Chapter 4 Multivariate Probability Distributions

4.1 Random Vectors and Joint Probability Distributions

- **Definition: Random Vector.** An n-dimensional *random vector* defined on a probability space $(S, \mathcal{B}, P)$, denoted as $\boldsymbol{Z} = (Z_1, Z_2, \cdots, Z_n)'$, is a $\mathcal{B}$-measurable function from the sample space S to $\mathbb{R}^n$, the n-dimensional Euclidean space. For each outcome $s \in S$, $\boldsymbol{Z}(s)$ is a n-dimensional real-valued vector and is called a *realization* of the random vector $\boldsymbol{Z}$.

- **Example: Bivariate Random Variable.** Consider a bivariate r.v. $\boldsymbol{Z} = (X, Y)$. For each outcome $s \in S$, $\boldsymbol{Z}(s)$ is a pair $(x, y) \in \mathbb{R}^2$.

4.1 Random Vectors and Joint Probability Distributions

- **Definition: Joint Cumulative Distribution Function.** The joint CDF of X and Y is defined as

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

for any pair $(x,y) \in \mathbb{R}^2$.

- **Lemma 1: Properties of $F_{XY}(x,y)$.**

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

(2) $F_{XY}(x,y)$ is non-decreasing in both x and y.

(3) $F_{X,Y}(x,y)$ is right continuous in both x and y.

4.1 Random Vectors and Joint Probability Distributions

- **Remark:** For all $x_1 \le x_2$ and $y_1 \le y_2$,

$$F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1) \ge 0.$$

Why? Because

$$P(x_1 < X \le x_2 \ , \ y_1 < Y \le y_2)$$
$$= F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1).$$

- **Theorem 1.** $F_X(x) = F_{XY}(x, +\infty)$ and $F_Y(y) = F_{XY}(+\infty, y)$.

- **Example: Bivariate Copula.** $U = F_X(X)$, $V = F_Y(Y)$. Then both U and V are $U[0,1]$ random variables. The joint CDF of (U, V)

$$C(u, v) = P(U \le u, V \le v)$$

is called the copula associated with the joint probability distribution of (X, Y).

4.1 Random Vectors and Joint Probability Distributions

The Discrete Case

- **Definition: Joint PMF.** Let X and Y be two DRV's. Then their joint PMF is defined as

$$f_{X,Y}(x,y) = P(X = x, Y = y)$$

for any point $(x,y) \in \mathbb{R}^2$.

- **Remarks:** The joint CDF of (X,Y) is

$$F_{XY}(x,y) = P(X \le x, Y \le Y) = \sum_{(u,v)\in\Omega_{XY}(x,y)} f_{XY}(u,v).$$

4.1 Random Vectors and Joint Probability Distributions

- **Lemma 2: Properties of $f_{XY}(x, y)$.**

  (1) $f_{X,Y}(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$.

  (2) $\sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} f_{XY}(x, y) = 1$, where Ω_X and Ω_Y are the support of X and Y respectively.

- **Definition: Support.** The support of two DRV's (X, Y) is defined as the set of all possible pairs of (x, y) which (X, Y) will take with strictly positive probability. That is,

$$\text{Support}(X, Y) = \Omega_{XY} = \{(x, y) \in \mathbb{R}^2 : f_{XY}(x, y) > 0\}.$$

- **Question:** $\Omega_{XY} = \Omega_X \times \Omega_Y$?

4.1 Random Vectors and Joint Probability Distributions

- **Example:** Let the experiment consist of three tosses of a coin, and let X="number of heads in all three tosses" and Y="number of tails in the last two tosses".
  - The sample space S and corresponding values of X and Y are

S	X	Y	S	X	Y
HHH	3	0	TTH	1	1
HHT	2	1	THT	1	1
HTH	2	1	HTT	1	2
THH	2	0	TTT	0	2

- – All possible values of (X, Y) and their probabilities are as follows

(X, Y)	0	1	2
0	0	0	$\frac{1}{8}$
1	0	$\frac{2}{8}$	$\frac{1}{8}$
2	$\frac{1}{8}$	$\frac{2}{8}$	0
3	$\frac{1}{8}$	0	0

  – In this example, $A = \{0, 1, 2, 3\}$ and $B = \{0, 1, 2\}$. $f_{XY}(x, y) > 0$ on a subset of $A \times B$.

