Stochastic Signal Processing

Lesson 3:

Functions of one and two random variables

Weize Sun

Review Examples from last week

1: Assume that the waiting time of a bus follows an $\lambda = 0.1$ exponential distribution. (Time: minutes) What is the probability that you will be able to get on the bus in 10 minutes?

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$F(x) = 1 - e^{-\lambda x},\ x \geqslant 0$$

Solution:

$$\Pr(X \leqslant 10) = F(10) = 1 - e^{-1} = 0.6321$$

Review Examples from last week

2: Let B represent the event $\{ a < X(\xi) \leq b\}$ with $b > a$. For a given $F_X(x)$, determine $F_X(x|B)$and $f_X(x|B)$

Solution:

$$F_X(x|B) = P\{ X(\xi) \leq x|B \} = \frac{P\{ (X(\xi) \leq x) \cap (a < X(\xi) \leq b) \}}{P(a < X(\xi) \leq b)}$$

$$= \frac{P\{(X(\xi) \leq x) \cap (a < X(\xi) \leq b)\}}{F_X(b) - F_X(a)}.$$

For $x < a$, we have $\{ X(\xi) \leq x\} \cap \{ a < X(\xi) \leq b\} = \emptyset$, hence $F_X(x|B) = 0$.

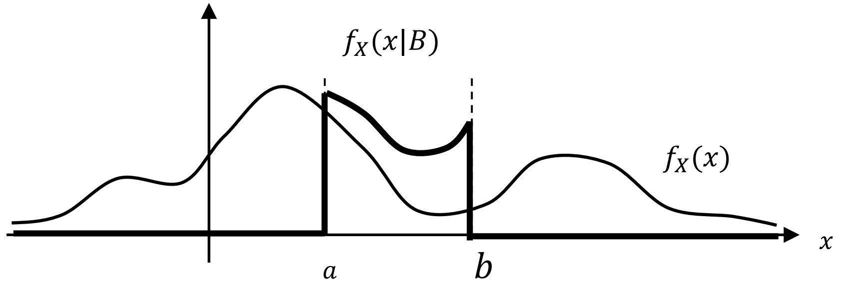
For $a \leq x < b$,we have $\{ X(\xi) \leq x\} \cap \{ a < X(\xi) \leq b\} = \{a < X(\xi) \leq x\}$, hence

$$F_X(x|B) = \frac{P(a < X(\xi) \leq x)}{F_X(b) - F_X(a)} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)}.$$

Review Examples from last week

2: Let B represent the event $\{\, a < X(\xi) \le b\}$ with$b > a$. For a given $F_X(x)$, determine $F_X(x|B)$ and $f_X(x|B)$.

For $x \ge b$, we have $\{\, X(\xi) \le x\} \cap \{\, a < X(\xi) \le b\} = \{\, a$



- The shape of the conditional pdf is the same as the shape of original pdf, for $a < x \le b$
- Otherwise, pdf becomes 0

The random variable $g(X)$

- Let X be a r.v and suppose $g(\boldsymbol{x})$ is a function of the variable $\boldsymbol{x}$. Define
$$Y = g(X)$$
- Clearly Y is a r.v (there are some conditions, see Page 87 of the text book, but here we will not discuss it), In particular
$$F_Y(y) = P(Y(\xi) \leq y) = P\big(g\big(X(\xi)\big) \leq y\big)$$
$$= P(X(\xi) \leq g^{-1}(-\infty, y])$$
- Thus the CDF and pdf of Y can be determined in terms of that of X.
- Usually, we start with the CDF $F_Y(y)$, then take the derivative of it to get the pdf.
- We will start with some examples

The random variable $g(X)$

- Quick Example 1: $Y = aX + b$, compute $f_Y(y)$.

The random variable $g(X)$

- Quick Example 1: $Y = aX + b$, compute $f_Y(y)$.

Suppose $a > 0$.

$$F_Y(y) = P(Y(\xi) \le y) = P(aX(\xi) + b \le y) = P(X(\xi) \le \frac{y-b}{a})$$
$$= F_X(\frac{y-b}{a})$$

and $f_Y(y) = \frac{1}{a} f_X(\frac{y-b}{a})$

On the other hand if $a < 0$, then

$$F_Y(y) = P(Y(\xi) \le y) = P(aX(\xi) + b \le y) = P\left(X(\xi) > \frac{y-b}{a}\right)$$
$$= 1 - F_X\left(\frac{y-b}{a}\right) \quad ,$$

and hence $f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right)$

Generally speaking, $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$ for all a

The random variable $g(X)$

- Quick Example 2: $Y = X^2$, compute $f_Y(y)$.

The random variable $g(X)$

- Quick Example 2: $Y = X^2$, compute $f_Y(y)$.

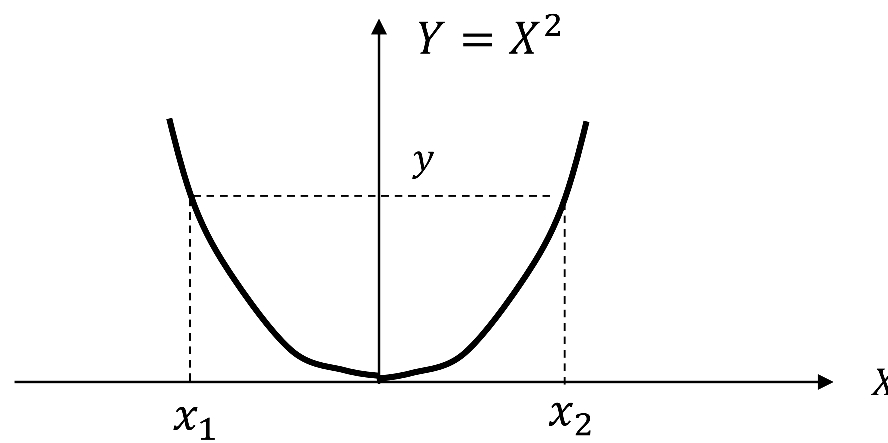
$$F_Y(y) = P(Y(\xi) \leq y) = P(X^2(\xi) \leq y)\,.$$

If $y < 0$, then the event $\{\, X^2(\xi) \leq y\} = \emptyset$, and hence

$$F_Y(y) = 0, \qquad y < 0.$$

For $y > 0$, from the below figure, the event $\{Y(\xi) \leq y\} = \{X^2(\xi) \leq y\}$ is equivalent to

$\{x_1 < X(\xi) \leq x_2\}$



The random variable $g(X)$

Hence

$$F_Y(y) = P(x_1 < X(\xi) \leq x_2) = F_X(x_2) - F_X(x_1)$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}), y > 0.$$

By direct differentiation, we get

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right), & y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Is there a general solution?

