Chapter 4. Multivariate Distributions

4.1 Bivariate Distributions.

For a pair r.v.s (X, Y), the Joint CDF is defined as

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

Obviously, the marginal distributions may be obtained easily from the joint distribution:

$$F_X(x) = P(X \le x) = P(X \le x, Y < \infty) = F_{X,Y}(x, \infty),$$

and $F_Y(y) = F_{X,Y}(\infty, y)$.

Covariance and correlation of X and Y:

$$\text{Cov}(X, Y) = E\{(X - EX)(Y - EY)\} = E(XY) - (EX)(EY),$$
$$\text{Corr}(X, Y) = \text{Cov}(X, Y)\big/\sqrt{\text{Var}(X)\text{Var}(Y)}.$$

Discrete bivariate distributions

If X takes discrete values $x_1, \cdots, x_m$ and Y takes discrete values $y_1, \cdots, y_n$, their joint probability function may be presented in a table:

$X \backslash Y$	y_1	y_2	$\cdots$	y_n	
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$p_{2\cdot}$
		$\cdots$	$\cdots$		
x_m	p_{m1}	p_{22}	$\cdots$	p_{mn}	$p_{m\cdot}$
	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot n}$	

where $p_{ij} = P(X = x_i, Y = y_j)$, and

$$p_{i\cdot} = P(X = x_i) = \sum_{j=1}^{n} P(X = x_i, Y = y_j) = \sum_j p_{ij},$$

$$p_{\cdot j} = P(Y = y_j) = \sum_{i=1}^{m} P(X = x_i, Y = y_j) = \sum_i p_{ij}.$$

In general, $p_{ij} \neq p_{i\cdot} \times p_{\cdot j}$. However if $p_{ij} = p_{i\cdot} \times p_{\cdot j}$ for all i and j, X and Y are *independent*, i.e.

$$P(X = x_i, Y = y_j) = P(X = x_i) \times P(Y = y_j), \quad \forall i, j.$$

For independent X and Y, $\mathrm{Cov}(X, Y) = 0$.

Example 1. Flip a fair coin two times. Let $X = 1$ if H occurs in the first flip, and 0 if T occurs in the first flip. Let $Y = 1$ if the outcomes in the two flips are the same, and 0 if the two outcomes are different. The joint probability function is

$X \backslash Y$	1	0	
1	1/4	1/4	1/2
0	1/4	1/4	1/2
	1/2	1/2	

It is easy to see that X and Y are independent, which is a bit counter-intuitive.

Continuous bivariate distribution

If the CDF $F_{X,Y}$ can be written as

$$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(u,v)dudv \qquad \text{for any } x \text{ and } y,$$

where $f_{X,Y} \geq 0$, (X,Y) has a continuous joint distribution, and $f_{X,Y}(x,y)$ is the joint PDF.

As $F_{X,Y}(\infty,\infty) = 1$, it holds that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(u,v)dudv = 1.$$

In fact,, any non-negative function satisfying this condition is a PDF. Furthermore for any subset A in R^2,

$$P\{(X,Y) \in A\} = \int_A f_{X,Y}(x,y)dxdy.$$

Also

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \int (x-EX)(y-EY)f_{X,Y}(x,y)dxdy \\ &= \int xyf_{X,Y}(x,y)dxdy - EX\,EY.\end{aligned}$$

Note that

$$F_X(x) = F_{X,Y}(x,\infty) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{X,Y}(u,v)dudv = \int_{-\infty}^{x}\left\{\int_{-\infty}^{\infty} f_{X,Y}(u,v)dv\right\}du,$$

hence the *marginal PDF* of X can be derived from the joint PDF as follows

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

Similarly, $f_y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$.

Note. Different from discrete cases, it is not always easy to work out marginal PDFs from joint PDFs, especially when PDFs are discontinuous.

When $f_{X,Y}(x, y) = f_X(x)f_y(y)$ for any x and y, X and Y are independent, as then

$$P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v)dudv = \int_{-\infty}^{y} \int_{-\infty}^{x} f_X(u)f_Y(v)dudv$$
$$= \int_{-\infty}^{x} f_X(u)du \int_{-\infty}^{y} f_Y(v)dv = P(X \leq x)P(Y \leq y),$$

and also $\text{Cov}(X, Y) = 0$.

