We tend to think that certainty is required to have true information. However, we make decisions every day using probabilistic information:

Engineering activities:

Many applications also arise in engineering:

- random disturbances
- random system fluctuations
- information theory
- system reliability

Ex: Suppose space shuttle has 10,000 critical sub-systems, each of which has an expected failure rate of 1 in 1,000,000 flights.

prob. of no system failure =

Definitions

random experiment — action that results in an outcome that is uncertain in advance.

outcome — any possible result of an experiment

trial — single performance of an experiment

event — an outcome or set of outcomes defined by the experiment

elementary event — equivalent to single outcome

composite event — equivalent to a set of outcomes

discrete outcomes — have one-to-one correspondence with integers

continuous outcomes — outcomes form a continuum (e.g., measuring resistance of resistor)

Relative frequency

- an intuitive approach to understanding probability

Suppose we have four possible outcomes — A, B, C, D. Let N_A = # of time A occurs. and N = # of trials.

Then $N_A + N_B + N_C + N_D = N$

Relative freq. of A is

$$r(A) + r(B) + r(C) + r(D) = ?$$

$$Pr(A) = \lim_{N \to \infty} r(A)$$

$$\Rightarrow Pr(A) + Pr(B) + Pr(C) + Pr(D) =$$

multiple events

- We need concept of joint probability
  e.g. probability of two heads in a row on coin toss

- conditional probability — probability of one event given that another has occurred (or is known)

In general,

$$Pr(A, B) = Pr(A|B)\,Pr(B) = Pr(B|A)\,Pr(A)$$

- joint probabity of A & B

For coin toss,

$$Pr(H_1, H_2) =$$

If $Pr(A|B) = Pr(A)$, then

Elementary Set Theory

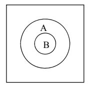
- A *set* is a collection of elements

  Ex: $A=\{1,2,3,4,5,6\}$

- set B is a *subset* of A if all elements of B are in A

  Ex: $B=\{1,2,3\}$

  We denote this as $B \subset A$



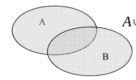
empty set: ϕ

equality: $A=B$ iff $A \subset B$ and $B \subset A$

space: S - set of all possible elements

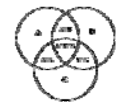
Elementary Set Theory cont'd

- *union* $A \cup B$
  - » *sum*
  - » Logical OR

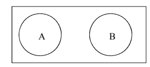


$A \cup B$

- *intersection* $A \cap B$
  - » *product*
  - » Logical AND



- *mutually exclusive* if $A \cap B = \phi$

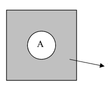


$$A \cup \phi = A,\ A \cap S = A,\ A \cup S = S,\ A \cap \phi = \phi$$

Elementary Set Theory cont'd

- *complement* $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$

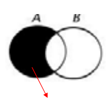
  $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$



$\overline{A}$

- *difference* $A - B = A \cap \overline{B}$

  $= A - (A \cap B)$



$A - B$

$A \cup \overline{A} = S$

where S is the space of all possible elements.

Elementary Set Theory cont'd

- Union is commutative and associative:

  $A \cup B = B \cup A$

  $(A \cup B) \cup C = A \cup (B \cup C)$

- Intersection is commutative and associative with respect to union (distributive):

  $A \cap B = B \cap A$

  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Axiomatic Approach

- A probability space S is the set of all possible outcomes of an experiment

  Ex: rolling a six-sided die

  S= $\{1,2,3,4,5,6\}$

- An event is a subset of S

  the space S is the *certain event*

  ϕ is the *impossible event*

  an event consisting of a single element is called an *elementary event*

Axiomatic Approach cont'd

- To each event, we assign a probability $\Pr(A)$

  Axioms of Probability

  1) $\Pr(A) \geq 0$

  2) $\Pr(S) = 1$

  3) If $A \cap B = \phi$, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

- All probability theory can be derived from these axioms.

Axiomatic approach

If $A \cap B = \phi$, then $Pr(A \cup B) = Pr(A) + Pr(B)$

$Pr(\phi) = ?$

$S \cap \phi = \phi$ $\qquad S \cup \phi = S$

$Pr(S \cup \phi) = Pr(S)$

$Pr(\bar{A}) = ?$

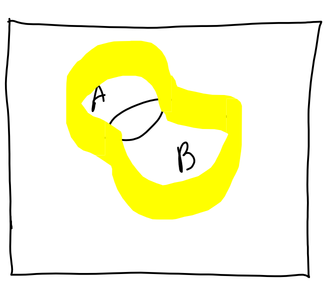
$A \cap \bar{A} =$ $\qquad A \cup \bar{A} =$

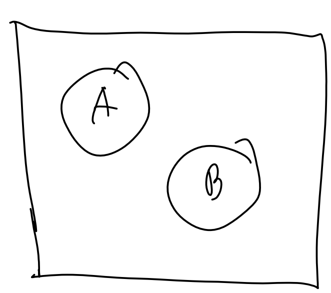
$Pr(A \cup \bar{A}) = Pr(S) = 1 =$

$\Rightarrow Pr(A) = 1 - Pr(\bar{A})$

$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

$\leq Pr(A) + Pr(B)$





Ex: rolling a die

$$Pr(\alpha_i) = \frac{1}{6}, \quad \alpha_i = 1, \ldots, 6$$

Let $A = \{1, 3\} = \{1\} \cup \{3\}$

$$Pr(A) = ?$$

$$B = \{3, 5\}$$

$$Pr(A \cup B) =$$

$$Pr(A \cup B) = Pr(\{1, 3, 5\}) = Pr(\{1\}) + Pr(\{3\}) + Pr(\{5\})$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

* rolling die twice – probability of at least one 6

$$\frac{1}{6} + \frac{1}{6} ?$$

S = two throws

A = 6 on 1st

B = 6 on 2nd

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$$

$$\Pr(A \cup \underbrace{B \cup C}) = \Pr(A) + \Pr(B \cup C) - \Pr(A \cap (B \cup C))$$

$$= \Pr(A) + \overbrace{\Pr(B) + \Pr(C) - \Pr(B \cap C)} - \Pr(A \cap (B \cup C))$$

$$= \Pr(A) + \Pr(B) + \Pr(C) - \Pr(B \cap C)$$

$$- \left[\Pr(A \cap B) + \Pr(A \cap C) - \Pr(\underbrace{(A \cap B) \cap (A \cap C)}_{A \cap B \cap C})\right]$$

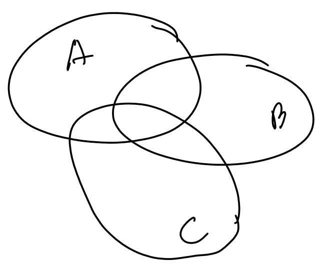
$A \cap (B \cup C)$

$= (A \cap B) \cup (A \cap C)$

$$= \Pr(A) + \Pr(B) + \Pr(C)$$

$$- \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C)$$

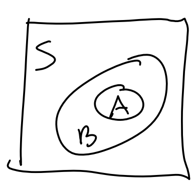
$$+ \Pr(A \cap B \cap C)$$



<u>Conditional probability</u>

$\Pr(A \mid B) =$

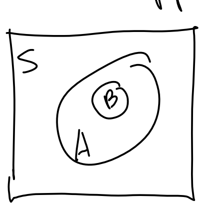
Suppose $A \subset B$ $\quad (A \cap B = \quad)$



$\Rightarrow \Pr(A \cap B) =$

$\Rightarrow \Pr(A \mid B) =$

Suppose $B \subset A$ $\quad (A \cap B = \quad)$



$\Rightarrow \Pr(A \mid B) =$

- Conditional probabilities satisfy the probability axioms, Therefore, they are valid probabilities.

Ex: rolling a die

$A = \{1, 2\}$, $B = \{2, 4, 6\}$

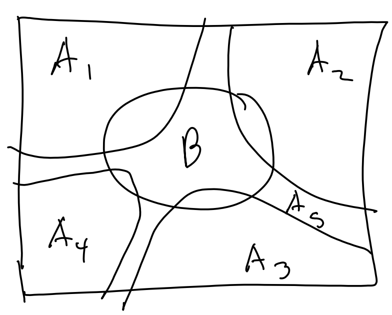
$Pr(A) =$ $Pr(B) =$

$Pr(A \cap B) = Pr(\{2\}) =$

$$Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

$Pr(B / A) =$

- Conditional probability is useful for computing total probabilities



$A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 = S$

$A_i \cap A_j = \emptyset$ for $i \neq j$

$(A_i \cap A_j) \cap B =$

$= (A_i \cap A_j) \cap (B \cap B)$

$= (A_i \cap B) \cap (A_j \cap B)$

$$B = B \cap S = B \cap (A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5)$$

$$= (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup (B \cap A_4) \cup (B \cap A_5)$$

$$Pr(B) = \sum_{i=1}^{5} Pr(B \cap A_i)$$

$$=$$

Ex:

then calc \ if major	engr	educ.	ag.
yes	0.8	0.05	0.15
no	0.2	0.95	0.85

$Pr(\text{engr.}) = 0.5$, $Pr(\text{educ.}) = 0.2$, $Pr(\text{ag}) = 0.3$

Choose a major w/ equal probability & then pick a student at random from that major. Prob. of this student carrying a calc.?

Now, pick a student at random out of the population —

What is probability you've picked an engr. student if student has a calculator?

$$\Pr(A_i \cap B) = \Pr(A_i | B)\Pr(B) = \Pr(B|A_i)\Pr(A_i)$$

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\Pr(B)}$$

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\qquad}$$

$\Pr(A_i|B)$ is called the <u>a posteriori probability</u> because it refers to probability after the experiment is performed.

<u>Conditional probability exercise</u>

Find $\Pr(\Sigma \leq 7)$

Ex: binary comm.

$$Pr(T_1) = 0.55$$

$$Pr(R_0 | T_1) = 0.1 \quad , \quad Pr(R_1 | T_0) = 0.2$$

$Pr(\text{error}) =$

$Pr(\text{received } 1 \mid \text{was } \cancel{\text{transmitted}} \text{ as a } 1)$

Independence

Two events A & B are independent iff

- In many physical situations, independence can be assumed; e.g., flipping coin twice
- sometimes not intuitive: $A = \{1, 2\}$
  $B = \{1, 3, 5\}$
  (die roll)

$A \cap B = \{1\}$

$Pr(A \cap B) = \frac{1}{6}$

$Pr(A)\,Pr(B) =$

What about three events?
The conditions for independence are:

$Pr(A \cap B) = Pr(A)\,Pr(B)$, $Pr(A \cap C) = Pr(A)\,Pr(C)$

$Pr(B \cap C) = Pr(B)\,Pr(C)$ $Pr(A \cap B \cap C) = Pr(A)\,Pr(B)\,Pr(C)$

Combined experiments

- a space S can represent probabilities associated with combined experiments

* we assume that individual experiments are independent

So, if event A_1 is in S_1
and event A_2 is in S_2

$S = S_1 \times S_2$, $A = A_1 \times A_2$

Then $Pr(A_1, A_2) =$

Let $A_1 = \{5, 6\}$ on die
$A_2 = \{H\}$ on coin

$Pr(A_1, A_2) =$

$Pr($no sixes in n throws$) =$

$Pr(\geq 1$ six in n throws$)$

two dice

pr. of no double-six

$Pr(\geq 1$ double-six in n throws$)$

Counting

How many ways to arrange n objects?

permutations of
n objects

How many arrangements of n objects if
m are indistinguishable?

$\leftarrow$ number of arrangements

$\leftarrow$ # of arrangements of
indistinguishable objects

Ex: $\{1, 2, 3, 3\}$

(# of distinguishable seq.'s of objects)

How many arrangements of k out of n objects?

- k permutation of n

How many <u>combinations</u> of k out of n objects?

- like k permutation of n except that order doesn't matter

- combinations of n objects taken k at a time

Ex: $\binom{6}{3} =$

$Pr(3 \text{ wrong questions}) =$

Ex: How many possible lab groups of five students can be formed from a class of 20?

$\binom{20}{5} =$

How many different lab groups could you end up in?

______

Repeated trials (Bernoulli trials)

Probability that A occurs exactly k out of n times, regardless of order?

$$p = Pr(A), \quad q = Pr(\bar{A})$$

$$\text{probability} = p^k q^{n-k} \times (\# \text{ of comb's})$$

$$=$$

Ex: 7 bits plus a parity bit

$Pr(\text{error in reading a bit}) = 0.001$

What is $Pr\{\text{single bit error in 8 bits}\}$?

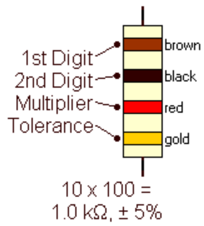
$Pr\{\text{two bit errors}\} =$

$Pr\{\text{no bit errors}\} =$

$0! = 1$ $\qquad =$

$Pr(<2 \text{ errors}) =$

Random Variables



$1\text{k}\Omega \pm 5\%$ resistor could have a resistance between $950\Omega - 1050\Omega$

Random Variables

- The value of the resistance, *R*, can be a range of values, but the exact value is not known in advance
  - » We call the resistance, *R*, a *random variable (RV)*
- Comparison with algebraic variables
  - » Algebraic variable: X represents an unknown or variable quantity.
  - » Random variable: *R* represents a range of values with certain properties.
    - – Anywhere between 950-1050Ω with equal probability
    - – An average value of 1000Ω with a deviation of 50Ω

Random Variables

- Since *R* can be any value in the continuous range [950,1050]Ω, we say it is a *continuous RV*.
  - » Outcome of rolling a fair die is an example of a *discrete RV*.
- It does not make sense to assign a probability to a single value for continuous RV's because we will see later the probability of any single value is zero.
- It only makes sense to talk about ranges of values for a continuous RV like 1000 and 1002Ω.