The random variable $g(X)$ – general solution – discrete type r.v

- Assume that there are n different function values x_i $(i = 1,2,\dots,n)$ map to the same value y, then

$$P(Y=y)=\sum_{i=1}^{n}P(X=x_n)=\sum_{i=1}^{n}P(X=g_n^{-1}(y))$$

- where $g_n^{-1}(y)$ represents different values x mapped to the same y value

- Example: Given $X = \{-2,-1,0,1,2\}$ with probabilities {0.1, 0.15, 0.25, 0.3, 0.2}, respectively, and the distribution of $Y = X^2$ is:

Y的取值为$\{0,1,4\}$ →

$$P(Y=0)=P(X=0)=0.25$$
$$P(Y=1)=P(X=-1)+P(X=1)=0.45,$$
$$P(Y=4)=P(X=2)+P(X=-2)=0.3$$

The random variable $g(X)$ – general solution – continuous type r.v

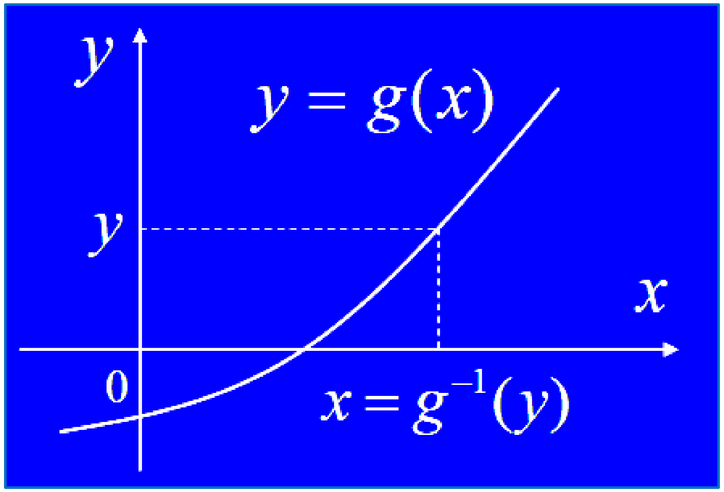
- Assume that $g(x)$ is a derivable monotone function (可导的单调函数) whose inverse function is $g^{-1}(y)$, and it is **monotonically increasing**

$\rightarrow \mathrm{F_Y(y) = P(Y \le y) = P\{g(X) \le y\} = P\{X \le g^{-1}(y)\} = F_X(g^{-1}(y))}$

- Take the derivative of both sides

$$f_Y(y) = f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} = f_X(x)\frac{dx}{dy}|_{x=g^{-1}(y)}$$

Note: almost always first CDF then pdf



The random variable $g(X)$ – general solution – continuous type r.v

- Assume that $g(x)$ is a derivable monotone function (可导的单调函数) whose inverse function is $g^{-1}(y)$, and it is **monotonically increasing**

$$\rightarrow F_Y(y) = P(Y \le y) = P\{g(X) \le y\} = P\{X \le g^{-1}(y)\} = F_X(g^{-1}(y))$$

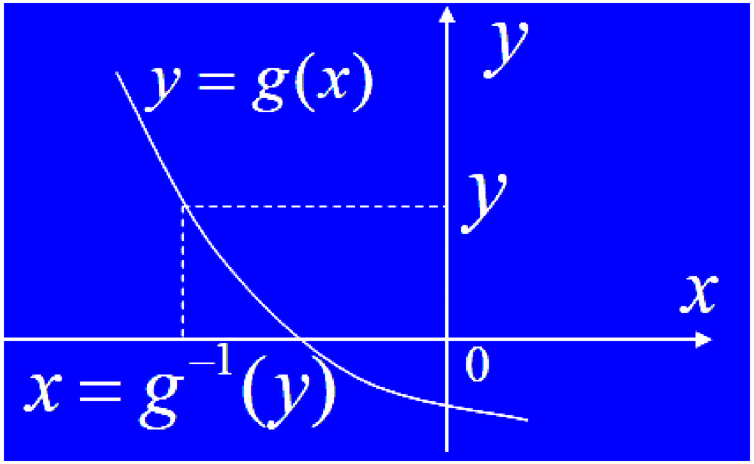
If $g(x)$ is **monotonically decreasing** →

Note the difference

$$F_Y(y) = P(Y \le y) = P\{g(X) \le y\} = \mathbf{P\{X \ge g^{-1}(y)\}}$$
$$= 1 - P\{X \le g^{-1}(y)\} = 1 - F_X(g^{-1}(y))$$

- Take the derivative of both sides

$$f_Y(y) = -f_X\left(g^{-1}(y)\right)\frac{dg^{-1}(y)}{dy}$$
$$= -f_X(x)\frac{dx}{dy}|_{x=g^{-1}(y)}$$
$$= f_X(x)|\frac{dx}{dy}|_{x=g^{-1}(y)}$$



The random variable $g(X)$ – general solution – continuous type r.v

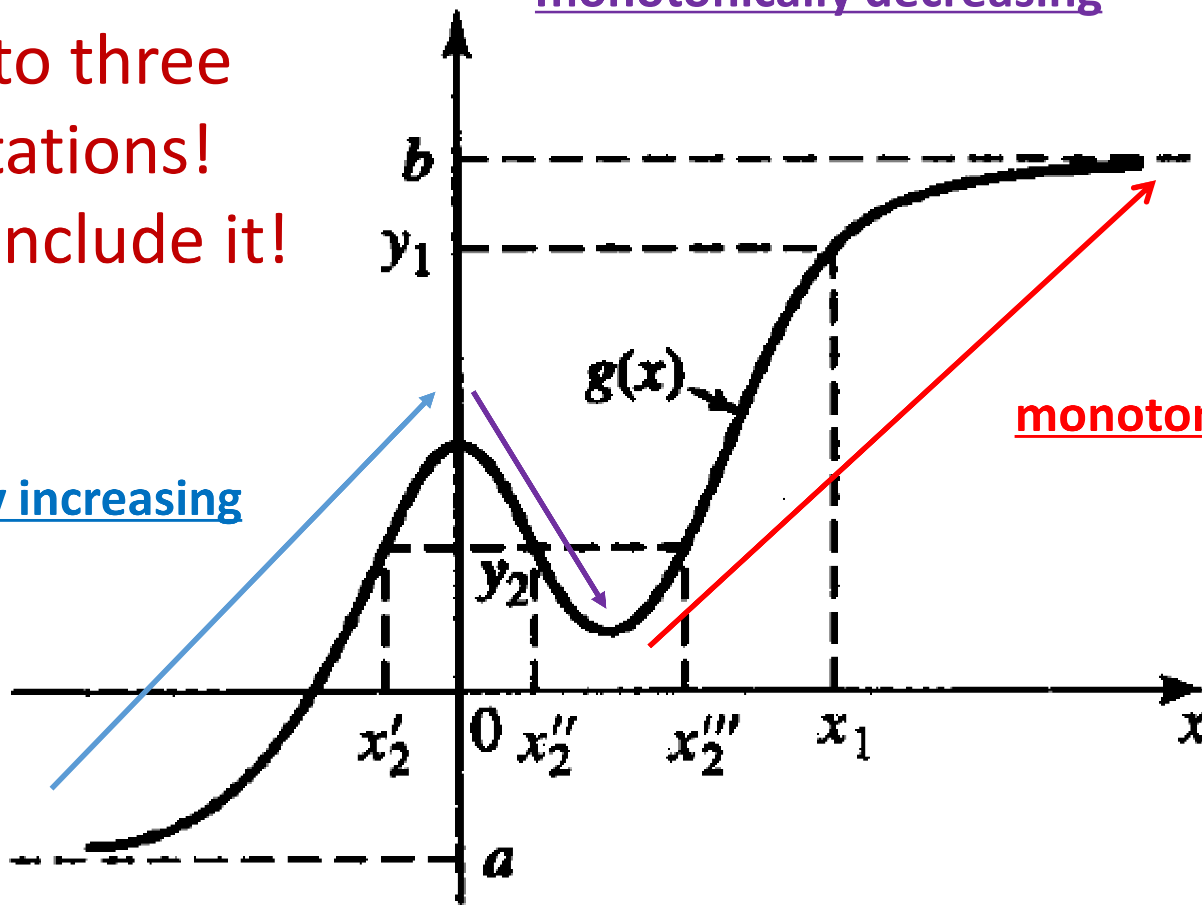
- How about a function like this?