Example 2. *Uniform distribution on unit square –* $U[0, 1]^2$.

$$f(x, y) = \begin{cases} 1 & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

This is well-defined PDF, as $f \geq 0$ and $\int\int f(x, y)dxdy = 1$. It is easy to see that X and Y are independent. Let us calculate some probabilities

$$\begin{aligned} P(X < 1/2, Y < 1/2) &= F(1/2, 1/2) = \int_{-\infty}^{1/2}\int_{-\infty}^{1/2} f_{X,Y}(x, y)dxdy \\ &= \int_0^{1/2}\int_0^{1/2} dxdy = 1/4. \end{aligned}$$

$$\begin{aligned} P(X + Y < 1) &= \int_{\{x+y<1\}} f_{X,Y}(x, y)dxdy = \int_{\{x>0,\ y>0,\ x+y<1\}} dxdy \\ &= \int_0^1 dy \int_0^{1-y} dx = \int_0^1 (1 - y)dy = 1/2. \end{aligned}$$

Example 3. Let (X, Y) have the joint PDF

$$f(x, y) = \begin{cases} x^2 + xy/3 & 0 \leq x \leq 1,\ 0 \leq y \leq 2, \\ 0 & \text{otherwsie.} \end{cases}$$

Calculate $P(0 < X < 1/2, 1/4 < Y < 3)$ and $P(X < Y)$. Are X and Y independent with each other?

$$P(0 < X < 1/2,\ 1/4 < Y < 3) = P(0 < X < 1/2,\ 1/4 < Y < 2)$$
$$= \int_{1/4}^{2} dy \int_{0}^{1/2} (x^2 + \frac{xy}{3}) dx = \int_{1/4}^{2} \frac{1+y}{24} dy = \frac{1.75}{24} + \frac{y^2}{48}\Big|_{1/4}^{2} = 0.155.$$

$$P(X < Y) = \int_0^1 dx \int_x^2 (x^2 + \frac{xy}{3}) dy = \int_0^1 (\frac{2}{3}x + 2x^2 - \frac{7}{6}x^3) dx = 17/24 = 0.708.$$

$$f_X(x) = \int_0^2 (x^2 + \frac{xy}{3}) dy = 2x^2 + \frac{2x}{3}, \quad f_Y(y) = \int_0^1 (x^2 + \frac{xy}{3}) dx = \frac{1}{3} + \frac{y}{6}.$$

Both $f_X(x)$ and $f_Y(y)$ are well-defined PDFs.

But $f(x, y) \neq f_X(x) f_Y(y)$, hence they are not independent.

4.2 Conditional Distributions

If X and Y are not independent, knowing X should be helpful in determining Y, as X may carry some information on Y. Therefore it makes sense to define the distribution of Y given, say, $X = x$. This is the concept of conditional distributions.

If both X and Y are discrete, the conditional probability function is simply a special case of conditional probabilities:

$$P(Y = y|X = x) = P(Y = y,\ X = x)\big/P(X = x).$$

However this definition does not extend to continuous r.v.s, as then $P(X = x) = 0$.

Definition (Conditional PDF). For continuous r.v.s X and Y, the conditional PDF of Y given $X = x$ is

$$f_{Y|X}(\cdot|x) = f_{X,Y}(x, \cdot)/f_X(x).$$

Remark. (i) As a function of y, $f_{Y|X}(y|x)$ is a PDF:

$$P(Y \in A|X = x) = \int_A f_{Y|X}(y|x)dy,$$

while x is treated as a constant (i.e. not a variable).

(ii) $E(Y|X = x) = \int y f_{Y|X}(y|x)dy$ is a function of x, and

$$\mathrm{Var}(Y|X = x) = \int \{y - E(Y|X = x)\}^2 f_{Y|X}(y|x)dy.$$

(iii) If X and Y are independent, $f_{Y|X}(y|x) = f_Y(y)$.

(iv) $f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x) = f_{X|Y}(x|y) f_Y(y)$, which offers alternative ways to determine the joint PDF.

(v) *$E\{E(Y|X)\} = E(Y)$ — This in fact holds for any r.v.s X and Y*. We give a proof here for continuous r.v.s only:

$$E\{E(Y|X)\} = \int \left\{ \int y f_{Y|X}(y|x) dy \right\} f_X(x) dx = \int \int y f_{X,Y}(x,y) dx dy$$
$$= \int y \left\{ \int f_{X,Y}(x,y) dx \right\} dy = \int y f_Y(y) dy = EY.$$

Example 4. Let $f_{X,Y}(x,y) = e^{-y}$ for $0 < x < y < \infty$, and 0 otherwise. Find $f_{Y|X}(y|x)$, $f_{X|Y}(x|y)$ and $\text{Cov}(X,Y)$.