4.1 Random Vectors and Joint Probability Distributions

- If the values of X is arranged as $x_1 < x_2 < \cdots$, and of Y as $y_1 < y_2 < \cdots$.
  Then, for $i > 1$, $j > 1$,

$$\begin{aligned} f_{XY}(x_i, y_j) &= \Delta_Y \Delta_X F_{XY}(x_i, y_j) \\ &= F_{XY}(x_i, y_j) - F_{XY}(x_i, y_{j-1}) - F_{XY}(x_{i-1}, y_j) + F_{XY}(x_{i-1}, y_{j-1}), \end{aligned}$$

  and

$$f_{XY}(x_i, y_j) = \begin{cases} F_{XY}(x_i, y_j) - F_{XY}(x_i, y_{j-1}), & \text{if } i = 1, j > 1, \\ F_{XY}(x_i, y_j) - F_{XY}(x_{i-1}, y_j), & \text{if } i > 1, j = 1, \\ F_{XY}(x_i, y_j), & \text{if } i = 1, j = 1. \end{cases}$$

4.1 Random Vectors and Joint Probability Distributions

The Continuous Case

- **Definition: Joint PDF.** Random variables (X, Y) are said to have continuous joint distribution if their joint CDF $F_{XY}(x, y)$ is absolutely continuous. There exists a nonnegative function $f_{XY}(x, y)$ such that for any subset $A \in \mathbb{R}^2$,

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

  The function $f_{XY}(x, y)$ is called *joint (or bivariate) probability density function* of (X, Y).

- **Lemma 3: Properties of Joint PDF.**

(1) $f_{XY}(x, y) \geq 0$ for all (x, y) in the xyplane;

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

- **Remarks:**
  - The definition implies for any Borel set $B \subset \mathbb{R}^2$ on the plane,
  $$P\{(X,Y) \in B\} = \int_B \int f_{XY}(x,y)\, dxdy.$$
  - Suppose $f_{XY}(x,y)$ is continuous on (x,y),
  $$f_{XY}(x,y) \simeq \frac{1}{\varepsilon\delta} P\{(X,Y) \in (x-0.5\varepsilon, x+0.5\varepsilon) \times (y-0.5\delta, y+0.5\delta)\}$$
  when ε, δ are small.
  - For continuous r.v. (X,Y), pdf $f_{XY}(x,y)$ can be recovered from cdf $F_{XY}(x,y)$ by
  $$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y).$$

- **Remarks:**

  – We can generalize the bivariate concepts in the continuous case to the multivariate concepts: The joint cdf in the continuous case is given by

  $$F(x_1, \cdots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(u_1, \cdots, u_n) du_1 \cdots du_n.$$

  Also, the joint pdf can be recovered from the joint cdf by

  $$f(x_1, \cdots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \cdots, x_n).$$

- **Support of (X, Y).** The support of (X, Y) is defined as

  $$\text{Support}(X, Y) = \{(x, y) \in \mathbb{R}^2 : f_{XY}(x, y) > 0\}.$$

- **Example:** Suppose (X, Y) has a joint cdf $F_{XY}(x, y) = \frac{1}{16}xy(x+y)$ for $0 \le x \le 2$ and $0 \le y \le 2$. Find (a) $f_{XY}(x, y)$; (b) $P(1 \le X \le 2, 1 \le Y \le 2)$.

- **Solution:**

(a) For $0 \le x \le 2$ and $0 \le y \le 2$,

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} = \frac{1}{8}(x+y).$$

(b) We have

$$P(1 \le X \le 2, 1 \le Y \le 2) = \int_1^2 \int_1^2 \frac{1}{8}(x+y)\, dx dy = \frac{3}{8},$$

or

$$\begin{aligned} P(1 \le X \le 2, 1 \le Y \le 2) &= F_{XY}(2,2) - F_{XY}(2,1) - F_{XY}(1,2) + F_{XY}(1,1) \\ &= \frac{3}{8}. \end{aligned}$$

- **Example:** Suppose $f_{XY}(x,y) = cx^2y$ for $x^2 \leq y \leq 1$. Find (a) the value of c; (b) $P(X > Y)$.

- **Solution:**

(a) $f_{XY}(x,y) \geq 0$ for $[-1,1] \times [x^2, 1]$,

$$\begin{aligned} 1 = \int_{-1}^{1} \int_{x^2}^{1} cx^2y \, dydx &= \int_{-1}^{1} \frac{c}{2} x^2 y^2 |_{x^2}^{1} dx \\ &= \frac{c}{2} \int_{-1}^{1} (x^2 - x^6) dy = c(\frac{1}{3} - \frac{1}{7}). \end{aligned}$$

Thus $c = \frac{21}{4}$.

(b)

$$P(X > Y) = \int_0^1 \int_{x^2}^{x} cx^2y \, dydx = \frac{c}{2}(\frac{1}{5} - \frac{1}{7}) = \frac{3}{20}.$$

4.2 Marginal Distributions

The Discrete Case

- **Definition: Discrete Marginal PMF's.** Suppose X and Y have joint discrete distribution with joint PMF $f_{XY}(x, y)$. Then the marginal PMF's of X and Y are defined as

$$f_X(x) = P(X = x) \; = \; \sum_{y \in \Omega_Y} f_{XY}(x, y), \text{ where } -\infty < x < \infty,$$
$$f_Y(y) = P(Y = y) \; = \; \sum_{x \in \Omega_X} f_{XY}(x, y), \text{ where } -\infty < y < \infty.$$

- **Lemma: Properties of $f_X(x)$ and $f_Y(y)$.**

(1) $f_X(x) \geq 0$ for all $x \in (-\infty, \infty)$;

(2) $\sum_{x \in \Omega_X} f_X(x) = 1$, where Ω_X is the support of X.