Random Variables

- Sometimes we want to describe how the values of an RV are distributed

Example – Exam Grades

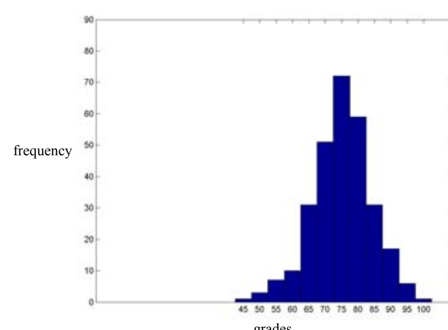
47	60	65	67	69	71	72	73	74	75	77	78	80	81	83	86	88
49	61	65	68	69	71	73	73	74	76	77	79	80	82	84	86	88
51	61	65	68	69	71	73	73	74	76	77	79	80	82	84	86	89
51	62	65	68	69	71	73	73	74	76	77	79	80	82	84	86	89
53	63	65	68	69	71	73	74	74	76	77	79	80	82	84	86	89
54	64	65	68	69	71	73	74	74	76	78	79	80	82	84	86	90
54	64	66	68	69	71	73	74	74	76	78	79	80	82	84	87	90
55	64	66	68	70	71	73	74	75	76	78	79	81	82	84	87	90
55	64	67	68	70	72	73	74	75	76	78	79	81	82	84	87	92
55	64	67	68	70	72	73	74	75	76	78	79	81	82	85	87	92
56	64	67	68	70	72	73	74	75	76	78	79	81	82	85	88	94
58	64	67	68	70	72	73	74	75	76	78	79	81	82	85	88	95
59	64	67	68	70	72	73	74	75	76	78	79	81	82	85	88	95
59	65	67	68	70	72	73	74	75	76	78	80	81	83	85	88	96
60	65	67	69	70	72	73	74	75	77	78	80	81	83	85	88	96
60	65	67	69	70	72	73	74	75	77	78	80	81	83	85	88	96
60	65	67	69	70	72	73	74	75	77	78	80	81	83	86	88	100

Total: 289

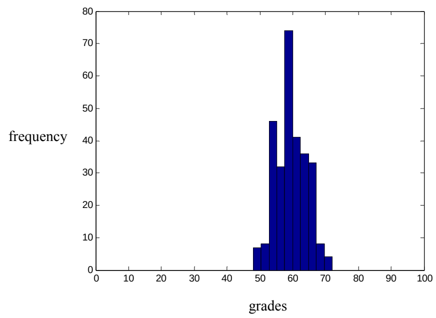
41-50: 2
51-60: 16
61-70: 67
71-80: 126
81-90: 69
91-100: 9

AVG: 74.9896
STD: 9.0245

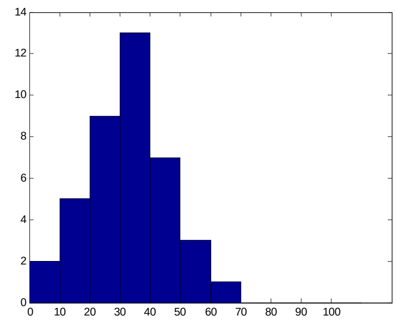
Histograms



Histograms



Typical ELEC 3800 Final ;-)



Distribution Functions

- Histograms describe how a set of data points are distributed
- For random variables, we use a *probability distribution function* or a *probability density function*

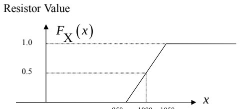
Distribution Functions

- Let ***X*** be an RV and *x* be an allowed value of ***X***.
- The *probability distribution function* of ***X*** is defined as:

$$F_X(x) = \Pr(X \le x)$$

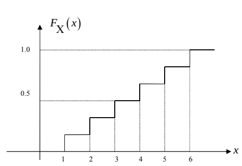
- For example, if the resistor in our example can take on values from 950 to 1050Ω with equal probability, then its distribution function is

Resistor Value



Distribution Functions

- If ***X*** is the number of dots on a fair die, its distribution function is



Distribution Functions cont'd

Probability distribution functions are probabilities, so they satisfy the axioms of probability:

1) $0 \le F_X(x) \le 1 \quad , \quad -\infty < x < \infty$

2) $F_X(-\infty) = 0 \qquad F_X(\infty) = 1$

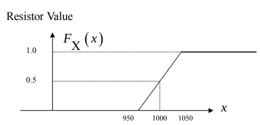
3) $F_X(x)$ non-decreasing as x increases

4) $\Pr(x_1 < X \le x_2) = F_X(x_2) - F_X(x_1)$

5) $\Pr(X > x) = 1 - F_X(x)$

Example

- 1000Ω 5% resistor

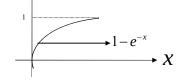


$$\Pr(X \le 970) = 0.2$$
$$\Pr(X > 970) = 1 - 0.2 = 0.8$$
$$\Pr(975 < X \le 1025) = F_X(1025) - F_X(975)$$
$$= 0.75 - 0.25 = 0.5$$

- Note!
  - » $\Pr(\boldsymbol{X}=970)=\Pr(970<\boldsymbol{X}\le 970)=F_X(970)-F_X(970)=0$

Example

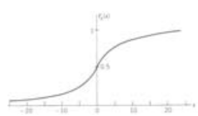
- Suppose $\boldsymbol{X}$ has the following distribution function



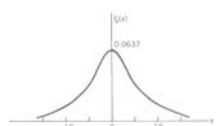
$$\Pr(X > 0.5) = 1 - F_X(0.5) = 1 - (1 - e^{-0.5}) = 1 - 0.3935 = 0.6065$$
$$\Pr(X \le 0.25) = F_X(0.25) = 1 - e^{-0.25} = 0.2212$$
$$\Pr(0.3 < X \le 0.7) = F_X(0.7) - F_X(0.3)$$
$$= 0.5034 - 0.2592 = 0.2442$$

Probability Density Functions

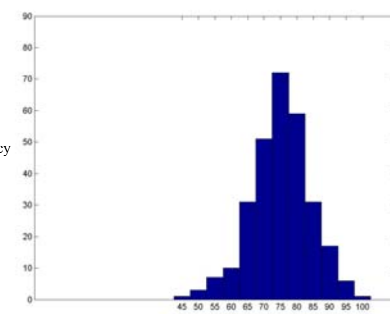
- Probability distribution functions
  - » Easy to work with because $F_X(x)$ is a probability.
  - » Useful in certain advanced applications
  - » Non-intuitive
    - – Avg value?
    - – Most likely value?
    - – More likely to be positive or negative?



- **Probability density functions** (PDF's) – $f_X(x)$
  - » A more intuitive way to describe a random variable
  - » Derivative of the probability distribution function



Histograms



Computing Probabilities with PDF's

- Integrate the probability density function to compute the probability that X is in a given range

$$\Pr(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x)dx$$

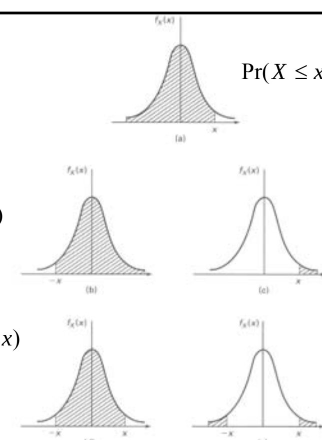
- Related to mass density
  - – Ex. A 0.25in dia. steel rod has density 0.034 kg/cm
    - How much does a 100cm section weigh?

$$\int_0^{100} 0.034\,dx = 3.4\,\text{kg}$$

$\Pr(X \le x)$

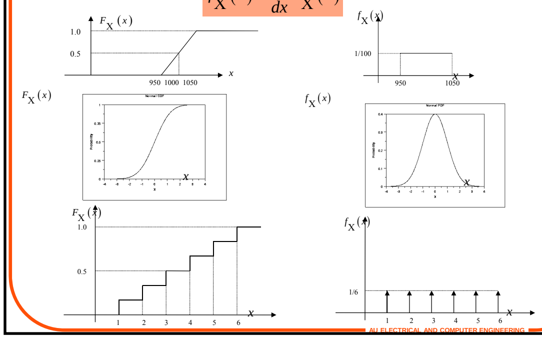
$\Pr(X > -x)$ $\Pr(X > x)$

$\Pr(-x < X \le x)$ $\Pr(|X| > x)$



Density Functions - Examples

$$f_X(x) = \frac{d}{dx}F_X(x)$$



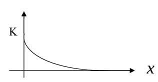
Properties of PDF's

- $f_X(x) \geq 0 \qquad -\infty < x < \infty$
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
- $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$
- $\int_{x_1}^{x_2} f_X(x)\,dx = \Pr(x_1 < X \leq x_2)$

Density Functions cont'd

EX:

$$f_X(x) = Ke^{-2x}u(x)$$



a) Value of K

$$1 = \int_{-\infty}^{\infty} Ke^{-2x}u(x)\,dx = \int_{0}^{\infty} Ke^{-2x}dx$$

$$1 = -\frac{K}{2}e^{-2x}\Big|_{0}^{\infty} = 0 - (-\frac{K}{2}) = \frac{K}{2} \qquad \Rightarrow K = 2$$

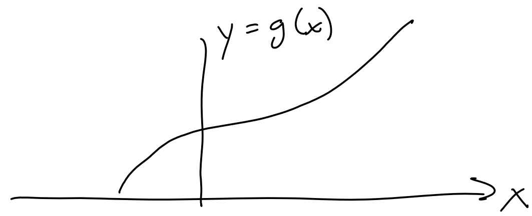
b) $\Pr(X > 1) = \int_{1}^{\infty} 2e^{-2x}u(x)dx = \int_{1}^{\infty} 2e^{-2x}dx = -e^{-2x}\Big|_{1}^{\infty} = e^{-2} = 0.1353$

c) $\Pr(X \leq 0.5) = \int_{-\infty}^{0.5} 2e^{-2x}u(x)dx = \int_{0}^{0.5} 2e^{-2x}dx = -e^{-2x}\,|_{0}^{0.5} = -e^{-2(0.5)} + 1 = 0.6321$

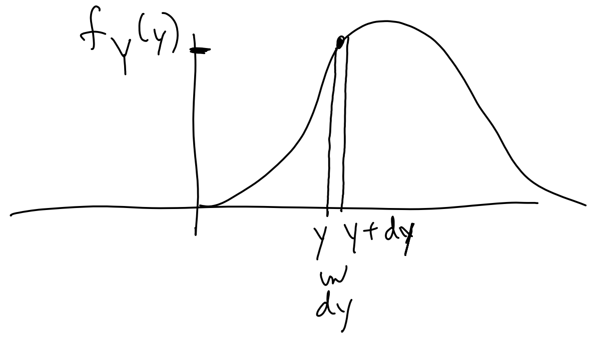
transformation of random variable

- random variables often undergo transformations
  - noise through amp
  - noise at receiver

Let $Y = g(X)$



We know $\Pr(x < X \le x + dx) =$



$\Pr(x < X \le x + dx) =$

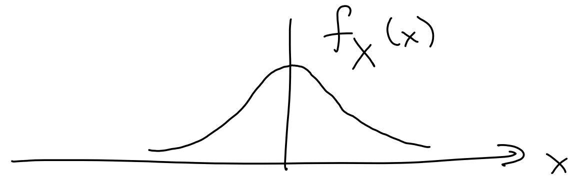
$$\Rightarrow f_X(x)\,dx = f_Y(y)\,dy$$

Ex: $Y = X^3$

$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

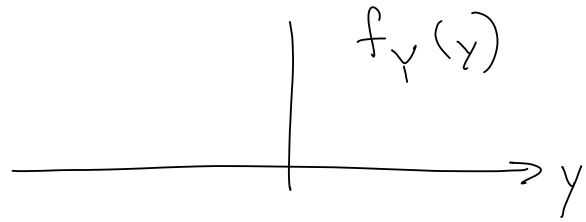
$(y = x^3)$

$\Rightarrow x =$



$\frac{dy}{dx} =$

$f_Y(y) = \frac{1}{|dy/dx|} f_X(x) \Big|_{x = y^{1/3}} = \frac{1}{\quad}$



Ex: $Y = X^2$

Note: There are two values of X for each Y.

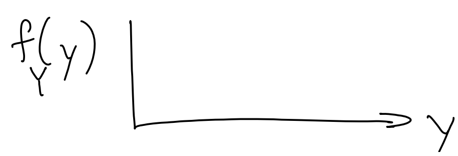
$\Longrightarrow$

$\frac{dy}{dx} =$ $\qquad \left|\frac{dx}{dy}\right| =$

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right], & y \geq 0 \\ 0 & , y < 0 \end{cases}$$

$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$\Longrightarrow f_Y(y) =$



Mean values + moments

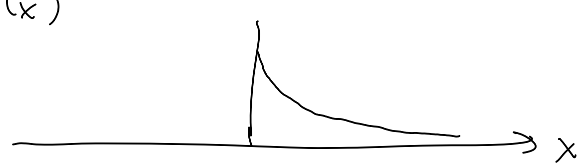
What is the "best guess" value of a random variable?

- mean

$$\sum_i x_i \Pr(x_i)$$

$$\lim_{\Delta x \to 0} \sum_i x_i \overbrace{f_x(x_i)\,\Delta x} =$$

Ex: $f_X(x) = e^{-x} u(x)$



$$\overline{X} = \int_{-\infty}^{\infty} x\, e^{-x} u(x)\, dx =$$

- mean of a function of x

$$E[g(X)] =$$

$E[(X-\alpha)^2]$: minimize with respect to α

$\Rightarrow \alpha = \overline{X}$

- moments

$\equiv$ mean of $g(x) = x^n$

$$\overline{X^n} = E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,dx$$

$$\overline{X^2} = E[X^2] =$$

mean-square value

- variance

$$\sigma^2 = \overline{(X-\overline{X})^2} =$$

$$\sigma^2 = \overline{(X-\overline{X})^2}$$

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

$$\sigma^2 = E[(X-\overline{X})^2] =$$

σ = standard deviation (in same units as X)

Ex: $f_X(x) = e^{-x} u(x)$

$$\overline{X} = \int_{-\infty}^{\infty} x e^{-x} u(x)\,dx =$$

$$\overline{X^2} = \int_0^\infty x^2 e^{-x}\,dx = -e^{-x}\left[x^2 + 2x + 2\right]\Big|_0^\infty$$

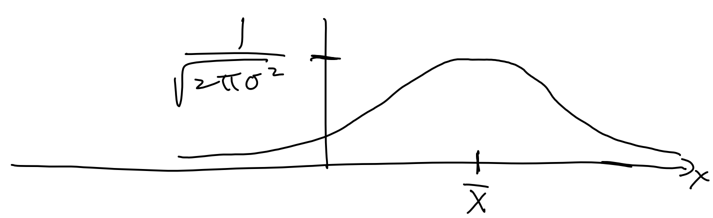
$$= 2$$

$$\sigma^2 = \overline{X^2} - (\bar{X})^2 =$$

Gaussian random variable

- the most important!
  - good model for many phenomena
  - easily handles multiple random variables

$f_X(x) =$



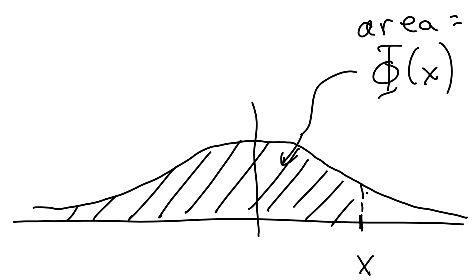
$F_X(x)$ cannot be expressed in closed form.

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx =$$

This function is usually tabulated for $\bar{X} = 0$ and $\sigma^2 = 1$.

$$\Phi(x) =$$

$$\Phi(-x) = 1 - \Phi(x)$$



$$F(x) =$$

$Q(x) = 1 - \Phi(x)$ is also tabulated frequently (good for extreme probabilities)

$$Q(-x) =$$

Ex: Gaussian with $\bar{X} = 1$ $(\mu = 1)$, $\sigma^2 = 2$, Find $\Pr\{X \le 0\}$.

Ex: Gaussian with $\overline{X} = 1$, $\sigma^2 = 2$.