We need to find $f_X(x)$, $f_Y(y)$ first:

$$f_X(x) = \int f_{X,Y}(x,y)dy = \int_x^{\infty} e^{-y}dy = e^{-x} \qquad x \in (0,\infty),$$

$$f_Y(y) = \int f_{X,Y}(x,y)dx = \int_0^{y} e^{-y}dx = ye^{-y} \qquad y \in (0,\infty).$$

Hence

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = e^{-(y-x)} \qquad y \in (x,\infty),$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = 1/y \qquad x \in (0,y).$$

Note that given $Y = y$, $X \sim U(0,\ y)$, i.e. the uniform distribution on $(0,\ y)$.

To find Cov(X, Y), we compute $EX,\ EY$ and $E(XY)$ first.

$$
\begin{aligned}
EX &= \int x f_X(x) dx = \int_0^\infty x e^{-x} dx = -e^{-x}(1+x)\Big|_0^\infty = 1, \\
EY &= \int y f_Y(y) dy = \int_0^\infty y^2 e^{-y} dy = -y^2 e^{-y}\Big|_0^\infty + 2\int_0^\infty y e^{-y} dy = 2, \\
E(XY) &= \int xy f_{X,Y}(x, y) dx dy = \int_0^\infty dy \int_0^y xy e^{-y} dx = \frac{1}{2}\int_0^\infty y^3 e^{-y} dy \\
&= -\frac{1}{2} y^3 e^{-y}\Big|_0^\infty + \frac{3}{2}\int_0^\infty y^2 e^{-y} dy = 3.
\end{aligned}
$$

Hence Cov$(X, Y) = E(XY) - (EX)(EY) = 3 - 2 = 1$.

4.3 Multivariate Distributions

Let $\mathbf{X} = (X_1, \cdots, X_n)'$ be a random vector (r.v.) consisting of n r.v.s. The joint CDF is defined as

$$F(x_1, \cdots, x_n) \equiv F_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = P(X_1 \leq x_1, \cdots, X_n \leq x_n).$$

If X is continuous, its PDF f satisfies

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

In general, the PDF admits the factorisation

$$f(x_1, \cdots, x_n) = f(x_1) f(x_2|x_1) f(x_3|x_1, x_2) \cdots f(x_n|x_1, \cdots, x_{n-1}),$$

where $f(x_j|x_1, \cdots, x_{j-1})$ denotes the conditional PDF of X_j given $X_1 = x_1, \cdots, X_{j-1} = x_j$.

However, when $X_1, \cdots, X_n$ are independent,

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

IID Samples. If $X_1, \cdots, X_n$ are independent and each has the same CDF F, we say that $X_1, \cdots, X_n$ are IID (independent and identically distributed) and write

$$X_1, \cdots, X_n \sim_{\text{iid}} F.$$

We also call $X_1, \cdots, X_n$ *a sample* or *a random sample.*

4.3 Two important multivariate distributions

Multinomial Distribution Multinomial$(n, p_1, \cdots, p_k)$ — an extension of Bin(n, p).

Suppose we threw a k-sided die n times, record X_i as the number of times ended with the i-th side, $i = 1, \cdots, k$. Then

$$(X_1, \cdots, X_k) \sim \text{Multinomial}(n, p_1, \cdots, p_k),$$

where p_i is the probability of the event that the i-th side occurs in one threw. Obviously $p_i \geq 0$ and $\sum_i p_i = 1$.

We may immediately make the following observation from the above definition.

(i) $X_1 + \cdots + X_k \equiv n$, therefore $X_1, \cdots, X_n$ are not independent.

(ii) $X_i \sim \text{Bin}(n, p_i)$, hence $EX_i = np_i$ and $\text{Var}(X_i) = np_i(1 - p_i)$.

The joint probability function for Multinomial$(n, p_1, \cdots, p_k)$:

For any $j_1, \cdots, j_k \geq 0$ and $j_1 + \cdots + j_k = n$,

$$P(X_1 = j_1, \cdots, X_k = j_k) = \frac{n!}{j_1! \cdots j_k!} p_1^{j_1} \cdots p_k^{j_k}.$$

Multivariate Normal Distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: a k-variable r.v. $\mathbf{X} = (X_1, \cdots, X_k)'$ is normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if its PDF is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\} \qquad \mathbf{x} \in R^k,$$

where $\boldsymbol{\mu}$ is k-vector, and $\boldsymbol{\Sigma}$ is a $k \times k$ positive-definite matrix.