Split it into three segmentations!
Now we conclude it!

monotonically decreasing

monotonically increasing

monotonically increas



*: see Figure 3-1 and section 3-2 in page 88 of text book (or the Chinese text book, page 96-97)

The random variable $g(X)$ – general solution – continuous type r.v

Define $J = dx/dy$ and we have:

- If $g(x)$ **monotonically increasing**→

$$f_Y(y) = f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} = f_X(x)\frac{dx}{dy}|_{x=g^{-1}(y)}$$

- If $g(x)$ **monotonically decreasing**→

$$f_Y(y) = -f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} = -f_X(x)\frac{dx}{dy}|_{x=g^{-1}(y)}$$

$$= \mathrm{f_X}(x)|\frac{\mathrm{d}x}{\mathrm{dy}}|_{\mathrm{x=g^{-1}(y)}}$$

- Which is, for both cases, generally speaking:

$$f_Y(y) = f_X(x)|J|_{X=g^{-1}(y)}$$

- In fact, the J is called the Jacobian, we will come back to this notation later when talking about two random variables

The random variable $g(X)$ – general solution – continuous type r.v

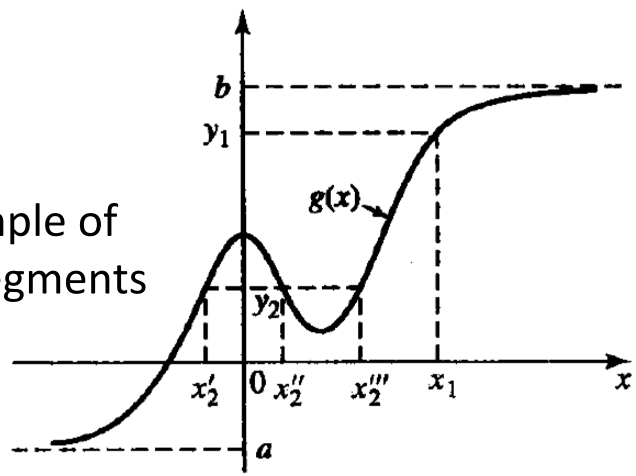
Define $J = dx/dy$ and we have:

$$f_Y(y) = f_X(x)|J|_{X=g^{-1}(y)}$$

- If $g(x)$ is not a monotonic function, then divided it into multiple monotonic segments: if there are n inverse functions for $y = g(x)$ as $x_i = h_i(y), i = 1,2, \ldots, n$, then

$$f_Y(y) = \sum_{i=1}^{n} f_X(x_i)|J_i|_{x_i=h_i(y)} \qquad J_i = \frac{dx_i}{dy}$$

This is an example of 3monotonic segments



The random variable $g(X)$

- Example 2.1: Assuming that the relationship between input and output is $Y = bX^2 (b > 0)$, find the pdf of Y. ($f_Y(y) = ???\, f_X(x) ???$)

Note: $f_Y(y) = \sum_{i=1}^{n} f_X(x_i)|J_i|_{x_i=h_i(y)}$ $\quad J_i = \frac{dx_i}{dy}$

Solution:

1. For $y < 0$, $f_Y(y) = 0$.
2. For $y > 0$, we have $x_1 = \sqrt{y/b}$, $\mathrm{x}_2 = -\sqrt{\mathrm{y/b}}$

Taking the derivative of x_1, x_2 for each respectively:

$$J_1 = \frac{dx_1}{dy} = \frac{1}{2\sqrt{by}}$$

$$J_2 = \frac{dx_2}{dy} = -\frac{1}{2\sqrt{by}}$$

$\rightarrow$

$$f_Y(y) = \frac{1}{2\sqrt{by}}[f_x(\sqrt{\frac{y}{b}}) + f_x(-\sqrt{\frac{y}{b}})]$$

When $b = 1$, it is example 2

The random variable $g(X)$

- Example 2.1: Assuming that the relationship between input and output is $Y = bX^2 (b > 0)$, find the pdf of Y.

It is called ‘power detector’ (there are might be some other names)

小信号平方率检波原理在民航通导设备中的应用

来自万方

收藏 引用 批量引用 报错 分享

作者 马有存 ， 康立学

摘要 摘 要：射频信号是民航空管通信导航监视各类设备板件中不可或缺的信号,同时测量射频信号特征也是工程师排除设备故障的必要方法,为使测量更加方便、排故更加准确,本文采用小信号平方率检波原理设计并制作了便携式调幅检波器.

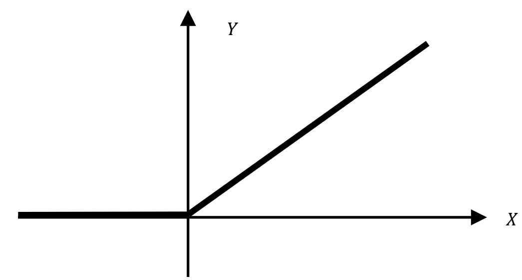
出版源 《军民两用技术与产品》，2017 (2) :38-39

被引量 0

The random variable $g(X)$

- Example 3: Half-wave rectifier(半波整流器), or the ReLU

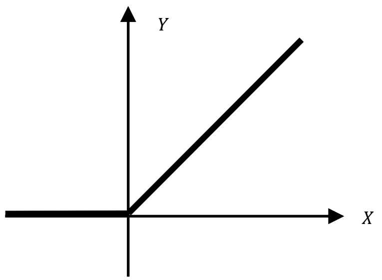
$$Y = g(X), \qquad g(x) = \begin{cases} x, & x > 0, \\ 0, & x \le 0. \end{cases}$$



The random variable $g(X)$

- Example 3: Half-wave rectifier(半波整流器), or the ReLU

$$Y = g(X), \qquad g(x) = \begin{cases} x, & x > 0, \\ 0, & x \le 0. \end{cases}$$



Solution: we have

$$P(Y = 0) = P(X(\xi) \le 0) = F_X(0).$$

and for $y > 0$ since $Y = X$,

$$F_Y(y) = P(Y(\xi) \le y) = P(X(\xi) \le y) = F_X(y)$$

Thus

$$f_Y(y) = \begin{cases} f_X(y), & y > 0, \\ F_X(0)\delta(y) & y = 0, \\ 0, & y < 0, \end{cases}$$

The random variable $g(X)$ – mean, variance and moments

- **Mean** or the **Expected Value** of a r.v *X* is defined as

$$\mu_X = \bar{X} = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)dx.$$

- If *X* is a discrete-type r.v:

$$\mu_X = \bar{X} = E(X) = \int x \sum_i p_i \delta(x - x_i)dx$$

$$= \sum_i x_i p_i \underbrace{\int \delta(x - x_i)dx}_{1} = \sum_i x_i p_i = \sum_i x_i\, P(X = x_i)\ .$$

- suppose $Y = g(X)$ defines a new r.v with pdf $f_Y(y)$, then Y has a mean:

$$\mu_Y = E(Y) = \int_{-\infty}^{+\infty} y\, f_Y(y)dy.$$

The random variable $g(X)$ – mean, variance and moments

- Mean of usually used distributions
  - Bernoulli distribution: $E(X) = p \cdot 1 + (1-p) \cdot 0 = p$
  - Binomial distribution: $E(X) = \sum_{x=0}^{\infty} x c_n^x p^x (1-p)^{n-x} = \mathrm{np}$
  - Poisson distribution: $E(X) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda$
  - Uniform distributed: $E(X) = \int_a^b x/(b-a)dx = (b+a)/2$
  - Exponential distribution: $E(\mathrm{X}) = \int_0^{\infty} x\lambda e^{-\lambda x} dx = 1/\lambda$
  - Normal distribution: $E(X) = \mu$

The random variable $g(X)$ – mean, variance and moments

- 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:

$$\sigma_X^2 = E[(X-\mu)^{\,2}] > 0.$$

This is the **variance** of r.v X. Its square root $\sigma_X = \sqrt{E}(X-\mu)^2$ is known as the standard deviation of the r.v X. Note that the standard deviation represents the root mean square spread of the r.v X around its mean.