$Pr\{1 < X \le 2\} =$

$Pr\{X > 4\} =$

Central limit theorem

- most random variables, when summed together, tend toward Gaussian r.v.
  - * adding at least 30 r.v.'s approaches Gaussian

other pdf's

distribution of random power

Suppose current I is randomly distributed as a Gaussian: $\overline{I} = 0$

$$f_I(i) = \frac{1}{\sqrt{2\pi\sigma_I^2}} \exp\left[\frac{-i^2}{2\sigma_I^2}\right]$$

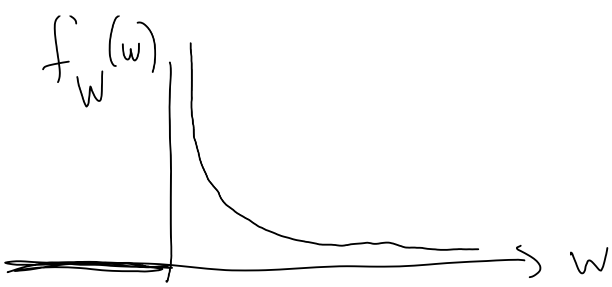
How is power distributed?

$$W = R I^2 \quad \Longrightarrow$$

$$\frac{dw}{di} =$$

$$f_W(w) = \left|\frac{di}{dw}\right| f_I(i) =$$

$$=$$



$$\overline{W} = E[W] =$$

$$\bar{I} = 0 \quad , \quad \sigma_I^2 =$$

$$\sigma_W^2 = \overline{W^2} - (\overline{W})^2 =$$

Ex: stereo — $8\,\Omega$ speaker, 50 W max.

Gaussian current w/ avg. pwr of 8 W $\qquad W = 8I^2$

Does power exceed 50 W w/ probability ≥ 0.10?

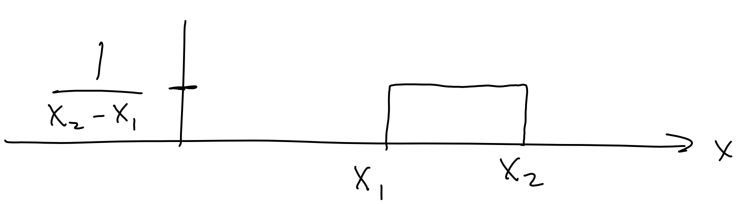
$Pr(W > 50) =$

$\sigma_I^2 = ?$

<u>uniform distribution</u>

- all values within a range are equally likely
  - quantization in A/D conversion
  - phase angle of sinusoid at receiver

$f(x) =$



$\overline{X} = \int x f_X(x)\,dx$

$$\overline{X^2} = \int_{x_1}^{x_2} \frac{1}{x_2 - x_1} x^2 \, dx = \frac{1}{3} \frac{x_2^3 - x_1^3}{x_2 - x_1}$$

$$= \frac{1}{3} \frac{(x_2 - x_1)(x_2^2 + x_1 x_2 + x_1^2)}{x_2 - x_1}$$

$$= \frac{x_2^2 + x_1 x_2 + x_1^2}{3}$$

$$\sigma^2 = \overline{X^2} - (\bar{X})^2 =$$

For the case of a centered quantizer:

$\bar{X} = 0$

$x_1 = -\frac{\Delta}{2}$, $x_2 = \frac{\Delta}{2}$ $\qquad \sigma^2 = \frac{1}{12} \Delta^2$

Ex: 16-bit quantizer

10 V_{p-p} signal

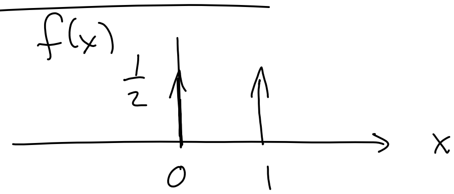
Δ = volts per level =

$$\sigma_x^2 = \frac{1}{12}\Delta^2 \Rightarrow \sigma_x =$$

$$SNR = 10 \log_{10} \frac{V^2}{\sigma_x^2}$$

$$=$$

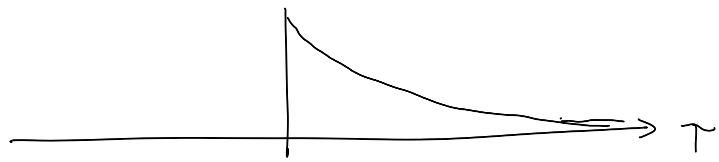
Delta distributions



delta distribution models pdf's that have discrete outcomes

Exponential distribution

$$f(\tau) = \begin{cases} \frac{1}{\bar{\tau}} \exp\left(\frac{-\tau}{\bar{\tau}}\right), & \tau \geq 0 \\ 0, & \tau < 0 \end{cases}$$



Ex: A PC CPU has an MTF of 8 yrs. What is the probability that the PC will outlive its useful life (3 yrs)?

$Pr(\tau > 3) =$

$Pr(\text{fail in } 1^{st} \text{ year}) =$

Conditional pdf's

- Conditional probability distribution for random X given event M is

$$F(x \mid M) = \Pr(X \le x \mid M)$$

$$=$$

- meets all characteristics of a distribution function

Ex: Let $M = \{X \le m\}$

Then $F(x \mid M) = \Pr(X \le x \mid X \le m) =$

If $x \ge m$, then $\Pr(X \le x, X \le m) =$

Thus, $F(x \mid M) =$, $x \ge m$

If $x \le m$, then $\Pr(X \le x, X \le m) =$

$F(x \mid M) =$, $x \le m$

$$F(x \mid X \le m) = \begin{cases} & x \le m \\ & x \ge m \end{cases}$$

- Conditional pdf is

$$f(x|M) =$$

This has all the standard pdf properties.

For previous example,

$$f(x|X \le m) = \begin{cases} & , \quad x \le m \\ & , \quad x \ge m \end{cases}$$

- conditional mean

$$E[g(X)|M] = \int_{-\infty}^{\infty} g(x)\, f(x|M)\, dx$$

$$E[X|M] =$$

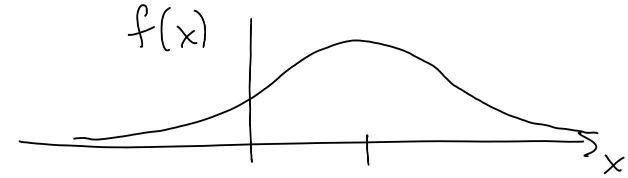
Ex: Let X be Gaussian with

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\bar{X})^2}{2\sigma^2}}$$

What is the mean value of the random variable given that we know $X \le \bar{X}$?

$$E[X|M] =$$

$$f(x|X \le \bar{X}) = \frac{f(x)}{Pr(X \le \bar{X})} =$$



$$E[X|X\le\bar X] = \int_{-\infty}^{\bar X} x\, 2\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\bar X)^2}{2\sigma^2}}\, dx$$

$$E[X|X\le\bar X] = 2\int_{-\infty}(u+\bar X)\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{u^2}{2\sigma^2}}\, du$$

$$= 2\bar X\int_{-\infty}^{0}\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{u^2}{2\sigma^2}}\, du + 2\int_{-\infty}^{0}\frac{1}{\sqrt{2\pi\sigma^2}}\, \underbrace{u\, e^{-\frac{u^2}{2\sigma^2}}}\, du$$

$=$

Ex: PC w/ mean time to failure (MTF) = 8 yrs. Already 5 yrs. old. What's pdf with this condition, and what is remaining life expectancy.

$$F\{\tau \mid T > 5\} = Pr(T \le \tau \mid T > 5)$$

$$f(\tau \mid T > 5) = \frac{d}{d\tau} F(\tau \mid T > 5) =$$

$$f(\tau) = \frac{1}{\bar{\tau}} e^{-\tau/\bar{\tau}} u(\tau)$$

$\bar{\tau} = 8$

$$E[T \mid T > 5] = ?$$

$$1 - F(5) = \int \frac{1}{8} e^{-\tau/8} d\tau =$$

$$f(\tau \mid T > 5) = \begin{cases} \\ \end{cases}$$

$$E[T \mid T > 5] = \int \tau f(\tau \mid T > 5)\, d\tau$$

$$= \int_5^\infty \tau\, e^{5/8}\, \frac{1}{8} e^{-\tau/8}\, d\tau$$

$\alpha = \tau - 5$

Ex: Suppose a space probe has three transmitters. The lifetime of each transmitter is a random variable described by

$$\frac{1}{\bar{\tau}} e^{-\tau/\bar{\tau}} u(\tau), \quad \text{where } \bar{\tau} = 3 \text{ yrs},$$

the MTF

What is the probability that at least one transmitter will work for five years?

Let p = probability of given transmitter surviving five years

$p =$

$$\sum_{k=1}^{3} Pr(k \text{ out of } 3 \text{ work}) =$$

Reliability

Reliability issues show up in several aspects of engineering! Ex: $Pr(0 \text{ out of } n \text{ transmitters}) \leq 0.01$

- design

Reliability competes with cost, complexity, size, weight, & maintainability.

Reliability draws together combinatorial probability and random variables.

Ex: A missile system has four independent critical microcontrollers, each of which has a probability of failure during a flight of $F_i = 0.01$. What is probability that the missile will fail?

$$\text{Let } R_i = 1 - F_i$$

$$F_S = Pr(M_f) =$$

- called <u>series</u> system -

In general,

$R_s =$

Ex: A drag racer has 2 small tires in front each with a probability of blowout in a race of 0.05. The two back tires have a probability of blowout of 0.02. The car won't finish if blowout.

$Pr(\text{no finish}) = ?$

$F_s =$

A space probe has 4 redundant hard drives in the computer system, $F_i = 0.01$ in one year. The probe is lost if there's no HD.

$$Pr(\text{probe lost in 1 yr}) = ?$$

$$P_f =$$

$$Pr(P_f) = F_p =$$

A parallel system is one in which the entire system will fail only if all components fail.

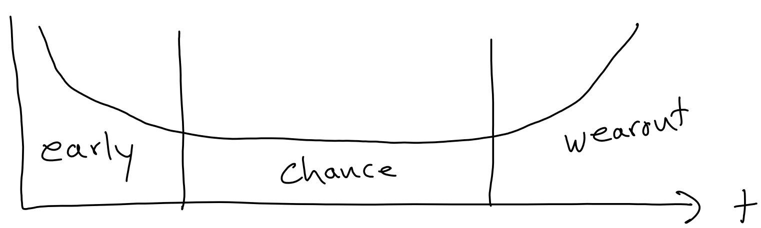
$$F_p =$$

Ex: A plane has two engines with $F_e = 10^{-4}$ and two redundant fuel delivery systems with $F_f = 10^{-5}$. The plane can land with one engine and either fuel system. What is probability of crash?

Failure-rate distributions

- Average probability of failure per unit time
  - called failure-rate function $z(t)$

typical $z(t)$



- early —
- chance —
- wearout

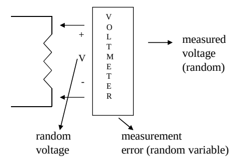
Usually assume $z(t) = \alpha$, a constant during useful life.

$\Longrightarrow$

Two Random Variables

- Previously, we only dealt with one random variable
- For the next few lectures, we will study two random variables
  - » How they are related to each other
  - » How we describe this relationship
- This is an intermediate step toward random signals

Example



Joint Probability Distribution

Joint Probability Distribution:
Consider two RV's X and Y

$$F(x, y) = \Pr(X \leq x, Y \leq y)$$

AND

Properties

1) $0 \leq F(x,y) \leq 1 \qquad -\infty < x < \infty \qquad -\infty < y < \infty$

2) $F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0$

3) $F(\infty, \infty) = 1$

Properties cont'd

4) F(x,y) is a nondecreasing function as either x or y, or both increase

5) $\left.\begin{aligned} F(\infty, y) &= F_Y(y) \\ F(x, \infty) &= F_X(x) \end{aligned}\right\}$ marginal distributions

$$F(\infty, y) = \Pr(X \leq \infty, Y \leq y) = \Pr(Y \leq y) = F_Y(y)$$

Joint Density Function

$f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$ order of differentiation is not important

$$f(x, y)dxdy = \Pr[x < X \leq x + dx, y < Y \leq y + dy]$$

1) $f(x,y) \geq 0 \qquad -\infty < x < \infty \qquad -\infty < y < \infty$

2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$

3) $F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dvdu$

Joint Density Function cont'd

4) $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

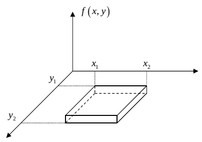
marginal densities

5) $\Pr[x_1 < X \leq x_2, y_1 < Y \leq y_2] = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f(x,y)dxdy$

Example

$$f(x,y)=\begin{cases}\dfrac{1}{(y_2-y_1)(x_2-x_1)} & \text{for} \quad x_1<x\le x_2,\ y_1<y\le y_2\\ 0 & \text{elsewhere}\end{cases}$$

$$\text{height}=\frac{1}{(y_2-y_1)(x_2-x_1)}$$



Expected Values

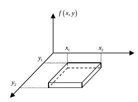
- Given two RV's *X* and *Y*, the expected value of *g*(*X*,*Y*) is

$$E\{g(X,Y)\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)f(x,y)\,dy\,dx$$

$$E\{XY\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)\,dy\,dx$$ is called the *correlation*.

Example

$$f(x,y)=\begin{cases}\dfrac{1}{(y_2-y_1)(x_2-x_1)} & \text{for} \quad x_1<x\le x_2,\ y_1<y\le y_2\\ 0 & \text{elsewhere}\end{cases}$$



$$E\{XY\}=\int_{x_1}^{x_2}\int_{y_1}^{y_2}xy\left(\frac{1}{(y_2-y_1)(x_2-x_1)}\right)dy\,dx$$

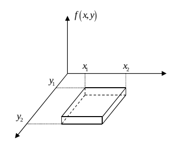
$$=\frac{1}{(y_2-y_1)(x_2-x_1)}\left[\frac{x^2}{2}\Big|_{x_1}^{x_2}\right]\left[\frac{y^2}{2}\Big|_{y_1}^{y_2}\right]=\frac{1}{4}(x_1+x_2)(y_1+y_2)$$

Example cont'd

$$f_X(x)=\int_{y_1}^{y_2}\frac{1}{(y_2-y_1)(x_2-x_1)}dy$$

$$=\frac{1}{(y_2-y_1)(x_2-x_1)}\left[y\big|_{y_1}^{y_2}\right]=\frac{1}{x_2-x_1}$$

$x_1<x\le x_2$



Similarly,

$$f_Y(y)=\int_{x_1}^{x_2}\frac{1}{(y_2-y_1)(x_2-x_1)}dx=\frac{1}{y_2-y_1} \qquad y_1<y\le y_2$$

Example

- Rectangular semiconductor substrate with dimensions having mean values of 1cm and 2cm.
  - » Actual dimensions are independent and uniformly distributed between ±0.01cm of their means

a) Probability that both dimensions are larger than their mean values by 0.005 cm.

$$\Pr(1.005<X,2.005<Y)=\int_{2.005}^{2.010}\int_{1.005}^{1.010}\frac{1}{0.020^2}dxdy$$

$$=\frac{0.005^2}{0.020^2}$$

$$=\frac{1}{16}$$

Example cont'd

b) Probability that the larger dimension is greater than its mean value by 0.005cm and the smaller dimension is less than its mean value by 0.005 cm.

$$\Pr(X\le 0.995,2.005<Y)=\int_{2.005}^{2.010}\int_{0.990}^{0.995}\frac{1}{0.020^2}dxdy$$

$$=\frac{0.005^2}{0.020^2}$$

$$=\frac{1}{16}$$

Example cont'd

c) The mean value of the area of the substrate

$$\begin{aligned} \overline{\mathrm{A}} &= E\{A\} \\ &= E\{XY\} \\ &= \int_{0.99}^{1.01}\int_{1.99}^{2.01} xy \frac{1}{(0.02)^2} dydx \\ &= \frac{1}{(0.02)^2}\left[\frac{x^2}{2}\Big|_{0.99}^{1.01}\right]\left[\frac{y^2}{2}\Big|_{1.99}^{2.01}\right] \\ &= 2 \end{aligned}$$

