Chapter 5, 6
Multiple Random Variables

ENCS6161 - Probability and Stochastic Processes

Concordia University

Vector Random Variables

- A vector r.v. X is a function $X : S \to R^n$, where S is the sample space of a random experiment.
- Example: randomly pick up a student name from a list. $S = \{$all student names on the list$\}$. Let ω be a given outcome, e.g. Tom

$$\left.\begin{array}{ll} H(\omega): & \text{height of student } \omega \\ W(\omega): & \text{weight of student } \omega \\ A(\omega): & \text{age of student } \omega \end{array}\right\} H, W, A \text{ are r.v.s.}$$

Let $X = (H, W, A)$, then X is a vector r.v.

Events

- Each event involving $X = (X_1, X_2, \cdots, X_n)$ has a corresponding region in R^n.
- Example: $X = (X_1, X_2)$ is a two-dimensional r.v.

$$A = \{X_1 + X_2 \leq 10\}$$

$$B = \{\min(X_1, X_2) \leq 5\}$$

$$C = \{X_1^2 + X_2^2 \leq 100\}$$

Pairs of Random Variables

- Pairs of discrete random variables
  - Joint probability mass function

$$P_{X;Y}(x_j; y_k) = P\{X = x_j \bigcap Y = y_k\} = P\{X = x_j; Y = y_k\}$$

Obviously $\sum_j \sum_k P_{X,Y}(x_j, y_k) = 1$.

  - Marginal Probability Mass Function

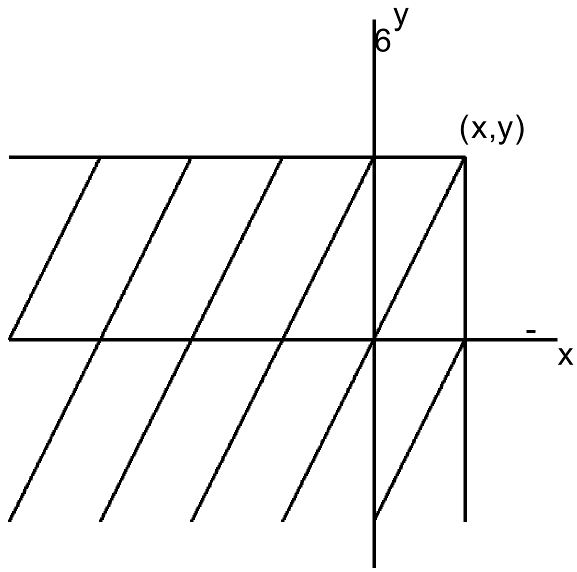
$$P_X(x_j) = P\{X = x_j\} = P\{X = x_j; Y = \text{anything}\} = \sum_{k=1}^{1} P_{X;Y}(x_j; y_k)$$

Similarly $P_Y(y_k) = \sum_{j=1}^{1} P_{X,Y}(x_j, y_k)$.

Pairs of Random Variables

- The joint CDF of X and Y (for both discrete and continuous r.v.s)

$$F_{X,Y}(x,y) = P\{X \leq x, Y \leq y\}$$



Pairs of Random Variables

- Properties of the joint CDF:
  1. $F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2)$, if $x_1 \le x_2, y_1 \le y_2$.
  2. $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$
  3. $F_{X,Y}(\infty, \infty) = 1$
  4. $F_X(x) = P\{X \le x\} = P\{X \le x, Y = \text{anything}\}$
     $= P\{X \le x, Y \le \infty\} = F_{X,Y}(X, \infty)$
     $F_Y(y) = F_{X,Y}(\infty, y)$
     $F_X(x), F_Y(y)$: Marginal cdf

Pairs of Random Variables

- The joint pdf of two jointly continuous r.v.s.

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

Obviously,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \mathrm{d}x \mathrm{d}y = 1$$

and

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x', y') \mathrm{d}y' \mathrm{d}x'$$

Pairs of Random Variables

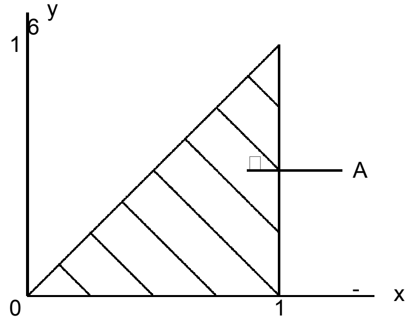
- The probability

$$P\{a \le X \le b, c \le Y \le d\} = \int_a^b \int_c^d f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x$$

In general,

$$P\{(X,Y) \in A\} = \int\int_A f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$