Some properties of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: Let $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_k)'$ and $\boldsymbol{\Sigma} \equiv (\sigma_{ij})$, then

(i) $E\mathbf{X} = \boldsymbol{\mu}$, and the covariance matrix

$$\mathrm{Cov}(\mathbf{X}) = E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\} = \boldsymbol{\Sigma},$$

and

$$\sigma_{ij} = \mathrm{Cov}(X_i, X_j) = E\{(X_i - \mu_i)(X_j - \mu_j)\}.$$

(ii) When $\sigma_{ij} = 0$ for all $i \neq j$, i.e. the components of $\mathbf{X}$ are *uncorrelated*, $\mathbf{\Sigma} = \text{diag}(\sigma_{11}, \cdots, \sigma_{kk})$, $|\mathbf{\Sigma}| = \prod_i \sigma_{ii}$. Hence the PDF admits a simple form

$$f(\mathbf{x}) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma_{ii}}} \exp\{-\frac{1}{2\sigma_{ii}}(x_i - \mu_i)^2\}.$$

Thus $X_1, \cdots, X_n$ are independent when $\sigma_{ij} = 0$ for all $i \neq j$.

(iii) Let $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, where $\mathbf{A}$ is a constant matrix and $\mathbf{b}$ is a constant vector. Then $\mathbf{Y} \sim N(\mathbf{A}\mu + \mathbf{b}, \ \mathbf{A\Sigma A'})$.

(iv) $X_i \sim N(\mu_i, \sigma_{ii})$. For any constant k-vector $\mathbf{a}$, $\mathbf{a'X}$ is a scale r.v. and $\mathbf{a'X} \sim N(\mathbf{a'}\boldsymbol{\mu}, \ \mathbf{a'\Sigma a})$.

(v). **Standard Normal Distribution**: $N(0, \mathbf{I_k})$, where $\mathbf{I}_k$ is the $k \times k$ identity matrix.

Example 5. Let X_1, X_2, X_3 be jointly normal with the common mean 0, variance 1 and

$$\mathrm{Corr}(X_i, X_j) = 0.5, \qquad 1 \le i \ne j \le 3.$$

Find the probability $P(|X_1| + |X_2| + |X_3| \le 2)$.

It is difficult to calculate this probability by the integration of the joint PDF. We provide an estimate by simulation. We solve a general problem first.

Let $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{X}$ has p component. For any set $A \subset R^p$, we may estimate the probability $P(\mathbf{X} \in A)$ by the relative frequency

$$\#\{1 \le i \le n : \mathbf{X}_i \in A\} \big/ n,$$

where n is a large integer, and $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are n vectors generated independently from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Note

$$\mathbf{X} = \mu + \mathbf{\Sigma}^{1/2}\mathbf{Z},$$

where $\mathbf{Z} \sim N(0, \mathbf{I_p})$ is standard normal, and $\mathbf{\Sigma}^{1/2} \geq 0$ and $\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma}$. We generate $\mathbf{Z}$ by `rnorm(p)`, and apply the above linear transformation to obtain $\mathbf{X}$.

$\mathbf{\Sigma}^{1/2}$ may be obtained by an eigenanalysis for $\mathbf{\Sigma}$ using *R*-function `eigen`. Since $\mathbf{\Sigma} \geq 0$, it holds that

$$\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}',$$

where $\mathbf{\Gamma}$ is an orthogonal matrix (i.e. $\mathbf{\Gamma}'\mathbf{\Gamma} = \mathbf{I}_p$), $\mathbf{\Lambda} = \text{diag}(\lambda_1, \cdots, \lambda_p)$ is a diagonal matrix. Then

$$\mathbf{\Sigma}^{1/2} = \mathbf{\Gamma}\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}', \quad \text{where} \;\; \mathbf{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_p}).$$