Joint Density Function cont'd

Ex: Two RV's X and Y $\quad f(x,y) = \begin{cases} Ae^{-(2x+3y)} & x \ge 0, y \ge 0 \\ 0 & x < 0, y < 0 \end{cases}$

a) Value of A

$$\begin{aligned} 1 &= \int_0^\infty\int_0^\infty Ae^{-(2x+3y)} dxdy \\ 1 &= A\int_0^\infty e^{-3y}\left[\int_0^\infty e^{-2x}dx\right]dy = A\int_0^\infty e^{-3y}\left[-\frac{1}{2}e^{-2x}\Big|_0^\infty\right]dy \\ 1 &= \frac{A}{2}\int_0^\infty e^{-3y}dy = \frac{A}{2\cdot 3} \\ A &= 6 \end{aligned}$$

Joint Density Function cont'd

b) $\Pr\left(X < \frac{1}{2}, Y < \frac{1}{4}\right) = 6\int_0^{\frac{1}{2}}\int_0^{\frac{1}{4}} e^{-(2x+3y)}dydx = 6[\frac{1}{2}(1-e^{-1})\frac{1}{3}(1-e^{-\frac{3}{4}})]$=0.3335

c) The expected value of XY.

$$\begin{aligned} E\{XY\} &= 6\int_0^\infty\int_0^\infty xye^{-(2x+3y)}dydx \\ &= 6\int_0^\infty xe^{-2x}\left[\int_0^\infty ye^{-3y}dy\right]dx \\ &= 6\times\frac{1}{4}\times\frac{1}{9} = \frac{1}{6} \end{aligned}$$

Ex: Let $f_X(x) = \frac{1}{4\sqrt{\pi}} e^{-\frac{x^2}{16}}, f_N(n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{n^2}{2}}$

$$\begin{aligned}
f_N(y-x)f_X(x) &= \frac{1}{4\sqrt{\pi}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{16}} e^{-\frac{(y-x)^2}{2}} \\
&= \frac{1}{4\pi\sqrt{2}} e^{-\frac{1}{16}[x^2+8y^2-16xy+8x^2]} \\
&= \frac{1}{4\pi\sqrt{2}} e^{-\frac{1}{16}[9x^2+8y^2-16xy]} \\
&= \frac{1}{4\pi\sqrt{2}} e^{-\frac{9}{16}[x^2+\frac{8}{9}y^2-\frac{16}{9}xy]} \\
&= \frac{1}{4\pi\sqrt{2}} e^{-\frac{9}{16}[(x-\frac{8}{9}y)^2+\frac{8}{81}y^2]} \\
&= \frac{1}{\sqrt{2\pi(9)}} \frac{1}{\sqrt{2\pi(8/9)}} e^{-\frac{y^2}{18}} e^{-\frac{(x-\frac{8}{9}y)^2}{16/9}}
\end{aligned}$$

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_N(y-x)f_X(x)dx \\
&= \frac{1}{\sqrt{2\pi(9)}} e^{-\frac{y^2}{18}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\frac{8}{9})}} e^{-\frac{(x-\frac{8}{9}y)^2}{16/9}} dx \\
&= \frac{1}{\sqrt{2\pi(9)}} e^{-\frac{y^2}{18}}
\end{aligned}$$

$$\begin{aligned}
f(x|y) &= \frac{f_N(y-x)f_X(x)}{f_Y(y)} \\
&= \frac{\frac{1}{\sqrt{2\pi(9)}} \frac{1}{\sqrt{2\pi(8/9)}} e^{-\frac{y^2}{18}} e^{-\frac{(x-\frac{8}{9}y)^2}{16/9}}}{\frac{1}{\sqrt{2\pi(9)}} e^{-\frac{y^2}{18}}} \\
&= \frac{1}{\sqrt{2\pi(\frac{8}{9})}} e^{-\frac{(x-\frac{8}{9}y)^2}{16/9}}
\end{aligned}$$

Generalizing conditional probability

We saw that $F_X(x \mid M) = \dfrac{\Pr(X \le x, M)}{\Pr(M)}$

What happens if $\Pr(M) = 0$?

$$F_X(x \mid Y = y) = \frac{\Pr(X \le x, Y = y)}{\Pr(Y = y)}$$

$$\underline{\underline{F_X(x \mid Y = y)}} = \frac{\Pr[X \le x, Y = y]}{\Pr(Y = y)}$$

- use limits

$\approx$

$=$

$$= \frac{[F(x, y + \Delta y) - F(x, y)] / \Delta y}{[F_Y(y + \Delta y) - F_Y(y)] / \Delta y}$$

- conditional density

$$f_X(x \mid Y=y) = \frac{\partial F_X(x \mid Y=y)}{\partial x} =$$

$$f_Y(y \mid X=x) =$$

$$f(x \mid y) = \qquad , \quad f(y \mid x) =$$

Bayes theorem for pdf's:

$$f(y \mid x) = \frac{f(x \mid y)\, f_Y(y)}{f_X(x)} = ?$$

Previously,

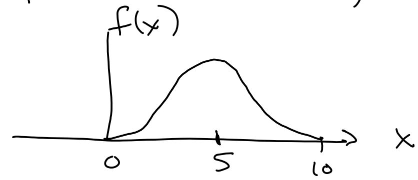
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy =$$

Estimating a signal in noise

- We want to measure X, but we can only measure a noisy version of X, which is Y:

$$Y = X + N \quad (N \text{ is random noise})$$

Find the pdf of X given the measurement Y — $f(x|y)$. From this, we can estimate the "best" guess for X.



From Bayes theorem,

$$f(x|y) =$$

For $f(y|x)$, N is the only random element since X is assumed known

$$f(y|x) =$$

$$f(x|y) = \frac{f_N(y-x)\, f_X(x)}{f_Y(y)} = ?$$

Since $f_Y(y) = \int_{-\infty}^{\infty} f(y|x) f_X(x)\,dx = \underbrace{\int_{-\infty}^{\infty} f_N(y-x)\, f_X(x)\, dx}_{\text{convolution of two pdf's}}$

* We're trying to estimate X. Note that $f_Y(y)$ is not a function of x and can be treated as a

constant scale factor.

Statistical independence

- When two random variables are statistically independent, knowledge of one random variable gives no information about the other.

$$f(x, y) =$$

If knowledge of y says nothing about x, then $f(x|y) = f_X(x)$.

$$f(x, y) =$$

- Correlation of statistically independent r.v.'s

$$E[XY] = \iint xy\, f(x,y)\, dx\, dy$$

$$= \iint xy\, f_X(x) f_Y(y)\, dx\, dy$$

- conditional density for stat. indep X, Y

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} =$$

and $f(y|x) =$

Ex: $f(x,y) = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}}$ stat. indep.?

Ex: $f(x,y) = \frac{1}{\sqrt{4\pi^2/3}} e^{-(x^2+xy+y^2)}$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx =$$

By the same reasoning,

$$f_X(x) =$$

$$f_X(x)\, f_Y(y) \stackrel{?}{=} f(x,y)$$

<u>Correlation between r.v.'s</u>

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dx\,dy$$

If X and Y have nonzero mean, then <u>covariance</u> may be more useful!

$$E\left[(X-\bar{X})(Y-\bar{Y})\right] =$$

The <u>correlation coefficient</u> is used to describe correlation without regard to magnitude of each variable:

$$\rho = E\left\{\left[\frac{X-\bar{X}}{\sigma_X}\right]\left[\frac{Y-\bar{Y}}{\sigma_Y}\right]\right\} =$$

$$= \frac{E[XY] - \bar{X}\bar{Y}}{\sigma_X \sigma_Y}$$

Properties:

1) $-1 \le \rho \le 1$

2) If X, Y are stat. indep., then $\rho = 0$.
   However, $\rho = 0 \not\Rightarrow X, Y$ are stat. indep.

Ex: $f(x,y) = \begin{cases} x+y, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$

$f_X(x) =$

$f_Y(y) =$

$\bar{X} =$

$\bar{Y} =$

$$\sigma_X^2 = \int_0^1 \left(x - \frac{7}{12}\right)^2 \left(x + \frac{1}{2}\right) dx = \int_0^1 \left(x^2 - \frac{7}{6}x + \frac{49}{144}\right)\left(x + \frac{1}{2}\right) dx$$

$$= \int_0^1 \left(x^3 - \frac{4}{6}x^2 - \frac{35}{144}x + \frac{49}{288}\right) dx$$

$$= \frac{1}{4}x^4 - \frac{2}{9}x^3 - \frac{35}{288}x^2 + \frac{49}{288}x \Big|_0^1$$

$$= \frac{1}{4} - \frac{2}{9} - \frac{35}{288} + \frac{49}{288} = \frac{11}{144}$$

$E[XY] =$

$$= \int_0^1 \left(\frac{1}{3}yx^3 + \frac{1}{2}y^2x^2\right)\Big|_0^1 dy$$

$$= \int_0^1 \left(\frac{1}{3}y + \frac{1}{2}y^2\right) dy$$

$$= \frac{1}{6}y^2 + \frac{1}{6}y^3 \Big|_0^1 = \frac{1}{3}$$

$$\rho = \frac{E[XY] - \bar{X}\,\bar{Y}}{\sigma_x \sigma_y} =$$

* ρ negative says that X & Y tend to move in opposite directions — larger X would imply smaller Y is more likely. (but only a little, since ρ is nearly 0.)

The corr. coef. is particularly useful to describe jointly Gaussian random variables.

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\bar{X})^2}{\sigma_x^2} + \frac{(y-\bar{Y})^2}{\sigma_y^2} - 2\rho\frac{(x-\bar{X})(y-\bar{Y})}{\sigma_x\sigma_y}\right]\right\}$$

If $\rho = 0$, we get

$f(x,y) =$

$\rho = 0 \implies$ stat. indep. in Gaussian case

Statistics

- Probability and Random Variables
  - » Assume we know how the RV behaves – either by assumption or physics of the problem.
- Statistics
  - » Analysis of data (specific values of an RV)

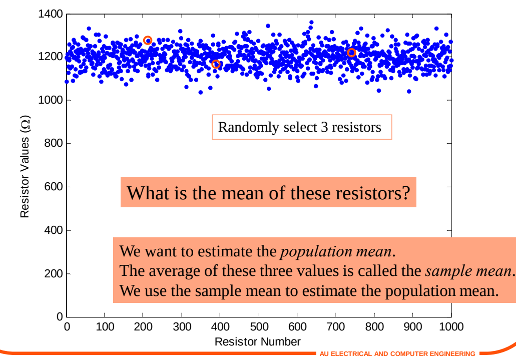
Applications of Statistics

- Sampling Theory
  - » How do we select samples from a large population?
  - » Examples
    - – Public opinion polls
    - – IC manufacturing
- Hypothesis Testing
  - » How do we decide which of two hypotheses are true?
  - » Does my electronic fuel injector really improve gas mileage?
- Linear Regression
  - » Derive a linear equation that describes the data.
  - » How much worse gas mileage do I get for each MPH faster that I drive?

Terminology

- Population (*N*)
  - » Collection of data being studied
    - – All registered voters
    - – All IC's from a manufacturing line
- Sample (*n*)
  - » Part of the population that is selected at random
    - – Voters contacted in a survey
    - – IC's selected for test
- Note:
  - » *N* = population size
  - » *n* = sample size

Example – A bin with 1000 Resistors



Example

- Let X be the value of a resistor from this bin of 1000 resistors
- Select 3 resistors at random

$$x_1 = 1133\Omega$$
$$x_2 = 1205\Omega$$
$$x_3 = 1250\Omega$$

- The mean of this sample (called the *sample mean*) is: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{3}\sum_{i=1}^{3} x_i = 1196\Omega$

Example

- How many resistors do we need in our sample to ensure that the sample mean is within ±1Ω of the true mean?

More Terminology / Notation

Notation: (capital letters; RV's, lower case: particular value)

X	random variable representing the population
X_i	an arbitrary sample taken from the population
x	a particular value of X
x_i	a particular sample taken from the population
$\bar{X}$	true (population) mean
$\hat{\bar{X}}$	theoretical sample mean
$\bar{x}$	actual sample mean (the mean of a particular sample)

Example

- Usually we are interested in the mean of an arbitrary set of samples, X_i

$$\hat{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

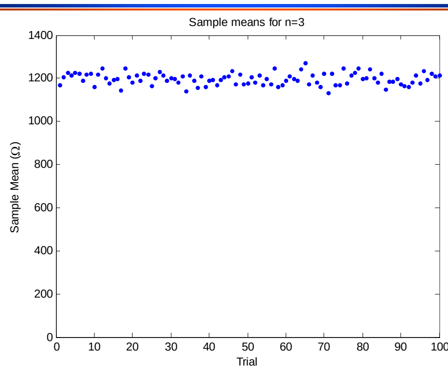
  » Note that $\hat{\bar{X}}$ is itself a random variable
- Suppose we picked another set of 3 resistors from the batch

$x_1 = 1275\Omega$
$x_2 = 1198\Omega$
$x_3 = 1185\Omega$

the mean of this sample is $\bar{x} = 1219.33\Omega$
(previous mean was 1196Ω)

- The mean of a particular sample depends on which 3 resistors you pick.

Example



Two Expressions for the Sample Mean

- Note we have two expressions for the sample mean
- $\hat{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i$ The mean of n arbitrary samples
  » A random variable used for theoretical analysis
- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ The mean of n actual samples
  » A number computed from the data
- $\bar{x}$ is a particular value of $\hat{\bar{X}}$
  » Same style of notation as in random variables, e.g. $\Pr(X<x)$

Mean of the Sample Mean

- What is the mean of the sample mean ?