- Example:



$$\int_0^1 \int_0^x f_{X,Y}(x',y')\mathrm{d}y'\mathrm{d}x'$$

Pairs of Random Variables

- Marginal pdf:

$$
\begin{aligned}
f_X(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X,Y}(x,\infty) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\left(\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(x',y')\mathrm{d}y'\mathrm{d}x'\right) \\
&= \int_{-\infty}^{\infty} f_{X,Y}(x,y')\mathrm{d}y' \\
f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x',y)\mathrm{d}x'
\end{aligned}
$$

Pairs of Random Variables

- Example:

$$f_{X;Y}(x;y) = \begin{cases} 1 & \text{if } 0 \le x \le 1; 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $F_{X;Y}(x;y)$

1) $x \le 0$ or $y \le 0$; $F_{X;Y}(x;y) = 0$

2) $0 \le x \le 1$; and $0 \le y \le 1$

$$F_{X;Y}(x;y) = \int_0^x \int_0^y 1\, dy^0 dx^0 = xy$$

3) $0 \le x \le 1$; and $y > 1$

$$F_{X;Y}(x;y) = \int_0^x \int_0^1 1\, dy^0 dx^0 = x$$

4) $x > 1$ and $0 \le y < 1$

$$F_{X;Y}(x;y) = y$$

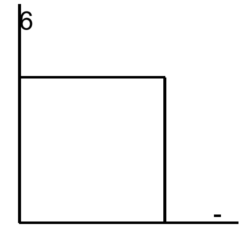
5) $x > 1$ and $y > 1$

$$F_{X;Y}(x;y) = 1$$

Independence

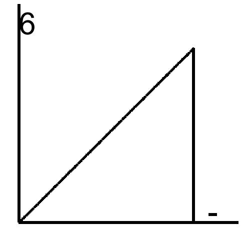
- $P_{X,Y}(x_j, y_k) = P_X(x_j)P_Y(y_k)$, for all x_j and y_k (discrete r.v.s)
  or $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for all x and y
  or $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all x and y
- Example:
  a)

$$f_{X,Y} = \begin{cases} 1 & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$



b)

$$f_{X,Y} = \begin{cases} 1 & 0 \le x \le \sqrt{2}, 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$



Conditional Probability

- If X is discrete,

$$F_Y(y \mid x) = \frac{P\{Y \le y; X = x\}}{P\{X = x\}} \quad \text{for } P\{X = x\} > 0$$

$$f_Y(y \mid x) = \frac{d}{dy} F_Y(y \mid x)$$

- If X is continuous, $P\{X = x\} = 0$

$$F_Y(y \mid x) = \lim_{h \to 0} F_Y(y \mid x < X \le x + h) = \lim_{h \to 0} \frac{P\{Y \le y; x < X \le x + h\}}{P\{x < X \le x + h\}}$$

$$= \lim_{h \to 0} \frac{\int_{-\infty}^{y} \int_{x}^{x+h} f_{X,Y}(x', y') dx' dy'}{\int_{x}^{x+h} f_X(x') dx'}$$

$$= \lim_{h \to 0} \frac{\int_{-\infty}^{y} f_{X,Y}(x, y') dy' \cdot h}{f_X(x) \cdot h} = \frac{\int_{-\infty}^{y} f_{X,Y}(x, y') dy'}{f_X(x)}$$

$$f_Y(y|x) = \frac{d}{dy} F_Y(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Conditional Probability

- If X, Y independent,

$$f_Y(y|x) = f_Y(y)$$

- Similarly,

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

So,

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

- Bayes Rule:

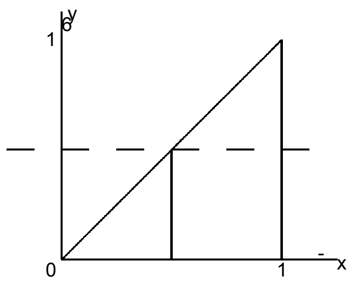
$$f_X(x|y) = \frac{f_X(x) \cdot f_Y(y|x)}{f_Y(y)}$$

Conditional Probability

- Example: A r.v. X is uniformly selected in $[0, 1]$, and then Y is selected uniformly in $[0, x]$. Find $f_Y(y)$
- Solution:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y|x) = 1 \cdot \frac{1}{x} = \frac{1}{x}$$

for $0 \le x \le 1, 0 \le y \le x$ and is 0 elsewhere.