Note that

$$D(X) = Var(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} (x^2 - 2x\mu + \mu^2) f_X(x)dx$$

$$= \int_{-\infty}^{+\infty} x^2 f_X(x)dx - 2\mu \int_{-\infty}^{+\infty} x\, f_X(x)dx + \mu^2$$

$$= E(X^2) - \mu^2 = E(X^2) - [E(X)]^{\,2} = \overline{X^2} - \bar{X}^2.$$

The random variable $g(X)$ – mean, variance and moments

- Variance of usually used distributions
  - Bernoulli distribution: $D(X) = p - p^2$
  - Binomial distribution: $D(\mathrm{X}) = np(1-p)$
  - Poisson distribution: $D(X) = \lambda = E(X)$

The mean and variance of a Poisson distribution are numerically equal

  - Uniform distributed: $D(X) = (b-a)^2/12$
  - Exponential distribution: $D(\mathrm{X}) = 1/\lambda^2$
  - Normal distribution: $D(X) = \sigma^2$

The random variable $g(X)$ – mean, variance and moments

- Properties of Variance
  - Linear transformation:
    $$D(aX + b) = a^2 D(X)$$
  - If X and Y independent:
    $$D(X \pm Y) = D(X) + D(Y)$$
  - For any two random variables X, Y(not required to be independent):
    $D(X \pm Y) = D(X) + D(Y) \pm 2cov(X, Y)$

The random variable $g(X)$ – mean, variance and moments

- **Moments**:

$$m_n = \overline{X^n} = E(X^n), \qquad n \geq 1$$

are known as the moments of the r.v *X*, and

$$\mu_n = E[(X-\mu)^n]$$

are known as the central moments of *X*.

- Clearly
  - the mean $\mu = m_1$,
  - and the variance $\sigma^2 = \mu_2$

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). Then

$$P(x_1 < X(\xi) \le x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)dx,$$

and

$$P(y_1 < Y(\xi) \le y_2) = F_Y(y_2) - F_Y(y_1) = \int_{y_1}^{y_2} f_Y(y)dy\,.$$

Two Random Variables

- What about the probability that the pair of r.vs (*X*,*Y*) belongs to an arbitrary region *D*? In other words, how to estimate, for example, $P[(x_1 < X(\xi) \leq x_2) \cap (y_1 <$

Two Random Variables

- **Properties**

(i) $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0, \quad F_{XY}(+\infty, +\infty) = 1.$

since $(X(\xi) \leq -\infty, Y(\xi) \leq y) \subset (X(\xi) \leq -\infty)$, we get $F_{XY}(-\infty, y) \leq P(X(\xi) \leq -\infty) = 0.$

Similarly $F_{XY}(\infty, \infty) = P(S) = 1$

(ii) $P(x_1 < X(\xi) \leq x_2, Y(\xi) \leq y) = F_{XY}(x_2, y) - F_{XY}(x_1, y)$
$P(X(\xi) \leq x, y_1 < Y(\xi) \leq y_2) = F_{XY}(x, y_2) - F_{XY}(x, y_1)$

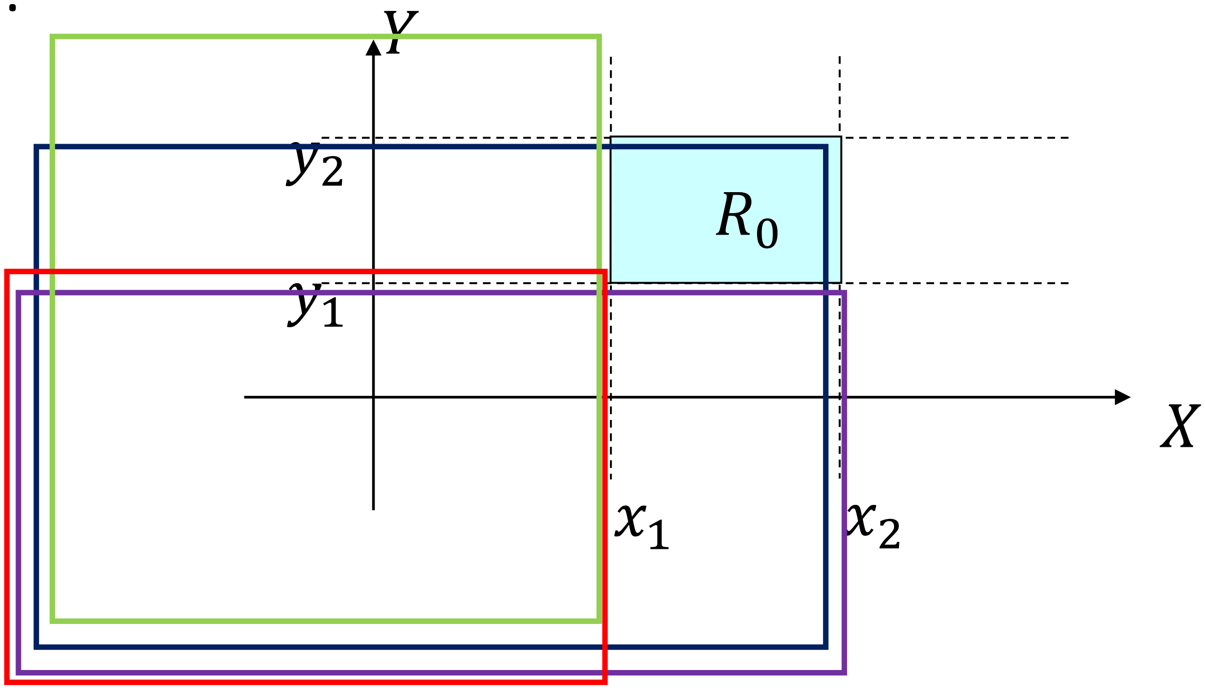
Two Random Variables

- **Properties**

(iii) $P(x_1 < X(\xi) \le x_2, y_1 < Y(\xi) \le y_2)$

$$\boxed{= F_{XY}(x_2, y_2)} - \boxed{F_{XY}(x_2, y_1)}$$
$$\boxed{-F_{XY}(x_1, y_2)} + \boxed{F_{XY}(x_1, y_1).}$$

This is the probability that (X,Y) belongs to the rectangle R_0 in the Fig.:



Two Random Variables - Joint pdf

- By definition, the joint pdf of *X* and *Y* is given by

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x\, \partial y}.$$

- and hence we obtain the useful forula

$$F_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)\ dudv.$$

we also get

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x,y)\ dxdy = 1.$$

and

$$P((X,Y) \in D) = \int \int_{(x,y)\in D} f_{XY}(x,y) dxdy.$$

Two Random Variables - 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 CDF of X, and $f_X(x)$ is the marginal pdf of X. It is interesting to note that all marginals can be obtained from the joint pdf:

$$F_X(x) = F_{XY}(x, +\infty), \qquad F_Y(y) = F_{XY}(+\infty, y).$$

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

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

Two Random Variables - Marginal Statistics

- If X and Y are discrete r.vs, then $p_{ij} = P(X = x_i, Y = y_j)$ represents their joint pdf, and their respective marginal pdfs are given by

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}$$

And

$$P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{ij}$$

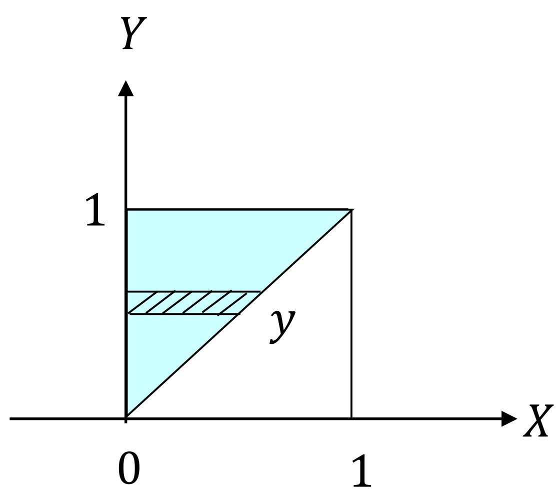
- Assuming that $P(X = x_i, Y = y_j)$ is written out in the form of a rectangular array, to obtain $P(X = x_i)$, one need to add up all entries in the i-th row.