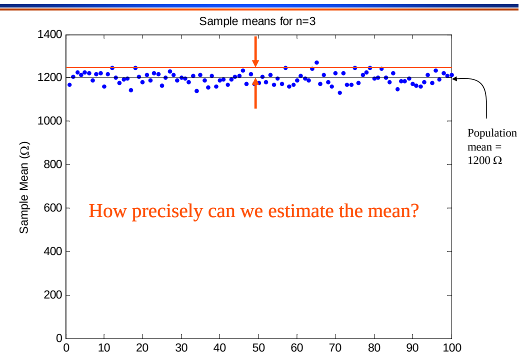
$$E\left\{\hat{\bar{X}}\right\} = E\left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\} = \frac{1}{n}\sum_{i=1}^{n} E\{X_i\} = \frac{1}{n}\sum_{i=1}^{n} \bar{X} = \bar{X}$$

population mean (true mean)

where X_i is an arbitrary sample from the population X

- On average, the sample mean equals the true mean
  » This is called an *unbiased* estimator

Example



Variance of the Sample Mean

- On average the sample mean equals the true mean
- How much does the sample mean vary about the true mean?
  - » How precise is the sample mean?
- We need to compute the variance of the sample mean

Variance of the Sample Mean

Assume $N>>n$

$$\mathrm{Var}\left(\hat{\bar{X}}\right)=E\left\{\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2\right\}-\left(\bar{X}\right)^2$$

$$=E\left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j\right\}-\left(\bar{X}\right)^2=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left\{X_iX_j\right\}-\left(\bar{X}\right)^2$$

assuming X_i and X_j are independent

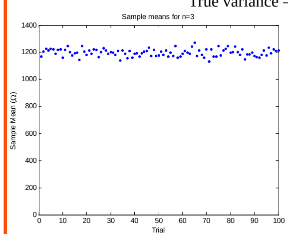
$$E\left\{X_iX_j\right\}=\begin{cases}\overline{X^2} & \text{for } i=j\\ \left(\bar{X}\right)^2 & \text{for } i\neq j\end{cases}$$

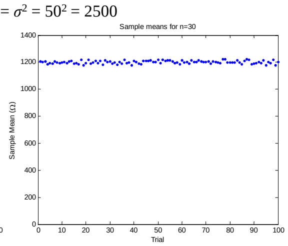
$$\mathrm{Var}\left(\hat{\bar{X}}\right)=\frac{1}{n^2}\left[n\overline{X^2}+\left(n^2-n\right)\left(\bar{X}\right)^2\right]-\left(\bar{X}\right)^2$$

$$=\frac{\overline{X^2}-\left(\bar{X}\right)^2}{n}=\frac{\sigma^2}{n}$$

Example

True variance = $\sigma^2 = 50^2 = 2500$





$$\mathrm{Var}(\hat{\bar{X}})=\frac{\sigma^2}{n}=\frac{2500}{3}=833.33\Omega^2$$

$$\mathrm{Std}(\hat{\bar{X}})=\sqrt{\mathrm{Var}(\hat{\bar{X}})}=28.87\Omega$$

$$\mathrm{Var}(\hat{\bar{X}})=\frac{\sigma^2}{n}=\frac{2500}{30}=83.33\Omega^2$$

$$\mathrm{Std}(\hat{\bar{X}})=\sqrt{\mathrm{Var}(\hat{\bar{X}})}=9.13\Omega$$

Variance of the Sample Mean

- The variance of the sample mean is

$$\mathrm{Var}\left(\hat{\bar{X}}\right)=\frac{\sigma^2}{n}$$

  - » The sample mean gets closer to the true mean as the sample size increases
- The above formula works when $N>>n$.
  - » If the population size is small, it also works when the samples can be returned to the population after testing.
    - – Called *sampling with replacement*.

Sampling Without Replacement

- When the population size is small and the samples are not replaced, the mean of the remaining population may be different from the original population. In this case

$$\mathrm{Var}\left(\hat{\bar{X}}\right)=\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

Example

<u>Ex:</u> endless production line produces diodes randomly tested for reverse current I_{-1} and forward current I_1. If I_{-1} has true mean of 10^{-6} A and variance of 10^{-12} A^2, how many diodes must be tested to get a sample mean within 5% of the true mean?

$N >> n$, so use $\mathrm{Var}(\hat{\bar{X}})=\dfrac{\sigma^2}{n}$

$$\sqrt{\mathrm{Var}\left(I_{-1}\right)}<(0.05)\left(10^{-6}\right)$$

$$\sqrt{\frac{10^{-12}}{n}}<5\times10^{-8}$$

$$n>400$$

Example

<u>Ex:</u> *A production line produces 30 diodes* randomly tested *without replacement* for reverse current I_{-1} and forward current I_1. If has true mean of 10^{-6} A and variance of 10^{-12} A^2, how many diodes must be tested to get a sample mean within 5% of the true mean?

N is small, so use $\mathrm{Var}(\hat{\hat{X}}) = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$

$$\sqrt{Var\left(I_{-1}\right)} < (0.05)\left(10^{-6}\right)$$

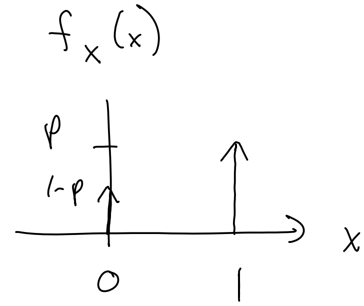
$$\sqrt{\frac{10^{-12}}{n}\left(\frac{30-n}{30-1}\right)} < 5\times10^{-8}$$

$$n > 27.97 \rightarrow 28$$

Note that when N=n, the variance of the sample mean is zero.

binomial

- opinion polls
- coin flips
- reliability

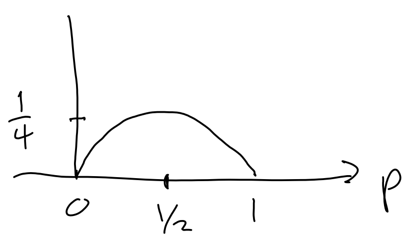


$\overline{X} = E[X] =$

$=$

$\overline{X^2} = E[X^2] =$

$\sigma^2 = \overline{X^2} - (\overline{X})^2 =$



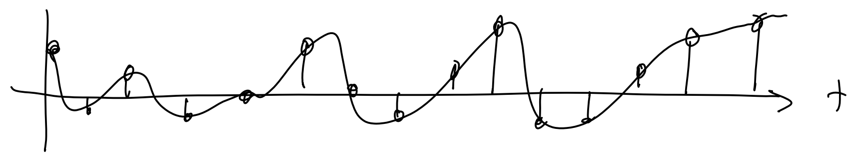
$\frac{d\sigma^2}{dp} =$

$var(\hat{\overline{X}}) = \frac{\sigma^2}{n} =$

std. dev $(\hat{\overline{X}}) \leq$

Suppose we want std. dev. ≤ 0.01

PDF of sample mean



Sample a random waveform at equally spaced intervals — assume samples are independent

- true mean of 10 V
- true var of 9 V^2

$$\mathrm{var}(\hat{X}) =$$

How many samples do we need to estimate the mean to within 1% of its true value?

Suppose we use $n = 900$, and we calculate $\hat{X}$.
Are we guaranteed that our estimate is within 1% of the true mean?

Since the sample size is large ($n = 900$), ______________ allows us to approximate the distribution of $\hat{X}$ as a Gaussian regardless of the distribution of X_i.

$Pr(9.9 \le \hat{X} \le 10.1) =$

* Gaussian R.V. has 68% chance of being within $\pm\sigma$ of its mean.

Ex: 100 bipolar transistors

What is mean value of the current gain β?

true mean & variance are $\bar{\beta} = 120$

$\sigma_\beta^2 = 25$

Suppose we want std. dev. of $\hat{\beta}$ to be within 1% of $\bar{\beta}$.

$var(\hat{\beta}) =$

$$\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

Because we have a small population, we use sampling without replacement formula.

What is the probability that the sample mean is within $\pm 1\%$ of the true mean?

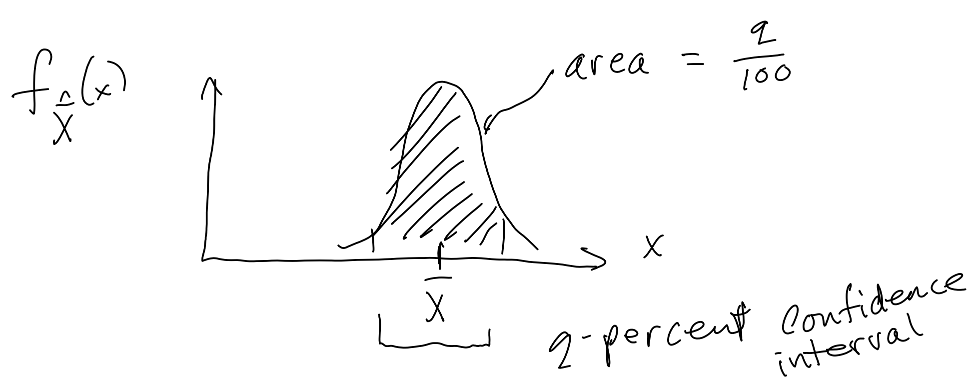
If the β are Gaussian,

If the β are <u>not</u> Gaussian,

<u>Confidence intervals</u>

Often in statistics, we want to know the <u>range</u> the sample mean (or any other estimate) will be in.

Define the <u>q-percent</u> <u>confidence</u> <u>interval</u>



The $q\%$- confidence interval for the sample mean is

σ : true std. dev. of each sample

: std. dev. of the sample mean

k is a constant that depends on q and density of $\hat{X}$.

$q =$

k is tabulated for common values of q.

$$Z = \frac{\hat{X} - \overline{X}}{\sigma/\sqrt{n}} \Rightarrow q = 100 \int_{-k}^{k} f_Z(z)\,dz$$

usually a Gaussian

$\Rightarrow$ can use $Q()$ table to infer k.

Ex: large population of resistors

$\overline{X} = 100\,\Omega$, $\sigma = 4\,\Omega$

Find 95% confidence limits for $n = 100$.

$q = 95 \quad \Rightarrow \quad k = \qquad$ (Table 4-1)

$$\frac{k\sigma}{\sqrt{n}} =$$

$$q = 100 \Pr\left[\bar{X} - k\frac{\sigma}{\sqrt{n}} \le \hat{X} \le \bar{X} + k\frac{\sigma}{\sqrt{n}}\right]$$

$$= 100\left[F\left(\bar{X} + k\frac{\sigma}{\sqrt{n}}\right) - F\left(\bar{X} - k\frac{\sigma}{\sqrt{n}}\right)\right]$$

$$= 100\left[\Phi\left(\frac{\bar{X} + k\frac{\sigma}{\sqrt{n}} - \bar{X}}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\bar{X} - k\frac{\sigma}{\sqrt{n}} - \bar{X}}{\sigma/\sqrt{n}}\right)\right]$$

$$=$$

Ex: randomly sample Gaussian waveform

$\bar{X} = 10$ 99% confidence interval:

$\sigma^2 = 9$

$n = 900$

Sample variance

- need varian   but not usually known

$$S^2 =$$

$$E[S^2] = \frac{n-1}{n}\sigma^2$$

$$\tilde{S}^2 = \frac{n}{n-1} S^2 \Rightarrow E[\tilde{S}^2] =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{X})^2$$

$\Rightarrow$ use $\tilde{S}^2$

Sample mean distributions

The sample mean is $\hat{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

What is the pdf of $\hat{X}$?

It is often convenient to work with a normalized r.v.:

$Z =$

which is a Gaussian zero-mean unit-variance r.v, if n is large enough.

If n is large ($n \geq 30$),

Case 1

If σ^2 is unknown and n is large,

$$Z = \frac{\hat{X} - \bar{X}}{\tilde{s}/\sqrt{n}}$$

Case 2

If n is small and σ^2 is unknown,

If $n < 30$ (and X_i are Gaussian), define the normalized sample mean r.v, as

$$T = \frac{\hat{X} - \bar{X}}{\tilde{s}/\sqrt{n}}$$

This r.v. is a Student's t-distribution with n-1 degrees of freedom.

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\;\Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{t^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

where $\nu = n-1$ (degrees of freedom)

$\Gamma(\cdot)$ = gamma function

$\Gamma(k+1) = k\,\Gamma(k)$ any k

$= k!$ integer k

$\Gamma(1) = \Gamma(2) = 1$, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

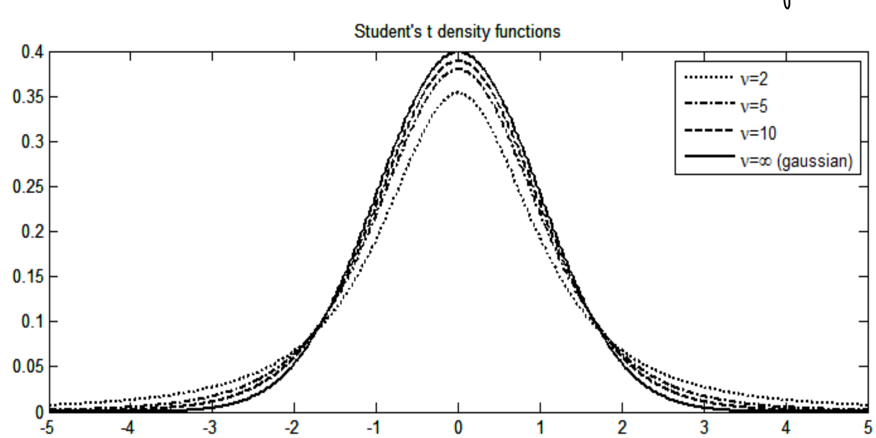
$\sigma^2 \backslash n$	small	large
known	Z	Z
unknown	T	Z with $\tilde{S}$

X_i Gaussian

$\sigma^2 \backslash n$	small	large
known	*	Z
unknown	*	Z with $\tilde{S}$

X_i not Gaussian

* special case



hypothesis testing

Ex: a particular car model gets 26 mpg with a std. dev. of 5 mpg.

A new fuel injection system is thought to improve gas mileage.

36 gas mileage tests are performed with $\bar{x} = 28.04$ mpg. Assume σ is unaffected.

Is gas mileage better with 95% confidence?

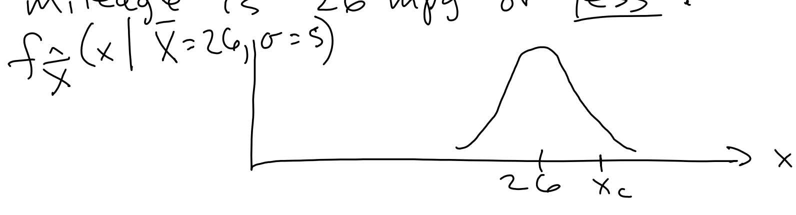
We will assume H_0 is true (which is why it needs the equality) and test if the data are consistent with this assumption.

If not consistent, we reject H_0 and accept H_1.

In other words, we'll say...

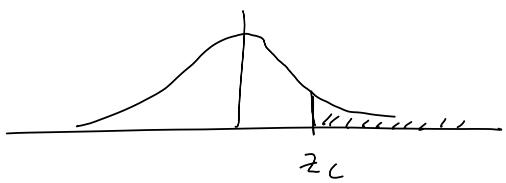
But how much larger?

In our example, we first assume the true mean mileage is 26 mpg or less.

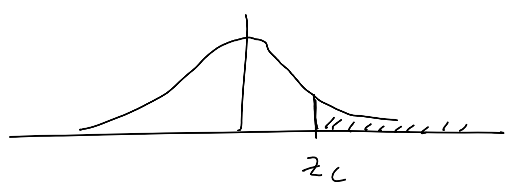


Compute a critical value x_c such that

$$\Pr(\hat{X} \le x_c) = 0.95$$



$$\Pr(Z \le z_c) = 0.95$$



$$\Rightarrow \Phi(z_c) = 0.95$$

$$1 - Q(z_c) = 0.95$$

$$Q(z_c) = 0.05$$

$$\Rightarrow z_c =$$

If the true gas mileage is unchanged, 95% of the sample means (from 36 samples) will be ≤ 27.36 mpg

Two possibilities: (1)

(2)

* called a one-sided test because the alternative only considers one side

Steps:

(1) construct 2 hypotheses H_0 & H_1:
- H_1 should be the complement of H_0
- equality should be with H_0.

(2) choose test statistic (normalized sample mean)

$$z = \frac{\bar{x} - \bar{X}}{\sigma/\sqrt{n}}$$

$$z \text{ or } t = \frac{\bar{x} - \bar{X}}{\tilde{s}/\sqrt{n}} \quad \text{if } \sigma \text{ is unknown}$$

(3) compute critical value of test statistic (z_c or t_c) based on desired confidence level.