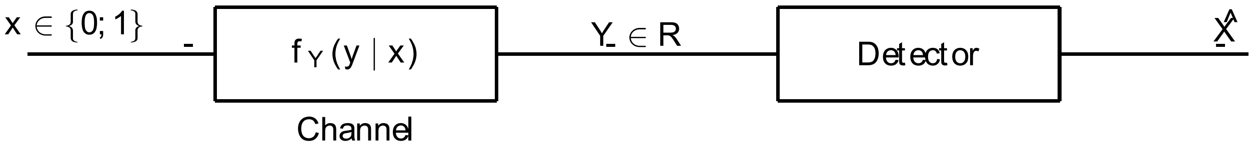


$$\begin{aligned} f_Y(y) &= \int_{\square 1}^{1} f_{X,Y}(x, y) \mathrm{d}x \\ &= \int_y^1 \frac{1}{x} \mathrm{d}x = -\ln y \end{aligned}$$

for $0 \le y \le 1$ and $f_Y(y) = 0$ elsewhere.

Conditional Probability

- Example:



$X = 0 \longrightarrow -A$ (volts), we assume $P\{X = 0\} = P_0$
$X = 1 \longrightarrow +A$ (volts), $P\{X = 1\} = P_1 = 1 - P_0$

- Decide $\hat{X} = 0$, if $P\{X = 0 \mid y\} \geq P\{X = 1 \mid y\}$
  Decide $\hat{X} = 1$, if $P\{X = 0 \mid y\} < P\{X = 1 \mid y\}$
- This is called the Maximum a posterior probability (MAP) detection.

Conditional Probability

- Binary communication over Additive White Gaussian Noise (AWGN) channel

$$f_Y(y|0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y+A)^2}{2\sigma^2}}$$

$$f_Y(y|1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-A)^2}{2\sigma^2}}$$

- Apply the MAP detection, we need to find $P\{X = 0|y\}$ and $P\{X = 1|y\}$. Note here X is discrete, Y is continuous.

Conditional Probability

- Use the similar approach (considering $x < X \le x + h$, and let $h \to 0$), we have

$$P\{X = 0|y\} = \frac{P\{X = 0\} f_Y(y|0)}{f_Y(y)}$$

$$P\{X = 1|y\} = \frac{P\{X = 1\} f_Y(y|0)}{f_Y(y)}$$

- Decide $\hat{X} = 0$, if

$$P\{X = 0|y\} \ge P\{X = 1|y\} \Rightarrow y \le \frac{\sigma^2}{2A} \ln \frac{p_0}{p_1}$$

  Decide $\hat{X} = 1$, if

$$P\{X = 0|y\} < P\{X = 1|y\} \Rightarrow y > \frac{\sigma^2}{2A} \ln \frac{p_0}{p_1}$$

- When $p_0 = p_1 = \frac{1}{2}$: Decide $\hat{X} = 0$, if $y \le 0$
  Decide $\hat{X} = 1$, if $y > 0$

Conditional Probability

- Prob of error:
  considering the special case $p_0 = p_1 = \frac{1}{2}$

$$
\begin{aligned}
P_\varepsilon &= P_0 P\{\hat{X} = 1 | X = 0\} + P_1 P\{\hat{X} = 0 | X = 1\} \\
&= P_0 P\{Y > 0 | X = 0\} + P_1 P\{Y \le 0 | X = 1\}
\end{aligned}
$$

$$P\{Y > 0 | X = 0\} = \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty e^{-\frac{(y+A)^2}{2\sigma^2}} \mathrm{d}y = Q\left(\frac{A}{\sigma}\right)$$

Similarly,

$$P\{Y \le 0 | X = 1\} = Q\left(\frac{A}{\sigma}\right)$$

$$\Rightarrow \; P_\varepsilon = Q\left(\frac{A}{\sigma}\right)$$

$$A \uparrow, P_\varepsilon \downarrow \qquad \sigma \uparrow, P_\varepsilon \uparrow$$

Conditional Expectation

- The conditional expectation

$$E[Y|x] = \int_{\Box 1}^{1} y f_Y(y|x) dy$$

In discrete case,

$$E[Y|x] = \sum_{y_i} y_i P_Y(y_i|x)$$

- An important fact:

$$E[Y] = E[E[Y|X]]$$

Proof:

$$E[E[Y|X]] = \int_{\Box 1}^{1} E[Y|x] f_X(x) dx = \int_{\Box 1}^{1} \int_{\Box 1}^{1} y f_Y(y|x) f_X(x) dy dx$$

$$= \int_{\Box 1}^{1} \int_{\Box 1}^{1} y f_{X;Y}(x;y) dy dx = E[Y]$$

In general:

$$E[h(Y)] = E[E[h(Y)|X]]$$

Multiple Random Variables

- Joint cdf

$$F_{X_1,\cdots,X_n}(x_1,\cdots x_n) = P[X_1 \le x_1,\cdots,X_1 \le x_n]$$

- Joint pdf

$$f_{X_1,\cdots,X_n}(x_1,\cdots x_n) = \frac{\partial^n}{\partial x_1,\cdots \partial x_n} F_{X_1,\cdots,X_n}(x_1,\cdots x_n)$$

