice that $E(X\,|\,Y=y)$ will be a function of y. Also

$$\mu_{Y|X}=E\big(Y\,|\,X=x\big)=\int_{-\infty}^{+\infty} y\, f_{Y|X}(y\,|\,x)\;dy. \tag{11-29}$$

In a similar manner, the conditional variance of X given $Y=y$ is given by

$$\begin{aligned} Var(X\,|\,Y)=\sigma^2_{X|Y} &= E\big(X^2\,|\,Y=y\big)-\big(E(X\,|\,Y=y)\big)^2 \\ &= E\big((X-\mu_{X|Y})^2\,|\,Y=y\big). \end{aligned} \tag{11-30}$$

we shall illustrate these calculations through an example.

Example 11.3: Let

$$f_{XY}(x,y)=\begin{cases}1, & 0<|y|<x<1,\\ 0, & \text{otherwise}.\end{cases} \tag{11-31}$$

Determine $E(X\,|\,Y)$ and $E(Y\,|\,X)$.

Solution: As Fig. 11.3 shows, $f_{XY}(x,y)=1$ in the shaded area, and zero elsewhere. From there

$$f_X(x)=\int_{-x}^{x} f_{XY}(x,y)dy=2x, \quad 0<x<1,$$

and

$$f_Y(y)=\int_{-|y|}^{1} 1\ dx = 1-|y|, \quad |y|<1,$$

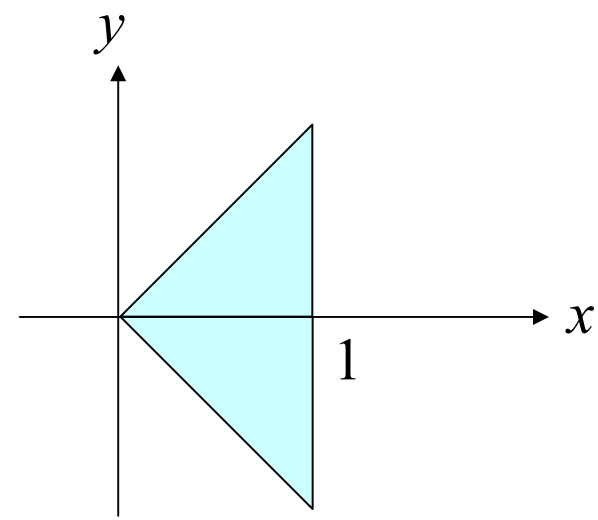


Fig. 11.3

This gives

$$f_{X|Y}(x\,|\,y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\frac{1}{1-|y|}, \quad 0<|y|<x<1, \tag{11-32}$$

and

$$f_{Y|X}(y\,|\,x)=\frac{f_{XY}(x,y)}{f_X(x)}=\frac{1}{2x}, \quad 0<|y|<x<1. \tag{11-33}$$

Hence $$E(X \mid Y)=\int x\ f_{X|Y}(x \mid y)dx=\int_{|y|}^{1}\frac{x}{(1-|y|)}dx$$

$$=\frac{1}{(1-|y|)}\left.\frac{x^2}{2}\right|_{|y|}^{1}=\frac{1-|y|^2}{2(1-|y|)}=\frac{1+|y|}{2},\quad |y|<1. \qquad (11\text{-}34)$$

$$E(Y \mid X)=\int y f_{Y|X}(y \mid x)dy=\int_{-x}^{x}\frac{y}{2x}dy=\frac{1}{2x}\left.\frac{y^2}{2}\right|_{-x}^{x}=0,\quad 0<x<1. \qquad (11\text{-}35)$$

It is possible to obtain an interesting generalization of the conditional mean formulas in (11-28) - (11-29). More generally, (11-28) gives

$$E\big(g(x) \mid y\big)=\int_{-\infty}^{+\infty} g(x) f_{X|Y}(x \mid y)dx\ . \qquad (11\text{-}36)$$

But

$$E\big(g(x)\big)=\int_{-\infty}^{+\infty} g(x) f_X(x)dx=\int_{-\infty}^{+\infty} g(x)\int_{-\infty}^{+\infty} f_{XY}(x,y)dydx$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x) f_{XY}(x,y)dxdy=\int_{-\infty}^{+\infty}\underbrace{\int_{-\infty}^{+\infty} g(x) f_{X|Y}(x \mid y)dx}_{E(g(x)|Y=y)}\ f_Y(y)dy$$

$$=\int_{-\infty}^{+\infty} E\big(g(x) \mid Y=y\big) f_Y(y)dy=E\{E\big(g(x) \mid Y=y\big)\}. \qquad (11\text{-}37)$$

Obviously, in the right side of (11-37), the inner expectation is with respect to X and the outer expectation is with respect to Y. Letting $g(X) = X$ in (11-37) we get the interesting identity

$$E(X) = E\{E(X \mid Y = y)\}, \tag{11-38}$$

where the inner expectation on the right side is with respect to X and the outer one is with respect to Y. Similarly, we have

$$E(Y) = E\{E(Y \mid X = x)\}. \tag{11-39}$$

Using (11-37) and (11-30), we also obtain

$$Var(X) = E\big(Var(X \mid Y = y)\big). \tag{11-40}$$

Conditional mean turns out to be an important concept in estimation and prediction theory. For example given an observation about a r.v X, what can be say about a related r.v Y? In other words what is the best predicted value of Y given that $X = x$? It turns out that if "best" is meant in the sense of minimizing the mean square error between Y and its estimate $\hat{y}$, then the conditional mean of Y given $X = x$, i.e., $E(Y \mid X = x)$ is the best estimate for Y (see Lecture 16 for more on Mean Square Estimation).

We conclude this lecture with yet another application of the conditional density formulation.

Example 11.4 : **Poisson sum of Bernoulli random variables**

Let $X_i, \quad i = 1, 2, 3, \cdots$ represent independent, identically distributed Bernoulli random variables with

$$P(X_i = 1) = p, \qquad P(X_i = 0) = 1 - p = q$$

and N a Poisson random variable with parameter λ that is independent of all X_i. Consider the random variables

$$Y = \sum_{i=1}^{N} X_i, \qquad Z = N - Y. \tag{11-41}$$

Show that Y and Z are independent Poisson random variables.

Solution : To determine the joint probability mass function of Y and Z, consider

$$\begin{aligned} P(Y=m,\ Z=n) &= P(Y=m,\ N-Y=n) = P(Y=m,\ N=m+n) \\ &= P(Y=m \mid N=m+n) P(N=m+n) \\ &= P(\sum_{i=1}^{N} X_i = m \mid N=m+n)\, P(N=m+n) \\ &= P(\sum_{i=1}^{m+n} X_i = m) P(N=m+n) \qquad (11\text{-}42) \end{aligned}$$

(Note that $\sum_{i=1}^{m+n} X_i \sim B(m+n,\ p)$ and $X_i s$ are independent of N)

$$= \left(\frac{(m+n)!}{m!n!} p^m q^n \right) \left(e^{-\lambda} \frac{\lambda^{m+n}}{(m+n)} \right)$$

$$= \left(e^{-p\lambda} \frac{(p\lambda)^m}{m!} \right) \left(e^{-q\lambda} \frac{(q\lambda)^n}{n!} \right)$$

$$= P(Y=m)P(Z=n). \tag{11-43}$$

Thus

$$Y \sim P(p\lambda) \quad \text{and} \quad Z \sim P(q\lambda) \tag{11-44}$$

and Y and Z are independent random variables. Thus if a bird lays eggs that follow a Poisson random variable with parameter λ , and if each egg survives with probability p, then the number of chicks that survive also forms a Poisson random variable with parameter $p\lambda$.

12. Principles of Parameter Estimation

The purpose of this lecture is to illustrate the usefulness of the various concepts introduced and studied in earlier lectures to practical problems of interest. In this context, consider the problem of estimating an unknown parameter of interest from a few of its noisy observations. For example, determining the daily temperature in a city, or the depth of a river at a particular spot, are problems that fall into this category.

Observations (measurement) are made on data that contain the desired nonrandom parameter θ and undesired noise. Thus, for example,

$$\text{Observation} = \text{signal (desired part)} + \text{noise}, \qquad (12\text{-}1)$$

or, the i th observation can be represented as

$$X_i = \theta + n_i, \quad i = 1, 2, \cdots, n. \qquad (12\text{-}2)$$

Here θ represents the unknown nonrandom desired parameter, and $n_i, \quad i = 1, 2, \cdots, n$ represent random variables that may be dependent or independent from observation to observation. Given n observations $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$, the estimation problem is to obtain the "best" estimator for the unknown parameter θ in terms of these observations.

Let us denote by $\hat{\theta}(X)$ the estimator for θ. Obviously $\hat{\theta}(X)$ is a function of only the observations. "Best estimator" in what sense? Various optimization strategies can be used to define the term "best".

Ideal solution would be when the estimate $\hat{\theta}(X)$ coincides with the unknown θ. This of course may not be possible, and almost always any estimate will result in an error given by

$$e = \hat{\theta}(X) - \theta . \tag{12-3}$$

One strategy would be to select the estimator $\hat{\theta}(X)$ so as to minimize some function of this error - such as - minimization of the mean square error (MMSE), or minimization of the absolute value of the error etc.

A more fundamental approach is that of the **principle of Maximum Likelihood** (ML).

The underlying assumption in any estimation problem is that the available data $X_1, X_2, \cdots, X_n$ has something to do with the unknown parameter θ. More precisely, we assume that the joint p.d.f of $X_1, X_2, \cdots, X_n$ given by $f_X(x_1, x_2, \cdots, x_n; \theta)$ depends on θ. The method of maximum likelihood assumes that the given sample data set is representative of the population $f_X(x_1, x_2, \cdots, x_n; \theta)$, and chooses that value for θ that most likely caused the observed data to occur, i.e., once observations $x_1, x_2, \cdots, x_n$ are given, $f_X(x_1, x_2, \cdots, x_n; \theta)$ is a function of θ alone, and the value of θ that maximizes the above p.d.f is the most likely value for θ, and it is chosen as the ML estimate $\hat{\theta}_{ML}(X)$ for θ (Fig. 12.1).

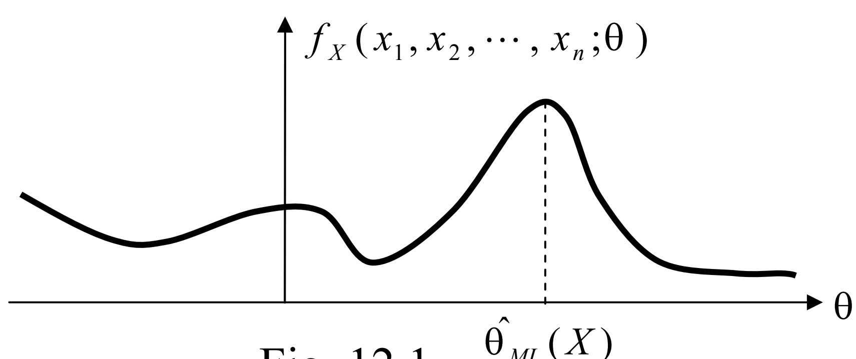


Fig. 12.1

Given $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n,$ the joint p.d.f $f_X(x_1, x_2, \cdots, x_n; \theta)$ represents the likelihood function, and the ML estimate can be determined either from the likelihood equation

$$\sup_{\hat{\theta}_{ML}} f_X(x_1, x_2, \cdots, x_n; \theta) \qquad (12\text{-}4)$$

or using the log-likelihood function (sup in (12-4) represents the supremum operation)

$$L(x_1, x_2, \cdots, x_n; \theta) \stackrel{\Delta}{=} \log f_X(x_1, x_2, \cdots, x_n; \theta). \qquad (12\text{-}5)$$

If $L(x_1, x_2, \cdots, x_n; \theta)$ is differentiable and a supremum $\hat{\theta}_{ML}$ exists in (12-5), then that must satisfy the equation

$$\left. \frac{\partial \log f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta} \right|_{\theta = \hat{\theta}_{ML}} = 0. \qquad (12\text{-}6)$$

We will illustrate the above procedure through several examples:

Example 12.1: Let $X_i = \theta + w_i, \ \ i = 1 \to n,$ represent n observations where θ is the unknown parameter of interest, and $w_i, \ \ i = 1 \to n,$ are zero mean independent normal r.vs with common variance σ^2. Determine the ML estimate for θ.

Solution: Since w_i are independent r.vs and θ is an unknown constant, we have X_i s are independent normal random variables. Thus the likelihood function takes the form

$$f_X(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta). \tag{12-7}$$

Moreover, each X_i is Gaussian with mean θ and variance σ^2 (Why?). Thus

$$f_{X_i}(x_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \theta)^2 / 2\sigma^2}. \tag{12-8}$$

Substituting (12-8) into (12-7) we get the likelihood function to be

$$f_X(x_1, x_2, \cdots, x_n; \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n} (x_i - \theta)^2 / 2\sigma^2}. \tag{12-9}$$

It is easier to work with the log-likelihood function $L(X;\theta)$ in this case. From (12-9)

$$L(X;\theta) = \ln f_X(x_1, x_2, \cdots, x_n; \theta) = \frac{n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(x_i - \theta)^2}{2\sigma^2}, \tag{12-10}$$

and taking derivative with respect to θ as in (12-6), we get

$$\left.\frac{\partial \ln f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta}\right|_{\theta = \hat{\theta}_{ML}} = 2\sum_{i=1}^{n} \left.\frac{(x_i - \theta)}{2\sigma^2}\right|_{\theta = \hat{\theta}_{ML}} = 0, \tag{12-11}$$

or

$$\hat{\theta}_{ML}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{12-12}$$

Thus (12-12) represents the ML estimate for θ, which happens to be a linear estimator (linear function of the data) in this case.

Notice that the estimator is a r.v. Taking its expected value, we get

$$E[\hat{\theta}_{ML}(x)] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \theta, \qquad (12\text{-}13)$$

i.e., the expected value of the estimator does not differ from the desired parameter, and hence there is no bias between the two. Such estimators are known as unbiased estimators. Thus (12-12) represents an unbiased estimator for θ. Moreover the variance of the estimator is given by

$$Var(\hat{\theta}_{ML}) = E[(\hat{\theta}_{ML} - \theta)^2] = \frac{1}{n^2} E\left\{\left(\sum_{i=1}^{n} X_i - \theta\right)^2\right\}$$

$$= \frac{1}{n^2}\left\{\sum_{i=1}^{n} E(X_i - \theta)^2 + \sum_{i=1}^{n}\ \sum_{j=1,\ i\neq j}^{n} E(X_i - \theta)(X_j - \theta)\right\}.$$

The later terms are zeros since X_i and X_j are independent r.vs.

Then

$$Var(\hat{\theta}_{ML}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \tag{12-14}$$

Thus

$$Var(\hat{\theta}_{ML}) \to 0 \quad \text{as} \quad n \to \infty, \tag{12-15}$$

another desired property. We say such estimators (that satisfy (12-15)) are consistent estimators.

Next two examples show that ML estimator can be highly nonlinear.

Example 12.2: Let $X_1, X_2, \cdots, X_n$ be independent, identically distributed uniform random variables in the interval $(0, \theta)$ with common p.d.f

$$f_{X_i}(x_i;\theta) = \frac{1}{\theta}, \quad 0 < x_i < \theta, \tag{12-16}$$

where θ is an unknown parameter. Find the ML estimate for θ.

Solution: The likelihood function in this case is given by

$$f_X(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n; \theta) = \frac{1}{\theta^n}, \quad 0 < x_i \le \theta, \quad i = 1 \to n$$

$$= \frac{1}{\theta^n}, \quad 0 \le \max(x_1, x_2, \cdots, x_n) \le \theta. \quad \text{(12-17)}$$

From (12-17), the likelihood function in this case is maximized by the minimum value of θ, and since $\theta \ge \max(X_1, X_2, \cdots, X_n)$, we get

$$\hat{\theta}_{ML}(X) = \max(X_1, X_2, \cdots, X_n) \quad \text{(12-18)}$$

to be the ML estimate for θ. Notice that (18) represents a nonlinear function of the observations. To determine whether (12-18) represents an unbiased estimate for θ, we need to evaluate its mean. To accomplish that in this case, it is easier to determine its p.d.f and proceed directly. Let

$$Z = \max(X_1, X_2, \cdots, X_n) \qquad (12\text{-}19)$$

with X_i as in (12-16). Then

$$F_Z(z) = P[\max(X_1, X_2, \cdots, X_n) \le z] = P(X_1 \le z, X_2 \le z, \cdots, X_n \le z)$$

$$= \prod_{i=1}^{n} P(X_i \le z) = \prod_{i=1}^{n} F_{X_i}(z) = \left(\frac{z}{\theta}\right)^n, \quad 0 < z < \theta, \qquad (12\text{-}20)$$

so that

$$f_Z(z) = \begin{cases} \dfrac{nz^{n-1}}{\theta^n}, & 0 < z < \theta, \\ 0, & \text{otherwise} \end{cases}. \qquad (12\text{-}21)$$

Using (12-21), we get

$$E[\hat{\theta}_{ML}(X)] = E(Z) = \int_0^\theta z\, f_Z(z)dz = \frac{n}{\theta^n}\int_0^\theta z^n dz = \frac{n}{n+1}\frac{\theta^{n+1}}{\theta^n} = \frac{\theta}{(1+1/n)}. \qquad (12\text{-}22)$$

In this case $E[\hat{\theta}_{ML}(X)] \ne \theta$, and hence the ML estimator is not an unbiased estimator for θ. However, from (12-22) as $n \to \infty$

$$\lim_{n\to\infty} E[\hat{\theta}_{ML}(X)] = \lim_{n\to\infty} \frac{\theta}{(1+1/n)} = \theta, \tag{12-23}$$

i.e., the ML estimator is an asymptotically unbiased estimator. From (12-21), we also get

$$E(Z^2) = \int_0^\theta z^2 f_Z(z)dz = \frac{n}{\theta^n}\int_0^\theta z^{n+1}dz = \frac{n\theta^2}{n+2} \tag{12-24}$$

so that

$$Var[\hat{\theta}_{ML}(X)] = E(Z^2) - [E(Z)]^2 = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} = \frac{n\theta^2}{(n+1)^2(n+2)}. \tag{12-25}$$

Once again $Var[\hat{\theta}_{ML}(X)] \to 0$ as $n \to \infty$, implying that the estimator in (12-18) is a consistent estimator.

Example 12.3: Let $X_1, X_2, \cdots, X_n$ be i.i.d Gamma random variables with unknown parameters α and β. Determine the ML estimator for α and β.

Solution: Here $x_i \geq 0,$ and

$$f_X(x_1, x_2, \cdots, x_n; \alpha, \beta) = \frac{\beta^{n\alpha}}{(\Gamma(\alpha))^n} \prod_{i=1}^{n} x_i^{\alpha-1} \; e^{-\beta \sum_{i=1}^{n} x_i}. \tag{12-26}$$

This gives the log-likelihood function to be

$$L(x_1, x_2, \cdots, x_n; \alpha, \beta) = \log f_X(x_1, x_2, \cdots, x_n; \alpha, \beta)$$

$$= n\alpha \log \beta - n \log \Gamma(\alpha) + (\alpha - 1)\left(\sum_{i=1}^{n} \log x_i\right) - \beta \sum_{i=1}^{n} x_i. \tag{12-27}$$

Differentiating L with respect to α and β we get

$$\frac{\partial L}{\partial \alpha} = n \log \beta - \frac{n}{\Gamma(\alpha)} \Gamma'(\alpha) + \sum_{i=1}^{n} \log x_i \Big|_{\alpha, \beta = \hat{\alpha}, \hat{\beta}} = 0, \tag{12-28}$$

$$\frac{\partial L}{\partial \beta} = \frac{n\alpha}{\beta} - \sum_{i=1}^{n} x_i \Big|_{\alpha, \beta = \hat{\alpha}, \hat{\beta}} = 0. \tag{12-29}$$

Thus from (12-29)

$$\hat{\beta}_{ML}(X) = \frac{\hat{\alpha}_{ML}}{\frac{1}{n}\sum_{i=1}^{n} x_i}, \tag{12-30}$$

and substituting (12-30) into (12-28), it gives

$$\log \hat{\alpha}_{ML} - \frac{\Gamma'(\hat{\alpha}_{ML})}{\Gamma(\hat{\alpha}_{ML})} = \log\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{12-31}$$

Notice that (12-31) is highly nonlinear in $\hat{\alpha}_{ML}$.

In general the (log)-likelihood function can have more than one solution, or no solutions at all. Further, the (log)-likelihood function may not be even differentiable, or it can be extremely complicated to solve explicitly (see example 12.3, equation (12-31)).

Best Unbiased Estimator:

Referring back to example 12.1, we have seen that (12-12) represents an unbiased estimator for θ with variance given by (12-14). It is possible that, for a given n, there may be other unbiased estimators to this problem with even lower variances. If such is indeed the case, those estimators will be naturally preferable compared to (12-12). In a given scenario, is it possible to determine the lowest possible value for the variance of *any* unbiased estimator? Fortunately, a theorem by Cramer and Rao (Rao 1945; Cramer 1948) gives a complete answer to this problem.

Cramer - Rao Bound: Variance of any unbiased estimator $\hat{\theta}$ based on observations $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$ for θ must satisfy the lower bound

$$Var(\hat{\theta}) \geq \frac{1}{E\left(\frac{\partial \ln f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta}\right)^2} = \frac{-1}{E\left(\frac{\partial^2 \ln f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta^2}\right)}. \quad (12\text{-}32)$$

This important result states that the right side of (12-32) acts as a lower bound on the variance of *all* unbiased estimator for θ, provided their joint p.d.f satisfies certain regularity restrictions. (see (8-79)-(8-81), Text).

Naturally any unbiased estimator whose variance coincides with that in (12-32), must be the best. There are no better solutions! Such estimates are known as *efficient* estimators. Let us examine whether (12-12) represents an efficient estimator. Towards this using (12-11)

$$\left(\frac{\partial \ln f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta} \right)^2 = \frac{1}{\sigma^4} \left(\sum_{i=1}^{n} (X_i - \theta) \right)^2; \tag{12-33}$$

and

$$E\left(\frac{\partial \ln f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta} \right)^2 = \frac{1}{\sigma^4} \left\{ \sum_{i=1}^{n} E[(X_i - \theta)^2] + \sum_{i=1}^{n} \sum_{j=1, i \neq j}^{n} E[(X_i - \theta)(X_j - \theta)] \right\}$$

$$= \frac{1}{\sigma^4} \sum_{i=1}^{n} \sigma^2 = \frac{n}{\sigma^2}, \tag{12-34}$$

and substituting this into the first form on the right side of (12-32), we obtain the Cramer - Rao lower bound for this problem to be

$$\frac{\sigma^2}{n}. \tag{12-35}$$

But from (12-14) the variance of the ML estimator in (12-12) is the same as (12-35), implying that (12-12) indeed represents an efficient estimator in this case, the best of all possibilities!

It is possible that in certain cases there are no unbiased estimators that are efficient. In that case, the best estimator will be an unbiased estimator with the lowest possible variance.
How does one find such an unbiased estimator?
Fortunately Rao-Blackwell theorem (page 335-337, Text) gives a complete answer to this problem.
Cramer-Rao bound can be extended to multiparameter case as well (see page 343-345,Text).

So far, we discussed nonrandom parameters that are unknown. What if the parameter of interest is a r.v with a-priori p.d.f $f_\theta(\theta)$? How does one obtain a good estimate for θ based on the observations $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$?

One technique is to use the observations to compute its a-posteriori probability density function $f_{\theta|X}(\theta \mid x_1, x_2, \cdots, x_n)$. Of course, we can use the Bayes' theorem in (11.22) to obtain this a-posteriori p.d.f. This gives

$$f_{\theta|X}(\theta \mid x_1, x_2, \cdots, x_n) = \frac{f_{X|\theta}(x_1, x_2, \cdots, x_n \mid \theta) f_\theta(\theta)}{f_X(x_1, x_2, \cdots, x_n)}. \qquad (12\text{-}36)$$

Notice that (12-36) is only a function of θ, since $x_1, x_2, \cdots, x_n$ represent given observations. Once again, we can look for the most probable value of θ suggested by the above a-posteriori p.d.f. Naturally, the most likely value for θ is that corresponding to the maximum of the a-posteriori p.d.f (see Fig. 12.2). This estimator - maximum of the a-posteriori p.d.f is known as the MAP estimator for θ.

It is possible to use other optimality criteria as well. Of course, that should be the subject matter of another course!

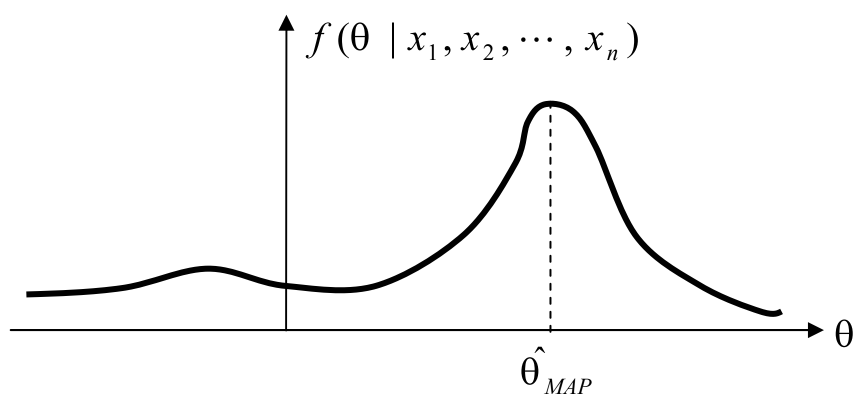


Fig. 12.2

13. The Weak Law and the Strong Law of Large Numbers

James Bernoulli proved the weak law of large numbers (WLLN) around 1700 which was published posthumously in 1713 in his treatise *Ars Conjectandi.* Poisson generalized Bernoulli's theorem around 1800, and in 1866 Tchebychev discovered the method bearing his name. Later on one of his students, Markov observed that Tchebychev's reasoning can be used to extend Bernoulli's theorem to dependent random variables as well.

In 1909 the French mathematician Emile Borel proved a deeper theorem known as the strong law of large numbers that further generalizes Bernoulli's theorem. In 1926 Kolmogorov derived conditions that were necessary and sufficient for a set of mutually independent random variables to obey the law of large numbers.

Let X_i be independent, identically distributed Bernoulli random Variables such that

$$P(X_i) = p, \qquad P(X_i = 0) = 1 - p = q,$$

and let $k = X_1 + X_2 + \cdots + X_n$ represent the number of "successes" in n trials. Then the weak law due to Bernoulli states that [see Theorem 3-1, page 58, Text]

$$P\left\{ \left| \tfrac{k}{n} - p \right| > \varepsilon \right\} \leq \frac{pq}{n\varepsilon^2}. \tag{13-1}$$

i.e., the ratio "total number of successes to the total number of trials" tends to p *in probability* as n increases.

A stronger version of this result due to Borel and Cantelli states that the above ratio k/n tends to p not only *in probability*, but *with probability* 1. This is the strong law of large numbers (SLLN).

What is the difference between the weak law and the strong law?

The strong law of large numbers states that if $\{\varepsilon_n\}$ is a sequence of positive numbers converging to zero, then

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{k}{n} - p\right| \geq \varepsilon_n\right\} < \infty. \tag{13-2}$$

From Borel-Cantelli lemma [see (2-69) Text], when (13-2) is satisfied the events $A_n \triangleq \left\{ \left|\frac{k}{n} - p\right| \geq \varepsilon_n \right\}$ can occur only for a finite number of indices *n* in an infinite sequence, or equivalently, the events $\left\{\left|\frac{k}{n} - p\right| < \varepsilon_n\right\}$ occur infinitely often, i.e., the event *k/n* converges to *p almost-surely*.

Proof: To prove (13-2), we proceed as follows. Since

$$\left|\frac{k}{n} - p\right| \geq \varepsilon \quad \Rightarrow \quad |k - np|^4 \geq \varepsilon^4 n^4$$

we have

$$\sum_{k=0}^{n}(k-np)^4 p_n(k) \geq \varepsilon^4 n^4 = \varepsilon^4 n^4 \Big(P\Big\{\left|\tfrac{k}{n}-p\right| \geq \varepsilon\Big\} + P\Big\{\left|\tfrac{k}{n}-p\right| < \varepsilon\Big\}\Big)$$

and hence

$$P\Big\{\left|\tfrac{k}{n}-p\right| \geq \varepsilon\Big\} \leq \frac{\sum_{k=0}^{n}(k-np)^4 p_n(k)}{\varepsilon^4 n^4} \tag{13-3}$$

where

$$p_n(k) = P\left\{\sum_{i=1}^{n} X_i = k\right\} = \binom{n}{k} p^k q^{n-k}$$

By direct computation

$$\sum_{k=0}^{n}(k-np)^4 p_n(k) = E\Big\{(\sum_{i=1}^{n} X_i - np)^4\Big\} = E\Big\{\Big(\sum_{i=1}^{n}(X_i - p)\Big)^4\Big\}$$

$$= E\{(\sum_{i=1}^{n} Y_i)^4\} = \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n} E(Y_i Y_k Y_j Y_l)$$

$i = 1 \rightarrow n$ can coincide with j, k or l, and the second variable takes (n-1) values

$$= \sum_{i=1}^{n} E(Y_i^4) + 4n(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} E(Y_i^3)\underbrace{E(Y_j)}_{0} + 3n(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} E(Y_i^2)E(Y_j^2)$$

$$= n(p^3+q^3)pq + 3n(n-1)(pq)^2 \le [n+3n(n-1)]pq$$

$$= 3n^2 pq, \qquad (13\text{-}4)$$

since

$$p^3+q^3 = (p+q)^3 - 3p^2q - 3pq^2 < 1, \quad pq \le 1/2 < 1$$

Substituting (13-4) also (13-3) we obtain

$$P\left\{\left|\tfrac{k}{n} - p\right| \ge \varepsilon\right\} \le \frac{3pq}{n^2\varepsilon^4}$$

Let $\varepsilon = \frac{1}{n^{1/8}}$ so that the above integral reads
and hence

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{k}{n} - p\right| \ge \frac{1}{n^{1/8}}\right\} \le 3pq\sum_{n=1}^{\infty}\frac{1}{n^{3/2}} \le 3pq\left(1 + \int_1^{\infty} x^{-3/2}dx\right)$$

$$= 3pq(1+2) = 9pq < \infty, \qquad (13\text{-}5)$$

thus proving the strong law by exhibiting a sequence of positive numbers $\varepsilon_n = 1/n^{1/8}$ that converges to zero and satisfies (13-2).

We return back to the same question: "What is the difference between the weak law and the strong law?."

The weak law states that for every *n* that is large enough, the ratio $(\sum_{i=1}^{n} X_i)/n = k/n$ is likely to be near *p* with certain probability that tends to 1 as *n* increases. However, it does not say that *k/n* is bound to stay near *p* if the number of trials is increased. Suppose (13-1) is satisfied for a given ε in a certain number of trials n_0. If additional trials are conducted beyond n_0, the weak law does not guarantee that the new *k/n* is bound to stay near *p* for such trials. In fact there can be events for which $k/n > p+\varepsilon$, for $n > n_0$ in some regular manner. The probability for such an event is the sum of a large number of very small probabilities, and the weak law is unable to say anything specific about the convergence of that sum.

However, the strong law states (through (13-2)) that not only all such sums converge, but the total number of all such events where $k/n > p+\varepsilon$ is in fact finite! This implies that the probability $\{|\frac{k}{n} - p| > \varepsilon\}$ of the events as n increases becomes and remains small, since with probability 1 only finitely many violations to the above inequality takes place as $n \to \infty$.

Interestingly, if it possible to arrive at the same conclusion using a powerful bound known as *Bernstein's inequality* that is based on the WLLN.

Bernstein's inequality : Note that

$$\left|\frac{k}{n} - p\right| > \varepsilon \qquad \Rightarrow \qquad k > n(p+\varepsilon)$$

and for any $\lambda > 0$, this gives $e^{\lambda(k-n(p+\varepsilon))} > 1.$

Thus

$$P\{\tfrac{k}{n} - p > \varepsilon\} = \sum_{k=\lceil n(p+\varepsilon)\rceil}^{n} \binom{n}{k} p^k q^{n-k}$$

$$\leq \sum_{k=\lceil n(p+\varepsilon)\rceil}^{n} e^{\lambda(k-n(p+\varepsilon))} \binom{n}{k} p^k q^{n-k}$$

$$\leq \sum_{k=0}^{n} e^{\lambda(k-n(p+\varepsilon))} \binom{n}{k} p^k q^{n-k}$$

$$P\{\tfrac{k}{n}-p>\varepsilon\}=e^{-\lambda n\varepsilon}\sum_{k=0}^{n}\binom{n}{k}(pe^{\lambda q})^k(qe^{-\lambda p})^{n-k}$$

$$=e^{-\lambda n\varepsilon}(pe^{\lambda q}+qe^{-\lambda p})^n. \tag{13-6}$$

Since $e^x \le x+e^{x^2}$ for any real x,

$$pe^{\lambda q}+qe^{-\lambda p}\le p(\lambda q+e^{\lambda^2q^2})+q(-\lambda p+e^{\lambda^2p^2})$$

$$=pe^{\lambda^2q^2}+qe^{\lambda^2p^2}\le e^{\lambda^2}. \tag{13-7}$$

Substituting (13-7) into (13-6), we get

$$P\{\tfrac{k}{n}-p>\varepsilon\}\le e^{\lambda^2n-\lambda n\varepsilon}.$$

But $\lambda^2 n-\lambda n\varepsilon$ is minimum for $\lambda=\varepsilon/2$ and hence

$$P\{\tfrac{k}{n}-p>\varepsilon\}\le e^{-n\varepsilon^2/4},\qquad \varepsilon>0. \tag{13-8}$$

Similarly

$$P\{\tfrac{k}{n}-p<-\varepsilon\}\le e^{-n\varepsilon^2/4}$$

and hence we obtain *Bernstein's inequality*

$$P\{\left|\tfrac{k}{n}-p\right|>\varepsilon\}\le 2\,e^{-n\varepsilon^2/4}. \qquad (13\text{-}9)$$

Bernstein's inequality is more powerful than Tchebyshev's inequality as it states that the chances for the relative frequency *k /n* exceeding its probability *p* tends to zero exponentially fast as $n\to\infty$.

Chebyshev's inequality gives the probability of *k /n* to lie between $p-\varepsilon$ and $p+\varepsilon$ for a specific *n*. We can use Bernstein's inequality to estimate the probability for *k /n* to lie between $p-\varepsilon$ and $p+\varepsilon$ for *all* large *n*.

Towards this, let

$$y_n=\{p-\varepsilon\le \tfrac{k}{n}<p+\varepsilon\}$$

so that

$$P(y_n^c)=P\{\left|\tfrac{n}{k}-p\right|>\varepsilon\}\le 2\,e^{-n\varepsilon^2/4}$$

To compute the probability of the event $\bigcap_{n=m}^{\infty} y_n,$ note that its complement is given by $(\bigcap_{n=m}^{\infty} y_n)^c=\bigcup_{n=m}^{\infty} y_n^c$

and using Eq. (2-68) Text,

$$P(\bigcup_{n=m}^{\infty} y_n^c) \leq \sum_{n=m}^{\infty} P(y_n^c) \leq \sum_{n=m}^{\infty} 2\, e^{-n\varepsilon^2/4} = \frac{2e^{-m\varepsilon^2/4}}{1-e^{-\varepsilon^2/4}}.$$

This gives

$$P(\bigcap_{n=m}^{\infty} y_n) = \{1 - P(\bigcup_{n=m}^{\infty} \overline{y}_n)\} \geq 1 - \frac{2\, e^{-m\varepsilon^2/4}}{1-e^{-\varepsilon^2/4}} \to 1 \quad \text{as} \quad m \to \infty$$

or,

$$P\{p - \varepsilon \leq \tfrac{k}{n} \leq p + \varepsilon, \;\; \textit{for all} \;\; n \geq m\} \to 1 \;\; \text{as} \;\; m \to \infty.$$

Thus k/n is bound to stay near p for all large enough n, in probability, a conclusion already reached by the SLLN.

Discussion: Let $\varepsilon = 0.1$. Thus if we toss a fair coin 1,000 times, from the weak law

$$P\left\{ \left| \tfrac{k}{n} - \tfrac{1}{2} \right| \geq 0.01 \right\} \leq \frac{1}{40}.$$

Thus on the average 39 out of 40 such events each with 1000 or more trials will satisfy the inequality $\left\{\left|\frac{k}{n}-\frac{1}{2}\right| \le 0.1\right\}$ or, it is quite possible that one out of 40 such events may not satisfy it. As a result if we continue the coin tossing experiment for an additional 1000 more trials, with k representing the total number of successes up to the current trial n, for $n = 1000 \to 2000,$ it is quite possible that for few such n the above inequality may be violated. This is still consistent with the weak law, but "not so often" says the strong law. According to the strong law such violations can occur only a finite number of times each with a finite probability in an infinite sequence of trials, and hence almost always the above inequality will be satisfied, i.e., the sample space of k/n coincides with that of p as $n \to \infty.$

Next we look at an experiment to confirm the strong law:

Example: $2n$ red cards and $2n$ black cards (all distinct) are shuffled together to form a single deck, and then split into half. What is the probability that each half will contain n red and n black cards?

Solution: From a deck of $4n$ cards, $2n$ cards can be chosen in $\binom{4n}{2n}$ different ways. To determine the number of favorable draws of n red and n black cards in each half, consider the unique draw consisting of $2n$ red cards and $2n$ black cards in each half. Among those $2n$ red cards, n of them can be chosen in $\binom{2n}{n}$ different ways; similarly for each such draw there are $\binom{2n}{n}$ ways of choosing n black cards. Thus the total number of favorable draws containing n red and n black cards in each half are $\binom{2n}{n}\binom{2n}{n}$ among a total of $\binom{4n}{2n}$ draws. This gives the desired probability p_n to be

$$p_n \approx \frac{\binom{2n}{n}\binom{2n}{n}}{\binom{4n}{2n}} = \frac{(2n\,!)^4}{(4n)!(n\,!)^4}.$$

For large n, using Stingling's formula we get

$$p_n \approx \frac{[\sqrt{2\pi(2n)}\,(2n)^{2n}e^{-2n}]^4}{\sqrt{2\pi(4n)}\,(4n)^{4n}e^{-4n}[\sqrt{2\pi\, n}\;n^n e^{-n}]^4} = \sqrt{\frac{2}{\pi\, n}}$$

For a full deck of 52 cards, we have $n = 13,$ which gives

$$p_n \approx 0.221$$

and for a partial deck of 20 cards (that contains 10 red and 10 black cards), we have $n = 5$ and $p_n \approx 0.3568.$

One summer afternoon, 20 cards (containing 10 red and 10 black cards) were given to a 5 year old child. The child split that partial deck into two equal halves and the outcome was declared a success if each half contained exactly 5 red and 5 black cards. With adult supervision (in terms of shuffling) the experiment was repeated 100 times that very same afternoon. The results are tabulated below in Table 13.1, and the relative frequency vs the number of trials plot in Fig 13.1 shows the convergence of k/n to p.

Table 13.1

Expt	Number of successes	Expt	Number of successes	Expt	Number of successes	Expt	Number of successes	Expt	Number of successes
1	0	21	8	41	14	61	23	81	29
2	0	22	8	42	14	62	23	82	29
3	1	23	8	43	14	63	23	83	30
4	1	24	8	44	14	64	24	84	30
5	2	25	8	45	15	65	25	85	30
6	2	26	8	46	16	66	25	86	31
7	3	27	9	47	17	67	25	87	31
8	4	28	10	48	17	68	25	88	32
9	5	29	10	49	17	69	26	89	32
10	5	30	10	50	18	70	26	90	32
11	5	31	10	51	19	71	26	91	33
12	5	32	10	52	20	72	26	92	33
13	5	33	10	53	20	73	26	93	33
14	5	34	10	54	21	74	26	94	34
15	6	35	11	55	21	75	27	95	34
16	6	36	12	56	22	76	27	96	34
17	6	37	12	57	22	77	28	97	34
18	7	38	13	58	22	78	29	98	34
19	7	39	14	59	22	79	29	99	34
20	8	40	14	60	22	80	29	100	35

The figure below shows results of an experiment of 100 trials.

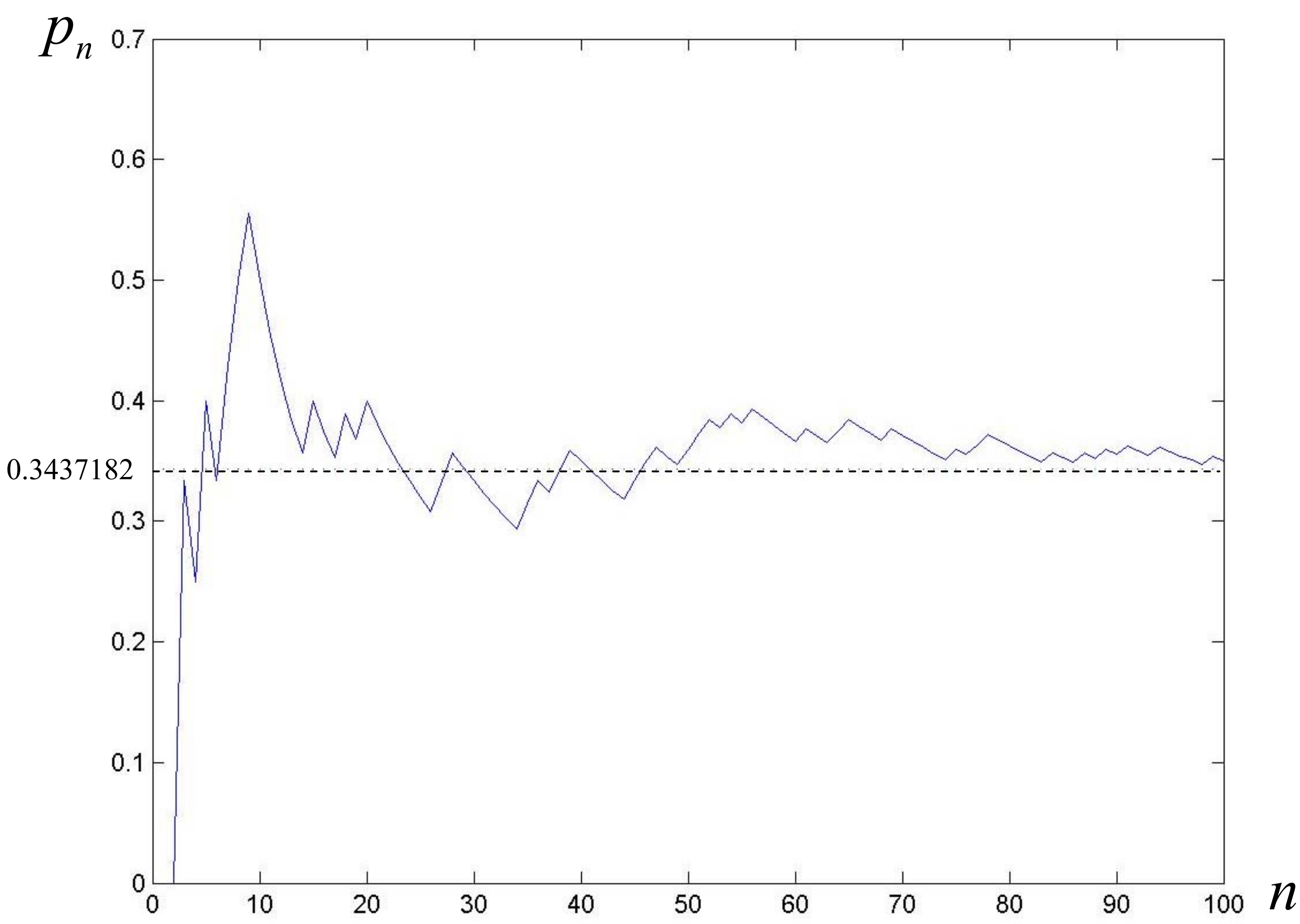


Fig 13.1

14. Stochastic Processes

Introduction

Let ξ denote the random outcome of an experiment. To every such outcome suppose a waveform $X(t,\xi)$ is assigned.

The collection of such waveforms form a stochastic process. The set of $\{\xi_k\}$ and the time index t can be continuous or discrete (countably infinite or finite) as well.

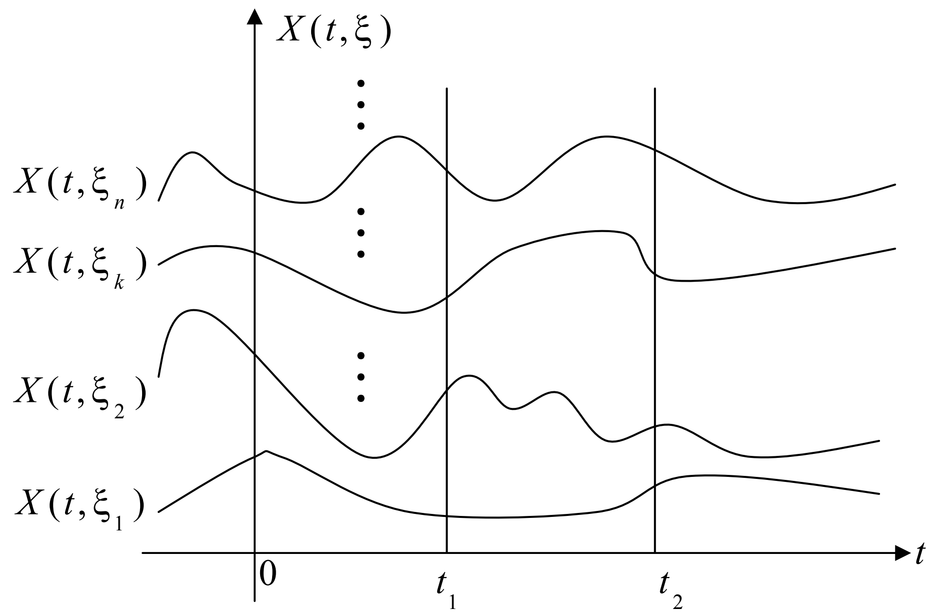


Fig. 14.1

For fixed $\xi_i \in S$ (the set of all experimental outcomes), $X(t,\xi)$ is a specific time function. For fixed t,

$$X_1 = X(t_1, \xi_i)$$

is a random variable. The ensemble of all such realizations $X(t,\xi)$ over time represents the stochastic process $X(t)$. (see Fig 14.1). For example

$$X(t) = a\cos(\omega_0 t + \varphi),$$

where φ is a uniformly distributed random variable in $(0, 2\pi)$, represents a stochastic process. Stochastic processes are everywhere: Brownian motion, stock market fluctuations, various queuing systems all represent stochastic phenomena.

If $X(t)$ is a stochastic process, then for fixed t, $X(t)$ represents a random variable. Its distribution function is given by

$$F_X(x,t) = P\{X(t) \le x\} \tag{14-1}$$

Notice that $F_X(x,t)$ depends on t, since for a different t, we obtain a different random variable. Further

$$f_X(x,t) \triangleq \frac{dF_X(x,t)}{dx} \tag{14-2}$$

represents the first-order probability density function of the process $X(t)$.

For $t = t_1$ and $t = t_2$, $X(t)$ represents two different random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$ respectively. Their joint distribution is given by

$$F_X(x_1, x_2, t_1, t_2) = P\{X(t_1) \le x_1, X(t_2) \le x_2\} \tag{14-3}$$

and

$$f_X(x_1, x_2, t_1, t_2) \triangleq \frac{\partial^2 F_X(x_1, x_2, t_1, t_2)}{\partial x_1 \ \partial x_2} \tag{14-4}$$

represents the second-order density function of the process $X(t)$. Similarly $f_X(x_1, x_2, \cdots x_n,\ t_1, t_2 \cdots, t_n)$ represents the n^{th} order density function of the process $X(t)$. Complete specification of the stochastic process $X(t)$ requires the knowledge of $f_X(x_1, x_2, \cdots x_n,\ t_1, t_2 \cdots, t_n)$ for all $t_i,\quad i = 1, 2, \cdots, n$ and for all n. (an almost impossible task in reality).

Mean of a Stochastic Process:

$$\mu(t) \stackrel{\Delta}{=} E\{X(t)\} = \int_{-\infty}^{+\infty} x f_X(x,t)dx \qquad (14\text{-}5)$$

represents the mean value of a process $X(t)$. In general, the mean of a process can depend on the time index t.

Autocorrelation function of a process $X(t)$ is defined as

$$R_{XX}(t_1,t_2) \stackrel{\Delta}{=} E\{X(t_1)X^*(t_2)\} = \int\int x_1 x_2^* f_X(x_1,x_2,t_1,t_2)dx_1 dx_2 \quad (14\text{-}6)$$

and it represents the interrelationship between the random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$ generated from the process $X(t)$.

Properties:

1. $R_{XX}(t_1,t_2) = R_{XX}^*(t_2,t_1) = [E\{X(t_2)X^*(t_1)\}]^*$ (14-7)

2. $R_{XX}(t,t) = E\{|X(t)|^2\} > 0.$ (Average instantaneous power)

3. $R_{XX}(t_1, t_2)$ represents a nonnegative definite function, i.e., for *any* set of constants $\{a_i\}_{i=1}^n$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* R_{XX}(t_i, t_j) \geq 0. \tag{14-8}$$

Eq. (14-8) follows by noticing that $E\{|Y|^2\} \geq 0$ for $Y = \sum_{i=1}^{n} a_i X(t_i)$.
The function

$$C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X^*(t_2) \tag{14-9}$$

represents the **autocovariance** function of the process $X(t)$.

Example 14.1

Let

$$z = \int_{-T}^{T} X(t)dt.$$

Then

$$E[|z|^2] = \int_{-T}^{T}\int_{-T}^{T} E\{X(t_1)X^*(t_2)\}dt_1 dt_2$$

$$= \int_{-T}^{T}\int_{-T}^{T} R_{XX}(t_1, t_2)dt_1 dt_2 \tag{14-10}$$

Example 14.2

$$X(t) = a\cos(\omega_0 t + \varphi), \qquad \varphi \sim U(0, 2\pi). \tag{14-11}$$

This gives

$$\begin{aligned}\mu_X(t) &= E\{X(t)\} = aE\{\cos(\omega_0 t + \varphi)\} \\ &= a\cos\omega_0 t\, E\{\cos\varphi\} - a\sin\omega_0 t\, E\{\sin\varphi\} = 0,\end{aligned} \tag{14-12}$$

since $E\{\cos\varphi\} = \frac{1}{2\pi}\int_0^{2\pi} \cos\varphi\, d\varphi = 0 = E\{\sin\varphi\}.$

Similarly

$$\begin{aligned}R_{XX}(t_1, t_2) &= a^2 E\{\cos(\omega_0 t_1 + \varphi)\cos(\omega_0 t_2 + \varphi)\} \\ &= \frac{a^2}{2} E\{\cos\omega_0(t_1 - t_2) + \cos(\omega_0(t_1 + t_2) + 2\varphi)\} \\ &= \frac{a^2}{2}\cos\omega_0(t_1 - t_2).\end{aligned} \tag{14-13}$$

Stationary Stochastic Processes

Stationary processes exhibit statistical properties that are invariant to shift in the time index. Thus, for example, second-order stationarity implies that the statistical properties of the pairs $\{X(t_1), X(t_2)\}$ and $\{X(t_1+c), X(t_2+c)\}$ are the same for *any* c. Similarly first-order stationarity implies that the statistical properties of $X(t_i)$ and $X(t_i+c)$ are the same for any c.

In strict terms, the statistical properties are governed by the joint probability density function. Hence a process is nth-order **Strict-Sense Stationary (S.S.S)** if

$$f_X(x_1, x_2, \cdots x_n, \; t_1, t_2 \cdots, t_n) \equiv f_X(x_1, x_2, \cdots x_n, \; t_1 + c, t_2 + c \cdots, t_n + c) \tag{14-14}$$

for *any* c, where the left side represents the joint density function of the random variables $X_1 = X(t_1), \; X_2 = X(t_2), \; \cdots, \; X_n = X(t_n)$ and the right side corresponds to the joint density function of the random variables $X_1' = X(t_1 + c), \; X_2' = X(t_2 + c), \; \cdots, \; X_n' = X(t_n + c).$ A process $X(t)$ is said to be **strict-sense stationary** if (14-14) is true for all $t_i, \;\; i = 1, 2, \cdots, n, \;\; n = 1, 2, \cdots$ and *any* c.

For a **first-order strict sense stationary process**, from (14-14) we have

$$f_X(x,t) \equiv f_X(x,t+c) \tag{14-15}$$

for any c. In particular $c = -t$ gives

$$f_X(x,t) = f_X(x) \tag{14-16}$$

i.e., the first-order density of $X(t)$ is independent of t. In that case

$$E[X(t)] = \int_{-\infty}^{+\infty} x f(x)dx = \mu, \ \ a\ constant. \tag{14-17}$$

Similarly, for a **second-order strict-sense stationary process** we have from (14-14)

$$f_X(x_1, x_2,\ t_1, t_2) \equiv f_X(x_1, x_2,\ t_1+c, t_2+c)$$

for any c. For $c = -t_2$ we get

$$f_X(x_1, x_2,\ t_1, t_2) \equiv f_X(x_1, x_2,\ t_1 - t_2) \tag{14-18}$$

i.e., the second order density function of a strict sense stationary process depends only on the difference of the time indices $t_1 - t_2 = \tau$. In that case the autocorrelation function is given by

$$\begin{aligned} R_{XX}(t_1, t_2) &\stackrel{\Delta}{=} E\{X(t_1)X^*(t_2)\} \\ &= \int\int x_1 x_2^* f_X(x_1, x_2, \tau = t_1 - t_2) dx_1 dx_2 \\ &= R_{XX}(t_1 - t_2) \stackrel{\Delta}{=} R_{XX}(\tau) = R_{XX}^*(-\tau), \end{aligned} \qquad (14\text{-}19)$$

i.e., the autocorrelation function of a second order strict-sense stationary process depends only on the difference of the time indices $\tau = t_1 - t_2$.

Notice that (14-17) and (14-19) are consequences of the stochastic process being first and second-order strict sense stationary.

On the other hand, the basic conditions for the first and second order stationarity – Eqs. (14-16) and (14-18) – are usually difficult to verify. In that case, we often resort to a looser definition of stationarity, known as **Wide-Sense Stationarity (W.S.S),** by making use of

(14-17) and (14-19) as the necessary conditions. Thus, a process $X(t)$ is said to be **Wide-Sense Stationary** if

(i) $E\{X(t)\} = \mu$ (14-20)

and

(ii) $E\{X(t_1)X^*(t_2)\} = R_{XX}(t_1 - t_2),$ (14-21)

i.e., for wide-sense stationary processes, the mean is a constant and the autocorrelation function depends only on the difference between the time indices. Notice that (14-20)-(14-21) does not say anything about the nature of the probability density functions, and instead deal with the average behavior of the process. Since (14-20)-(14-21) follow from (14-16) and (14-18), strict-sense stationarity always implies wide-sense stationarity. However, the converse is *not true* in general, the only exception being the Gaussian process.

This follows, since if $X(t)$ is a Gaussian process, then by definition $X_1 = X(t_1), X_2 = X(t_2), \cdots, X_n = X(t_n)$ are jointly Gaussian random variables for any $t_1, t_2 \cdots, t_n$ whose joint characteristic function is given by

$$\phi_{\underline{X}}(\omega_1,\omega_2,\cdots,\omega_n) = e^{j\sum_{k=1}^{n}\mu(t_k)\omega_k - \sum_{l,k}^{n}\sum C_{XX}(t_i,t_k)\omega_i\omega_k/2} \tag{14-22}$$

where $C_{XX}(t_i,t_k)$ is as defined on (14-9). If $X(t)$ is wide-sense stationary, then using (14-20)-(14-21) in (14-22) we get

$$\phi_{\underline{X}}(\omega_1,\omega_2,\cdots,\omega_n) = e^{j\sum_{k=1}^{n}\mu\omega_k - \frac{1}{2}\sum_{1=1}^{n}\sum_{k=1}^{n}C_{XX}(t_i-t_k)\omega_i\omega_k} \tag{14-23}$$

and hence if the set of time indices are shifted by a constant c to generate a new set of jointly Gaussian random variables $X_1' = X(t_1+c)$, $X_2' = X(t_2+c),\cdots,X_n' = X(t_n+c)$ then their joint characteristic function is identical to (14-23). Thus the set of random variables $\{X_i\}_{i=1}^n$ and $\{X_i'\}_{i=1}^n$ have the same joint probability distribution for all n and all c, establishing the strict sense stationarity of Gaussian processes from its wide-sense stationarity.

To summarize if $X(t)$ is a Gaussian process, then

wide-sense stationarity (w.s.s) $\Rightarrow$ strict-sense stationarity (s.s.s).

Notice that since the joint p.d.f of Gaussian random variables depends only on their second order statistics, which is also the basis for wide sense stationarity, we obtain strict sense stationarity as well. From (14-12)-(14-13), (refer to Example 14.2), the process $X(t) = a\cos(\omega_0 t + \varphi)$, in (14-11) is wide-sense stationary, but not strict-sense stationary.

Similarly if $X(t)$ is a zero mean wide sense stationary process in Example 14.1, then σ_z^2 in (14-10) reduces to

$$\sigma_z^2 = E\{|z|^2\} = \int_{-T}^{T}\int_{-T}^{T} R_{XX}(t_1 - t_2)dt_1 dt_2.$$

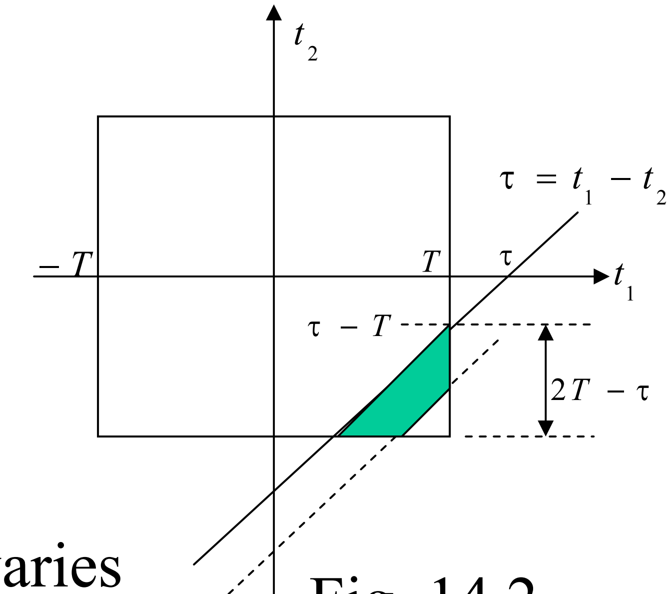


Fig. 14.2

As t_1, t_2 varies from $-T$ to $+T$, $\tau = t_1 - t_2$ varies from $-2T$ to $+2T$. Moreover $R_{XX}(\tau)$ is a constant over the shaded region in Fig 14.2, whose area is given by $(\tau > 0)$

$$\frac{1}{2}(2T-\tau)^2 - \frac{1}{2}(2T-\tau-d\tau)^2 = (2T-\tau)d\tau$$

and hence the above integral reduces to

$$\sigma_z^2 = \int_{-2t}^{2T} R_{XX}(\tau)(2T - |\tau|)d\tau = \tfrac{1}{2T}\int_{-2t}^{2T} R_{XX}(\tau)(1 - \tfrac{|\tau|}{2T})d\tau. \tag{14-24}$$

Systems with Stochastic Inputs

A deterministic system[1] transforms each input waveform $X(t,\xi_i)$ into an output waveform $Y(t,\xi_i) = T[X(t,\xi_i)]$ by operating only on the time variable t. Thus a set of realizations at the input corresponding to a process $X(t)$ generates a new set of realizations $\{Y(t,\xi)\}$ at the output associated with a new process $Y(t)$.