$$\begin{array}{c|cccccc}
 & & & & \sum_i p_{ij} & & \\
\hline
 & p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1n} \\
 & p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2n} \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_j p_{ij} & p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{in} \\
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & p_{m1} & p_{m2} & \cdots & p_{mj} & \cdots & p_{mn}
\end{array}$$

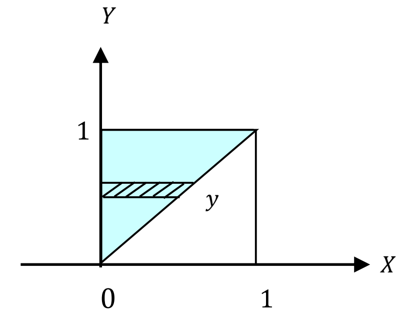
Two Random Variables - Marginal Statistics

- Example 4: Calculate $f_X(x)$ and $f_Y(y)$ Given

$$f_{XY}(x, y) = \begin{cases} \mathrm{k}, & 0 < x < y < 1, \\ 0, & \text{otherwise} \end{cases} .$$



Two Random Variables - Marginal Statistics



- Example 4: Calculate $f_X(x)$ and $f_Y(y)$ Given

$$f_{XY}(x,y) = \begin{cases} \mathrm{k}, & 0 < x < y < 1, \\ 0, & \text{otherwise} \ . \end{cases}$$

Solution: It is given that the joint pdf $f_{XY}(x,y)$ is a constant in the shaded region in the figure. Note that the integral of joint pdf is 1:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = \int_{y=0}^{1}\left(\int_{x=0}^{y} k\cdot dx\right)dy =$$

$$\int_{y=0}^{1} kydy = \left.\frac{ky^2}{2}\right|_0^1 = \frac{k}{2} = 1$$

- Thus k = 2. Then

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \int_{y=x}^{1} 2dy = 2(1-x), \qquad 0 < x < 1\,,$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx = \int_{x=0}^{y} 2dx = 2y, \qquad 0 < y < 1$$

注：最基本的二次项和指数项的求导、求积分要复习一下，以及最基本的分部积分法也是，默认大家都会；此外的，不作要求

Two Random Variables – 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 sets A and B in x and y axes respectively.

- If the r.vs X and Y are independent, then

$$P((X(\xi) \leq x) \cap (Y(\xi) \leq y)) = P(X(\xi) \leq x)P(Y(\xi) \leq y)$$

Which is $F_{XY}(x, y) = F_X(x)F_Y(y)$

or equivalently $f_{XY}(x, y) = f_X(x)f_Y(y)$.

- If X and Y are discrete-type r.vs then their independence implies

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \text{for all} \quad i, j$$

Two Random Variables – Independence of r.vs

- Example 5: Given

$$f_{XY}(x,y) = \begin{cases} xy^2 e^{-y}, & 0 < y < \infty,\ 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Determine whether X and Y are independent.

注：最基本的二次项和指数项的求导、求积分要复习一下，以及最基本的分部积分法也是，默认大家都会；此外的，不作要求

Two Random Variables – Independence of r.vs

- Example 5: Given

$$f_{XY}(x,y) = \begin{cases} xy^2e^{-y}, & 0 < y < \infty,\ 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Determine whether *X* and *Y* are independent.

Solution: $f_X(x) = \int_0^{+\infty} f_{XY}(x,y)dy = x\int_0^{\infty} y^2e^{-y}dy$

$$= x\left(-2ye^{-y}\Big|_0^{\infty} + 2\int_0^{\infty} ye^{-y}dy\right) = 2x, 0 < x < 1$$

Similarly

$$f_Y(y) = \int_0^1 f_{XY}(x,y)dx = \frac{y^2}{2}e^{-y}, \qquad 0 < y < \infty$$

In this case: $f_{XY}(x,y) = f_X(x)f_Y(y)$,

and hence *X* and *Y* are independent random variables.

Two Random Variables – functions of two r.vs

- Assume that there are two r.vs (X_1, X_2) and (Y_1, Y_2) that

$$Y_1 = g_1(X_1, X_2)$$
$$Y_2 = g_2(X_1, X_2)$$

- Then we have $f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2}(x_1, x_2)|J|$

$$J = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Jacobian determinant
(雅可比行列式)

Two Random Variables – functions of two r.vs

- Quick Example 6: Given two r.vs (X_1, X_2), calculate the pdf of their summation Y_1 under $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$

Note that we have: $x_1 = (y_1 + y_2)/2$ $x_2 = (y_1 - y_2)/2$

Two Random Variables – functions of two r.vs

- Quick Example 6: Given two r.vs (X_1, X_2), calculate the pdf of their summation Y_1 under $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$

Note that we have: $x_1 = (y_1 + y_2)/2$ $x_2 = (y_1 - y_2)/2$

Solution: Cal. The J

$$J = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = -\frac{1}{2}$$

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}(x_1, x_2)|J|$$

$$= \frac{1}{2} f_{X_1X_2}((y_1 + y_2)/2, (y_1 - y_2)/2)$$

$$f_{Y_1}(y_1) = \int_{-\infty}^{+\infty} f_{Y_1Y_2}(y_1, y_2)dy_2$$

$$= \frac{1}{2}\int_{-\infty}^{+\infty} f_{X_1X_2}(\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2))dy_2$$

Two Random Variables – functions of two r.vs

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}(x_1, x_2)|J|$$
$$= \frac{1}{2} f_{X_1X_2}((y_1 + y_2)/2, (y_1 - y_2)/2)$$
$$f_{Y_1}(y_1) = \int_{-\infty}^{+\infty} f_{Y_1Y_2}(y_1, y_2) dy_2$$
$$= \frac{1}{2} \int_{-\infty}^{+\infty} f_{X_1X_2}(\frac{1}{2}(y_1 + y_2), \frac{1}{2}(y_1 - y_2)) dy_2$$

Let $u = (y_1 + y_2)/2$ $\Longrightarrow$ $f_{Y_1}(y_1) = \int_{-\infty}^{+\infty} f_{X_1X_2}(u, y_1 - u) du$

When (X_1, X_2) independent:

$$f_{Y_1}(y_1) = \int_{-\infty}^{+\infty} f_{X_1}(u) f_{X_2}(y_1 - u) du = f_{X_1}(y) \otimes f_{X_2}(y_1)$$

which is, the pdf of two independent r.vs' summation is the convolution of the pdfs of the two r.vs

Two Random Variables – Joint Moments

- Given two r.vs *X* and *Y* and a function $g(x, y)$, define the r.v

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

- we can define the mean of *Z* to be

$$\mu_Z = E(Z) = \int_{-\infty}^{+\infty} z\, f_Z(z) dz\,.$$

- It is possible to express the mean of $Z = g(X, Y)$in terms of $f_{XY}(x, y)$ *without* computing $f_Z(z)$.

$$E(Z) = E[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{XY}(x, y) dx dy\,.$$

If *X* and *Y* are discrete-type r.vs, then

$$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j).$$

Two Random Variables – Joint Moments

- 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, we get:

$$E[g(X)h(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y)f_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{+\infty} g(x)f_X(x)dx \int_{-\infty}^{+\infty} h(y)f_Y(y)dy = E[g(X)]E[h(Y)] \quad \text{(3-1)}$$

- However it is in general not true (X and Y not always independent).
- In the case of one random variable, 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 to **Covariance**.