  If discrete, joint pmf

$$P_{X_1,\cdots,X_n}(x_1,\cdots x_n) = P[X_1 = x_1,\cdots X_n = x_n]$$

- Marginal pdf

$$f_{X_i}(x_i) = \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{\text{all } x_1,\cdots,x_n \text{ except } x_i} f(x_1,\cdots x_n)dx_1\cdots dx_n$$

Independence

- $X_1, \cdots X_n$ are independent iff

$$F_{X_1,\cdots X_n}(x_1, \cdots x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

for <u>all</u> $x_1, \cdots, x_n$

- If we use pdf,

$$f_{X_1,\cdots X_n}(x_1, \cdots x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

for <u>all</u> $x_1, \cdots, x_n$

Functions of Several r.v.s

- One function of several r.v.s

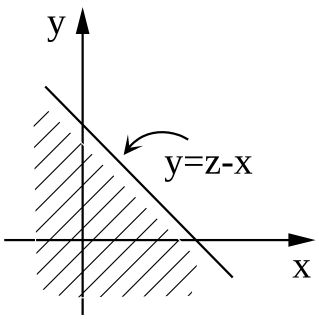
$$Z = g(X_1, \cdots X_n)$$

Let $R_z = \{\underline{x} = (x_1, \cdots, x_n)$ s.t. $g(\underline{x}) \le z\}$ then

$$\begin{aligned} F_z(z) &= P\{\underline{X} \in R_z\} \\ &= \int_{\underline{x} \in R_z} \cdots \int f_{X_1,\cdots,X_n}(x_1, \cdots, x_n) dx_1 \cdots dx_n \end{aligned}$$

Functions of Several r.v.s

- Example: $Z = X + Y$, find $F_z(z)$ and $f_z(z)$ in terms of $f_{X,Y}(x,y)$



$$
\begin{aligned}
Z &= X + Y \le z \ \Rightarrow\ Y \le z - X \\
F_z(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x,y)\,dy\,dx \\
f_z(z) &= \frac{d}{dz} F_z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx
\end{aligned}
$$

If X and Y are independent

$$
f_z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx
$$

Functions of Several r.v.s

- Example: let $Z = X/Y$. Find the pdf of Z if X and Y are independent and both exponentially distributed with mean one.
- Can use the similar approach as previous example, but complicated. Fix $Y = y$, then $Z = X/y$ and $f_Z(z|y) = |y| f_X(yz|y)$. So

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_{Z,Y}(z,y)dy = \int_{-\infty}^{\infty} f_Z(z|y) f_Y(y)dy \\
&= \int_{-\infty}^{\infty} |y| \, f_X(yz|y) f_Y(y)dy = \int_{-\infty}^{\infty} |y| \, f_{X,Y}(yz,y)dy
\end{aligned}
$$

Functions of Several r.v.s

- Since X, Y are indep. exponentially distributed

$$
\begin{aligned}
f_Z(z) &= \int_0^{1} y\, f_X(yz) f_Y(y) dy = \int_0^{1} y e^{\square yz} e^{\square y} dy \\
&= \frac{1}{(1+z)^2} \quad \text{for } z > 0
\end{aligned}
$$

Transformation of Random Vectors

- Transformation of Random Vectors

$$Z_1 = g_1(X_1 \cdots X_n) \; Z_2 = g_2(X_1 \cdots X_n) \; \cdots \; Z_n = g_n(X_1 \cdots X_n)$$

The joint CDF of $\underline{Z}$ is

$$\begin{aligned} F_{Z_1 \cdots Z_n}(z_1 \cdots z_n) &= P\{Z_1 \le z_1, \cdots, Z_n \le z_n\} \\ &= \underset{\underline{x}: g_k(\underline{x}) \le z_k}{\int \cdots \int} f_{X_1 \cdots X_n}(x_1 \cdots x_n) dx_1 \cdots dx_n \end{aligned}$$

pdf of Linear Transformations

- If $\underline{Z} = A\underline{X}$, where A is a $n \times n$ invertible matrix.

$$
\begin{aligned}
f_{\underline{Z}}(\underline{z}) &= f_{Z_1 \cdots Z_n}(z_1, \cdots, z_n) \\
&= \left. \frac{f_{X_1 \cdots X_n}(x_1, \cdots, x_n)}{|A|} \right|_{\underline{x} = A^{-1}\underline{z}} = \frac{f_{\underline{X}}(A^{-1}\underline{z})}{|A|}
\end{aligned}
$$