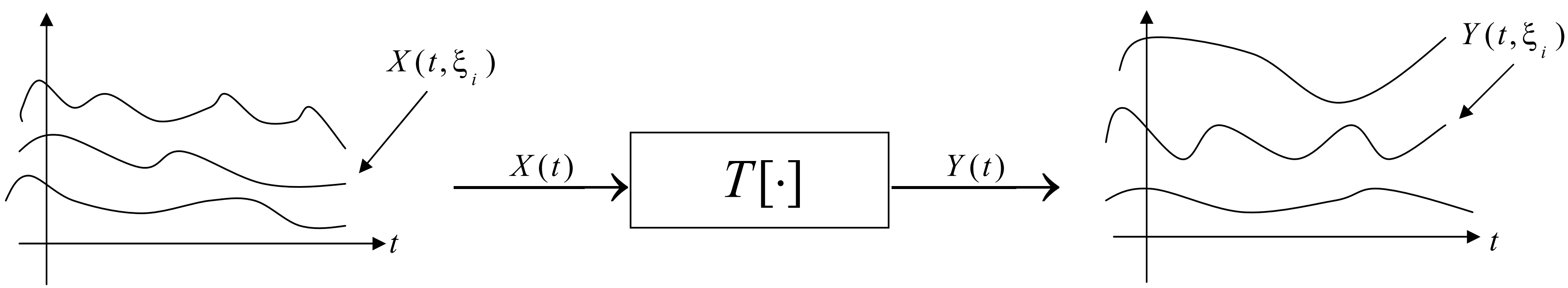


Fig. 14.3

Our goal is to study the output process statistics in terms of the input process statistics and the system function.

---

[1]A stochastic system on the other hand operates on both the variables t and ξ.

Deterministic Systems

Memoryless Systems

$$Y(t) = g[X(t)]$$

Systems with Memory

Time-varying systems

Time-Invariant systems

Linear systems

$$Y(t) = L[X(t)]$$

Fig. 14.3

Linear-Time Invariant (LTI) systems

$$X(t) \longrightarrow \boxed{h(t)} \longrightarrow Y(t) = \int_{-\infty}^{+\infty} h(t-\tau)X(\tau)\,d\tau$$

$$= \int_{-\infty}^{+\infty} h(\tau)X(t-\tau)\,d\tau.$$

LTI system

Memoryless Systems:

The output $Y(t)$ in this case depends only on the present value of the input $X(t)$. i.e.,

$$Y(t) = g\{X(t)\} \qquad (14\text{-}25)$$

Input		System		Output
Strict-sense stationary input	$\rightarrow$	Memoryless system	$\rightarrow$	Strict-sense stationary output.

(see (9-76), Text for a proof.)

Input		System		Output
Wide-sense stationary input	$\rightarrow$	Memoryless system	$\rightarrow$	Need *not* be stationary in any sense.
$X(t)$ stationary Gaussian with $R_{XX}(\tau)$	$\rightarrow$	Memoryless system	$\rightarrow$	$Y(t)$ stationary,but *not* Gaussian with $R_{XY}(\tau) = \eta R_{XX}(\tau)$. (see (14-26)).

Fig. 14.4

Theorem: If $X(t)$ is a zero mean stationary Gaussian process, and $Y(t) = g[X(t)]$, where $g(\cdot)$ represents a nonlinear memoryless device, then

$$R_{XY}(\tau) = \eta R_{XX}(\tau), \qquad \eta = E\{g'(X)\}. \tag{14-26}$$

Proof:

$$R_{XY}(\tau) = E\{X(t)Y(t-\tau)\} = E[X(t)g\{X(t-\tau)\}]$$
$$= \int\int x_1 g(x_2) f_{X_1X_2}(x_1,x_2)dx_1 dx_2 \tag{14-27}$$

where $X_1 = X(t), \quad X_2 = X(t-\tau)$ are jointly Gaussian random variables, and hence

$$f_{X_1X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{|A|}} e^{-\underline{x}^* A^{-1} \underline{x}/2}$$

$$\underline{X} = (X_1, X_2)^T, \qquad \underline{x} = (x_1, x_2)^T$$

$$A = E\{\underline{X}\,\underline{X}^*\} = \begin{pmatrix} R_{XX}(0) & R_{XX}(\tau) \\ R_{XX}(\tau) & R_{XX}(0) \end{pmatrix} \triangleq LL^*$$

where L is an upper triangular factor matrix with positive diagonal entries. i.e.,

$$L = \begin{pmatrix} l_{11} & l_{12} \\ 0 & l_{22} \end{pmatrix}.$$

Consider the transformation

$$\underline{Z} \triangleq L^{-1}\underline{X} = (Z_1,\, Z_2)^T, \qquad \underline{z} \triangleq L^{-1}\underline{x} = (z_1,\, z_2)^T$$

so that

$$E\{\underline{Z}\,\underline{Z}^*\} = L^{-1}E\{\underline{X}\,X^*\}L^{*^{-1}} = L^{-1}AL^{*^{-1}} = I$$

and hence Z_1, Z_2 are zero mean independent Gaussian random variables. Also

$$\underline{x} = L\underline{z} \;\Rightarrow\; x_1 = l_{11}z_1 + l_{12}z_2, \quad x_2 = l_{22}z_2$$

and hence

$$\underline{x}^* A^{-1}\underline{x} = \underline{z}^* L^* A^{-1} L\underline{z} = \underline{z}^*\,\underline{z} = z_1^2 + z_2^2.$$

The Jacobaian of the transformation is given by

$$|J| = |L^{-1}| = |A|^{-1/2}.$$

Hence substituting these into (14-27), we obtain

$$R_{XY}(\tau) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(l_{11}z_1 + l_{12}z_2)g(l_{22}z_2)\tfrac{1}{|J|}\cdot\frac{1}{2\pi|A|^{1/2}}e^{-z_1^2/2}e^{-z_2^2/2}$$

$$= l_{11}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} z_1 g(l_{22}z_2) f_{z_1}(z_1) f_{z_2}(z_2) dz_1 dz_2$$

$$+ l_{12}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} z_2 g(l_{22}z_2) f_{z_1}(z_1) f_{z_2}(z_2) dz_1 dz_2$$

$$= l_{11}\overbrace{\int_{-\infty}^{+\infty} z_1 f_{z_1}(z_1) dz_1}^{0}\int_{-\infty}^{+\infty} g(l_{22}z_2) f_{z_2}(z_2) dz_2$$

$$+ l_{12}\int_{-\infty}^{+\infty} z_2 g(l_{22}z_2)\underbrace{f_{z_2}(z_2)}_{\frac{1}{\sqrt{2\pi}}e^{-z_2^2/2}} dz_2$$

$$= \frac{l_{12}}{l_{22}^2}\int_{-\infty}^{+\infty} u g(u)\frac{1}{\sqrt{2\pi}}e^{-u^2/2l_{22}^2} du,$$

where $u = l_{22}z_2$. This gives

$$R_{XY}(\tau) = l_{12}l_{22}\int_{-\infty}^{+\infty} g(u) \underbrace{\frac{u}{l_{22}^2} \overbrace{\frac{1}{\sqrt{2\pi\ l_{22}^2}} e^{-u^2/2l_{22}^2}}^{f_u(u)}}_{-\frac{df_u(u)}{du} = -f_u'(u)} du$$

$$= -R_{XX}(\tau)\int_{-\infty}^{+\infty} g(u) f_u'(u) du,$$

since $A = LL^*$ gives $l_{12}l_{22} = R_{XX}(\tau)$. Hence

$$R_{XY}(\tau) = R_{XX}(\tau)\{-\underbrace{g(u) f_u(u) \big|_{-\infty}^{+\infty}}_{0} + \int_{-\infty}^{+\infty} g'(u) f_u(u) du\}$$

$$= R_{XX}(\tau) E\{g'(X)\} = \eta R_{XX}(\tau),$$

the desired result, where $\eta = E[g'(X)]$. Thus if the input to a memoryless device is stationary Gaussian, the cross correlation function between the input and the output is proportional to the input autocorrelation function.

Linear Systems: $L[\cdot]$ represents a linear system if

$$L\{a_1 X(t_1) + a_2 X(t_2)\} = a_1 L\{X(t_1)\} + a_2 L\{X(t_2)\}. \tag{14-28}$$

Let

$$Y(t) = L\{X(t)\} \tag{14-29}$$

represent the output of a linear system.

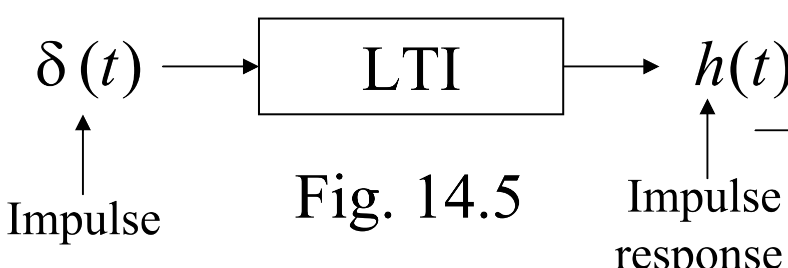
Time-Invariant System: $L[\cdot]$ represents a time-invariant system if

$$Y(t) = L\{X(t)\} \Rightarrow L\{X(t-t_0)\} = Y(t-t_0) \tag{14-30}$$

i.e., shift in the input results in the same shift in the output also.

If $L[\cdot]$ satisfies both (14-28) and (14-30), then it corresponds to a linear time-invariant (LTI) system.

LTI systems can be uniquely represented in terms of their output to a delta function



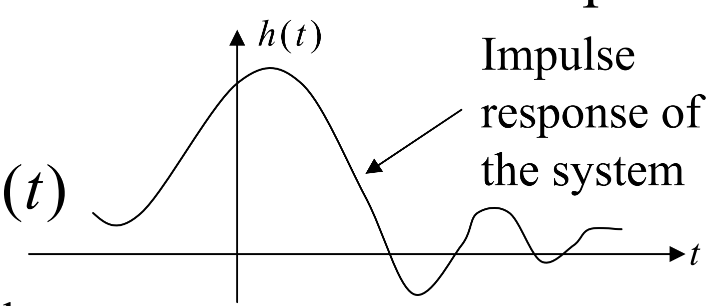


Fig. 14.5

then

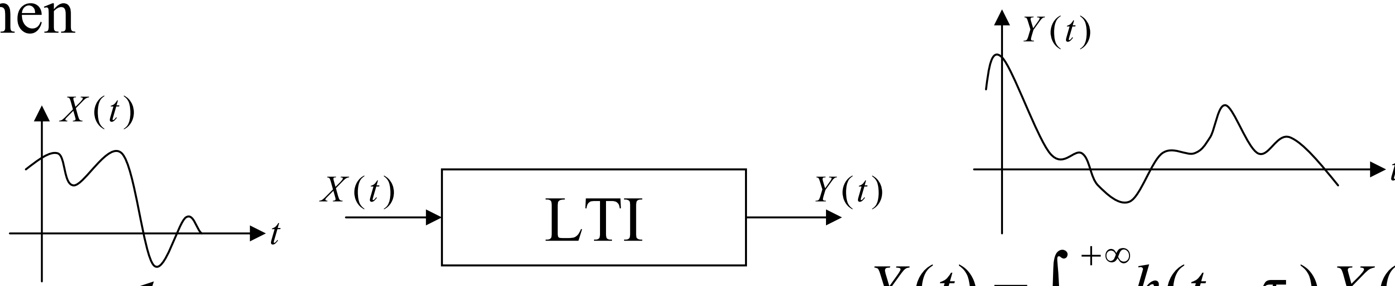


Fig. 14.6

$$Y(t) = \int_{-\infty}^{+\infty} h(t-\tau)X(\tau)d\tau$$

$$= \int_{-\infty}^{+\infty} h(\tau)X(t-\tau)d\tau \quad (14\text{-}31)$$

Eq. (14-31) follows by expressing $X(t)$ as

$$X(t) = \int_{-\infty}^{+\infty} X(\tau)\delta(t-\tau)d\tau \qquad (14\text{-}32)$$

and applying (14-28) and (14-30) to $Y(t) = L\{X(t)\}$. Thus

$$Y(t) = L\{X(t)\} = L\{\int_{-\infty}^{+\infty} X(\tau)\delta(t-\tau)d\tau\}$$

$$= \int_{-\infty}^{+\infty} L\{X(\tau)\delta(t-\tau)d\tau\}$$

By Linearity

$$= \int_{-\infty}^{+\infty} X(\tau)L\{\delta(t-\tau)\}d\tau$$

By Time-invariance

$$= \int_{-\infty}^{+\infty} X(\tau)h(t-\tau)d\tau = \int_{-\infty}^{+\infty} h(\tau)X(t-\tau)d\tau. \quad (14\text{-}33)$$

Output Statistics: Using (14-33), the mean of the output process is given by

$$\mu_Y(t) = E\{Y(t)\} = \int_{-\infty}^{+\infty} E\{X(\tau)h(t-\tau)d\tau\}$$
$$= \int_{-\infty}^{+\infty} \mu_X(\tau)h(t-\tau)d\tau = \mu_X(t) * h(t). \qquad (14\text{-}34)$$

Similarly the cross-correlation function between the input and output processes is given by

$$R_{XY}(t_1,t_2) = E\{X(t_1)Y^*(t_2)\}$$
$$= E\{X(t_1)\int_{-\infty}^{+\infty} X^*(t_2-\alpha)h^*(\alpha)d\alpha\}$$
$$= \int_{-\infty}^{+\infty} E\{X(t_1)X^*(t_2-\alpha)\}h^*(\alpha)d\alpha$$
$$= \int_{-\infty}^{+\infty} R_{XX}(t_1,t_2-\alpha)h^*(\alpha)d\alpha$$
$$= R_{XX}(t_1,t_2) * h^*(t_2). \qquad (14\text{-}35)$$

Finally the output autocorrelation function is given by

$$
\begin{aligned}
R_{YY}(t_1,t_2) &= E\{Y(t_1)Y^*(t_2)\} \\
&= E\{\int_{-\infty}^{+\infty} X(t_1-\beta)h(\beta)d\beta \; Y^*(t_2)\} \\
&= \int_{-\infty}^{+\infty} E\{X(t_1-\beta)Y^*(t_2)\}h(\beta)d\beta \\
&= \int_{-\infty}^{+\infty} R_{XY}(t_1-\beta,t_2)h(\beta)d\beta \\
&= R_{XY}(t_1,t_2) * h(t_1),
\end{aligned}
\tag{14-36}
$$

or

$$
R_{YY}(t_1,t_2) = R_{XX}(t_1,t_2) * h^*(t_2) * h(t_1). \tag{14-37}
$$

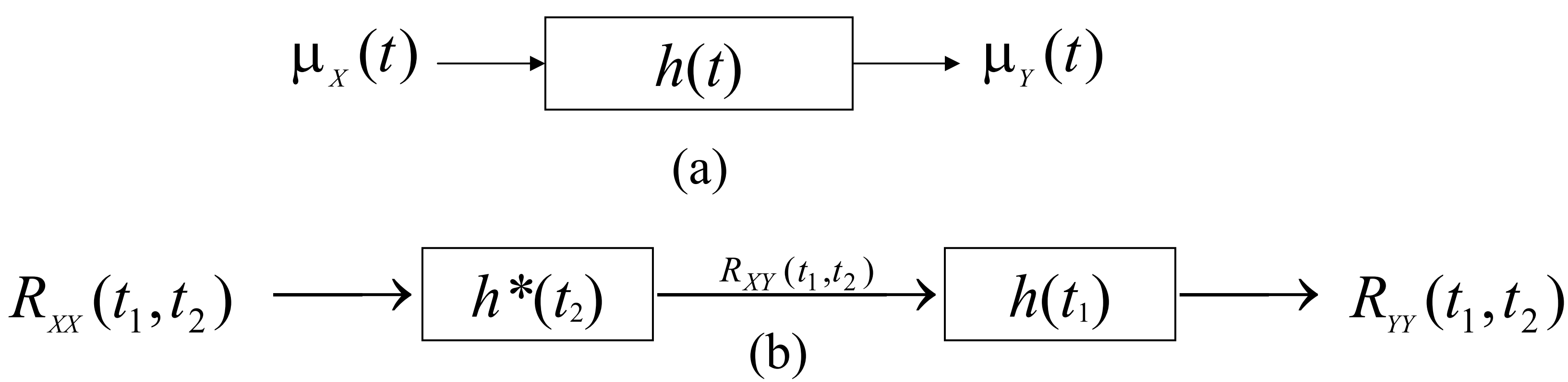


Fig. 14.7

In particular if $X(t)$ is wide-sense stationary, then we have $\mu_X(t) = \mu_X$ so that from (14-34)

$$\mu_Y(t) = \mu_X \int_{-\infty}^{+\infty} h(\tau)\,d\tau = \mu_X c, \quad a\ constant. \tag{14-38}$$

Also $R_{XX}(t_1, t_2) = R_{XX}(t_1 - t_2)$ so that (14-35) reduces to

$$\begin{aligned} R_{XY}(t_1, t_2) &= \int_{-\infty}^{+\infty} R_{XX}(t_1 - t_2 + \alpha) h^*(\alpha)\,d\alpha \\ &= R_{XX}(\tau) * h^*(-\tau) \stackrel{\Delta}{=} R_{XY}(\tau), \quad \tau = t_1 - t_2. \end{aligned} \tag{14-39}$$

Thus $X(t)$ and $Y(t)$ are jointly w.s.s. Further, from (14-36), the output autocorrelation simplifies to

$$\begin{aligned} R_{YY}(t_1, t_2) &= \int_{-\infty}^{+\infty} R_{XY}(t_1 - \beta - t_2) h(\beta)\,d\beta, \quad \tau = t_1 - t_2 \\ &= R_{XY}(\tau) * h(\tau) = R_{YY}(\tau). \end{aligned} \tag{14-40}$$

From (14-37), we obtain

$$R_{YY}(\tau) = R_{XX}(\tau) * h^*(-\tau) * h(\tau). \tag{14-41}$$

From (14-38)-(14-40), the output process is also wide-sense stationary. This gives rise to the following representation

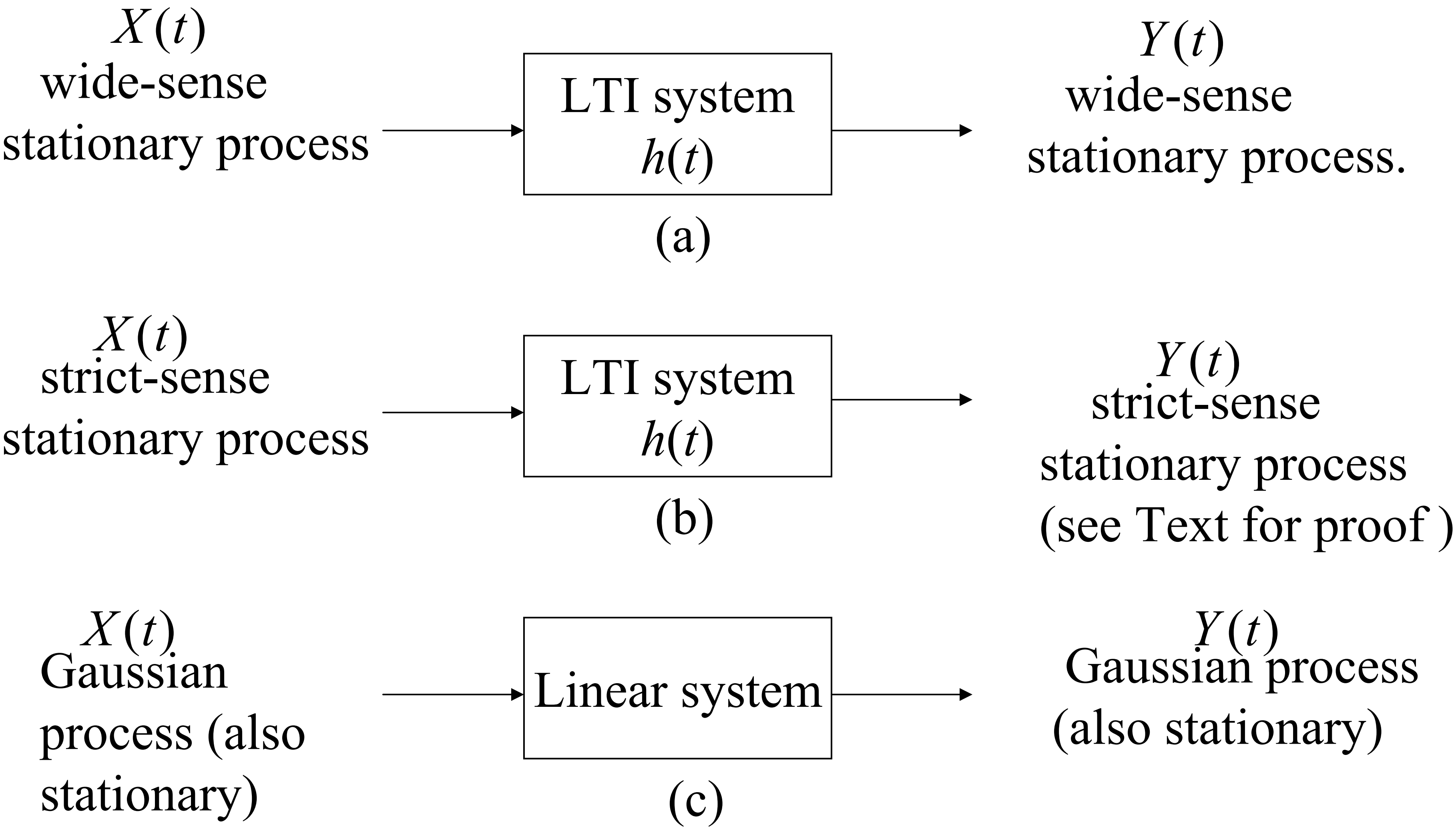


Fig. 14.8

White Noise Process:

$W(t)$ is said to be a white noise process if

$$R_{WW}(t_1, t_2) = q(t_1)\delta(t_1 - t_2), \tag{14-42}$$

i.e., $E[W(t_1)\, W^*(t_2)] = 0$ unless $t_1 = t_2$.
$W(t)$ is said to be wide-sense stationary (w.s.s) white noise if $E[W(t)]$ = constant, and

$$R_{WW}(t_1, t_2) = q\delta(t_1 - t_2) = q\delta(\tau). \tag{14-43}$$

If $W(t)$ is also a Gaussian process (white Gaussian process), then all of its samples are independent random variables (why?).

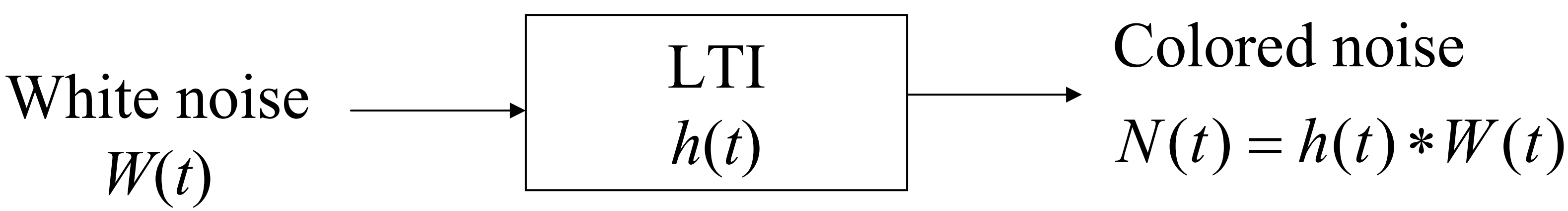


Fig. 14.9

For w.s.s. white noise input $W(t)$, we have

$$E[N(t)] = \mu_W \int_{-\infty}^{+\infty} h(\tau)\,d\tau, \quad a\ constant \tag{14-44}$$

and

$$\begin{aligned} R_{nn}(\tau) &= q\delta(\tau) * h^*(-\tau) * h(\tau) \\ &= qh^*(-\tau) * h(\tau) = q\rho(\tau) \end{aligned} \tag{14-45}$$

where

$$\rho(\tau) = h(\tau) * h^*(-\tau) = \int_{-\infty}^{+\infty} h(\alpha)h^*(\alpha+\tau)\,d\alpha. \tag{14-46}$$

Thus the output of a white noise process through an LTI system represents a (colored) noise process.

Note: White noise need not be Gaussian.

"White" and "Gaussian" are two different concepts!

Upcrossings and Downcrossings of a stationary Gaussian process:

Consider a zero mean stationary Gaussian process $X(t)$ with autocorrelation function $R_{XX}(\tau)$. An upcrossing over the mean value occurs whenever the realization $X(t)$ passes through zero with positive slope. Let $\rho\Delta t$ represent the probability of such an upcrossing in the interval $(t, t+\Delta t)$. We wish to determine ρ.

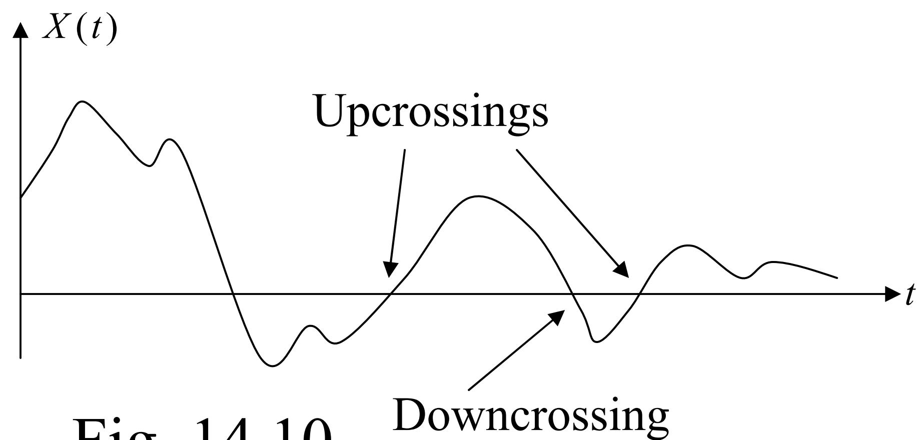


Fig. 14.10

Since $X(t)$ is a stationary Gaussian process, its derivative process $X'(t)$ is also zero mean stationary Gaussian with autocorrelation function $R_{X'X'}(\tau) = -R''_{XX}(\tau)$ (see (9-101)-(9-106), Text). Further $X(t)$ and $X'(t)$ are jointly Gaussian stationary processes, and since (see (9-106), Text)

$$R_{XX'}(\tau) = -\frac{dR_{XX}(\tau)}{d\tau},$$

we have

$$R_{XX'}(-\tau) = -\frac{dR_{XX}(-\tau)}{d(-\tau)} = \frac{dR_{XX}(\tau)}{d\tau} = -R_{XX'}(\tau) \qquad (14\text{-}47)$$

which for $\tau = 0$ gives

$$R_{XX'}(0) = 0 \quad \Rightarrow \quad E[X(t)X'(t)] = 0 \qquad (14\text{-}48)$$

i.e., the jointly Gaussian zero mean random variables

$$X_1 = X(t) \quad \text{and} \quad X_2 = X'(t) \qquad (14\text{-}49)$$

are uncorrelated and hence *independent* with variances

$$\sigma_1^2 = R_{XX}(0) \quad \text{and} \quad \sigma_2^2 = R_{X'X'}(0) = -R''_{XX}(0) > 0 \qquad (14\text{-}50)$$

respectively. Thus

$$f_{X_1X_2}(x_1, x_2) = f_X(x_1) f_X(x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\left(\frac{x_1^2}{2\sigma_1^2} + \frac{x_1^2}{2\sigma_2^2}\right)}. \qquad (14\text{-}51)$$

To determine ρ, the probability of upcrossing rate, we argue as follows: In an interval $(t, t+\Delta t)$, the realization moves from $X(t) = X_1$ to $X(t+\Delta t) = X(t) + X'(t)\Delta t = X_1 + X_2 \Delta t$, and hence the realization intersects with the zero level somewhere in that interval if

$$X_1 < 0, \;\; X_2 > 0, \quad \text{and} \quad X(t+\Delta t) = X_1 + X_2\Delta t > 0 \qquad (14\text{-}52)$$

i.e., $X_1 > -X_2 \Delta t.$
Hence the probability of upcrossing in $(t, t+\Delta t)$ is given by

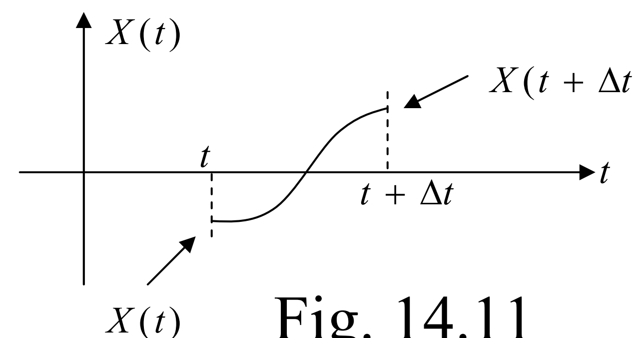


Fig. 14.11

$$\rho\Delta t = \int_{x_2=0}^{\infty}\int_{x_1=-x_2\Delta t}^{0} f_{X_1X_2}(x_1, x_2)\,dx_1 dx_2$$

$$= \int_0^{\infty} f_{X_2}(x_2)\,dx_2 \int_{-x_2\Delta t}^{\infty} f_{X_1}(x_1)\,dx_1. \qquad (14\text{-}53)$$

Differentiating both sides of (14-53) with respect to Δt, we get

$$\rho = \int_0^{\infty} f_{X_2}(x_2) x_2 f_{X_1}(-x_2\Delta t)\,dx_2 \qquad (14\text{-}54)$$

and letting $\Delta t \to 0$, Eq. (14-54) reduce to

$$\rho = \int_0^{\infty} x_2 f_X(x_2) f_X(0) dx_2 = \frac{1}{\sqrt{2\pi R_{XX}(0)}} \int_0^{\infty} x_2 f_X(x_2) dx_2$$

$$= \frac{1}{\sqrt{2\pi R_{XX}(0)}} \frac{1}{2} (\sigma_2 \sqrt{2/\pi}) = \frac{1}{2\pi} \sqrt{\frac{-R''_{XX}(0)}{R_{XX}(0)}} \tag{14-55}$$

[where we have made use of (5-78), Text]. There is an equal probability for downcrossings, and hence the total probability for crossing the zero line in an interval $(t, t+\Delta t)$ equals $\rho_0 \Delta t,$ where

$$\rho_0 = \frac{1}{\pi} \sqrt{-R''_{XX}(0)/R_{XX}(0)} \; > \; 0. \tag{14-56}$$

It follows that in a long interval T, there will be approximately $\rho_0 T$ crossings of the mean value. If $-\ R''_{XX}(0)$ is large, then the autocorrelation function $R_{XX}(\tau)$ decays more rapidly as τ moves away from zero, implying a large random variation around the origin (mean value) for $X(t)$, and the likelihood of zero crossings should increase with increase in $-\ R''_{XX}(0),$ agreeing with (14-56).

Discrete Time Stochastic Processes:

A discrete time stochastic process $X_n = X(nT)$ is a sequence of random variables. The mean, autocorrelation and auto-covariance functions of a discrete-time process are gives by

$$\mu_n = E\{X(nT)\} \tag{14-57}$$

$$R(n_1, n_2) = E\{X(n_1 T) X^*(n_2 T)\} \tag{14-58}$$

and

$$C(n_1, n_2) = R(n_1, n_2) - \mu_{n_1} \mu_{n_2}^* \tag{14-59}$$

respectively. As before strict sense stationarity and wide-sense stationarity definitions apply here also.
For example, $X(nT)$ is wide sense stationary if

$$E\{X(nT)\} = \mu, \quad a\ constant \tag{14-60}$$

and

$$E[X\{(k+n)T\} X^*\{(k)T\}] = R(n) = r_n \triangleq r_{-n}^* \tag{14-61}$$

i.e., $R(n_1, n_2) = R(n_1 - n_2) = R^*(n_2 - n_1)$. The positive-definite property of the autocorrelation sequence in (14-8) can be expressed in terms of certain Hermitian-Toeplitz matrices as follows:

Theorem: A sequence $\{r_n\}_{-\infty}^{+\infty}$ forms an autocorrelation sequence of a wide sense stationary stochastic process if and only if every Hermitian-Toeplitz matrix T_n given by

$$T_n = \begin{pmatrix} r_0 & r_1 & r_2 & \cdots & r_n \\ r_1^* & r_0 & r_1 & \cdots & r_{n-1} \\ & & \vdots & & \\ r_n^* & r_{n-1}^* & \cdots & r_1^* & r_0 \end{pmatrix} = T_n^* \tag{14-62}$$

is non-negative (positive) definite for $n = 0,\ 1,\ 2, \cdots, \infty.$

Proof: Let $\underline{a} = [a_0, a_1, \cdots, a_n]^T$ represent an arbitrary constant vector. Then from (14-62),

$$\underline{a}^* T_n \underline{a} = \sum_{i=0}^{n} \sum_{k=0}^{n} a_i a_k^* r_{k-i} \tag{14-63}$$

since the Toeplitz character gives $(T_n)_{i,k} = r_{k-i}$. Using (14-61), Eq. (14-63) reduces to

$$\underline{a}^* T_n \underline{a} = \sum_{i=0}^{n} \sum_{k=0}^{n} a_i a_k^* E\{X(kT)X^*(iT)\} = E\left\{ \left| \sum_{k=0}^{n} a_k^* X(kT) \right|^2 \right\} \geq 0. \quad (14\text{-}64)$$

From (14-64), if $X(nT)$ is a wide sense stationary stochastic process then T_n is a non-negative definite matrix for every $n = 0, 1, 2, \cdots, \infty.$ Similarly the converse also follows from (14-64). (see section 9.4, Text)

If $X(nT)$ represents a wide-sense stationary input to a discrete-time system $\{h(nT)\}$, and $Y(nT)$ the system output, then as before the cross correlation function satisfies

$$R_{XY}(n) = R_{XX}(n) * h^*(-n) \tag{14-65}$$

and the output autocorrelation function is given by

$$R_{YY}(n) = R_{XY}(n) * h(n) \tag{14-66}$$

or

$$R_{YY}(n) = R_{XX}(n) * h^*(-n) * h(n). \tag{14-67}$$

Thus wide-sense stationarity from input to output is preserved for discrete-time systems also.

Auto Regressive Moving Average (ARMA) Processes

Consider an input – output representation

$$X(n) = -\sum_{k=1}^{p} a_k X(n-k) + \sum_{k=0}^{q} b_k W(n-k), \tag{14-68}$$

where $X(n)$ may be considered as the output of a system $\{h(n)\}$ driven by the input $W(n)$.

$$W(n) \longrightarrow \boxed{h(n)} \longrightarrow X(n)$$

Fig.14.12

Z – transform of (14-68) gives

$$X(z)\sum_{k=0}^{p} a_k z^{-k} = W(z)\sum_{k=0}^{q} b_k z^{-k}, \quad a_0 \equiv 1 \tag{14-69}$$

or

$$H(z) = \sum_{k=0}^{\infty} h(k) z^{-k} = \frac{X(z)}{W(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_q z^{-q}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p}} \triangleq \frac{B(z)}{A(z)} \tag{14-70}$$

represents the transfer function of the associated system response $\{h(n)\}$ in Fig 14.12 so that

$$X(n)=\sum_{k=0}^{\infty} h(n-k)W(k). \qquad (14\text{-}71)$$

Notice that the transfer function $H(z)$ in (14-70) is rational with p poles and q zeros that determine the model order of the underlying system. From (14-68), the output undergoes regression over p of its previous values and at the same time a moving average based on $W(n), W(n-1), \cdots, W(n-q)$ of the input over $(q + 1)$ values is added to it, thus generating an **Auto Regressive Moving Average** (ARMA (p, q)) process $X(n)$. Generally the input $\{W(n)\}$ represents a sequence of uncorrelated random variables of zero mean and constant variance σ_W^2 so that

$$R_{WW}(n)=\sigma_W^2\delta(n). \qquad (14\text{-}72)$$

If in addition, $\{W(n)\}$ is normally distributed then the output $\{X(n)\}$ also represents a strict-sense stationary normal process.

If $q = 0$, then (14-68) represents an AR(p) process (all-pole process), and if $p = 0$, then (14-68) represents an MA(q)

process (all-zero process). Next, we shall discuss AR(1) and AR(2) processes through explicit calculations.

AR(1) process: An AR(1) process has the form (see (14-68))

$$X(n) = aX(n-1) + W(n) \tag{14-73}$$

and from (14-70) the corresponding system transfer

$$H(z) = \frac{1}{1 - az^{-1}} = \sum_{n=0}^{\infty} a^n z^{-n} \tag{14-74}$$

provided $|a| < 1$. Thus

$$h(n) = a^n, \qquad |a| < 1 \tag{14-75}$$

represents the impulse response of an AR(1) stable system. Using (14-67) together with (14-72) and (14-75), we get the output autocorrelation sequence of an AR(1) process to be

$$R_{XX}(n) = \sigma_W^2 \delta(n) * \{a^{-n}\} * \{a^n\} = \sigma_W^2 \sum_{k=0}^{\infty} a^{|n|+k} a^k = \sigma_W^2 \frac{a^{|n|}}{1-a^2} \tag{14-76}$$

where we have made use of the discrete version of (14-46). The normalized (in terms of $R_{XX}(0)$) output autocorrelation sequence is given by

$$\rho_X(n) = \frac{R_{XX}(n)}{R_{XX}(0)} = a^{|n|}, \quad |n| \geq 0. \tag{14-77}$$

It is instructive to compare an AR(1) model discussed above by superimposing a random component to it, which may be an error term associated with observing a first order AR process $X(n)$. Thus

$$Y(n) = X(n) + V(n) \tag{14-78}$$

where $X(n) \sim$ AR(1) as in (14-73), and $V(n)$ is an uncorrelated random sequence with zero mean and variance σ_V^2 that is also uncorrelated with $\{W(n)\}$. From (14-73), (14-78) we obtain the output autocorrelation of the observed process $Y(n)$ to be

$$\begin{aligned} R_{YY}(n) &= R_{XX}(n) + R_{VV}(n) = R_{XX}(n) + \sigma_V^2 \delta(n) \\ &= \sigma_W^2 \frac{a^{|n|}}{1-a^2} + \sigma_V^2 \delta(n) \end{aligned} \tag{14-79}$$

so that its normalized version is given by

$$\rho_Y(n) \triangleq \frac{R_{YY}(n)}{R_{YY}(0)} = \begin{cases} 1 & n = 0 \\ c\, a^{|n|} & n = \pm 1, \ \pm 2, \cdots \end{cases} \tag{14-80}$$

where

$$c = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_V^2 (1 - a^2)} < 1. \tag{14-81}$$

Eqs. (14-77) and (14-80) demonstrate the effect of superimposing an error sequence on an AR(1) model. For non-zero lags, the autocorrelation of the observed sequence $\{Y(n)\}$ is reduced by a constant factor compared to the original process $\{X(n)\}$.

From (14-78), the superimposed error sequence $V(n)$ only affects the corresponding term in $Y(n)$ (term by term). However, a particular term in the "input sequence" $W(n)$ affects $X(n)$ and $Y(n)$ as well as *all* subsequent observations.

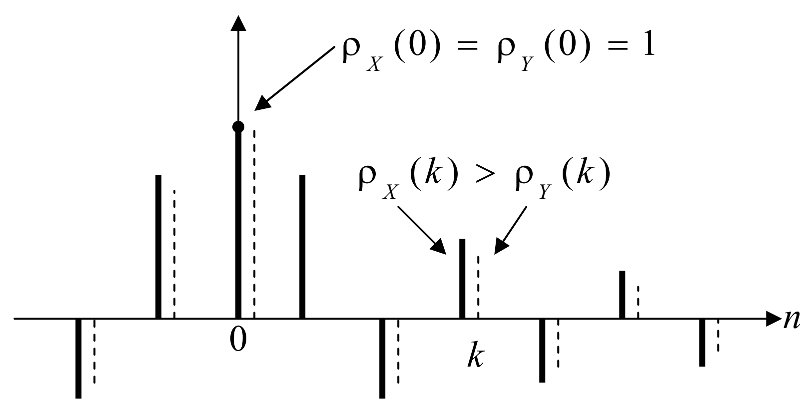


Fig. 14.13

AR(2) Process: An AR(2) process has the form

$$X(n) = a_1 X(n-1) + a_2 X(n-2) + W(n) \tag{14-82}$$

and from (14-70) the corresponding transfer function is given by

$$H(z) = \sum_{n=0}^{\infty} h(n) z^{-n} = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}} = \frac{b_1}{1 - \lambda_1 z^{-1}} + \frac{b_2}{1 - \lambda_2 z^{-1}} \tag{14-83}$$

so that

$$h(0) = 1,\; h(1) = a_1,\; h(n) = a_1 h(n-1) + a_2 h(n-2), \quad n \ge 2 \tag{14-84}$$

and in term of the poles λ_1 and λ_2 of the transfer function, from (14-83) we have

$$h(n) = b_1 \lambda_1^n + b_2 \lambda_2^n, \qquad n \ge 0 \tag{14-85}$$

that represents the impulse response of the system.
From (14-84)-(14-85), we also have $b_1 + b_2 = 1, \qquad b_1\lambda_1 + b_2\lambda_2 = a_1.$
From (14-83),

$$\lambda_1 + \lambda_2 = a_1, \qquad \lambda_1\lambda_2 = -a_2, \tag{14-86}$$

and $H(z)$ stable implies $|\lambda_1|<1, \; |\lambda_2|<1.$

Further, using (14-82) the output autocorrelations satisfy the recursion

$$\begin{aligned} R_{XX}(n) &= E\{X(n+m)X^*(m)\} \\ &= E\{[a_1X(n+m-1)+a_2X(n+m-2)]X^*(m)\} \\ &\quad + E\{\overset{0}{W(n+m)X^*(m)}\} \\ &= a_1R_{XX}(n-1)+a_2R_{XX}(n-2) \end{aligned} \qquad (14\text{-}87)$$

and hence their normalized version is given by

$$\rho_X(n) \triangleq \frac{R_{XX}(n)}{R_{XX}(0)} = a_1\rho_X(n-1)+a_2\rho_X(n-2). \qquad (14\text{-}88)$$

By direct calculation using (14-67), the output autocorrelations are given by

$$\begin{aligned} R_{XX}(n) &= R_{WW}(n)*h^*(-n)*h(n) = \sigma_W^2 h^*(-n)*h(n) \\ &= \sigma_W^2 \sum_{k=0}^{\infty} h^*(n+k)*h(k) \\ &= \sigma_W^2 \left(\frac{|b_1|^2(\lambda_1^*)^n}{1-|\lambda_1|^2} + \frac{b_1^*b_2(\lambda_1^*)^n}{1-\lambda_1^*\lambda_2} + \frac{b_1b_2^*(\lambda_2^*)^n}{1-\lambda_1\lambda_2^*} + \frac{|b_2|^2(\lambda_2^*)^n}{1-|\lambda_2|^2} \right) \end{aligned} \qquad (14\text{-}89)$$

where we have made use of (14-85). From (14-89), the normalized output autocorrelations may be expressed as

$$\rho_X(n) = \frac{R_{XX}(n)}{R_{XX}(0)} = c_1 \lambda_1^{*n} + c_2 \lambda_2^{*n} \tag{14-90}$$

where c_1 and c_2 are appropriate constants.

Damped Exponentials: When the second order system in (14-83)-(14-85) is real and corresponds to a damped exponential response, the poles are complex conjugate which gives $a_1^2 + 4a_2 < 0$ in (14-83). Thus

$$\lambda_1 = r\, e^{-j\theta}, \qquad \lambda_2 = \lambda_1^*, \qquad r < 1. \tag{14-91}$$

In that case $c_1 = c_2^* = c\, e^{j\varphi}$ in (14-90) so that the normalized correlations there reduce to

$$\rho_X(n) = 2\,\mathrm{Re}\{c_1 \lambda_1^{*n}\} = 2cr^n \cos(n\theta + \varphi). \tag{14-92}$$

But from (14-86)

$$\lambda_1 + \lambda_2 = 2r\cos\theta = a_1, \qquad r^2 = -a_2 < 1, \tag{14-93}$$

and hence $2r\sin\theta = \sqrt{-(a_1^2 + 4a_2)} > 0$ which gives

$$\tan\theta = \frac{\sqrt{-(a_1^2 + 4a_2)}}{a_1}. \qquad (14\text{-}94)$$

Also from (14-88)

$$\rho_X(1) = a_1\rho_X(0) + a_2\rho_X(-1) = a_1 + a_2\rho_X(1)$$

so that

$$\rho_X(1) = \frac{a_1}{1-a_2} = 2cr\cos(\theta+\varphi) \qquad (14\text{-}95)$$

where the later form is obtained from (14-92) with $n = 1$. But $\rho_X(0) = 1$ in (14-92) gives

$$2c\cos\varphi = 1, \quad \text{or} \quad c = 1/2\cos\varphi. \qquad (14\text{-}96)$$

Substituting (14-96) into (14-92) and (14-95) we obtain the normalized output autocorrelations to be

$$\rho_X(n) = (-a_2)^{n/2} \frac{\cos(n\theta + \varphi)}{\cos\varphi}, \qquad -a_2 < 1 \tag{14-97}$$

where φ satisfies

$$\frac{\cos(\theta + \varphi)}{\cos\theta} = \frac{a_1}{1 - a_2} \frac{1}{\sqrt{-a_2}}. \tag{14-98}$$

Thus the normalized autocorrelations of a damped second order system with real coefficients subject to random uncorrelated impulses satisfy (14-97).

More on ARMA processes

From (14-70) an ARMA (p, q) system has only $p + q + 1$ independent coefficients, $(a_k, k = 1 \rightarrow p,\ b_i, i = 0 \rightarrow q)$, and hence its impulse response sequence $\{h_k\}$ also must exhibit a similar dependence among them. In fact according to P. Dienes (*The Taylor series*, 1931), an old result due to Kronecker[1] (1881) states that the necessary and sufficient condition for $H(z)=\sum_{k=0}^{\infty} h_k z^{-k}$ to represent a rational system (ARMA) is that

$$\det H_n = 0, \qquad n \geq N \quad \text{(for all sufficiently large } n\text{)}, \qquad (14\text{-}99)$$

where

$$H_n \triangleq \begin{pmatrix} h_0 & h_1 & h_2 & \cdots & h_n \\ h_1 & h_2 & h_3 & \cdots & h_{n+1} \\ \vdots & & & & \\ h_n & h_{n+1} & h_{n+2} & \cdots & h_{2n} \end{pmatrix}. \qquad (14\text{-}100)$$

i.e., In the case of rational systems for all sufficiently large *n*, the Hankel matrices H_n in (14-100) all have the same rank.

The necessary part easily follows from (14-70) by cross multiplying and equating coefficients of like powers of z^{-k}, $k=0,1,2,\cdots$.

---

[1]Among other things "*God created the integers and the rest is the work of man.*" (Leopold Kronecker)

This gives

$$
\begin{aligned}
b_0 &= h_0 \\
b_1 &= h_0 a_1 + h_1 \\
&\vdots \\
b_q &= h_0 a_q + h_1 a_{q-1} + \cdots + h_m
\end{aligned}
\tag{14-101}
$$

$$0 = h_0 a_{q+i} + h_1 a_{q+i-1} + \cdots + h_{q+i-1} a_1 + h_{q+i}, \;\; i \geq 1. \tag{14-102}$$

For systems with $q \leq p-1$, letting $i = p-q, p-q+1, \cdots, 2p-q$ in (14-102) we get

$$
\begin{aligned}
& h_0 a_p + h_1 a_{p-1} + \cdots + h_{p-1} a_1 + h_p = 0 \\
& \;\vdots \\
& h_p a_p + h_{p+1} a_{p-1} + \cdots + h_{2p-1} a_1 + h_{2p} = 0
\end{aligned}
\tag{14-103}
$$

which gives det $H_p = 0$. Similarly $i = p-q+1, \cdots$ gives

$$
\begin{aligned}
&h_0 a_{p+1} + h_1 a_p + \cdots + h_{p+1} = 0 \\
&h_1 a_{p+1} + h_2 a_p + \cdots + h_{p+2} = 0 \\
&\quad \vdots \\
&h_{p+1} a_{p+1} + h_{p+2} a_p + \cdots + h_{2p+2} = 0,
\end{aligned}
\qquad (14\text{-}104)
$$

and that gives det $H_{p+1} = 0$ etc. (Notice that $a_{p+k} = 0, \ k = 1, 2, \cdots$)
(For sufficiency proof, see Dienes.)
It is possible to obtain similar determinantial conditions for ARMA systems in terms of Hankel matrices generated from its output autocorrelation sequence.

Referring back to the ARMA (p, q) model in (14-68), the input white noise process $w(n)$ there is uncorrelated with its own past sample values as well as the past values of the system output. This gives

$$
E\{w(n)w^*(n-k)\} = 0, \quad k \geq 1 \qquad (14\text{-}105)
$$

$$
E\{w(n)x^*(n-k)\} = 0, \quad k \geq 1. \qquad (14\text{-}106)
$$

Together with (14-68), we obtain

$$\begin{aligned} r_i &= E\{x(n)x^*(n-i)\} \\ &= -\sum_{k=1}^{p} a_k \{x(n-k)x^*(n-i)\} + \sum_{k=0}^{q} b_k \{w(n-k)w^*(n-i)\} \\ &= -\sum_{k=1}^{p} a_k r_{i-k} + \sum_{k=0}^{q} b_k \{w(n-k)x^*(n-i)\} \end{aligned} \tag{14-107}$$

and hence in general

$$\sum_{k=1}^{p} a_k r_{i-k} + r_i \neq 0, \qquad i \leq q \tag{14-108}$$

and

$$\sum_{k=1}^{p} a_k r_{i-k} + r_i = 0, \qquad i \geq q+1. \tag{14-109}$$

Notice that (14-109) is the same as (14-102) with $\{h_k\}$ replaced by $\{r_k\}$ and hence the Kronecker conditions for rational systems can be expressed in terms of its output autocorrelations as well. Thus if $X(n) \sim$ ARMA (p, q) represents a wide sense stationary stochastic process, then its output autocorrelation sequence $\{r_k\}$ satisfies

$$\text{rank } D_{p-1} = \text{rank } D_{p+k} = p, \qquad k \geq 0, \tag{14-110}$$

where

$$D_k \triangleq \begin{pmatrix} r_0 & r_1 & r_2 & \cdots & r_k \\ r_1 & r_2 & r_3 & \cdots & r_{k+1} \\ \vdots & & & & \\ r_k & r_{k+1} & r_{k+2} & \cdots & r_{2k} \end{pmatrix} \tag{14-111}$$

represents the $(k+1)\times(k+1)$ Hankel matrix generated from $r_0, r_1, \cdots, r_k, \cdots, r_{2k}$. It follows that for ARMA (p, q) systems, we have

$$\det D_n = 0, \qquad \text{for all sufficiently large } n. \tag{14-112}$$

15. Poisson Processes

In Lecture 4, we introduced Poisson arrivals as the limiting behavior of Binomial random variables. (Refer to Poisson approximation of Binomial random variables.)

From the discussion there (see (4-6)-(4-8) Lecture 4)

$$P\left\{\begin{array}{l}\text{"}k\text{ arrivals occur in an} \\ \text{interval of duration }\Delta\text{"}\end{array}\right\} = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \cdots \tag{15-1}$$

where

$$\lambda = np = \mu T \cdot \frac{\Delta}{T} = \mu\Delta \tag{15-2}$$

Fig. 15.1

It follows that (refer to Fig. 15.1)

$$P\left\{\begin{array}{l}\text{"}k \text{ arrivals occur in an} \\ \text{ interval of duration } 2\Delta\text{"}\end{array}\right\} = e^{-2\lambda}\frac{(2\lambda)^k}{k!}, \qquad k = 0,\ 1,\ 2, \cdots, \tag{15-3}$$

since in that case

$$np_1 = \mu T \cdot \frac{2\Delta}{T} = 2\mu\Delta = 2\lambda. \tag{15-4}$$

From (15-1)-(15-4), Poisson arrivals over an interval form a Poisson random variable whose parameter depends on the duration of that interval. Moreover because of the Bernoulli nature of the underlying basic random arrivals, events over nonoverlapping intervals are independent. We shall use these two key observations to define a Poisson process formally. (Refer to Example 9-5, Text)

Definition: $X(t) = n(0, t)$ represents a Poisson process if

(i) the number of arrivals $n(t_1, t_2)$ in an interval (t_1, t_2) of length $t = t_2 - t_1$ is a Poisson random variable with parameter λt. Thus

$$P\{n(t_1,t_2)=k\}=e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \quad k=0,1,2,\cdots, \quad t=t_2-t_1 \tag{15-5}$$

and

(ii) If the intervals (t_1, t_2) and (t_3, t_4) are nonoverlapping, then the random variables $n(t_1, t_2)$ and $n(t_3, t_4)$ are independent.

Since $n(0, t) \sim P(\lambda t)$, we have

$$E[X(t)]=E[n(0,t)]=\lambda t \tag{15-6}$$

and

$$E[X^2(t)]=E[n^2(0,t)]=\lambda t+\lambda^2 t^2. \tag{15-7}$$

To determine the autocorrelation function $R_{XX}(t_1,t_2)$, let $t_2 > t_1$, then from (ii) above $n(0, t_1)$ and $n(t_1, t_2)$ are independent Poisson random variables with parameters λt_1 and $\lambda(t_2-t_1)$ respectively. Thus

$$E[n(0,t_1)n(t_1,t_2)]=E[n(0,t_1)]E[n(t_1,t_2)]=\lambda^2 t_1(t_2-t_1). \tag{15-8}$$

But

$$n(t_1, t_2) = n(0, t_2) - n(0, t_1) = X(t_2) - X(t_1)$$

and hence the left side if (15-8) can be rewritten as

$$E[X(t_1)\{X(t_2) - X(t_1)\}] = R_{XX}(t_1, t_2) - E[X^2(t_1)]. \tag{15-9}$$

Using (15-7) in (15-9) together with (15-8), we obtain

$$\begin{aligned} R_{XX}(t_1, t_2) &= \lambda^2 t_1 (t_2 - t_1) + E[X^2(t_1)] \\ &= \lambda t_1 + \lambda^2 t_1\, t_2, \quad t_2 \geq t_1 . \end{aligned} \tag{15-10}$$

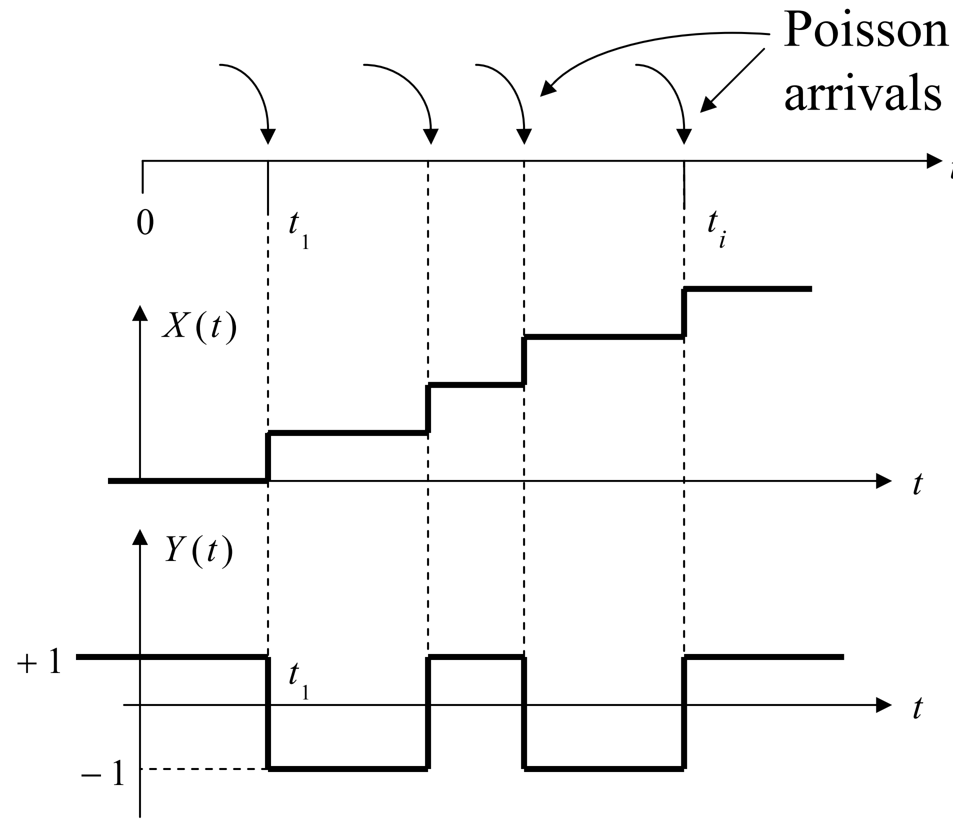
Similarly

$$R_{XX}(t_1, t_2) = \lambda t_2 + \lambda^2 t_1\, t_2, \quad t_2 < t_1 . \tag{15-11}$$

Thus

$$R_{XX}(t_1, t_2) = \lambda^2 t_1\, t_2 + \lambda \min(t_1, t_2). \tag{15-12}$$

From (15-12), notice that the Poisson process $X(t)$ *does not* represent a wide sense stationary process.



Define a binary level process

Fig. 15.2

$$Y(t) = (-1)^{X(t)} \tag{15-13}$$

that represents a telegraph signal (Fig. 15.2). Notice that the transition instants $\{t_i\}$ are random. (see Example 9-6, Text for the mean and autocorrelation function of a telegraph signal). Although $X(t)$ does not represent a wide sense stationary process, its derivative $X'(t)$ *does* represent a wide sense stationary process.

$$X(t) \longrightarrow \boxed{\dfrac{d(\cdot)}{dt}} \longrightarrow X'(t)$$

Fig. 15.3 (Derivative as a LTI system)

To see this, we can make use of Fig. 14.7 and (14-34)-(14-37). From there

$$\mu_{X'}(t) = \frac{d\mu_X(t)}{dt} = \frac{d\lambda t}{dt} = \lambda, \quad \textit{a constant} \tag{15-14}$$

and

$$R_{XX'}(t_1, t_2) = \frac{\partial R_{XX}(t_1, t_2)}{\partial t_2} = \begin{cases} \lambda^2 t_1 & t_1 \le t_2 \\ \lambda^2 t_1 + \lambda & t_1 > t_2 \end{cases}$$

$$= \lambda^2 t_1 + \lambda\, U(t_1 - t_2) \tag{15-15}$$

and

$$R_{X'X'}(t_1, t_2) = \frac{\partial R_{XX'}(t_1, t_2)}{\partial t_1} = \lambda^2 + \lambda\, \delta(t_1 - t_2). \tag{15-16}$$

From (15-14) and (15-16) it follows that $X'(t)$ is a wide sense stationary process. Thus nonstationary inputs to linear systems *can* lead to wide sense stationary outputs, an interesting observation.