Two Random Variables – Joint Moments

- **Covariance**: Given any two r.vs X and Y, define

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]\,.$$

Expanding and simplifying the right hand side of the above equation we get

$$Cov(X,Y) = E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y) = \overline{XY} - \overline{X}\,\overline{Y}. \quad \text{(3-2)}$$

Properties:

- $\mathrm{cov}(X,X) = \mathrm{E(XX) - E(X)E(X)} = D(X)$
- $\mathrm{cov}(aX,bY) = ab\mathrm{cov(X,Y)}$
- $|Cov(X,Y)| \le \sqrt{Var(X)Var(Y)}\,.$
- $co\,v(X,Y\pm Z) = E(XY\pm XZ) - E(X)E(Y\pm Z)$

$$= \{E(XY) - E(X)E(Y)\} \pm \{E(XZ) - E(X)E(Z)\}$$

$$= cov(X,Y) \pm cov(X,Z))$$

Two Random Variables – Joint Moments

- The range of **Covariance** is $(-\infty, +\infty)$, thus is hard to tell whether the value is large or small. Is a **Covariance** 10 highly correlated? Is a **Covariance** 0.1 weakly correlated? No one knows
- Therefore, we normalize it:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, -1 \le \rho_{XY} \le 1$$

it represents the correlation coefficient between *X* and *Y*.

And we have: $Cov(X,Y) = \rho_{XY}\sigma_X\sigma_Y$

Two Random Variables – Joint Moments

- **Uncorrelated r.vs**: If $\rho_{XY} = 0$, then X and Y are said to be uncorrelated r.vs. note that in this case, $Cov(X,Y) = \rho_{XY}\sigma_X\sigma_Y = 0$

Therefore, according to (3-2), if X and Y are uncorrelated, then

$$E(XY) = E(X)E(Y).$$

- **Orthogonality**: X and Y are said to be orthogonal if

$$E(XY) = 0.$$

Note that, if either X or Y has zero mean, then orthogonality implies uncorrelatedness also and vice-versa.

Expanding and simplifying the right hand side of the above equation we get

$$Cov(X,Y) = E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y) = \overline{XY} - \overline{X}\,\overline{Y}. \quad (3\text{-}2)$$

Two Random Variables – Joint Moments

- Suppose X and Y are independent r.vs. Then from (3-1) with $g(X)=X, \quad h(Y)=Y$,we get

$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ which is $E(XY) = E(X)E(Y)$

we conclude that if two r.vs **independence**, they are **uncorrelated**.

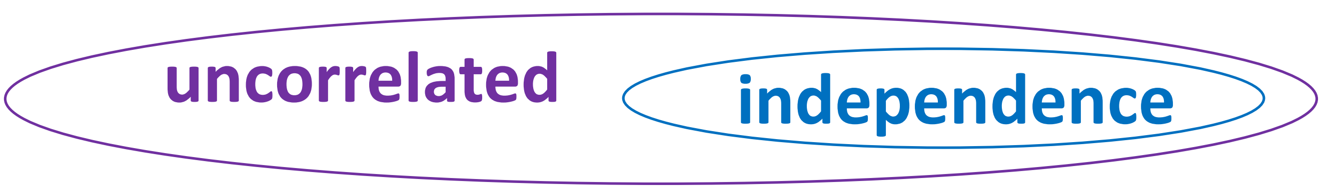
Note that, **uncorrelated** does not means **independence**.

- **Uncorrelated r.vs**: If $\rho_{XY}=0$, then X and Y are said to be uncorrelated r.vs. note that in this case, $Cov(X,Y)=\rho_{XY}\sigma_X\sigma_Y=0$

Therefore, according to (3-2), if X and Y are uncorrelated, then

$$E(XY)=E(X)E(Y).$$

$$E[g(X)h(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y)f_X(x)f_Y(y)dxdy$$
$$= \int_{-\infty}^{+\infty} g(x)f_X(x)dx \int_{-\infty}^{+\infty} h(y)f_Y(y)dy = E[g(X)]E[h(Y)] \qquad (3\text{-}1$$



Two Random Variables – Joint Moments

- Example 7: Calculate the correlation coefficient between Y and X
  1. $Y = aX + b$
  2. $Y = -aX + b$
  3. $Y = X^2, \; X \in [-1,1]$ uniform distributed

  For $a > 0$

$$\boxed{cov(X,Y) = \mathrm{E(XY) - E(X)E(Y)}}$$

Two Random Variables – Joint Moments

- Example 7: Calculate the correlation coefficient between Y and X
  1. $Y = aX + b$
  2. $Y = -aX + b$
  3. $Y = X^2, \ X \in [-1,1]$ uniform distributed

  For $a > 0$

$\boxed{cov(X,Y) = \mathrm{E(XY) - E(X)E(Y)}}$

Solutions:

$$D(Y) = D(aX + b) = a^2 D(X)$$

1. $$\mathrm{cov}(X,Y) = E(X(aX + b)) - E(X)E(aX + b) = aE(X^2) - a(E(X))^2 = aD(X)$$
$$\rho(X,Y) = \frac{aD(x)}{\sqrt{D(Y)D(X)}} = 1$$

2. Similarly as 1, get the result -1

3. $$\mathrm{cov}(X,Y) = E(X^3) - E(X)E(X^2) = E(X^3) = \int_{-1}^{1} x^3 \frac{1}{2} dx = 0$$

The (linear) correlation coefficient is 0, i.e. no linear relationship between X and Y, they are uncorrelated, but clearly, X and Y are not independent to each other

Multidimensional Random Variables – the covariance matrix

- For multidimensional random variables $X_1, X_2, \dots, X_n$, the covariance matrix is usually used to describe the relationship between random variables：

$$K = \begin{bmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & \ddots & \vdots \\ k_{n1} & \cdots & k_{nn} \end{bmatrix}$$

Where $k_{ij} = \operatorname{cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\}$

See the review example 2 for pratice

More on two dimensional Gaussian r.vs

- Define two r.vs X_1 and X_2, if their joint pdf is：

$$f_{X_1X_2}(x_1,x_2)$$

$$=\frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x_1-m_{X_1})^2}{\sigma_{X_1}^2}-\frac{2\rho(x_1-m_{X_1})(x_2-m_{X_2})}{\sigma_{X_1}\sigma_{X_2}}+\frac{(x_2-m_{X_2})^2}{\sigma_{X2}^2}\}$$

相关系数

They follows the joint Gaussian/Normal distribution, and their marginal pdf also follow Gaussian/Normal distribution

$$f_{X_1}(x_1)=\frac{1}{\sqrt{2\pi}\sigma_{X_1}}\exp[-\frac{(x_1-m_{X_1})^2}{2\sigma_{X_1}^2}];\ f_{X_2}(x_2)=\frac{1}{\sqrt{2\pi}\sigma_{X_2}}\exp[-\frac{(x_2-m_{X_2})^2}{2\sigma_{X_2}^2}]$$

If $\rho=0$(uncorrelated), then $f_{X_1X_2}(x_1,x_2)=f_{X_1}(x_1)f_{X_2}(x_2)$, which is independent.