$|A|$ is the absolute value of the determinant of A.

e.g if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $|A| = |ad - bc|$

pdf of General Transformations

- $Z_1 = g_1(\underline{X}),\ Z_2 = g_2(\underline{X}), \cdots, Z_n = g_n(\underline{X})$ where $\underline{X} = (X_1, \cdots, X_n)$
- We assume that the set of equations:
$$z_1 = g_1(\underline{x}), \cdots, z_n = g_n(\underline{x})$$
has a unique solution given by
$$x_1 = h_1(\underline{z}), \cdots, x_n = h_n(\underline{z})$$
- The joint pdf of $\underline{Z}$ is given by
$$f_{Z_1 \cdots Z_n}(z_1 \cdots z_n) = \frac{f_{X_1 \cdots X_n}(h_1(\underline{z}), \cdots, h_n(\underline{z}))}{|J(x_1, \cdots, x_n)|}$$
$$= f_{X_1 \cdots X_n}(h_1(\underline{z}), \cdots, h_n(\underline{z}))\, |J(z_1, \cdots, z_n)| \qquad (*)$$
where $J(x_1, \cdots, x_n)$ is called the Jacobian of the transformation.

pdf of General Transformations

- The Jacobian of the transformation

$$J(x_1, \cdots, x_n) = det \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{bmatrix}$$

and

$$J(z_1, \cdots, z_n) = det \begin{bmatrix} \frac{\partial h_1}{\partial z1} & \cdots & \frac{\partial h_1}{\partial z_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial h_n}{\partial z_1} & \cdots & \frac{\partial h_n}{\partial z_n} \end{bmatrix} = \frac{1}{J(x_1 \cdots x_n)}\Bigg|_{\underline{x} = \underline{h}(\underline{z})}$$

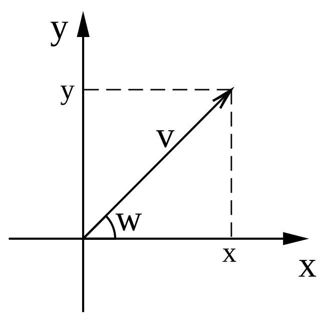
- Linear transformation is a special case of $(*)$

pdf of General Transformations

- Example: let X and Y be zero-mean unit-variance independent Gaussian r.vs. Find the joint pdf of V and W defined by:

$$\begin{cases} V = (X^2 + Y^2)^{\frac{1}{2}} \\ W = \backslash\ (X, Y) = \arctan(Y/X) \quad W \in [0, 2\pi) \end{cases}$$

- This is a transformation from Cartesian to Polar coordinates. The inverse transformation is:



$$\begin{cases} x = v\cos(w) \\ y = v\sin(w) \end{cases}$$

pdf of General Transformations

- The Jacobian

$$J(v,w) = \begin{bmatrix} \cos w & -v\sin w \\ \sin w & v\cos w \end{bmatrix} = v\cos^2 w + v\sin^2 w = v$$

- Since X and Y are zero-mean unit-variance independent Gaussian r.v.s,

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

- The joint pdf of V, W is then

$$f_{V,W}(v,w) = v \cdot \frac{1}{2\pi} e^{-\frac{(v^2\cos^2 w + v^2 \sin^2 w)}{2}} = \frac{v}{2\pi} e^{-\frac{v^2}{2}}$$

for $v \geq 0$ and $0 \leq w < 2\pi$

pdf of General Transformations

- The marginal pdf of V and W

$$f_V(v) = \int_{-\infty}^{\infty} f_{V,W}(v,w)dw = \int_0^{2\pi} \frac{v}{2\pi} e^{-\frac{v^2}{2}} dw = ve^{-\frac{v^2}{2}}$$

for $v \geq 0$. This is called the Rayleigh Distribution.

$$f_W(w) = \int_{-\infty}^{\infty} f_{V,W}(v,w)dv = \frac{1}{2\pi} \int_0^{\infty} ve^{-\frac{v^2}{2}} dv = \frac{1}{2\pi}$$

for $0 \leq w < 2\pi$

- Since

$$f_{V,W}(v,w) = f_V(v) f_W(w)$$

$\Rightarrow V, W$ are independent.