- **Sum of Poisson Processes:**

  If $X_1(t)$ and $X_2(t)$ represent two independent Poisson processes, then their sum $X_1(t) + X_2(t)$ is also a Poisson process with parameter $(\lambda_1 + \lambda_2)t$. (Follows from (6-86), Text and the definition of the Poisson process in (i) and (ii)).

- **Random selection of Poisson Points:**

  Let $t_1, t_2, \cdots, t_i, \cdots$ represent random arrival points associated with a Poisson process $X(t)$ with parameter λt, and associated with each arrival point, define an independent Bernoulli random variable N_i, where

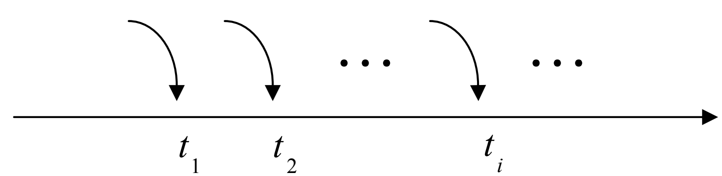


Fig. 15.4

$$P(N_i = 1) = p, \qquad P(N_i = 0) = q = 1 - p. \qquad (15\text{-}17)$$

Define the processes

$$Y(t)=\sum_{i=1}^{X(t)} N_i \quad ; \quad Z(t)=\sum_{i=1}^{X(t)} (1-N_i)=X(t)-Y(t) \qquad (15\text{-}18)$$

we claim that both $Y(t)$ and $Z(t)$ are independent Poisson processes with parameters λpt and λqt respectively.

Proof:

$$Y(t)=\sum_{n=k}^{\infty} P\{Y(t)=k \mid X(t)=n\} P\{X(t)=n)\}. \qquad (15\text{-}19)$$

But given $X(t)=n,$ we have $Y(t)=\sum_{i=1}^{n} N_i \ \sim \ B(n,p)$ so that

$$P\{Y(t)=k \mid X(t)=n\}=\binom{n}{k} p^k q^{n-k}, \qquad 0 \le k \le n, \qquad (15\text{-}20)$$

and

$$P\{X(t)=n\}=e^{-\lambda t} \frac{(\lambda t)^n}{n!}. \qquad (15\text{-}21)$$

Substituting (15-20)-(15-21) into (15-19) we get

$$P\{Y(t)=k\} = e^{-\lambda t}\sum_{n=k}^{\infty}\frac{n!}{(n-k)!k!}p^k q^{n-k}\frac{(\lambda t)^n}{n!} = \frac{p^k e^{-\lambda t}}{k!}(\lambda t)^k \underbrace{\sum_{n=k}^{\infty}\frac{(q\lambda t)^{n-k}}{(n-k)!}}_{e^{q\lambda t}}$$

$$= (\lambda pt)^k \frac{e^{-(1-q)\lambda t}}{k!} = e^{-\lambda pt}\frac{(\lambda pt)^k}{k!}, \qquad k = 0, 1, 2, \cdots$$

$$\sim \quad P(\lambda pt). \tag{15-22}$$

More generally,

$$P\{Y(t)=k, Z(t)=m\} = P\{Y(t)=k, X(t)-Y(t)=m\}$$

$$= P\{Y(t)=k, X(t)=k+m\}$$

$$= P\{Y(t)=k \mid X(t)=k+m\}P\{X(t)=k+m\}$$

$$= \binom{k+m}{k}p^k q^m \cdot e^{-\lambda t}\frac{(\lambda t)^{k+m}}{(k+m)!} = \underbrace{e^{-\lambda pt}\frac{(\lambda pt)^n}{k!}}_{P(Y(t)=k)} \quad \underbrace{e^{-\lambda qt}\frac{(\lambda qt)^n}{m!}}_{P(Z(t)=m)}$$

$$= P\{Y(t)=k\}P\{Z(t)=m\}, \tag{15-23}$$

which completes the proof.

Notice that $Y(t)$ and $Z(t)$ are generated as a result of random Bernoulli selections from the original Poisson process $X(t)$ (Fig. 15.5), where each arrival gets tossed over to either $Y(t)$ with probability p or to $Z(t)$ with probability q. Each such sub-arrival stream is also a Poisson process. Thus random selection of Poisson points preserve the Poisson nature of the resulting processes. However, as we shall see deterministic selection from a Poisson process destroys the Poisson property for the resulting processes.

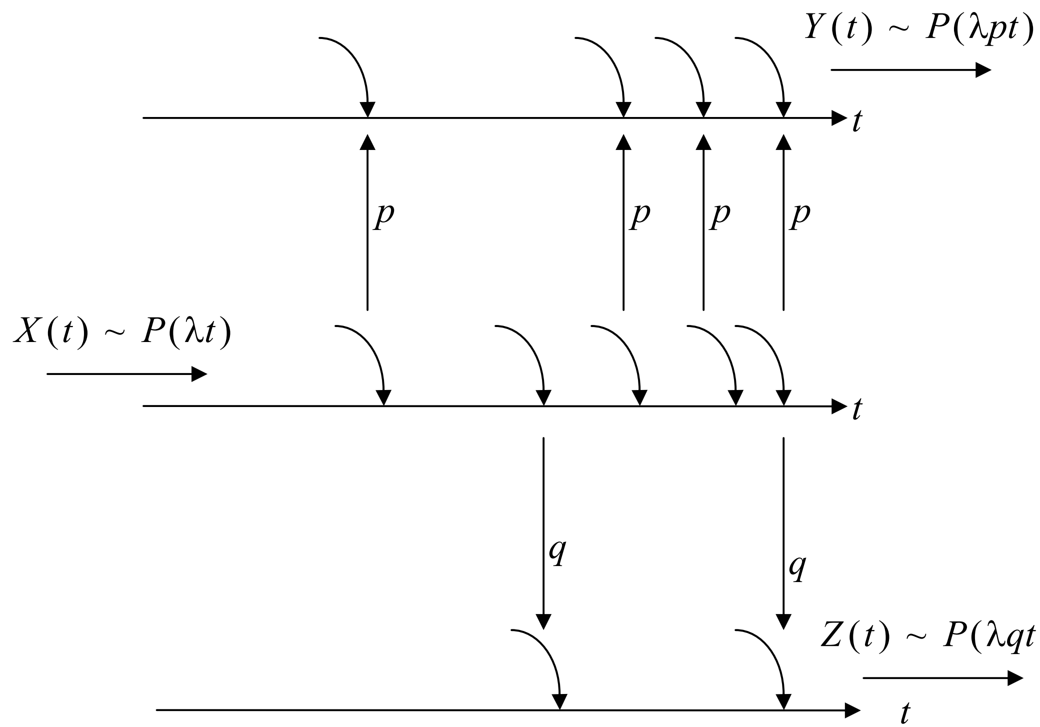


Fig. 15.5

Inter-arrival Distribution for Poisson Processes

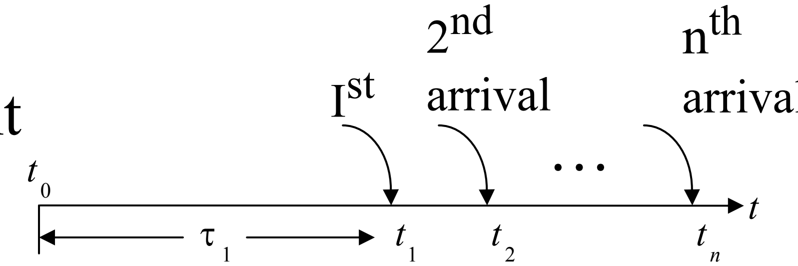


Fig. 15.6

Let τ_1 denote the time interval (delay) to the first arrival from *any* fixed point t_0. To determine the probability distribution of the random variable τ_1, we argue as follows: Observe that the event $"\tau_1 > t"$ is the same as "$n(t_0, t_0+t) = 0$", or the complement event $"\tau_1 \le t"$ is the same as the event "$n(t_0, t_0+t) > 0$" . Hence the distribution function of τ_1 is given by

$$F_{\tau_1}(t) \triangleq P\{\tau_1 \le t\} = P\{X(t) > 0\} = P\{n(t_0, t_0 + t) > 0\}$$

$$= 1 - P\{n(t_0, t_0 + t) = 0\} = 1 - e^{-\lambda t} \qquad (15\text{-}24)$$

(use (15-5)), and hence its derivative gives the probability density function for τ_1 to be

$$f_{\tau_1}(t) = \frac{dF_{\tau_1}(t)}{dt} = \lambda e^{-\lambda t}, \qquad t \ge 0 \qquad (15\text{-}25)$$

i.e., τ_1 is an exponential random variable with parameter λ so that $E(\tau_1) = 1/\lambda$.

Similarly, let t_n represent the n^{th} random arrival point for a Poisson process. Then

$$F_{t_n}(t) \triangleq P\{t_n \le t\} = P\{X(t) \ge n\}$$

$$= 1 - P\{X(t) < n\} = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad (15\text{-}26)$$

and hence

$$f_{t_n}(x) = \frac{dF_{t_n}(x)}{dx} = -\sum_{k=1}^{n-1} \frac{\lambda(\lambda x)^{k-1}}{(k-1)!} e^{-\lambda x} + \sum_{k=0}^{n-1} \frac{\lambda(\lambda x)^k}{k!} e^{-\lambda x}$$

$$= \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, \qquad x \ge 0 \qquad (15\text{-}27)$$

which represents a gamma density function. i.e., the waiting time to the n^{th} Poisson arrival instant has a gamma distribution. Moreover

$$t_n = \sum_{i=1}^{n} \tau_i$$

where τ_i is the random inter-arrival duration between the $(i-1)^{th}$ and i^{th} events. Notice that τ_i s are independent, identically distributed random variables. Hence using their characteristic functions, it follows that all inter-arrival durations of a Poisson process are independent exponential random variables with common parameter λ. i.e.,

$$f_{\tau_i}(t) = \lambda e^{-\lambda t}, \qquad t \geq 0. \tag{15-28}$$

Alternatively, from (15-24)-(15-25), we have τ_1 is an exponential random variable. By repeating that argument after shifting t_0 to the new point t_1 in Fig. 15.6, we conclude that τ_2 is an exponential random variable. Thus the sequence $\tau_1, \tau_2, \cdots, \tau_n, \cdots$ are independent exponential random variables with common p.d.f as in (15-25).

Thus if we systematically tag every m^{th} outcome of a Poisson process $X(t)$ with parameter λt to generate a new process $e(t)$, then the inter-arrival time between any two events of $e(t)$ is a gamma random variable.

Notice that

$$E[e(t)] = m/\lambda, \text{ and if } \lambda = m\mu, \text{ then } E[e(t)] = 1/\mu.$$

The inter-arrival time of $e(t)$ in that case represents an Erlang-m random variable, and $e(t)$ an Erlang-m process (see (10-90), Text). In summary, if Poisson arrivals are randomly redirected to form new queues, then each such queue generates a new Poisson process (Fig. 15.5). However if the arrivals are systematically redirected (Ist arrival to Ist counter, 2nd arrival to 2nd counter, $\cdots$, m^{th} to m^{th}, $(m+1)^{st}$ arrival to Ist counter, $\cdots$), then the new subqueues form Erlang-m processes.

Interestingly, we can also derive the key Poisson properties (15-5) and (15-25) by starting from a simple axiomatic approach as shown below:

Axiomatic Development of Poisson Processes:

The defining properties of a Poisson process are that in any "small" interval Δt, one event can occur with probability that is proportional to Δt. Further, the probability that two or more events occur in that interval is proportional to $o(\Delta t)$, (higher powers of Δt), and events over nonoverlapping intervals are independent of each other. This gives rise to the following axioms.

Axioms:

$$\left.\begin{array}{l}
\text{(i) } P\{n(t, t+\Delta t)=1\}=\lambda \Delta t+o(\Delta t) \\[2ex]
\text{(ii) } P\{n(t, t+\Delta t)=0\}=1-\lambda \Delta t+o(\Delta t) \\[2ex]
\text{(iii) } P\{n(t, t+\Delta t) \geq 2\}=o(\Delta t) \\
\text{and} \\
\text{(iv) } n(t, t+\Delta t) \text{ is independent of } n(0, t)
\end{array}\right\} \qquad (15\text{-}29)$$

Notice that axiom (iii) specifies that the events occur singly, and axiom (iv) specifies the randomness of the entire series. Axiom(ii) follows from (i) and (iii) together with the axiom of total probability.

We shall use these axiom to rederive (15-25) first:
Let t_0 be *any* fixed point (see Fig. 15.6) and let $t_0 + \tau_1$ represent the time of the first arrival after t_0 . Notice that the random variable τ_1 is independent of the occurrences prior to the instant t_0 (Axiom (iv)). With $F_{\tau_1}(t) = P\{\tau_1 \le t\}$ representing the distribution function of τ_1, as in (15-24) define $Q(t) \triangleq 1 - F_{\tau_1}(t) = P\{\tau_1 > t\}$. Then for $\Delta t > 0$

$$
\begin{aligned}
Q(t+\Delta t) &= P\{\tau_1 > t + \Delta t\} \\
&= P\{\tau_1 > t, \text{ and no event occurs in } (t_0 + t,\ t_0 + t + \Delta t)\} \\
&= P\{\tau_1 > t,\ n(t_0 + t,\ t_0 + t + \Delta t) = 0\} \\
&= P\{n(t_0 + t,\ t_0 + t + \Delta t) = 0 \mid \tau_1 > t\} P\{\tau_1 > t\}.
\end{aligned}
$$

From axiom (iv), the conditional probability in the above expression is not affected by the event $\{\tau_1 > t\}$ which refers to $\{n(t_0, t_0 + t) = 0\}$, i.e., to events before $t_0 + t$, and hence the unconditional probability in axiom (ii) can be used there. Thus

$$Q(t+\Delta t) = [1 - \lambda \Delta t + o(\Delta t)]Q(t)$$

or

$$\lim_{\Delta t \to 0} \frac{Q(t+\Delta t) - Q(t)}{\Delta t} = Q'(t) = -\lambda Q(t) \quad \Rightarrow \quad Q(t) = ce^{-\lambda t}.$$

But $c = Q(0) = P\{\tau_1 > 0\} = 1$ so that

$$Q(t) = 1 - F_{\tau_1}(t) = e^{-\lambda t}$$

or

$$F_{\tau_1}(t) = 1 - e^{-\lambda t}, \qquad t \geq 0$$

which gives

$$f_{\tau_1}(t) = \frac{dF_{\tau_1}(t)}{dt} = \lambda e^{-\lambda t}, \qquad t \geq 0 \tag{15-30}$$

to be the p.d.f of τ_1 as in (15-25).

Similarly (15-5) can be derived from axioms (i)-(iv) in (15-29) as well. To see this, let

$$p_k(t) \triangleq P\{n(0,t) = k)\}, \quad k = 0,\ 1,\ 2, \cdots$$

represent the probability that the total number of arrivals in the interval (0, t) equals k. Then

$$p_k(t + \Delta t) = P\{n(0, t + \Delta t) = k)\} = P\{X_1 \cup X_2 \cup X_3\}$$

where the events

$$X_1 \triangleq \text{"}n(0,t) = k, \ \text{ and } \ n(t,\ t+\Delta t) = 0\text{"}$$
$$X_2 \triangleq \text{"}n(0,t) = k-1, \ \text{ and } \ n(t,\ t+\Delta t) = 1\text{"}$$
$$X_3 \triangleq \text{"}n(0,t) = k-i, \ \text{ and } \ n(t,\ t+\Delta t) = i \geq 2\text{"}$$

are mutually exclusive. Thus

$$p_k(t+\Delta t) = P(X_1) + P(X_2) + P(X_3).$$

But as before

$$\begin{aligned} P(X_1) &= P\{n(t,t+\Delta t) = 0 \mid n(0,t) = k\} P\{n(0,t) = k\} \\ &= P\{n(t,t+\Delta t) = 0\} P\{n(0,t) = k\} \\ &= (1-\lambda \Delta t)\, p_k(t) \end{aligned}$$

$$\begin{aligned} P(X_2) &= P\{n(t,t+\Delta t) = 1 \mid n(0,t) = k-1\} P\{n(0,t) = k-1\} \\ &= \lambda \Delta t \; p_{k-1} \Delta t \end{aligned}$$

and

$$P(X_3) = 0$$

where once again we have made use of axioms (i)-(iv) in (15-29). This gives

$$p_k(t+\Delta t) = (1-\lambda\Delta t)\, p_k(t) + \lambda\Delta t p_{k-1}(t)$$

or with

$$\lim_{\Delta t \to 0} \frac{p_k(t+\Delta t) - p_k(t)}{\Delta t} = p_k'(t)$$

we get the differential equation

$$p_k'(t) = -\lambda p_k(t) - \lambda p_{k-1}(t), \qquad k = 0, 1, 2, \cdots$$

whose solution gives (15-5). Here $p_{-1}(t) \equiv 0$. [Solution to the above differential equation is worked out in (16-36)-(16-41), Text]. This is completes the axiomatic development for Poisson processes.

Poisson Departures between Exponential Inter-arrivals

Let $X(t) \sim P(\lambda t)$ and $Y(t) \sim P(\mu t)$ represent two independent Poisson processes called *arrival* and *departure* processes.

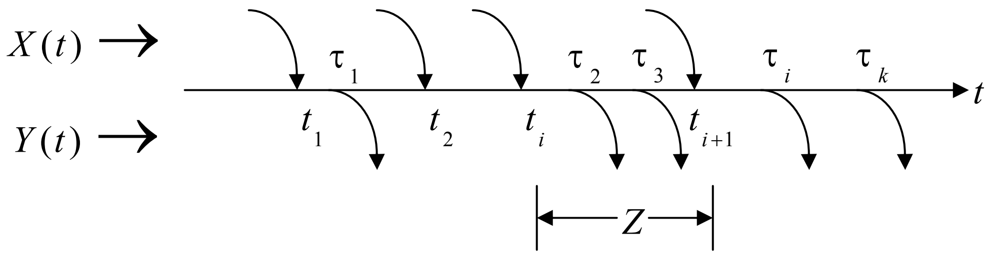


Fig. 15.7

Let Z represent the random interval between *any* two successive arrivals of $X(t)$. From (15-28), Z has an exponential distribution with parameter λ. Let N represent the number of "departures" of $Y(t)$ between *any* two successive arrivals of $X(t)$. Then from the Poisson nature of the departures we have

$$P\{N = k \mid Z = t\} = e^{-\mu t} \frac{(\mu t)^k}{k!}.$$

Thus

$$
\begin{aligned}
P\{N=k\} &= \int_0^{\infty} P\{N=k \mid Z=t\} f_Z(t)\,dt \\
&= \int_0^{\infty} e^{-\mu t} \frac{(\mu t)^k}{k!} \lambda e^{-\lambda t}\,dt \\
&= \frac{\lambda}{k!} \int_0^{\infty} (\mu t)^k e^{-(\lambda+\mu)t}\,dt \\
&= \frac{\lambda}{\lambda+\mu} \left(\frac{\mu}{\lambda+\mu} \right)^k \frac{1}{k!} \underbrace{\int_0^{\infty} x^k e^{-x}\,dx}_{k!} \\
&= \left(\frac{\lambda}{\lambda+\mu} \right) \left(\frac{\mu}{\lambda+\mu} \right)^k, \qquad k = 0, 1, 2, \cdots \qquad (15\text{-}31)
\end{aligned}
$$

i.e., the random variable N has a geometric distribution. Thus if customers come in and get out according to two independent Poisson processes at a counter, then the number of arrivals between any two departures has a geometric distribution. Similarly the number of departures between *any* two arrivals also represents another geometric distribution.

Stopping Times, Coupon Collecting, and Birthday Problems

Suppose a cereal manufacturer inserts a sample of one type of coupon randomly into each cereal box. Suppose there are n such distinct types of coupons. One interesting question is that how many boxes of cereal should one buy on the average in order to collect at least one coupon of each kind?

We shall reformulate the above problem in terms of Poisson processes. Let $X_1(t), X_2(t), \cdots, X_n(t)$ represent n *independent* identically distributed Poisson processes with common parameter λt. Let $t_{i1}, t_{i2}, \cdots$ represent the first, second, $\cdots$ random arrival instants of the process $X_i(t),\ \ i = 1, 2, \cdots, n.$ They will correspond to the first, second, $\cdots$ appearance of the i^{th} type coupon in the above problem. Let

$$X(t) \triangleq \sum_{i=1}^{n} X_i(t), \tag{15-32}$$

so that the sum $X(t)$ is also a Poisson process with parameter μt, where

$$\mu = n\lambda. \tag{15-33}$$

From Fig. 15.8, $1/\lambda$ represents The average inter-arrival duration between any two arrivals of $X_i(t), i = 1, 2, \cdots, n,$ whereas $1/\mu$ represents the average inter-arrival time for the combined sum process $X(t)$ in (15-32).

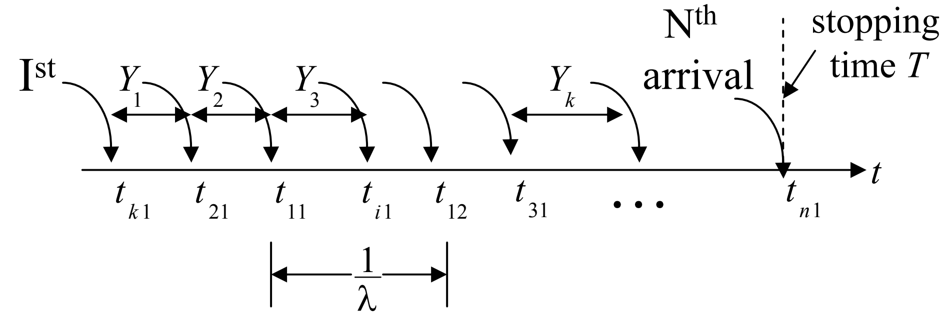


Fig. 15.8

Define the **stopping time** T to be that random time instant by which at least one arrival of $X_1(t), X_2(t), \cdots, X_n(t)$ has occurred. Clearly, we have

$$T = \max\,(t_{11}, t_{21}, \cdots, t_{i1}, \cdots, t_{n1}). \qquad (15\text{-}34)$$

But from (15-25), $t_{i1},\ \ i = 1, 2, \cdots, n$ are independent exponential random variables with common parameter λ. This gives

$$\begin{aligned} F_T(t) &= P\{T \le t\} = P\{\max\,(t_{11}, t_{21}, \cdots, t_{n1}) \le t\} \\ &= P\{t_{11} \le t,\ t_{21} \le t,\ \cdots,\ t_{n1} \le t\} \\ &= P\{t_{11} \le t\}P\{t_{21} \le t\}\cdots P\{t_{n1} \le t\} = [F_{ti}(t)]^n. \end{aligned}$$

Thus

$$F_T(t) = (1 - e^{-\lambda t})^n, \qquad t \geq 0 \tag{15-35}$$

represents the probability distribution function of the stopping time random variable T in (15-34). To compute its mean, we can make use of Eqs. (5-52)-(5-53) Text, that is valid for nonnegative random variables. From (15-35) we get

$$P(T > t) = 1 - F_T(t) = 1 - (1 - e^{-\lambda t})^n, \qquad t \geq 0$$

so that

$$E\{T\} = \int_0^\infty P(T > t)dt = \int_0^\infty \{1 - (1 - e^{-\lambda t})^n\}dt. \tag{15-36}$$

Let $1 - e^{-\lambda t} = x$, so that $\lambda\, e^{-\lambda t} dt = dx$, or $dt = \frac{dx}{\lambda(1-x)}$, and

$$E\{T\} = \frac{1}{\lambda}\int_0^1 (1 - x^n)\frac{dx}{1 - x} = \frac{1}{\lambda}\int_0^1 \frac{1-x^n}{1-x}\, dx$$

$$= \frac{1}{\lambda}\int_0^1 (1 + x + x^2 + \cdots + x^{n-1})dx = \frac{1}{\lambda}\sum_{k=1}^{n} \frac{x^k}{k}\Bigg|_0^1$$

$$E\{T\} = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) = \frac{n}{\mu}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)$$

$$\simeq \frac{1}{\mu} n(\ln n + \gamma) \tag{15-37}$$

where $\gamma \simeq 0.5772157\cdots$ is the Euler's constant[1].

Let the random variable N denote the total number of all arrivals up to the stopping time T, and $Y_k,\ k = 1, 2, \cdots, N$ the inter-arrival random variables associated with the sum process $X(t)$ (see Fig 15.8).

---

[1] **Euler's constant:** The series $\{1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln n\}$ converges, since

$$u_n \triangleq \int_0^1 \frac{x}{n(n+x)} dx = \int_0^1 \left(\frac{1}{n} - \frac{1}{n+x}\right) dx = \frac{1}{n} - \ln \frac{n+1}{n} > 0 \tag{1}$$

and $u_n \le \int_0^1 \frac{x}{n^2} dx \le \int_0^1 \frac{1}{n^2} dx = \frac{1}{n^2}$ so that $\sum_{n=1}^{\infty} u_n < \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} < \infty.$ Thus the series $\{u_n\}$ converges to some number $\gamma > 0$. From (1) we obtain also $\sum_{k=1}^{n} u_k = \sum_{k=1}^{n} \frac{1}{k} - \ln(n+1)$ so that

$$\lim_{n\to\infty} \{1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln n\} = \lim_{n\to\infty} \left\{\sum_{k=1}^{n} u_k + \ln \frac{n+1}{n}\right\} = \sum_{k=1}^{\infty} u_k = \gamma = 0.5772157\cdots.$$

Then we obtain the key relation

$$T = \sum_{i=1}^{N} Y_i \tag{15-38}$$

so that

$$E\{T \mid N = n\} = E\{\sum_{i=1}^{n} Y_i \mid N = n\} = E\{\sum_{i=1}^{n} Y_i\} = nE\{Y_i\} \tag{15-39}$$

since $\{Y_i\}$ and N are independent random variables, and hence

$$E\{T\} = E[E\{T \mid N = n\}] = E\{N\}E\{Y_i\} \tag{15-40}$$

But $Y_i \sim \text{Exponential}(\mu)$, so that $E\{Y_i\} = 1/\mu$ and substituting this into (15-37), we obtain

$$E\{N\} \simeq n(\ln n + \gamma). \tag{15-41}$$

Thus on the average a customer should buy about $n \ln n$, or slightly more, boxes to guarantee that at least one coupon of each type has been collected.

Next, consider a slight generalization to the above problem: What if two kinds of coupons (each of n type) are mixed up, and the objective is to collect one complete set of coupons of either kind?

Let $X_i(t)$ and $Y_i(t)$, $i = 1, 2, \cdots, n$ represent the two kinds of coupons (independent Poisson processes) that have been mixed up to form a single Poisson process $Z(t)$ with normalized parameter unity. i.e.,

$$Z(t) = \sum_{i=1}^{n} [X_i(t) + Y_i(t)] \sim P(t). \tag{15-42}$$

As before let $t_{i1}, t_{i2}, \cdots$ represent the first, second, $\cdots$ arrivals of the process $X_i(t)$, and $\tau_{i1}, \tau_{i2}, \cdots$ represent the first, second, arrivals of the process $Y_i(t)$, $i = 1, 2, \cdots, n.$

The stopping time T_1 in this case represents that random instant at which *either* all X – type *or* all Y – type have occurred at least once. Thus

$$T_1 = \min\{X, Y\} \tag{15-43}$$

where

$$X \triangleq \max\ (t_{11}, t_{21}, \cdots, t_{n1}) \tag{15-44}$$

and

$$Y \triangleq \max\ (\tau_{11}, \tau_{21}, \cdots, \tau_{n1}). \tag{15-45}$$

Notice that the random variables X and Y have the same distribution as in (15-35) with λ replaced by $1/2n$ (since $\mu = 1$ and there are $2n$ independent and identical processes in (15-42)), and hence

$$F_X(t) = F_Y(t) = (1 - e^{-t/2n})^n, \qquad t \geq 0. \tag{15-46}$$

Using (15-43), we get

$$
\begin{aligned}
F_{T_1}(t) &= P(T_1 \le t) = P(\min\{X,Y\} \le t) \\
&= 1 - P(\min\{X,Y\} > t) = 1 - P(X > t, Y > t) \\
&= 1 - P(X > t)P(Y > t) \\
&= 1 - (1 - F_X(t))(1 - F_Y(t)) \\
&= 1 - \{1 - (1 - e^{-t/2n})^n\}^2, \qquad t \ge 0 \qquad\qquad (15\text{-}47)
\end{aligned}
$$

to be the probability distribution function of the new stopping time T_1. Also as in (15-36)

$$
\begin{aligned}
E\{T_1\} &= \int_0^\infty P(T_1 > t)dt = \int_0^\infty \{1 - F_{T_1}(t)\}dt \\
&= \int_0^\infty \{1 - (1 - e^{-t/2n})^n\}^2 dt.
\end{aligned}
$$

Let $1 - e^{-t/2n} = x$, or $\frac{1}{2n} e^{-t/2n} dt = dx$, $dt = \frac{2n dx}{1-x}$.

$$
\begin{aligned}
E\{T_1\} &= 2n \int_0^1 (1 - x^n)^2 \frac{dx}{1-x} = 2n \int_0^1 \left(\frac{1 - x^n}{1 - x} \right) (1 - x^n)\, dx \\
&= 2n \int_0^1 (1 + x + x^2 + \cdots + x^{n-1})(1 - x^n) dx
\end{aligned}
$$

$$E\{T_1\} = 2n\int_0^1 (\sum_{k=0}^{n-1} x^k - \sum_{k=0}^{n-1} x^{n+k})\, dx$$

$$= 2n\left\{ \left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right) - \left(\frac{1}{n+1}+\frac{1}{n+2}+\cdots+\frac{1}{2n}\right) \right\}$$

$$\simeq 2n(\ln(n/2)+\gamma). \qquad (15\text{-}48)$$

Once again the total number of random arrivals N up to T_1 is related as in (15-38) where $Y_i \sim$ Exponential (1), and hence using (15-40) we get the average number of total arrivals up to the stopping time to be

$$E\{N\} = E\{T_1\} \simeq 2n(\ln(n/2)+\gamma). \qquad (15\text{-}49)$$

We can generalize the stopping times in yet another way:

Poisson Quotas

Let

$$X(t) = \sum_{i=1}^{n} X_i(t) \sim P(\mu t) \qquad (15\text{-}50)$$

where $X_i(t)$ are independent, identically distributed Poisson processes with common parameter $\lambda_i t$ so that $\mu = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. Suppose integers $m_1, m_2, \cdots, m_n$ represent the preassigned number of arrivals (quotas) required for processes $X_1(t), X_2(t), \cdots, X_n(t)$ in the sense that when m_i arrivals of the process $X_i(t)$ have occurred, the process $X_i(t)$ satisfies its "quota" requirement.

The **stopping time** T in this case is that random time instant at which *any* r processes have met their quota requirement where $r \le n$ is given. The problem is to determine the probability density function of the stopping time random variable T, and determine the mean and variance of the total number of random arrivals N up to the stopping time T.

Solution: As before let $t_{i1}, t_{i2}, \cdots$ represent the first, second, $\cdots$ arrivals of the i^{th} process $X_i(t)$, and define

$$Y_{ij} = t_{ij} - t_{i,j-1} \tag{15-51}$$

Notice that the inter-arrival times Y_{ij} and independent, exponential random variables with parameter λ_i, and hence

$$t_{i,m_i} = \sum_{j=1}^{m_i} Y_{ij} \sim \text{Gamma}(m_i, \lambda_i)$$

Define T_i to the **stopping time** for the i^{th} process; i.e., the occurrence of the m_i^{th} arrival equals T_i. Thus

$$T_i = t_{i,m_i} \sim \text{Gamma}(m_i, \lambda_i), \quad i = 1, 2, \cdots, n \tag{15-52}$$

or

$$f_{T_i}(t) = \frac{t^{m_i - 1}}{(m_i - 1)!} \lambda_i^{m_i} e^{-\lambda_i t}, \quad t \geq 0. \tag{15-53}$$

Since the n processes in (15-49) are independent, the associated stopping times $T_i, \quad i = 1, 2, \cdots, n$ defined in (15-53) are also *independent* random variables, a key observation.

Given *independent* gamma random variables $T_1, T_2, \cdots, T_n$, in (15-52)-(15-53) we form their order statistics: This gives

$$T_{(1)} < T_{(2)} < \cdots < T_{(r)} < \cdots < T_{(n)}. \tag{15-54}$$

Note that the two extremes in (15-54) represent

$$T_{(1)} = \min(T_1, T_2, \cdots, T_n) \tag{15-55}$$

and

$$T_{(n)} = \max(T_1, T_2, \cdots, T_n). \tag{15-56}$$

The desired stopping time T when r processes have satisfied their quota requirement is given by the r^{th} order statistics $T_{(r)}$. Thus

$$T = T_{(r)} \tag{15-57}$$

where $T_{(r)}$ is as in (15-54). We can use (7-14), Text to compute the probability density function of T. From there, the probability density function of the stopping time random variable T in (15-57) is given by

$$f_T(t) = \frac{n!}{(r-1)!(n-r)!} F_{T_i}^{k-1}(t)[1 - F_{T_i}(t)]^{n-k} f_{T_i}(t) \tag{15-58}$$

where $F_{T_i}(t)$ is the distribution of the i.i.d random variables T_i and $f_{T_i}(t)$ their density function given in (15-53). Integrating (15-53) by parts as in (4-37)-(4-38), Text, we obtain

$$F_{T_i}(t) = 1 - \sum_{k=0}^{m_i - 1} \frac{(\lambda_i t)^k}{k!} e^{-\lambda_i t}, \qquad t \ge 0. \tag{15-59}$$

Together with (15-53) and (15-59), Eq. (15-58) completely specifies the density function of the stopping time random variable T, where r types of arrival quota requirements have been satisfied.

If N represents the total number of *all* random arrivals up to T, then arguing as in (15-38)-(15-40) we get

$$T = \sum_{i=1}^{N} Y_i \qquad (15\text{-}60)$$

where Y_i are the inter-arrival intervals for the process $X(t)$, and hence

$$E\{T\} = E\{N\}E\{Y_i\} = \frac{1}{\mu} E\{N\} \qquad (15\text{-}61)$$

with normalized mean value $\mu(=1)$ for the sum process $X(t)$, we get

$$E\{N\} = E\{T\}. \qquad (15\text{-}62)$$

To relate the higher order moments of N and T we can use their characteristic functions. From (15-60)

$$E\{e^{j\omega T}\} = E\{e^{j\omega \sum_{i=1}^{N} Y_i}\} = E\{\underbrace{E[e^{j\omega \sum_{i=1}^{n} Y_i} \mid N = n]}_{[E\{e^{j\omega Y_i}\}]^n}\}$$

$$= E[\{E\{e^{j\omega Y_i}\} \mid N = n\}^n]. \qquad (15\text{-}63)$$

But $Y_i \sim$ Exponential (1) and independent of N so that

$$E\{e^{j\omega Y_i} \mid N=n\} = E\{e^{j\omega Y_i}\} = \frac{1}{1-j\omega}$$

and hence from (15-63)

$$E\{e^{j\omega T}\} = \sum_{n=0}^{\infty} [E\{e^{j\omega Y_i}\}]^n P(N=n) = E\left\{ \left(\tfrac{1}{1-j\omega}\right)^N \right\} = E\{(1-j\omega)^{-N}\}$$

which gives (expanding both sides)

$$\sum_{k=0}^{\infty} \frac{(j\omega)^k}{k!} E\{T^k\} = \sum_{k=0}^{\infty} \frac{(j\omega)^k}{k!} E\{N(N+1)\cdots(N+k-1)\}$$

or

$$E\{T^k\} = E\{N(N+1)\cdots(N+k-1)\}, \qquad (15\text{-}64)$$

a key identity. From (15-62) and (15-64), we get

$$\operatorname{var}\{N\} = \operatorname{var}\{T\} - E\{T\}. \qquad (15\text{-}65)$$

As an application of the Poisson quota problem, we can reexamine *the birthday pairing* problem discussed in Example 2-20, Text.

"Birthday Pairing" as Poisson Processes:

In the birthday pairing case (refer to Example 2-20, Text), we may assume that $n = 365$ possible birthdays in an year correspond to n independent identically distributed Poisson processes each with parameter $1/n$, and in that context each individual is an "arrival" corresponding to his/her particular "birth-day process". It follows that the birthday pairing problem (i.e., two people have the same birth date) corresponds to the first occurrence of the 2nd return for *any* one of the 365 processes. Hence

$$m_1 = m_2 = \cdots m_n = 2, \qquad r = 1 \tag{15-66}$$

so that from (15-52), for each process

$$T_i \sim \text{Gamma}\,(2, 1/n). \tag{15-67}$$

Since $\lambda_i = \mu / n = 1/n$, and from (15-57) and (15-66) the stopping time in this case satisfies

$$T = \min\,(T_1, T_2, \cdots, T_n). \tag{15-68}$$

Thus the distribution function for the "birthday pairing" stopping time turns out to be

$$
\begin{aligned}
F_T(t) &= P\{T \le t\} = 1 - P\{T > t\} \\
&= 1 - P\{\min\ (T_1, T_2, \cdots, T_n) > t\} \\
&= 1 - [P\{T_i > t\}]^n = 1 - [1 - F_{T_i}(t)]^n \\
&= 1 - (1 + \tfrac{t}{n})^n e^{-t}
\end{aligned}
\qquad (15\text{-}69)
$$

where we have made use of (15-59) with $m_i = 2$ and $\lambda_i = 1/n.$

As before let N represent the number of random arrivals up to the stopping time T. Notice that in this case N represents the number of people required in a crowd for at least two people to have the same birth date. Since T and N are related as in (15-60), using (15-62) we get

$$
\begin{aligned}
E\{N\} &= E\{T\} = \int_0^\infty P(T > t)dt \\
&= \int_0^\infty \{1 - F_T(t)\}dt = \int_0^\infty (1 + \tfrac{t}{n})^n e^{-t} dt.
\end{aligned}
\qquad (15\text{-}70)
$$

To obtain an approximate value for the above integral, we can expand $\ln(1 + \frac{t}{n})^n = n \ln(1 + \frac{t}{n})$ in Taylor series. This gives

$$\ln\left(1+\frac{t}{n}\right)=\frac{t}{n}-\frac{t^2}{2n^2}+\frac{t^3}{3n^3}$$

and hence

$$\left(1+\frac{t}{n}\right)^n=e^{(t-\frac{t^2}{2n}+\frac{t^3}{3n^2})}$$

so that

$$\left(1+\frac{t}{n}\right)^n e^{-t}=e^{-\frac{t^2}{2n}}\,e^{\frac{t^3}{3n^2}}\approx e^{-\frac{t^2}{2n}}\left(1+\frac{t^3}{3n^2}\right) \qquad (15\text{-}71)$$

and substituting (15-71) into (15-70) we get the mean number of people in a crowd for a two-person birthday coincidence to be

$$\begin{aligned} E\{N\} &\approx \int_0^\infty e^{-t^2/2n}\,dt+\frac{1}{3n^2}\int_0^\infty t^3 e^{-t^2/2n}\,dt \\ &=\frac{1}{2}\sqrt{2\pi\, n}+\frac{2n^2}{3n^2}\int_0^\infty x e^{-x}\,dx=\sqrt{\frac{\pi\, n}{2}}+\frac{2}{3} \\ &=24.612. \end{aligned} \qquad (15\text{-}72)$$

On comparing (15-72) with the mean value obtained in Lecture 6 (Eq. (6-60)) using entirely different arguments $(E\{X\}\approx 24.44)$, we observe that the two results are essentially equal. Notice that the probability that there will be a coincidence among 24-25 people is about 0.55. To compute the variance of T and N we can make use of the expression (see (15-64))

$$E\{N(N+1)\} = E\{T^2\} = \int_0^\infty 2tP(T>t)dt$$

$$= \int_0^\infty 2t\left(1+\frac{t}{n}\right)^n e^{-t}dt \approx 2\int_0^\infty t\; e^{-t^2/2n}dt + \frac{2}{3n^2}\int_0^\infty t^4\; e^{-t^2/2n}dt$$

$$= 2n\int_0^\infty e^{-x}dx + \frac{2}{3n^2}\sqrt{\pi}\;\frac{3}{8}(2n)^2\sqrt{2n} = 2n + \sqrt{2\pi\, n} \qquad (15\text{-}73)$$

which gives (use (15-65))

$$\sigma_T \approx 13.12, \qquad \sigma_N \approx 12.146. \qquad (15\text{-}74)$$

The high value for the standard deviations indicate that in reality the crowd size could vary considerably around the mean value.

Unlike Example 2-20 in Text, the method developed here can be used to derive the distribution and average value for a variety of "birthday coincidence" problems.

Three person birthday-coincidence:

For example, if we are interested in the average crowd size where three people have the same birthday, then arguing as above, we obtain

$$m_1 = m_2 = \cdots = m_n = 3, \qquad r = 1, \tag{15-75}$$

so that

$$T_i \sim \text{Gamma}\,(3, 1/n) \tag{15-76}$$

and T is as in (15-68), which gives

$$F_T(t) = 1 - [1 - F_{T_i}(t)]^n = 1 - \left(1 + \frac{t}{n} + \frac{t^2}{2n^2}\right)^n e^{-t}, \qquad t \geq 0 \tag{15-77}$$

(use (15 - 59) with $m_i = 3,\ \lambda_i = 1/n$) to be the distribution of the stopping time in this case. As before, the average crowd size for three-person birthday coincidence equals

$$E\{N\} = E\{T\} = \int_0^\infty P(T > t)dt = \int_0^\infty \left(1 + \frac{t}{n} + \frac{t^2}{2n^2}\right)^n e^{-t} dt.$$

By Taylor series expansion

$$\ln\left(1+\frac{t}{n}+\frac{t^2}{2n^2}\right)=\left(\frac{t}{n}+\frac{t^2}{2n^2}\right)-\frac{1}{2}\left(\frac{t}{n}+\frac{t^2}{2n^2}\right)^2+\frac{1}{3}\left(\frac{t}{n}+\frac{t^2}{2n^2}\right)^3$$

$$\approx \frac{t}{n}-\frac{t^3}{6n^3}$$

so that

$$E\{N\}=\int_0^\infty e^{-t^3/6n^2}\,dt=\frac{6^{1/3}n^{2/3}}{3}\int_0^\infty x^{\left(\frac{1}{3}-1\right)}e^{-x}\,dx$$

$$=6^{1/3}\,\Gamma(4/3)\;n^{2/3}\approx 82.85. \qquad (15\text{-}78)$$

Thus for a three people birthday coincidence the average crowd size should be around 82 (which corresponds to 0.44 probability).

Notice that other generalizations such as "two distinct birthdays to have a pair of coincidence in each case" (m_i =2, r = 2) can be easily worked in the same manner.

We conclude this discussion with the other extreme case, where the crowd size needs to be determined so that "all days in the year are birthdays" among the persons in a crowd.

All days are birthdays:

Once again from the above analysis in this case we have

$$m_1 = m_2 = \cdots = m_n = 1, \qquad r = n = 365 \tag{15-79}$$

so that the stopping time statistics T satisfies

$$T = \max(T_1, T_2, \cdots, T_n), \tag{15-80}$$

where T_i are independent exponential random variables with common parameter $\lambda = \frac{1}{n}$. This situation is similar to the *coupon collecting problem* discussed in (15-32)-(15-34) and from (15-35), the distribution function of T in (15-80) is given by

$$F_T(t) = (1 - e^{-t/n})^n, \qquad t \geq 0 \tag{15-81}$$

and the mean value for T and N are given by (see (15-37)-(15-41))

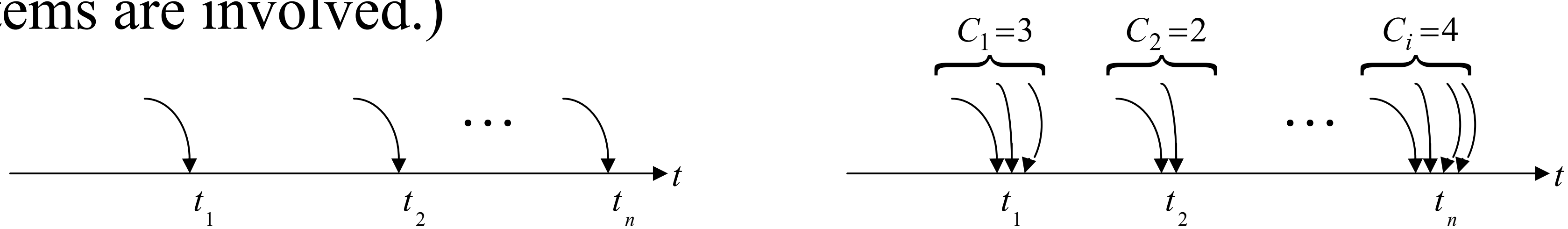
$$E\{N\} = E\{T\} \approx n(\ln n + \gamma) = 2{,}364.14. \tag{15-82}$$

Thus for "everyday to be a birthday" for someone in a crowd, the average crowd size should be 2,364, in which case there is 0.57 probability that the event actually happens.

For a more detailed analysis of this problem using Markov chains, refer to Examples 15-12 and 15-18, in chapter 15, Text. From there (see Eq. (15-80), Text) to be quite certain (with 0.98 probability) that all 365 days are birthdays, the crowd size should be around 3,500.

Bulk Arrivals and Compound Poisson Processes

In an ordinary Poisson process $X(t)$, only one event occurs at any arrival instant (Fig 15.9a). Instead suppose a random number of events C_i occur simultaneously as a cluster at every arrival instant of a Poisson process (Fig 15.9b). If $X(t)$ represents the total number of all occurrences in the interval $(0, t)$, then $X(t)$ represents a **compound Poisson process**, or a **bulk arrival process**. Inventory orders, arrivals at an airport queue, tickets purchased for a show, etc. follow this process (when things happen, they happen in a bulk, or a bunch of items are involved.)



(a) Poisson Process (b) Compound Poisson Process

Fig. 15.9

Let

$$p_k = P\{C_i = k\}, \quad k = 0, 1, 2, \cdots \qquad (15\text{-}83)$$

represent the common probability mass function for the occurrence in any cluster C_i. Then the compound process $X(t)$ satisfies

$$X(t) = \sum_{i=1}^{N(t)} C_i, \tag{15-84}$$

where $N(t)$ represents an ordinary Poisson process with parameter λ. Let

$$P(z) = E\{z^{C_i}\} = \sum_{k=0}^{\infty} p_k z^k \tag{15-85}$$

represent the moment generating function associated with the cluster statistics in (15-83). Then the moment generating function of the compound Poisson process $X(t)$ in (15-84) is given by

$$\begin{aligned}
\phi_X(z) &= \sum_{n=0}^{\infty} z^n P\{X(t)=n\} = E\{z^{X(t)}\} \\
&= E\{E[z^{X(t)} \mid N(t)=k]\} = E[E\{z^{\sum_{i=1}^{k} C_i} \mid N(t)=k\}] \\
&= \sum_{k=0}^{\infty} (E\{z^{C_i}\})^k P\{N(t)=k\} \\
&= \sum_{k=0}^{\infty} P^k(z) e^{-\lambda t} \frac{(\lambda t)^k}{k!} = e^{-\lambda t(1-P(z))}
\end{aligned} \tag{15-86}$$

If we let

$$P^k(z) \triangleq \left(\sum_{n=0}^{\infty} p_n z^k \right)^k = \sum_{n=0}^{\infty} p_n^{(k)} z^n \qquad (15\text{-}87)$$

where $\{p_n^{(k)}\}$ represents the k fold convolution of the sequence $\{p_n\}$ with itself, we obtain

$$P\{X(t)=n\} = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} p_n^{(k)} \qquad (15\text{-}88)$$

that follows by substituting (15-87) into (15-86). Eq. (15-88) represents the probability that there are n arrivals in the interval $(0, t)$ for a compound Poisson process $X(t)$.

Substituting (15-85) into (15-86) we can rewrite $\phi_X(z)$ also as

$$\phi_X(z) = e^{-\lambda_1 t(1-z)} e^{-\lambda_2 t(1-z^2)} \cdots e^{-\lambda_k t(1-z^k)} \cdots \qquad (15\text{-}89)$$

where $\lambda_k = p_k \lambda,$ which shows that the compound Poisson process can be expressed as the sum of integer-secaled independent Poisson processes $m_1(t), m_2(t), \cdots$. Thus

$$X(t) = \sum_{k=1}^{\infty} k\, m_k(t). \tag{15-90}$$

More generally, every linear combination of independent Poisson processes represents a compound Poisson process. (see Eqs. (10-120)-(10-124), Text).
Here is an interesting problem involving compound Poisson processes and coupon collecting: Suppose a cereal manufacturer inserts *either* one *or* two coupons randomly – from a set consisting of n types of coupons – into every cereal box. How many boxes should one buy on the average to collect at least one coupon of each type? We leave it to the reader to work out the details.

16. Mean Square Estimation

Given some information that is related to an unknown quantity of interest, the problem is to obtain a good estimate for the unknown in terms of the observed data.

Suppose $X_1, X_2, \cdots, X_n$ represent a sequence of random variables about whom one set of observations are available, and Y represents an unknown random variable. The problem is to obtain a good estimate for Y in terms of the observations $X_1, X_2, \cdots, X_n$. Let

$$\hat{Y} = \varphi(X_1, X_2, \cdots, X_n) = \varphi(\underline{X}) \tag{16-1}$$

represent such an estimate for Y.

Note that $\varphi(\cdot)$ can be a linear or a nonlinear function of the observation $X_1, X_2, \cdots, X_n$. Clearly

$$\varepsilon(\underline{X}) = Y - \hat{Y} = Y - \varphi(\underline{X}) \tag{16-2}$$

represents the error in the above estimate, and $|\varepsilon|^2$ the square of the error. Since ε is a random variable, $E\{|\varepsilon|^2\}$ represents the mean square error. One strategy to obtain a good estimator would be to minimize the mean square error by varying over all possible forms of $\varphi(\cdot)$, and this procedure gives rise to the Minimization of the Mean Square Error (MMSE) criterion for estimation. Thus under MMSE criterion,the estimator $\varphi(\cdot)$ is chosen such that the mean square error $E\{|\varepsilon|^2\}$ is at its minimum.

Next we show that the conditional mean of Y given $\underline{X}$ is the best estimator in the above sense.

Theorem1: Under MMSE criterion, the best estimator for the unknown Y in terms of $X_1, X_2, \cdots, X_n$ is given by the conditional mean of Y gives $\underline{X}$. Thus

$$\hat{Y} = \varphi(\underline{X}) = E\{Y \mid \underline{X}\}. \tag{16-3}$$

Proof : Let $\hat{Y} = \varphi(\underline{X})$ represent an estimate of Y in terms of $\underline{X} = (X_1, X_2, \cdots, X_n)$.Then the error $\varepsilon = Y - \hat{Y}$, and the mean square error is given by

$$\sigma_\varepsilon^2 = E\{|\varepsilon|^2\} = E\{|Y - \hat{Y}|^2\} = E\{|Y - \varphi(\underline{X})|^2\} \tag{16-4}$$

Since

$$E[z] = E_X[E_z\{z \mid \underline{X}\}] \tag{16-5}$$

we can rewrite (16-4) as

$$\sigma_\varepsilon^2 = E\{\underbrace{|Y - \varphi(\underline{X})|^2}_{z}\} = E_{\underline{X}}[E_Y\{\underbrace{|Y - \varphi(\underline{X})|^2}_{z} | \underline{X}\}]$$

where the inner expectation is with respect to Y, and the outer one is with respect to $\underline{X}$.

Thus

$$\sigma_\varepsilon^2 = E[E\{|Y - \varphi(\underline{X})|^2 \,|\underline{X}\}]$$

$$= \int_{-\infty}^{+\infty} E\{|Y - \varphi(\underline{X})|^2 \,|\underline{X}\} f_X(\underline{X})dx. \tag{16-6}$$

To obtain the best estimator φ, we need to minimize σ_ε^2 in (16-6) with respect to φ. In (16-6), since $f_{\underline{X}}(\underline{X}) \geq 0$, $E\{|Y - \varphi(\underline{X})|^2 \,|\underline{X}\} \geq 0$, and the variable φ appears only in the integrand term, minimization of the mean square error σ_ε^2 in (16-6) with respect to φ is equivalent to minimization of $E\{|Y - \varphi(\underline{X})|^2 \,|\underline{X}\}$ with respect to φ.

Since $\underline{X}$ is fixed at some value, $\varphi(\underline{X})$ is no longer random, and hence minimization of $E\{|Y-\varphi(\underline{X})|^2 \mid \underline{X}\}$ is equivalent to

$$\frac{\partial}{\partial\varphi} E\{|Y-\varphi(\underline{X})|^2 \mid \underline{X}\} = 0. \tag{16-7}$$

This gives

$$E\{|Y-\varphi(\underline{X})| \mid \underline{X}\} = 0$$

or

$$E\{Y \mid \underline{X}\} - E\{\varphi(\underline{X}) \mid \underline{X}\} = 0. \tag{16-8}$$

But

$$E\{\varphi(\underline{X}) \mid \underline{X}\} = \varphi(\underline{X}), \tag{16-9}$$

since when $\underline{X} = \underline{x}, \varphi(\underline{X})$ is a fixed number $\varphi(\underline{x})$. Using (16-9) in (16-8) we get the desired estimator to be

$$\hat{Y} = \varphi(\underline{X}) = E\{Y \mid \underline{X}\} = E\{Y \mid X_1, X_2, \cdots, X_n\}. \tag{16-10}$$

Thus the conditional mean of Y given $X_1, X_2, \cdots, X_n$ represents the best estimator for Y that minimizes the mean square error.

The minimum value of the mean square error is given by

$$\sigma^2_{\min} = E\{|Y - E(Y \mid X)|^2\} = E[\underbrace{E\{|Y - E(Y \mid X)|^2 \mid \underline{X}\}}_{\mathrm{var}(Y \mid \underline{X})}]$$

$$= E\{\mathrm{var}(Y \mid \underline{X})\} \geq 0. \tag{16-11}$$

As an example, suppose $Y = X^3$ is the unknown. Then the best MMSE estimator is given by

$$\hat{Y} = E\{Y \mid X\} = E\{X^3 \mid X\} = X^3. \tag{16-12}$$

Clearly if $Y = X^3$, then indeed $\hat{Y} = X^3$ is the best estimator for Y

in terms of X. Thus the best estimator can be nonlinear.

Next, we will consider a less trivial example.

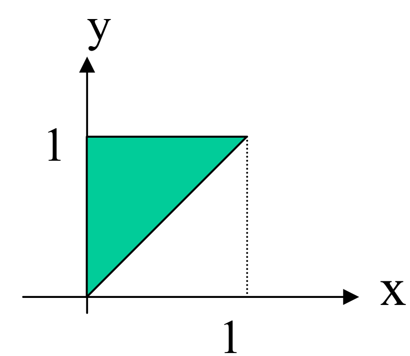
Example : Let

$$f_{X,Y}(x,y)=\begin{cases} kxy, & 0<x<y<1 \\ 0 & \text{otherwise,} \end{cases}$$

where $k>0$ is a suitable normalization constant. To determine the best estimate for Y in terms of X, we need $f_{Y|X}(y|x)$.

$$f_X(x)=\int_x^1 f_{X,Y}(x,y)dy=\int_x^1 kxydy$$

$$=\left.\frac{kxy^2}{2}\right|_x^1=\frac{kx(1-x^2)}{2}, \quad 0<\text{x}<1.$$



Thus

$$f_{Y|X}(y|x)=\frac{f_{X,Y}(x,y)}{f_X(x)}=\frac{kxy}{kx(1-x^2)/2}=\frac{2y}{1-x^2}; \quad 0<x<y<1. \tag{16-13}$$

Hence the best MMSE estimator is given by

$$\hat{Y} = \varphi(X) = E\{Y \mid \underline{X}\} = \int_x^1 y\, f_{Y|X}(y \mid x)dy$$

$$= \int_x^1 y \tfrac{2y}{1-x^2}\, dy = \tfrac{2}{1-x^2}\int_x^1 y^2 dy$$

$$= \frac{2}{3}\frac{y^3}{1-x^2}\bigg|_x^1 = \frac{2}{3}\frac{1-x^3}{1-x^2} = \frac{2}{3}\frac{(1+x+x^2)}{1-x^2}. \tag{16-14}$$

Once again the best estimator is nonlinear. In general the best estimator $E\{Y \mid \underline{X}\}$ is difficult to evaluate, and hence next we will examine the special subclass of best linear estimators.

Best Linear Estimator

In this case the estimator $\hat{Y}$ is a linear function of the observations $X_1, X_2, \cdots, X_n$. Thus

$$\hat{Y}_l = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n = \sum_{i=1}^{n} a_i X_i. \tag{16-15}$$

where $a_1, a_2, \cdots, a_n$ are unknown quantities to be determined. The mean square error is given by $(\varepsilon = Y - \hat{Y}_l)$

$$E\{|\varepsilon|^2\} = E\{|Y - \hat{Y}_l|^2\} = E\{|Y - \sum a_i X_i|^2\} \qquad (16\text{-}16)$$

and under the MMSE criterion $a_1, a_2, \cdots, a_n$ should be chosen so that the mean square error $E\{|\varepsilon|^2\}$ is at its minimum possible value. Let σ_n^2 represent that minimum possible value. Then

$$\sigma_n^2 = \min_{a_1, a_2, \cdots, a_n} E\{|Y - \sum_{i=1}^{n} a_i X_i|^2\}. \qquad (16\text{-}17)$$

To minimize (16-16), we can equate

$$\frac{\partial}{\partial a_k} E\{|\varepsilon|^2\} = 0, \quad k = 1, 2, \cdots, n. \qquad (16\text{-}18)$$

This gives

$$\frac{\partial}{\partial a_k} E\{|\varepsilon|^2\} = E\left\{\frac{\partial |\varepsilon|^2}{\partial a_k}\right\} = 2E\left[\varepsilon \left\{\frac{\partial \varepsilon}{\partial a_k}\right\}^*\right] = 0. \qquad (16\text{-}19)$$

But

$$\frac{\partial \varepsilon}{\partial a_k} = \frac{\partial (Y - \sum_{i=1}^{n} a_i X_i)}{\partial a_k} = \frac{\partial Y}{\partial a_k} - \frac{\partial (\sum_{i=1}^{n} a_i X_i)}{\partial a_k} = -X_k. \tag{16-20}$$

Substituting (16-19) in to (16-18), we get

$$\frac{\partial E\{|\varepsilon|^2\}}{\partial a_k} = -2E\{\varepsilon\ X_k^*\} = 0,$$

or the best linear estimator must satisfy

$$E\{\varepsilon\ X_k^*\} = 0, \quad k = 1, 2, \cdots, n. \tag{16-21}$$

Notice that in (16-21), ε represents the estimation error $(Y - \sum_{i=1}^{n} a_i X_i)$, and X_k, $k = 1 \to n$ represents the data. Thus from (16-21), the error ε is orthogonal to the data X_k, $k = 1 \to n$ for the best linear estimator. This is the **orthogonality principle**.