Similar results for $f_Y(y)$ also hold.

4.2 Marginal Distributions

- **Example:** Tossing a coin three times. Let X=“number of heads in all three tosses” and Y=“number of tails in the last two tosses”.

- The joint distribution and marginal distributions of (X, Y) are as follows

(X,Y)	0	1	2	X
0	0	0	$\frac{1}{8}$	$\frac{1}{8}$
1	0	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
2	$\frac{1}{8}$	$\frac{2}{8}$	0	$\frac{3}{8}$
3	$\frac{1}{8}$	0	0	$\frac{1}{8}$
Y	$\frac{2}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	

4.2 Marginal Distributions

- **Remark:** For bivariate random variable (X, Y), the joint distribution contains more information than the marginal distributions of X and Y separately. Given a joint distribution, we are able to recover the marginal distributions of X and of Y, but not conversely.

- **Example:** Toss a coin twice. Let $X =$"number of heads in the first toss", $Y_1 =$"number of tails in the first toss", and $Y_2 =$"number of tails in the second toss". The marginal distributions of $(X, Y = Y_1)$ and $(X, Y = Y_2)$ are the same, but their joint distributions are different.

4.2 Marginal Distributions

The Continuous Case

- **Definition: Continuous Marginal Distribution.** If (X, Y) is a pair of continuous random variables with bivariate probability density function $f_{XY}(x, y)$, then the functions

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x, y)dy \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x, y)dx \end{aligned}$$

are called the *marginal pdf* of variables X and Y, respectively.

4.2 Marginal Distributions

- **Example:** A man shoots at a circular target. Assume that he is certain to hit the target. However, he is unable to aim with any more precision, so that the probability of hitting a particular part of the target is proportional to the area of this part. Let X and Y be the horizontal and vertical distances from the center of the target to the point of impact. What is the the distribution of X?

- **Solution:**

  – The joint distribution of (X, Y) is

$$f_{XY}(x, y) = \begin{cases} \frac{1}{\pi R^2} & x^2 + y^2 \leq R^2, \\ 0 & \text{otherwise.} \end{cases}$$

4.2 Marginal Distributions

- – The marginal distribution of X is

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} f(x,y)dy \\ &= \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}dy \\ &= \frac{2\sqrt{R^2-x^2}}{\pi R^2}. \end{aligned}$$

- **Example:** Suppose $f_{XY}(x,y) = cy^2$ for $x^2 \le y \le 1$. Find (a) $f_X(x)$; (b) $f_Y(y)$.

- **Solution:**

  – For $-1 \le x \le 1$,

$$f_X(x) = \int_{x^2}^{1} cy^2 dy = \frac{c}{3}(1 - x^6).$$

  Because $\int_{-1}^{1} f_X(x)dx = 1$, so $c = \frac{7}{4}$, $f_X(x) = \frac{7}{12}(1 - x^6)$.

  – For $0 \le y \le 1$,

$$f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} cy^2 dx = \frac{7}{2} y^{5/2}.$$

4.2 Marginal Distributions

- **Example:** Suppose (X, Y) follows bivariate Normal distribution $N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2, \rho)$, *i.e.*,

$$
\begin{aligned}
& f_{XY}(x, y) \\
& = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\
& \quad exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho \left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 \right] \right\},
\end{aligned}
$$

then

$$
\begin{aligned}
f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma_1} exp\left\{ -\frac{(x-\mu_1)^2}{2\sigma_1^2} \right\}, \\
f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma_2} exp\left\{ -\frac{(y-\mu_2)^2}{2\sigma_2^2} \right\}.
\end{aligned}
$$

4.2 Marginal Distributions

- **Multivariate Marginal Distribution:**
  - When there are more than two random variables, we can define not only the marginal distributions of individual random variables but also the joint marginal distributions of any subset of random variables.
  - Let $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ be a n-dimensional random vector. Divide the random variables in $\boldsymbol{X}$ into $\boldsymbol{Y}$ and $\boldsymbol{Z}$. In the discrete case, the *marginal distribution* of $\boldsymbol{Y}$ is given by

$$f_Y(\boldsymbol{y}) = P(\boldsymbol{Y} = \boldsymbol{y}) = \sum_{\boldsymbol{z}} P(\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{Z} = \boldsymbol{z}).$$

In the continuous case, the *marginal density* of $\boldsymbol{Y}$ is given by

$$f_Y(\boldsymbol{y}) = \int_{\boldsymbol{z}} f_{YZ}(\boldsymbol{y}, \boldsymbol{z})\, d\boldsymbol{z}.$$

4.3 Conditional Distributions

- **Definition: Discrete Conditional Distribution.** Let (X, Y) be a discrete bivariate random variable. For y such that the marginal probability $P(Y = y) > 0$,

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

Similarly, for x such that the marginal probability $P(X = x) > 0$,

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

- **Remarks:**
  - For y such that $f_Y(y) = 0$, conditional pmf $f_{X|Y}(x \mid y)$ is not defined.
  - For y such that $f_Y(y) > 0$, we have $f_{X|Y}(x \mid y) \geq 0$ and $\sum_x f_{X|Y}(x \mid y) = 1$. $f_{X|Y}(x \mid y)$ can be considered as a pmf.

