Chapter 4: Multiple Random Variables

We study the joint distribution of more than two random variables, called a *random vector*, such that $(X, Y), (X, Y, Z)$, $(X_1, \cdots, X_n)$, and the distribution of their functions like $X + Y$, XYZ, or $X_1 + X_2 + \cdots + X_n$.

1 Bivariate Random Variables

Assume both X and Y are random. We treat (X, Y) as a two-dimensional random vector and study their relationship.

1.1 Discrete Case

Assume that both X and Y are discrete random variables, with the sample space $\mathcal{X}$ and $\mathcal{Y}$ respectively.

Joint pmf:

$$f_{X,Y}(x, y) = P(X = x, Y = y), \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}.$$

Properties:

- $f_{X,Y}(x, y) \geq 0$;
- $\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} f_{X,Y}(x, y) = 1$.

The probability of a set A is given by

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

Marginal pmf: If the joint distribution of (X, Y) is known, their marginal pmf are

$$f_X(x) = P(X = x) = \sum_{y\in\mathcal{Y}} f_{X,Y}(x, y).$$

$$f_Y(y) = P(Y = y) = \sum_{x\in\mathcal{X}} f_{X,Y}(x, y)$$

Example 1 Two fair dice thrown. Let X=maximum, Y=sum.

Possible values:

X: 1, 2, 3, 4, 5, 6.

Y: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

Can write the probabilities in a table.

Remark:

- Joint distribution determines the marginal distribution.
- Marginals do not determine the joint distribution.

Example: Define the joint pmf by

$$f(0,0) = f(0,1) = \frac{1}{6}; \quad f(1,0) = f(1,1) = \frac{1}{3}; \quad f(x,y) = 0, \text{ otherwise.}$$

Consider another joint pmf by

$$f(0,0) = \frac{1}{12}; \quad f(1,0) = \frac{5}{12}; \quad f(0,1) = f(1,1) = \frac{3}{12}; \quad f(x,y) = 0, \text{ otherwise.}$$

They share the same marginal distributions, but not the same joint distribution!

1.2 Continuous Case

Assume that both X and Y are continuous random variables.

Joint pdf: A function $f_{X,Y}(x,y)$ is called a *joint probability density function* of (X,Y) if

$$P((X,Y) \in A) = \int \int_{(x,y)\in A} f(x,y)dxdy, \quad \forall A \in R^2.$$

The joint pdf satisfies:

- $f(x,y) \geq 0$,
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$.

Joint cdf: The joint distribution of (X,Y) can be completely described with their joint cdf

$$F(x,y) = P(X \leq x, Y \leq y), \quad \forall (x,y) \in R^2.$$

Relationship between joint pdf and joint cdf: if F is differentiable with respect to both x and y, then

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)dudv,$$

$$\frac{\partial^2}{\partial x \partial y} F(x,y) = f(x,y).$$

Marginal pdf: If the joint pdf of (X,Y) is given, the marginal pdfs of X and Y are given by

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

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

Review on Double Integration:

Compute $\int\int_{\mathcal{D}} f(x,y)dxdy$ using *iterated integrals*

Example. Check whether the following function a valid pdf

$$f(x, y) = ye^{-(x+y)} I\{0 < x < y\}.$$

Example. Show that $f(x, y) = 2I\{0 \le x \le y \le 1\}$ is a valid pdf.

Example. Assume $f(x, y) = e^{-y} I\{0 < x < y\}$.

(i) Show that $f(x, y)$ is a valid pdf.

(ii) What is the marginal distribution of X?

(iii) What is the marginal distribution of Y?

(iv) Compute $P(X + Y \geq 1)$.

1.3 Expectation of Functions of Random Vector

Assume g is a real-valued function of two random variables $g(X, Y)$.

If X and Y are both discrete, then

$$E(g(X,Y)) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} g(x,y) f_{X,Y}(x,y).$$

If X and Y are both continuous, then

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

Properties:

- $E(aX + bY + c) = aE(X) + bE(Y) + c$.
- $E(ag_1(X,Y) + bg_2(X,Y) + c) = aE(g_1(X,Y)) + bE(g_2(X,Y)) + c$.
- In general, $E(XY) \neq E(X)E(Y)$ unless X and Y are independent.

Joint mgf: $M_{X,Y}(t,s) = E(e^{tX+sY})$. Note

$$M_{X,Y}(t,0) = M_X(t), \quad M_{X,Y}(0,s) = M_Y(s),$$

$$\frac{\partial^k M_{X,Y}(t,s)}{\partial t^k}|_{(0,0)} = E(X^k), \quad \frac{\partial^k M_{X,Y}(t,s)}{\partial s^k}|_{(0,0)} = E(Y^k).$$

Discrete Example: Two fair dice thrown. Let X=maximum, Y=sum. Compute $E(XY)$.