In other words, in the linear estimator (16-15), the unknown constants $a_1, a_2, \cdots, a_n$ must be selected such that the error

$\varepsilon = Y - \sum_{i=1}^{n} a_i X_i$ is orthogonal to every data $X_1, X_2, \cdots, X_n$ for the best linear estimator that minimizes the mean square error.

Interestingly a general form of the orthogonality principle holds good in the case of nonlinear estimators also.

Nonlinear Orthogonality Rule: Let $h(\underline{X})$ represent *any* functional form of the data and $E\{Y \mid \underline{X}\}$ the best estimator for Y given $\underline{X}$. With $e = Y - E\{Y \mid \underline{X}\}$ we shall show that

$$E\{eh(\underline{X})\} = 0, \tag{16-22}$$

implying that

$$e = Y - E\{Y \mid \underline{X}\} \quad \perp \quad h(\underline{X}).$$

This follows since

$$\begin{aligned} E\{eh(\underline{X})\} &= E\{(Y - E[Y \mid \underline{X}])h(\underline{X})\} \\ &= E\{Yh(\underline{X})\} - E\{E[Y \mid \underline{X}]h(\underline{X})\} \\ &= E\{Yh(\underline{X})\} - E\{E[Yh(\underline{X}) \mid \underline{X}]\} \\ &= E\{Yh(\underline{X})\} - E\{Yh(\underline{X})\} = 0. \end{aligned}$$

Thus in the nonlinear version of the orthogonality rule the error is orthogonal to *any* functional form of the data.

The orthogonality principle in (16-20) can be used to obtain the unknowns $a_1, a_2, \cdots, a_n$ in the linear case.

For example suppose $n = 2$, and we need to estimate Y in terms of X_1 and X_2 linearly. Thus

$$\hat{Y}_l = a_1 X_1 + a_2 X_2$$

From (16-20), the orthogonality rule gives

$$E\{\varepsilon\ \mathrm{X}_1^*\} = E\{(Y - a_1 X_1 - a_2 X_2) X_1^*\} = 0$$

$$E\{\varepsilon\ \mathrm{X}_2^*\} = E\{(Y - a_1 X_1 - a_2 X_2) X_2^*\} = 0$$

Thus

$$E\{| X_1 |^2\} a_1 + E\{X_2 X_1^*\} a_2 = E\{Y X_1^*\}$$

$$E\{X_1 X_2^*\} a_1 + E\{| X_2 |^2\} a_2 = E\{Y X_2^*\}$$

or

$$\begin{pmatrix} E\{|X_1|^2\} & E\{X_2X_1^*\} \\ E\{X_1X_2^*\} & E\{|X_2|^2\} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} E\{YX_1^*\} \\ E\{YX_2^*\} \end{pmatrix} \tag{16-23}$$

(16-23) can be solved to obtain a_1 and a_2 in terms of the cross-correlations.

The minimum value of the mean square error σ_n^2 in (16-17) is given by

$$\sigma_n^2 = \min_{a_1,a_2,\cdots,a_n} E\{|\varepsilon|^2\}$$

$$= \min_{a_1,a_2,\cdots,a_n} E\{\varepsilon\varepsilon^*\} = \min_{a_1,a_2,\cdots,a_n} E\{\varepsilon(Y - \sum_{i=1}^n a_i X_i)^*\}$$

$$= \min_{a_1,a_2,\cdots,a_n} E\{\varepsilon\, Y^*\} - \min_{a_1,a_2,\cdots,a_n} \sum_{i=1}^n a_i E\{\varepsilon\, X_l^*\}. \tag{16-24}$$

But using (16-21), the second term in (16-24) is zero, since the error is orthogonal to the data X_i, where $a_1, a_2, \cdots, a_n$ are chosen to be optimum. Thus the minimum value of the mean square error is given by

$$\sigma_n^2 = E\{\varepsilon\, Y^*\} = E\{(Y - \sum_{i=1}^{n} a_i X_i) Y^*\}$$

$$= E\{|\,Y\,|^2\} - \sum_{i=1}^{n} a_i E\{X_i Y^*\} \qquad (16\text{-}25)$$

where $a_1, a_2, \cdots, a_n$ are the optimum values from (16-21).

Since the linear estimate in (16-15) is only a special case of the general estimator $\varphi(\underline{X})$ in (16-1), the best linear estimator that satisfies (16-20) cannot be superior to the best nonlinear estimator $E\{Y \mid \underline{X}\}$. Often the best linear estimator will be inferior to the best estimator in (16-3).

This raises the following question. Are there situations in which the best estimator in (16-3) also turns out to be linear ? In those situations it is enough to use (16-21) and obtain the best linear estimators, since they also represent the best global estimators. Such is the case if Y and $X_1, X_2, \cdots, X_n$ are distributed as jointly Gaussian

We summarize this in the next theorem and prove that result.

Theorem2: If $X_1, X_2, \cdots, X_n$ and Y are jointly Gaussian zero mean random variables, then the best estimate for Y in terms of $X_1, X_2, \cdots, X_n$ is always linear.

Proof : Let

$$\hat{Y} = \varphi(X_1, X_2, \cdots, X_n) = E\{Y \mid \underline{X}\} \tag{16-26}$$

represent the best (possibly nonlinear) estimate of Y, and

$$\hat{Y}_l = \sum_{i=1}^{n} a_i X_i \tag{16-27}$$

the best linear estimate of Y. Then from (16-21)

$$\varepsilon \triangleq Y - Y_l = Y - \sum_{i=1}^{n} a_i X_i \tag{16-28}$$

is orthogonal to the data $X_k, \quad k = 1 \to n.$ Thus

$$E\{\varepsilon \, \mathrm{X}_{\mathrm{k}}^{*}\} = 0, \quad k = 1 \to n. \tag{16-29}$$

Also from (16-28),

$$E\{\varepsilon\} = E\{Y\} - \sum_{i=1}^{n} a_i E\{X_i\} = 0. \tag{16-30}$$

Using (16-29)-(16-30), we get

$$E\{\varepsilon\ X_k^*\} = E\{\varepsilon\}E\{X_k^*\} = 0, \quad k = 1 \to n. \tag{16-31}$$

From (16-31), we obtain that ε and X_k are zero mean uncorrelated random variables for $k = 1 \to n.$ But ε itself represents a Gaussian random variable, since from (16-28) it represents a linear combination of a set of jointly Gaussian random variables. Thus ε and $\underline{X}$ are jointly Gaussian and uncorrelated random variables. As a result, ε and $\underline{X}$ are independent random variables. Thus from their independence

$$E\{\varepsilon \mid \underline{X}\} = E\{\varepsilon\}. \tag{16-32}$$

But from (16-30), $E\{\varepsilon\} = 0,$ and hence from (16-32)

$$E\{\varepsilon \mid \underline{X}\} = 0. \tag{16-33}$$

Substituting (16-28) into (16-33), we get

$$E\{\varepsilon \mid \underline{X}\} = E\{Y - \sum_{i=1}^{n} a_i X_i \mid \underline{X}\} = 0$$

or

$$E\{Y \mid \underline{X}\} = E\{\sum_{i=1}^{n} a_i X_i \mid \underline{X}\} = \sum_{i=1}^{n} a_i X_i = Y_l. \tag{16-34}$$

From (16-26), $E\{Y \mid \underline{X}\} = \varphi(\underline{x})$ represents the best possible estimator, and from (16-28), $\sum_{i=1}^{n} a_i X_i$ represents the best linear estimator. Thus the best linear estimator is also the best possible overall estimator in the Gaussian case.

Next we turn our attention to prediction problems using linear estimators.

Linear Prediction

Suppose $X_1, X_2, \cdots, X_n$ are known and X_{n+1} is unknown. Thus $Y = X_{n+1}$, and this represents a one-step prediction problem. If the unknown is X_{n+k}, then it represents a *k*-step ahead prediction problem. Returning back to the one-step predictor, let $\hat{X}_{n+1}$ represent the best linear predictor. Then

$$\hat{X}_{n+1} \triangleq -\sum_{i=1}^{n} a_i X_i, \tag{16-35}$$

where the error

$$\begin{aligned}
\varepsilon_n &= X_{n+1} - \hat{X}_{n+1} = X_{n+1} + \sum_{i=1}^{n} a_i X_i \\
&= a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + X_{n+1} \\
&= \sum_{i=1}^{n+1} a_i X_i, \qquad\qquad a_{n+1} = 1,
\end{aligned} \tag{16-36}$$

is orthogonal to the data, i.e.,

$$E\{\varepsilon_n X_k^*\} = 0, \qquad k = 1 \to n. \tag{16-37}$$

Using (16-36) in (16-37), we get

$$E\{\varepsilon_n X_k^*\} = \sum_{i=1}^{n+1} a_i E\{X_i X_k^*\} = 0, \qquad k = 1 \to n. \tag{16-38}$$

Suppose X_i represents the sample of a wide sense stationary stochastic process $X(t)$ so that

$$E\{X_i X_k^*\} = R(i-k) = r_{i-k} = r_{k-i}^* \tag{16-39}$$

Thus (16-38) becomes

$$E\{\varepsilon_n X_k^*\} = \sum_{i=1}^{n+1} a_i r_{i-k} = 0, \quad a_{n+1} = 1, \quad k = 1 \to n. \tag{16-40}$$

Expanding (16-40) for $k = 1, 2, \cdots, n,$ we get the following set of linear equations.

$$\begin{aligned}
&a_1 r_0 + a_2 r_1 + a_3 r_2 + \cdots + a_n r_{n-1} + r_n = 0 \leftarrow k = 1 \\
&a_1 r_1^* + a_2 r_0 + a_3 r_1 + \cdots + a_n r_{n-2} + r_{n-1} = 0 \leftarrow k = 2 \\
&\vdots \\
&a_1 r_{n-1}^* + a_2 r_{n-2}^* + a_3 r_{n-3}^* + \cdots + a_n r_0 + r_1 = 0 \leftarrow k = n.
\end{aligned} \tag{16-41}$$

Similarly using (16-25), the minimum mean square error is given by

$$
\begin{aligned}
\sigma_n^2 &= E\{|\varepsilon|^2\} = E\{\varepsilon_n Y^*\} = E\{\varepsilon_n X_{n+1}^*\} \\
&= E\{(\sum_{i=1}^{n+1} a_i X_i) X_{n+1}^*\} = \sum_{i=1}^{n+1} a_i r_{n+1-i}^* \\
&= a_1 r_n^* + a_2 r_{n-1}^* + a_3 r_{n-2}^* + \cdots + a_n r_1 + r_0.
\end{aligned}
\tag{16-42}
$$

The n equations in (16-41) together with (16-42) can be represented as

$$
\begin{pmatrix}
r_0 & r_1 & r_2 & \cdots & r_n \\
r_1^* & r_0 & r_1 & \cdots & r_{n-1} \\
r_2^* & r_1^* & r_0 & \cdots & r_{n-2} \\
 & & \vdots & & \\
r_{n-1}^* & r_{n-2}^* & \cdots & r_0 & r_1 \\
r_n^* & r_{n-1}^* & \cdots & r_1^* & r_0
\end{pmatrix}
\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \\ 1 \end{pmatrix}
=
\begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \sigma_n^2 \end{pmatrix}.
\tag{16-43}
$$

Let

$$T_n = \begin{pmatrix} r_0 & r_1 & r_2 & \cdots & r_n \\ r_1^* & r_0 & r_1 & \cdots & r_{n-1} \\ & & \vdots & & \\ r_n^* & r_{n-1}^* & \cdots & r_1^* & r_0 \end{pmatrix}. \tag{16-44}$$

Notice that T_n is Hermitian Toeplitz and positive definite. Using (16-44), the unknowns in (16-43) can be represented as

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \\ 1 \end{pmatrix} = T_n^{-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \sigma_n^2 \end{pmatrix} = \sigma_n^2 \begin{pmatrix} \text{Last} \\ \text{column} \\ \text{of} \\ T_n^{-1} \end{pmatrix} \tag{16-45}$$

Let

$$T_n^{-1} = \begin{pmatrix} T_n^{11} & T_n^{12} & \cdots & T_n^{1,n+1} \\ T_n^{21} & T_n^{22} & \cdots & T_n^{2,n+1} \\ & & \vdots & \\ T_n^{n+1,1} & T_n^{n+1,2} & \cdots & T_n^{n+1,n+1} \end{pmatrix}. \tag{16-46}$$

Then from (16-45),

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ 1 \end{pmatrix} = \sigma_n^2 \begin{pmatrix} T_n^{1,n+1} \\ T_n^{2,n+1} \\ \vdots \\ T_n^{n+1,n+1} \end{pmatrix}. \tag{16-47}$$

Thus

$$\sigma_n^2 = \frac{1}{T_n^{n+1,n+1}} > 0, \tag{16-48}$$

and

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \frac{1}{T_n^{n+1,n+1}} \begin{pmatrix} T_n^{1,n+1} \\ T_n^{2,n+1} \\ \vdots \\ T_n^{n+1,n+1} \end{pmatrix}. \tag{16-49}$$

Eq. (16-49) represents the best linear predictor coefficients, and they can be evaluated from the last column of T_n in (16-45). Using these, The best one-step ahead predictor in (16-35) taken the form

$$\hat{X}_{n+1} = -\left(\frac{1}{T_n^{n+1,n+1}}\right)\sum_{i=1}^{n} (T_n^{i,n+1}) X_i. \tag{16-50}$$

and from (16-48), the minimum mean square error is given by the $(n+1,\, n+1)$ entry of T_n^{-1}.

From (16-36), since the one-step linear prediction error

$$\varepsilon_n = X_{n+1} + a_n X_n + a_{n-1} X_{n-1} + \cdots + a_1 X_1, \tag{16-51}$$

we can represent (16-51) formally as follows

$$X_{n+1} \rightarrow \boxed{1 + a_n z^{-1} + a_{n-1} z^{-2} + \cdots + a_1 z^{-n}} \rightarrow \varepsilon_n$$

Thus, let

$$A_n(z) = 1 + a_n z^{-1} + a_{n-1} z^{-2} + \cdots + a_1 z^{-n}, \tag{16-52}$$

them from the above figure, we also have the representation

$$\varepsilon_n \rightarrow \boxed{\frac{1}{A_n(z)}} \rightarrow X_{n+1}.$$

The filter

$$H(z) = \frac{1}{A_n(z)} = \frac{1}{1 + a_n z^{-1} + a_{n-1} z^{-2} + \cdots + a_1 z^{-n}} \tag{16-53}$$

represents an *AR*(n) filter, and this shows that linear prediction leads to an auto regressive (*AR*) model.

The polynomial $A_n(z)$ in (16-52)-(16-53) can be simplified using (16-43)-(16-44). To see this, we rewrite $A_n(z)$ as

$$A_n(z) = a_1 z^{-n} + a_2 z^{-(n-1)} + \cdots + a_{n-1} z^{-2} + a_n z^{-1} + 1$$

$$= [z^{-n}, z^{-(n-1)}, \cdots, z^{-1}, 1] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ 1 \end{bmatrix} = [z^{-n}, z^{-(n-1)}, \cdots, z^{-1}, 1]\, T_n^{-1} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \sigma_n^2 \end{bmatrix} \tag{16-54}$$

To simplify (16-54), we can make use of the following matrix identity

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -AB \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & D - CA^{-1}B \end{bmatrix}. \tag{16-55}$$

Taking determinants, we get

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| \, |D - CA^{-1}B|. \tag{16-56}$$

In particular if $D \equiv 0$, we get

$$|CA^{-1}B| = \frac{(-1)^n}{|A|} \begin{vmatrix} A & B \\ C & 0 \end{vmatrix}. \tag{16-57}$$

Using (16-57) in (16-54), with

$$C = [z^{-n}, z^{-(n-1)}, \cdots, z^{-1}, 1], \;\; A = T_n, \;\; B = \begin{bmatrix} 0 \\ \vdots \\ \sigma_n^2 \end{bmatrix}$$

we get

$$A_n(z) = \frac{(-1)^n}{|T_n|}\left|\begin{array}{c:c} T_n & \begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \\ \sigma_n^2 \end{array} \\ \hdashline z^{-n},\cdots,z^{-1},1 & 0 \end{array}\right| = \frac{\sigma_n^2}{|T_n|}\begin{vmatrix} r_0 & r_1 & r_2 & \cdots & r_n \\ r_1^* & r_0 & r_1 & \cdots & r_{n-1} \\ & & \vdots & & \\ r_{n-1}^* & r_{n-2}^* & \cdots & r_0 & r_1 \\ z^{-n} & z^{-(n-1)} & \cdots & z^{-1} & 1 \end{vmatrix}. \quad (16\text{-}58)$$

Referring back to (16-43), using Cramer's rule to solve for $a_{n+1}(=1)$, we get

$$a_{n+1} = \frac{\sigma_n^2 \begin{vmatrix} r_0 & \cdots & r_{n-1} \\ & \vdots & \\ r_{n-1} & \cdots & r_0 \end{vmatrix}}{|T_n|} = \sigma_n^2 \frac{|T_{n-1}|}{|T_n|} = 1$$

or

$$\sigma_n^2 = \frac{|T_n|}{|T_{n-1}|} > 0. \qquad (16\text{-}59)$$

Thus the polynomial (16-58) reduces to

$$A_n(z) = \frac{1}{|T_{n-1}|} \begin{vmatrix} r_0 & r_1 & r_2 & \cdots & r_n \\ r_1^* & r_0 & r_1 & \cdots & r_{n-1} \\ & & \vdots & & \\ r_{n-1}^* & r_{n-2}^* & \cdots & r_0 & r_1 \\ z^{-n} & z^{-(n-1)} & \cdots & z^{-1} & 1 \end{vmatrix} \qquad (16\text{-}60)$$

$$= 1 + a_n z^{-1} + a_{n-1} z^{-2} + \cdots + a_1 z^{-n}.$$

The polynomial $A_n(z)$ in (16-53) can be alternatively represented as in (16-60), and $H(z) = \dfrac{1}{A_n(z)} \sim AR(n)$ in fact represents a stable

AR filter of order n, whose input error signal ε_n is white noise of constant spectral height equal to $|T_n|/|T_{n-1}|$ and output is X_{n+1}. It can be shown that $A_n(z)$ has all its zeros in $|z|>1$ provided $|T_n|>0$ thus establishing stability.

Linear prediction Error

From (16-59), the mean square error using n samples is given by

$$\sigma_n^2 = \frac{|T_n|}{|T_{n-1}|} > 0. \qquad (16\text{-}61)$$

Suppose one more sample from the past is available to evaluate X_{n+1} (i.e., $X_n, X_{n-1}, \cdots, X_1, X_0$ are available). Proceeding as above the new coefficients and the mean square error σ_{n+1}^2 can be determined. From (16-59)-(16-61),

$$\sigma_{n+1}^2 = \frac{|T_{n+1}|}{|T_n|}. \qquad (16\text{-}62)$$

Using another matrix identity it is easy to show that

$$|T_{n+1}| = \frac{|T_n|^2}{|T_{n-1}|}(1-|s_{n+1}|^2). \tag{16-63}$$

Since $|T_k| > 0$, we must have $(1-|s_{n+1}|^2) > 0$ or $|s_{n+1}| < 1$ for every n. From (16-63), we have

$$\underbrace{\frac{|T_{n+1}|}{|T_n|}}_{\sigma_{n+1}^2} = \underbrace{\frac{|T_n|}{|T_{n-1}|}}_{\sigma_n^2}(1-|s_{n+1}|^2)$$

or

$$\sigma_{n+1}^2 = \sigma_n^2 (1-|s_{n+1}|^2) < \sigma_n^2, \tag{16-64}$$

since $(1-|s_{n+1}|^2) < 1$. Thus the mean square error decreases as more and more samples are used from the past in the linear predictor. In general from (16-64), the mean square errors for the one-step predictor form a monotonic nonincreasing sequence

$$\sigma_n^2 \geq \sigma_{n+1}^2 \geq \cdots \sigma_k^2 > \cdots \to \sigma_\infty^2 \tag{16-65}$$

whose limiting value $\sigma_\infty^2 \geq 0.$

Clearly, $\sigma_\infty^2 \geq 0$ corresponds to the irreducible error in linear prediction using the entire past samples, and it is related to the power spectrum of the underlying process $X(nT)$ through the relation

$$\sigma_\infty^2 = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{+\pi} \ln S_{XX}(\omega)d\omega\right] \geq 0. \tag{16-66}$$

where $S_{XX}(\omega) \geq 0$ represents the power spectrum of $X(nT)$.
For any finite power process, we have

$$\int_{-\pi}^{+\pi} S_{XX}(\omega)d\omega < \infty,$$

and since $(S_{XX}(\omega) \geq 0), \quad \ln S_{XX}(\omega) \leq S_{XX}(\omega).$ Thus

$$\int_{-\pi}^{+\pi} \ln S_{XX}(\omega)d\omega \leq \int_{-\pi}^{+\pi} S_{XX}(\omega)d\omega < \infty. \tag{16-67}$$

Moreover, if the power spectrum is strictly positive at every Frequency, i.e.,

$$S_{XX}(\omega) > 0, \qquad \text{in } -\pi < \omega < \pi, \tag{16-68}$$

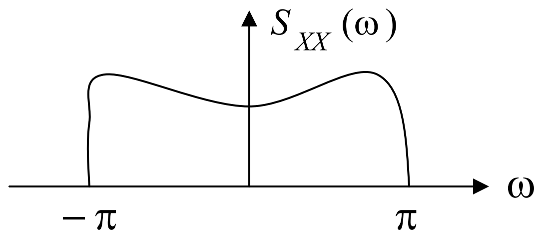
then from (16-66)

$$\int_{-\pi}^{+\pi} \ln S_{XX}(\omega)\, d\omega > -\infty. \tag{16-69}$$

and hence

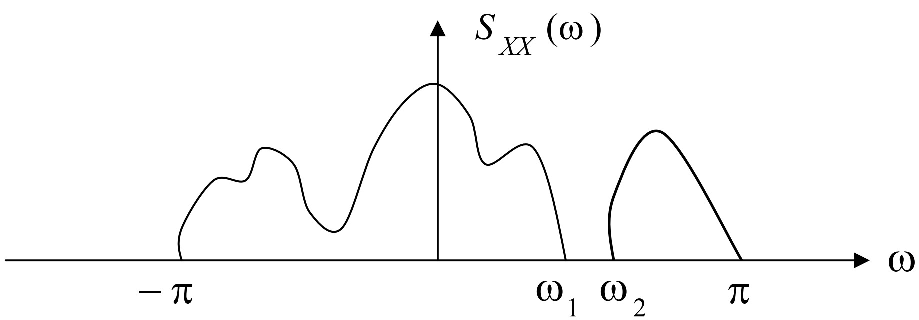
$$\sigma_\infty^2 = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{+\pi} \ln S_{XX}(\omega)\, d\omega\right] > e^{-\infty} = 0 \tag{16-70}$$

i.e., For processes that satisfy the strict positivity condition in (16-68) almost everywhere in the interval $(-\pi,\pi)$, the final minimum mean square error is strictly positive (see (16-70)). i.e., Such processes are not completely predictable even using their entire set of past samples, or they are inherently stochastic, since the next output contains information that is not contained in the past samples. Such processes are known as *regular* stochastic processes, and their power spectrum is strictly positive.



Power Spectrum of a regular stochastic Process

Conversely, if a process has the following power spectrum,

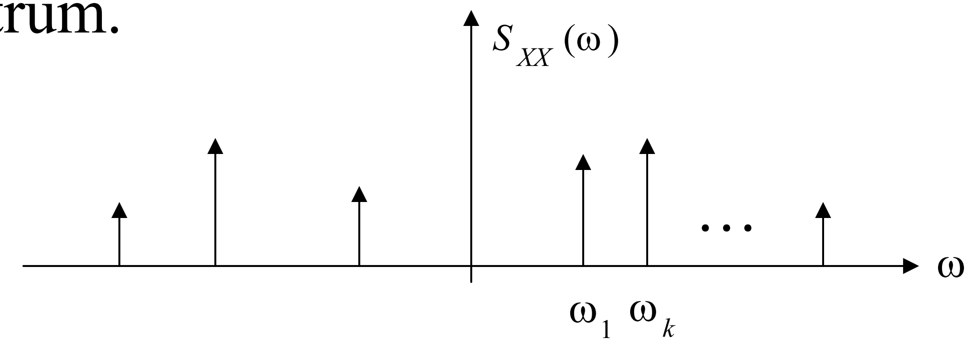


such that $S_{XX}(\omega)=0$ in $\omega_1<\omega<\omega_2$ then from (16-70), $\sigma_\infty^2=0.$

Such processes are completely predictable from their past data samples. In particular

$$X(nT) = \sum_k a_k \cos(\omega_k t + \phi_k) \qquad (16\text{-}71)$$

is completely predictable from its past samples, since $S_{XX}(\omega)$ consists of line spectrum.



$X(nT)$ in (16-71) is a shape deterministic stochastic process.

17. Long Term Trends and Hurst Phenomena

From ancient times the Nile river region has been known for its peculiar long-term behavior: long periods of dryness followed by long periods of yearly floods. It seems historical records that go back as far as 622 AD also seem to support this trend. There were long periods where the high levels tended to stay high and other periods where low levels remained low[1].

An interesting question for hydrologists in this context is how to devise methods to regularize the flow of a river through reservoir so that the outflow is uniform, there is no overflow at any time, and in particular the capacity of the reservoir is ideally as full at time $t+t_0$ as at *t*. Let $\{y_i\}$ denote the annual inflows, and

$$s_n = y_i + y_2 + \cdots + y_n \tag{17-1}$$

---

[1] A reference in the Bible says "*seven years of great abundance are coming throughout the land of Egypt, but seven years of famine will follow them*" (*Genesis*).

their cumulative inflow up to time n so that

$$\bar{y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{s_N}{N} \tag{17-2}$$

represents the overall average over a period N. Note that $\{y_i\}$ may as well represent the internet traffic at some specific local area network and $\bar{y}_N$ the average system load in some suitable time frame.

To study the long term behavior in such systems, define the "extermal" parameters

$$u_N = \max_{1\le n\le N}\{s_n - n\bar{y}_N\}, \tag{17-3}$$

$$v_N = \min_{1\le n\le N}\{s_n - n\bar{y}_N\}, \tag{17-4}$$

as well as the sample variance

$$D_N = \frac{1}{N}\sum_{n=1}^{N}(y_n - \bar{y}_N)^2. \tag{17-5}$$

In this case

$$R_N = u_N - v_N \tag{17-6}$$

defines the *adjusted range statistic* over the period N, and the dimensionless quantity

$$\frac{R_N}{\sqrt{D_N}} = \frac{u_N - v_N}{\sqrt{D_N}} \tag{17-7}$$

that represents the *readjusted range statistic* has been used extensively by hydrologists to investigate a variety of natural phenomena.

To understand the long term behavior of $R_N / \sqrt{D_N}$ where $y_i,\ i = 1, 2, \cdots N$ are independent identically distributed random variables with common mean μ and variance σ^2, note that for large N by the strong law of large numbers

$$s_n \xrightarrow{d} N(n\mu, n\sigma^2), \tag{17-8}$$

$$\bar{y}_N \xrightarrow{d} N(\mu, \sigma^2 / N) \to \mu \tag{17-9}$$

and

$$D_N \xrightarrow{d} \sigma^2 \tag{17-10}$$

with probability 1. Further with $n = Nt$, where $0 < t < 1$, we have

$$\lim_{N\to\infty} \frac{s_n - n\mu}{\sqrt{N}\sigma} = \lim_{N\to\infty} \frac{s_{\lfloor Nt \rfloor} - \lfloor Nt \rfloor \mu}{\sqrt{N}\sigma} \xrightarrow{d} B(t) \tag{17-11}$$

where $B(t)$ is the standard Brownian process with auto-correlation function given by *min* (t_1, t_2).To make further progress note that

$$\begin{aligned} s_n - n\bar{y}_N &= s_n - n\mu - n(\bar{y}_N - \mu) \\ &= (s_n - n\mu) - \frac{n}{N}(s_N - N\mu) \end{aligned} \tag{17-12}$$

so that

$$\frac{s_n - n\bar{y}_N}{\sqrt{N}\sigma} = \frac{s_n - n\mu}{\sqrt{N}\sigma} - \frac{n}{N}\frac{s_N - N\mu}{\sqrt{N}\sigma} \xrightarrow{d} B(t) - tB(1), \quad 0 < \mathrm{t} < 1. \tag{17-13}$$

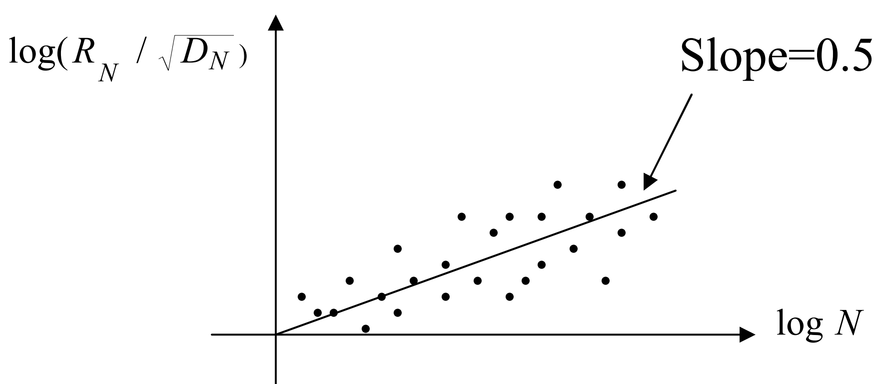
Hence by the functional central limit theorem, using (17-3) and (17-4) we get

$$\frac{u_N - v_N}{\sqrt{N}\sigma} \xrightarrow{d} \max_{0<t<1}\{B(t) - tB(1)\} - \min_{0<t<1}\{B(t) - tB(1)\} \equiv Q, \qquad (17\text{-}14)$$

where Q is a strictly positive random variable with finite variance. Together with (17-10) this gives

$$\frac{R_N}{\sqrt{D_N}} \to \frac{u_N - v_N}{\sigma} \xrightarrow{d} \sqrt{N}Q, \qquad (17\text{-}15)$$

a result due to Feller. Thus in the case of i.i.d. random variables the rescaled range statistic $R_N / \sqrt{D_N}$ is of the order of $O(N^{1/2})$. It follows that the plot of $\log(R_N / \sqrt{D_N})$ versus $\log N$ should be linear with slope $H = 0.5$ for independent and identically distributed observations.



The hydrologist Harold Erwin Hurst (1951) generated tremendous interest when he published results based on water level data that he analyzed for regions of the Nile river which showed that Plots of $\log(R_N / \sqrt{D_N})$ versus log N are linear with slope $H \approx 0.75.$ According to Feller's analysis this must be an anomaly if the flows are i.i.d. with finite second moment.

The basic problem raised by Hurst was to identify circumstances under which one may obtain an exponent $H > 1/2$ for N in (17-15). The first positive result in this context was obtained by Mandelbrot and Van Ness (1968) who obtained $H > 1/2$ under a strongly dependent stationary Gaussian model. The Hurst effect appears for independent and non-stationary flows with finite second moment also. In particular, when an appropriate slow-trend is superimposed on a sequence of i.i.d. random variables the Hurst phenomenon reappears. To see this, we define the Hurst exponent fora data set to be H if

$$\frac{R_N}{\sqrt{D_N}\,N^H} \xrightarrow{\ d\ } Q, \quad N \to \infty, \tag{17-16}$$

where Q is a nonzero real valued random variable.

IID with slow Trend

Let $\{X_n\}$ be a sequence of i.i.d. random variables with common mean μ and variance σ^2, and g_n be an arbitrary real valued function on the set of positive integers setting a deterministic trend, so that

$$y_n = x_n + g_n \tag{17-17}$$

represents the actual observations. Then the partial sum in (17-1) becomes

$$s_n = y_1 + y_2 + \cdots + y_n = x_1 + x_2 + \cdots + x_n + \sum_{i=1}^{n} g_i$$

$$= n\,(\bar{x}_n + \bar{g}_n) \tag{17-18}$$

where $\bar{g}_n = 1/n \sum_{i=1}^{n} g_i$ represents the running mean of the slow trend. From (17-5) and (17-17), we obtain

$$
\begin{aligned}
D_N &= \frac{1}{N}\sum_{n=1}^{N}(y_n - \overline{y}_N)^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}(x_n - \overline{x}_N)^2 + \frac{1}{N}\sum_{n=1}^{N}(g_n - \overline{g}_N)^2 + \frac{2}{N}\sum_{n=1}^{N}(x_n - \overline{x}_N)(g_n - \overline{g}_N) \\
&= \hat{\sigma}_X^2 + \frac{1}{N}\sum_{n=1}^{N}(g_n - \overline{g}_N)^2 + \frac{2}{N}\sum_{n=1}^{N}(x_n - \overline{x}_N)(g_n - \overline{g}_N). \qquad (17\text{-}19)
\end{aligned}
$$

Since $\{x_n\}$ are i.i.d. random variables, from (17-10) we get $\hat{\sigma}_X^2 \xrightarrow{d} \sigma^2$. Further suppose that the deterministic sequence $\{g_n\}$ converges to a finite limit c. Then their Caesaro means $\frac{1}{N}\sum_{n=1}^{N} g_n = \overline{g}_N$ also converges to c. Since

$$
\frac{1}{N}\sum_{n=1}^{N}(g_n - \overline{g}_N)^2 = \frac{1}{N}\sum_{n=1}^{N}(g_n - c)^2 - (\overline{g}_N - c)^2, \qquad (17\text{-}20)
$$

applying the above argument to the sequence $(g_n - c)^2$ and $(\overline{g}_N - c)^2$ we get (17-20) converges to zero. Similarly, since

$$\frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x}_N)(g_n-\bar{g}_N)=\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)(g_n-c)-(\bar{x}_N-\mu)(\bar{g}_N-c), \tag{17-21}$$

by Schwarz inequality, the first term becomes

$$\left|\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)(g_n-c)\right|^2 \le \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2\,\frac{1}{N}\sum_{n=1}^{N}(g_n-c)^2. \tag{17-22}$$

But $\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2 \to \sigma^2$ and the Caesaro means $\frac{1}{N}\sum_{n=1}^{N}(g_n-c)^2 \to 0.$ Hence the first term (17-21) tends to zero as $N\to\infty,$ and so does the second term there. Using these results in (17-19), we get

$$g_n \to c \quad \Rightarrow \quad D_N \xrightarrow{d} \sigma^2. \tag{17-23}$$

To make further progress, observe that

$$\begin{aligned} u_N &= \max\{s_n - n\bar{g}_N\} \\ &= \max\{n(\bar{x}_n-\bar{x}_N)+n(\bar{g}_n-\bar{g}_N)\} \\ &\le \max_{0<n<N}\{n(\bar{x}_n-\bar{x}_N)\}+\max_{0<n<N}\{n(\bar{g}_n-\bar{g}_N)\} \end{aligned} \tag{17-24}$$

and

$$\begin{aligned} v_N &= \min\{s_n - n\bar{g}_N\} \\ &= \min\{n\,(\bar{x}_n - \bar{x}_N) + n\,(\bar{g}_n - \bar{g}_N)\} \\ &\geq \min_{0<n<N}\{n\,(\bar{x}_n - \bar{x}_N)\} + \min_{0<n<N}\{n\,(\bar{g}_n - \bar{g}_N)\}. \end{aligned} \tag{17-25}$$

Consequently, if we let

$$r_N = \max_{0<n<N}\{n\,(\bar{x}_n - \bar{x}_N)\} - \min_{0<n<N}\{n\,(\bar{x}_n - \bar{x}_N)\} \tag{17-26}$$

for the i.i.d. random variables, then from (17-6),(17-24) and (17-25) (17-26), we obtain

$$R_N = u_N - v_N \leq r_N + G_N \tag{17-27}$$

where

$$G_N = \max_{0<n<N}\{n\,(\bar{g}_n - \bar{g}_N)\} - \min_{0<n<N}\{n\,(\bar{g}_n - \bar{g}_N)\} \tag{17-28}$$

From (17-24) – (17-25), we also obtain

$$u_N \ge \min_{0<n<N}\{n(\bar{x}_n - \bar{x}_N)\} + \max_{0<n<N}\{n(\bar{g}_n - \bar{g}_N)\}, \tag{17-29}$$

$$v_N \le \max_{0<n<N}\{n(\bar{x}_n - \bar{x}_N)\} + \min_{0<n<N}\{n(\bar{g}_n - \bar{g}_N)\}, \tag{17-30}$$

[use $\max_i\{(x_i + y_i)\} \ge \max_i\{(\min_i x_i) + y_i\} = \min_i(x_i) + \max_i(y_i)$] and hence

$$R_N \ge G_N - r_N. \tag{17-31}$$

From (17-27) and (17-31) we get the useful estimates

$$|R_N - G_N| \le r_N,$$

and (17-32)

$$|R_N - r_N| \le G_N. \tag{17-33}$$

Since $\{x_n\}$ are i.i.d. random variables, using (17-15) in (17-26) we get

$$\frac{r_N}{\hat{\sigma}_X^2\sqrt{N}} \to \frac{r_N}{\sigma\sqrt{N}} \to Q, \quad \textit{in probability} \tag{17-34}$$

a positive random variable, so that

$$\frac{r_N}{\sigma} \to \sqrt{N}Q \quad \textit{in probability.} \tag{17-35}$$

Consequently for the sequence $\{y_n\}$ in (17-17) using (17-23) in (17-32)-(17-34) we get

$$\frac{|R_N - G_N|}{\sqrt{D_N}\,N^H} \rightarrow \frac{r_N}{\sigma\ N^H} \rightarrow \frac{Q/\sigma}{N^{H-1/2}} \rightarrow 0 \qquad (17\text{-}36)$$

if $H > 1/2$. To summarize, if the slow trend $\{g_n\}$ converges to a finite limit, then for the observed sequence $\{y_n\}$, for every H > 1/2

$$\left|\frac{R_N}{\sqrt{D_N}\,N^H} - \frac{G_N}{\sqrt{D_N}\,N^H}\right| \rightarrow \left|\frac{R_N}{\sqrt{D_N}\,N^H} - \frac{G_N}{\sigma\ N^H}\right| \rightarrow 0 \qquad (17\text{-}37)$$

in probability as $N \rightarrow \infty.$

In particular it follows from (17-16) and (17-36)-(17-37) that the Hurst exponent $H > 1/2$ holds for a sequence $\{y_n\}$ if and only if the slow trend sequence $\{g_n\}$ satisfies

$$\lim_{N\rightarrow\infty} \frac{G_N}{N^H} = c_0 > 0, \qquad H > 1/2. \qquad (17\text{-}38)$$

In that case from (17-37), for that H > 1/2 we obtain

$$\frac{R_N}{\sqrt{D_N}\,N^H} \to c_0/\sigma \qquad \textit{in probability as} \qquad N \to \infty, \tag{17-39}$$

where c_0 is a positive number.

Thus if the slow trend $\{g_n\}$ satisfies (17-38) for some H > 1/2, then from (17-39)

$$\log \frac{R_N}{\sqrt{D_N}} \to H \log N + c, \qquad \text{as} \qquad N \to \infty. \tag{17-40}$$

Example: Consider the observations

$$y_n = x_n + a + bn^{\alpha}, \qquad n \ge 1 \tag{17-41}$$

where x_n are i.i.d. random variables. Here $g_n = a + bn^{\alpha}$, and the sequence converges to a for $\alpha < 0$, so that the above result applies. Let

$$M_n = n\,(\bar{g}_n - \bar{g}_N) = b\left(\sum_{k=1}^{n} k^{\alpha} - \frac{n}{N}\sum_{k=1}^{N} k^{\alpha}\right). \tag{17-42}$$

To obtain its *max* and *min*, notice that

$$M_n - M_{n-1} = b\left(n^{\alpha} - \frac{1}{N}\sum_{k=1}^{N} k^{\alpha} \right) > 0$$

if $n < (\frac{1}{N}\sum_{k=1}^{N} k^{\alpha})^{1/\alpha}$, and negative otherwise. Thus max M_N is achieved at

$$n_0 = \left(\frac{1}{N}\sum_{k=1}^{N} k^{\alpha} \right)^{1/\alpha} \tag{17-43}$$

and the minimum of $M_N = 0$ is attained at N=0. Hence from (17-28) and (17-42)-(17-43)

$$G_N = b\left(\sum_{k=1}^{n_0} k^{\alpha} - \frac{n_0}{N}\sum_{k=1}^{N} k^{\alpha} \right). \tag{17-44}$$

Now using the Reimann sum approximation, we may write

$$\frac{1}{N}\sum_{k=1}^{N} k^{\alpha} \approx \frac{1}{N}\int_0^N x^{\alpha}\,dx = \begin{cases} (1+\alpha)^{-1} N^{\alpha}, & \alpha > -1 \\ \dfrac{\log N}{N}, & \alpha = -1 \\ \dfrac{\sum_{k=1}^{\infty} k^{\alpha}}{N}, & \alpha < -1 \end{cases} \tag{17-45}$$

so that

$$n_0 \approx \begin{cases} (1+\alpha)^{-1/\alpha} N, & \alpha > -1 \\ \dfrac{N}{\log N}, & \alpha = -1 \\ \dfrac{\sum_{k=1}^{\infty} k^{\alpha}}{N^{1/\alpha}}, & \alpha < -1 \end{cases} \tag{17-46}$$

and using (17-45)-(17-46) repeatedly in (17-44) we obtain

$$G_n = bn_0\left(\frac{1}{n_0}\sum_{k=1}^{n_0} k^{\alpha} - \frac{1}{N}\sum_{k=1}^{N} k^{\alpha}\right)$$

$$\approx \begin{cases} \dfrac{bn_0}{1+\alpha}(n_0^{\alpha} - N^{\alpha}) \approx c_1 N^{1+\alpha}, & \alpha > -1 \\ bn_0\left(\dfrac{1}{n_0}\log n_0 - \dfrac{1}{N}\log N\right) \approx c_2 \log N, & \alpha = -1 \\ b\left(1 - \dfrac{n_0}{N}\right)\left(\displaystyle\sum_{k=1}^{\infty} k^{\alpha}\right) \approx c_3, & \alpha < -1 \end{cases} \qquad (17\text{-}47)$$

where c_1, c_2, c_3 are positive constants independent of N. From (17-47), notice that if $-1/2 < \alpha < 0$, then

$$G_n \sim c_1 N^H,$$

where $1/2 < H < 1$ and hence (17-38) is satisfied. In that case

$$\frac{R_N}{\sqrt{D_N}\,N^{(1+\alpha)}} \to c_1 \quad \textit{in probability as} \quad N \to \infty. \tag{17-48}$$

and the Hurst exponent $H = 1+\alpha > 1/2.$

Next consider $\alpha < -1/2.$ In that case from the entries in (17-47) we get $G_N = o(N^{1/2}),$ and diving both sides of (17-33) with $\sqrt{D_N}\,N^{1/2},$

$$\frac{|R_N - r_N|}{\sqrt{D_N}\,N^{1/2}} \sim \frac{o(N^{1/2})}{\sigma\ N^{1/2}} \to 0 \quad \textit{in probability}$$

so that

$$\frac{R_N}{\sqrt{D_N}\,N^{1/2}} \sim \frac{r_N}{\sigma\ N^{1/2}} \to Q \tag{17-49}$$

where the last step follows from (17-15) that is valid for i.i.d. observations. Hence using a limiting argument the Hurst exponent

$H = 1/2$ if $\alpha \le -1/2$. Notice that $\alpha = 0$ gives rise to i.i.d. observations, and the Hurst exponent in that case is 1/2. Finally for $\alpha > 0$, the slow trend sequence $\{g_n\}$ does not converge and (17-36)-(17-40) does not apply. However direct calculation shows that D_N in (17-19) is dominated by the second term which for large N can be approximated as $\frac{1}{N}\int_0^N x^{2\alpha} \approx N^{2\alpha}$ so that

$$\sqrt{D_N} \to c_4 N^{\alpha} \qquad as \qquad N \to \infty \tag{17-50}$$

From (17-32)

$$\frac{|R_N - G_N|}{\sqrt{D_N}\,N} \approx \frac{r_N}{\sqrt{D_N}\,N} \to \frac{\sqrt{N}\sigma\ Q}{c_4 N^{1+\alpha}} \to 0$$

where the last step follows from (17-34)-(17-35). Hence for $\alpha > 0$ from (17-47) and (17-50)

$$\frac{R_N}{\sqrt{D_N}\,N} \approx \frac{G_N}{\sqrt{D_N}\,N} \approx \frac{c_1 N^{1+\alpha}}{c_4 N^{1+\alpha}} \rightarrow \frac{c_1}{c_4} \tag{17-51}$$

as $N \rightarrow \infty$. Hence the Hurst exponent is 1 if $\alpha > 0$. In summary,

$$H(\alpha) = \begin{cases} 1 & \alpha > 0 \\ 1/2 & \alpha = 0 \\ 1+\alpha & 0 > \alpha > \text{-}1/2 \\ 1/2 & \alpha < \text{-}1/2 \end{cases} \tag{17-52}$$

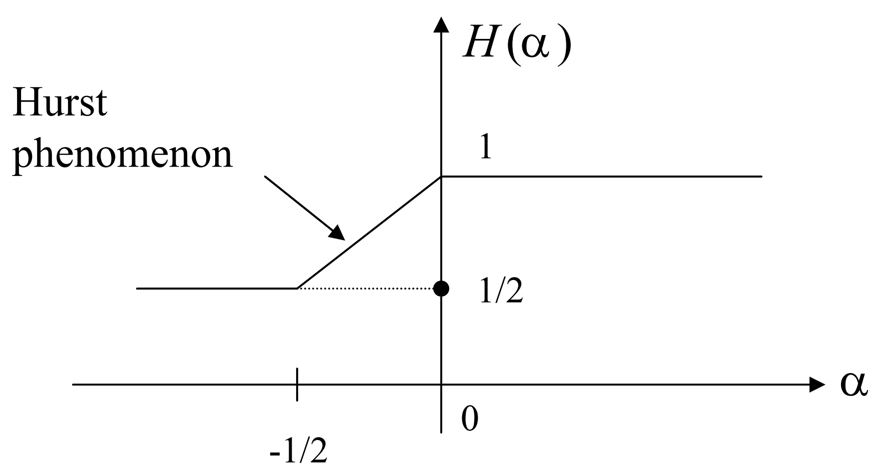


Fig.1 Hurst exponent for a process with superimposed slow trend

18. Power Spectrum

For a deterministic signal x(t), the spectrum is well defined: If $X(\omega)$ represents its Fourier transform, i.e., if

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt, \tag{18-1}$$

then $|X(\omega)|^2$ represents its energy spectrum. This follows from Parseval's theorem since the signal energy is given by

$$\int_{-\infty}^{+\infty} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |X(\omega)|^2 \, d\omega = E. \tag{18-2}$$

Thus $|X(\omega)|^2 \Delta\omega$ represents the signal energy in the band $(\omega, \omega + \Delta\omega)$ (see Fig 18.1).

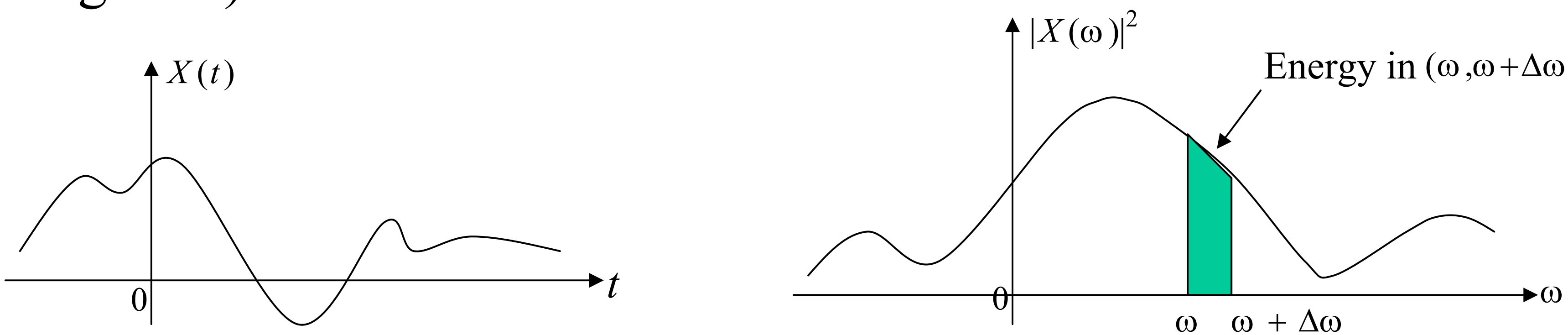


Fig 18.1

However for stochastic processes, a direct application of (18-1) generates a sequence of random variables for every ω. Moreover, for a stochastic process, $E\{|X(t)|^2\}$ represents the ensemble average power (instantaneous energy) at the instant t.

To obtain the spectral distribution of power versus frequency for stochastic processes, it is best to avoid infinite intervals to begin with, and start with a finite interval $(-T,\ T)$ in (18-1). Formally, partial Fourier transform of a process $X(t)$ based on $(-T,\ T)$ is given by

$$X_T(\omega) = \int_{-T}^{T} X(t) e^{-j\omega t} dt \tag{18-3}$$

so that

$$\frac{|X_T(\omega)|^2}{2T} = \frac{1}{2T}\left| \int_{-T}^{T} X(t) e^{-j\omega t} dt \right|^2 \tag{18-4}$$

represents the power distribution associated with that realization based on $(-T,\ T)$. Notice that (18-4) represents a random variable for every ω, and its ensemble average gives, the average power distribution based on $(-T,\ T)$. Thus

$$P_T(\omega) = E\left\{\frac{|X_T(\omega)|^2}{2T}\right\} = \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} E\{X(t_1)X^*(t_2)\}e^{-j\omega(t_1-t_2)}dt_1dt_2$$

$$= \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} R_{XX}(t_1,t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2 \qquad (18\text{-}5)$$

represents the power distribution of $X(t)$ based on $(-T, T)$. For wide sense stationary (w.s.s) processes, it is possible to further simplify (18-5). Thus if $X(t)$ is assumed to be w.s.s, then $R_{XX}(t_1,t_2) = R_{XX}(t_1 - t_2)$ and (18-5) simplifies to

$$P_T(\omega) = \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} R_{XX}(t_1 - t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2.$$

Let $\tau = t_1 - t_2$ and proceeding as in (14-24), we get

$$P_T(\omega) = \frac{1}{2T}\int_{-2T}^{2T} R_{XX}(\tau)e^{-j\omega\tau}(2T - |\tau|)d\tau$$

$$= \int_{-2T}^{2T} R_{XX}(\tau)e^{-j\omega\tau}\left(1 - \tfrac{|\tau|}{2T}\right)d\tau \;\geq\; 0 \qquad (18\text{-}6)$$

to be the power distribution of the w.s.s. process $X(t)$ based on $(-T, T)$. Finally letting $T \to \infty$ in (18-6), we obtain

$$S_{XX}(\omega) = \lim_{T \to \infty} P_T(\omega) = \int_{-\infty}^{+\infty} R_{XX}(\tau) e^{-j\omega\tau} d\tau \geq 0 \qquad (18\text{-}7)$$

to be the *power spectral density* of the w.s.s process $X(t)$. Notice that

$$R_{XX}(\omega) \xleftrightarrow{\text{F}\cdot\text{T}} S_{XX}(\omega) \geq 0. \qquad (18\text{-}8)$$

i.e., the autocorrelation function and the power spectrum of a w.s.s Process form a Fourier transform pair, a relation known as the **Wiener-Khinchin Theorem**. From (18-8), the inverse formula gives

$$R_{XX}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{XX}(\omega) e^{j\omega\tau} d\omega \qquad (18\text{-}9)$$

and in particular for $\tau = 0,$ we get

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{XX}(\omega) d\omega = R_{XX}(0) = E\{| X(t) |^2\} = P, \quad \textit{the total power.} \qquad (18\text{-}10)$$

From (18-10), the area under $S_{XX}(\omega)$ represents the total power of the process $X(t)$, and hence $S_{XX}(\omega)$ truly represents the power spectrum. (Fig 18.2).

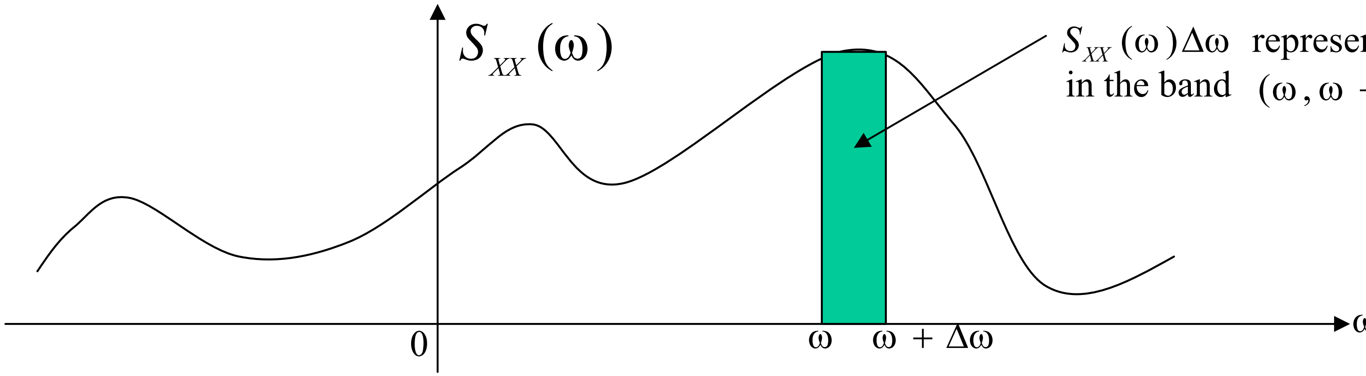


Fig 18.2

The nonnegative-definiteness property of the autocorrelation function in (14-8) translates into the "nonnegative" property for its Fourier transform (power spectrum), since from (14-8) and (18-9)

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* R_{XX}(t_i - t_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{XX}(\omega) e^{j\omega(t_i - t_j)} d\omega$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{XX}(\omega) \left| \sum_{i=1}^{n} a_i e^{j\omega t_i} \right|^2 d\omega \geq 0. \qquad (18\text{-}11)$$

From (18-11), it follows that

$$R_{XX}(\tau) \ \textit{nonnegative-definite} \ \Leftrightarrow \ S_{XX}(\omega) \geq 0. \qquad (18\text{-}12)$$

If $X(t)$ is a real w.s.s process, then $R_{XX}(\tau) = R_{XX}(-\tau)$ so that

$$
\begin{aligned}
S_{XX}(\omega) &= \int_{-\infty}^{+\infty} R_{XX}(\tau) e^{-j\omega\tau}\, d\tau \\
&= \int_{-\infty}^{+\infty} R_{XX}(\tau) \cos\omega\tau\, d\tau \\
&= 2\int_{0}^{\infty} R_{XX}(\tau) \cos\omega\tau\, d\tau = S_{XX}(-\omega) \geq 0 \qquad (18\text{-}13)
\end{aligned}
$$

so that the power spectrum is an even function, (in addition to being real and nonnegative).

Power Spectra and Linear Systems

If a w.s.s process $X(t)$ with autocorrelation function $R_{XX}(\tau) \;\leftrightarrow\; S_{XX}(\tau) \geq 0$ is applied to a linear system with impulse response $h(t)$, then the cross correlation function $R_{XY}(\tau)$ and the output autocorrelation function $R_{YY}(\tau)$ are given by (14-40)-(14-41). From there

$$X(t) \longrightarrow \boxed{h(t)} \longrightarrow Y(t)$$

Fig 18.3

$$R_{XY}(\tau) = R_{XX}(\tau) * h^*(-\tau), \quad R_{YY}(\tau) = R_{XX}(\tau) * h^*(-\tau) * h(\tau). \tag{18-14}$$

But if

$$f(t) \;\leftrightarrow\; F(\omega), \quad g(t) \;\leftrightarrow\; G(\omega) \tag{18-15}$$

Then

$$f(t) * g(t) \;\leftrightarrow\; F(\omega) G(\omega) \tag{18-16}$$

since

$$\mathcal{F}\{f(t) * g(t)\} = \int_{-\infty}^{+\infty} f(t) * g(t) e^{-j\omega t} dt$$

$$\mathcal{F}\{f(t) * g(t)\} = \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} f(\tau) g(t-\tau) d\tau \right\} e^{-j\omega t} dt$$

$$= \int_{-\infty}^{+\infty} f(\tau) e^{-j\omega\tau} d\tau \ \int_{-\infty}^{+\infty} g(t-\tau) e^{-j\omega(t-\tau)} d(t-\tau)$$

$$= F(\omega) G(\omega). \tag{18-17}$$

Using (18-15)-(18-17) in (18-14) we get

$$S_{XY}(\omega) = \mathcal{F}\{R_{XX}(\omega) * h^*(-\tau)\} = S_{XX}(\omega) H^*(\omega) \tag{18-18}$$

since

$$\int_{-\infty}^{+\infty} h^*(-\tau) e^{-j\omega\tau} d\tau = \left(\int_{-\infty}^{+\infty} h(t) e^{-j\omega t} dt \right)^* = H^*(\omega),$$

where

$$H(\omega) = \int_{-\infty}^{+\infty} h(t) e^{-j\omega t} dt \tag{18-19}$$

represents the transfer function of the system, and

$$S_{YY}(\omega) = \mathcal{F}\{R_{YY}(\tau)\} = S_{XY}(\omega) H(\omega)$$

$$= S_{XX}(\omega) \mid H(\omega) \mid^2 . \tag{18-20}$$

From (18-18), the cross spectrum need not be real or nonnegative; However the output power spectrum is real and nonnegative and is related to the input spectrum and the system transfer function as in (18-20). Eq. (18-20) can be used for system identification as well.

W.S.S White Noise Process: If $W(t)$ is a w.s.s white noise process, then from (14-43)

$$R_{WW}(\tau) = q\delta(\tau) \quad \Rightarrow \quad S_{WW}(\omega) = q. \tag{18-21}$$

Thus the spectrum of a white noise process is flat, thus justifying its name. Notice that a white noise process is unrealizable since its total power is indeterminate.