4.3 Conditional Distributions

- **Example:** Let the experiment consist of two tosses of a die. We let X_1 and X_2 denote the result of the first and the second toss, respectively, and put $Z = |X_1 - X_2|$. The joint distribution of X_1 and Z are as follows

(X_1, Z)	0	1	2	3	4	5	X_1
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
Z	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	

- $P(X_1 \text{ odd} \mid Z = 5) = 1/2$, $P(Z > 2 \mid X_1 = 5) = 1/3$.

4.3 Conditional Distributions

Continuous Case

- **Definition: Conditional Probability Density Function.** For continuous bivariate random variable (X, Y), the *conditional probability density functions* $f_{X|Y}$ and $f_{Y|X}$ are defined by

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

provided that $f_Y(y) > 0$, and

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

provided that $f_X(x) > 0$, where $f_{XY}(x, y)$ is the joint density of (X, Y) and f_X and f_Y are the marginal densities of X and Y, respectively.

- **Remark:** For given y such that $f_Y(y) > 0$, $f_{X|Y}(x \mid y)$ can be considered as a pdf; similarly, for given x such that $f_X(x) > 0$, $f_{Y|X}(y \mid x)$ can be considered as a pdf.

- **Remarks:**
  - Assume for simplicity that $f_{XY}(x,y)$, and hence also $f_X(x)$, $f_Y(y)$, are continuous. Then for any Borel sets $A, B \subset \mathbb{R}$,

$$\begin{aligned}
\lim_{h\to 0} P(X \in A \mid y \le Y \le y+h) &= \frac{\int_A \int_y^{y+h} f_{XY}(u,v)\,dv\,du}{\int_y^{y+h} f_Y(v)dv} \\
&= \frac{\int_A f_{XY}(u,y)\,du}{f_Y(y)} = \int_A f_{X|Y}(u \mid y)du,
\end{aligned}$$

$$\begin{aligned}
\lim_{h\to 0} P(Y \in B \mid x \le X \le x+h) &= \frac{\int_B \int_x^{x+h} f_{XY}(u,v)\,du\,dv}{\int_x^{x+h} f_X(u)du} \\
&= \frac{\int_B f_{XY}(x,v)\,dv}{f_X(x)} = \int_B f_{Y|X}(v \mid x)dv.
\end{aligned}$$

- **Example:** Let the joint density of (X, Y) be given by

$$f_{XY}(x,y) = \begin{cases} cxy^2 & 0 \le x \le 1,\, 0 \le y \le 1,\, x + y \ge 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find $f_{X|Y}(x \mid y)$ and $f_{Y|X}(y \mid x)$.

- **Solution:**

  – For $0 \le x \le 1$,

$$f_X(x) = \int_{1-x}^{1} cxy^2 dy = \frac{c}{3}[x - x(1-x)^3] = \frac{c}{3}(3x^2 - 3x^3 + x^4),$$

and for $0 \le y \le 1$,

$$f_Y(y) = \int_{1-y}^{1} cxy^2 dx = \frac{c}{2}[y^2 - y^2(1-y)^2] = \frac{c}{2}(2y^3 - y^4).$$

4.3 Conditional Distributions

- – For given $0 < y < 1$,

$$\begin{aligned} f_{X|Y}(x \mid y) &= f_{XY}(x,y)/f_Y(y) \\ &= \frac{2x}{2y - y^2} \end{aligned}$$

  for $1 - y \leq x \leq 1$; $f_{X|Y}(x \mid y) = 0$ otherwise.

  – For given $0 < x < 1$,

$$\begin{aligned} f_{Y|X}(y \mid x) &= f_{XY}(x,y)/f_X(x) \\ &= \frac{3y^2}{3x - 3x^2 + x^3} \end{aligned}$$

  for $1 - x \leq y \leq 1$; $f_{Y|X}(y \mid x) = 0$ otherwise.

4.3 Conditional Distributions

- **Example:** A point X is chosen at random from the interval $[A, B]$ according to the uniform distribution, and then a point Y is chosen at random, again with uniform distribution, from the interval $[X, B]$. Find the marginal distribution of Y.