Expected Value of Functions of r.v.s

- Let $Z = g(X_1, X_2, \cdots, X_n)$ then
  $$E[Z] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \cdots, x_n) f_{X_1 \cdots X_n}(x_1, \cdots, x_n) dx_1 \cdots dx_n$$
  For discrete case,
  $$E[Z] = \underbrace{\sum \cdots \sum}_{\text{all possible } \underline{x}} g(x_1, \cdots, x_n) P_{X_1 \cdots X_n}(x_1, \cdots, x_n)$$

- Example: $Z = X_1 + X_2 + \cdots + X_n$
  $$\begin{aligned} E[Z] &= E[X_1 + X_2 + \cdots + X_n] \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_1 + \cdots + x_n) f_{X_1 \cdots X_n}(x_1 \cdots x_n) dx_1 \cdots dx_n \\ &= E[X_1] + \cdots + E[X_n] \end{aligned}$$

Expected Value of Functions of r.v.s

- Example: $Z = X_1 X_2 \cdots X_n$

  $$E[Z] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1 \cdots x_n f_{X_1 \cdots X_n}(x_1 \cdots x_n) dx_1 \cdots dx_n$$

  If $X_1, X_2, \cdots, X_n$ are indep.

  $$E[Z] = E[X_1 X_2 \cdots X_n] = E[X_1]E[X_2] \cdots E[X_n]$$

- The (j,k)-th moment of two r.v.s $X \& Y$ is

  $$E[X^j Y^k] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^j y^k f_{X,Y}(x,y) dx dy$$

  If $j = k = 1$, it is called the correlation.

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

  If $E[XY] = 0$, we call $X \& Y$ are orthogonal.

Expected Value of Functions of r.v.s

- The (j,k)-th *central moment* of X, Y is
$$E\left[(X - E(X))^j (Y - E(Y))^k\right]$$
when $j = 2,\ k = 0,\ \Rightarrow\ Var(X)$
$j = 0,\ k = 2,\ \Rightarrow\ Var(Y)$
- When $j = k = 1$, it is called the *covariance* of X, Y
$$\begin{aligned} Cov(X,Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E[XY] - E[X]E[Y] = Cov(Y,X) \end{aligned}$$
- The *correlation coefficient* of X and Y is defined as
$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$
where $\sigma_X = \sqrt{Var(X)}$ and $\sigma_Y = \sqrt{Var(Y)}$.

Expected Value of Functions of r.v.s

- The correlation coefficient $-1 \leq \rho_{X,Y} \leq 1$
  Proof:
$$\begin{aligned} 0 &\leq E\left\{\left(\frac{X-E[X]}{\sigma_X} \pm \frac{Y-E[Y]}{\sigma_Y}\right)^2\right\} \\ &= 1 \pm 2\rho_{X,Y} + 1 = 2(1 \pm \rho_{X,Y}) \end{aligned}$$
- If $\rho_{X,Y} = 0$, X, Y are said to be *uncorrelated*.
- If X, Y are independent,
  $E[XY] = E[X]E[Y] \Rightarrow Cov(X,Y) = 0 \Rightarrow \rho_{X,Y} = 0.$
  Hence, X, Y are uncorrelated.
- The converse is not always true. It is true in the case of Gaussian r.v.s (will be discussed later)

Expected Value of Functions of r.v.s

- Example: θ is uniform in $[0, 2\pi)$.
  Let $X = \cos\theta$ and $Y = \sin\theta$
  X and Y are not independent, since $X^2 + Y^2 = 1$.
  However

$$
\begin{aligned}
E[XY] &= E[\sin\theta\cos\theta] = E[\frac{1}{2}\sin(2\theta)] \\
&= \int_0^{2\pi} \frac{1}{2\pi}\frac{1}{2} sin(2\theta) d\theta = 0
\end{aligned}
$$

  We can also show $E[X] = E[Y] = 0$. So

$$
Cov(X, Y) = E[XY] - E[X]E[Y] = 0 \Rightarrow \rho_{X,Y} = 0
$$

  X, Y are uncorrelated but not independent.

Joint Characteristic Function

- Joint Characteristic Function

$$\Phi_{X_1 \cdots X_n}(w_1, \cdots, w_n) = E\left[e^{j(w_1 X_1 + \cdots + w_n X_n)}\right]$$

- For two variables

$$\Phi_{X,Y}(w_1, w_2) = E\left[e^{j(w_1 X + w_2 Y)}\right]$$

- Marginal characteristic function

$$\Phi_X(w) = \Phi_{X,Y}(w, 0) \quad \Phi_Y(w) = \Phi_{X,Y}(0, w)$$

- If X, Y are independent

$$\begin{aligned} \Phi_{X,Y}(w_1, w_2) &= E[e^{jw_1 X + jw_2 Y}] \\ &= E[e^{jw_1 X}]E[e^{jw_2 Y}] = \Phi_X(w_1)\Phi_Y(w_2) \end{aligned}$$