(4) If z inside confidence interval, accept H_0

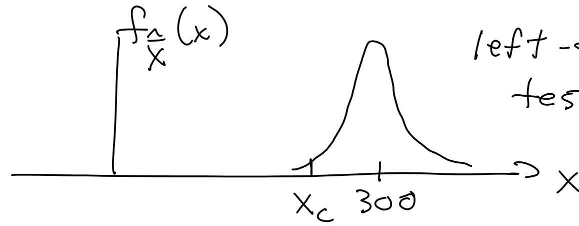
Ex: a manufacturer claims the breakdown voltage of their capacitors is 300 V or greater.

100 capacitors tested $\bar{x} = 290$ V, $\tilde{s} = 40$ V

Is the claim valid at the 99% confidence level?

① H_0:

H_1:



left-sided test

②

③

④

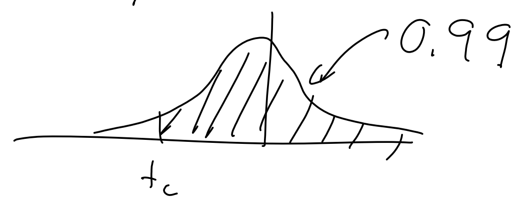
Ex: same as before but with $n = 9$

($\bar{x} = 290$, $\tilde{s} = 40$)

sample size is small + σ^2 is unknown $\Rightarrow T_n$

$\nu = n - 1 = \underline{\underline{8}}$ $\qquad t_8 = \dfrac{290 - 300}{40/\sqrt{9}} = -0.75$

$\Pr(T_8 > t_c) = 0.99$



$1 - F_{T_8}(t_c) = 0.99$

$1 - [1 - F_{T_8}(-t_c)] = 0.99$

$F_{T_8}(-t_c) = 0.99 \implies -t_c = 2.896$

$t_c = -2.896$

$t = -0.75 > t_c = -2.896$

$\implies$ accept null hypothesis (claim is valid)

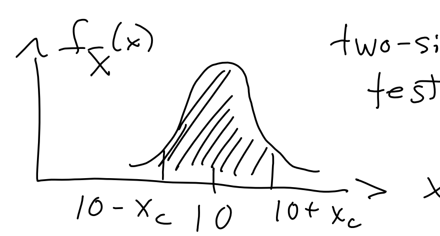
Ex: Zener diodes with mean breakdown voltage of 10 V. Deviations on either side are undesirable because they're used as voltage regulators.

test 100 diodes $\quad \bar{x} = 10.3\text{ V} \quad \tilde{S} = 1.2\text{ V}$

Is claim valid @ 95% confidence level?

① H_0 :

H_1 :



two-sided test

②

(3)

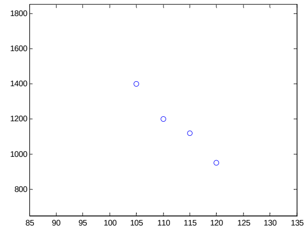
(4)

Ex: same as before but $n = 36$

Curve Fitting and Linear Regression

Ex: Four light bulbs are tested to determine the relationship between lifetime and operation voltage.

i	V	Hrs
1	105	1400
2	110	1200
3	115	1120
4	120	950
	x_i	y_i



Curve Fitting and Linear Regression

- We would like a mathematical relationship between the operation voltage and lifetime so we could estimate lifetime for other operation voltages.
  - » For example, 90V, 112V, 130V
- One way to do this is to fit a straight line to the data. This is called *linear regression*.
  - » We can fit other curves as well polynomials, Fourier series.

Curve Fitting and Linear Regression

- Our data model is $y_i = mx_i + b$
- Only need two points to solve for m and b, but we would like to use all the data available.
- Define $d_i = y_i - mx_i - b$
- Choose m and b to minimize the *sum of squared differences*

$$SSD = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} \left[y_i - (mx_i + b) \right]^2$$

$$\frac{\partial}{\partial m} SSD = \sum_{i=1}^{n} \frac{\partial}{\partial m} \left[\left[y_i - (mx_i + b) \right]^2 \right] = 0 \qquad \frac{\partial}{\partial b} SSD = \sum_{i=1}^{n} \frac{\partial}{\partial b} \left[\left[y_i - (mx_i + b) \right]^2 \right] = 0$$

$$-\sum_{i=1}^{n} x_i y_i + m \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i = 0 \qquad \sum_{i=1}^{n} y_i + m \sum_{i=1}^{n} x_i + bn = 0$$

$$\sum_{i=1}^{n} x_i y_i = m \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i \qquad \sum_{i=1}^{n} y_i = m \sum_{i=1}^{n} x_i + nb$$

Curve Fitting and Linear Regression

Solving for m and b yields:

slope ⟵
$$m = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}$$

y-intercept ⟵
$$b = \overline{y} - m\overline{x}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Curve Fitting and Linear Regression cont’d

For our example:

$n = 4$

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{4}\times 450 \qquad \overline{x} = 112.5 \qquad \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{4}\times 50750$$

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{4}\times 4670 \qquad \overline{y} = 1167.5 \qquad \frac{1}{n}\sum_{i=1}^{n} y_i^2 = \frac{1}{4}\times 5556900$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i = \frac{1}{4}\times 521800$$

$$m = \frac{\frac{1}{4}(521800) - \left(\frac{1}{4}\times 450\right)\left(\frac{1}{4}\times 4670\right)}{\frac{1}{4}(50750) - \left(\frac{1}{4}\times 450\right)^2} = \text{-}28.6$$

$$b = 1167.5 - (-28.6)(112.5) = 4385$$

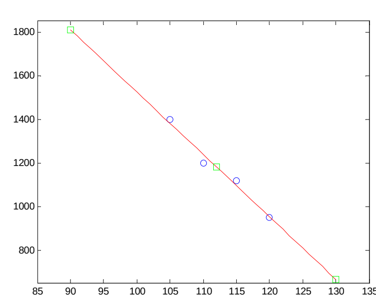
$$y = -28.6x + 4385$$

90V $\rightarrow$ -28.6(90) + 4385 = 1811 hrs

112V $\rightarrow$ -28.6(112) + 4385 = 1181.8 hrs

130V $\rightarrow$ -28.6(130) + 4385 = 667 hrs

Data, Line, and Predicted Values



Correlation Coefficient

Can compute correlation coefficient for data

How correlated is operation voltage with lifetime?

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2\right)}}$$

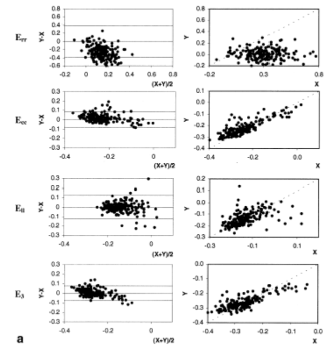
Compare to the theoretical formula:

$$\rho = \frac{E\{XY\} - \overline{X}\,\overline{Y}}{\sigma_X \sigma_Y}$$

highly negatively correlated

For our example: $r = \frac{\frac{1}{4}(521800) - \left(\frac{1}{4}\times 450\right)\left(\frac{1}{4}\times 4670\right)}{\sqrt{\left(\frac{1}{4}(50750) - \left(\frac{1}{4}\times 450\right)^2\right)\left(\frac{1}{4}(5556900) - \left(\frac{1}{4}\times 4670\right)^2\right)}} = -0.9883$

Example



$r = -0.05$

$r = 0.83$

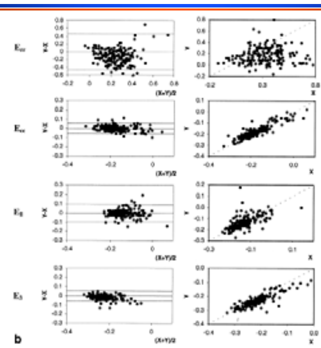
FIG. 6. Bland-Altman (left) and correlation (right) plots for comparing UNTETHER (Y) and user-supervised (X) strain measurements in the (a) subendocardium, (b) mid-wall, and (c) subepicardium. In the Bland-Altman plots, lines above and below the difference = 0 line represent ±2 SD of the difference. In the correlation plots, the dashed line represents perfect correlation.

$r = 0.50$

$r = 0.80$

From Denney Jr., TS, et al, *Magnetic Resonance in Medicine*, 2003

Example



$r = 0.09$

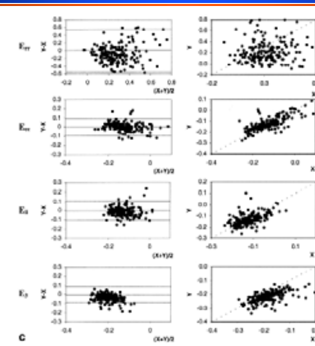
$r = 0.91$

$r = 0.63$

$r = 0.86$

FIG. 6. Continued.

Example



$r = 0.08$

$r = 0.74$

$r = 0.58$

$r = 0.60$

FIG. 6. Continued.

Random processes

* A r.p. is a random time function. $\{x(t)\}$ represents the entire collection of time functions

* R.P.'s show up in a large variety of engineering applications — any time a random signal is observed over time!

Classifications

— values of the sample function can take on any values within a specific range.

— sample function can only take on a discrete or isolated set of values.

— some discrete and some continuous elements

Continuous

discrete

mixed

stationary vs. nonstationary

If no marginal or joint pdf's of $X(t)$ depend on the choice of the time origin, the r.p. is...

Analysis of systems with nonstationary input is very difficult.

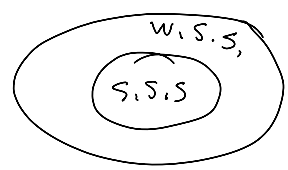
- If $\overline{X(t_1)}$ is independent of t_1 and $\overline{X(t_1)X(t_2)}$ only depends on $t_2 - t_1$, the r.p. is called

These conditions are not as strict as the previous definition of stationarity.

wss $\Longrightarrow$

Ex: Let $f_{X(t)}(x(t)) = \dfrac{1}{\sqrt{2\pi e^{-|t|}}} e^{-x^2(t)/2e^{-|t|}}$

Is it stationary?



Ex: Consider a random process defined by the number of people in the grocery store at a given time.

Stationary?

Ergodic vs. nonergodic

- Every member of the ensemble exhibits the same statistical behavior as the whole ensemble

$$X(t) = Y \cos \omega_0 t$$

where Y is random.

An important property of ergodic processes is that means and moments can be determined by time averages as well as ensemble (statistical) averages.

$$\overline{X^n} = \int_{-\infty}^{\infty} x^n f(x)\, dx =$$

- Ergodicity is usually assumed for physical processes so that time averages can be used to estimate ensemble averages.

Correlation functions

In many engineering applications, a complete probability description is either unnecessary or impossible to obtain. Instead, we can characterize random processes by certain average values.

$$E[X_1 X_2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2)\, dx_1\, dx_2$$

This is a function of t_1, t_2 — called the <u>autocorrelation function</u> and denoted by

$$R_x(t_1, t_2) =$$

For stationary random processes, correlation is independent of time origin. Therefore,

$$R_x(t_1 - t_1, t_2 - t_1) =$$

Let $\tau = t_2 - t_1$, and suppress the "0" argument:

$$R_x(\tau) =$$

Since the choice of t_1 has no effect on $R_x(\tau)$, we write

$$R_x(\tau) =$$

We can also define a time autocorrelation function:

$$\mathcal{R}_x(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\, x(t+\tau)\, dt$$

$$= \langle x(t)\, x(t+\tau) \rangle$$

If the process is ergodic, then

What does autocorrelation tell us?

Ex: Is there an echo or ghost in signal?

Let $X(t)$ be zero-mean and stationary and let $Y(t) = X(t) - \rho X(t+\tau)$.

($Y(t)$ is also zero mean.)

The mean-square value of $Y(t)$ measures how much of $X(t+\tau)$ is contained in $X(t)$.

$E\{[Y(t)]^2\} =$

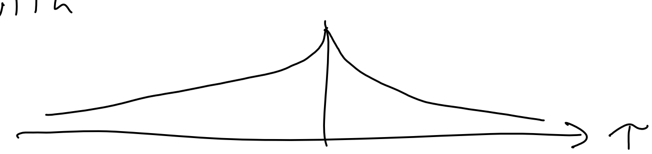
We can find the true value of ρ by minimizing the expected squared difference between $X(t)$ and $X(t+\tau)$:

$$\frac{d\sigma_Y^2}{d\rho} =$$

$$\sigma_Y^2 =$$

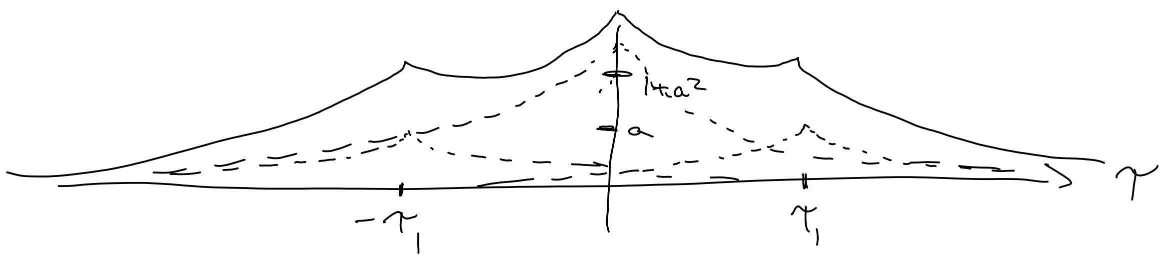
$\Rightarrow$ look for peak in $R_X(\tau)$, $\tau \neq 0$
- this reveals shift of the ghost signal

Let $Z(t) = X(t) + aX(t+\tau_1)$
where $X(t)$ is a sample function from a stationary r.p. with
$R_X(\tau) = e^{-|\tau|}$



Find $R_Z(\tau)$.

$R_Z(\tau) = E[Z(t)\, Z(t+\tau)]$



$E[X(t)X(t+\tau)] =$

Autocorrelation properties

1) $R_x(0) = E[X(t)\,X(t)] = \overline{X^2}$

Note that if the r.p. has $\bar{X} = 0$, then $R_x(0) = \sigma_x^2$.

2) $R_x(\tau) = R_x(-\tau)$ (an even function of τ)
(symm. about vertical axis)

$R_x(-\tau) = E[X(t)\,X(t-\tau)]$

$t = t' + \tau$ $\quad = E[X(t'+\tau)\,X(t')]$

$= E[X(t)\,X(t+\tau)]$

$= R_x(\tau)$

3) $|R_X(\tau)| \leq R_X(0)$

$$E[(X_1 \pm X_2)^2] = E[X_1^2 + X_2^2 \pm 2X_1X_2] \geq 0$$

$$E[X_1^2 + X_2^2] \geq \mp E[2X_1X_2]$$

$$2R_X(0) \geq 2|R_X(\tau)|$$

$$R_X(0) \geq |R_X(\tau)|$$

4) If $\bar{X}(t) \neq 0$, then $R_X(\tau)$ will have a constant component.