- **Solution:**

  – For $A \leq x \leq y \leq B$,

$$f_{XY}(x, y) = f_X(x) f_{Y|X}(y \mid x) = \frac{1}{(B-A)(B-x)}.$$

  – For $A \leq y \leq B$,

$$\begin{aligned} f_Y(y) &= \int_A^y \frac{1}{(B-A)(B-x)}\, dx \\ &= \frac{1}{B-A} \log\left(\frac{B-A}{B-y}\right). \end{aligned}$$

- **Example:** Assume that Y takes one of n integer values with $P(Y = i) = p_i$ and $p_1 + \cdots + p_n = 1$. For a given $Y = i$, the random variable X has a continuous distribution with density $\phi_i(x)$. Find the marginal distribution of X.

- **Solution:**

$$
\begin{aligned}
P(X \leq x) &= \sum_{i=1}^{n} P(X \leq x \mid Y = i) P(Y = i) \\
&= \sum_{i=1}^{n} p_i \int_{-\infty}^{x} \phi_i(u) du.
\end{aligned}
$$

The pdf of X is $f_X(x) = \sum_{i=1}^{n} p_i \phi_i(x)$.

- **Multivariate Case.** When we are dealing with more than two random variables, we can consider various kinds of conditional distribution. E.g. we can define

$$f_{X_i|X^{i-1}}(x_i|x^{i-1}) = \frac{f_{X^i}(x^i)}{f_{X^{i-1}}(x^{i-1})}$$

if $f_{X^{i-1}}(x^{i-1}) > 0$, where $X^{i-1} = (X_1, X_2, ..., X_{i-1})'$ and $x^{i-1} = (x_1, x_2, ..., x_{i-1})'$. We can also define

$$f_{(X_1,X_2)|(X_3,X_4)}(x_1, x_2|x_3, x_4) = \frac{f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4)}{f_{X_3,X_4}(x_3, x_4)}$$

if $f_{X_3X_4}(x_3, x_4) > 0$.

4.4 Independence

- **Definition: Independence.** We say that random variables X and Y are *independent* if events $\{X \in A\}$ and $\{Y \in B\}$ are independent for all Borel sets A, B, that is

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

- **Definition: Independence.** The random variables X and Y are independent if their joint and marginal cdf's satisfy the following condition: For every $x, y \in \mathbb{R}$, we have

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

where $F_X(x)$ and $F_Y(y)$ are marginal cdf's of X and Y, respectively.

4.4 Independence

- **Remarks:**
  - For discrete bivariate random variable (X, Y), independence is implied by

$$P(X = x, Y = y) = P(X = x)P(Y = y),$$

    for every (x, y).
  - For continuous bivariate random variable (X, Y), suppose the joint pdf $f_{XY}(x, y)$ is continuous on its support. Then independence is implied by

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

    where $f_{XY}(x, y)$ is the joint density, $f_X(x)$ and $f_Y(y)$ are the marginal densities of X and Y, respectively.

4.4 Independence

- **Theorem.** Two discrete random variables (X, Y) are independent if and only if

$$f_{XY}(x,y) = f_X(x)f_Y(y) \text{ for } \text{ all pairs of } (x,y) \in \mathbb{R}^2,$$

where $f_{XY}(x,y)$, $f_X(x)$, $f_Y(y)$ are the joint and marginal PMF's.

- **Theorem.** Suppose X and Y are two continuous random variables. Then X and Y are independent if and only if

$$f_{XY}(x,y) = f_X(x)f_Y(y) \text{ for } \text{ all } (x,y) \in \mathbb{R}^2,$$

where $f_{XY}(x,y)$, $f_X(x)$, $f_Y(y)$ are the joint and marginal PDF's.

4.4 Independence

- **Theorem (Factorization):** The two random variables X and Y are independent if and only if the joint pmf/pdf can be written as

$$f_{XY}(x, y) = g(x)h(y), \quad \text{for all } -\infty < x, y < \infty.$$

- **Remark:** $g(x)$ and $h(y)$ are not necessary PDF/PMF.

- **Definition.** The random variables $X_1, X_2, ..., X_n$ are mutually independent if the joint CDF is equal to the product of their marginal CDF's, namely,

$$F_{X^n}(x^n) = \Pi_{i=1}^n F_{X_i}(x_i) \text{ for all } -\infty < x_1, ..., x_n < \infty.$$

where $X^n = (X_1, ..., X_n)'$ and $x^n = (x_1, ..., x_n)'$.

- **Example:** Let the experiment consist of three tosses of a coin, and let X="number of heads in all three tosses" and Y="number of tails in the last two tosses". The joint and marginal distributions of (X, Y) are

(X,Y)	0	1	2	X
0	0	0	$\frac{1}{8}$	$\frac{1}{8}$
1	0	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
2	$\frac{1}{8}$	$\frac{2}{8}$	0	$\frac{3}{8}$
3	$\frac{1}{8}$	0	0	$\frac{1}{8}$
Y	$\frac{2}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	

$$P(X=3, Y=2) = 0 \neq P(X=3)P(Y=2).$$

X and Y are not independent.