Joint Characteristic Function

- If $Z = aX + bY$

$$\Phi_Z(w) = E[e^{jw(aX+bY)}] = \Phi_{X,Y}(aw, bw)$$

- If $Z = X + Y$, X and Y are independent

$$\Phi_Z(w) = \Phi_{X,Y}(w, w) = \Phi_X(w)\Phi_Y(w)$$

Jointly Gaussian Random Variables

- Consider a vector of random variables $\underline{X} = (X_1, X_2, \cdots X_n)$. Each with mean $m_i = E[X_i]$, for $i = 1, \cdots, n$ and the covariance matrix

$$K = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ \vdots & \vdots & \cdots & \vdots \\ Cov(X_n, X_1) & \cdots & \cdots & Var(X_n) \end{bmatrix}$$

Let $\underline{m} = [m_1, \cdots, m_n]^T$ be the mean vector and $\underline{x} = [x_1, \cdots, x_n]^T$ where $(\cdot)^T$ denotes transpose. Then $X_1, X_2, \cdots X_n$ are said to be the jointly Gaussian if their joint pdf is:

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |K|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\underline{x} - \underline{m})^T K^{-1} (\underline{x} - \underline{m}) \right\}$$

where $|K|$ is the determinant of K.

Jointly Gaussian Random Variables

- For $n = 1$, let $Var(X_1) = \sigma^2$, then $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$
- For $n = 2$, denote the r.v.s by X and Y. Let

$$\underline{m} = \begin{bmatrix} m_X \\ m_Y \end{bmatrix} \quad K = \begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

Then $|K| = \sigma_X^2 \sigma_Y^2 (1 - \rho_{XY}^2)$

$$K^{-1} = \frac{1}{\sigma_X^2 \sigma_Y^2 (1 - \rho_{XY}^2)} \begin{bmatrix} \sigma_Y^2 & -\rho_{XY}\sigma_X\sigma_Y \\ -\rho_{XY}\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}$$

and

$$\begin{aligned} f_{X,Y}(x,y) \quad = \quad & \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left\{ -\frac{1}{2(1-\rho_{XY}^2)} \left[\left(\frac{x-m_x}{\sigma_x}\right)^2 \right.\right. \\ & \left.\left. -2\rho_{XY}\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{y-m_y}{\sigma_y}\right) + \left(\frac{y-m_y}{\sigma_y}\right)^2 \right] \right\} \end{aligned}$$

Jointly Gaussian Random Variables

- The marginal pdf of X :

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

The marginal pdf of Y :

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}$$

- If $\rho_{XY} = 0 \Rightarrow X, Y$ are independent.

- The conditional pdf.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\exp\left\{-\frac{1}{2(1-\rho_{XY}^2)\sigma_X^2}\left[x-\rho_{XY}\frac{\sigma_X}{\sigma_Y}(y-m_y)-m_x\right]^2\right\}}{\sqrt{2\pi\sigma_X^2(1-\rho_{XY}^2)}}$$

$$\sim N(\underbrace{\rho_{XY}\frac{\sigma_X}{\sigma_Y}(y-m_y)+m_x}_{\Downarrow \; E[X|Y]} ; \underbrace{\sigma_X^2(1-\rho_{XY}^2)}_{\Downarrow \; Var(X|Y)})$$

Linear Transformation of Gaussian r.v.s

- Let $\underline{X} \sim N(\underline{m}; K)$, $\underline{Y} = A\underline{X}$ then
  $\underline{Y} \sim N(\underline{\hat{m}}; C)$, where $\hat{m} = A\underline{m}$ and $C = AKA^T$
  proof:

$$f_{\underline{Y}}(\underline{y}) = \frac{f_{\underline{X}}(A^{-1}\underline{y})}{|A|} = \frac{1}{(2\pi)^{\frac{n}{2}}|A||K|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(A^{-1}\underline{y}-\underline{m})^T K^{-1}(A^{-1}\underline{y}-\underline{m})\right\}$$

Note that $A^{-1}\underline{y}-\underline{m} = A^{-1}(\underline{y}-A\underline{m}) = A^{-1}(\underline{y}-\underline{\hat{m}})$, so

$$\begin{aligned}(A^{-1}\underline{y}-\underline{m})^T K^{-1}(A^{-1}\underline{y}-\underline{m}) &= (\underline{y}-\underline{\hat{m}})^T (A^{-1})^T K^{-1} A^{-1}(\underline{y}-\underline{\hat{m}}) \\ &= (\underline{y}-\underline{\hat{m}})^T C^{-1}(\underline{y}-\underline{\hat{m}})\end{aligned}$$

and $|A||K|^{\frac{1}{2}} = \left(|A|^2|K|\right)^{\frac{1}{2}} = \left(|AKA^T|\right)^{\frac{1}{2}} = |C|^{\frac{1}{2}}$

$$f_{\underline{Y}}(\underline{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}|C|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\underline{y}-\underline{\hat{m}})^T C^{-1}(\underline{y}-\underline{\hat{m}})\right\} \sim N(\underline{\hat{m}}; C)$$

Linear Transformation of Gaussian r.v.s

- Since K is symmetric, it is always possible to find a matrix A s.t.