Suppose $X(t) = \bar{X} + N(t)$, where $N(t)$ is zero mean

$$R_X(\tau) = E\{[\bar{X} + N(t)][\bar{X} + N(t+\tau)]\}$$

$$= E\{(\bar{X})^2 + \bar{X}N(t) + \bar{X}N(t+\tau) + N(t)N(t+\tau)\}$$

$$= (\bar{X})^2 + R_N(\tau)$$

5) If $X(t)$ has a periodic component, then $R_X(\tau)$ will also have a periodic component with the same period.

Ex: Let $X(t) = A\cos(\omega t + \theta)$, where A + ω are constant and θ is uniformly distributed in $[0, 2\pi)$.

$$f(\theta) = \begin{cases} \frac{1}{2\pi}, & 0 \leq \theta < 2\pi \\ 0, & \text{otherwise} \end{cases}$$

$R_X(\tau) =$

If $X(t) = A\cos(\omega t + \Theta) + N(t)$, where $N(t)$ has no periodic component,

then $R_X(\tau) = \frac{A^2}{2}\cos\omega\tau + R_N(\tau)$

6) If $\{X(t)\}$ is ergodic and zero-mean, then

$$\lim_{|\tau|\to\infty} R_X(\tau) = 0$$

7) $R_X(\tau)$ cannot have arbitrary shape, even with symmetry and max @ $\tau = 0$.

$$\mathcal{F}\{R_X(\tau)\} = \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}\,d\tau \geq 0 \text{ for all } \omega.$$

$\Rightarrow$ no flat tops or discontinuities

<u>measurement</u>

- autocorr. used to model signals + systems, such as in controls, speech waveforms in DSP, images.
- autocorr. is not usually known in advance, so we must estimate it from samples of the r.p.

If $x(t)$ is observed from 0 to T,

In practice, we usually just sample $x(t)$ and approximate the integral with a summation:
$N =$ length of sampled $x(t)$, $\Delta t =$ sampling period

Since the values $x(k\Delta t)$ are just random variables, $\hat{R}_x(n\Delta t)$ is also a random variable.

How good is $\hat{R}_x(n\Delta t)$ as an estimate of $R_x(n\Delta t)$?

Let $X_k = x(k\Delta t)$

$$E[\hat{R}_x(n\Delta t)] = E\left[\frac{1}{N-n+1}\sum_{k=0}^{N-n} X_k X_{k+n}\right]$$

$X_N = x(N\Delta t)$

$$= \frac{1}{N-n+1}\sum_{k=0}^{N-n} E[X_k X_{k+n}]$$

$N\Delta t = T$

$$= \frac{1}{N-n+1}(N-n+1)\, R_x(n\Delta t)$$

$$= R_x(n\Delta t), \quad n = 0, \dots, M$$

* this is an unbiased estimate of $R_x(n\Delta t)$

* A more common estimator deals with the problem of variance better:

$$\hat{R}_x(n\Delta t) = \frac{1}{N+1} \sum_{k=0}^{N-n} X_k X_{k+n}$$

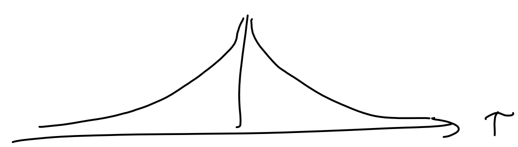
* This estimator is biased:

$$E[\hat{R}_x(n\Delta t)] =$$

$$\text{var}[\hat{R}_x(n\Delta t)] \le \frac{2}{N} \sum_{k=-N}^{N} R_x^2(k\Delta t)$$

$$\text{or} \quad \le \frac{2}{N\Delta t} \int_{-\infty}^{\infty} R_x^2(\tau)\, d\tau$$

Ex: $R_x(\tau) = 10^{-|\tau|}$



What range of τ values must be estimated to get all values of $R_x(\tau)$ such that

$$|R_x(\tau)| \ge 0.01 R_x(0)$$

* we only need $\tau \geq 0$ to calculate $R_x(\tau)$.

What if we want RMS error $< 10\%$ of $R_x(\tau)$?

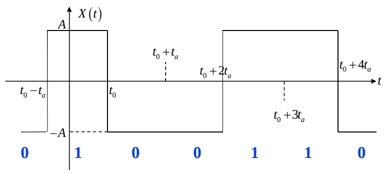
How many samples do we need?

use $\Delta t = 0.05$

note that $R_x(\tau)$ has no constant component.

$$\text{var}(\hat{R}_x(n\Delta t)) =$$

Random Binary Waveform



- Waveform changes sign every t_a seconds with equal probability
  - » Each pulse (bit) is independent

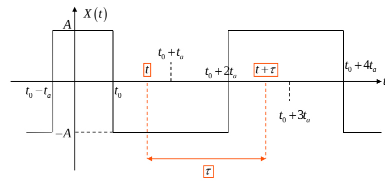
A: pulse amplitude

t_a : pulse width

t_0 : random time shift uniform on $[0, t_a]$

$E\{X(t)\}=0$

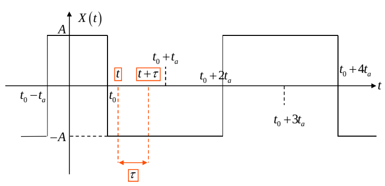
Random Binary Waveform



If $\tau > t_a$, then $X(t)$ and $X(t+\tau)$ are independent and $R_X(\tau)=0$

If $\tau < t_a$, then we need to look at the probability that t and t+τ are in the same t_a interval

Random Binary Waveform

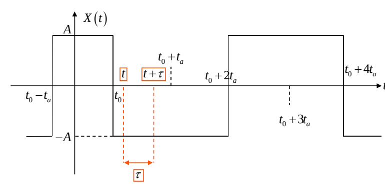


If t and t+τ are in the same interval (for $\tau \geq 0$):

$$t_0 \leq t \quad \text{and} \quad t+\tau < t_0+t_a$$
$$t+\tau-t_a < t_0$$

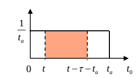
Combining these two inequalities yields: $t+\tau-t_a < t_0 \leq t$

Random Binary Waveform



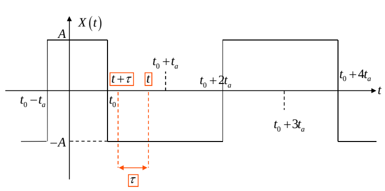
$P(t \text{ and t+}\tau \text{ in the same interval}) = P[t+\tau-t_a < t_0 \leq t]$

recall that t_0 is uniform on $[0, t_a]$



So, $P[t+\tau-t_a < t_0 \leq t] = \frac{1}{t_a}[(t)-(t+\tau-t_a)] = \frac{1}{t_a}[t_a-\tau]$

Random Binary Waveform

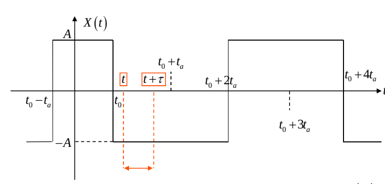


For $\tau < 0$

$$t_0 \leq t+\tau \quad \text{and} \quad t < t_0+t_a$$
$$t-t_a < t_0 \leq t+\tau$$

$P(t \text{ and t+}\tau \text{ in the same interval}) = \frac{1}{t_a}[t_a+\tau]$

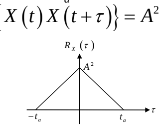
Random Binary Waveform



In general, $P(t \text{ and t+}\tau \text{ in the same interval}) = \frac{t_a-|\tau|}{t_a}$

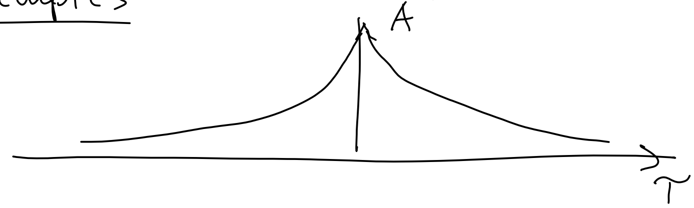
If t and t+τ in the same interval, then $E\{X(t)X(t+\tau)\}=A^2$

Therefore, $R_X(\tau)=\begin{cases} A^2\left[1-\frac{|\tau|}{t_a}\right] & 0 \leq |\tau| \leq t_a \\ 0 & |\tau| > t_a \end{cases}$

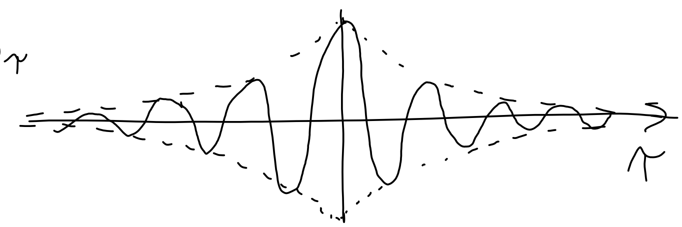


Autocorrelation examples

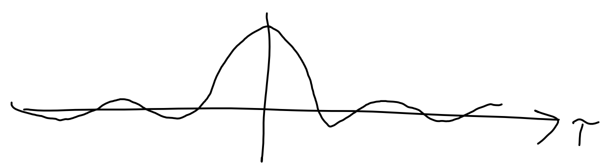
$$R_x(\tau) = A^2 e^{-\alpha|\tau|}$$



$$R_x(\tau) = A^2 e^{-\alpha|\tau|} \cos\beta\tau$$



$$R_x(\tau) = \frac{A^2 \sin \pi \gamma \tau}{\pi \gamma \tau}$$



Crosscorrelation functions

- sometimes we need to compare two <u>different</u> signals:

$$X_1 = X(t_1) \quad , \quad Y_2 = Y(t_1 + \tau)$$

crosscorr:

$$= R_{xy}(\tau) =$$

* Note that order of subscripts is important:

$$R_{YX}(\tau) = E[Y_1 X_2], \quad \text{where } Y_1 = Y(t_1)$$
$$X_2 = X(t_1 + \tau)$$

$R_{XY}(\tau)$ assumes X and Y are jointly stationary r.p.'s — $R_{XY}(\tau)$ only depends on time difference τ.

time crosscorr:

$$\mathcal{R}_{XY}(\tau) =$$

If jointly ergodic,

$$\mathcal{R}_{XY}(\tau) = R_{XY}(\tau)$$

Properties:

1) $R_{XY}(0)$ and $R_{YX}(0)$ are <u>not</u> mean-square values, but $R_{XY}(0) = R_{YX}(0)$.

2) Not even function in general, but

3) Max not necessarily at $\tau = 0$, but

$$|R_{XY}(\tau)| \le [R_X(0)\, R_Y(0)]^{1/2}$$

4) If X, Y are independent,

$$R_{XY}(\tau) = E[X_1 Y_2] =$$

Ex: Let $Y(t) = aX(t-T) + N(t)$

This models a problem in which $X(t)$ is known, and we want to find how much $X(t)$ has been delayed in measured signal $Y(t)$.

$R_{XY}(\tau) = E[X(t)\,Y(t+\tau)]$

If $N(t)$ is zero-mean random noise that is independent of $X(t)$, then

$R_{XN}(\tau) =$

$R_{XY}(\tau) =$

- Find τ that maximizes $R_{XY}(\tau)$

ELEC 3800: Spectrum Calculation Example

Let $X(t) = A + B\cos(\omega_1 t + \theta)$, where θ is random, and

$$f(\theta) = \begin{cases} \frac{1}{2\pi}, & 0 \le \theta < 2\pi \\ 0, & \text{elsewhere} \end{cases}$$

$$\mathcal{F}\{X(t)\} = 2\pi A\delta(\omega)$$

$$\begin{aligned} F_{X_T}(\omega) &= \int_{-T}^{T} [A + B\cos(\omega_1 t + \theta)] e^{-j\omega t} dt \\ &= \frac{2A\sin(\omega T)}{\omega} + B\left[\frac{e^{j\theta}\sin(\omega - \omega_1)T}{(\omega - \omega_1)} + \frac{e^{-j\theta}\sin(\omega + \omega_1)T}{(\omega + \omega_1)}\right] \end{aligned}$$

$$\begin{aligned} |F_{X_T}(\omega)|^2 = \frac{4A^2\sin^2\omega T}{\omega^2} &+ B^2\left[\frac{\sin^2(\omega - \omega_1)T}{(\omega - \omega_1)^2} + \frac{\sin^2(\omega + \omega_1)T}{(\omega + \omega_1)^2}\right] \\ &+ C(\omega)e^{j\theta} + C(-\omega)e^{-j\theta} + D(\omega)e^{j2\theta} + D(-\omega)e^{-j2\theta} \end{aligned}$$

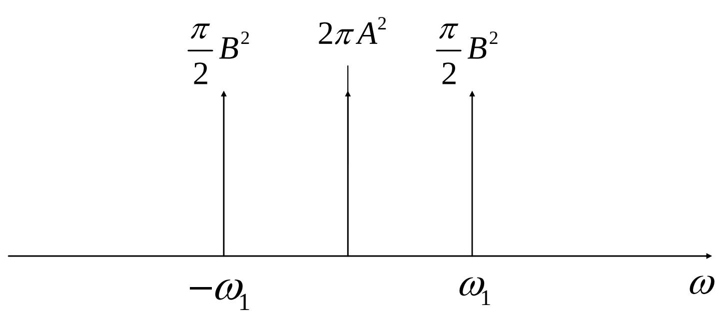
for some $C(\omega)$ and $D(\omega)$.

$$E[|F_{X_T}(\omega)|^2] = 4A^2\frac{\sin^2\omega T}{\omega^2} + B^2\left[\frac{\sin^2(\omega - \omega_1)T}{(\omega - \omega_1)^2} + \frac{\sin^2(\omega + \omega_1)T}{(\omega + \omega_1)^2}\right]$$

$$S_X(\omega) = \lim_{T\to\infty} \frac{E[|F_X(\omega)|^2]}{2T}$$

$$\lim_{T\to\infty} \frac{\sin^2\omega T}{\omega^2 T} = \pi\delta(\omega)$$

$$S_X(\omega) = 2\pi A^2\delta(\omega) + \frac{\pi}{2}B^2[\delta(\omega - \omega_1) + \delta(\omega + \omega_1)]$$



Spectral density

- frequency-domain representation of signals is a very important tool in analyzing signals & systems
- How do we represent random signals in frequency domain?
  - Fourier transform usually doesn't exist for stationary random processes, since we must have $\int_{-\infty}^{\infty} |x(t)|\, dt < \infty$

  - Instead, we define a finite-duration signal:

    $X_T(t) =$

Let $F_{X_T}(\omega) =$

Phase of $F_{X_T}(\omega)$ is usually random. Therefore, $|F_{X_T}(\omega)|^2$ is more meaningful.

However, we are interested in freq.-domain representation of $X(t)$, not $X_T(t)$.

However, if we average over the range of the integral $(2T)$, then it doesn't blow up.

$S_X(\omega) =$

- called spectral density or power density spectrum of $X(t)$

$$S_X(\omega) = \lim_{T \to \infty} \frac{E\{|F_{X_T}(\omega)|^2\}}{2T}$$

- called spectral density or power density spectrum of $X(t)$

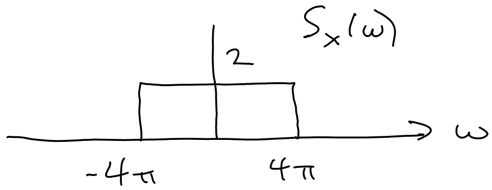
Properties:

1) $\overline{X^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega)\, d\omega$

2) $S_X(\omega)$ is real.