4.4 Independence

- **Example:** Consider the case of three tosses of a coin. Let U be "the number of heads in the first two tosses", and let V be "the number of tails in the last toss". The sample points are

S	U	V	S	U	V
HHH	2	0	HTT	1	1
HHT	2	1	THT	1	1
HTH	1	0	TTH	0	0
THH	1	0	TTT	0	1

- – The joint and marginal probability distributions are

(U,V)	0	1	U
0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
1	$\frac{2}{8}$	$\frac{2}{8}$	$\frac{1}{2}$
2	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
V	$\frac{1}{2}$	$\frac{1}{2}$	

  – U and V are independent.

- **Example:** The joint density of random variables X and Y is

$$f(x,y) = \begin{cases} cx^n e^{-\alpha x - \beta y} & x > 0,\ y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent?

- **Example:** The joint density of random variables X and Y is

$$f(x,y) = \begin{cases} cx^n e^{-\alpha x - \beta y} & x > 0,\ y > 0,\ x + y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent?

- **Example:** A man makes two attempts at some goal. His performance X at the first attempt, measured on the scale from 0 to 1, is a random variable with density $f_X(x) = 12x^2(1-x)$. His performance Y at the second attempt is independent of X; its density is $f_Y(y) = 6y(1-y)$, $0 \le y \le 1$. What is the probability that the man exceeds level 0.75 in the better of the two attempts?

- **Solution:**

  –

$$f_{XY}(x,y) = \begin{cases} 12x^2(1-x)6y(1-y) & 0 \le x \le 1,\, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

4.4 Independence

- –

$$\begin{aligned} P(\{X > 0.75\} \cup \{Y > 0.75\}) &= 1 - P(\{X \le 0.75\} \cap \{Y \le 0.75\}) \\ &= 1 - \int_0^{0.75} \int_0^{0.75} f_{XY}(x, y)\, dxdy \\ &= 1 - (4x^3 - 3x^4)|_0^{0.75} (3y^2 - 2y^3)|_0^{0.75} \\ &= 0.3771. \end{aligned}$$

4.4 Independence

- **Theorem:** Suppose two c.r.v. X and Y are independent, then

$$f_{X|Y}(x \mid y) = f_X(x), \quad \text{for all } x \text{ and } y,$$

and

$$f_{Y|X}(y \mid x) = f_Y(y), \quad \text{for all } x \text{ and } y.$$

- **Theorem:** Suppose X and Y are independent, and $U = g(X)$ and $V = h(Y)$ for some Borel-measurable functions $g(\cdot)$ and $h(\cdot)$. Then U and V are independent.

- **Remark:** When $g(\cdot)$ and $h(\cdot)$ are 1-1 function such that their inverse functions exist, independence of U and V also implies independence of X and Y.

- **Definition: Jacobian Matrix and Jacobian.** Consider the bivariate transformation: $U = g_1(X, Y)$, $V = g_2(X, Y)$, where g_1, g_2 are continuously differentiable with respect to x, y. Then the 2×2 matrix

$$J_{UV}(x, y) = \begin{pmatrix} \frac{\partial g_1(X,Y)}{\partial x} & \frac{\partial g_1(X,Y)}{\partial y} \\ \frac{\partial g_2(X,Y)}{\partial x} & \frac{\partial g_2(X,Y)}{\partial y} \end{pmatrix}$$

is called the *Jacobian matrix* of (U, V). Its determinant is called the *Jacobian.*

4.5 Bivariate Transformation

- **Definition: Inverse Function.** Suppose A and B are subsets of $\mathbb{R}^2$. Suppose transformation $(U, V) = \big(g_1(X, Y), g_2(X, Y)\big)$ is continuously differentiable with the determinant of $J_{UV}(x, y)$ not zero on A. Then the following functions h_1, h_2 exist: $(X, Y) = \big(h_1(U, V), h_2(U, V)\big)$, where $(h_1, h_2) : B \to A$ are continuously differentiable on B such that

$$h_1\big(g_1(x, y), g_2(x, y)\big) = x,$$
$$h_2\big(g_1(x, y), g_2(x, y)\big) = y.$$

We call (h_1, h_2) the *inverse functions* of (g_1, g_2).