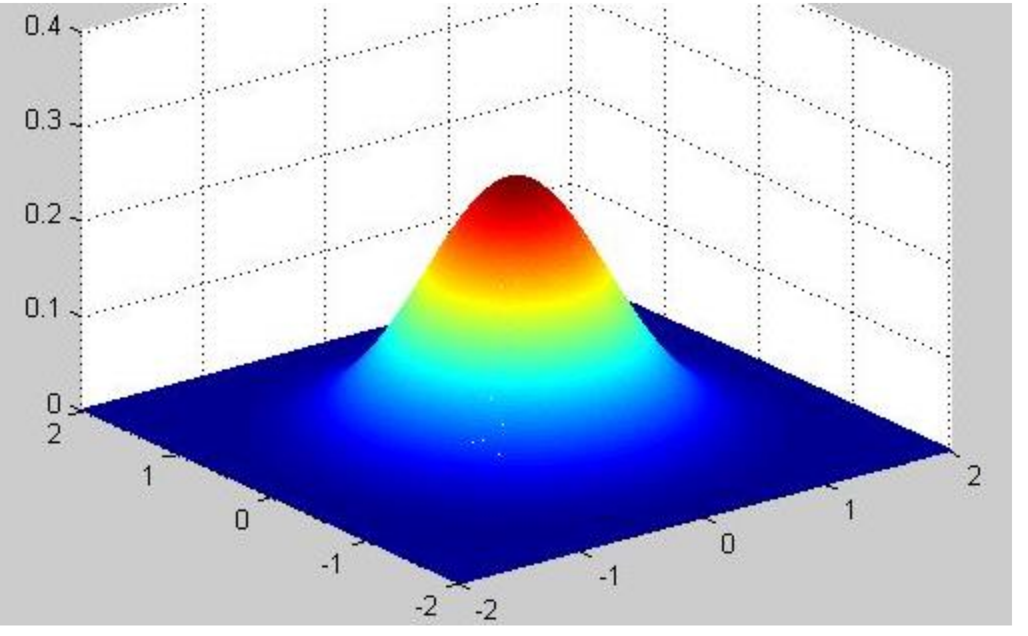
For two dimensional Gaussian r.vs, uncorrelated = independent

More on two dimensional Gaussian r.vs

- Define two r.vs X_1 and X_2, if their joint pdf is：

$$f_{X_1X_2}(x_1,x_2)$$
$$=\frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x_1-m_{X_1})^2}{\sigma_{X_1}^2}-\frac{2\rho(x_1-m_{X_1})(x_2-m_{X_2})}{\sigma_{X_1}\sigma_{X_2}}+\frac{(x_2-m_{X_2})^2}{\sigma_{X2}^2}\}$$

When $\rho = 0$



More on two dimensional Gaussian r.vs

- Define two r.vs X_1 and X_2, if their joint pdf is：

$$f_{X_1X_2}(x_1,x_2)=\frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[\frac{\left(x_1-m_{X_1}\right)^2}{\sigma_{X_1}^2}-\frac{2\rho(x_1-m_{X_1})(x_2-m_{X_2})}{\sigma_{X_1}\sigma_{X_2}}+\frac{(x_2-m_{X_2})^2}{\sigma_{X2}^2}$$

Write the joint pdf in matrix form:

$$f_{\mathrm{X}}(\boldsymbol{x})=\frac{1}{2\pi|\boldsymbol{K}|^{1/2}}\exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{K}^{-1}(\boldsymbol{x}-\boldsymbol{m})]$$

Where $\boldsymbol{x}=[\mathrm{x}_1,\mathrm{x}_2]^{\mathrm{T}}$ $\boldsymbol{m}=[m_{x_1},m_{x_2}]^T$

$$\boldsymbol{K}=\begin{bmatrix}\mathrm{E}[(\mathrm{X}_1-\mathrm{m}_{\mathrm{x}_1})^2] & \mathrm{E}[(\mathrm{X}_1-\mathrm{m}_{\mathrm{x}_1})(\mathrm{X}_2-\mathrm{m}_{\mathrm{x}_2})]\\ \mathrm{E}[(\mathrm{X}_2-\mathrm{m}_{\mathrm{x}_2})(\mathrm{X}_1-\mathrm{m}_{\mathrm{x}_1})] & \mathrm{E}[(\mathrm{X}_2-\mathrm{m}_{\mathrm{x}_2})^2]\end{bmatrix}$$

$$=\begin{bmatrix}\sigma_{\mathrm{X}_1}^2 & r\sigma_{\mathrm{X}_1}\sigma_{\mathrm{X}_2}\\ r'\sigma_{\mathrm{X}_1}\sigma_{\mathrm{X}_2} & \sigma_{\mathrm{X}_2}^2\end{bmatrix}$$

More on two dimensional Gaussian r.vs

- Define two r.vs X_1 and X_2, their joint pdf is：

$$f_{\mathrm{X}}(\boldsymbol{x}) = \frac{1}{2\pi|\boldsymbol{K}|^{1/2}} \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{K}^{-1}(\boldsymbol{x}-\boldsymbol{m})]$$

- Similarly, for multi-dimensional case (*):

Inverse matrix

$$f_X(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{K}|^{\frac{1}{2}}} \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T\mathbf{K}^{-1}(\boldsymbol{x}-\boldsymbol{m})]$$

Determinant

The covariance matrix

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}, \boldsymbol{K} = \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1n} \\ K_{21} & K_{22} & & K_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \cdots & K_{nn} \end{bmatrix}$$

Conditional distributions

- As we have seen in the previous lecture 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 7, lesson 2).
- 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.
- Recall that the distribution function of *X* given an event *B* is

$$F_X(x|B) = P(X(\xi) \le x|B) = \frac{P((X(\xi) \le x) \cap B)}{P(B)}$$

- Here we use a r.v $Y = y$ to represent the event B and get

$$F_{X|Y}(x|Y = y) = \frac{\int_{-\infty}^{x} f_{XY}(u, y)du}{f_Y(y)}.$$

- The pdf is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}.$$

For discrete-type r.vs, the pmf is

$$P\left(X = x_i|Y = y_j\right) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

Conditional distributions

- Conditional CDF and pdf

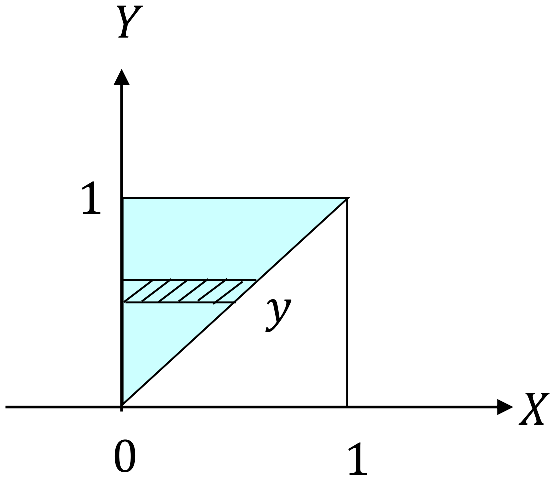
$$F_{X|Y}(x|Y=y) = \frac{\int_{-\infty}^{x} f_{XY}(u,y)du}{f_Y(y)}; \quad f_{X|Y}(x|Y=y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

- Similarly $$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \qquad (3\text{-}3)$$
- If the r.vs X and Y are independent, then $f_{XY}(x,y) = f_X(x) f_Y(y)$ and it becomes

$$f_{X|Y}(x|y) = f_X(x), \qquad f_{Y|X}(y|x) = f_Y(y),$$

implying that the conditional pdfs coincide with their unconditional pdfs. 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.

Conditional distributions



- Example 4.2 (see example 4 for review): Given

$$f_{XY}(x,y) = \begin{cases} \mathrm{k}, & 0 < x < y < 1, \\ 0, & \text{otherwise} \end{cases} .$$

determine $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$.

Solution: Note that the integral of joint pdf is 1:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = \int_{y=0}^{1}\left(\int_{x=0}^{y} k\cdot dx\right)dy = \int_{y=0}^{1} kydy = \frac{ky^2}{2}\bigg|_0^1 = \frac{k}{2}$$
$$= 1.$$

Then

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \int_{y=x}^{1} 2dy = 2(1-x), \qquad 0 < x < 1,$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx = \int_{x=0}^{y} 2dx = 2y, \qquad 0 < y < 1$$

Now calculate $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ yourself

Conditional distributions

Therefore $f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{y}, \qquad 0 < x < y < 1$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{1-x}, \qquad 0 < x < y < 1$$

Conditional distributions

- Conditional CDF and pdf

$$F_{X|Y}(x|Y=y)=\frac{\int_{-\infty}^{x} f_{XY}(u,y)du}{f_Y(y)}; \quad f_{X|Y}(x|Y=y)=\frac{f_{XY}(x,y)}{f_Y(y)}$$

- Similarly $$f_{Y|X}(y|x)=\frac{f_{XY}(x,y)}{f_X(x)} \qquad (3\text{-}3)$$

- We can use (3-3) to derive an important result as:

$$f_{XY}(x,y)=f_{X|Y}(x|y)f_Y(y)=f_{Y|X}(y|x)f_X(x)$$

Or
$$f_{Y|X}(y|x)=\frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}.$$

But

$$f_X(x)=\int_{-\infty}^{+\infty} f_{XY}(x,y)dy=\int_{-\infty}^{+\infty} f_{X|Y}(x|y)f_Y(y)dy$$

Therefore we get

$$f_{Y|X}(y|x)=\frac{f_{X|Y}(x|y)f_Y(y)}{\int_{-\infty}^{+\infty} f_{X|Y}(x|y)f_Y(y)dy} \qquad (3-4)$$

This equation represents the pdf version of Bayes' theorem. Here we use an example to illustrate the significance of it.