From (18-20), if the input to an unknown system in Fig 18.3 is a white noise process, then the output spectrum is given by

$$S_{YY}(\omega) = q\,|\,H(\omega)\,|^2 \tag{18-22}$$

Notice that the output spectrum captures the system transfer function characteristics entirely, and for rational systems Eq (18-22) may be used to determine the pole/zero locations of the underlying system.

Example 18.1: A w.s.s white noise process $W(t)$ is passed through a low pass filter (LPF) with bandwidth $B/2$. Find the autocorrelation function of the output process.

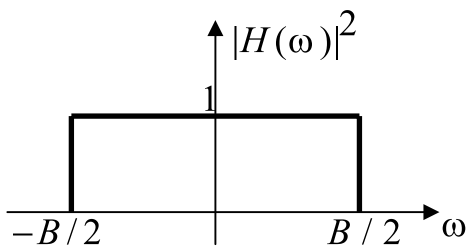
Solution: Let $X(t)$ represent the output of the LPF. Then from (18-22)

$$S_{XX}(\omega) = q\,|H(\omega)|^2 = \begin{cases} q, & |\omega| \le B/2 \\ 0, & |\omega| > B/2 \end{cases}. \tag{18-23}$$

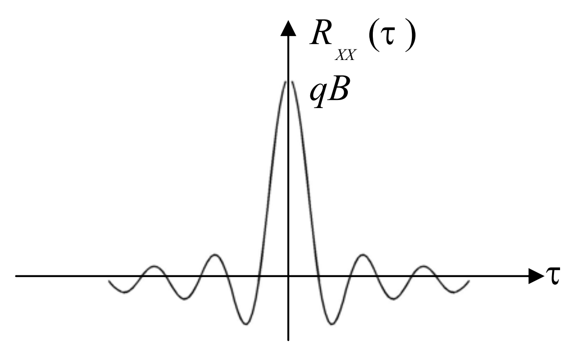
Inverse transform of $S_{XX}(\omega)$ gives the output autocorrelation function to be

$$R_{XX}(\tau) = \int_{-B/2}^{B/2} S_{XX}(\omega) e^{j\omega\tau}\, d\omega = q \int_{-B/2}^{B/2} e^{j\omega\tau}\, d\omega$$

$$= qB \frac{\sin(B\tau/2)}{(B\tau/2)} = qB \,\text{sinc}(B\tau/2) \tag{18-24}$$



(a) LPF



(b)

Fig. 18.4

Eq (18-23) represents colored noise spectrum and (18-24) its autocorrelation function (see Fig 18.4).

Example 18.2: Let

$$Y(t) = \frac{1}{2T}\int_{t-T}^{t+T} X(\tau)\,d\tau \tag{18-25}$$

represent a "smoothing" operation using a moving window on the input process $X(t)$. Find the spectrum of the output $Y(t)$ in term of that of $X(t)$.

Solution: If we define an LTI system with impulse response $h(t)$ as in Fig 18.5, then in term of $h(t)$, Eq (18-25) reduces to

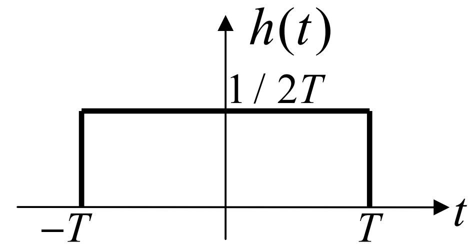


Fig 18.5

$$Y(t) = \int_{-\infty}^{+\infty} h(t-\tau) X(\tau)\,d\tau = h(t) * X(t) \tag{18-26}$$

so that

$$S_{YY}(\omega) = S_{XX}(\omega)\,|\,H(\omega)\,|^2 \,. \tag{18-27}$$

Here

$$H(\omega) = \int_{-T}^{+T} \tfrac{1}{2T} e^{-j\omega t} dt = \text{sinc}(\omega T) \tag{18-28}$$

so that

$$S_{YY}(\omega) = S_{XX}(\omega)\,\text{sinc}^2(\omega T). \tag{18-29}$$

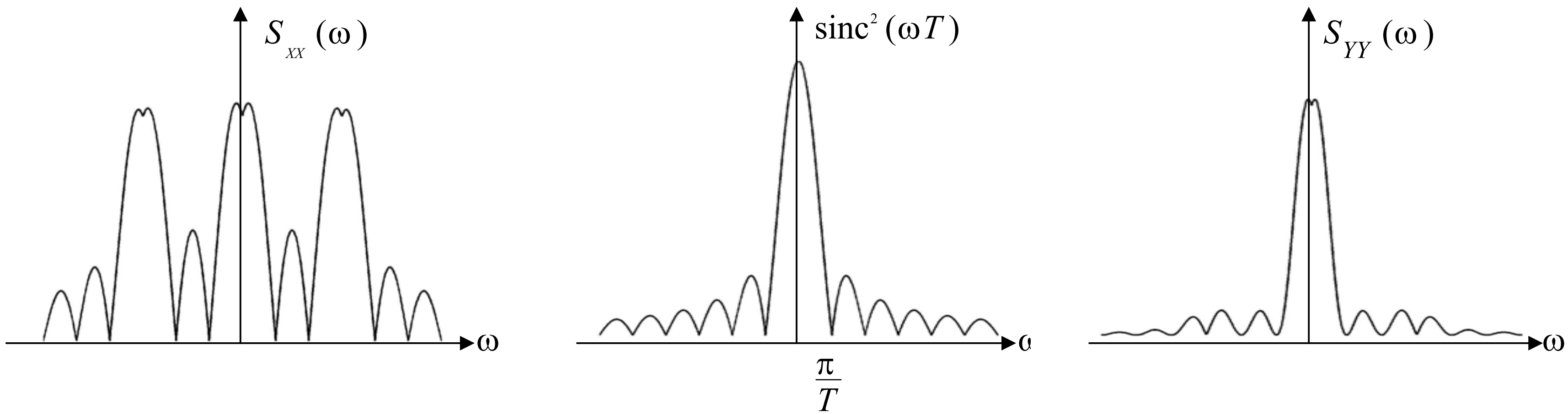


Fig 18.6

Notice that the effect of the smoothing operation in (18-25) is to suppress the high frequency components in the input (beyond π/T), and the equivalent linear system acts as a low-pass filter (continuous-time moving average) with bandwidth $2\pi/T$ in this case.

Discrete – Time Processes

For discrete-time w.s.s stochastic processes $X(nT)$ with autocorrelation sequence $\{r_k\}_{-\infty}^{+\infty}$, (proceeding as above) or formally defining a continuous time process $X(t)=\sum_n X(nT)\delta(t-nT)$, we get the corresponding autocorrelation function to be

$$R_{XX}(\tau)=\sum_{k=-\infty}^{+\infty} r_k \delta(\tau-kT).$$

Its Fourier transform is given by

$$S_{XX}(\omega)=\sum_{k=-\infty}^{+\infty} r_k e^{-j\omega T} \geq 0, \tag{18-30}$$

and it defines the power spectrum of the discrete-time process $X(nT)$. From (18-30),

$$S_{XX}(\omega)=S_{XX}(\omega+2\pi/T) \tag{18-31}$$

so that $S_{XX}(\omega)$ is a periodic function with period

$$2B=\frac{2\pi}{T}. \tag{18-32}$$

This gives the inverse relation

$$r_k = \frac{1}{2B}\int_{-B}^{B} S_{XX}(\omega)e^{jk\omega T}d\omega \tag{18-33}$$

and

$$r_0 = E\{|X(nT)|^2\} = \frac{1}{2B}\int_{-B}^{B} S_{XX}(\omega)d\omega \tag{18-34}$$

represents the total power of the discrete-time process $X(nT)$. The input-output relations for discrete-time system $h(nT)$ in (14-65)-(14-67) translate into

$$S_{XY}(\omega) = S_{XX}(\omega)H^*(e^{j\omega}) \tag{18-35}$$

and

$$S_{YY}(\omega) = S_{XX}(\omega)\,|\,H(e^{j\omega})\,|^2 \tag{18-36}$$

where

$$H(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} h(nT)\,e^{-j\omega nT} \tag{18-37}$$

represents the discrete-time system transfer function.

Matched Filter

Let $r(t)$ represent a deterministic signal $s(t)$ corrupted by noise. Thus

$$r(t) = s(t) + w(t), \qquad 0 < t < t_0 \tag{18-38}$$

where $r(t)$ represents the observed data, and it is passed through a receiver with impulse response $h(t)$. The output $y(t)$ is given by

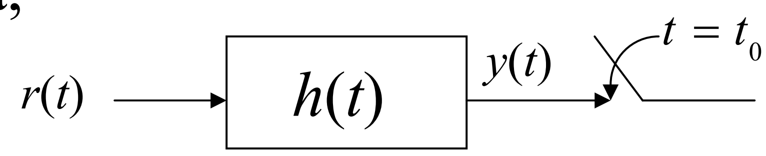


Fig 18.7 Matched Filter

$$y(t) \stackrel{\Delta}{=} y_s(t) + n(t) \tag{18-39}$$

where

$$y_s(t) = s(t) * h(t), \qquad n(t) = w(t) * h(t), \tag{18-40}$$

and it can be used to make a decision about the presence of absence of $s(t)$ in $r(t)$. Towards this, one approach is to require that the receiver output signal to noise ratio $(SNR)_0$ at time instant t_0 be maximized. Notice that

$$(SNR)_0 \triangleq \frac{\text{Output signal power at } t = t_0}{\text{Average output noise power}} = \frac{|\,y_s(t_0)\,|^2}{E\{|\,n(t)\,|^2\}}$$

$$= \frac{|\,y_s(t_0)\,|^2}{\frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{nn}(\omega)d\omega} = \frac{\left|\frac{1}{2\pi}\int_{-\infty}^{+\infty} S(\omega)H(\omega)e^{j\omega t_0}d\omega\right|^2}{\frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{WW}(\omega)\,|\,H(\omega)\,|^2\,d\omega} \tag{18-41}$$

represents the output SNR, where we have made use of (18-20) to determine the average output noise power, and the problem is to maximize $(\text{SNR})_0$ by optimally choosing the receiver filter $H(\omega)$.

Optimum Receiver for White Noise Input: The simplest input noise model assumes $w(t)$ to be white noise in (18-38) with spectral density N_0, so that (18-41) simplifies to

$$(SNR)_0 = \frac{\left|\int_{-\infty}^{+\infty} S(\omega)H(\omega)e^{j\omega t_0}d\omega\right|^2}{2\pi N_0 \int_{-\infty}^{+\infty} |\,H(\omega)\,|^2\,d\omega} \tag{18-42}$$

and a direct application of Cauchy-Schwarz' inequality in (18-42) gives

$$(SNR)_0 \le \frac{1}{2\pi N_0}\int_{-\infty}^{+\infty} |S(\omega)|^2 \, d\omega = \frac{\int_0^{+\infty} s(t)^2 \, dt}{N_0} = \frac{E_s}{N_0} \qquad (18\text{-}43)$$

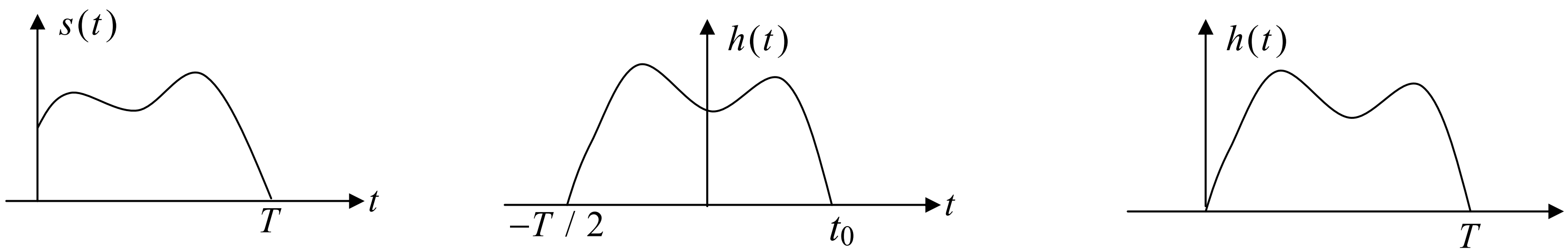
and equality in (18-43) is guaranteed if and only if

$$H(\omega) = S^*(\omega) e^{-j\omega t_0} \qquad (18\text{-}44)$$

or

$$h(t) = s(t_0 - t). \qquad (18\text{-}45)$$

From (18-45), the optimum receiver that maximizes the output SNR at $t = t_0$ is given by (18-44)-(18-45). Notice that (18-45) need not be causal, and the corresponding SNR is given by (18-43).



(a) (b) t_0=T/2 (c) t_0=T

Fig 18.8

Fig 18-8 shows the optimum $h(t)$ for two different values of t_0. In Fig 18.8 (b), the receiver is noncausal, whereas in Fig 18-8 (c) the receiver represents a causal waveform.

If the receiver is not causal, the optimum causal receiver can be shown to be

$$h_{opt}(t) = s(t_0 - t)u(t) \tag{18-46}$$

and the corresponding maximum $(SNR)_0$ in that case is given by

$$(SNR_0) = \frac{1}{N_0}\int_0^{t_0} s^2(t)dt \tag{18-47}$$

Optimum Transmit Signal: In practice, the signal $s(t)$ in (18-38) may be the output of a target that has been illuminated by a transmit signal $f(t)$ of finite duration T. In that case

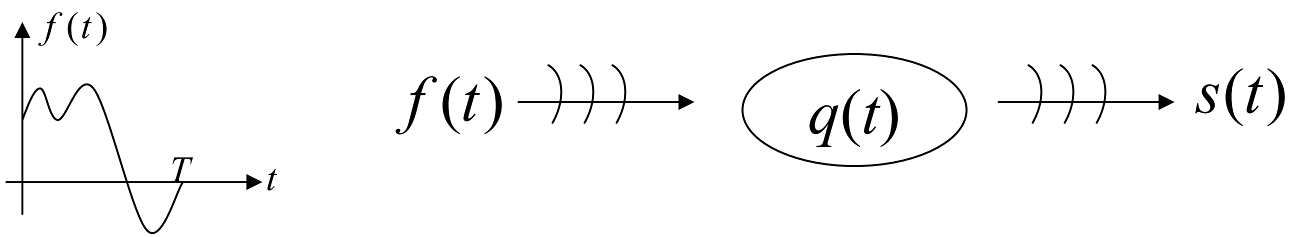


Fig 18.9

$$s(t) = f(t) * q(t) = \int_0^T f(\tau)q(t-\tau)d\tau, \tag{18-48}$$

where $q(t)$ represents the target impulse response. One interesting question in this context is to determine the optimum transmit signal $f(t)$ with normalized energy that maximizes the receiver output SNR at $t = t_0$ in Fig 18.7. Notice that for a given $s(t)$, Eq (18-45) represents the optimum receiver, and (18-43) gives the corresponding maximum $(\text{SNR})_0$. To maximize $(\text{SNR})_0$ in (18-43), we may substitute (18-48) into (18-43). This gives

$$(SNR)_0 = \int_0^\infty |\{\int_0^T q(t-\tau_1) f(\tau_1) d\tau_1\}|^2 \, dt$$

$$= \frac{1}{N_0} \int_0^T \int_0^T \underbrace{\int_0^\infty q(t-\tau_1) q^*(t-\tau_2) dt}_{\Omega(\tau_1,\tau_2)} f(\tau_2) d\tau_2 f(\tau_1) d\tau_1$$

$$= \frac{1}{N_0} \int_0^T \{\int_0^T \Omega(\tau_1,\tau_2) f(\tau_2) d\tau_2\} f(\tau_1) d\tau_1 \le \lambda_{\max} / N_0 \qquad (18\text{-}49)$$

where $\Omega(\tau_1,\tau_2)$ is given by

$$\Omega(\tau_1,\tau_2) = \int_0^\infty q(t-\tau_1) q^*(t-\tau_2) dt \qquad (18\text{-}50)$$

and $\lambda_{\max}$ is the largest eigenvalue of the integral equation

$$\int_0^T \Omega(\tau_1,\tau_2) f(\tau_2) d\tau_2 = \lambda_{\max} f(\tau_1), \qquad 0 < \tau_1 < T. \qquad (18\text{-}51)$$

and

$$\int_0^T f^2(t)dt = 1. \tag{18-52}$$

Observe that the kernal $\Omega(\tau_1, \tau_2)$ in (18-50) captures the target characteristics so as to maximize the output SNR at the observation instant, and the optimum transmit signal is the solution of the integral equation in (18-51) subject to the energy constraint in (18-52). Fig 18.10 show the optimum transmit signal and the companion receiver pair for a specific target with impulse response $q(t)$ as shown there .

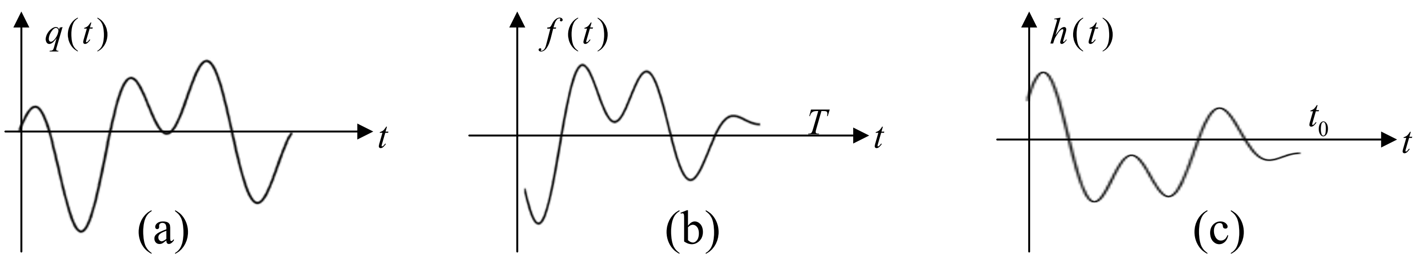


Fig 18.10

If the causal solution in (18-46)-(18-47) is chosen, in that case the kernel in (18-50) simplifies to

$$\Omega(\tau_1, \tau_2) = \int_0^{t_0} q(t-\tau_1) q^*(t-\tau_2) dt. \tag{18-53}$$

and the optimum transmit signal is given by (18-51). Notice that in the causal case, information beyond $t = t_0$ is not used.

What if the additive noise in (18-38) is not white?

Let $S_{WW}(\omega)$ represent a (non-flat) power spectral density. In that case, what is the optimum matched filter?

If the noise is *not* white, one approach is to *whiten* the input noise first by passing it through a whitening filter, and then proceed with the whitened output as before (Fig 18.7).

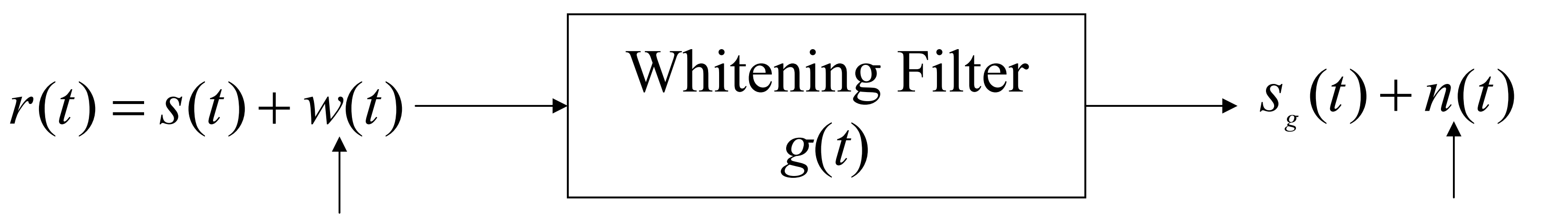


Fig 18.11

Notice that the signal part of the whitened output $s_g(t)$ equals

$$s_g(t) = s(t) * g(t) \qquad (18\text{-}54)$$

where $g(t)$ represents the whitening filter, and the output noise $n(t)$ is white with unit spectral density. This interesting idea due to

Wiener has been exploited in several other problems including prediction, filtering etc.

Whitening Filter: What is a whitening filter? From the discussion above, the output spectral density of the whitened noise process $S_{nn}(\omega)$ equals unity, since it represents the normalized white noise by design. But from (18-20)

$$1 = S_{nn}(\omega) = S_{WW}(\omega)\,|\,G(\omega)\,|^2,$$

which gives

$$|\,G(\omega)\,|^2 = \frac{1}{S_{WW}(\omega)}. \tag{18-55}$$

i.e., the whitening filter transfer function $G(\omega)$ satisfies the magnitude relationship in (18-55). To be useful in practice, it is desirable to have the whitening filter to be *stable* and *causal* as well. Moreover, at times its inverse transfer function also needs to be implementable so that it needs to be stable as well. How does one obtain such a filter (if any)? [See section 11.1 page 499-502, (and also page 423-424), Text for a discussion on obtaining the whitening filters.].

From there, any spectral density that satisfies the finite power constraint

$$\int_{-\infty}^{+\infty} S_{XX}(\omega)\,d\omega < \infty \tag{18-56}$$

and the Paley-Wiener constraint (see Eq. (11-4), Text)

$$\int_{-\infty}^{+\infty} \frac{|\log S_{XX}(\omega)|}{1+\omega^2}\,d\omega < \infty \tag{18-57}$$

can be factorized as

$$S_{XX}(\omega) = |H(j\omega)|^2 = H(s)H(-s)\big|_{s=j\omega} \tag{18-58}$$

where $H(s)$ together with its inverse function $1/H(s)$ represent two filters that are both analytic in Re $s > 0$. Thus $H(s)$ and its inverse $1/H(s)$ can be chosen to be s*table* and *causal* in (18-58). Such a filter is known as the Wiener factor, and since it has all its poles and zeros in the left half plane, it represents a *minimum phase factor*. In the rational case, if $X(t)$ represents a real process, then $S_{XX}(\omega)$ is even and hence (18-58) reads

$$0 \le S_{XX}(\omega^2) = \tilde{S}_{XX}(-s^2)\,|_{s=j\omega} = H(s)H(-s)\,|_{s=j\omega}\,. \tag{18-59}$$

Example 18.3: Consider the spectrum

$$S_{XX}(\omega) = \frac{(\omega^2+1)(\omega^2-2)^2}{(\omega^4+1)}$$

which translates into

$$\tilde{S}_{XX}(-s^2) = \frac{(1-s^2)(2+s^2)^2}{1+s^4}.$$

The poles ($\times$) and zeros ($\circ$) of this function are shown in Fig 18.12. From there to maintain the symmetry condition in (18-59), we may group together the left half factors as

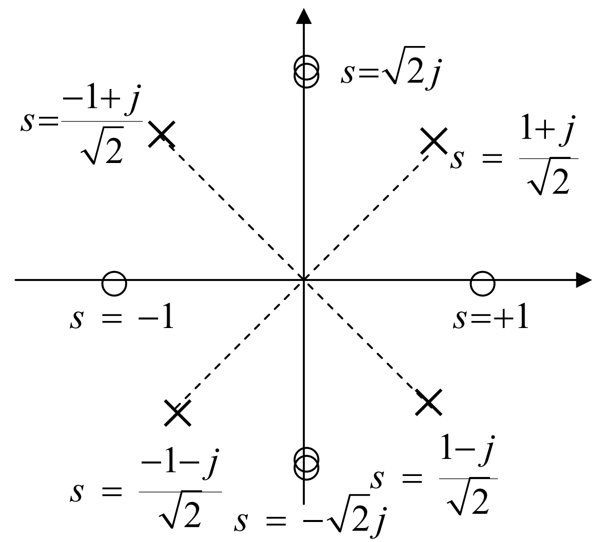


Fig 18.12

$$H(s) = \frac{(s+1)(s-\sqrt{2}j)(s+\sqrt{2}j)}{\left(s+\frac{1+j}{\sqrt{2}}\right)\left(s+\frac{1-j}{\sqrt{2}}\right)} = \frac{(s+1)(s^2+2)}{s^2+\sqrt{2}s+1}$$

and it represents the Wiener factor for the spectrum $S_{XX}(\omega)$ above. Observe that the poles and zeros (if any) on the $j\omega-$axis appear in even multiples in $S_{XX}(\omega)$ and hence half of them may be paired with H(s) (and the other half with $H(-s)$) to preserve the factorization condition in (18-58). Notice that $H(s)$ is stable, and so is its inverse.

More generally, if $H(s)$ is minimum phase, then ln $H(s)$ is analytic on the right half plane so that

$$H(\omega) = A(\omega)e^{-j\varphi(\omega)} \tag{18-60}$$

gives

$$\ln H(\omega) = \ln A(\omega) - j\varphi(\omega) \triangleq \int_0^{+\infty} b(t)e^{-j\omega t}dt.$$

Thus

$$\ln A(\omega) = \int_0^t b(t)\cos\omega t\ dt$$

$$\varphi(\omega) = \int_0^t b(t)\sin\omega t\ dt$$

and since $\cos\omega t$ and $\sin\omega t$ are Hilbert transform pairs, it follows that the phase function $\varphi(\omega)$ in (18-60) is given by the Hilbert transform of $\ln A(\omega)$. Thus

$$\varphi(\omega) = \mathcal{H}\{\ln A(\omega)\}. \tag{18-61}$$

Eq. (18-60) may be used to generate the unknown phase function of a minimum phase factor from its magnitude.

For discrete-time processes, the factorization conditions take the form (see (9-203)-(9-205), Text)

$$\int_{-\pi}^{\pi} S_{XX}(\omega)d\omega < \infty \tag{18-62}$$

and

$$\int_{-\pi}^{\pi} \ln S_{XX}(\omega)d\omega > -\infty. \tag{18-63}$$

In that case

$$S_{XX}(\omega) = |H(e^{j\omega})|^2$$

where the discrete-time system

$$H(z) = \sum_{k=0}^{\infty} h(k)z^{-k}$$

is analytic together with its inverse in $|z| > 1$. This unique minimum phase function represents the Wiener factor in the discrete-case.

Matched Filter in Colored Noise:

Returning back to the matched filter problem in colored noise, the design can be completed as shown in Fig 18.13.

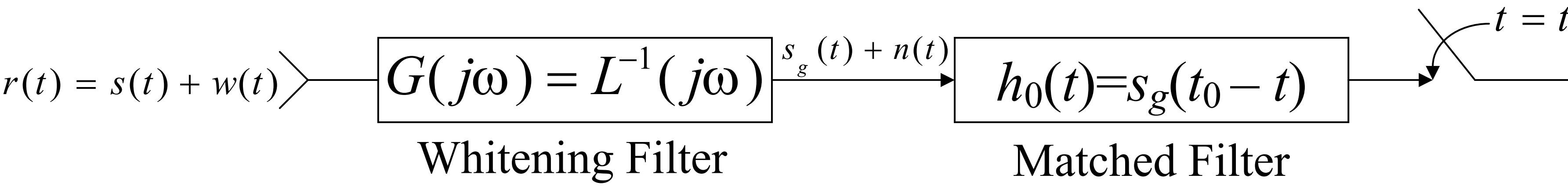


Fig 18.13

(Here $G(j\omega)$ represents the whitening filter associated with the noise spectral density $S_{WW}(\omega)$ as in (18-55)-(18-58). Notice that $G(s)$ is the inverse of the Wiener factor $L(s)$ corresponding to the spectrum $S_{WW}(\omega)$. i.e.,

$$L(s)L(-s)\,|_{s=j\omega} = |\,L(j\omega)\,|^2 = S_{WW}(\omega). \qquad (18\text{-}64)$$

The whitened output $s_g(\mathrm{t}) + n(t)$ in Fig 18.13 is similar to (18-38), and from (18-45) the optimum receiver is given by

$$h_0(t) = s_g(t_0 - t)$$

where

$$s_g(t) \leftrightarrow S_g(\omega) = G(j\omega)S(\omega) = L^{-1}(j\omega)S(\omega).$$

If we insist on obtaining the receiver transfer function $H(\omega)$ for the original colored noise problem, we can deduce it easily from Fig 18.14

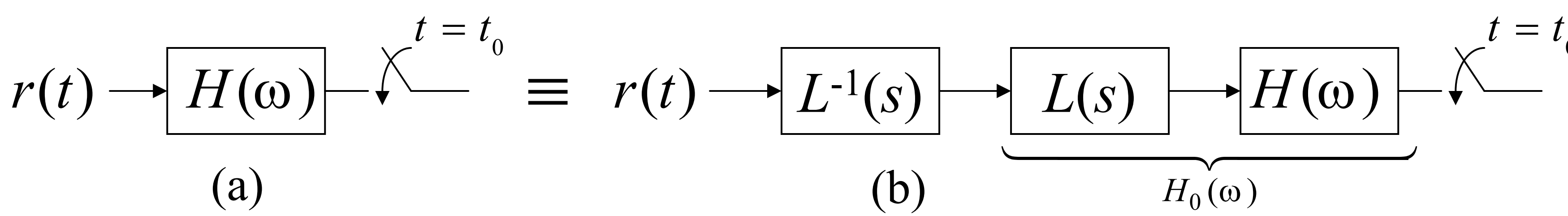


Fig 18.14

Notice that Fig 18.14 (a) and (b) are equivalent, and Fig 18.14 (b) is equivalent to Fig 18.13. Hence (see Fig 18.14 (b))

$$H_0(\omega) = L(j\omega)H(\omega)$$

or

$$H(\omega) = L^{-1}(j\omega)H_0(\omega) = L^{-1}(\omega)S_g^*(\omega)e^{-j\omega t_0}$$
$$= L^{-1}(\omega)\{L^{-1}(\omega)S(\omega)\}^* e^{-j\omega t_0} \qquad (18\text{-}65)$$

turns out to be the overall matched filter for the original problem. Once again, transmit signal design can be carried out in this case also.

AM/FM Noise Analysis:

Consider the noisy AM signal

$$X(t) = m(t)\cos(\omega_0 t + \theta) + n(t), \qquad (18\text{-}66)$$

and the noisy FM signal

$$X(t) = A\cos(\omega_0 t + \varphi(t) + \theta) + n(t), \qquad (18\text{-}67)$$

where

$$\varphi(t) = \begin{cases} c\int_0^t m(\tau)\,d\tau & \text{FM} \\ cm(t) & \text{PM.} \end{cases} \qquad (18\text{-}68)$$

Here $m(t)$ represents the message signal and θ a random phase jitter in the received signal. In the case of FM, $\omega(t) = \varphi'(t) = c\, m(t)$ so that the instantaneous frequency is proportional to the message signal. We will assume that both the message process $m(t)$ and the noise process $n(t)$ are w.s.s with power spectra $S_{mm}(\omega)$ and $S_{nn}(\omega)$ respectively. We wish to determine whether the AM and FM signals are w.s.s, and if so their respective power spectral densities.

Solution: AM signal: In this case from (18-66), if we assume $\theta \sim U(0, 2\pi)$, then

$$R_{XX}(\tau) = \frac{1}{2} R_{mm}(\tau)\cos\omega_0\tau + R_{nn}(\tau) \tag{18-69}$$

so that (see Fig 18.15)

$$S_{XX}(\omega) = \frac{S_{XX}(\omega-\omega_0) + S_{XX}(\omega+\omega_0)}{2} + S_{nn}(\omega). \tag{18-70}$$

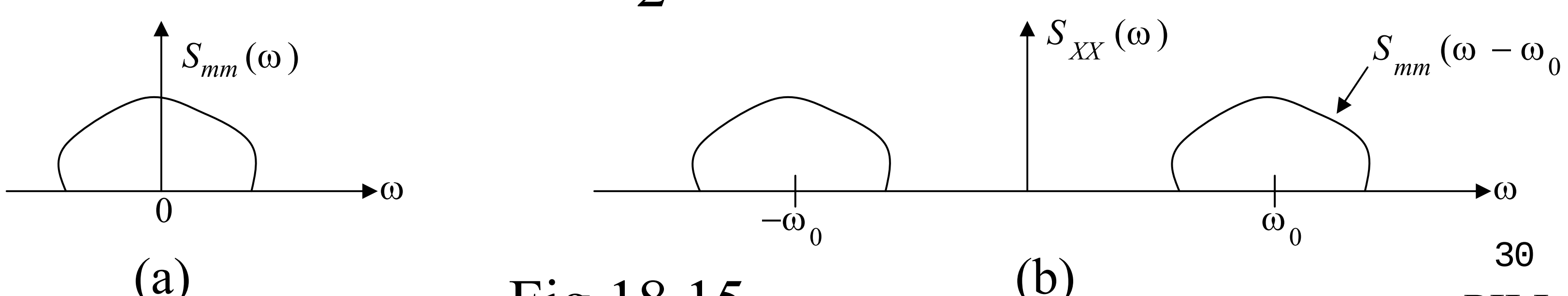


Fig 18.15

Thus AM represents a stationary process under the above conditions. What about FM?

FM signal: In this case (suppressing the additive noise component in (18-67)) we obtain

$$
\begin{aligned}
R_{XX}(t+\tau/2, t-\tau/2) &= A^2 E\{\cos(\omega_0(t+\tau/2)+\varphi(t+\tau/2)+\theta)\times \\
&\qquad \cos(\omega_0(t-\tau/2)+\varphi(t-\tau/2)+\theta)\} \\
&= \frac{A^2}{2} E\{\cos[\omega_0\tau+\varphi(t+\tau/2)-\varphi(t-\tau/2)] \\
&\qquad +\cos[2\omega_0 t+\varphi(t+\tau/2)+\varphi(t-\tau/2)+2\theta]\} \\
&= \frac{A^2}{2} [E\{\cos(\varphi(t+\tau/2)-\varphi(t-\tau/2))\}\cos\omega_0\tau \\
&\qquad -E\{\sin(\varphi(t+\tau/2)-\varphi(t-\tau/2))\}\sin\omega_0\tau]
\end{aligned}
\tag{18-71}
$$

since

$$
\begin{aligned}
&E\{\cos(2\omega_0 t+\varphi(t+\tau/2)+\varphi(t-\tau/2)+2\theta)\} \\
&= E\{\cos(2\omega_0 t+\varphi(t+\tau/2)+\varphi(t-\tau/2))\}\overset{0}{E\{\cos 2\theta\}} \\
&-E\{\sin(2\omega_0 t+\varphi(t+\tau/2)+\varphi(t-\tau/2))\}\overset{0}{E\{\sin 2\theta\}} = 0.
\end{aligned}
$$

Eq (18-71) can be rewritten as

$$R_{XX}(t+\tau/2, t-\tau/2) = \frac{A^2}{2}[a(t,\tau)\cos\omega_0\tau - b(t,\tau)\sin\omega_0\tau] \qquad (18\text{-}72)$$

where

$$a(t,\tau) \stackrel{\Delta}{=} E\{\cos(\varphi(t+\tau/2) - \varphi(t-\tau/2))\} \qquad (18\text{-}73)$$

and

$$b(t,\tau) \stackrel{\Delta}{=} E\{\sin(\varphi(t+\tau/2) - \varphi(t-\tau/2))\} \qquad (18\text{-}74)$$

In general $a(t,\tau)$ and $b(t,\tau)$ depend on both t and τ so that noisy FM is *not* w.s.s in general, even if the message process $m(t)$ is w.s.s.
In the special case when m(t) is a stationary Gaussian process, from (18-68), $\varphi(t)$ is also a stationary Gaussian process with autocorrelation function

$$R_{\varphi'\varphi'}(\tau) = c^2 R_{mm}(\tau) = \frac{-d^2 R_{\varphi\varphi}(\tau)}{d\tau^2} \qquad (18\text{-}75)$$

for the FM case. In that case the random variable

$$Y \stackrel{\Delta}{=} \varphi(t+\tau/2) - \varphi(t-\tau/2) \sim N(0,\sigma_Y^2) \tag{18-76}$$

where

$$\sigma_Y^2 = 2(R_{\varphi\varphi}(0) - R_{\varphi\varphi}(\tau)). \tag{18-77}$$

Hence its characteristic function is given by

$$E\{e^{j\omega Y}\} = e^{-\omega^2 \sigma_Y^2/2} = e^{-(R_{\varphi\varphi}(0) - R_{\varphi\varphi}(\tau))\omega^2} \tag{18-78}$$

which for $\omega = 1$ gives

$$E\{e^{jY}\} = E\{\cos Y\} + jE\{\sin Y\} = a(t,\tau) + jb(t,\tau), \tag{18-79}$$

where we have made use of (18-76) and (18-73)-(18-74). On comparing (18-79) with (18-78) we get

$$a(t,\tau) = e^{-(R_{\varphi\varphi}(0) - R_{\varphi\varphi}(\tau))} \tag{18-80}$$

and

$$b(t,\tau) \equiv 0 \tag{18-81}$$

so that the FM autocorrelation function in (18-72) simplifies into

$$R_{XX}(\tau) = \frac{A^2}{2} e^{-(R_{\varphi\varphi}(0) - R_{\varphi\varphi}(\tau))} \cos\omega_0\tau. \tag{18-82}$$

Notice that for stationary Gaussian message input $m(t)$ (or $\varphi(t)$), the nonlinear output $X(t)$ is indeed strict sense stationary with autocorrelation function as in (18-82).

Narrowband FM: If $R_{\varphi\varphi}(0) \ll 1$, then (18-82) may be approximated as $(e^{-x} \simeq 1 - x, \ |x| << 1)$

$$R_{XX}(\tau) = \frac{A^2}{2} \{(1 - R_{\varphi\varphi}(0)) + R_{\varphi\varphi}(\tau)\} \cos\omega_0\tau \tag{18-83}$$

which is similar to the AM case in (18-69). Hence narrowband FM and ordinary AM have equivalent performance in terms of noise suppression.

Wideband FM: This case corresponds to $R_{\varphi\varphi}(0) > 1$. In that case a Taylor series expansion or $R_{\varphi\varphi}(\tau)$ gives

$$R_{\varphi\varphi}(\tau) = R_{\varphi\varphi}(0) + \frac{1}{2} R''_{\varphi\varphi}(0)\tau^2 + \cdots \; = R_{\varphi\varphi}(0) - \frac{c^2}{2} R_{mm}(0)\tau^2 + \cdots \quad (18\text{-}84)$$

and substituting this into (18-82) we get

$$R_{XX}(\tau) = \frac{A^2}{2}\left\{ e^{-\frac{c^2}{2} R_{mm}(0)\tau^2 + \cdots} \right\} \cos\omega_0\tau \qquad (18\text{-}85)$$

so that the power spectrum of FM in this case is given by

$$S_{XX}(\omega) = \tfrac{1}{2}\{ S(\omega - \omega_0) + S(\omega + \omega_0) \} \qquad (18\text{-}86)$$

where

$$S(\omega) \simeq \frac{A^2}{2} e^{-\omega^2 / 2c^2 R_{mm}(0)}. \qquad (18\text{-}87)$$

Notice that $S_{XX}(\omega)$ always occupies infinite bandwidth irrespective of the actual message bandwidth (Fig 18.16)and this capacity to spread the message signal across the entire spectral band helps to reduce the noise effect in any band.

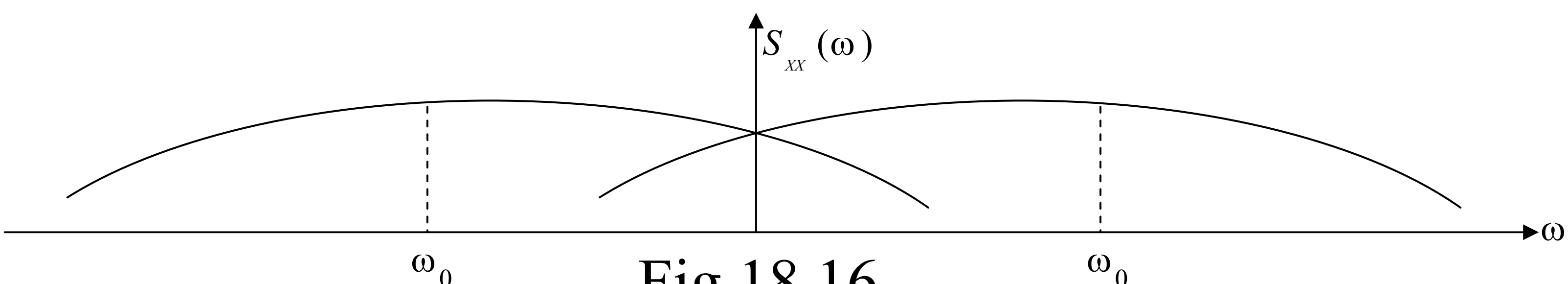


Fig 18.16

Spectrum Estimation / Extension Problem

Given a finite set of autocorrelations $r_0, r_1, \cdots, r_n,$ one interesting problem is to extend the given sequence of autocorrelations such that the spectrum corresponding to the overall sequence is nonnegative for all frequencies. i.e., given $r_0, r_1, \cdots, r_n,$ we need to determine $r_{n+1}, r_{n+2}, \cdots$ such that

$$S(\omega) = \sum_{k=-\infty}^{\infty} r_k e^{-jk\omega} \geq 0. \qquad (18\text{-}88)$$

Notice that from (14-64), the given sequence satisfies $T_n > 0$, and at every step of the extension, this nonnegativity condition must be satisfied. Thus we must have

$$T_{n+k} > 0, \qquad k = 1,\ 2, \cdots. \qquad (18\text{-}89)$$

Let $r_{n+1} = x.$ Then

$$
T_{n+1} = \begin{vmatrix} & & x \\ & & r_n \\ & T_n & \vdots \\ & & r_1 \\ x^*, r_n^*, & \cdots\; r_1^* & r_0 \end{vmatrix}
$$

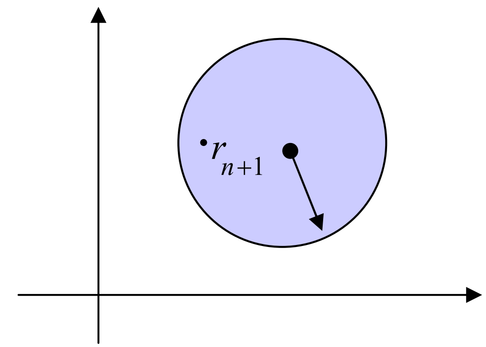


Fig 18.17

so that after some algebra

$$
\Delta_{n+1} = \det T_{n+1} = \frac{\Delta_n^2 - \Delta_{n-1}^2 \,|\, x - \xi_n \,|^2}{\Delta_{n-1}} > 0 \qquad (18\text{-}90)
$$

or

$$
|\, r_{n+1} - \xi_n \,|^2 \quad < \quad \left(\frac{\Delta_n}{\Delta_{n-1}} \right)^2, \qquad (18\text{-}91)
$$

where

$$
\xi_n \triangleq \underline{a}^T T_{n-1}^{-1} \underline{b}, \quad \underline{a} \triangleq [r_1, r_2, \cdots, r_n]^T, \quad \underline{b} \triangleq [r_n, r_{n-1}, \cdots, r_1]^T. \qquad (18\text{-}92)
$$

Eq. (18-91) represents the interior of a circle with center ξ_n and radius Δ_n / Δ_{n-1} as in Fig 18.17, and geometrically it represents the admissible set of values for r_{n+1}. Repeating this procedure for $r_{n+2}, r_{n+3}, \cdots,$ it follows that the class of extensions that satisfy (18-85) are infinite.

It is possible to parameterically represent the class of all admissible spectra. Known as the trigonometric moment problem, extensive literature is available on this topic.

[See section 12.4 "Youla's Parameterization", pages 562-574, Text for a reasonably complete description and further insight into this topic.].

19. Series Representation of Stochastic Processes

Given information about a stochastic process $X(t)$ in $0 \le t \le T$, can this continuous information be represented in terms of a countable set of random variables whose relative importance decrease under some arrangement?

To appreciate this question it is best to start with the notion of a **Mean-Square periodic process**. A stochastic process $X(t)$ is said to be mean square (M.S) periodic, if for some $T > 0$

$$E[|X(t+T) - X(t)|^2] = 0 \quad \text{for all } t. \tag{19-1}$$

i.e $X(t) = X(t+T)$ with *probability* 1 for all t.
Suppose $X(t)$ is a W.S.S process. Then

$X(t)$ is mean-square perodic $\Leftrightarrow$ $R(\tau)$ is periodic in the ordinary sense, where $R(\tau) = E[X(t)X^*(t+T)]$

Proof: $(\Rightarrow)$ suppose $X(t)$ is M.S. periodic. Then

$$E[|X(t+T)-X(t)|^2]=0. \qquad (19\text{-}2)$$

But from Schwarz' inequality

$$\left|E[X(t_1)\{X(t_2+T)-X(t_2)\}^*]\right|^2 \le E[|X(t_1)|^2]\underbrace{E[|X(t_2+T)-X(t_2)|^2]}_{0}$$

Thus the left side equals

$$E[X(t_1)\{X(t_2+T)-X(t_2)\}^*]=0$$

or

$$E[X(t_1)X^*(t_2+T)]=E[X(t_1)X^*(t_2)] \quad \Rightarrow \quad R(t_2-t_1+T)=R(t_2-t_1)$$

$$\Rightarrow \quad R(\tau+T)=R(\tau) \qquad \text{for any } \tau$$

i.e., $R(\tau)$ is periodic with period T. (19-3)

$(\Leftarrow)$ Suppose $R(\tau)$ is periodic. Then

$$E[|X(t+\tau)-X(t)|^2]=2R(0)-R(\tau)-R^*(\tau)=0$$

i.e., $X(t)$ is mean square periodic.

Thus if $X(t)$ is mean square periodic, then $R(\tau)$ is periodic and let

$$R(\tau) = \sum_{-\infty}^{+\infty} \gamma_n e^{jn\omega_0 \tau}, \qquad \omega_0 = \frac{2\pi}{T} \tag{19-4}$$

represent its Fourier series expansion. Here

$$\gamma_n = \frac{1}{T} \int_0^T R(\tau) e^{-jn\omega_0 \tau} d\tau. \tag{19-5}$$

In a similar manner define

$$c_k = \frac{1}{T} \int_0^T X(t) e^{jk\omega_0 t} dt \tag{19-6}$$

Notice that $c_k, \quad k = -\infty \to +\infty$ are random variables, and

$$E[c_k c_m^*] = \frac{1}{T^2} E[\int_0^T X(t_1) e^{jk\omega_0 t_1} dt_1 \int_0^T X^*(t_2) e^{-jm\omega_0 t_2} dt_2]$$

$$= \frac{1}{T^2} \int_0^T \int_0^T R(t_2 - t_1) e^{jk\omega_0 t_1} e^{-jm\omega_0 t_2} dt_1 dt_2$$

$$= \frac{1}{T} \int_0^T \underbrace{[\frac{1}{T} \int_0^T R \underbrace{(t_2 - t_1)}_{\tau} e^{-jm\omega_0 \overbrace{(t_2 - t_1)}^{\tau}} d \underbrace{(t_2 - t_1)}_{\tau}]}_{\gamma_m} e^{-j(m-k)\omega_0 t_1} dt_1$$

$$E[c_k c_m^*] = \gamma_m \underbrace{\{\tfrac{1}{T}\int_0^T e^{-j(m-k)\omega_0 t_1} dt_1\}}_{\delta_{m,k}} = \begin{cases} \gamma_m > 0, & k = m \\ 0 & k \neq m. \end{cases} \tag{19-7}$$

i.e., $\{c_n\}_{n=-\infty}^{n=+\infty}$ form a sequence of uncorrelated random variables, and, further, consider the partial sum

$$\tilde{X}_N(t) = \sum_{K=-N}^{N} c_k e^{-jk\omega_0 t}. \tag{19-8}$$

We shall show that $\tilde{X}_N(t) = X(t)$ in the mean square sense as $N \to \infty$. i.e.,

$$E[|X(t) - \tilde{X}_N(t)|^2] \to 0 \quad \text{as} \quad N \to \infty. \tag{19-9}$$

Proof:

$$E[|X(t) - \tilde{X}_N(t)|^2] = E[|X(t)|^2] - 2\,\mathrm{Re}[E(X^*(t)\tilde{X}_N(t)]$$
$$+E[|\tilde{X}_N(t)|^2]. \tag{19-10}$$

But

$$E[|X(t)|^2] = R(0) = \sum_{k=-\infty}^{+\infty} \gamma_k,$$

and

$$E[X^*(t)\tilde{X}_N(t)] = E[\sum_{k=-N}^{N} c_k e^{-jk\omega_0 t} X^*(t)]$$

$$= \frac{1}{T}\sum_{k=-N}^{N} E[\int_0^T X(\alpha) e^{-jk\omega_0 (t-\alpha)} X^*(t) d\alpha]$$

$$= \sum_{k=-N}^{N} \underbrace{[\frac{1}{T}\int_0^T R(t-\alpha) e^{-jk\omega_0 (t-\alpha)} d(t-\alpha)]}_{\gamma_k} = \sum_{k=-N}^{N} \gamma_k. \qquad (19\text{-}12)$$

Similarly

$$E[|\tilde{X}_N(t)|^2] = E[\sum_k \sum_m c_k c_m^* e^{j(k-m)\omega_0 t} = \sum_k \sum_m E[c_k c_m^*] e^{j(k-m)\omega_0 t} = \sum_{k=-N}^{N} \gamma_k.$$

$$\Rightarrow E[|X(t) - \tilde{X}_N(t)|^2] = 2(\sum_{k=-\infty}^{+\infty} \gamma_k - \sum_{k=-N}^{N} \gamma_k) \to 0 \quad \text{as} \quad N \to \infty \qquad (19\text{-}13)$$

i.e.,

$$X(t) \doteq \sum_{k=-\infty}^{+\infty} c_k e^{-jk\omega_0 t}, \qquad -\infty < t < +\infty. \qquad (19\text{-}14)$$

Thus mean square periodic processes can be represented in the form of a series as in (19-14). The stochastic information is contained in the random variables $c_k, \quad k = -\infty \to +\infty.$ Further these random variables are uncorrelated $(E\{c_k c_m^*\} = \gamma_k \delta_{k,m})$ and their variances $\gamma_k \to 0$ as $k \to \infty.$ This follows by noticing that from (19-14)

$$\sum_{k=-\infty}^{+\infty} \gamma_k = R(0) = E[|X(t)|^2] = P < \infty.$$

Thus if the power P of the stochastic process is finite, then the positive sequence $\sum_{k=-\infty}^{+\infty} \gamma_k$ converges, and hence $\gamma_k \to 0$ as $k \to \infty.$ This implies that the random variables in (19-14) are of relatively less importance as $k \to \infty,$ and a finite approximation of the series in (19-14) is indeed meaningful.

The following natural question then arises: What about a general stochastic process, that is *not* mean square periodic? Can it be represented in a similar series fashion as in (19-14), if not in the whole interval $-\infty < t < \infty,$ say in a finite support $0 \le t \le T$?

Suppose that it is indeed possible to do so for any arbitrary process $X(t)$ in terms of a certain sequence of orthonormal functions.

i.e.,

$$\tilde{X}(t) = \sum_{n=1}^{\infty} c_k \varphi_k(t) \tag{19-15}$$

where

$$c_k \triangleq \int_0^T X(t)\varphi_k^*(t)dt \tag{19-16}$$

$$\int_0^T \varphi_k(t)\varphi_n^*(t)dt = \delta_{k,n}, \tag{19-17}$$

and in the mean square sense

$$\tilde{X}(t) \doteq X(t) \quad \text{in} \quad 0 \le t \le T.$$

Further, as before, we would like the c_k s to be uncorrelated random variables. If that should be the case, then we must have

$$E[c_k c_m^*] = \lambda_m \delta_{k,m}. \tag{19-18}$$

Now

$$\begin{aligned} E[c_k c_m^*] &= E[\int_0^T X(t_1)\varphi_k^*(t_1)dt_1 \int_0^T X^*(t_2)\varphi_m(t_2)dt_2] \\ &= \int_0^T \varphi_k^*(t_1)\int_0^T E\{X(t_1)X^*(t_2)\}\ \varphi_m(t_2)dt_2 dt_1 \\ &= \int_0^T \varphi_k^*(t_1)\{\int_0^T R_{XX}(t_1,t_2)\ \varphi_m(t_2)dt_2\}dt_1 \qquad (19\text{-}19) \end{aligned}$$

and

$$\lambda_m \delta_{k,m} = \lambda_m \int_0^T \varphi_k^*(t_1)\varphi_m(t_1)dt_1. \tag{19-20}$$

Substituting (19-19) and (19-20) into (19-18), we get

$$\int_0^T \varphi_k^*(t_1)\{\int_0^T R_{XX}(t_1,t_2)\,\varphi_m(t_2)dt_2 - \lambda_m \varphi_m(t_1)\}dt_1 = 0. \tag{19-21}$$

Since (19-21) should be true for *every* $\varphi_k(t), \ \ k = 1 \to \infty,$ we must have

$$\int_0^T R_{XX}(t_1,t_2)\varphi_m(t_2)dt_2 - \lambda_m \varphi_m(t_1) \equiv 0,$$

or

$$\int_0^T R_{XX}(t_1,t_2)\varphi_m(t_2)dt_2 = \lambda_m \varphi_m(t_1), \quad 0 < t_1 < T, \quad m = 1 \to \infty. \tag{19-22}$$

i.e., the desired uncorrelated condition in (19-18) gets translated into the integral equation in (19-22) and it is known as the *Karhunen-Loeve* or K-L. integral equation. The functions $\{\varphi_k(t)\}_{k=1}^{\infty}$ are *not arbitrary* and they must be obtained by solving the integral equation in (19-22). They are known as the eigenvectors of the autocorrelation function of $R_{XX}(t_1, t_2)$. Similarly the set $\{\lambda_k\}_{k=1}^{\infty}$ represent the eigenvalues of the autocorrelation function. From (19-18), the eigenvalues λ_k represent the variances of the uncorrelated random variables c_k, $k = 1 \to \infty$. This also follows from Mercer's theorem which allows the representation

$$R_{XX}(t_1, t_2) = \sum_{k=1}^{\infty} \mu_k \phi_k(t_1) \phi_k^*(t_2), \quad 0 < t_1, t_2 < T, \qquad (19\text{-}23)$$

where

$$\int_0^T \phi_k(t) \phi_m^*(t) dt = \delta_{k,m}.$$

Here $\phi_k(t)$ and μ_k, $k = 1 \to \infty$ are known as the eigenfunctions and eigenvalues of $R_{XX}(t_1, t_2)$ respectively. A direct substitution and simplification of (19-23) into (19-22) shows that

$$\varphi_k(t) = \phi_k(t), \quad \lambda_{\text{k}} = \mu_k, \quad k = 1 \to \infty. \qquad (19\text{-}24)$$

Returning back to (19-15), once again the partial sum

$$\tilde{X}_N(t) = \sum_{k=1}^{N} c_k \varphi_k(t) \xrightarrow[N \to \infty]{} X(t), \quad 0 \le t \le T \qquad (19\text{-}25)$$

in the mean square sense. To see this, consider

$$E[|X(t)-\tilde{X}_N(t)|^2] = E[|X(t)|^2] - E[X(t)\tilde{X}_N^*(t)] - E[X^*(t)\tilde{X}_N(t)] + E[|\tilde{X}_N(t)|^2]. \tag{19-26}$$

We have

$$E[|X(t)|^2] = R(t,t). \tag{19-27}$$

Also

$$\begin{aligned} E[X(t)\tilde{X}_N^*(t)] &= \sum_{k=1}^{N} X(t) c_k^* \varphi_k^*(t) \\ &= \sum_{k=1}^{N} \int_0^T E[X(t)X^*(\alpha)]\, \varphi_k^*(t)\varphi_k(\alpha)\,d\alpha \\ &= \sum_{k=1}^{N} \left(\int_0^T R(t,\alpha)\varphi_k(\alpha)\,d\alpha\right)\varphi_k^*(t) \\ &= \sum_{k=1}^{N} \lambda_k \varphi_k(t)\varphi_k^*(t) = \sum_{k=1}^{N} \lambda_k\, |\varphi_k(t)|^2. \end{aligned} \tag{19-28}$$

Similarly

$$E[X^*(t)\tilde{X}_N(t)] = \sum_{k=1}^{N} \lambda_k\, |\varphi_k(t)|^2 \tag{19-29}$$

and

$$E[|\tilde{X}_N(t)|^2] = \sum_k \sum_m E[c_k c_m^*]\varphi_k(t)\varphi_m^*(t) = \sum_{k=1}^{N} \lambda_k\, |\varphi_k(t)|^2. \qquad (19\text{-}30)$$

Hence (19-26) simplifies into

$$E[|X(t) - \widetilde{X}_N(t)|^2] = R(t,t) - \sum_{k=1}^{N} \lambda_k\, |\varphi_k(t)|^2 \to 0 \ \text{ as } \to \infty. \qquad (19\text{-}31)$$

i.e.,

$$X(t) \doteq \sum_{k=1}^{\infty} c_k \varphi_k(t), \qquad 0 \le t \le T, \qquad (19\text{-}32)$$

where the random variables $\{c_k\}_{k=1}^{\infty}$ are uncorrelated and faithfully represent the random process $X(t)$ in $0 \le t \le T$, provided $\varphi_k(t)$, $k = 1 \to \infty$, satisfy the K-L. integral equation.

Example 19.1: If $X(t)$ is a w.s.s white noise process, determine the sets $\{\varphi_k, \lambda_k\}_{k=1}^{\infty}$ in (19-22).

Solution: Here

$$R_{XX}(t_1, t_2) = q\delta(t_1 - t_2) \qquad (19\text{-}33)$$

and

$$\int_0^T R_{XX}(t_1, t_2)\varphi_k(t_2)\,dt_1 = q\int_0^T \delta(t_1 - t_2)\varphi_k(t_2)\,dt_1$$
$$= q\,\varphi_k(t_1) \stackrel{\Delta}{=} \lambda_k\,\varphi_k(t_1) \qquad (19\text{-}34)$$

$\Rightarrow \varphi_k(t)$ can be arbitrary so long as they are orthonormal as in (19-17) and $\lambda_k = q, \quad k = 1 \to \infty.$ Then the power of the process

$$P = E[|X(t)|^2] = R(0) = \sum_{k=1}^{\infty} \lambda_k = \sum_{k=1}^{\infty} q = \infty$$

and in that sense white noise processes are *unrealizable*. However, if the received waveform is given by

$$r(t) = s(t) + n(t), \qquad 0 < t < T \qquad (19\text{-}35)$$

and $n(t)$ is a w.s.s white noise process, then since *any* set of orthonormal functions is sufficient for the white noise process representation, they can be chosen solely by considering the other signal $s(t)$. Thus, in (19-35)

$$R_{rr}(t_1 - t_2) = R_{ss}(t_1 - t_2) + q\delta(t_1 - t_2) \qquad (19\text{-}36)$$

and if

$$R_{ss}(t_1 - t_2) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t_1)\phi_k^*(t_2) \tag{19-37}$$

Then it follows that

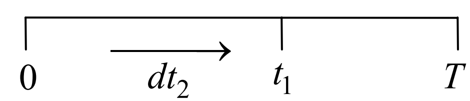
$$R_{rr}(t_1 - t_2) = \sum_{k=1}^{\infty} (\lambda_k + q)\phi_k(t_1)\phi_k^*(t_2). \tag{19-38}$$

Notice that the eigenvalues of $R_{ss}(t_1 - t_2)$ get incremented by q.