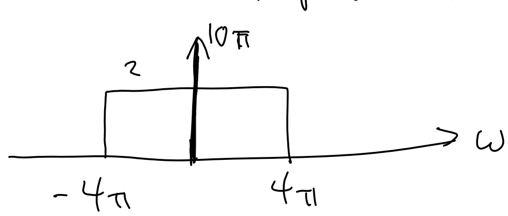
3) $S_X(\omega) \geq 0$ (nonnegative)

4) $S_X(\omega) = S_X(-\omega)$ (even)



$\overline{X^2} =$

5) $(\overline{X})^2 = \frac{1}{2\pi} S_0$, where S_0 is the scale factor on an impulse at $\omega = 0$.



Ex: Find mean + var. of $X(t)$.

6) $S_X(\omega) =$

$$R_X(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega) e^{j\omega\tau}\, d\omega$$

Ex: Let $R_X(\tau) = e^{-|\tau|}$

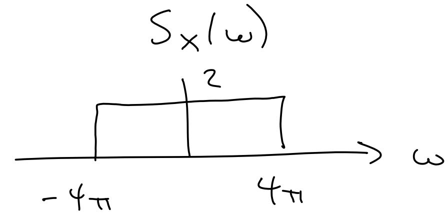
$$S_X(\omega) = \int_{-\infty}^{\infty} e^{-|\tau|} e^{-j\omega\tau}\, d\tau = \int_0^{\infty} e^{-\tau} e^{-j\omega\tau}\, d\tau + \int_{-\infty}^{0} e^{\tau} e^{-j\omega\tau}\, d\tau$$

$$= \int_0^{\infty} e^{-(1+j\omega)\tau}\, d\tau + \int_{-\infty}^{0} e^{(1-j\omega)\tau}\, d\tau$$

$$= \frac{-1}{1+j\omega} e^{-(1+j\omega)\tau}\Big|_0^{\infty} + \frac{1}{1-j\omega} e^{(1-j\omega)\tau}\Big|_{-\infty}^{0}$$

$=$

Ex:



$R_x(\tau) =$

$R_x(0) = \overline{X^2} =$

$(\overline{X})^2 = \qquad \Rightarrow \sigma^2 =$

Note: $R_x(0) = \overline{X^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)\,d\omega$

White noise

- white light is a combination of all wavelengths (or freq's) of light.

  White noise has

$S_x(\omega) = \qquad$ (white noise)

$\delta(\tau) \longleftrightarrow 1$

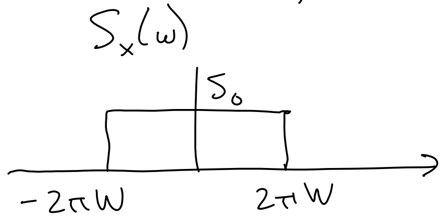
$R_x(\tau) = \mathcal{F}^{-1}\{S_x(\omega)\} =$

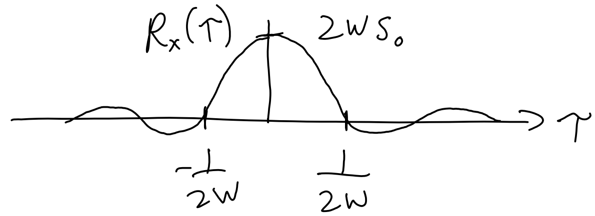
What is $\overline{X^2}$?

- bandlimited white noise

$$S_x(\omega) = \begin{cases} S_0 \ , & |\omega| \le 2\pi W \\ 0 \ , & |\omega| > 2\pi W \end{cases}$$

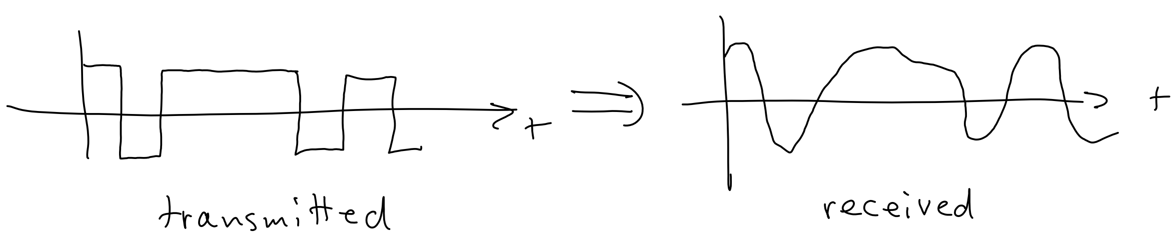
$$R_x(\tau) =$$





Linear systems with random input —

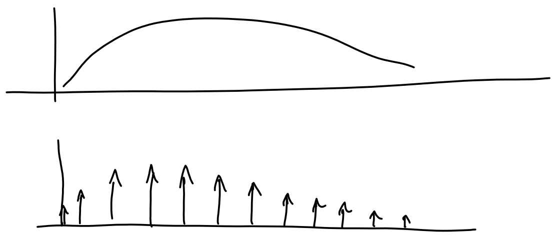
Examples:



Only statistical description of input is available

$\Rightarrow$ must settle for statistical description of output

• linear, time-invariant systems



$$x(t) \simeq \sum x(k)\,\delta(t - kT)$$

In the limit,

$x(t) =$: sifting property

$$x(t) \longrightarrow \boxed{T[\cdot]} \longrightarrow y(t)$$

$T[\cdot]$ is linear time-invariant

$$\delta(t) \longrightarrow \boxed{T[\cdot]} \longrightarrow h(t),$$

impulse response

$$y(t) = T[x(t)] = T\left[\int x(\lambda)\,\delta(t-\lambda)\,d\lambda\right]$$

$=$ by linearity

$=$ by time-invariance

$=$

If $h(t) = T[\delta(t)]$ and $T[\cdot]$ is time-invariant,
then $T[\delta(t-\lambda)] = h(t-\lambda)$

$$x(t) = e^{j\omega t}, \quad y(t) = x(t) * h(t)$$
$$= h(t) * x(t)$$
$$= \int h(\lambda)\, x(t-\lambda)\, d\lambda$$

$\mathcal{F}\{h(t)\} = H(\omega)$, the freq. response of the system

nonperiodic: $x(t) =$: inverse FT

periodic: $x(t) =$: Fourier series expansion

$$y(t) = \frac{1}{2\pi}\int Y(\omega)\, e^{j\omega t}\, d\omega$$
$$= T[x(t)]$$
$$= \frac{1}{2\pi}\int X(\omega)\, T[e^{j\omega t}]\, d\omega$$
$$= \frac{1}{2\pi}\int X(\omega)\, H(\omega)\, e^{j\omega t}\, d\omega$$

$$\Rightarrow \quad Y(\omega) = X(\omega)\, H(\omega)$$

Ex: $h(t) = \begin{cases} e^{-t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$

$$X(t) = A + B\cos(\omega_1 t + \theta), \quad \text{where}$$

$$f(\theta) = \begin{cases} \frac{1}{2\pi}, & 0 \leq \theta \leq 2\pi \\ 0, & \text{otherwise} \end{cases}$$

$$R_x(\tau) = A^2 + \frac{B^2}{2}\cos\omega_1 t$$

$$S_x(\omega) = 2\pi A^2 \delta(\omega) + \frac{\pi}{2} B^2 \left[\delta(\omega - \omega_1) + \delta(\omega + \omega_1)\right]$$

$$Y(t) = X(t) * h(t)$$

$$= \int_{-\infty}^{\infty} X(\lambda)\, h(t-\lambda)\, d\lambda$$

$t - \lambda \geq 0$

$t \geq \lambda$

$$= \int_{-\infty}^{t} \left[A + B\cos(\omega_1 \lambda + \theta)\right] e^{-(t-\lambda)}\, d\lambda$$

$$= e^{-t}\left[A e^{\lambda} + B\frac{e^{\lambda}}{1+\omega_1^2}\left(\cos(\omega_1\lambda + \theta) + \omega_1 \sin(\omega_1 \lambda + \theta)\right)\right]\Bigg|_{-\infty}^{t}$$

$$= e^{-t}\left[A e^{t} + B\frac{e^{t}}{1+\omega_1^2}\left(\cos(\omega_1 t + \theta) + \omega_1 \sin(\omega_1 t + \theta)\right)\right]$$

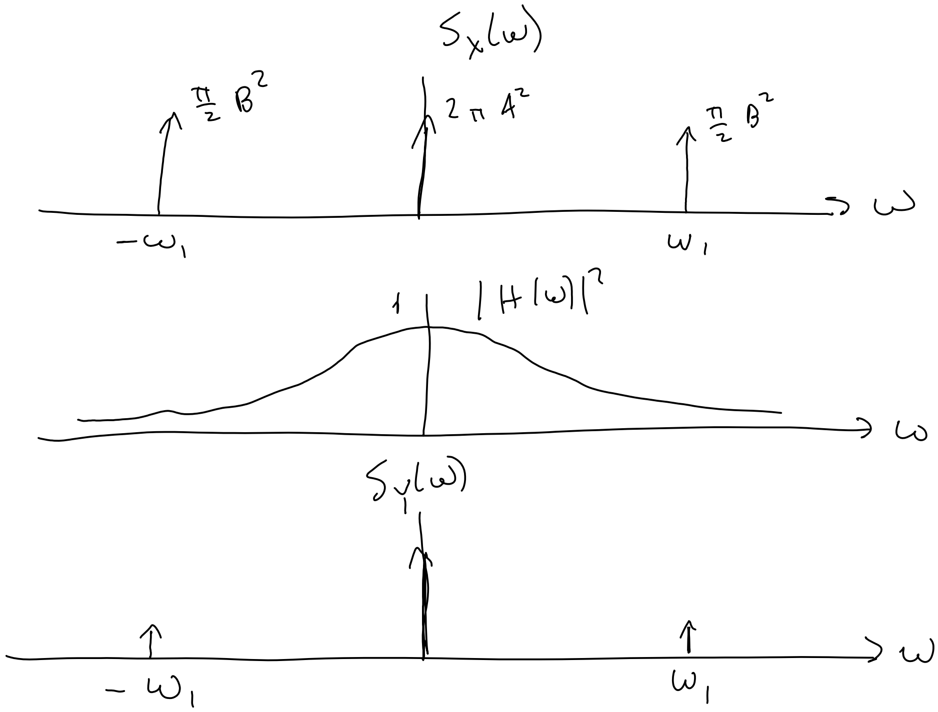
$$= A + \frac{B}{1+\omega_1^2}\left[\cos(\omega_1 t + \theta) + \omega_1 \sin(\omega_1 t + \theta)\right]$$

$$|H(\omega)|^2 = H(\omega)H^*(\omega) = \frac{1}{(1+j\omega)(1-j\omega)}$$

$$= \frac{1}{1+\omega^2}$$

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega)$$

$$Y(\omega) = H(\omega)X(\omega)$$



$$X(t) \longrightarrow \boxed{\;h(t)\;} \longrightarrow Y(t)$$

$$Y(t) = X(t) * h(t)$$

$$\bar{Y} = E\{X(t) * h(t)\}$$

$$= E\left\{\int_{-\infty}^{\infty} h(\lambda)\, X(t-\lambda)\, d\lambda\right\}$$

$$= \int_{-\infty}^{\infty} h(\lambda)\, E\{X(t-\lambda)\}\, d\lambda$$

$$R_Y(\tau) = E\{Y(t)\, Y(t+\tau)\}$$

$$= E\left[\int_{-\infty}^{\infty} h(\lambda_1)\, X(t-\lambda_1)\, d\lambda_1 \int_{-\infty}^{\infty} h(\lambda_2)\, X(t+\tau-\lambda_2)\, d\lambda_2\right]$$

$$= \iint E\{X(t-\lambda_1)\, X(t+\tau-\lambda_2)\}\, h(\lambda_1)\, h(\lambda_2)\, d\lambda_1\, d\lambda_2$$

$$= \iint R_X(\lambda_2-\lambda_1-\tau)\, h(\lambda_1)\, h(\lambda_2)\, d\lambda_1\, d\lambda_2$$

$$= \iint R_x(t-\tau)\, h(\lambda_1)\, h(t+\lambda_1)\, d\lambda_1\, dt$$

$$= \int_{-\infty}^{\infty} R_X(t-\tau)\left[\int_{-\infty}^{\infty} h(\lambda_1)\, h(\lambda_1+t)\, d\lambda_1\right] dt$$

$$=$$

$=$

$$\mathcal{R}_h(t) = \int_{-\infty}^{\infty} h(\lambda_1)\, h(\lambda_1 + t)\, d\lambda_1$$

$$\mathcal{F}\{\mathcal{R}_h(t)\} = |H(\omega)|^2$$

By conv. theorem,

$$S_Y(\omega) =$$

$$R_{XY}(\tau) = E[X(t)\,Y(t+\tau)]$$

$$= E\left[X(t) \int h(\lambda)\, X(t+\tau-\lambda)\, d\lambda\right]$$

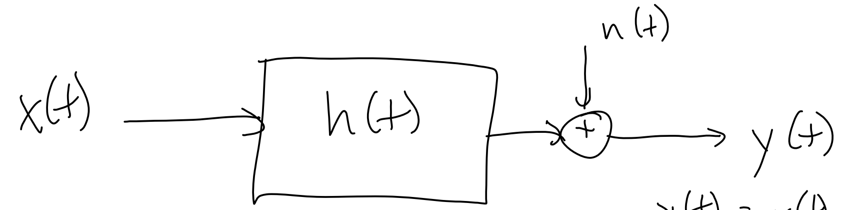
$$= \int h(\lambda)\, E\{X(t)\, X(t+\tau-\lambda)\}\, d\lambda$$

$=$

$=$

$$R_{YX}(\tau) = R_{XY}(-\tau) =$$

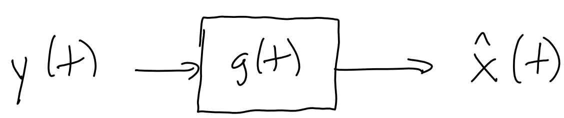
Wiener filter



$y(t) = x(t) * h(t) + n(t)$

How can we recover $x(t)$?

$\hat{X}(\omega) =$



- estimate $x(t)$ by minimizing expected squared error in $\hat{x}(t)$.

$$\Rightarrow R_Y(\tau) * g(\tau) = R_{YX}(\tau)$$

$$R_Y(\tau) = R_h(\tau) * R_X(\tau) + R_N(\tau)$$

assuming $x(t)$ & $n(t)$ are independent & $\bar{N} = 0$

$$\Rightarrow R_{XN}(\tau) =$$

$$R_{YX}(\tau) =$$

$$\left[R_h(\tau) * R_X(\tau) + R_N(\tau)\right] * g(t) = h(-\tau) * R_X(\tau)$$

$$G(\omega) =$$

Ex: Let $h(t) = \delta(t) + a\,\delta(t - t_0)$ (multipath)

what is Wiener filter if

$S_X(\omega) = e^{-\omega^2}$ and $S_N(\omega) = \sigma_n^2$?

$$H(\omega) =$$

$$|H(\omega)|^2 =$$

$$\hat{X}(\omega) =$$