Conditional distributions

- Example 8: 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 $f(\theta|r)$.

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{\int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y) dy} \qquad (3\text{–}4)$$

Conditional distributions

- Example 8: 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 $f(\theta|r)$.

Solution:

Initially almost nothing about the r.v θ is known, so that we assume its a-priori pdf 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

$$f(r|\boldsymbol{\theta} = \theta) \sim N(\theta, \sigma^2)$$

since it is given that $\boldsymbol{\theta} = \theta$ is a constant, $r = \theta + n$ behaves like *n*. Using (3-4) can gives the a-posteriori pdf of θ given *r*

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{\int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y) dy} \qquad (3\text{–}4)$$

Conditional distributions

- Using (3-4) can gives the a-posteriori pdf of θ given *r,* note that $f_\theta(\theta) = 1/2\pi$ and $f(r|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-\theta)^2}{2\sigma^2}}$:

$$f(\theta|r) = \frac{f(r|\theta)f_\theta(\theta)}{\int_0^{2\pi} f(r|\theta)f_\theta(\theta)d\theta}$$

$$= \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-(r-\theta)^2/2\sigma^2} \frac{1}{2\pi}}{\int_0^{2\pi} \frac{1}{\sqrt{2\pi}\sigma} e^{-(r-\theta)^2/2\sigma^2} \frac{1}{2\pi} d\theta}$$

$$= \frac{e^{-(r-\theta)^2/2\sigma^2}}{\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta} = \phi(r) e^{-(\theta-r)^2/2\sigma^2}, 0 < \theta < 2\pi,$$

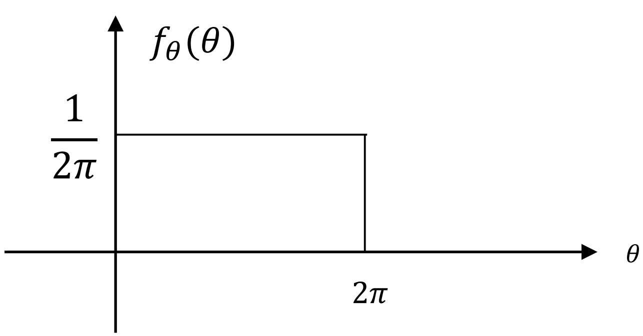
Where $\phi(r) = \frac{1}{\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta}$

Conditional distributions

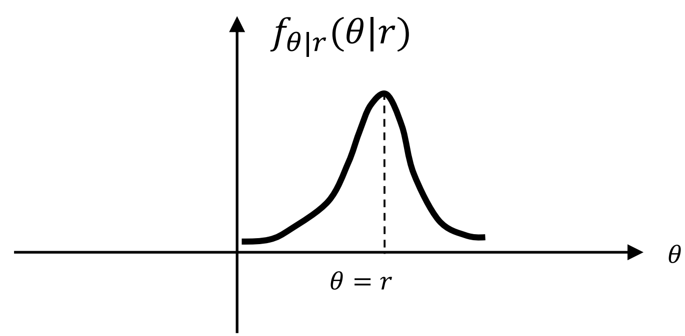
$$f(\theta|r) \ = \phi(r)e^{-(\theta-r)^2/2\sigma^2}, 0 < \theta < 2\pi,$$

Where $\phi(r) \ = \dfrac{1}{\int_0^{2\pi} e^{-(r-\theta)^2/2\sigma^2} d\theta}$

- Notice that the knowledge about the observation *r* is reflected in the a-posteriori pdf of θ in Fig. 3.1 (b). It is no longer flat as the a-priori pdf in Fig. 3.1 (a), and it shows higher probabilities in the neighborhood of $\theta = r$.



(a) a-priori pdf of θ



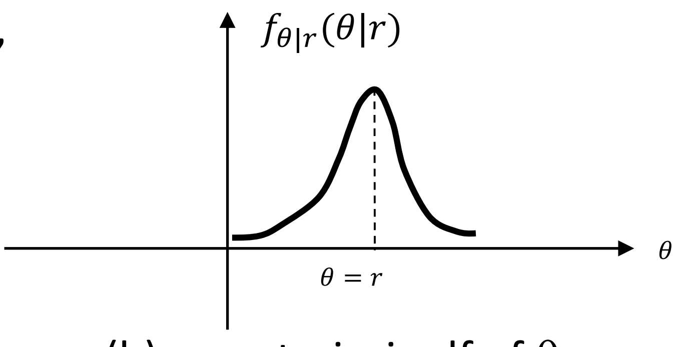
(b) a-posteriori pdf of θ

Fig. 3.1

Conditional distributions

- A very straight forward application:
  - Suppose a r.v θ is uniformly distributed in the interval $(50{,}80)$, and $r = \theta + n$, where n is noise. Note that r is also a r.v
  - We observe the r.v r 10 times, and get the result:

  {40,68,72,95,38,85,67,99,63,51}
  - What is a good estimation of θ?
- Based on the a-posteriori pdf of θ, we can see that
  - The observed pdf should be Gaussian
  - Gaussian Distribution has a mean $\theta = r$
  - Therefore, the mean of {40,68,72,95,38,85,67,99,63,51} is a good estimation of θ!
  - This is the basic of 'estimation theory'



(b) a-posteriori pdf of θ

Conditional mean

- We can use the conditional pdfs to define the conditional mean:

$$\mu_{X|Y} = E(X|Y = y) = \int_{-\infty}^{+\infty} x\, f_{X|Y}(x|y)\, dx$$

to be the conditional mean of *X* given *Y* = *y*. Notice that $E(X|Y = y)$ will be a function of *y*. Also

$$\mu_{Y|X} = E(Y|X = x) = \int_{-\infty}^{+\infty} y\, f_{Y|X}(y|x)\; dy\,.$$

- similarly, the conditional variance of *X* given *Y* = *y* is given by

$$\begin{aligned} Var(X|Y) = \sigma^2_{X|Y} &= E(X^2|Y = y) - (E(X|Y = y))^2 \\ &= E\big((X - \mu_{X|Y})^2|Y = y\big)\,. \end{aligned}$$

Lesson 3 – reading

- This week: Text book, Chapters 3-4
- Next week: Text book, pages 207-228 of Chapters 5

Lesson 3 – review examples for next week

1. Assume that X is a r.v follows exponential distribution with parameter$\lambda > 1$, find the expectation of $Y = e^X$

Note that for exponential distribution:

- Mean: $\mathrm{E(X)} = \int_0^{\infty} x\lambda e^{-\lambda x} dx = 1/\lambda$
- Variance:$\mathrm{D(X)} = 1/\lambda^2$
- pdf:$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

2. Given $Y = 0.5X + 1, Z = -0.5X + 1 \ and \ D(X) = 1$, calculate the covariance matrix of the vector (X, Y, Z)

Note: $\mathrm{cov}(X, Y) = E\left[\left(X - \overline{X}\right)\left(Y - \overline{Y}\right)\right] = E(XY) - E(X)E(Y)$