  $\Lambda = AKA^T$ is diagonal. $\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$

  so

$$
\begin{aligned}
f_{\underline{Y}}(\underline{y}) &= \frac{1}{(2\pi)^{\frac{n}{2}}|\Lambda|^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{y}-\underline{\hat{m}})^T \Lambda^{-1}(\underline{y}-\underline{\hat{m}})} \\
&= \frac{1}{\sqrt{2\pi\lambda_1}} e^{-\frac{(y_1-\hat{m}_1)^2}{2\lambda_1}} \cdots \frac{1}{\sqrt{2\pi\lambda_n}} e^{-\frac{(y_n-\hat{m}_n)^2}{2\lambda_n}}
\end{aligned}
$$

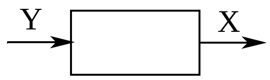
  That is, we can transform $\underline{X}$ into n independent Gaussian r.v.s $Y_1 \cdots Y_n$ with means $\hat{m}_i$ and variance λ_i.

Mean Square Estimation

- We use $g(X)$ to estimate Y, write as $\hat{Y} = g(X)$. The cost associated with the estimation error is $C(Y - \hat{Y})$.

  e.g.

$$\begin{aligned} C(Y - \hat{Y}) &= (Y - \hat{Y})^2 \\ &= (Y - g(X))^2 \end{aligned}$$



The mean square error

$$E[C] = E[(Y - g(X))^2]$$

- Case 1: if $\hat{Y} = a$

$$E[(Y-a)^2] = E[Y^2] - 2aE[Y] + a^2 = f(a)$$

$$\frac{df}{da} = 0 \Rightarrow a^{\square} = E[Y] = \mu_Y$$

The mean square error: $E[(Y - \mu_Y)^2] = Var[Y]$

Mean Square Estimation

- Case 2: if $\hat{Y} = aX + b$, then $E[(Y - aX - b)^2] = f(a, b)$

$$\begin{cases} \frac{\partial f}{\partial a} = 0 \\ \frac{\partial f}{\partial b} = 0 \end{cases} \Rightarrow a^{\square} = \rho_{XY}\frac{\sigma_Y}{\sigma_X},\ b^{\square} = \mu_Y - a^{\square}\mu_X$$

$$\Rightarrow \hat{Y} = \rho_{XY}\frac{\sigma_Y}{\sigma_X}(X - \mu_X) + \mu_Y$$

  This is called the MMSE linear estimation.

- The mean square error:

$$E[(Y - \hat{Y})^2] = \sigma_Y^2(1 - \rho_{XY}^2)$$

  If $\rho_{XY} = 0$, $\hat{Y} = \mu_Y$, error$= \sigma_Y^2$, reduces to case 1.

  If $\rho_{XY} = \pm 1$, $\hat{Y} = \pm\frac{\sigma_Y}{\sigma_X}(X - \mu_X) + \mu_Y$, and error$= 0$

Mean Square Estimation

- Case 3: $\hat{Y}$ is a general function of X.

  $$E[(Y-\hat{Y})^2] = E[(Y-g(X))^2] = E[E[(Y-g(X))^2|X]]$$
  $$= \int_{\square 1}^{1} E[(Y-g(x))^2|x] f_X(x)dx$$

  For any x, choose $g(x)$ to minimize $E[(Y-g(x))^2|X=x]$

  $$\Rightarrow g^{\square}(x) = E[Y|X=x]$$

  This is called the MMSE estimation.

- Example: X, Y joint Gaussian.

  $$E[Y|X] = \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}(X-\mu_X)+\mu_Y$$

  The MMSE estimation is linear for Gaussian.

Mean Square Estimation

- Example: $X \sim$ uniform$(-1, 1)$, $Y = X^2$. We have

$$E[X] = 0$$

$$\rho_{XY} = E[XY] - E[X]E[Y] = E[XY] = E[X^3] = 0$$

So, the MMSE linear estimation:

$$\hat{Y} = \mu_Y$$

and the error is σ_Y^2.

The MMSE estimation:

$$g^{\square}(x) = E[Y|X = x] = E[X^2|X = x] = x^2$$

So $\hat{Y} = X^2$ and the error is 0.