- **Theorem:** The Jacobian matrix of (X, Y) satisfies

$$J_{XY}(u, v) = \begin{pmatrix} \frac{\partial h_1(u,v)}{\partial u} & \frac{\partial h_1(u,v)}{\partial v} \\ \frac{\partial h_2(u,v)}{\partial u} & \frac{\partial h_2(u,v)}{\partial v} \end{pmatrix} = J_{UV}(x, y)^{-1};$$

$$det[J_{XY}(u, v)] = 1/det[J_{UV}(h_1(u, v), h_2(u, v))].$$

4.5 Bivariate Transformation

- **Theorem: Bivariate Transformation.** Let (X, Y) be a bivariate continuous random variable with PDF $f_{XY}(x, y)$, and and let $\Omega_{XY} = \{(x, y) \in \mathbb{R}^2 : f_{XY}(x, y) > 0\}$ be the support of (X, Y). Define $U = g_1(X, Y)$, $V = g_2(X, Y)$, where $(g_1, g_2) : \Omega_{XY} \to \mathbb{R}^2$ is a 1-1 and continuously differentiable function on Ω_{XY}, with $det[J_{UV}(x, y)] \neq 0$ for all $(x, y) \in \Omega_{XY}$. Let (h_1, h_2) be the inverse function of (g_1, g_2), that is $X = h_1(U, V)$, $Y = h_2(U, V)$. Then the joint PDF of (U, V) is given by

$$
\begin{aligned}
f_{UV}(u, v) &= \begin{cases} f_{XY}(h_1(u, v), h_2(u, v))|det[J_{XY}(u, v)]| & \text{for all } (u, v) \in \Omega_{UV}, \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} \frac{f_{XY}(h_1(u,v), h_2(u,v))}{|det[J_{U,V}(h_1(u,v), h_2(u,v))]|} & \text{for all } (u, v) \in \Omega_{UV}, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

Where $\Omega_{UV} = \{(u, v) \in \mathbb{R}^2 : u = g_1(x, y), v = g_2(x, y) \text{for all} (x, y) \in \Omega_{XY}.\}$

- **Example: Sum of Random Variables.** Assume (X, Y) have joint pdf $f_{XY}(x, y)$, find the pdf of $U = X + Y$.

- **Solution:**

  – Let $V = Y$, then $|det[J_{UV}]| = 1$.

  – The pdf of (U, V) is

$$f_{UV}(u, v) = f_{XY}(u - v, v).$$

  – Therefore, the marginal pdf of U is

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u - v, v) dv.$$

- **Example: Sum of Two Uniform Random Variables.** Let the joint density of (X, Y) be

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

  Find the pdf of $U = X + Y$.

- **Solution:**

  – For $0 < u \leq 1$,

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u-v, v)\, dv = \int_0^u 1\, dv = u.$$

  – For $1 < u < 2$,

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u-v, v)\, dv = \int_{u-1}^1 1\, dv = 2 - u.$$

- **Example: Sum of Exponential Random Variables.** Let the joint density of (X, Y) be

$$f_{XY}(x, y) = \begin{cases} \frac{1}{\beta^2} e^{-(x+y)/\beta} & x \geq 0,\ y \geq 0, \\ 0 & \text{otherwise.} \end{cases}.$$

Find the pdf of $U = X + Y$.

- **Solution:** For $u > 0$,

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{\infty} f_{XY}(u - v, v)\, dv \\ &= \int_0^u \frac{1}{\beta^2} e^{-u/\beta}\, dv \\ &= \frac{1}{\beta^2}\, u e^{-u/\beta}. \end{aligned}$$

$U = X + Y$ follows a Gamma(2, β) distribution.

- **Example: Product of Two Random Variables.** Assume (X, Y) have joint pdf $f_{XY}(x, y)$, find the pdf of $U = XY$.

- **Solution:**

  – Let $V = X$. Then $X = V$, $Y = U/V$. The Jacobian of this transformation is

$$det[J_{XY}(u, v)] = \begin{vmatrix} 0 & 1 \\ \frac{1}{v} & -\frac{u}{v^2} \end{vmatrix} = -\frac{1}{v}.$$

  – The joint pdf of (U, V) is

$$f_{UV}(u, v) = \frac{1}{|v|} f_{XY}(v, u/v).$$

  – The marginal pdf of U is

$$f_U(u) = \int_{-\infty}^{\infty} \frac{1}{|v|} f_{XY}(v, u/v) dv.$$

- **Example:** X and Y are independent following $N(0, \sigma^2)$. Let $U = X^2 + Y^2$, $V = X/\sqrt{X^2 + Y^2}$. (a) Find $f_{UV}(u, v)$. (b) Show that U and V are independent.

- **Solution:**

  – Let $W = Y^2$. Then

  $$f_W(w) = \frac{1}{\sqrt{2\pi w}\sigma} e^{-\frac{w}{2\sigma^2}}$$

  for $w \geq 0$.

  – $(U, V) = (X^2 + W, X/\sqrt{X^2 + W})$ is a 1-1 mapping from $(-\infty, \infty) \times [0, \infty)$ to $[0, \infty) \times [-1, 1]$. Its inverse function is $X = \sqrt{U}V$, $W = U - UV^2$.

- – The joint pdf of (U, V) is

$$\begin{aligned} f_{UV}(u,v) &= f_{XW}(x,w) \times |det[J_{XW}(u,v)]| \\ &= \frac{1}{2\pi\sigma^2\sqrt{w}}\, e^{-\frac{x^2+w}{2\sigma^2}} \times \sqrt{u} \\ &= \frac{1}{2\pi\sigma^2\sqrt{1-v^2}}\, e^{-\frac{u}{2\sigma^2}}. \end{aligned}$$

  – U and V are independent.