Example19.2: $X(t)$ is a Wiener process with

$$R_{XX}(t_1, t_2) = \alpha \min(t_1, t_2) = \begin{cases} \alpha t_2 & t_1 > t_2 \\ \alpha t_1 & t_1 \le t_2 \end{cases}, \qquad \alpha > 0 \tag{19-39}$$

In that case Eq. (19-22) simplifies to



$$\int_0^T R_{XX}(t_1, t_2)\varphi_k(t_2)dt_2 = \int_0^{t_1} R_{XX}(t_1, t_2)\varphi_k(t_2)dt_2 + \int_{t_1}^T R_{XX}(t_1, t_2)\varphi_k(t_2)dt_2 = \lambda_k \varphi_k(t_1),$$

and using (19-39) this simplifies to

$$\int_0^{t_1} \alpha t_2 \, \varphi_k(t_2)dt_2 + \int_{t_1}^T \alpha t_1 \, \varphi_k(t_2)dt_2 = \lambda_k \varphi_k(t_1). \tag{19-40}$$

Derivative with respect to t_1 gives [see Eqs. (8-5)-(8-6), Lecture 8]

$$\alpha\, t_1 \varphi_k(t_1) + (-1)\alpha\, t_1 \varphi_k(t_1) + \alpha \int_{t_1}^{T} \varphi_k(t_2)\,dt_2 = \lambda_k \dot{\varphi}_k(t_1)$$

or

$$\alpha \int_{t_1}^{T} \varphi_k(t_2)\,dt_2 = \lambda_k \dot{\varphi}_k(t_1). \qquad (19\text{-}41)$$

Once again, taking derivative with respect to t_1, we obtain

$$\alpha(-1)\varphi_k(t_1) = \lambda_k \ddot{\varphi}_k(t_1)$$

or

$$\frac{d^2 \varphi_k(t_1)}{dt_1^2} + \frac{\alpha}{\lambda_k}\varphi_k(t_1) = 0, \qquad (19\text{-}42)$$

and its solution is given by

$$\varphi_k(t) = A_k \cos\sqrt{\tfrac{\alpha}{\lambda_k}}\,t + B_k \sin\sqrt{\tfrac{\alpha}{\lambda_k}}\,t.$$

But from (19-40)

$$\varphi_k(0) = 0, \qquad (19\text{-}43)$$

and from (19-41)

$$\dot{\varphi}_k(T) = 0. \tag{19-44}$$

This gives

$$\begin{aligned} \varphi_k(0) &= A_k = 0, \qquad k = 1 \to \infty, \\ \dot{\varphi}_k(t) &= B_k \sqrt{\tfrac{\alpha}{\lambda_k}} \cos \sqrt{\tfrac{\alpha}{\lambda_k}}\, t, \end{aligned} \tag{19-45}$$

and using (19-44) we obtain

$$\dot{\varphi}_k(T) = B_k \sqrt{\tfrac{\alpha}{\lambda_k}} \cos \sqrt{\tfrac{\alpha}{\lambda_k}}\, T = 0$$

$$\Rightarrow \sqrt{\tfrac{\alpha}{\lambda_k}}\; T = (2k-1)\frac{\pi}{2} \tag{19-46}$$

$$\Rightarrow \quad \lambda_k = \frac{\alpha T^2}{(k-\frac{1}{2})^2 \pi^2}, \qquad k = 1 \to \infty. \tag{19-47}$$

Also

$$\varphi_k(t) = B_k \sin\sqrt{\tfrac{\alpha}{\lambda_k}}\ t, \qquad 0 \le t \le T. \tag{19-48}$$

Further, orthonormalization gives

$$\int_0^T \varphi_k^2(t)dt = B_k^2 \int_0^T \left(\sin\sqrt{\tfrac{\alpha}{\lambda_k}}\ t\right)^2 dt = B_k^2\left[\int_0^T \left(\frac{1-\cos 2\sqrt{\frac{\alpha}{\lambda_k}}\ t}{2}\right)dt\right]$$

$$= B_k^2\left(\frac{T}{2} - \frac{1}{2}\,\frac{\sin 2\sqrt{\frac{\alpha}{\lambda_k}}\ t}{2\sqrt{\frac{\alpha}{\lambda_k}}}\Bigg|_0^T\right) = B_k^2\left(\frac{T}{2} - \frac{\sin(2k-1)\pi - 0}{4\sqrt{\frac{\alpha}{\lambda_k}}}\right) = B_k^2\ \frac{T}{2} = 1$$

$$\Rightarrow B_k = \sqrt{2/T}.$$

Hence

$$\varphi_k(t) = \sqrt{\tfrac{2}{T}}\ \sin\left(\sqrt{\tfrac{\alpha}{\lambda_k}}\ t\right) = \sqrt{\tfrac{2}{T}}\ \sin\left(k - \tfrac{1}{2}\right)\frac{\pi t}{T}, \tag{19-49}$$

with λ_k as in (19-47) and c_k as in (19-16),

$X(t) = \sum_{k=1}^{\infty} c_k \varphi_k(t)$ is the desired series representation.

Example 19.3: Given

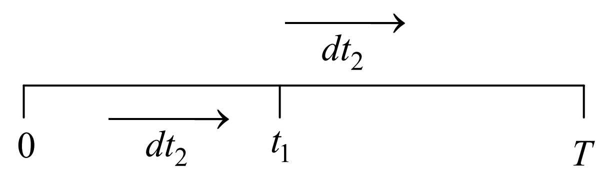
$$R_{XX}(\tau) = e^{-\alpha|\tau|}, \qquad \alpha > 0, \tag{19-50}$$

find the orthonormal functions for the series representation of the underlying stochastic process $X(t)$ in $0 < t < T$.

Solution: We need to solve the equation

$$\int_0^T e^{-\alpha|t_1 - t_2|} \varphi_n(t_2) dt_2 = \lambda_n \varphi_n(t_1). \tag{19-51}$$

Notice that (19-51) can be rewritten as,



$$\int_0^{t_1} e^{-\alpha \overbrace{(t_1 - t_2)}^{\geq 0}} \varphi_n(t_2) dt_2 + \int_{t_1}^{T} e^{-\alpha \overbrace{(t_2 - t_1)}^{\geq 0}} \varphi_n(t_2) dt_2 = \lambda_n \varphi_n(t_1) \tag{19-52}$$

Differentiating (19-52) once with respect to t_1, we obtain

$$\varphi_n(t_1)+\int_0^{t_1}(-\alpha)e^{-\alpha(t_1-t_2)}\varphi_n(t_2)dt_2-\varphi_n(t_1)+\int_{t_1}^{T}\alpha e^{-\alpha(t_2-t_1)}\varphi_n(t_2)dt_2$$

$$=\lambda_n\frac{d\varphi_n(t_1)}{dt_1}$$

$$\Rightarrow\ -\int_0^{t_1}e^{-\alpha(t_1-t_2)}\varphi_n(t_2)dt_2+\int_{t_1}^{T}e^{-\alpha(t_2-t_1)}\varphi_n(t_2)dt_2=\frac{\lambda_n}{\alpha}\frac{d\varphi_n(t_1)}{dt_1} \qquad (19\text{-}53)$$

Differentiating (19-53) again with respect to t_1, we get

$$-\varphi_n(t_1)-\int_0^{t_1}(-\alpha)e^{-\alpha(t_1-t_2)}\varphi_n(t_2)dt_2$$

$$-\varphi_n(t_1)+\int_{t_1}^{T}\alpha e^{-\alpha(t_2-t_1)}\varphi_n(t_2)dt_2=\frac{\lambda_n}{\alpha}\frac{d^2\varphi_n(t_1)}{dt_1^2}$$

or

$$-2\varphi_n(t_1)+\alpha\underbrace{\left[\int_0^{t_1}e^{-\alpha(t_1-t_2)}\varphi_n(t_2)dt_2+\int_{t_1}^{T}e^{-\alpha(t_2-t_1)}\varphi_n(t_2)dt_2\right]}_{\lambda_n\varphi_n(t_1)\quad \{\text{use (19-52)}\}}$$

$$=\frac{\lambda_n}{\alpha}\frac{d^2\varphi_n(t_1)}{dt_1^2}$$

or

$$(\alpha\lambda_n - 2)\varphi_n(t_1) = \frac{\lambda_n}{\alpha}\frac{d^2\varphi_n(t_1)}{dt_1^2}$$

or

$$\frac{d^2\varphi_n(t_1)}{dt_1^2} = \left(\frac{\alpha(\alpha\lambda_n - 2)}{\lambda_n}\right)\varphi_n(t_1). \tag{19-54}$$

Eq.(19-54) represents a second order differential equation. The solution for $\varphi_n(t)$ depends on the value of the constant $\alpha(\alpha\lambda_n - 2)/\lambda_n$ on the right side. We shall show that solutions exist in this case only if $\alpha\lambda_n < 2$, or

$$0 < \lambda_n < \frac{2}{\alpha}. \tag{19-55}$$

In that case $\alpha(\alpha\lambda_n - 2)/\lambda_n < 0.$

Let

$$\omega_n^2 \triangleq \frac{\alpha(2 - \alpha\lambda_n)}{\lambda_n} > 0, \tag{19-56}$$

and (19-54) simplifies to

$$\frac{d^2\varphi_n(t_1)}{dt_1^2} = -\omega_n^2\,\varphi_n(t_1). \tag{19-57}$$

General solution of (19-57) is given by

$$\varphi_n(t_1) = A_n \cos\omega_n t_1 + B_n \sin\omega_n t_1. \tag{19-58}$$

From (19-52)

$$\varphi_n(0) = \frac{1}{\lambda_n}\int_0^T e^{-\alpha t_2}\varphi_n(t_2)dt_2 \tag{19-59}$$

and

$$\varphi_n(T) = \frac{1}{\lambda_n}\int_0^T e^{-\alpha t_2}\varphi_n(T-t_2)dt_2. \tag{19-60}$$

Similarly from (19-53)

$$\dot{\varphi}_n(0) = \left.\frac{d\varphi_n(t_1)}{dt_1}\right|_{t_1=0} = \frac{\alpha}{\lambda_n}\int_0^T e^{-\alpha t_2}\varphi_n(t_2)dt_2 = \alpha\varphi_n(0) \tag{19-61}$$

and

$$\dot{\varphi}_n(T) = -\frac{\alpha}{\lambda_n}\int_0^T e^{-\alpha t_2}\varphi_n(T-t_2)dt_2 = -\alpha\varphi_n(T). \tag{19-62}$$

Using (19-58) in (19-61) gives

$$B_n\omega_n = \alpha A_n$$

or

$$\frac{A_n}{B_n} = \frac{\omega_n}{\alpha}, \tag{19-63}$$

and using (19-58) in (19-62), we have

$$-A_n\omega_n \sin\omega_n T + B_n\omega_n \cos\omega_n T = -\alpha(A_n \cos\omega_n T + B_n \sin\omega_n T),$$

$$\Rightarrow \quad (A_n\alpha + B_n\omega_n)\cos\omega_n T = (A_n\omega_n - B_n\alpha)\sin\omega_n T$$

or

$$\tan\omega_n T = \frac{2A_n\alpha}{A_n\omega_n - B_n\alpha} = \frac{2A_n\alpha / B_n\alpha}{\frac{A_n\omega_n}{B_n\alpha} - 1} = \frac{2A_n / B_n}{\frac{A_n}{B_n}\frac{\omega_n}{\alpha} - 1} = \frac{2(\omega_n/\alpha)}{(\frac{\omega_n}{\alpha})^2 - 1}.$$

Thus ω_n s are obtained as the solution of the transcendental equation

$$\tan\omega_n T = \frac{2(\omega_n/\alpha)}{(\omega_n/\alpha)^2 - 1}, \tag{19-64}$$

which simplifies to

$$\tan(\omega_n T/2) = -\frac{\omega_n}{\alpha}. \tag{19-65}$$

In terms of ω_n s from (19-56) we get

$$\lambda_n = \frac{2\alpha}{\alpha^2 + \omega_n^2} > 0. \tag{19-66}$$

Thus the eigenvalues are obtained as the solution of the transcendental equation (19-65). (see Fig 19.1). For each such λ_n (or ω_n^2), the corresponding eigenvector is given by (19-58). Thus

$$\begin{aligned} \varphi_n(t) &= A_n \cos\omega_n t + B_n \sin\omega_n t \\ &= c_n \sin(\omega_n t - \theta_n) = c_n \sin\omega_n (t - \tfrac{T}{2}), \qquad 0 < t < T \end{aligned} \tag{19-67}$$

since from (19-65)

$$\theta_n = \tan^{-1}\left(-\frac{A_n}{B_n}\right) = \tan^{-1}\left(-\frac{\omega_n}{\alpha}\right) = \omega_n T/2, \tag{19-68}$$

and c_n is a suitable normalization constant.

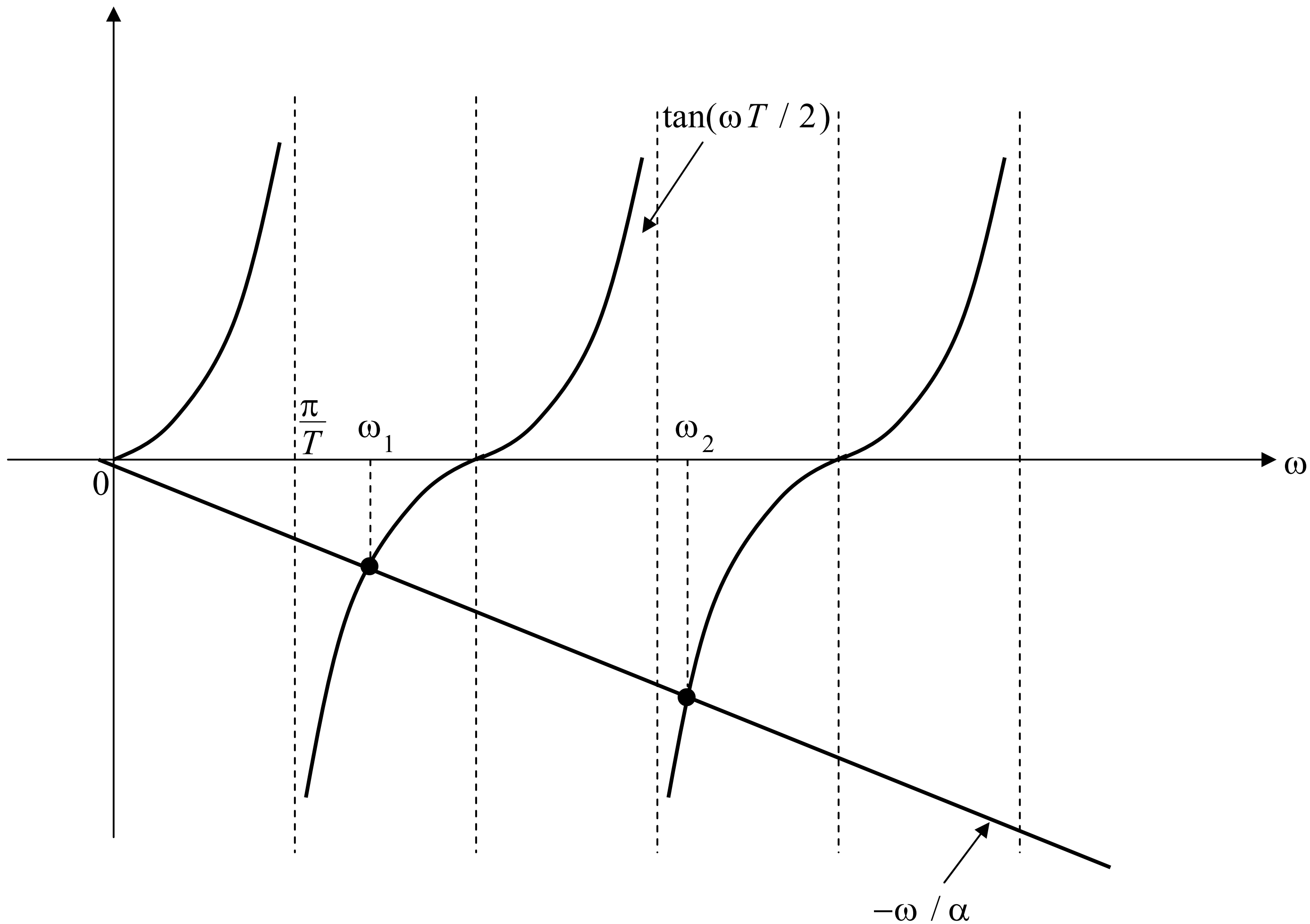


Fig 19.1 Solution for Eq.(19-65).

Karhunen – Loeve Expansion for Rational Spectra

[The following exposition is based on Youla's classic paper "The solution of a Homogeneous Wiener-Hopf Integral Equation occurring in the expansion of Second-order Stationary Random Functions," IRE Trans. on Information Theory, vol. 9, 1957, pp 187-193. Youla is tough. Here is a friendlier version. Even this may be skipped on a first reading. (Youla can be made only so much friendly.)]

Let $X(t)$ represent a w.s.s zero mean real stochastic process with autocorrelation function $R_{XX}(\tau) = R_{XX}(-\tau)$ so that its power spectrum

$$S_{XX}(\omega) = \int_{-\infty}^{+\infty} R_{XX}(\tau) e^{-j\omega t} dt = 2\int_0^{\infty} R_{XX}(\tau) \cos\tau \, d\tau \qquad (19\text{-}69)$$

is nonnegative and an even function. If $S_{XX}(\omega)$ is rational, then the process $X(t)$ is said to be rational as well. $S_{XX}(\omega)$ rational and even implies

$$S_{XX}(\omega) = \frac{N(\omega^2)}{D(\omega^2)} \geq 0. \qquad (19\text{-}70)$$

The total power of the process is given by

$$P = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{XX}(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{N(\omega^2)}{D(\omega^2)}d\omega \tag{19-71}$$

and for P to be finite, we must have

(i) The degree $\delta(D) = 2n$ of the denominator polynomial $D(\omega^2)$ must exceed the degree $\delta(N) = 2m$ of the numerator polynomial $N(\omega^2)$ by at least two,

and

(ii) $D(\omega^2)$ must not have any zeros on the real-frequency $(s = j\omega)$ axis.

The s-plane $(s = \sigma + j\omega)$ extension of $S_{XX}(\omega)$ is given by

$$S_{XX}(\omega)\,|_{s=j\omega} \triangleq S(s^2) = \frac{N(-s^2)}{D(-s^2)}. \tag{19-72}$$

Thus

$$D(-s^2) = \prod_k (s^2 - \mu_k^2)^{k_i} \tag{19-73}$$

and the Laplace inverse transform of $(s^2 - \alpha^2)^{-k}$ is given by

$$\frac{1}{(s^2-\alpha^2)^k} \quad \leftrightarrow \quad \frac{(-1)^k}{(k-1)!}e^{-\alpha|\tau|}\sum_{j=1}^{k}\frac{(k+j-2)!}{(j-1)!(k-j)!}\,\frac{|\tau\;|^{k-j}}{(2\alpha)^{k+j-1}} \qquad (19\text{-}74)$$

Let $\pm\mu_1,\ \pm\mu_2,\cdots,\ \pm\mu_n$ represent the roots of $D(-s^2)$. Then

$$0<\operatorname{Re}\mu_1\le\operatorname{Re}\mu_2\le\cdots\le\operatorname{Re}\mu_n \qquad (19\text{-}75)$$

Let $D^+(s)$ and $D^-(s)$ represent the left half plane (LHP) and the right half plane (RHP) products of these roots respectively. Thus

$$D(-s^2)=D^+(s)D^-(s), \qquad (19\text{-}76)$$

where

$$D^+(s)=\prod_k(s+\mu_k)(s+\mu_k^*)=\sum_{k=0}^{n}d_k s^k=D^-(-s). \qquad (19\text{-}77)$$

This gives

$$S(s^2)=\frac{N(-s^2)}{D(-s^2)}=\frac{C_1(s)}{D^+(s)}+\frac{C_2(s)}{D^-(s)} \qquad (19\text{-}78)$$

Notice that $\frac{C_1(s)}{D^+(s)}$ has poles only on the LHP and its inverse (for all $t>0$) converges only if the strip of convergence is to the right of *all* its poles. Similarly $C_2(s)/D^-(s)$ has poles only on the RHP and its inverse will converge only if the strip is to the left of all those poles. In that case, the inverse exists for $t < 0$. In the case of $R_{XX}(\tau)$, from (19-78) its transform $N(s^2)/D(-s^2)$ is defined only for $-\operatorname{Re}\mu_1 < \operatorname{Re} s < \operatorname{Re}\mu_1$ (see Fig 19.2). In particular, for $\tau > 0$, from the above discussion it follows that $R_{XX}(\tau)$ is given by the inverse transform of $C_1(s)/D^+(s)$. We need the solution to the integral equation

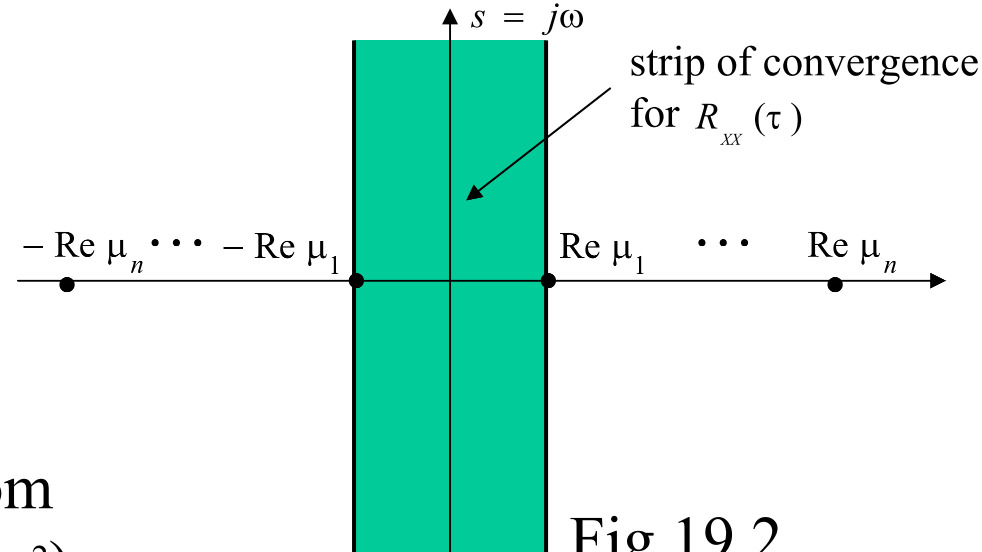


Fig 19.2

$$\varphi(t) = \lambda \int_0^T R_{XX}(t-\tau)\varphi(\tau)\,d\tau, \qquad 0 < t < T \tag{19-79}$$

that is valid *only* for $0 < t < T$. (Notice that λ in (19-79) is the reciprocal of the eigenvalues in (19-22)). On the other hand, the right side (19-79) can be defined for every t. Thus, let

$$g(t) \triangleq \int_0^T R_{XX}(t-\tau)\varphi(\tau)\,d\tau, \qquad -\infty < t < +\infty \tag{19-80}$$

and to confirm with the same limits, define

$$\phi(t) = \begin{cases} \varphi(t) & 0 < t < T \\ 0 & \text{otherwise} \end{cases}. \tag{19-81}$$

This gives

$$g(t) = \int_{-\infty}^{+\infty} R_{XX}(t-\tau)\phi(\tau)\,d\tau \tag{19-82}$$

and let

$$f(t) = \phi(t) - \lambda g(t) = \phi(t) - \lambda \int_{-\infty}^{+\infty} R_{XX}(t-\tau)\phi(\tau)\,d\tau. \tag{19-83}$$

Clearly

$$f(t) = 0, \qquad 0 < t < T \tag{19-84}$$

and for $t > T$

$$D^{+}\left(\tfrac{d}{dt}\right) f(t) = -\lambda \int_{-\infty}^{+\infty} \{\overbrace{D^{+}\left(\tfrac{d}{dt}\right) R_{XX}(t-\tau)}^{D^{+}(-\mu_k)=0}\}\phi(\tau)\,d\tau = 0, \tag{19-85}$$

since $R_{XX}(t)$ is a sum of exponentials $\sum_k a_k e^{-\mu_k t}$, for $t > 0$. Hence it follows that for $t > T$, the function $f(t)$ must be a sum of exponentials $\sum_k a_k e^{-\mu_k t}$. Similarly for $t < 0$

$$D^-\left(\frac{d}{dt}\right)f(t) = -\lambda\int_{-\infty}^{+\infty}\overbrace{\{D^-\left(\frac{d}{dt}\right)R_{XX}(t-\tau)\}}^{D^-(\mu_k)=0}\phi(\tau)d\tau = 0,$$

and hence $f(t)$ must be a sum of exponentials $\sum_k b_k e^{\mu_k t}$, for $t<0$. Thus the overall Laplace transform of $f(t)$ has the form

$$F(s) = \underbrace{\frac{P(s)}{D^-(s)}}_{\substack{\text{contributions} \\ \text{in } t<0}} \underbrace{- e^{-sT}\frac{Q(s)}{\underbrace{D^+(s)}_{\text{contributes to } t>0}}}_{\text{contributions in } t>T} \tag{19-86}$$

where $P(s)$ and $Q(s)$ are polynomials of degree $n-1$ at most. Also from (19-83), the bilateral Laplace transform of $f(t)$ is given by

$$F(s) = \Phi(s)\left[1-\lambda\frac{N(-s^2)}{D(-s^2)}\right], \qquad -\operatorname{Re}\mu_1 < \operatorname{Re}s < \operatorname{Re}\mu_1 \tag{19-87}$$

Equating (19-86) and (19-87) and simplifying, Youla obtains the key identity

$$\Phi(s) = \frac{P(s)D^+(s) - e^{-sT}Q(s)D^-(s)}{D(-s^2)-\lambda N(-s^2)}. \tag{19-88}$$

Youla argues as follows: The function $\Phi(s) = \int_0^T \phi(t)e^{-st}dt$ is an entire function of s, and hence it is free of *poles* on the *entire* finite s-plane $(-\infty < \operatorname{Re} s < +\infty)$. However, the denominator on the right side of (19-88) is a polynomial and its roots contribute to poles of $\Phi(s)$. Hence all such poles must be cancelled by the numerator. As a result the numerator of $\Phi(s)$ in (19-88) must possess exactly the *same set* of zeros as its denominator to the respective order at least.

Let $\pm\omega_1(\lambda),\ \pm\omega_2(\lambda),\cdots,\ \pm\omega_n(\lambda)$ be the (distinct) zeros of the denominator polynomial $D(-s^2)-\lambda N(-s^2)$. Here we assume that λ is an eigenvalue for which all ω_k's are distinct. We have

$$0 < \operatorname{Re}\omega_1(\lambda) < \operatorname{Re}\omega_2(\lambda) < \cdots < \operatorname{Re}\omega_n(\lambda) < \infty. \tag{19-89}$$

These ω_k's also represent the zeros of the numerator polynomial $P(s)D^+(s) - e^{-sT}Q(s)D^-(s)$. Hence

$$D^+(\omega_k)P(\omega_k) = e^{-\omega_k T} D^-(\omega_k)Q(\omega_k) \tag{19-90}$$

and

$$D^+(-\omega_k)P(-\omega_k) = e^{\omega_k T} D^-(-\omega_k)Q(-\omega_k) \tag{19-91}$$

which simplifies into

$$D^-(\omega_k)P(-\omega_k) = e^{\omega_k T} D^+(\omega_k)Q(-\omega_k). \tag{19-92}$$

From (19-90) and (19-92) we get

$$P(\omega_k)P(-\omega_k) = Q(\omega_k)Q(-\omega_k), \qquad k = 1,\ 2,\ \cdots,\ n \tag{19-93}$$

i.e., the polynomial

$$L(s) = P(s)P(-s) - Q(s)Q(-s) \tag{19-94}$$

which is at most of degree $n - 1$ in s^2 vanishes at $\omega_1^2,\ \omega_2^2, \cdots,\ \omega_n^2$ (for n distinct values of s^2). Hence

$$L(s^2) \equiv 0 \tag{19-95}$$

or

$$P(s)P(-s) = Q(s)Q(-s). \tag{19-96}$$

Using the linear relationship among the coefficients of $P(s)$ and $Q(s)$ in (19-90)-(19-91) it follows that

$$P(s) = \pm Q(s) \quad \text{or} \quad P(s) = \pm Q(-s) \tag{19-97}$$

are the only solutions that are consistent with each of those equations, and together we obtain

$$P(s) = \pm Q(-s) \tag{19-98}$$

as the *only solution* satisfying both (19-90) and (19-91). Let

$$P(s) = \sum_{i=0}^{n-1} p_i \, s^i . \tag{19-99}$$

In that case (19-90)-(19-91) simplify to (use (19-98))

$$\begin{aligned} & P(\omega_k) D^+(\omega_k) \mp e^{-\omega_k T} D^-(\omega_k) P(-\omega_k) \\ & = \sum_{i=0}^{n-1} \{1 \mp (-1)^i a_k\} \omega_k^i \, p_i = 0, \qquad k = 1, 2, \cdots, n \end{aligned} \tag{19-100}$$

where

$$a_k = \frac{D^-(\omega_k)}{D^+(\omega_k)} e^{-\omega_k T} = \frac{D^+(-\omega_k)}{D^+(\omega_k)} e^{-\omega_k T} . \tag{19-101}$$

For a nontrivial solution to exist for $p_0, p_1, \cdots, p_{n-1}$ in (19-100), we must have

$$\Delta_{1,2} = \begin{vmatrix} (1 \mp a_1) & (1 \pm a_1)\omega_1 & \cdots & (1 \mp (-1)^{n-1} a_1)\omega_1^{n-1} \\ (1 \mp a_2) & (1 \pm a_2)\omega_2 & \cdots & (1 \mp (-1)^{n-1} a_2)\omega_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ (1 \mp a_n) & (1 \pm a_n)\omega_n & \cdots & (1 \mp (-1)^{n-1} a_n)\omega_n^{n-1} \end{vmatrix} = 0. \quad (19\text{-}102)$$

The two determinant conditions in (19-102) must be solved together to obtain the eigenvalues λ_i's that are implicitly contained in the a_i's and ω_i's (Easily said than done!).

To further simplify (19-102), one may express a_k in (19-101) as

$$a_k = e^{-2\theta_k}, \qquad k = 1,\ 2,\ \cdots\ ,n \qquad (19\text{-}103)$$

so that

$$\tanh \theta_k = \frac{e^{\theta_k} - e^{-\theta_k}}{e^{\theta_k} + e^{-\theta_k}} = \frac{1-a_k}{1+a_k} = \frac{D^+(\omega_k) - e^{-\omega_k T} D^+(-\omega_k)}{D^+(\omega_k) + e^{-\omega_k T} D^+(-\omega_k)}$$

$$= \frac{e^{\omega_k T/2} D^+(\omega_k) - e^{-\omega_k T/2} D^+(-\omega_k)}{e^{\omega_k T/2} D^+(\omega_k) + e^{-\omega_k T/2} D^+(-\omega_k)} \qquad (19\text{-}104)$$

Let

$$D^+(s) = d_0 + d_1 s + \cdots + d_n s^n \tag{19-105}$$

and substituting these known coefficients into (19-104) and simplifying we get

$$\tanh\, \theta_k = \frac{(d_0 + d_2\omega_k^2 + \cdots)\tanh\,(\omega_k T/2) + (d_1\omega_k + d_3\omega_k^3 + \cdots)}{(d_0 + d_2\omega_k^2 + \cdots) + (d_1\omega_k + d_3\omega_k^3 + \cdots)\tanh\,(\omega_k T/2)} \tag{19-106}$$

and in terms of $\tanh\theta_k$, Δ_2 in (19-102) simplifies to

$$\begin{vmatrix} 1 & \omega_1 \tanh\theta_1 & \omega_1^2 & \omega_1^3 \tanh\theta_1 & \cdots & \omega_1^{n-1} \tanh\theta_1 \\ 1 & \omega_2 \tanh\theta_2 & \omega_2^2 & \omega_2^3 \tanh\theta_2 & \cdots & \omega_2^{n-1} \tanh\theta_2 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega_n \tanh\theta_n & \omega_n^2 & \omega_n^3 \tanh\theta_n & \cdots & \omega_n^{n-1} \tanh\theta_n \end{vmatrix} = 0 \tag{19-107}$$

if n is even (if n is odd the last column in (19-107) is simply $[\omega_1^{n-1},\ \omega_2^{n-1}, \cdots,\ \omega_n^{n-1}]^T$). Similarly Δ_1 in (19-102) can be obtained by replacing $\tanh\theta_k$ with $\coth\theta_k$ in (19-107).

To summarize determine the roots $\omega_k\,'s$ with $\mathrm{Re}(\omega_i) > 0$ that satisfy

$$D(-\omega_k^2) - \lambda N(-\omega_k^2) = 0, \qquad k = 1,\ 2,\ \cdots\ ,n \tag{19-108}$$

in terms of λ, and for every such ω_k, determine θ_k using (19-106). Finally using these $\omega_k\ s$ and $\tanh\ \theta_k\ s$ in (19-107) and its companion equation Δ_1, the eigenvalues $\lambda_k\ s$ are determined. Once $\lambda_k\ s$ are obtained, $p_k\ s$ can be solved using (19-100), and using that $\Phi_i(s)$ can be obtained from (19-88).

Thus

$$\Phi_i(s) = \frac{D^+(s)P(s,\lambda_i) - e^{-sT}D^-(s)Q(s,\lambda_i)}{D(-s^2) - \lambda_i N(-s^2)} \tag{19-109}$$

and

$$\phi_i(t) = L^{-1}\{\Phi_i(s)\}. \tag{19-110}$$

Since $\Phi_i(s)$ is an entire function in (19-110), the inverse Laplace transform in (19-109) can be performed through *any* strip of convergence in the s-plane, and in particular if we use the strip

$\operatorname{Re} s > \operatorname{Re}(\omega_n)$ (to the right of all $\operatorname{Re}(\omega_i)$), then the two inverses

$$L^{-1}\left\{\frac{D^+(s)P(s)}{D(-s^2)-\lambda N(-s^2)}\right\}, \qquad L^{-1}\left\{\frac{D^-(s)Q(s)}{D(-s^2)-\lambda N(-s^2)}\right\} \qquad (19\text{-}111)$$

obtained from (19-109) will be causal. As a result $L^{-1}\left\{e^{-sT}\frac{D^-(s)Q(s)}{D(-s^2)-\lambda N(-s^2)}\right\}$ will be nonzero only for $t > T$ and using this in (19-109)-(19-110) we conclude that $\phi_i(t)$ for $0 < t < T$ has contributions only from the first term in (19-111). Together with (19-81), finally we obtain the desired eigenfunctions to be

$$\varphi_k(t) = L^{-1}\left\{\frac{D^+(s)P(s,\lambda_k)}{D(-s^2)-\lambda_k N(-s^2)}\right\}, \quad 0 < t < T, \qquad (19\text{-}112)$$

$$\operatorname{Re} s > \operatorname{Re}\omega_n > 0, \quad k = 1, 2, \cdots, n$$

that are orthogonal by design. Notice that in general (19-112) corresponds to a sum of modulated exponentials.

Next, we shall illustrate this procedure through some examples. First, we shall re-do Example 19.3 using the method described above.

Example 19.4: Given $R_{XX}(\tau) = e^{-\alpha|\tau|},$ we have

$$S_{XX}(\omega) = \frac{2\alpha}{\alpha^2 + \omega^2} = \frac{N(\omega^2)}{D(\omega^2)}.$$

This gives $D^+(s) = \alpha + s,\ D^-(s) = \alpha - s$ and $P(s)$, $Q(s)$ are constants here. Moreover since $n = 1$, (19-102) reduces to $1 \pm a_1 = 0,$ or $a_1 = \pm 1$ and from (19-101), ω_1 satisfies

$$e^{\omega_1 T} = \frac{D^-(\omega_1)}{D^+(\omega_1)} = \frac{\alpha - \omega_1}{\alpha + \omega_1} \tag{19-113}$$

or ω_1 is the solution of the s-plane equation

$$e^{sT} = \frac{\alpha - s}{\alpha + s} \tag{19-114}$$

But $|e^{sT}| > 1$ on the RHP, whereas $\left|\frac{\alpha - s}{\alpha + s}\right| < 1$ on the RHP. Similarly $|e^{sT}| < 1$ on the LHP, whereas $\left|\frac{\alpha - s}{\alpha + s}\right| > 1$ on the LHP.

Thus in (19-114) the solution s must be purely imaginary, and hence ω_1 in (19-113) is purely imaginary. Thus with $s = j\omega_1$ in (19-114) we get

$$e^{j\omega_1 T} = \frac{\alpha - j\omega_1}{\alpha + j\omega_1}$$

or

$$\tan(\omega_1 T/2) = -\frac{\omega_1}{\alpha} \tag{19-115}$$

which agrees with the transcendental equation (19-65). Further from (19-108), the λs satisfy

$$D(-s^2) - \lambda_n N(-s^2)\Big|_{s=j\omega_n} = \alpha^2 + \omega_n^2 - 2\alpha\lambda_n = 0$$

or

$$\lambda_n = \frac{\alpha^2 + \omega_n^2}{2\alpha} > 0. \tag{19-116}$$

Notice that the λ_n in (19-66) is the inverse of (19-116) because as noted earlier λ in (19-79) is the inverse of that in (19-22).

Finally from (19-112)

$$\varphi_n(t) = L^{-1}\left\{\frac{s+\alpha}{s^2+\omega_n^2}\right\} = A_n \cos\omega_n t + B_n \sin\omega_n t, \qquad 0 < t < T \qquad (19\text{-}117)$$

which agrees with the solution obtained in (19-67). We conclude this section with a less trivial example.

Example 19.5

$$R_{XX}(\tau) = e^{-\alpha|\tau|} + e^{-\beta|\tau|}. \qquad (19\text{-}118)$$

In this case

$$S_{XX}(\omega) = \frac{2\alpha}{\omega^2+\alpha^2} + \frac{2\beta}{\omega^2+\beta^2} = \frac{2(\alpha+\beta)(\omega^2+\alpha\beta)}{(\omega^2+\alpha^2)(\omega^2+\beta^2)}. \qquad (19\text{-}119)$$

This gives $D^+(s) = (s+\alpha)(s+\beta) = s^2 + (\alpha+\beta)s + \alpha\beta$. With $n = 2$, (19-107) and its companion determinant reduce to

$$\omega_2 \tanh\,\theta_2 = \omega_1 \tanh\,\theta_1$$

$$\omega_2 \coth\,\theta_2 = \omega_1 \coth\,\theta_1$$

or

$$\tanh\ \theta_1 = \pm \tanh\ \theta_2. \tag{19-120}$$

From (19-106)

$$\tanh\ \theta_i = \frac{(\alpha\beta + \omega_i^2)\tanh\ (\omega_i T/2) + (\alpha+\beta)\omega_i}{(\alpha\beta + \omega_i^2) + (\alpha+\beta)\omega_i \tanh\ (\omega_i T/2)}, \quad i = 1,2 \tag{19-121}$$

Finally ω_1^2 and ω_2^2 can be parametrically expressed in terms of λ using (19-108) and it simplifies to

$$\begin{aligned} D(-s^2) - \lambda N(-s^2) &= s^4 - (\alpha^2 + \beta^2 - 2\lambda(\alpha+\beta))s^2 \\ &\quad + \alpha^2\beta^2 - 2\lambda(\alpha+\beta)\alpha\beta \\ &\triangleq s^4 - bs^2 + c = 0. \end{aligned}$$

This gives

$$\omega_1^2 = \frac{b(\lambda) + \sqrt{b^2(\lambda) - 4c(\lambda)}}{2}$$

and

and

$$\omega_2^2 = \frac{b(\lambda) - \sqrt{b^2(\lambda) - 4c(\lambda)}}{2} = \omega_1^2 - \sqrt{b^2(\lambda) - 4c(\lambda)}$$

and substituting these into (19-120)-(19-121) the corresponding transcendental equation for $\lambda_i s$ can be obtained. Similarly the eigenfunctions can be obtained from (19-112).

20. Extinction Probability for Queues and Martingales

(Refer to section 15.6 in text (**Branching processes)** for discussion on the extinction probability).

20.1 Extinction Probability for Queues:

- A customer arrives at an empty server and immediately goes for service initiating a busy period. During that service period, other customers may arrive and if so they wait for service. The server continues to be busy till the last waiting customer completes service which indicates the end of a busy period. An interesting question is whether the busy periods are bound to terminate at some point ? Are they ?

Do busy periods continue forever? Or do such queues come to an end sooner or later? If so, how ?

• **Slow Traffic ($\rho \le 1$)**
Steady state solutions exist and the probability of extinction equals 1. (Busy periods are bound to terminate with probability 1. Follows from sec 15.6, theorem 15-9.)

• **Heavy Traffic ($\rho > 1$)**
Steady state solutions do not exist, and such queues can be characterized by their probability of extinction.

• Steady state solutions exist if the traffic rate $\rho < 1$. Thus

$$p_k = \lim_{n\to\infty} P\{X(nT) = k\} \text{ exists if } \rho < 1.$$

• What if too many customers rush in, and/or the service rate is slow ($\rho \ge 1$) ? How to characterize such queues ?

Extinction Probability (π_0) for Population Models

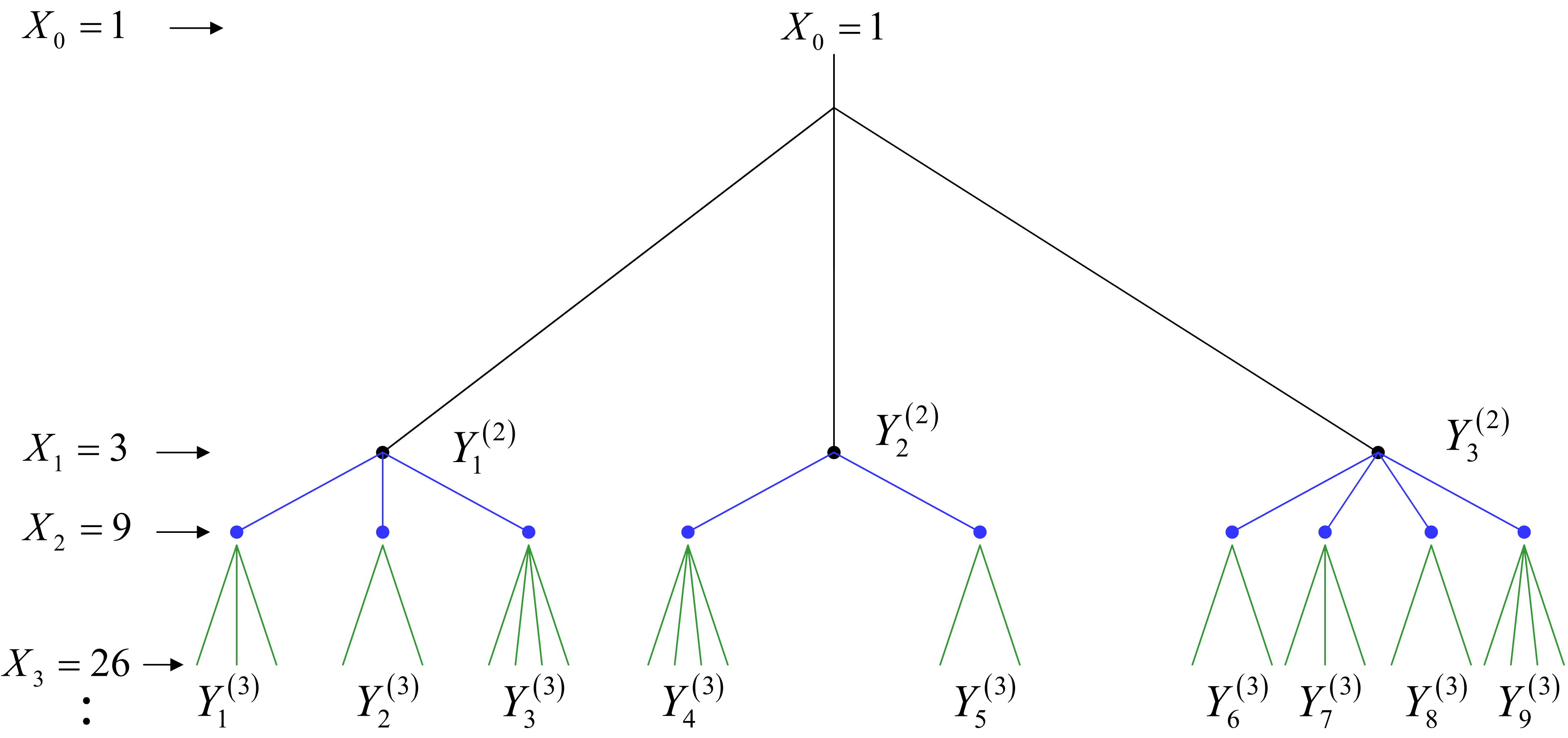


Fig 20.1

Queues and Population Models

- **Population models**

X_n : Size of the n$^{\text{th}}$ generation

$Y_i^{(n)}$: Number of offspring for the i^{th} member of the n$^{\text{th}}$ generation. From Eq.(15-287), Text

$$X_{n+1} = \sum_{k=1}^{X_n} Y_k^{(n)}$$

Let

$$a_k \triangleq P\{Y_i^{(n)} = k\}$$

- Offspring moment generating function:

$$P(z) = \sum_{k=0}^{\infty} a_k z^k \qquad (20\text{-}1)$$

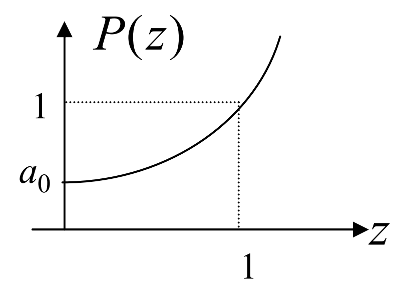


Fig 20.2

$$P_{n+1}(z) = \sum_{k=0}^{\infty} P\{X_{n+1} = k\}\, z^k = E\{z^{X_{n+1}}\}$$

$$= E(E\{z^{X_{n+1}} \mid X_n = j\}) = E\left(E\left\{ z^{\sum_{i=1}^{j} Y_i} \,\middle|\, X_n = j \right\} \right)$$

$$= E\{[P(z)]^j\} = \sum_j \{P(z)\}^j P\{X_n = j\} = P_n(P(z)) \qquad \text{(20-2)}$$

$$P_{n+1}(z) = P_n(P(z)) = P(P_n(z)) \qquad \text{(20-3)}$$

$$P\{X_n = k\} = ?$$

Extinction probability $= \lim_{n\to\infty} P\{X_n = 0\} = \pi_o = ?$

Extinction probability π_0 satisfies the equation $P(z) = z$ which can be solved iteratively as follows:

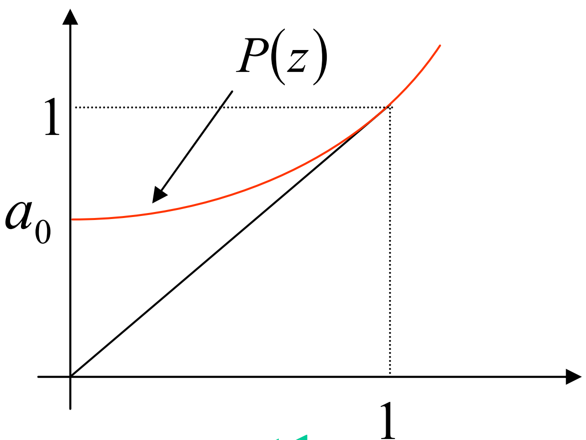
$$z_0 = P(0) \stackrel{\Delta}{=} a_0 \qquad \text{(20-4)}$$

and

$$z_k = P(z_{k-1}), \qquad k = 1, 2, \cdots \qquad \text{(20-5)}$$

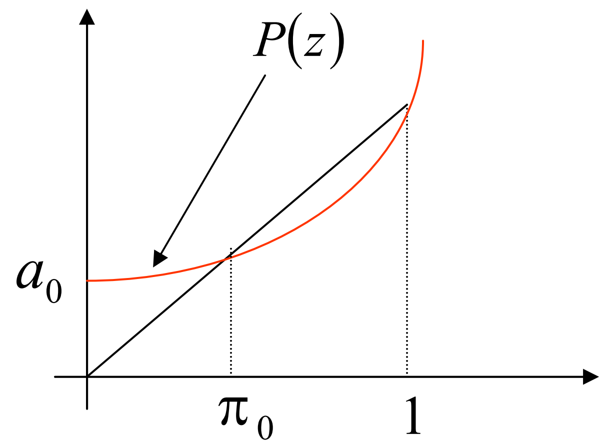
- Review Theorem 15-9 (Text)

Let $\quad \rho = E(Y_i) = P'(1) = \sum_{k=0}^{\infty} k \; P\{Y_i = k\} = \sum_{k=0}^{\infty} k a_k > 0 \qquad (20\text{-}6)$



(a) $\rho \le 1$

Fig 20.3



(b) $\rho > 1$

$$\left.\begin{aligned} &\rho \le 1 \Rightarrow \pi_0 = 1 \\ &\rho > 1 \Rightarrow \pi_0 \text{ is the unique solution of } P(z) = z\,, \\ &a_0 < \pi_0 < 1 \end{aligned}\right\} \qquad (20\text{-}7)$$

- Left to themselves, in the long run, populations either die out completely with probability π_0, or explode with probability $1\text{-}\pi_0$. (Both unpleasant conclusions).

Queues :

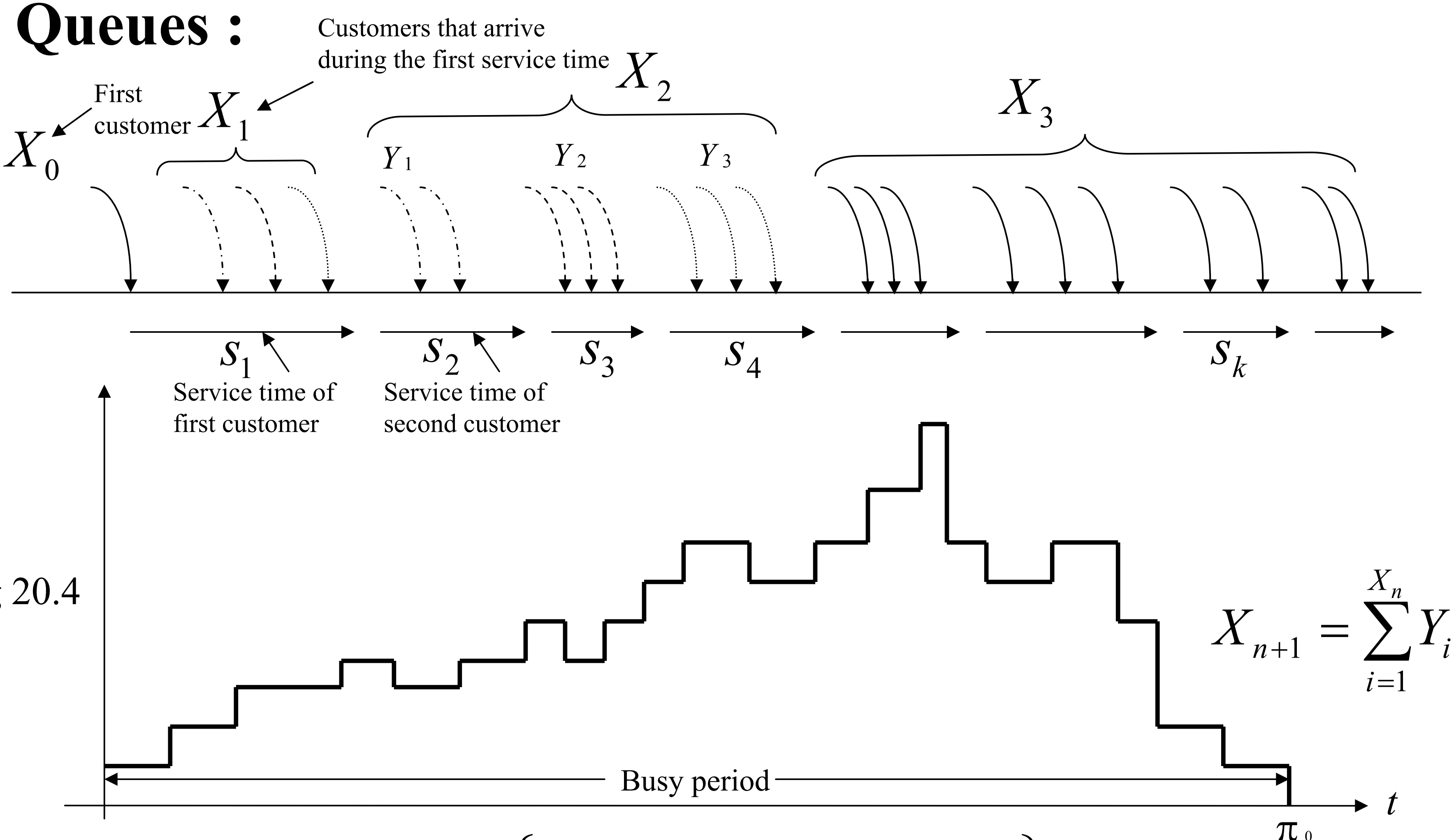


Fig 20.4

$$a_k = P(Y_i = k) = P\left\{\begin{array}{c} k \text{ arrivals between two} \\ \text{successive departures} \end{array}\right\} \geq 0.$$

Note that the statistics $\{a_k\}$ depends both on the arrival as well as the service phenomena.

- $P(z) = \sum_{k=0}^{\infty} a_k z^k$: Inter-departure statistics generated by arrivals
- $\rho = P'(1) = \sum_{k=1}^{\infty} k a_k$: Traffic Intensity $\left.\begin{matrix} \le 1 \\ > 1 \end{matrix}\right\} \Rightarrow$ $\begin{matrix}\text{Steady state} \\ \text{Heavy traffic}\end{matrix}$
- Termination of busy periods corresponds to extinction of queues. From the analogy with population models the extinction probability π_0 is the unique root of the equation $P(z) = z$
- Slow Traffic : $\rho \le 1 \Rightarrow \pi_0 = 1$

Heavy Traffic : $\rho > 1 \Rightarrow 0 < \pi_0 < 1$

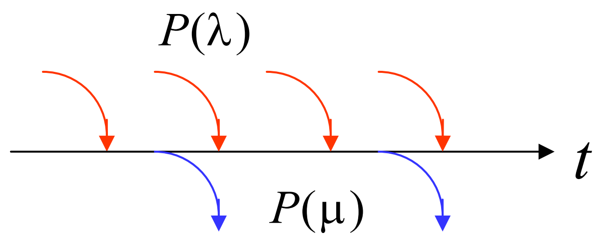
i.e., unstable queues $(\rho > 1)$ either terminate their busy periods with probability $\pi_0 < 1$, or they will continue to be busy with probability 1- π_0. Interestingly, there is a finite probability of busy period termination even for unstable queues.

π_0 : Measure of stability for unstable queues.

Example 20.1 : *M/M/* 1 queue

From (15-221), text,we have

$$a_k = \frac{\mu}{\lambda+\mu}\left(\frac{\lambda}{\lambda+\mu}\right)^k = \frac{1}{1+\rho}\left(\frac{\rho}{1+\rho}\right)^k, \qquad k = 0, 1, 2, \cdots \tag{20-8}$$



Number of arrivals between any two departures follows a geometric random variable.

$$P(z) = \frac{1}{[1+\rho(1-z)]}, \qquad \rho = \frac{\lambda}{\mu} \tag{20-9}$$

$$P(z) = z \Rightarrow \rho\, z^2 - (\rho+1)z + 1 = (z-1)\,(\rho\, z - 1) = 0$$

$$\pi_0 = \begin{cases} 1, & \rho \le 1 \\ \dfrac{1}{\rho}, & \rho > 1 \end{cases} \tag{20-10}$$

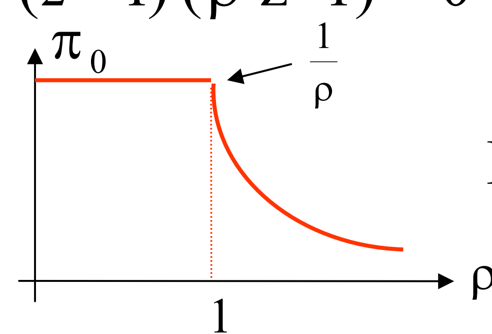


Fig 20.5

Example 20.2 : Bulk Arrivals $M^{[x]}/M/1$ queue

Compound Poisson Arrivals : Departures are exponential random variables as in a Poisson process with parameter μ. Similarly arrivals are Poisson with parameter λ. However each arrival can contain multiple jobs.

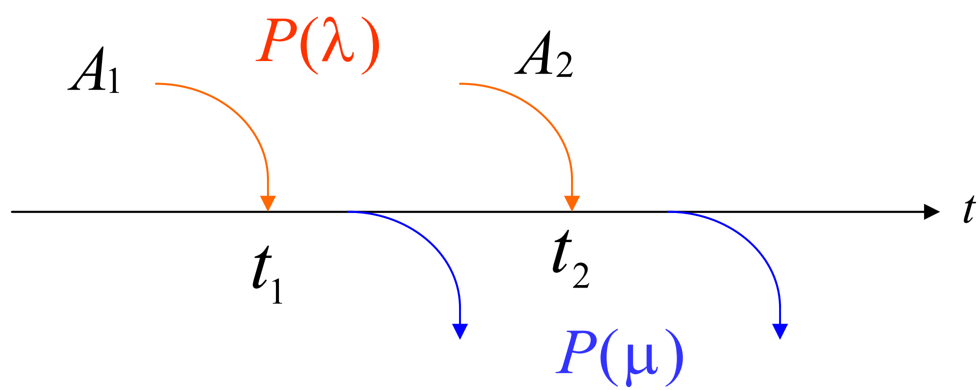


Fig 20.6

A_i : Number of items arriving at instant t_i

Let $\quad P\{A_i = \mathrm{k}\} = c_k, \quad k = 0, 1, 2, \cdots$

and $\quad C(z) = E\{z^{A_i}\} = \sum_{k=0}^{\infty} c_k z^k$ represents the bulk arrival statistics.