The *R* function `rMNorm` below generate random vectors from $N(\boldsymbol{\mu}, \mathbf{\Sigma})$.

```r
rMNorm <- function(n, p, mu, Sigma) {
   # generate n p-vectors from N(mu, Sigma)
   # mu is p-vector of mean, Sigma >=0 is pxp matrix
t <- eigen(Sigma, symmetric=T) # eigenanalysis for Sigma
ev <- sqrt(t$values) # square-roots of the eigenvalues
G <- as.matrix(t$vectors) # line up eigenvectors into a matrix G
D <- G*0; for(i in 1:p) D[i,i] <- ev[i]; # D is diagonal matrix
P <- G%*%D%*%t(G) # P=GDG' is the required transformation matrix
Z <- matrix(rnorm(n*p), byrow=T, ncol=p)
  # Z is nxp matrix with elements drawn from N(0,1)
Z <- Z%*%P # Now each row of Z is N(0, Sigma)
X <- matrix(rep(mu, n), byrow=T, ncol=p) + Z
   # each row of X is N(mu, Sigma)
}
```

This function is saved in the file ‘rMNorm.r’. We may use it to perform the required task:

```r
source("rMNorm.r")
```

```r
mu <- c(0, 0, 0)
Sigma <- matrix(c(1,0.5,0.5,0.5,1,0.5,0.5,0.5,1), byrow=T, ncol=3)
X <- rMNorm(20000, 3, mu, Sigma)
dim(X)  # check the size of X
t <- abs(X[,1]) + abs(X[,2]) + abs(X[,3])
cat("Estimated probability:", length(t[t<=2])/20000, "\n")
```

It returned the value:

```
Estimated probability:  0.446
```

I repeated it a few more times and obtained the estimates 0.439, 0.445, 0.441 etc.

4.4 Transformations of random variables

Let a random vector $\mathbf{X}$ have PDF $f_{\mathbf{X}}$. We are interested in the distribution of a scalar function of $\mathbf{X}$, say, $Y = r(\mathbf{X})$. We introduce a general procedure first.

Three steps to find the PDF of $Y = r(\mathbf{X})$:

(i) For each y, find the set $A_y = \{\mathbf{x} : \ r(\mathbf{x}) \leq y\}$
(ii) Find the CDF

$$F_Y(y) = P(Y \leq y) = P\{r(\mathbf{X}) \leq y\} = \int_{A_y} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

(iii) $f_Y(y) = \frac{d}{dy} F_Y(y)$.

Example 6. Let $X \sim f_X(x)$ (X is a scalar). Find the PDF of $Y = e^X$.

$A_y = \{x : e^x \leq y\} = \{x : x \leq \log y\}$. Hence

$$F_Y(y) = P(Y \leq y) = P\{e^X \leq y\} = P(X \leq \log y) = F_X(\log y).$$

Hence

$$f_Y(y) = \frac{d}{dy} F_X(\log y) = f_X(\log y) \frac{d \log y}{dy} = y^{-1} f_X(\log y).$$

Note that $y = e^x$ and $\log y = x$, $\frac{dy}{dx} = e^x = y$. The above result can be written as

$$f_Y(y) = f_X(x) \Big/ \frac{dy}{dx}, \quad \text{or} \quad f_Y(y)dy = f_X(x)dx.$$

For 1-1 transformation $Y = r(X)$ (i.e. the inverse function $X = r^{-1}(Y)$ is uniquely defined), it holds that

$$f_Y(y) = f_X(x)/|r'(x)| = f_X(x)\left|\frac{dx}{dy}\right|.$$

Note. You should replace all x in the above by $x = r^{-1}(y)$.

Example 7. Let $X \sim U(-1, 3)$. Find the PDF of $Y = X^2$. Now this is not a `1-1` transformation. We have to use the general 3-step procedure.

Note that Y takes values in $(0, 9)$. Consider two cases:

(i) For $y \in (0, 1)$, $A_y = (-\sqrt{y}, \sqrt{y})$, $F_y(y) = \int_{A_y} f_X(x)dx = 0.5\sqrt{y}$. Hence $f_Y(y) = F'_Y(y) = 0.25/\sqrt{y}$.

(ii) For $y \in [1, 9)$, $A_y = (-1, \sqrt{y})$, $F_y(y) = \int_{A_y} f_X(x)dx = 0.25(\sqrt{y} + 1)$. Hence $f_Y(y) = F'_Y(y) = 0.125/\sqrt{y}$.

Collectively we have

$$f_Y(y) = \begin{cases} 0.25/\sqrt{y} & 0 < y < 1 \\ 0.125/\sqrt{y} & 1 \le y < 9 \\ 0 & \text{otherwise.} \end{cases}$$