- **Theorem.** Suppose $U = g_1(X)$ and $V = g_2(Y)$ are some continuously differentiable 1-1 measurable functions. Then X and Y are independent if and only if U and V are independent.

- **Definition: Bivariate Normal Distribution.** (X, Y) are jointly normally distributed, denoted as $BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, where $|\rho| \leq 1$, if their joint PDF

$$\begin{aligned} f_{XY}(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times exp\{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2 \\ &+ (\frac{y-\mu_2}{\sigma_2})^2 - 2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2})]\}. \end{aligned}$$

When $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) = (0, 0, 1, 1, 0)$, it is called the standard bivariate normal distribution.

- **Remark.** An alternative representation of $f_{XY}(x, y)$ is

$$f_{XY}(x, y) = \frac{1}{\sqrt{(2\pi)^2 det(\Sigma)}} \exp\Big[-\frac{1}{2}(z-\mu)'\Sigma^{-1}(z-\mu)\Big],$$

where $z = (x, y)'$, $\mu = (\mu_1, \mu_2)'$, and

$$\begin{aligned} \Sigma &= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}. \end{aligned}$$

- **Marginal PDF of X and Y.**

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}} e^{\frac{(y-\mu)^2}{2\sigma_2^2(1-\rho^2)}} dy = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}.$$

- **Conditional PDF.** With $f_X(x)$, the conditional PDF of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}} e^{-\frac{(y-\mu)^2}{2\sigma_2^2(1-\rho^2)}},$$

where, as before, $\mu = \mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1)$. It is a normal distribution, $N(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1-\rho^2))$. Similarly, the conditional distribution of X given $Y = y$ is $N(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y - \mu_2), \sigma_1^2(1-\rho^2))$.

- **Multivariate Normal Distribution.** The random vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ follow a joint normal distribution if their joint pdf

$$f_X(\boldsymbol{x}) = \frac{1}{\sqrt{det[2\pi\Sigma]}} exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\},$$

where $\boldsymbol{x} = (x_1, \cdots, x_n)'$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)'$, and Σ is a $n \times n$ symmetric and positive definite matrix.

- **Theorem:** The MGF of multivariate normal distribution $N(\boldsymbol{\mu}, \Sigma)$ is

$$M_X(t_1, \cdots, t_n) = exp\{\boldsymbol{\mu}'\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\},$$

where $\boldsymbol{t} = (t_1, \cdots, t_n)'$.

- **Proof.**

  –

$$
\begin{aligned}
M_X(\boldsymbol{t}) &= \int e^{\boldsymbol{t}'\boldsymbol{x}} f_X(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int \frac{1}{\sqrt{det[2\pi\Sigma]}}\, exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) + \boldsymbol{t}'\boldsymbol{x}\right\} d\boldsymbol{x} \\
&= \int \frac{1}{\sqrt{det[2\pi\Sigma]}} \\
&\quad \times exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}-\Sigma\boldsymbol{t})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}-\Sigma\boldsymbol{t}) + \boldsymbol{\mu}'\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right\} d\boldsymbol{x} \\
&= exp\left\{\boldsymbol{\mu}'\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right\}.
\end{aligned}
$$

4.6 Bivariate Normal Distribution

- **Remarks:** Suppose random variables $(X_1, \cdots, X_n)$ follow multivariate normal distribution $N(\boldsymbol{\mu}, \Sigma)$. Then
  - Marginal distributions of $(X_1, \cdots, X_n)$ are normal.
  - $E(X_i) = \mu_i$.
  - $\mathrm{Var}(X_i) = \Sigma_{i,i}$.
  - $\mathrm{Cov}(X_i, X_j) = \Sigma_{i,j}$.
  - Pairwise independence is equivalent to joint independence in multivariate normal distribution.

- **Example:** Suppose X and Y are independent following $N(0, \sigma^2)$. Let $U = X + Y$, $V = X - Y$. Find the joint pdf $f_{UV}(u, v)$.

- **Solution:**

  – The joint pdf of (U, V) is

$$
\begin{aligned}
f_{UV}(u, v) &= f_{XY}(x, y) \times |det[J_{XY}(u, v)]| \\
&= \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \times \frac{1}{2} \\
&= \frac{1}{4\pi\sigma^2} e^{-\frac{u^2+v^2}{4\sigma^2}}.
\end{aligned}
$$

  – U and V are independent.

4.6 Bivariate Normal Distribution

- **Example:** Let X and Y are random variables with joint pdf

$$f_{XY}(x,y) = \begin{cases} 2f_X(x)f_Y(y) & \text{if } xy > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \\ f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}. \end{aligned}$$

Then X and Y are normal random variables, but they are not jointly normally distributed.