Inter-departure Statistics of Arrivals

$$P(z)=\sum_{k=0}^{\infty} P\{Y=\mathrm{k}\}\, z^k = E\{z^Y\}$$

$$=\sum_{n=0}^{\infty}\int_0^{\infty} E\{z^{(A_1+A_2+\cdots+A_n)}\ \big|\, n \text{ arrivals in } (0,t)\}\, P\{n \text{ arrivals in } (0,t)\}\, f_s(\mathrm{t})\, dt$$

$$=\sum_{n=0}^{\infty}\int_0^{\infty} [E\{z^{A_i}\}]^n\ e^{-\lambda t}\frac{(\lambda t)^n}{n!}\mu\ e^{-\lambda t} dt\ , \qquad \rho=\frac{\lambda}{\mu}$$

$$=\sum_{n=0}^{\infty}\mu\int_0^{\infty} e^{-(\lambda+\mu)t}\frac{[\lambda t\ C(z)]^n}{n!}=\frac{1}{1+\rho\{1-C(z)\}} \tag{20-11}$$

$$P(z)=[1+\rho\{1-C(z)\}]^{-1},\ \text{Let}\ \ c_k=(1-\alpha)\,\alpha^k, \qquad k=0,\ 1,\ 2,\ \cdots$$

$$C(z)=\frac{1-\alpha}{1-\alpha z}, \qquad P(z)=\frac{1-\alpha\ z}{1+\alpha\rho-\alpha(1+\rho)z} \tag{20-12}$$

$$\text{Traffic Rate}=P'(1)=\frac{\alpha\rho}{1-\alpha}>1$$

$$P(z)=z \ \Rightarrow\ (\mathrm{z}-1)[\alpha(1+\rho)\mathrm{z}-1]=0 \quad \pi_0=1/(\alpha(1+\rho))$$

Bulk Arrivals (contd)

- Compound Poisson arrivals with geometric rate

$$\pi_0 = \frac{1}{\alpha(1+\rho)}, \qquad \rho > \frac{1-\alpha}{\alpha}. \quad (20\text{-}13)$$

For $\alpha = \frac{2}{3}$, we obatain

$$\pi_0 = \frac{3}{2\,(1+\rho)}, \qquad \rho > \frac{1}{2} \qquad (20\text{-}14)$$

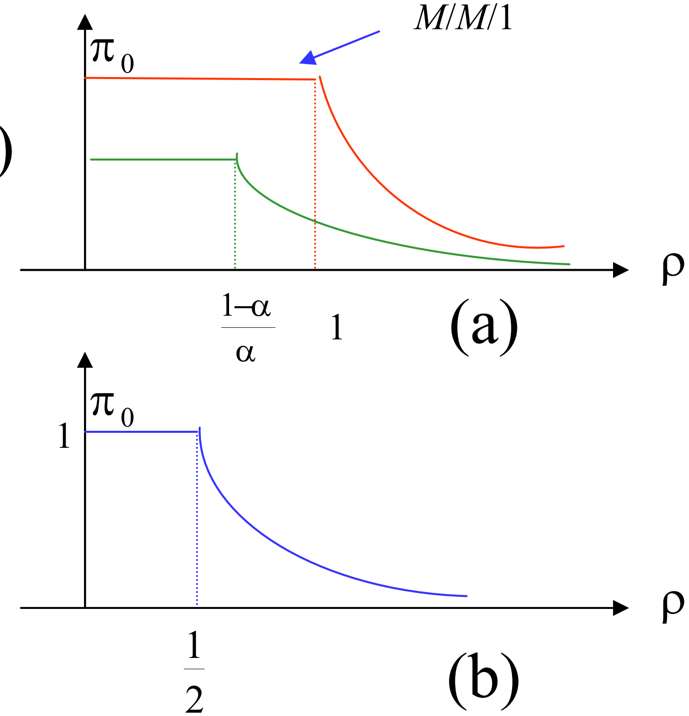


Fig 20.7

- **Doubly Poisson arrivals** $\Rightarrow$ $C(z) = e^{-\mu(1-z)}$ gives

$$P(z) = \frac{1}{1+\rho[1-C(z)]} = z \qquad (20\text{-}15)$$

Example 20.3 : *M/E*$_n$ / 1 queue (*n*-phase exponential service)

From (16-213)

$$P(z) = \left(1 + \frac{\rho}{n}(1-z)\right)^{-n} \qquad (20\text{-}16)$$

$$P(z) = z \Rightarrow \left(x^n + x^{n-1} + \cdots + x - \frac{n}{\rho}\right) = 0, \quad x = z^{\frac{1}{n}}$$

$$n = 2 \Rightarrow \pi_0 = \left(\frac{-1 + \sqrt{1 + 8/\rho}}{2}\right)^2, \quad \rho > 1$$

$$\pi_0 \approx (1 + \rho / n)^{-n}, \quad n >> 1 \qquad (20\text{-}17)$$

$\rho = 2$	
$M/M/1 \rightarrow n = 1$	$\pi_0 = 0.5$
$M/E_2/1 \rightarrow n = 2$	$\pi_0 = 0.38$

Example 20.4 : *M/D/*1 queue

Letting $m \rightarrow \infty$ in (16.213),text, we obtain $P(z) = e^{-\rho(1-z)}$,
so that

$$\pi_0 \approx e^{-\rho(1-e^{-\rho})}, \quad \rho > 1. \qquad (20\text{-}18)$$

20.2 Martingales

Martingales refer to a specific class of stochastic processes that maintain a form of "stability" in an overall sense. Let $\{X_i, i \geq 0\}$ refer to a discrete time stochastic process. If n refers to the present instant, then in any realization the random variables $X_0, X_1, \cdots, X_n$ are known, and the future values $X_{n+1}, X_{n+2}, \cdots$ are unknown. The process is "stable" in the sense that conditioned on the available information (past and present), no change is expected on the average for the future values, and hence the conditional expectation of the immediate future value is the same as that of the present value. Thus, if

$$E\{X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1, X_0\} = X_n \qquad (20\text{-}19)$$

for all n, then the sequence $\{X_n\}$ represents a **Martingale**.

Historically *martingales* refer to the "doubling the stake" strategy in gambling where the gambler doubles the bet on every loss till the almost sure win occurs eventually at which point the entire loss is recovered by the wager together with a modest profit. Problems 15-6 and 15-7, chapter 15, Text refer to examples of martingales. [Also refer to section 15-5, Text].

If $\{X_n\}$ refers to a Markov chain, then as we have seen, with

$$p_{ij} = P\{X_{n+1} = j \mid X_n = i\},$$

Eq. (20-19) reduces to the simpler expression [Eq. (15-224), Text]

$$\sum_j j\, p_{ij} = i. \qquad (20\text{-}20)$$

For finite chains of size N, interestingly, Eq. (20-20) reads

$$P x_2 = x_2, \qquad x_2 = [1,\ 2,\ 3,\ \cdots,\ N]^T \tag{20-21}$$

implying that x_2 is a right-eigenvector of the $N \times N$ transition probability matrix $P = (p_{ij})$ associated with the eigenvalue 1. However, the "all one" vector $x_1 = [1,\ 1,\ 1,\ \cdots,\ 1]^T$ is always an eigenvector for any P corresponding to the unit eigenvalue [see Eq. (15-179), Text], and from Perron's theorem and the discussion there [Theorem 15-8, Text] it follows that, for finite Markov chains that are also martingales, P *cannot* be a *primitive* matrix, and the corresponding chains are in fact *not irreducible*. Hence every finite state martingale has at least two closed sets embedded in it. (The closed sets in the two martingales in Example 15-13, Text correspond to two absorbing states. Same is true for the Branching Processes discussed in the next example also. Refer to remarks following Eq. (20-7)).

Example 20.5: As another example, let $\{X_n\}$ represent the branching process discussed in section 15-6, Eq. (15-287), Text. Then Z_n given by

$$Z_n = \pi_0^{X_n}, \qquad X_n = \sum_{i=1}^{X_{n-1}} Y_i \tag{20-22}$$

is a martingale, where Y_i s are independent, identically distributed random variables, and π_0 refers to the extinction probability for that process [see Theorem 15.9, Text].

To see this, note that

$$E\{Z_{n+1} \mid Z_n, \cdots, Z_0\} = E\{\pi_0^{X_{n+1}} \mid X_n, \cdots, X_0\}$$

$$= E\{\pi_0^{\sum_{i=0}^{k} Y_i} \mid \underbrace{X_n = k}_{\substack{\text{since } \{X_n\} \text{ is} \\ \text{a Markov chain}}} \} = \prod_{i=1}^{X_n = k} [E\{\pi_0^{Y_i}\}] = [P(\pi_0)]^{X_n} = \pi_0^{X_n} = Z_n, \tag{20-23}$$

since Y_i s are independent of X_n

use (15-2)

where we have used the Markov property of the chain, the common moment generating function $P(z)$ of $Y_i\,s$, and Theorem 15-9, Text.

Example 20.6 (DeMoivre's Martingale): The gambler's ruin problem (see Example 3-15, Text) also gives rise to various martingales. (see problem 15-7 for an example).

From there, if S_n refers to player A's cumulative capital at stage n, (note that $S_0 = \$\, a$), then as DeMoivre has observed

$$Y_n = \left(\frac{q}{p}\right)^{S_n} \tag{20-24}$$

generates a martingale. This follows since

$$S_{n+1} = S_n + Z_{n+1} \tag{20-25}$$

where the instantaneous gain or loss given by Z_{n+1} obeys

$$P\{Z_{n+1} = 1\} = p, \qquad P\{Z_{n+1} = -1\} = q, \tag{20-26}$$

and hence

$$E\{Y_{n+1} \mid Y_n, Y_{n-1}, \cdots, Y_0\} = E\{\left(\frac{q}{p}\right)^{S_{n+1}} \mid S_n, S_{n-1}, \cdots, S_0\}$$

$$= E\{\left(\frac{q}{p}\right)^{S_n + Z_{n+1}} \mid S_n\},$$

since $\{S_n\}$ generates a Markov chain.

Thus

$$E\{Y_{n+1} \mid Y_n, Y_{n-1}, \cdots, Y_0\} = \left(\tfrac{q}{p}\right)^{S_n} \left(\tfrac{q}{p} \cdot p + \left(\tfrac{q}{p}\right)^{-1} \cdot q\right) = \left(\tfrac{q}{p}\right)^{S_n} = Y_n \quad (20\text{-}27)$$

i.e., Y_n in (20-24) defines a martingale!

Martingales have excellent convergence properties in the long run. To start with, from (20-19) for any *given* n, taking expectations on both sides we get

$$E\{X_{n+1}\} = E\{X_n\} = E\{X_0\}. \quad (20\text{-}28)$$

Observe that, as stated, (20-28) is true only when n is *known* or n is a *given* number.

As the following result shows, martingales do not fluctuate wildly. There is in fact only a small probability that a large deviation for a martingale from its initial value will occur.

Hoeffding's inequality: Let $\{X_n\}$ represent a martingale and $\sigma_1, \sigma_2, \cdots,$ be a sequence of real numbers such that the random variables

$$Y_i \triangleq \frac{X_i - X_{i-1}}{\sigma_i} \le 1 \qquad \textit{with probability one.} \qquad (20\text{-}29)$$

Then

$$P\{|X_n - X_0| \ge x\} \le 2e^{-(x^2/2\sum_{i=1}^{n}\sigma_i^2)} \qquad (20\text{-}30)$$

Proof: Eqs. (20-29)-(20-30) state that so long as the martingale increments remain bounded almost surely, then there is only a very small chance that a large deviation occurs between X_n and X_0. We shall prove (20-30) in three steps.

(i) For any convex function $f(x)$, and $0 < \alpha < 1,$ we have (Fig 20.8)

$$\alpha f(x_1) + (1-\alpha) f(x_2) \ge f(\alpha x_1 + (1-\alpha)x_2), \qquad (20\text{-}31)$$

which for $\alpha = \frac{1-a}{2},\ 1-\alpha = \frac{1+a}{2}$, $|a|<1,\ x_1 = -1,\ x_2 = 1$ and $f(x) = e^{\varphi x},\ \varphi > 0$ gives

$$\frac{1}{2}(1-a)e^{-\varphi} + \frac{1}{2}(1+a)e^{\varphi} \geq e^{a\varphi}, \qquad |a|<1. \qquad (20\text{-}32)$$

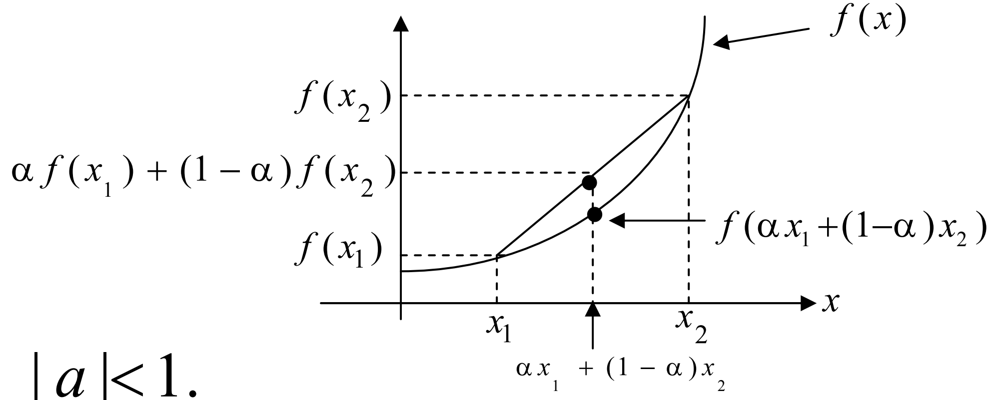


Fig 20.8

Replacing a in (20-32) with any zero mean random variable Y that is bounded by unity almost everywhere, and taking expected values on both sides we get

$$E\{e^{\varphi Y}\} \leq \frac{1}{2}(e^{\varphi} + e^{-\varphi}) \leq e^{\varphi^2/2} \qquad (20\text{-}33)$$

Note that the right side is independent of Y in (20-33). On the other hand, from (20-29)

$$E\{Y_i \mid X_i, \cdots, X_1, X_0\} = E(X_i \mid X_{i-1}) - X_{i-1} = X_{i-1} - X_{i-1} = 0 \qquad (20\text{-}34)$$

and since Y_i s are bounded by unity, from (20-32) we get (as in (20-33))

$$E\{e^{\varphi Y_i} \mid X_{i-1}, \cdots, X_1, X_0\} \le e^{\varphi^2/2} \qquad (20\text{-}35)$$

(ii) To make use of (20-35), referring back to the Markov inequality in (5-89), Text, it can be rewritten as

$$P\{X \ge \alpha\} \le e^{-\theta\alpha} E\{e^{\theta X}\}, \quad \theta > 0 \qquad (20\text{-}36)$$

and with $X = X_n - X_0$ and $\alpha = x,$ we get

$$P\{X_n - X_0 \ge x\} \le e^{-\theta x} E\{e^{\theta(X_n - X_0)}\} \qquad (20\text{-}37)$$

But

use (20-29)

$$\begin{aligned} E\{e^{\theta(X_n - X_0)}\} &= E\{e^{\theta(X_n - X_{n-1}) + \theta(X_{n-1} - X_0)}\} \\ &= E[E\{e^{\theta(X_{n-1} - X_0)} e^{\theta\sigma_n Y_n} \mid X_{n-1}, \cdots, X_1, X_0\}] \\ &= E[e^{\theta(X_{n-1} - X_0)} \underbrace{E\{e^{\theta\sigma_n Y_n} \mid X_{n-1}, \cdots, X_1, X_0\}}_{\le e^{\theta^2\sigma_n^2/2} \text{ using (20-35)}}] \\ &\le E\{e^{\theta(X_{n-1} - X_0)}\} e^{\theta^2\sigma_n^2/2} \le e^{\theta^2 \sum_{i=1}^n \sigma_i^2/2}. \end{aligned} \qquad (20\text{-}38)$$

Substituting (20-38) into (20-37) we get

$$P\{X_n - X_0 \ge x\} \le e^{-(\theta x - \theta^2 \sum_{i=1}^{n} \sigma_i^2 /2)} \tag{20-39}$$

(iii) Observe that the exponent on the right side of (20-39) is minimized for $\theta = x / \sum_{i=1}^{n} \sigma_i^2$ and hence it reduces to

$$P\{X_n - X_0 \ge x\} \le e^{-x^2 / 2\sum_{i=1}^{n} \sigma_i^2}, \qquad x > 0. \tag{20-40}$$

The same result holds when $X_n - X_0$ is replaced by $X_0 - X_n$, and adding the two bounds we get (20-30), the Hoeffding's inequality.

From (20-28), for any fixed n, the mean value $E\{X_n\}$ equals $E\{X_0\}$. Under what conditions is this result true if we replace n by a *random time* T? i.e., if T is a random variable, then when is

$$E\{X_T\} \overset{?}{=} E\{X_0\}. \tag{20-41}$$

The answer turns out to be that T has to be a *stopping time*. What is a **stopping time**?

A stochastic process may be known to assume a particular value, but the time at which it happens is in general unpredictable or random. In other words, the nature of the outcome is fixed but the timing is random. When that outcome actually occurs, the time instant corresponds to a *stopping time*. Consider a gambler starting with \$a and let T refer to the time instant at which his capital becomes \$1. The random variable T represents a stopping time. When the capital becomes zero, it corresponds to the gambler's ruin and that instant represents another stopping time (Time to go home for the gambler!)

Recall that in a Poisson process the occurrences of the first, second, $\cdots$ arrivals correspond to stopping times $T_1, T_2, \cdots$. Stopping times refer to those random instants at which there is sufficient information to decide whether or not a specific condition is satisfied.

Stopping Time: The random variable T is a stopping time for the process $X(t)$, if for all $t \geq 0$, the event $\{T \leq t\}$ is a function of the values $\{X(\tau) \mid \tau > 0, \tau \leq t\}$ of the process up to t, i.e., it should be possible to decide whether T has occurred or not by the time t, knowing *only* the value of the process $X(t)$ up to that time t. Thus the Poisson arrival times T_1 and T_2 referred above are stopping times; however $T_2 - T_1$ is not a stopping time.

A key result in martingales states that so long as

T is a stopping time (under some additional mild restrictions)

$$E\{X_T\} = E\{X_0\}. \tag{20-42}$$

Notice that (20-42) generalizes (20-28) to certain random time instants (stopping times) as well.

Eq. (20-42) is an extremely useful tool in analyzing martingales. We shall illustrate its usefulness by rederiving the gambler's ruin probability in Example 3-15, Eq. (3-47), Text.

From Example 20.6, Y_n in (20-24) refer to a martingale in the gambler's ruin problem. Let T refer to the random instant at which the game ends; i.e., the instant at which either player A loses all his wealth and P_a is the associated probability of ruin for player A, or player A gains all wealth \$$(a + b)$ with probability $(1 - P_a)$. In that case, T is a stopping time and hence from (20-42), we get

$$E\{Y_T\} = E\{Y_0\} = \left(\frac{q}{p}\right)^a \tag{20-43}$$

since player A starts with \$a in Example 3.15. But

$$E\{Y_T\} = \left(\frac{q}{p}\right)^0 P_a + \left(\frac{q}{p}\right)^{a+b}(1-P_a)$$

$$= P_a + \left(\frac{q}{p}\right)^{a+b}(1-P_a). \tag{20-44}$$

Equating (20-43)-(20-44) and simplifying we get

$$P_a = \frac{1-\left(\frac{p}{q}\right)^b}{1-\left(\frac{p}{q}\right)^{a+b}} \tag{20-45}$$

that agrees with (3-47), Text. Eq. (20-45) can be used to derive other useful probabilities and advantageous plays as well. [see Examples 3-16 and 3-17, Text].

Whatever the advantage, it is worth quoting the master Gerolamo Cardano (1501-1576) on this: "*The greatest advantage in gambling comes from not playing at all*."

Errata for Papoulis/Pillai's Probability, Random Variables and Stochastic Processes, 4e

Page	Line	Instead of	Read
165	Prob. 5–17 (first line)	$Y = X^2$	$Y = \sqrt{X}$
166	Prob. 5–38 (a)	$(1 - \beta e^{j\omega})^{-\alpha}$	$(1 - j\beta\omega)^{-\alpha}$
166	Prob. 5–38 (b)	$(1 - 2e^{j\omega})^{-n/2}$	$(1 - j2\omega)^{-n/2}$
236	Prob. 6–8 last line	$f_{xy}(x, y)$	$f_z(z)$
246	4 (from bottom)	α	λ
398	7 (from bottom)	(9–142)	(9–146)
719	$10, 15$	(16–166)	(15–125)
719	11	(16–163)	(15–120)
719	12	(16–165)	(15–124)
719	$13, 15$	(16–167)	(15–126)
719	13	(16–159)	(15–114)
719	16	(16–173)	(15–131)
720	1	(16–176)	(15–133)
720	1	(16–182)	(15–135)
720	1	(16–185)	(15–136)
720	3	(16–165)	(15–124)
720	$3, 5, 11$	(16–186)	(15–137)
720	4	(16–168)	(15–127)
720	5	(16–159)	(15–114)
720	9	(16–169)	(15–128)

Page	Line	Instead of	Read
720	12	(16–187)	(15–138)
720	12	(16–170)	(15–129)
720	19	(16–181)– (16–186)	(15–134)– (15–137)
720	21	(16–171)	(15–130)
721	16	(16–200)	(15–144)
722	14, 16	(16–213)	(15–147)
722	14, 21	(16–214)	(15–148)
722	18	(16–212)	(15–146)
722	20	(16–216)	(15–149)
723	8	(16–219)	(15–152)
723	12, 18	(16–221)	(15–154)
723	10 (from bottom)	(16–169)	(15–128)
723	3 (from bottom)	(16–218)	(15–151)
724	6	(16–239)	(15–156)
725	16	(16–166)	(15–125)
725	18	(16–240)	(15–157)
725	20	(16–163)	(15–120)
726	11 (from bottom)	(16–156)	(15–110)
810	12, 14	Theorem 15–8	Theorem 15–9
813	2 (from bottom)	arriving	originated
817	3, 8, 15 (from bottom)	Theorem 15–9	Theorem 15–10
820	5	Theorem 15–8	Theorem 15–9
821	Left Margin	Nyquist Theorem	Burke′s Theorem
822	Eq.(16–256)	$\lambda\, F_{n-1}(t)$	$\lambda\, \Delta t\, F_{n-1}(t)$

Errata, Hints, and Problem Solutions

to accompany the text

Probability, Random Variables, and Stochastic Processes, Fourth Edition,
by Athanasios Papoulis and S. Unnikrishna Pillai,
McGraw Hill, 2002

prepared by Gary Matchett

Northrop Grumman
55 Walkers Brook Drive
Reading MA 01867-3297

also associated with
Northeastern University

Objective

This volume was prepared by the instructor for students of Northeastern University course ECE 3211, Applied Probability and Stochastic Processes, to help in dealing with the course Text. As the title indicates, there are three parts: Text errata, followed by hints for solving the problems, followed by problem solutions. The course does not cover the entire Text; nor do these notes. Chapters 1 through 7, Chapters 9 through 11, and the first Section of Chapter 12 are covered here. The course does not cover all the topics in Chapters 10 and 11, but all the problems in those chapters are included here.

The main part of this work is the problem solutions. The glory of the Text is its problems. Many of them contain key parts of important developments in the field, and are widely usable. Doing the problems is where real learning happens. There are so many problems and some are so difficult as to daunt all but the most intrepid student. I cannot read and correct homework, and I am often not available when help is needed, especially in getting started, so this volume is an attempt to fill the void.

Reading the solution to a problem without first seriously attempting to solve it yourself is a mistake. You may well learn something, but it will not have the staying power of an answer you developed. There are exceptions. Some of my students do not have enough training in mathematics to approach a few of the more mathematically oriented problems. The hints section identifies those problems that (I think) are important or difficult, as well as those few that have typographical errors, missing background material, incomplete assumptions, or indemonstrable results.

The attempt here is not to provide the typical "answer book." Rather, it is an attempt to turn each problem into a worked example so that the student may follow the solution and see it flow from the body of the Text. As an instructor, I have access to the publisher-provided solutions manuals. Following my own advice, I have not used those manuals in developing the solutions here. That means, sometimes, that an easier road to the solution of a particular problem might well exist in the official manual, although those are not generally available to students.

For the Second and Third Editions of the Text, this volume was reproduced from hand-written masters. For the Fourth Edition, it has been newly typeset. Typographical (and other) errors are likely to exist, and are especially confusing to students. I would be grateful for any corrections or comments, which might be addressed to

GMatchett@northropgrumman.com

Note that this volume is split into three parts for electronic dissemination as PDF files. The first part contains just the errata and the hints. The second part has the solutions for Chapters 2 through 6, and the third part has the solutions for Chapters 7, 9, 10, 11, and the first section of Chapter 12.

ERRATA & COMMENTS ON THE TEXT

(Fourth Edition)

The Fourth Edition of the Text is (almost) new at this writing. The errata here are for the first printing. Errors are sometimes corrected with new printings. The printing number is the first integer remaining in the sequence 1234567890 on the reverse of the title page, above the ISBN line. Since this Text is newly revised, I have not thoroughly examined it, and have likely overlooked some errors. I would appreciate notification of any errors not listed here (or any comments on errors listed here) to GMatchett@northropgrumman.com.

CHAPTER 2

- p. 19, first paragraph of Section 2-2, line 2, should read "... and certain of its subsets events."
- p. 19, first paragraph of Section 2-2, line 4, should read "... ζ_i is an elementary event, if, in fact, $\{\zeta_i\}$ is an event at all."
- p. 29, following Eq. (2-38), should read "...results involving probabilities hold also...".
- p. 32, following Eq. (2-42), note that "This result" refers to Eq. (2-41), and not to Eq. (2-42).

CHAPTER 3

- p. 53, Fig. 3-2 (a) should read $\frac{2}{3\sqrt{2\pi}}e^{-(x-4.5)^2/4.5}$.
- p. 53, Fig. 3-2 (b) should read $\frac{1}{\sqrt{4\pi}}e^{-(x-3.0)^2/4}$.
- p. 71, Problem 3-7 should read "... net gain or loss exceeds...".
- p. 71, Problem 3-5 should read "We pick at random $n \le N$ components ...".

CHAPTER 4

- p. 79, 3. Proof, should read "Suppose that $\mathbf{x}(\zeta) > 0$ for every ζ."
- p. 91, Eq. (4-48) should read $0 < x < 1$.
- p. 91, Eq. (4-49), the upper limit on the second integral should be $\pi/2$.
- p. 96, Watch out! There are two different definitions for both the negative binomial distribution and for the geometric distribution. When these distributions are specified, as in the problems, it is not always clear which of the two is meant.
- p. 99, line 4 (in Example 4-14) should read $20 \le x < 40$.
- p. 100, lines 5 and 6 should read $b < x \le a$.
- p. 100, line 10 should read $b < x < a$.
- p. 102, Example 4-17, the equation for the density is incorrect. The G symbols should

be $\boldsymbol{g}$ symbols (which were dropped from the Fourth Edition). Note that

$$\boldsymbol{g}(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \quad \text{so that} \quad \boldsymbol{G}(x) = \int_{-\infty}^{x} \boldsymbol{g}(\xi)d\xi$$

(compare with Eq. (4-27), and see Notational Pecularities of the Hints section).

- p. 106, Table 4-1, note that the Text definition of $\mathrm{erf}(x)$ is nonstandard. The relationship is

$$\mathrm{erf}(x) = \frac{1}{2}\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right) = \frac{1}{2}\Phi\left(\frac{x}{\sqrt{2}}\right)$$

where the first term is the Text definition, the second term is the standard definition, and the third term is the “Gradshteyn & Ryzhik” notation.

- p. 107, first line after Eq. (4-93), the restriction $x > 0$ is not needed.
- p. 119, Problem 4-3 should read “Using Table 4-1 ...”, and, in part (b), “$\boldsymbol{x}$ is $N(\eta, \sigma^2)$”.
- p. 120, Problem 4-8 should read“$\boldsymbol{x}$ is $N(10, 1)$”.
- p. 120, Problem 4-10 should read “$\boldsymbol{x}$ is $N(0, 4)$”.
- p. 120, Problem 4-12 should read “$\boldsymbol{x}$ is $N(1000, 400)$”.
- p. 120, Problem 4-14, the reference to Eq. (4-34) is not correct. The actual equation intended was removed in going to the Fourth Edition. Eq. (4-90) is a close equivalent to what was intended.
- p. 120, Problem 4-15 should read “... then $F(x) = 1$ for $x \geq b$ ...”.
- p. 121, Problem 4-26 should read “A system has 1000 components .”
- p. 121, Problem 4-28, the last “>” should be a “<“ in the Hint.
- p. 122, Problem 4-35 should say in addition “Assume also: $k_1, k_2 \ll n, k_3$”.

CHAPTER 5

- p. 133, equation after Eq. (5-24), the last term should be $U(x)$.
- p. 145, Example 5-22, next to last line, should read “... that $E\{(\boldsymbol{x} - \eta)^2\} = \sigma^2$ and ...”.
- p. 151, equation following Eq. (5-88), should read

$$P\{|\boldsymbol{x} - \eta| \geq \varepsilon\} = \int_{-\infty}^{\eta - \varepsilon} f(x)dx + \ldots$$

- p. 153, Example 5-28, should read “... function of an $N(\eta, \sigma^2)$ random variable ...”.
- p. 155, two lines after Eq. (5-111), the right hand equation should have a double prime on the phi on the left hand side.
- p. 164, Problem 5-1, should read $N(5, 4)$.
- p. 165, Problem 5-17, should read $\boldsymbol{y} = \sqrt{\boldsymbol{x}}$.
- p. 166, Problem 5-38, (a) should read $\Phi_x(\omega) = (1 - j\omega\beta)^{-\alpha}$, and (b) should read

  $\Phi_x(\omega) = (1 - j2\omega)^{-n/2}$.
- p. 168, Problem 5-51 (b), the equation line should end $k \leq \min(M, n)$.

CHAPTER 6

- p. 182, Eq. (6-43), lower limit on the first integral should read $y = -\infty$.
- p. 208, Eq. (6-157), see Special Note 3 in the Hints section of this volume.
- p. 219, Example 6-36, line 5, should begin "$\phi(s_1, s_2) = e^A$".
- p. 236, Problem 6-8, second centered equation, left side should read $f_z(z) =$.
- p. 237, Problem 6-25, see Special Note 2 in the Hints section of this volume It is easiest if it reads "... exceeds $2/\lambda$.", and "... original component by $1/\lambda$?"
- p. 238, Problem 6-36, should ask to show that w is an exponential random variable.
- p. 239, Problem 6-43 has "excess" information. See the hint for this problem.
- p. 241, Problem 6-67, the first display equation should read

$$E\{\boldsymbol{z}\} = \sum_n p_n E\{g(x_n, \boldsymbol{y}) | x_n\}$$

- p. 242, Problem 6-76, should read "..., $\beta_y(t) = f_y(t|(\boldsymbol{y} > t))$ and ...".

CHAPTER 7

- p. 243, Eq. (7-3), should not have commas in the denominator of the fraction.
- p. 260, line 17, should read "as in Example 7-5. ..."
- p. 276, second line of the section "Ergodicity" should refer to Sec. 12-1.
- p. 279, Example 7-15, should read "... in the interval $(0, T)$."
- p. 283, Eq. (7-136), should read

$$\int_{-\infty}^{\infty} |x|^{\alpha} f_i(x) dx < K < \infty \quad \text{for all } i$$

- p. 283, Eq. (7-138), should read

$$E\{|\boldsymbol{x}_i|^3\} \le c\sigma_i^2 \quad \text{all } i$$

- p. 291, Example 7-21, in two places the reference to (7-15) should be to (7-156).
- p. 302, Problem 7-32, should begin "... are normal, uncorrelated with zero mean ...".

CHAPTER 9

- p. 378, Example 9-5, results are valid for positive times only. The right hand term of the equation following Eq. (9-14) is not correct if $t_1 = t_2 = t$, where it produces $2\lambda t$. Correctly, $C(t, t) = \lambda t$.
- p. 379 & 380, equations following Eq. (9-18) are only true if $t_2 \le t_1$. In case $t_1 \le t_2$, all the t_2's in these equations should be replaced by t_1's.
- p. 384, Eq. 9-37, should read $E\{|\boldsymbol{s} - \eta_s|^2\} = \int_a^b \dots$
- p. 391, first line of "Proof" near page bottom should read "...[See (6-242)]"
- p. 397, Example 9-17 should assume a SSS process.
- p. 397, equation after Eq. (9-83) should have $g(x) = +c$ when $x > c$, and $g(x) = -c$

when $x < -c$.

- p. 399, Eqs. (9-92) and (9-93) are only true for a real system. When the system is complex, the conjugates L_2^* and h^* should appear.
- p. 400, equation before Example 9-18 should read "$\ldots h(\alpha)h^*(\beta)\ldots$".
- p. 404, Eq. (9-111) should begin "$a_n \mathbf{y}^{(n)}(t) + \ldots$".
- p. 412, next to last line, should read "... cases of (9-92) and (9-93)." Note that $H(\omega)$ is used here without any introduction. It is the *system function*, where

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t}dt \qquad\qquad h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)e^{j\omega t}d\omega$$

- p. 412, Eq. (9-150) uses $\rho(\tau)$ without calling it the *deterministic autocorrelation* of $h(t)$, and the next line (on p. 413) should read " ... white noise with intensity q.
- p. 413, two lines before Eq. (9-155) should read "... It is thus a simple low-pass filter."
- p. 414, Eq. (9-157) should read

$$S_{yy}(\omega) = |j\omega|^{2n}S_{xx}(\omega) \qquad\qquad R_{yy}(\tau) = (-1)^n R_{xx}^{(2n)}(\tau)$$

- p. 415, Example 9-28, line 3, should read "... equals $\frac{1}{\pi t}$".
- p. 419, Eq. (9-179) should read "$E\{|\mathbf{x}(t+\tau_1) - \mathbf{x}(t)|^2\} = 0$".
- p. 412, Eq. (9-194), the sum should begin at $m = 1$.
- p. 425, Example 9-34, note that a is presumed real, and line 5 should read

$$R_{yy}[m] = \frac{q}{1-a^2}a^{|m|}$$

- p. 430, Problem 9-6, should read

$$E\{\mathbf{w}^2(t)\} = \int_0^t (t-\tau)^2 q(\tau)d\tau$$

- p. 433, Problem 9-41, should read

$$S_y(\omega) = 2\pi R_x^2(0)\delta(\omega) + \frac{1}{\pi}S_x(\omega)*S_x(\omega)$$

- p. 433, Problem 9-51, should read "$R[0]R[2] \geq 2R^2[1] - R^2[0]$"
- p. 433, Problem 9-52, should read "$\mathbf{x}[n] = Ae^{jn\omega}$." ($\omega$ is a random variable here; I cannot make bold greek symbols.)

CHAPTER 10

- p. 447, line after Eq. (10-58) should read "$k = 1.37 \times 10^{-23}$ Joules/degree-K".
- p. 450, first word, should read "inertia has the...".
- p. 461, four lines after Eq. (10-103), should read "... and $E\{\mathbf{n}^2(t)\} = \lambda^2 t^2 + \lambda t$."
- p. 466, Eq. (10-142) (second part) should read "$R_{x\hat{x}}(-\tau) = -R_{x\hat{x}}(\tau)$".
- p. 469, line three, should read "... that $S_{zz}(\omega)$ is specified. ...".

- p. 473, two lines before Eq. (10-173), should read "SSCS process with period T and ...
- p. 474, Eq. (10-180), should read

$$\bar{S}_x(\omega) = \frac{1}{T}\mathsf{S}_c(e^{j\omega T})|H(\omega)|^2 \text{ ''}$$

- p. 474, Eq. (10-182) is improper. To fix this, define $\boldsymbol{w}(t) = \sum_{n=0}^{\infty} \boldsymbol{c}_n U(t-nT)$ for $t \ge 0$, with an appropriate alternate definition for $t < 0$. Now $\boldsymbol{w}(0^-) = 0$.
- p. 474, last equation should have the term $R_c[n-r]$, not $R_c(n-r)$.
- p. 475, Eq. (10-183), should read

$$\ldots = \sum_{m=-\infty}^{\infty} R_c[m] \sum_{r=-\infty}^{\infty} \delta[t+\tau-(m+r)T]\delta(t-rT)$$

- p. 475, equation after Eq. (10-183) should read

$$\sum_{r=-\infty}^{\infty} \int_0^T \delta[t+\tau-(m+r)T]\delta(t-rT)dt = \delta(\tau-mT)$$

- p. 475, Eq. (10-185) should read

$$\bar{S}_z(\omega) = \frac{1}{T}\sum_{m=-\infty}^{\infty} R_c[m]e^{-jm\omega T} = \frac{1}{T}S_c(\omega T)$$

- p. 476, Eq. (10-189) should indicate equality in mean square.
- p. 477, Eq. (10-194) presumes the process is real.
- p. 480, five lines before Eq. (10-203), should read "$T_0 \le \pi/\sigma$ ...".
- p. 493, line before Eq. (10A-1) should read "... for any $c > 0$".
- p. 494, line 3 is incorrect. See the solution to Problem 10-23.
- p. 496, Problem 10-13, equation should read "$S_x(\omega) = \frac{2\pi}{T^2}\left|\int_0^T \ldots\right.$"
- p. 497, Problem 10-21, needs the assumption that $\boldsymbol{x}(t)$ is independent of all $\boldsymbol{t}_i$, and should read "$\boldsymbol{X}_c(\omega) = \frac{1}{\lambda}\sum_{|\boldsymbol{t}_i|<c} \boldsymbol{x}(\boldsymbol{t}_i)e^{-j\omega\boldsymbol{t}_i}$". It is not needed that $E\{\boldsymbol{x}(t)\} = 0$.
- p. 497, Problem 10-25, should read "$\boldsymbol{y}(t) = B\cos(\omega_0 t + \varphi) + \boldsymbol{y}_n(t)$".

CHAPTER 11

- p. 506, two lines after Eq. (12-30), should read "... system of Fig. 11-4 is ...".
- p. 507, line after Eq, (11-37), should read "$\alpha_i = \gamma_i \mathsf{L}(1/z_i)$". (I cannot duplicate the Text fonts, but there is a problem of consistency here.)
- p. 508, line before Eq. (11-44), should refer to Example 9-32.

- p. 515, line before Eq. (11-78), should read "$\mathbf{B}(-\omega) = -\mathbf{B}(\omega)$".
- p. 515, line after Eq. (11-79a) should refer to Eq. (11-70), not (11-9).
- p. 521, Problem 11-7, should read

$$\beta_n = \left(a + \frac{\lambda_n}{2}\right)^{-1/2} \qquad \beta_n' = \left(a + \frac{\lambda_n'}{2}\right)^{-1/2}$$

- p. 522, Problem 11-10, should read "$E\{\boldsymbol{x}_n \boldsymbol{x}_k^*\} = \ldots$".

CHAPTER 12

- p. 527 Eq. (12-9), should read "$\int_0^T |C(\tau)| d\tau < \infty$".

PROBLEM HINTS AND COMMENTS

1) Introduction

Several symbols are used to comment on the problems, often subjectively. They are

- **I**, an important problem, likely one whose results will be needed later
- **M**, a moderately difficult problem
- **D**, a difficult problem
- **F**, a flawed problem, generally containing a typographical error, a missing assumption, or an indemonstrable result
- **B**, a problem presuming background information not presented in the Text up to the point of the problem, or not presented at all

Notational Peculiarities

Where the Text uses $\{\varnothing\}$ for the null, or empty, set, we use $\varnothing$.

The symbol used here for the Gaussian distribution function, defined in Eq. 4-27 of the Text, is $\mathbb{G}(x)$.

The Fourth Edition of the Text does not use the symbol $\boldsymbol{g}(x)$ (except in Problem 4-28), used in the Third Edition, and occasionally here, for the Gaussian density function. The Fourth Edition thinks that no symbol is necessary for this function, since
$\boldsymbol{g}(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

The notation $\ln x$ means the natural logarithm of x. The Text sometimes uses $\log x$.

Special Notes

1. The Text changed notation in going from the Third Edition to the Fourth Edition. In the Third edition, a normal random variable with mean η and standard deviation σ was denoted as $N(\eta;\ \sigma)$, which was sloppily written as $N(\eta,\ \sigma)$ from time to time. In the Fourth Edition, the official notation is $N(\eta,\ \sigma^2)$. Unfortunately, not all instances of the older notation were changed. This is particularly troubling when actual numbers are used for the parameters, as they are in some problems. Does the notation $N(100,\ 25)$ mean a normal random variable with a standard

deviation of 25 or of 5? One cannot be sure. These hints will provide my guess, from comparison of the two editions, and from a reading of the official Solutions Manual.

2. Eq. 4-30, p. 85, defines the distribution of an exponential random variable with parameter λ to have the p.d.f. $f_x(x) = \lambda e^{-\lambda x} U(x)$. The parameter has units that are the inverse of the units of the random variable itself (often $\boldsymbol{x}$ has units of time, and λ has units of (1/time)). It makes, perhaps, more sense to use the inverse of λ as the parameter, so that the parameter and the variable share the same units, and the p.d.f. is $f_x(x) = (1/\lambda)e^{-x/\lambda} U(x)$, but that was not done, except in some of the problems. When a problem states that $\boldsymbol{x}$ is an exponential random variable with parameter λ, one cannot be sure which of the two possibilities is meant, except by implication. For example, if it asks: What is the probability that $\boldsymbol{x}$ exceeds 2λ? (see Problem 6-25), then it must be that $\boldsymbol{x}$ and λ have the same units, so the inverse parameter is indicated.

3. Eqs. 4-49 and 4-51 define and evaluate the beta function, which is denoted as $B(\alpha, \beta)$. In some places (e.g., Eq. 6-157) this function is denoted as $\beta(m, n)$. For further confusion, the binomial distribution is sometimes denoted as $B(n, p)$ (see Theorem 5-2 or Example 5-30).

References

The problems are more easily done with three tools: a scientific calculator, a set of good function tables, and a table of integrals and sums. I use and recommend:

AS — Milton Abramowitz and Irene A. Stegun, Editors, *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series 55, June 1964 (now available as a Dover paperback).

GR — I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series, and Products*, Fourth Edition, Academic Press, 1965 (Sixth Edition, 2000, now available).

A excellent book for general reference here is:

FI — William Feller, *An Introduction to Probability Theory and Its Applications*, Volume I, Second Edition, Wiley & Sons, 1957 (Third Edition, 1990, is current).

2) Chapter 2 (there are no problems in Chapter 1)

2-1 Begin with DeMorgan's laws, Eq. 2-5.

2-2 No hint.

2-3 Try to find some set C such that $\bar{B} = A + C$, and $AC = \varnothing$. Then apply Eq. 2-10.

2-4 (a) Use Eq. 2-10. (b) Use Eqs. 2-12 and 2-13.

2-5 Use Eq. 2-13, repeatedly.

2-6 **M**, **B** Use the fact that a set is countable if it is empty or is the range of some sequence. Show that any subset of a countable set is countable. Then use the countable union property of Borel fields to show that every subset of S is an event.

2-7 List all subsets of S. Beginning with the list $\varnothing$, S, $\{1\}$, and $\{2, 3\}$, form compliments and unions among the list items to find new subsets that must be in the Borel field and add them to the list. This process stops when nothing new can be found. Note that any finite Borel field must have 2^n elements for some integer n.

2-8 Use the definition of conditional probability, Eq. 2-33.

2-9 No hint.

2-10 **I** No hint.

2-11 **M** See the solution for help.

2-12 Presume classical probability theory with the probability of an interval of points being proportional to its length.

2-13 **D**, **B** For an easier problem, assume that $P\{t \le t_1\} = F(t_1)$ is a continuous, differentiable function of t_1, and assume $F(0) = 0$.

2-14 **I** No hint.

2-15 **I**, **M** Enrich this problem by letting B_i be any of the sets A_i, $\overline{A_i}$, S, or $\varnothing$.

2-16 **M**, **B** Problems 2-16 through 2-19 and 2-21 are problems in combinatorics that the Text has not yet considered. It would be good to at least solve Problem 2-26 first, to begin the topic. Let an outcome of the experiment here be a k-element sequence of distinct numbers selected from the set 1 to n. Find out how many outcomes there are. (a) Next, find out how many outcomes there are that contain no number larger than m. Call this result M_m. Then notice that the number of outcomes that have m as the largest number is $N_m = M_m - M_{m-1}$. (b) The number of outcomes with a largest number less than or equal to m is the number of outcomes with no number larger than m.

2-17 B Work Problem 2-16 first. The difference here is that the outcomes are k-element sequences of numbers that are not necessarily distinct.

2-18 B Work Problem 2-26 first.

2-19 B Number the black balls from 1 to n, and number the white balls from $n+1$ to $n+m$. Now the problem is: what is the probability that, if k balls are drawn, the highest numbered will be $n+1$ or more, and Problem 2-16 will be useful.

2-20 Consider the outcome of the experiment to be the point where the center of the penny lands, and consider that events are sets of points with areas, and that probability is proportional to area.

2-21 B (a) This can be done by appealing to Problem 2-16 again. (b) Be sure you understand Problem 2-26.

2-22 Find out how many subsets with two or more elements there are of a set of n elements. Relate each of these subsets to an equation needed for independence.

2-23 Use Bayes' theorem, Eq. 2-44.

2-24 Use total probability, Eq. 2-41.

2-25 Draw a diagram, something like Fig. 2-12c. Note that the area of the diagonal strip equals the area of the square less the area of the corner triangles.

2-26 I, **M** First, count the different <u>sequences</u> of k distinct elements taken from a set of n elements. Then, consider how many different <u>subsets</u> of k distinct elements taken from a set of n elements there are.

2-27 Use Bayes' theorem, Eq. 2-44.

3) Chapter 3

3-1 A occurs two or more times if it does not occur zero or one time.

3-2 A simple application of Problem 3-1b.

3-3 Find the probability that seven will not show at all.

3-4 Write down the binomial theorem expansions of $(q+p)^n$ and $(q-p)^n$, then add them together.

3-5 Deduce that the number of ways to take n items from N items, so that a subset of k of the n items come from a subset of K of the N items, is the product of the number of ways to take k items from K items with the number of ways to take $n-k$ items from $N-K$ items.

3-6 Apply Problem 3-1.

3-7 First, find out how many (what range of) wins are needed to bound the amount won or lost to the amounts specified, then compute the probability of having that many wins. The unnumbered equation on p. 57, just after the "**Proof.**" heading, will be useful in speeding up the computations, if you are doing them on a hand calculator.

3-8 Deduce that having r successes in all n trials, including a success on the ith trial, is the same thing as having $r-1$ successes in the $n-1$ trials that exclude the ith trial, along with a success on the ith trial.

3-9 F The problem is not well stated. Does it ask for the probability that any one (or more) of the four players will have all 13 cards of any one suit? Restate the problem so that it asks what is the probability that a specified one of the players, will have a perfect hand.

3-10 D, **B** What does the "average duration" of such a game mean? As yet, it has no meaning. What is meant is the expected value of the total number of games. The expected value is a concept introduced in Section 5-3. You might put this problem off until the concept of an expected value, and that of a conditional expected value, are clear. Define $\boldsymbol{n}_k$ to be the random variable that is the number of games until either A or B is ruined, supposing that A starts with capital k. Deduce the total probability theorem for random variables, which in this case is (see Problem 5-27)

$$N_k = E\{\boldsymbol{n}_k\} = E\{\boldsymbol{n}_k | H\}P(H) + E\{\boldsymbol{n}_k | \overline{H}\}P(\overline{H})$$

Let H be the event that A succeeds in the next game. This will establish the hint given in the problem. Solving the difference equation of the hint is somewhat difficult. Define the increments $\Delta N_k = N_k - N_{k-1}$, find the difference equation for the increments, and note that $\Delta N_1 = N_1$, where N_1 is temporarily unknown. Use the solution for the increments (in terms of the unknown N_1) to solve for the average durations themselves, then use the fact that $N_{a+b} = 0$ to find N_1 and all the other average durations.

3-11 D, **F** The concept of stakes may not be clear. What is meant is that, on each play, A

bets an amount α, while B bets an amount β, so that A wins β, or he loses α. What remains unclear is the concept of ruin in this game. Is A ruined when he has no money left? Or, is A ruined when he has less than his stake, α, left so that he may no longer play? The latter is a more reasonable definition. Even with this clarification, the problem is too difficult to work in general. Try a specific example: let $\alpha = 2$, $\beta = 3$, and $a + b = 8$. Then change to $a + b = 9$ and see how the problem changes character.

3-12 B Again, the premature use of the term "expected," a term not properly used until Section 5-3. Here, the expected loss is the sum of the various possible losses, each multiplied by its probability.

4) Chapter 4

4-1 I, **F** This problem cannot be done unless it is assumed that $F(x)$ is invertible. See the solution for a lengthy discussion.

4-2 F Comments here are similar to those for Problem 4-1.

4-3 F (a) A typo: the reference should be to Table 4-1. This problem points up the need for better tables than Table 4-1 of the Text. The reference AS (Tables 26.5 and 26.6) is a good source. Use Eq. 4-26. (b) Another typo. See Special Note 1.

4-4 The same remarks apply here as for Problem 4-3.

4-5 No hint.

4-6 Use Eq. 4-8.

4-7 See p. 88 for the Erlang random variable.

4-8 Follow Example 4-15.

4-9 Note that $\frac{d}{dx}U(x-c) = \delta(x-c)$.

4-10 F See Special Note 1. The problem should read $\boldsymbol{x} \sim N(0, 4)$. (b) Use Eq. 2-33.

4-11 I Show $\{t|\boldsymbol{x}(t) \le x\} = \{t|t \le G(x)\}$ for G increasing.

4-12 F See Special Note 1. The problem should read $\boldsymbol{x} \sim N(1000, 400)$. (a) Use Eq. 4-1. (b) Use Eq. 4-67.

4-13 Equations which were in the Third Edition of the Text, but have been removed, are for the density and distribution function of the binomial random variable. They are

$$f(x) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \delta(x-k)$$

$$F(x) = \sum_{k=0}^{m} \binom{n}{k} p^k q^{n-k}, \quad \text{where } m \le x < m+1 \text{ defines } m \text{, given } x.$$

4-14 **F** There is a typo here. The reference is to old Eq. 4-34, which no longer exists. It said for a binomial random variable that $F(x) \approx \mathbb{G}\left(\frac{x-np}{\sqrt{npq}}\right)$. This may be deduced from (new) Eq. 4-90. Use the equation for $f(x)$ in the hint for 4-13 above for the exact result.

4-15 No hint.

4-16 **I** Use set theory reasoning.

4-17 **B** This problem is seriously out of place. You need Eq. 6-224 to see how the conditional failure rate function is related to the distribution function for the nonnegative random variable $\boldsymbol{x}$. Then use the Rayleigh density from Eq. 4-44.

4-18 Use Eq. 2-41 (total probability) with $M = \{\zeta | \boldsymbol{x}(\zeta) \le x\}$ and $\overline{M}$.

4-19 Use Eqs. 4-67 and 2-33.

4-20 **M** Use Eq. 4-80 with $M = A \cap \{\zeta | \boldsymbol{x}(\zeta) \le x_0\}$.

4-21 Use Eq. 4-80, noting that $\int_0^1 p^j (1-p)^k dp = \frac{j!k!}{(j+k+1)!}$.

4-22 Do Problem 4-21 first. Use Eq. 4-82. Leave your answer in integral form. For a numerical evaluation, see the solution.

4-23 Use Eq. 4-96.

4-24 Use Eq. 4-96.

4-25 Use Eq. 4-96. For large x, approximate $\mathbb{G}(x) \approx 1 - \frac{1}{x}\boldsymbol{g}(x)$ from Problem 4-28.

4-26 **F**, **M** First fix the typo: The system should have 1000 components, not 100. Next, note that Eq. 4-100 does not apply. Use Eq. 4-89 and Stirling's formula, Eq. 4-111, for numerical results.

4-27 Deduce that a head must have come up at the nth flip, so that $k-1$ heads must have come up in the first $n-1$ flips.

4-28 F Note the typo: the second inequality of the hint should contain a "less than" sign rather than a "greater than" sign. No hint.

4-29 Find the probability that A does not happen in n trials.

4-30 Use Eq. 4-107. To add insight, consider a slightly different problem, as well. Consider that accidents are a Poisson point process, with an average rate of 0.02 accidents per month, and ask then what is the exact probability that a driver will have 3 accidents in 100 months.

4-31 Use Eq. 4-102.

4-32 No hint.

4-33 I Deduce that if eleven does not show on the first roll, then the probability that Y will win is the same as the original (before the first roll) probability that X will win.

4-34 I (b) **M** the solution.

4-35 F, **M** It is necessary to assume $k_1 \ll k_3$ and $k_2 \ll k_3$.

4-36 Use Eq. 4-115 and 4-116.

5) Chapter 5

5-1 F See Special Note 1. Consider this a typo. The problem should say that $\mathbf{x} \sim N(5, 4)$. Use Eq. 5-18.

5-2 Use Eq. 5-18, Examples 4-12 and 5-1b.

5-3 See Example 5-3.

5-4 Use Eqs. 5-22 and 4-16.

5-5 Follow Example 5-4 to find $F_y(y)$. Differentiate to find $f_y(y)$, being careful to consider discontinuities in the distribution function.

5-6 Use the "fundamental theorem," Eq. 5-16.

5-7 Begin with Eq. 4-6, then use Eqs. 4-116 and 4-117.

5-8 Use Eqs. 5-16, 4-30, and 4-44.

5-9 Use Eq. 5-16, but remember that $F_x(x)$ may not be continuous. In part (a), there could be a discontinuity at $x = 0$. In part (b), the value $\boldsymbol{y} = 0$ corresponds to an entire range of $\boldsymbol{x}$ values.

5-10 (a) The value $\boldsymbol{y} = 0$ corresponds to an entire range of $\boldsymbol{x}$ values.

5-11 F Assume that the symmetry point of the Cauchy density—the value μ in Eq. 4-52—is zero. The symbol μ is often used for the mean value of a random variable, but a Cauchy random variable has no mean value.

5-12 No hint.

5-13 Presume a function $\boldsymbol{y} = y(\boldsymbol{x})$, where $y = y(x)$ has one root in the range $x \in (-1, 1)$ for all $y \geq 0$. Find a differential equation for $y(x)$ and solve it.

5-14 No hint.

5-15 (a) See the hint for Problem 4-13. (b) Some (individual) values of $\boldsymbol{y}$ correspond to more than one value of $\boldsymbol{x}$.

5-16 The beta distribution is defined on p. 91. Note the typo in Eq. 4-48: the symbol b should be 1.

5-17 First, correct the typo. It should be that $\boldsymbol{y} = \sqrt{\boldsymbol{x}}$. Use Eqs. 4-39 and 5-16.

5-18 Use Eqs. 5-16 and 4-39.

5-19 Use Eqs. 5-16, 4-30, and 4-43.

5-20 D Try an easier problem: Assume $f_x(t)$ is continuous at $t = 0$. Use Illustration 7, on page 134.

5-21 Use the development on p. 100 of the Text.

5-22 No hint.

5-23 Use Eq. 5-76.

5-24 Use Eq. 5-73.

5-25 I Use Eq. 4-56 for the binomial random variable. (a) Use Eq. 5-46 for a direct approach, which is somewhat tricky. Make use of the binomial theorem,

$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$, good for all integer n and all p and q. Differentiate the theorem equation with respect to the variable p. An easier approach is to take advantage of Example 5-30, Eq. 5-117, and the moment theorem, Eq. 5-115.

5-26 (a) Use the Chebyshev inequality, Eq. 5-88. (b) Use the moment theorem, Eq. 5-115.

5-27 **I** Use Eqs. 4-74, 5-44, and 5-47.

5-28 Use the Markov inequality, Eq. 5-89.

5-29 No hint.

5-30 No hint.

5-31 No hint.

5-32 **M** (a) Use $|x-m| = \begin{cases} (a-x)+(m-a) & x \le m \\ (x-a)-(m-a) & x \ge m \end{cases}$.

(b) Deduce that $\int_a^m (x-a)f(x)dx \ge 0$ for any a.

5-33 **M** Write $E\{|\mathbf{x}|\} = \frac{1}{\sigma\sqrt{2\pi}}\left\{\int_0^{\infty} xe^{-(x-\eta)^2/(2\sigma^2)} - \int_{-\infty}^{0} xe^{-(x-\eta)^2/(2\sigma^2)}\right\}$. Then substitute $z = (x-\eta)/\sigma$ in both integrals. Explicitly integrate what you can and relate the rest to $\mathbb{G}(\eta/\sigma)$.

5-34 **M** Use the basic inequality $\ln z \le z-1$.

5-35 No hint.

5-36 See the Lyapunov inequality, p. 152.

5-37 **M** (a) Either assume that $\mu = 0$ in Eq. 4-52 defining the Cauchy density, or, better, deduce a slightly different result than the problem asks. Use Cauchy's residue theorem for contour integration to find the value of the integral. Or, just use a good table of integrals, such as reference GR.

5-38 This is a workhorse problem. (a) **F** Note the typo: The characteristic function should read $\Phi(\omega) = (1-j\beta\omega)^{-\alpha}$. (See Table 5-2.) (c) See Problem 5-25. (d) **F** The section on the negative binomial random variable is confusing. Two different distributions share this name. One is described by Eq. 4-62 or 4-63, and the other is described by Eq. 4-64. This duality is reinforced by Table 5-2. Comparing the Table with the answer wanted in this problem leads to the conclusion that Eq. 4-64 should be chosen here. While the problem does not request it, it is illustrative to find $E\{\boldsymbol{x}\}$, to see how sums may be manipulated in the same ways as integrals.

5-39 Use Eqs. 5-113 and 5-116.

5-40 Note that $(1-y)^{-n} = \sum_{k=0}^{\infty}\binom{-n}{k}(-y)^k = \sum_{k=0}^{\infty}\binom{n+k-1}{k}y^k$. See the solution for more on binomial coefficients.

5-41 **M** It is clear from the problem that we are discussing the negative binomial random variable of Eq. 4-63. Techniques developed for Problem 4-35 will be useful here.

5-42 Note that $e^{s(\boldsymbol{x}-\eta)} = \sum_{k=0}^{\infty}\frac{s^k(\boldsymbol{x}-\eta)^k}{k!}$.

5-43 Split the integral in the hint in the Text into real and imaginary parts.

5-44 Extend the development following Eq. 5-111 to higher derivatives.

5-45 No hint.

5-46 No hint.

5-47 **I, M** Consider, for small ε, $E\{g(\boldsymbol{x}-\eta-\varepsilon)\}$, by expanding the function $f(\eta+x+\varepsilon)$ in a Taylor series in ε about the point $\eta + x$. Use the obvious symmetry and maximum and limiting properties of $g(x)$ and $f(\eta + x)$ that are deduced from the graphs.

5-48 (a) **D** Regard the density of $\boldsymbol{x}$ as a function of both x and v. Show that $\frac{\partial^2 f(x,v)}{\partial x^2} = 2\frac{\partial f(x,v)}{\partial v}$. Integration by parts will then prove the theorem. (b) Use part (a) with $g(\boldsymbol{x}) = \boldsymbol{x}^n$.

5-49 **I** Use the fundamental theorem of Fourier series.

5-50 **F** The description of the experiment is confusing. What is the length of the run? The problem assumes that the first tossing is not included in the run, so that the run may end on the tossing following the first tossing, in which case $\boldsymbol{x} = 1$, not 2. Note also that the p.m.f. refers to the moment function?

5-51 How is it possible that two items are identical, yet one is defective and one is not? Any difference makes them not identical. Perhaps they are merely similar. Forget this cavil. (b) **F, D** There is a typo: the equation line should end with $\min(M, n)$ and not with $\min(M, N)$. See Table 5-2 for the answers here. To compute the expected value of $\boldsymbol{x}$ you will need Vandermonde's identity, which is

$$\binom{n+m}{k} = \sum_{j=0}^{k}\binom{n}{j}\binom{m}{k-j}$$

The computation of $\mathrm{Var}(\boldsymbol{x})$ is even more difficult; a reference is provided.

(c) Notice that, as $M, N \to \infty$ while $M/N = p$ is fixed, $\dfrac{M!}{(M-k)!}\dfrac{(N-k)!}{N!} \to \left(\dfrac{M}{N}\right)^k = p^k$

5-52 (a) To get the r'th white ball on the k'th draw implies that $r-1$ white balls are drawn in the first $k-1$ draws, and a white ball is drawn on the k'th draw.

(b) **D** Consider the number of ways of ordering all the balls. Find out how many of these ways are favorable to the desired outcome.

(c) Find the limit of the result in part (b) in a manner similar to that used in Problem 5-51c.

6) Chapter 6

6-1 A workhorse problem, but parts have already been done in examples. (a) Use Eq. 6-45. (b) Use Eq. 6-54. (c) **D** The Text omits the prime example where $\boldsymbol{z} = \boldsymbol{xy}$. You have three options: 1) Work out this case generically, using Example 6-10 as a point of departure. 2) Work out this case for positive random variables only, where it simplifies somewhat. 3) Skip ahead to Eq. 6-148. Once the generic result is established, the specific result will be an integral difficult to evaluate. Leave the answer in integral form. (d) Use Eq. 6-60. (e) Use Eq. 6-82. (f) Use Eq. 6-79. (g) Use Eq. 6-84.

6-2 (a) Use Eq. 6-60. (b) One way to work this is to notice that $\{\boldsymbol{z} \le z\} = \left\{\dfrac{\boldsymbol{x}}{\boldsymbol{y}} \ge \dfrac{1-z}{z}\right\}$, so if $\boldsymbol{w} = \dfrac{\boldsymbol{x}}{\boldsymbol{y}}$, and $F_w(w)$ is computed from Eq. 6-60, then $F_z(z) = 1 - F_w\left(\dfrac{1-z}{z}\right)$. (c) Consider graphically (as a function of z) where $|x-y| \le z$ inside the square $0 \le x \le a$, $0 \le y \le a$ in the xy-plane. Figure 6-13 is a starting point.

6-3 Begin with a diagram of the region of integration—the graphical approach.

6-4 Why is this a problem at all? Part (a) is Example 6-15; part (b) is example 6-14; and part (c) is obvious from the discussion of joint normality on p.202.

6-5 See Problem 6-1, part (c), or go directly to Eq. 6-148.

6-6 Use Eq. 6-148.

6-7 (a) Use a graphical approach. (b) Use Eq. 6-148. (c) Use Eq. 6-60. (d) Back to a graphical approach.

6-8 Use a graphical approach.

6-9 (a) Eq. 6-60 could be used, but it is easier to use a graphical approach. (b) Use Eq. 6-148.

6-10 The graphical approach is best. Be sure to get the correct triangle in the xy-plane. It has unit area.

6-11 This would be a challenging problem, except that it is (mostly) done in Example 6-27. There, the claim is made about the marginal densities of $\boldsymbol{x} + \boldsymbol{y}$ and $\boldsymbol{x}/\boldsymbol{y}$ that is not quite demonstrated. The best way to attack part (a) is to use the Convolution Theorem on p. 216 along with the characteristic function for a gamma random variable from Table 5-2. (b) may then be attacked with Example 6-27. (c) From Problem 6-2(b) we learned that if $\boldsymbol{u} = \boldsymbol{x}/(\boldsymbol{x}+\boldsymbol{y})$ and $\boldsymbol{w} = \boldsymbol{y}/\boldsymbol{x}$, then $F_u(u) = 1 - F_w\left(\frac{1-u}{u}\right)$.

6-12 Use Eq. 6-119, plus a graphical approach to find where the density is nonzero in the zw-plane.

6-13 Use Eqs. 4-44 and 6-60.

6-14 Use Eq. 6-43.

6-15 Use Eq. 6-43.

6-16 Do not attempt to use the function $g^{-1}(.)$; it may not exist.

6-17 (a) **I**, **M** Use Eq. 6-43 to work the general problem of the sum of independent normal random variables with zero means. (b) Use Eq. 6-62.

6-18 **I**, **M** Start with Eq. 6-148 and reverse the roles of x and y.

6-19 Use Eq. 6-59. The development of the third absolute moment of a zero-mean normal random variable on p. 148 will be useful in evaluating the integral.

6-20 (a) Use Eqs. 5-18 and 6-43. (b) Use Eq. 6-54. (c) Use Eq. 6-60. (d) Use the equation following Eq. 6-78. (e) Use (part of) Example 6-18.

6-21 M First solve the general problem of finding the density of $\boldsymbol{z} = |\boldsymbol{x} - \boldsymbol{y}|$.

6-22 (a) **I**, **M** Compare with Problem 6-17. (b) **I**, **M** Use Eq. 6-43. The integral here is evaluated with Cauchy's residue theorem.

6-23 Use Eqs. 4-39, 5-18, 5-60, and 4-35.

6-24 This is simply Example 6-21 in very thin disguise.

6-25 F See Special Note 2. There is confusion here equivalent to a typographical error. The easiest way to rectify this is to change the problem to find the probability that the combined lifetime exceeds $2/\lambda$ (instead of 2λ) and the probability that the excess lifetime of the second bulb over that of the first exceeds $1/\lambda$ (instead of λ).

6-26 (a) Note that $\boldsymbol{r} = |\boldsymbol{x} - \boldsymbol{y}|$, and Problem 6-2c applies. (b) Note that $\boldsymbol{s} = \boldsymbol{x} + \boldsymbol{y}$.

6-27 (a) Note that $\boldsymbol{z} = \boldsymbol{y}/\boldsymbol{x}$ if $\boldsymbol{y} < \boldsymbol{x}$, or $\boldsymbol{z} = 1$ if $\boldsymbol{y} \geq \boldsymbol{x}$, and be sure to explicitly consider the discontinuity at $z = 1$. (b) Use an approach similar to that for part (a).

6-28 See the hint for Problem 6-11c. Problem 6-1d may also be useful.

6-29 Use Example 6-18 to find $f_z(z)$. Note that $\boldsymbol{w} = |\boldsymbol{x} - \boldsymbol{y}|$, and, starting with Problem 6-2c, find $f_w(w)$. Then use Eq. 6-115 to find $f_{zw}(z, w)$. Finally, test Eq. 6-20.

6-30 (a) Define $\boldsymbol{u} = \boldsymbol{x} + \boldsymbol{y}$. Use Eq. 6-45. Consider two cases: $0 < u < \beta$ and $\beta < u < 2\beta$. In the first case, use the definition of the beta function, Eq. 4-49, to simplify the result. In the second case, simplification is not practical, so leave the result in integral form. (b) Start from Example 6-21. (c) Find the joint density of $\boldsymbol{v}$ and $\boldsymbol{w}$ from Eq. 6-115. Then, from the joint density, find the marginal densities, and show that the joint density is the product of the marginal densities.

6-31 A workhorse problem. (a) Make use of Examples 6-27 and 6-12. Problem 6-11 comes close. (b) Use Eq. 6-115 to find the joint density and show that it is the product of the marginal densities from part (a). (c) Use Theorem 6-1.

6-32 (a) Let $\boldsymbol{z} = \boldsymbol{x}/|\boldsymbol{y}|$. Use a graphical approach to develop an iterated integral expression for $F_z(z)$, as in Example 6-10. Differentiate the double integral expression with respect to z to get a single integral expression for $f_z(z)$, also as in the cited example, then solve the integral. Repeat this process for $\boldsymbol{w} = |\boldsymbol{x}|/|\boldsymbol{y}|$. (b) Use Eq. 6-43 to show that $\boldsymbol{u} \sim N(0, 2)$. From Example 6-14 see that $\boldsymbol{v}$ is exponential with parameter $1/2$. Then use Eq. 6-115 to find the joint density, and see if it is the product of the marginal densities.

6-33 Use Eq. 6-121.

6-34 (a) & (b) **M** Transform first to $\boldsymbol{r} = \sqrt{\boldsymbol{x}^2 + \boldsymbol{y}^2}$ and $\theta = \tan^{-1}(\boldsymbol{y}/\boldsymbol{x})$, using Example 6-22. Then transform from $\boldsymbol{r}$ and θ to $\boldsymbol{u}$ and $\boldsymbol{v}$ to find the joint p.d.f. (c) Express this random variable in terms of $\boldsymbol{u}$ and $\boldsymbol{v}$.

6-35 Eq. 6-157 defines the F distribution (see Special Note 3). Use Example 5-10. (b) Use Eq. 5-16 to find the p.d.f. of $\boldsymbol{w} = m\boldsymbol{z}/(m\boldsymbol{z} + n)$, and compare the result to Eq. 4-48.

6-36 **F** Use Eq. 6-115 to find the joint density of $\boldsymbol{z}$ and $\boldsymbol{w}$, then use Eq. 6-10 to find the marginal density of $\boldsymbol{z}$, which will reveal that $\boldsymbol{z}$ is not, in fact, exponential. Go on to find the marginal density of $\boldsymbol{w}$, which is exponential.

6-37 Use Eq. 6-115 to find the joint density of $\boldsymbol{z}$ and $\boldsymbol{w}$, then find the marginal densities from the joint density.

6-38 Use Eqs. 5-33 and 6-148.

6-39 First let $\boldsymbol{v} = a\cos\boldsymbol{y}$; thenUse Eqs. 5-33 and 6-43.

6-40 No hint.

6-41 Use Eqs. 6-43 and 6-148.

6-42 (a) Use the result in Problem 6-40. (b) Let $\boldsymbol{w} = \boldsymbol{x} - \boldsymbol{y}$, and develop $P\{\boldsymbol{w} = n\}$. Treat the cases $n < 0$ and $n \geq 0$ separately.

6-43 This problem is somewhat confused. The given information is that two conditional probabilities are equal to $1/(k+1)$. It is only necessary to assume that the two conditional probabilities are equal; it is then possible to show that they must equal $1/(k+1)$.

6-44 Way 1: Proceed directly, using Problem 6-42 and the identity $\sum_{j=0}^{k} \binom{n}{j}\binom{m}{k-j} = \binom{n+m}{k}$, or Way 2: Use the moment generating functions of Eq. 5-117.

6-45 (a) Define $\boldsymbol{z} = \min(\boldsymbol{x}, \boldsymbol{y})$ and $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{y}$. Find the joint p.m.f. of $\boldsymbol{z}$ and $\boldsymbol{u}$, then find the two marginal p.m.f.s from the joint p.m.f., and show that the joint p.m.f. is the product of the marginal p.m.f.s. (b) Repeat the procedure of part (a) for $\boldsymbol{z}$ and $\boldsymbol{w}$.

6-46 A direct approach is workable here. Use Eqs. 4-57 and 4-56, along with the results of Problem 6-40.

6-47 **I** Use Example 6-30 to relate μ_{12} and μ_{21} to σ_1, σ_2, and r.

6-48 While it is possible to find $f_z(z)$, where $\boldsymbol{z} = \boldsymbol{xy}$ and work the problem in this way, that is unnecessarily difficult. Instead, just focus on the two quadrants of the xy-plane where $\boldsymbol{z}$ is negative, and use the independence of $\boldsymbol{x}$ and $\boldsymbol{y}$ to compute their probabilities.

6-49 Let $\boldsymbol{w} = \boldsymbol{x} - \boldsymbol{y}$, then use Eq. 5-74.

6-50 Use Eq. 6-159.

6-51 **I**, **M** A fundamental problem. (a) This is the Schwarz inequality for an inner product space. Those familiar with linear algebra know the standard trick to demonstrate it. Show first that $|E\{\boldsymbol{xy}^*\}| \le \sqrt{E\{|\boldsymbol{x}|^2\}E\{|\boldsymbol{y}|^2\}}$. Start with $E\{|a\boldsymbol{x} - \boldsymbol{y}|^2\} \ge 0$, and pick the arbitrary constant a artfully. (b) This is the triangle inequality. Note that $E\{\boldsymbol{xy}^*\} + E\{\boldsymbol{x}^*\boldsymbol{y}\} = 2\Re(E\{\boldsymbol{xy}^*\})$, where $\Re$ denotes the real part, and $\Re(E\{\boldsymbol{xy}^*\}) \le |E\{\boldsymbol{xy}^*\}|$, then use part (a).

6-52 **F**, **M** You cannot show that $\boldsymbol{y} = a\boldsymbol{x} + b$ for each and every outcome of the experiment, for it may not be true. You can show that the set of outcomes for which this is true has probability one. Use Eq. 5-89, the Markov inequality.

6-53 This result follows from Problem 6-52.

6-54 **M** There are two ways to proceed here, one more direct than the other. For the direct method, begin with $f_n(n) = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\delta(n-k)$. For the less direct method, begin by using the results of Problem 5-37 to find $\Phi_x(\omega)$, and then find $\Phi_z(\omega)$ by using Eq. 6-242.

6-55 Note that $\boldsymbol{y}$ is a function of $\boldsymbol{x}$, so that $\boldsymbol{z}$ may be written as a function of $\boldsymbol{x}$ alone.

6-56 Use Eqs. 5-100, 5-99, and 6-194.

6-57 Characteristic functions provide the easier approach here. With effort and care, the direct approach is also workable.

6-58 All basic stuff.

6-59 Use Eqs. 5-100, 5-119, 6-193, and 6-194.

6-60 Use Eq. 6-159 directly.

6-61 (a) This is Problem 6-4. Another method is available now: Use Eq. 6-55. (b) Use Eq. 6-205. (c) Use Eq. 6-44.

6-62 It helps to know here that $\Gamma(1/2) = \sqrt{\pi}$. Use the string of equations: 4-39, 5-20, 4-25, 6-205, 6-10, and 4-52.

6-63 **I** (a) Use Eqs. 6-179 and 6-166. (b) **D** A continuation or extension of Problem 6-51. See the solution.

6-64 Use Example 6-41.

6-65 **M** Rederive the result in note 4 on p. 337 of the Text. Fix the typo there, where the second equation should read $\mathrm{Var}\{E\{\boldsymbol{x}|\boldsymbol{y}\}\} = E\{(E\{\boldsymbol{x}|\boldsymbol{y}\})^2\} - (E\{E\{\boldsymbol{x}|\boldsymbol{y}\}\})^2$. Use Eq. 6-241 as a part of this derivation.

6-66 Use Eq. 6-180.

6-67 **F** The correct result is: $E\{\boldsymbol{z}\} = \Sigma_n E\{g(x_n, \boldsymbol{y})|x_n\}p_n$. Define $f(y|x_n) = f_n(y)$, and see that $f(x, y) = \Sigma_n p_n f_n(y)\delta(x - x_n)$ to begin the solution.

6-68 (a) Use Eqs. 6-211, 6-26, 5-55, and 4-29. (b) Use Eqs. 6-238 and 6-243.

6-69 **M** Let $I(\mu) = E\{|\boldsymbol{xy}|\}$ and use Price's theorem (Eqs. 6-200 and 6-201) to show that $\frac{\partial I(\mu)}{\partial \mu} = E\{\mathrm{sgn}\boldsymbol{x}\ \mathrm{sgn}\boldsymbol{y}\}$, then use Eq. 6-64 to find this explicitly. Also find $I(0)$ explicitly. Then use $I(\mu) = I(0) + \int_0^{\mu} \frac{\partial I(\mu)}{\partial \mu} d\mu$.

6-70 See Example 6-41.

6-71 Use the unnumbered equation just above Example 4-45.

6-72 Consider $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$, and $\boldsymbol{w} = \boldsymbol{x}$ as a variable transformation.

6-73 To simplify the computation of the Jacobian of the transformation from $\boldsymbol{x}, \boldsymbol{y}$ to $\boldsymbol{z}, \boldsymbol{w}$, note that $\frac{\partial z}{\partial y} = 0$, so that $\frac{\partial w}{\partial x}$ is unimportant.

6-74 Use Bayes' theorem, Eq. 2-44.

6-75 **M** Start with Eq. 6-152, and use Eq. 5-55. Then make some inspired substitutions in the integral to get to the definition of the beta function, Eq. 4-49. Use Eq. 4-51 to relate to the gamma function, then Eq. 4-36 to evaluate the gamma functions. Note that the result is only valid, of course, for $n > 2$.

6-76 Note the typo: the expression $\beta_y(t|\mathbf{y} > t)$ should read $\beta_y(t) = f_y(t|\mathbf{y} > t)$. Use the unnumbered equation prior to Eq. 6-225.

6-77 Use the Markov inequality, Eq. 5-89.

6-78 Use Eq. 6-159.

6-79 Use Eq. 6-227, Example 6-38, and Eqs. 6-228 and 6-205.

7) Chapter 7

7-1 Rewrite the probability $P\{x_1 < \mathbf{x} < x_2,\ y_1 < \mathbf{y} < y_2,\ z_1 < \mathbf{z} < z_2\}$ in terms of $F(x, y, z)$.

7-2 The useful term "zero-one random variable" was removed from the Fourth Edition, except for this problem, but see Example 4-8. Use Problem 2-15 to show that the zero-one random variables are independent if and only if the events they are associated with are independent.

7-3 **I** Extend the argument in the Text that results in Eqs. 7-57 and 7-58 to the case with nonzero means to find $f(x, y, z)$.

7-4 Expand the polynomial and consider each term.

7-5 (a) Extend the argument leading to Eq. 6-31 to three dimensions. (b) See the bottom of page 259; ultimately use Eq. 5-78.

7-6 **I**, **M** Construct the covariance matrix; it must have a nonnegative determinent.

7-7 **D** Keeping track of the random variables here is hard. Define the three functions: $g(x_3) = E\{\mathbf{x}_1\mathbf{x}_2|x_3\}$, $h(x_2, x_3) = E\{\mathbf{x}_1\mathbf{x}_2|x_2, x_3\}$, $p(x_3) = E\{h(\mathbf{x}_2, \mathbf{x}_3)|x_3\}$. Show that $g(x_3) = p(x_3)$, and interpret.

7-8 **M** Solve for a_1 and a_2 in terms of the expectations of $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{y}$. Let $\hat{E}\{\mathbf{y}|x_1\} = bx_1$, and solve for b in terms of the expectations of $\mathbf{x}_1$, and $\mathbf{y}$. Let $\hat{E}\{a_1\mathbf{x}_1 + a_2\mathbf{x}_2|x_1\} = cx_1$, and solve for c in terms of a_1 and a_2 and expectations involving $\mathbf{x}_1$ and $\mathbf{x}_2$. Then show $b = c$.

7-9 Use $E\{\mathbf{s}^2\} = E\{E\{\mathbf{s}^2|\mathbf{n}\}\}$. Use Problem 6-51a to limit $(E\{\mathbf{x}_i\mathbf{x}_j\})^2$. The fact that

$\mathbf{x}_i \geq 0$ is not needed.

7-10 M Let event A_n be that heads <u>first</u> appears on the n'th tossing. Note that $[A_1, A_2, \ldots]$ is a partition. Use total probability to find $E\{\mathbf{x}_1\}$. Next, establish $E\{\mathbf{x}_m|x_{m-1}\}$ by a similar argument, where event B_n is that the first appearance of heads after toss x_{m-1} is on toss $(x_{m-1} + n)$. Finally, use $E\{\mathbf{x}_m\} = E\{E\{\mathbf{x}_m|\mathbf{x}_{m-1}\}\}$.

7-11 Use $\Phi(\omega) = E\{e^{j\omega\mathbf{m}}\} = E\{E\{e^{j\omega\mathbf{m}}|\mathbf{n}\}\}$ to find the characteristic function of $\mathbf{m}$ (using the hint in the Text). Compare the result with the characteristic function of a Poisson process, obtained from Eq. 5-119.

7-12 Use the semi-discrete version of Eq. 6-213 (i.e., $f(s) = \Sigma_n f(s|n)p_n$), plus Example 7-1 and Eq. 7-5 (generalized) to find $f(s|n)$.

7-13 No hint.

7-14 Use Example 7-2.

7-15 Note that the event $\{z < \mathbf{z} \leq z + dz, w < \mathbf{w} \leq w + dw\}$ happens if the events $\{\mathbf{x} \leq w\}$ and $\{\mathbf{x} > z + dz\}$ do not happen <u>and</u> the events $\{w < \mathbf{x} \leq w + dw\}$ and $\{z < \mathbf{x} \leq z + dz\}$ happen once, <u>and</u> the event $\{w + dw < \mathbf{x} \leq z\}$ happens $(n-2)$ times.

7-16 This follows directly from Eq. 5-96, with independence.

7-17 **M** Show that $(\bar{\mathbf{x}}, \mathbf{x}_1 - \bar{\mathbf{x}}, \ldots, \mathbf{x}_n - \bar{\mathbf{x}})$ are jointly normal and that $\bar{\mathbf{x}}$ is uncorrelated with each $\mathbf{x}_i - \bar{\mathbf{x}}$. Hence, $\bar{\mathbf{x}}$ is independent of the group $(\mathbf{x}_1 - \bar{\mathbf{x}}, \ldots, \mathbf{x}_n - \bar{\mathbf{x}})$, and $\bar{\mathbf{x}}$ is independent of $\mathbf{s}^2$.

7-18 Use Eq. 7-87 to determine α_0, α_1, and α_2. Use Eq. 7-83 to determine the α_1 and α_2 in $\hat{E}\{\mathbf{s} - \eta_s|\mathbf{x}_1 - \eta_1, \mathbf{x}_2 - \eta_2\}$, after you figure out just what this is. See Eq. 7-90.

7-19 Denote $\hat{\mathbf{y}} = \hat{E}\{\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2\} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$. Now $\hat{E}\{\hat{E}\{\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2\}|\mathbf{x}_1\} = \hat{E}\{\hat{\mathbf{y}}|\mathbf{x}_1\} = a\mathbf{x}_1$. Also, $\hat{E}\{\mathbf{y}|\mathbf{x}_1\} = b\mathbf{x}_1$. The problem here is to show that $a = b$, which follows from the orthogonality principle.

7-20 **M** Use Example 7-2 for $F_x(x)$. Use Problem 7-14 for $F_y(y)$. Use Example 6-21 to obtain $F_{xy}(x, y)$.

7-21 **I**, **D** No conceptual difficulties here, but this is a tedious problem. First deduce that

$\sigma_{\bar{v}}^2$ is unchanged if the means of the $\boldsymbol{x}_i$ s are varied, and thus assume that the $\boldsymbol{x}_i$ s have zero means. Use brute force to find, in order

$$E\{\boldsymbol{x}_i\},\ E\{\boldsymbol{x}_i\boldsymbol{x}_j\},\ E\{\boldsymbol{x}_i^2\boldsymbol{x}_j^2\},\ E\{\bar{\boldsymbol{x}}\boldsymbol{x}_i\},\ E\{\bar{\boldsymbol{x}}^2\},\ E\{\boldsymbol{x}_k\boldsymbol{x}_i^2\boldsymbol{x}_j\},\ E\{\bar{\boldsymbol{x}}\boldsymbol{x}_i^2\boldsymbol{x}_j\},\ E\{\bar{\boldsymbol{x}}^2\boldsymbol{x}_i\boldsymbol{x}_j\},$$
$$E\{\bar{\boldsymbol{x}}^3\boldsymbol{x}_i\},\ E\{\bar{\boldsymbol{x}}^4\}$$

Expand $E\{\bar{\boldsymbol{v}}^2\}$ in terms of the above and solve. Note that $\sigma_{\bar{v}}^2 = E\{\bar{\boldsymbol{v}}^2\} - \sigma^4$.

7-22 Use Problem 6-49.

7-23 D Note that for (not necessarily square) matrices A and B, where the product AB is square, that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, where $\mathrm{tr}(\ )$ is the matrix trace function. For a scalar product x, $\mathrm{tr}(x) = x$. Another approach is to note that $R = ADA^t$, where $A^t = A^{-1}$ and D is a diagonal matrix with positive diagonal entries. Define $\boldsymbol{Y} = \boldsymbol{X}AD^{-1/2}A^t$, and show that $E\{\boldsymbol{Y}\boldsymbol{Y}^t\} = E\{\boldsymbol{X}R^{-1}\boldsymbol{X}^t\} = \mathrm{tr}(E\{\boldsymbol{Y}^t\boldsymbol{Y}\})$.

7-24 F This problem cannot be solved as set; more information about the $\boldsymbol{x}_i$ s is needed. Assume the sequence $\{\boldsymbol{x}_i\}$ is such that the central limit theorem holds. Among other things, this assures that $\sigma_1^2 + \ldots + \sigma_n^2 \to \infty$ as $n \to \infty$. Use the discussion prior to Eq. 5-35.

7-25 M If you are a mathematician, there is no trouble here. If not, this problem is dificult because you likely have little experience with "ε, N" limit proofs. You must show that, given any $\varepsilon > 0$, there exists an N such that $E\{|\boldsymbol{x}_n - a|^2\} < \varepsilon$ for all $n \geq N$. Since $a_n \to a$, we know that, given any ε_1, there exists an N_1 such that $|a_n - a| < \varepsilon_1$ for all $n \geq N_1$. Since $E\{|\boldsymbol{x}_n - a_n|^2\} \to 0$, we know that, given any ε_2, there exists an N_2 such that $E\{|\boldsymbol{x}_n - a_n|^2\} < \varepsilon_2$ for all $n \geq N_2$. Use the triangle inequality, Problem 6-51b, to relate what is known to what is needed.

7-26 F, **M** Prove only that if the limit of $E\{\boldsymbol{x}_n\boldsymbol{x}_m\}$ exists, then $\boldsymbol{x}_n$ converges in the mean square. The converse cannot be shown.

7-27 Use the Cauchy criterion, Eq. 7-117. Define $\alpha_n = \sum_{k=1}^{n} \sigma_k^2$. Show that $\{\alpha_n\}$ converges, and use this fact to show that $\{\boldsymbol{y}_n\}$ converges in the mean square sense.

7-28 Let $f_n(y)$ be the density of $\boldsymbol{y}_n = \boldsymbol{y}_{n-1} + \boldsymbol{x}_n$. Use Eq. 6-43 to relate $f_n(\ )$ to $f_{n-1}(\ )$ and to $f_x(\ )$. Solve for the first few densities, compare with the Erlang density (Eq. 4-37) and guess the general result. Then confirm the general result by induction.

7-29 Use Problem 7-28 and the central limit theorem.

7-30 No hint.

7-31 Use Problem 6-22b to show that the sum of Cauchy random variables is Cauchy and never becomes gaussian.

7-32 Note that $\boldsymbol{x}$ and $\boldsymbol{y}$ are presumed to be normal. Use Goodman's theorem, Eq. 7-62, even though $\boldsymbol{x}$ and $\boldsymbol{y}$ are scalars.

8) Chapter 8

Chapter 8 is omitted here.

9) Chapter 9

9-1 (a) From Eqs. 4-74, 5-44, and 5-47, deduce that if $[A_1, \ldots, A_n]$ is a partition, then $E\{\boldsymbol{x}\} = E\{\boldsymbol{x}|A_1\}P(A_1) + \ldots + E\{\boldsymbol{x}|A_n\}P(A_n)$. (b) Since $\boldsymbol{x}(t)$ has only two values (for any fixed t), then $P\{\boldsymbol{x}(t) \le x\}$ can have only three different values, including zero and one.

9-2 Find $F(x, t)$ and differentiate to get $f(x, t)$.

9-3 See Example 9-5. In part (c), notice that an event like $\{\boldsymbol{x}(2) = 2, \boldsymbol{x}(4) = 4\}$ is the same as the event $\{\boldsymbol{x}(4) - \boldsymbol{x}(2) = 2, \boldsymbol{x}(2) = 2\}$, and that the two sub-events that make up this latter form are independent, as they count Poisson points in nonoverlapping intervals. The computation is tedious.

9-4 (b) **M** See: Papoulis, A., *The Fourier Integral and Its Applications*, McGraw Hill, 1962, Appendix I, Eq. I-29.

9-5 Use Eq. 6-63.

9-6 **I**, **M**, **F** You need to show that $\boldsymbol{w}(t) = \int_0^t (t-s)\boldsymbol{v}(s)ds$, for $t \ge 0$. See the solution for how this is done. Note the typo. The correct result is $E\{\boldsymbol{w}^2(t)\} = \int_0^t (t-s)^2 q(s)ds$.

9-7 (a) Use inequality 5-89. (b) Let $\boldsymbol{x}_1 = \boldsymbol{x}(t+\tau)$, $\boldsymbol{x}_2 = \boldsymbol{x}(t)$. Find the region D_x of the

$x_1 x_2$-plane where $|x_1 - x_2| \geq a$, and integrate the second-order density over D_x.

9-8 No hint.

9-9 Do not forget the complex conjugate in Eq. 9-30.

9-10 Use Eq. 7-61.

9-11 **B**, **I**, **M** The Text does not fairly prepare you to solve this problem. In the discussion of linear, constant-coefficient, differential equations, beginning on page 404, the Text notes that such equations are not uniquely solvable without initial conditions. To assure a unique solution and to assure the linearity condition (Eq. 9-86) is satisfied, the Text says that it will presume a solution with "zero" initial conditions at $t = 0$. The trouble with this approach is that the differential equation then only holds for $t \geq 0$, and there is no way that $\mathbf{y}(t)$ can be WSS. By clear implication of this problem, the equation is to hold for all time, and $\mathbf{y}(t)$ will be WSS.

The way to accomplish these goals together is tricky. Suppose we set some arbitrary initial conditions at $t = t_0$, and presume that the differential equation holds for $t \geq t_0$. Now we let $t_0 \to -\infty$. This approach accomplishes all the objectives, provided that the differential equation has a (unique) solution under such assumptions, and the effect of the initial conditions at t_0 on the solution at some fixed t declines to zero as $t_0 \to -\infty$.

This is true for <u>stable</u> differential equations. The differential equation

$$a_n y^{(n)}(t) + a_{n-1} y^{(n-1)}(t) + \ldots + a_1 y'(t) + a_0 y(t) = x(t)$$

is stable if and only if the associated polynomial equation in s

$$a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0 = 0$$

has only roots with negative real parts (roots in the left half of the s-plane). When this is true, we define the Laplace version of the system function as

$$H(s) = \frac{1}{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0}$$

The Fourier version of the system function is given by $H(\omega) = H(j\omega)$, and the impulse response function of this linear system is the inverse Fourier transform of the system function, or

$$h(t) \leftrightarrow H(\omega), \quad h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(\omega)e^{j\omega t}d\omega$$

Also, in this problem assume $E\{\boldsymbol{v}(t)\} = 0$. To make things more manageable, define the zero-mean process $\boldsymbol{z}(t) = \boldsymbol{y}(t) - 2$, and begin work with it.

9-12 Just note that $f(t_1)g(t_2)\delta(t_1 - t_2) = f(t_1)g(t_1)\delta(t_1 - t_2) = f(t_2)g(t_2)\delta(t_1 - t_2)$.

9-13 **M** If $\varphi(\tau)$ is the phase angle of $R_{xy}(\tau)$, consider $E\{|\boldsymbol{x}(t+\tau) \pm e^{j\varphi(\tau)}\boldsymbol{y}(t)|^2\}$.

9-14 Show that $\boldsymbol{x}(t) = \boldsymbol{y}(t)$ in the mean square sense. Then use the Schwarz inequality, Eq. 9-176, twice to show $R_{xx}(\tau) = R_{xy}(\tau)$ and $R_{yy}(\tau) = R_{xy}(\tau)$.

9-15 No hint.

9-16 Presume φ to be real. Use $\Phi(1) = 0$ to establish $E\{\cos\varphi\} = E\{\sin\varphi\} = 0$, and use $\Phi(2) = 0$ to establish $E\{\cos 2\varphi\} = E\{\sin 2\varphi\} = 0$. Then show $\eta(t) = 0$, and $R(t_1, t_2) = \frac{1}{2}\cos[\omega(t_1 - t_2)]$.

9-17 (a) **B** There is a missing definition here. The stochastic process $\boldsymbol{x}(t)$ has *orthogonal increments* if, for $t_a \le t_b \le t_c \le t_d$, $E\{[\boldsymbol{x}(t_d) - \boldsymbol{x}(t_c)][\boldsymbol{x}(t_b) - \boldsymbol{x}(t_a)]\} = 0$. Using this definition, substitute $t_d = t_2$, $t_c = t_b = t_1$, and $t_a = 0$. (b) **F** You need to show that $\eta_y(t)$ is constant, and you cannot do this unless you assume that $\eta_x(t)$ is constant. Use part (a) to find $R_{yy}(t_1, t_2)$ as a function of $|t_1 - t_2|$.

9-18 No hint.

9-19 See Problem 5-14.

9-20 **I** Note that $f_y(y_1, \ldots, y_n; t_1, \ldots, t_n) = \int_{-\infty}^{\infty} f_{y|\varepsilon}(y_1, \ldots, y_n; t_1, \ldots, t_n|\varepsilon)f_\varepsilon(\varepsilon)d\varepsilon$, and that $f_{y|\varepsilon}(y_1, \ldots, y_n; t_1, \ldots, t_n|\varepsilon) = f_x(y_1, \ldots, y_n; t_1 - \varepsilon, \ldots, t_n - \varepsilon)$.

9-21 Use Eqs. 9-101 and 9-106.

9-22 (b) Note that $f_x(x, t) = f_x(x)$. Note that $\boldsymbol{z}$ and $\boldsymbol{w}$ are jointly normal. Use Eq. 9-46 to find their correlation coefficient.

9-23 **F** You must assume here that $\boldsymbol{x}(t)$ is WSS (or, equivalently, that $\boldsymbol{x}(t)$ has a constant mean). The discussion following Eq. 9-87 assures that $\mathbf{x}'(t)$ is normal. Eq. 9-106 will help to find $\sigma_{x'}^2$.

9-24 **M** See the hint in the problem, which is: Use Eq. 9-80, and establish the Fourier series $\sin^{-1} z = \sum_{n=1}^{\infty} \frac{1}{n}[J_0(n\pi) - (-1)^n]\sin n\pi z$.

9-25 Presume, as is implied by the problem, that $\mathbf{x}(t)$ is WSS. There are two ways to proceed. One way is direct, but algebraically demanding. The other way is less direct, but easier. The direct method simply relies on the relationships

$$E\{g(\mathbf{x}(t))\} = \int_{-\infty}^{\infty} g(x)f(x, t)dx$$

$$E\{g(\mathbf{x}(t_1), \mathbf{x}(t_2))\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2)f(x_1, x_2; t_1, t_2)dx_1dx_2$$

which are obvious extensions of Eqs. 9-7 and 9-8. The first- and second-order density function are obtained from the fact that $\mathbf{x}(t)$ is normal. The integration is difficult, but doable. The less direct method relies on the close connection between the desired expected values and the first- and second-order characteristic functions of $\mathbf{x}(t)$, which are fully developed for a normal process in Eqs. 5-100 and 6-195. In either case, one must note that $\sigma^2(t) = R_x(0)$, and $r(t+\tau, t) = r(\tau) = R_x(\tau)/R_x(0)$.

9-26 (b) **F** Assume $\tau R_x(\tau) \to 0$ as $\tau \to 0$, (not just $R_x(\tau) \to 0$, as in the problem statement).

9-27 Write $\mathbf{y}(t_1)\mathbf{y}(t_2)$ as a double integral and take expected values.

9-28 (a) Write $\mathbf{y}^2(t)$ as a double integral and take expected values. (b) **I**, **M** Find the general, formal, solution to this differential equation (an equation which is not time-invariant). Then proceed as in part (a).

9-29 (a) Write a formal solution for $\mathbf{y}(t)$ to find the impulse response function $h(t)$. Then use Eq. 9-99.

9-30 No hint.

9-31 **F** Assume that $\mathbf{x}(t)$ is WSS. Use a method analogous to that used to develop Eq. 9-59.

9-32 (a) Use the solution to Problem 9-29, or otherwise find the impulse response function. Use Eqs. 9-93 and 9-95. (b) Use Example 9-18.

9-33 (a) **M** See Example 5-28.

(b) Use $\cos\omega_0\tau\cos\omega\tau = [\cos(\omega_0-\omega)\tau + \cos(\omega_0+\omega)\tau]/2$, with part (a).

9-34 Assume $x(t)$ is real.

9-35 Use the identities $e^{j2a\omega} + e^{-j2a\omega} = 2\cos 2a\omega$, and $1 - \cos 2a\omega = 2\sin^2 2a\omega$.

9-36 No hint.

9-37 **M** Show $\eta_y = I$. Use Eq. 7-61 from Example 7-9 to find R_y, then use Table 9-1 to find S_y.

9-38 Since $S(\omega) \ge 0$, clearly $A = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)\left|\Sigma_i a_i e^{j\omega\tau_i}\right|^2 d\omega \ge 0$.

9-39 (a) **M** Use Cauchy's residue theorem to find $R(\tau)$ via contour integration, as in the solution to Problem 5-37. (b) **M** Use the same approach. The pole here is of order two, and the residue is given by $r = \frac{d}{d\omega}[(\omega-\omega_p)^2 h(\omega)]\Big|_{\omega=\omega_p}$, where ω_p is the pole and $h(\omega)$ is the integrand.

9-40 **D** $H(\omega)$ is a complex function of the real variable ω. $H(s)$ is a complex function of the complex variable s. To take the complex conjugate of $H(s)$ you must conjugate both the function $H(\ )$ and the variable s. To assist in the notation, define, in the first and second parts of the problem

$$W(s) = H^*(-s^*) = \int_{-\infty}^{\infty} h^*(t)e^{st}dt, \quad W(z) = H^*(1/z^*) = \Sigma_n h^*[n]z^n$$

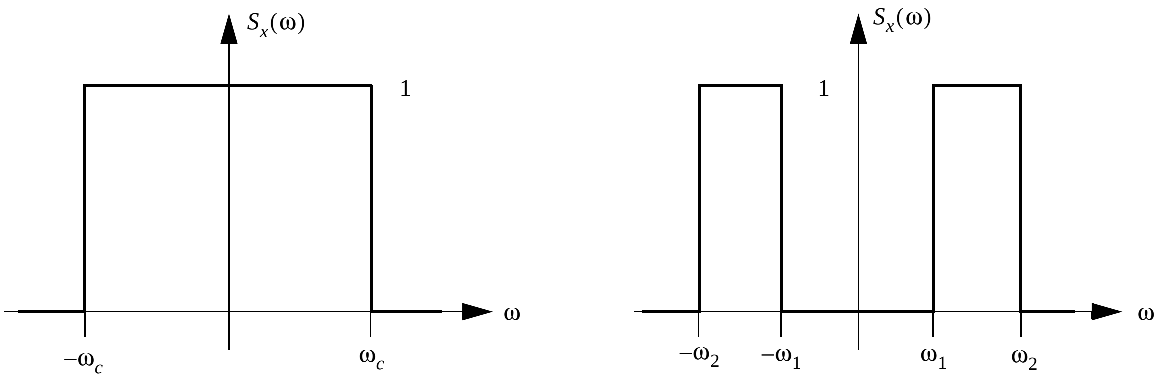
9-41 **B** To solve this problem, it is best to use the general convolution theorem. There are two versions, and both assume three pairs of Fourier transforms: $f(\tau) \leftrightarrow F(\omega)$, $g(\tau) \leftrightarrow G(\omega)$, and $h(\tau) \leftrightarrow H(\omega)$.

I) If $H(\omega) = F(\omega)G(\omega)$, then $h(\tau) = f(\tau) * g(\tau)$

II) If $h(\tau) = f(\tau)g(\tau)$, then $H(\omega) = \frac{1}{2\pi}F(\omega) * G(\omega)$

Notice the factor of two pi in the second version of the theorem. Apparently this factor was neglected in the setting of the problem, leading to the typo. The correct result is $S_y(\omega) = 2\pi R_x^2(0)\delta(\omega) + \frac{1}{\pi}S_x(\omega) * S_x(\omega)$.

Also, you need to know what ideal LP (low pass) and BP (band pass) spectra are. They are simply unit intensity white noise processes put through an idealized filter. They are illustrated below.



Low Pass Band Pass

9-42 Use Eq. 9-106 to find $R_{xx'}(\tau)$ and $R_{x'x'}(\tau)$. Note that $R_{xx'}(\tau)$ is discontinuous at the origin, leading to a $\delta(\tau)$ term in $R_{x'x'}(\tau)$. Use Table 9-1 to convert from $R_{yy}(\tau)$ to $S_{yy}(\omega)$.

9-43 (a) Use Example 9-27a. (b) Use Eq. 9-115.

9-44 Note that $R(0)$ must be real and nonnegative, so that it must be that $R(\tau_1) = R(0)e^{j\varphi}$, for some phase angle φ. Define $\omega = \varphi/\tau_1$.

9-45 (a) **M** The presumption is that $\mathbf{x}(t)$ is real. Express $E\{\mathbf{x}(t)\hat{\mathbf{x}}(t)\}$ in terms of $S_{xx}(\omega)$ usinf Eq. 9-160. Attempting to use $R_{xx}(\tau)$ will fail.

(b) **D** Express $\tilde{\mathbf{x}}(t)$ (my symbol for double upside-down hats which I lack) in terms of $\mathbf{x}(t)$ and $h(t)$ via an iterated convolution. Since $H^2(\omega) = -1$, it follows that $\rho(t) = -\delta(t)$. An innovative alternate approach involves the Fourier transform of $\mathbf{x}(t)$ itself, but this method lacks generality.

9-46 Find some $G(\omega)$ such that $S_{yy}(\omega) = S_{xx}(\omega)G(\omega)$. Then find that $\omega = \pm\omega_0$ maximizes $G(\omega)$, and put all the available energy of $S_{xx}(\omega)$ at these frequencies.

9-47 Use inequality 9-181 to show that $S_{xy}(\omega) = 0$ for $\omega \neq \omega_0$. Deduce that

$S_{xy}(\omega) = 2\pi B\delta(\omega - \omega_0)$.

9-48 (a) Use convolution integrals directly. Note R_{yx} instead of R_{xy}.

9-49 Presume that over sufficiently small intervals of the frequency axis that the functions $S_{xy}(\omega)$, $S_{xx}(\omega)$, and $S_{yy}(\omega)$ are essentially constant. Use Eq. 9-181.

9-50 Presume that $\mathbf{x}(t)$ is real. Use the cosine inequality, Eq. 6-169.

9-51 Proceed as in Problem 9-50. Note the typo: the greater-than sign should be a greater-than-or-equal sign.

9-52 Express $R[m]$ in terms of $f(\omega)$. Compare with Eq. 9-193.

9-53 (a) **I**, **M** Find the solutions to the homogeneous differential equation. Use the variation of parameters method to find the solution to the non-homogeneous differential equation. Write $\mathbf{y}^2(t)$ as a double integral and take expected values.

(b) **I**, **D** Use the same approach as above for the difference equation.

9-54 (a) **I** Find the general solution to the difference equation.

(b) **I** Modify the development of part (a).

10) Chapter 10

10-1 (a) Use Eq. 9-13. (b) Use Eq. 10-52 and Example 5-28.

10-2 Use Eq. 10-52 for $f_x(x, t)$ and $f_y(y, t)$. Then use Eq. 6-70 for $f_z(z, t)$.

10-3 **I** Find the differential equation relating $\mathbf{v}(t)$ and $\mathbf{n}_e(t)$. Use Laplace transforms to find $H(s)$, and then solve for $|H(\omega)|^2$. Use Eq. 10-74 to find $S_v(\omega)$. A similar process works for the current case.

10-4 Note that $e^{-2\alpha t}U(t) \leftrightarrow \dfrac{1}{(j\omega + 2\alpha)}$, and, if $h(t) \leftrightarrow H(\omega)$, then $h'(t) \leftrightarrow j\omega H(\omega)$.

10-5 Use Example 7-12 and Eq. 9-106.

10-6 Presume that the Wiener process, $\mathbf{w}(t)$, is a normal stochastic process with zero mean. Use Eq. 6-199 to find R_y and Problem 6-69 to find R_z.

10-7 Use Campbell's theorem, Eq. 10-102. Note that $\mathbf{s}(7)$ can be zero only if no Poisson points arrive in the interval prior to $t_0 = 7$ when $h(7-t)$ is not zero.

10-8 Show that $S_{xy}(\omega)$ is purely imaginary. Then use Eqs. 9-147 and 9-149 to confine $H(\omega)$.

10-9 D A brute force approach works here, but it is both lengthy and demanding.

10-10 F As the Text points out on the bottom of page 465, given just $\mathbf{x}(t)$ there is no unambiguous complex envelope. You must assume Rice's representation, so that the $\mathbf{y}(t)$ in Eq. 10-132 is $\hat{\mathbf{x}}(t)$.

10-11 Differentiate the relationship of Eq. 9-134 to find $R_{xx}''(0)$. Use Eqs. 10-148 and 10-150.

10-12 F You must assume that $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are real, jointly WSS processes, and that you are given the constant ω.

10-13 M First note the typo. The correct answer is

$$S_{xx}(\omega) = \frac{2\pi}{T^2}\left|\int_0^T f(t)e^{-j\omega t}dt\right|^2 \sum_{m=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi m}{T}\right)$$

Begin by showing that the stochastic process $\mathbf{y}(t) = f(t)$ is SSCS. Then show $\mathbf{x}(t)$ is SSS, and use Eq. 10-177 for $R_{xx}(\tau)$. Find $S_{xx}(\omega)$ directly from Eq. 9-133, then split up the infinite time integral into a sum over consecutive spans of length T. Finally, use Eq. 10A-2 to get the desired result.

10-14 M Mimic Eq. 10-197 and define

$$\mathbf{y}_N(t) = \mathbf{x}(t+\tau) - \sum_{n=-N}^{N} \mathbf{x}(t+nT)\frac{\sin\sigma(\tau-nT)}{\sigma(\tau-nT)}, \quad T = \pi/\sigma$$

Note that $\varepsilon_N(t) = \mathbf{y}_N(0)$.

10-15 Assume $\mathbf{x}(t)$ is real, and use the proof at the bottom of page 477.

10-16 Use Eq. 10-196, Eq. 10-198 with $\omega = 0$, and Table 9-1.

10-17 See hint 10-16 above.

10-18 Note that in the range $|\tau| < \frac{\pi}{2\sigma}$ and $\omega \le \sigma$, then $\cos\omega\tau \ge \cos\sigma\tau$. Also, assume that $\boldsymbol{x}(t)$ is real, as usual in BL problems.

10-19 M This is a straightforward application of the "Papoulis Sampling Expansion" on Text page 480, but difficult because it involves considerable manipulation. For some help, note that $P_1(\omega, \tau) = 1 - \frac{\omega}{\sigma}(e^{j\sigma\tau} - 1)$, $P_2(\omega, \tau) = \frac{1}{j\sigma}(e^{j\sigma\tau} - 1)$, and $p_1(\tau) = \frac{1}{\sigma}\int_{-\sigma}^{0}\left[1 - \frac{\omega}{\sigma}(e^{j\sigma\tau} - 1)\right]e^{j\omega\tau}d\omega$, $p_2(\tau) = \frac{1}{j\sigma^2}(e^{j\sigma\tau} - 1)\int_{-\sigma}^{0} e^{j\omega\tau}d\omega$. Use Eq. 10-206 with $t_0 = 0$, noting that $\boldsymbol{y}_1(n\bar{T}) = \boldsymbol{y}_1(n\Delta) = \boldsymbol{x}(n\Delta)$, $\boldsymbol{y}_2(n\bar{T}) = \boldsymbol{y}_2(n\Delta) = \boldsymbol{x}'(n\Delta)$.

10-20 Use the method of Eq. 10-212, with $f(t) = \cos\omega_0 t \cos\omega t$, and with the integral limits from $-a$ to a. Get equivalent formulas to those following Eq. 10-214.

10-21 F First correct the typo. The sum should be $\boldsymbol{X}_c(\omega) = \frac{1}{\lambda}\sum_{\substack{i \\ |t_i| < a}} \boldsymbol{x}(\boldsymbol{t}_i)e^{-j\omega \boldsymbol{t}_i}$. Next, you must assume that $\boldsymbol{x}(t)$ is independent of the process $\boldsymbol{z}(t) = \sum_i \delta(t - \boldsymbol{t}_i)$. The integral transform of Eq. 9-59, valid for any function $C(\tau)$, is also useful.

10-22 No hint.

10-23 Consider $I = \sum_{i=1}^{n}\sum_{j=1}^{n}\left|a_i b_j^* - a_j b_i^*\right|^2 \ge 0$.

10-24 (a) **D** The treatment of discrete processes in the Text is too brief to even give a useful hint here, except to mimic the matched filter discussion for a continuous process. Assume real processes.

(b) **D** Note that $\boldsymbol{v}[n]$ is not presumed to be white here. To maximize $r = \frac{y_f^2[0]}{E\{\boldsymbol{y}_v^2[n]\}}$, you should minimize the denominator while holding the numerator constant. Use the method of Lagrange multipliers.

10-25 M Find the impulse response function $h(t)$ (see Example 10-2). From the deterministic input, $f(t) = A\cos\omega_0 t$, find the deterministic output in the form $y_f(t) = B\cos(\omega_0 t + \varphi)$ to relate B to A, α, and ω_0. Find the spectrum of the noise output, and from that find the autocorrelation of the noise output, using Table 9-1. Use Eq. 9-152 to find the average power of the noise output.

10-26 M (a) It helps to define the vectors and matrices: $\vec{a}^T = (a_0, a_1, \ldots, a_m)$,

$$\vec{f}^T = (f_0, f_1, \ldots, f_m) = (f(t_0), f(t_0 - T), \ldots, f(t_0 - mT)), \text{ and } R = \begin{bmatrix} R_{00} & R_{01} & \ldots & R_{0m} \\ R_{10} & R_{11} & \ldots & R_{1m} \\ \ldots & \ldots & & \ldots \\ R_{m0} & R_{m1} & \ldots & R_{mm} \end{bmatrix},$$

where $R_{ij} = R_v(iT - jT) = E\{\mathbf{v}(t_0 - iT)\mathbf{v}(t_0 - jT)\}$. Then, show

$y_f = y_f(t_0) = \sum_{i=0}^{m} a_i f_i = \vec{a}^T \vec{f}$, and $E = E\{\mathbf{y}_v^2(t_0)\} = \vec{a}^T R \vec{a}$. Maximize r by minimizing E subject to the constraint of a given value of $y_f > 0$ by the method of Lagrange multipliers.

(b) The maximum value of r is $\sqrt{\vec{f}^T R^{-1} \vec{f}} = \sqrt{y_f / k}$.

10-27 This follows directly from Eq. 10-243 using (repeatedly) the identity $\int_{-\infty}^{\infty} e^{-j\omega\tau} d\tau = 2\pi\delta(\omega)$.

10-28 M Several facts are useful here. If $t_1 < t_2 < t_3$, then $\tilde{\mathbf{x}}(t_3) = \tilde{\mathbf{x}}(t_1) + [\tilde{\mathbf{x}}(t_2) - \tilde{\mathbf{x}}(t_1)] + [\tilde{\mathbf{x}}(t_3) - \tilde{\mathbf{x}}(t_2)]$, and the three random variables $\tilde{\mathbf{x}}(t_1)$, $\tilde{\mathbf{x}}(t_2) - \tilde{\mathbf{x}}(t_1)$, and $\tilde{\mathbf{x}}(t_3) - \tilde{\mathbf{x}}(t_2)$ are mean zero and independent, because they count Poisson points in nonoverlapping intervals. From the discussion of Poisson random variables on page 149 of the Text, it follows that $E\{\tilde{\mathbf{x}}^3(t_1)\} = \lambda t_1$. Finally, ignore the hint in the problem and use instead the three identities

$$\min(t_1, t_2, t_3) = \min(t_1, \min(t_2, t_3)), \qquad \frac{\partial \min(t_a, t_b)}{\partial t_a} = U(t_b - t_a)$$

$$U(\min(t_2, t_3) - t_1) = U(t_2 - t_1)U(t_3 - t_1)$$

10-29 Follow the outline in the problem.

11) Chapter 11

11-1 D Finding the whitening filter is easy, especially if you note that

$\cos 2\omega = (z^2 + z^{-2})/2$. Finding the autocorrelation sequence $R_{xx}[m]$ is hard—so hard that no hint is given here, see the solution.

11-2 Factor $S(s) = \dfrac{N(s)N(-s)}{D(s)D(-s)}$, then set $L(s) = \dfrac{N(s)}{D(s)}$.

11-3 Use the convolution sum: $\mathbf{s}[n] = \sum_{k=0}^{\infty} l_s[k]\mathbf{i}[n-k]$.

11-4 (a) **I** Note that $R_{yx}'(\tau) = E\{\mathbf{y}'(t+\tau)\mathbf{x}^*(t)\}$, etc.

(b) **I**, **M** Use the discussion following Eq. 11-25. Note that $S_{yx}^+(s) = S_{yx}(s) = q/D(s)$ to find $R_{yx}(\tau)$ for $\tau > 0$. Note that $S_{yy}^-(s) = S_{yy}^+(-s)$, and $S_{yy}(s) = \dfrac{q}{D(s)D(-s)}$ to find $S_{yy}^+(s)$ and, hence, $R_{yy}(\tau)$ for $\tau > 0$.

11-5 Show $R_{xx}[m] = R_{ss}[m] + R_{vv}[m]$ and $R_{vv}[m] = q\delta[m]$. Deduce that $S_{ss}(z) = \dfrac{1}{D(z)D(1/z)}$, where $D(z)$ has all its roots z_i with $|z_i| < 1$. Conclude that $S_{xx}(z)$ is also rational, with the same poles as $S_{ss}(z)$, and that $S_{xx}(z) = S_{xx}(1/z)$.

11-6 Define $\mathbf{s}(t) = \dfrac{1}{n}\sum_{k=1}^{n} \mathbf{x}(t+kT)$, and regard $\mathbf{s}(t)$ as the output of a linear system with input $\mathbf{x}(t)$. Find the impulse response function, the system function, and use Eq. 9-152.

11-7 **D**, **F** Presume $\mathbf{x}(t)$ is WSS, so the origin of the time scale may be shifted so that the interval $(0, T)$ in the old time scale corresponds to the interval $(-a, a)$ in the new time scale. Particularize the integral equation 11-58. Differentiate the integral equation twice to obtain a differential equation. Find the general solutions to the differential equation and put them back into the integral equation to learn more about them. Use the normalization equation, Eq. 11-56, to scale the $\varphi(t)$ functions. Note that the results should read: $\beta_n = \left(a + \dfrac{\lambda_n}{2}\right)^{-1/2}$ and $\beta_n' = \left(a + \dfrac{\lambda_n'}{2}\right)^{-1/2}$, not the results given in the Text.

11-8 Write $E\{|\mathbf{X}(\omega)|^2\}$ as a double integral. Transform to a single integral as in Eq. 9-59. Differentiate with respect to T.

11-9 No hint.

11-10 No hint.

11-11 **M** Use Eq. 11-51. (a) Show

$$E\{\boldsymbol{x}(t)\boldsymbol{x}^*(t)\} = E\{\boldsymbol{x}(t)\hat{\boldsymbol{x}}^*(t)\} = E\{\hat{\boldsymbol{x}}(t)\boldsymbol{x}^*(t)\} = E\{\hat{\boldsymbol{x}}(t)\hat{\boldsymbol{x}}^*(t)\} = R(0)$$

In order to do part (a), it is useful to do part (b). (c) Substitute $s = t - \alpha$ in the $\beta_n(\alpha)$ expression.

11-12 As is usual in these cases, show that $E\{\boldsymbol{x}(t)\} = 0$ and that $E\{\boldsymbol{x}(t)\boldsymbol{x}^*(s)\}$ is a function of $t - s$.

11-13 Deduce that if $\boldsymbol{A}$ and $\boldsymbol{B}$ satisfy Eqs. 11-79, then it must be that $E\{\boldsymbol{A}(u)\boldsymbol{A}(v)\} = E\{\boldsymbol{B}(u)\boldsymbol{B}(v)\} = Q(u)\delta(u-v)$ and $E\{\boldsymbol{A}(u)\boldsymbol{B}(v)\} = 0$.

11-14 **M** Note that $\int_{-T}^{T} f(t)e^{-j\omega t}dt = \int_{-\infty}^{\infty} f(t)p_T(t)e^{-j\omega t}dt$, where

$$p_T(t) = U(t-T) - U(T-t) = \begin{cases} 1 & -T < t < T \\ 0 & \text{otherwise} \end{cases}$$

Use the frequency convolution theorem: if $F_1(\omega) = \int_{-\infty}^{\infty} f_1(t)e^{-j\omega t}dt$, and

$F_2(\omega) = \int_{-\infty}^{\infty} f_2(t)e^{-j\omega t}dt$, then

$$F_{12}(\omega) = \int_{-\infty}^{\infty} f_1(t)f_2(t)e^{-j\omega t}dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} F_1(y)F_2(\omega - y)dy$$

Use Eq. 11-82. Note that $\int_0^{\infty} \frac{\sin^2 \alpha T}{\alpha^2} d\alpha = \frac{T\pi}{2}$.

12) Chapter 12

12-1 No hint.

12-2 **I**, **D** Note that $\boldsymbol{x}(t)$ is WSS. Use Eq. 12-35 to deduce that $f(x, x; \tau) \to f^2(x)$ as $|\tau| \to \infty$. Find both sides of this limit explicitly, and show that the limit requires

$r(\tau) \to 0$. Use Eq. 12-10.

12-3 Use Eq. 12-27.

12-4 Show that $C_{zz}(\tau)$ does not depend on τ, so that Eq. 12-5 cannot be satisfied.

12-5 I Recall that $\lim_{T \to \infty} \boldsymbol{R}_T = R_{xy}(\lambda)$ if and only if both $E\{\boldsymbol{R}_T\} = R_{xy}(\lambda)$, and $\sigma^2_{R_T} \to 0$.

12-6 F, **D**, **I** The condition in this problem should read: $R(t+\tau, t) \to \eta(t+\tau)\eta(t)$ as $|\tau| \to \infty$, uniformly in t. Also, you must assume $C(t, t) < \sigma^2$ for some σ and all t. Ignore the hint in the Text. Define the average mean as $\bar{\eta} = \frac{1}{T}\int_0^T \eta(t)dt$. Show that $\lim_{c \to \infty} \frac{1}{2c}\int_{-c}^{c} \eta(t)dt = \bar{\eta}$. Use this result to show that $\lim_{c \to \infty} E\{(\eta_c - \bar{\eta})^2\} = 0$ is equivalent to the revised condition above. Do Problem 13-7 first to understand this latter result.

12-7 I, **D**! Assume that $C(t, t) < \sigma^2$ for some σ and all t. Use Eq. 9-37 to get $\sigma_T^2 = \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} C(t_1, t_2)dt_1 dt_2$. Change variables to $t = t_2$, $\tau = t_1 - t_2$. Break up the integral in the $t\tau$-plane into portions where $|\tau| \geq T_2$ and $|\tau| < T_2$, and find ways to limit both parts.