Ex. $f(x, y) = e^{-y} I\{0 < x < y\}$. Compute $E(X), E(Y), E(XY), M_{X,Y}(t, s)$.

2 Conditional Distributions

Oftentimes (X, Y) are related. For example, let X be a person's height and Y be a person's weight. Knowledge about the value of X gives us some information about the value of Y. It turns out conditional probabilities of Y given knowledge of X can be computed from their joint distribution $f_{X,Y}(x, y)$.

2.1 Discrete Case

Assume both X and Y are discrete. For any x such that $P(X = x) > 0$, the conditional pmf of Y given $X = x$ is defined as

$$f_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}, \quad \forall y \in \mathcal{Y}.$$

We can define $f_{X|Y}(x|y)$ similarly.

REMARK: The function $f(y|x)$ is indeed a pmf, since for any fixed x it satisfies

- $f_{Y|X}(y|x) \geq 0$ for any y.
- $\sum_y f_{Y|X}(y|x) = 1$.

Proof:

EXAMPLE. The two dice example, X=maximum, Y=sum.

$f_{Y|X}(y|3)$.

$f_{X|Y}(x|7)$.

2.2 Continuous Case

Assume both X and Y are continuous. For any x such that $f_X(x) > 0$, the conditional pdf of Y given $X = x$ is defined as

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \quad \forall y \in \mathcal{Y}.$$

We can define $f_{X|Y}(x|y)$ similarly.

REMARK: The function $f_{Y|X}(y|x)$ is indeed a pdf, since for any fixed x it satisfies

- $f_{Y|X}(y|x) \geq 0$ for any y.
- $\int_y f_{Y|X}(y|x)dy = 1$.

EXAMPLE. Assume $f(x,y) = e^{-y}I\{0 < x < y\}$. Compute $f_{Y|X}(y|x)$.

2.3 Conditional Mean and Variance

For discrete random variables:

$$\begin{aligned} E(Y|X=x) &= \sum_y y f_{Y|X}(y|x), \\ \text{Var}(Y|X=x) &= \sum_y \{y - E(Y|X=x)\}^2 f_{Y|X}(y|x). \end{aligned}$$

For continuous random variables:

$$\begin{aligned} E(Y|X=x) &= \int y f_{Y|X}(y|x) dy, \\ \text{Var}(Y|X=x) &= \int \{y - E(Y|X=x)\}^2 f_{Y|X}(y|x) dy. \end{aligned}$$

REMARK 1: As before, we have

$$\text{Var}(Y|X=x) = E(Y^2|X=x) - \{E(Y|X=x)\}^2.$$

EXAMPLE. Two dice example, X=max, Y=sum. Compute $E(Y|X=3)$.

EX. $f(x,y) = e^{-y} I\{0 < x < y\}$. Find $\text{E}(Y|X=x)$ and $\text{Var}(Y|X=x)$.

Remark 2: Note $E(Y|X = x)$ is a function of x. Therefore, $E(Y|X)$ is a random variable as a function of X.

- $E(g(X)|X) = g(X)$.

Theorem:

- Conditional Expectation Identity

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

- Conditional Variance Identity

$$\mathrm{Var}(Y) = E(\mathrm{Var}(Y|X)) + \mathrm{Var}(E(Y|X)).$$

Remark 3:

- $E(g(X, Y)) = E(E(g(X, Y)|Y)) = E(E(g(X, Y)|X))$.
- Conditional expectation as projection:

$$E(Y - E(Y|X))^2 \leq E(Y - g(X))^2, \quad \forall\, g \ \text{ function}$$

So $E(Y|X)$ is "closest" (in above sense) to Y among all the functions of X.

3 Independence

Def: Let (X, Y) be a bivariate random vector with joint pdf/pmf $f_{X,Y}(x, y)$, Then X and Y are called *independent* random variables if for every $x, y \in R$

$$f(x, y) = f_X(x) f_Y(y).$$

EXAMPLE. Consider the discrete bivariate random vector (X, Y) with joint pmf given by

$$f(10, 1) = f(20, 1) = f(20, 2) = \frac{1}{10}, \quad f(10, 2) = f(10, 3) = \frac{1}{5}, \quad f(20, 3) = \frac{3}{10}.$$

Are X and Y are independent?

Lemma. Let (X, Y) be a bivariate random vector with joint pdf or pmf $f_{X,Y}(x, y)$, Then X and Y are independent random variables **if and only if** there exist functions $g(x)$ and $h(y)$ such that for every $x \in R$ and $y \in R$,

$$f(x, y) = g(x) h(y).$$

In other words, the joint pmf/pdf is factorizable. (We do not need to compute marginal pdfs).

EXAMPLE. Consider the continuous bivariate random vector (X, Y) with joint pdf given by

$$f(x, y) = \frac{1}{384} x^2 y^4 e^{-y-(x/2)}, \quad x > 0, \ y > 0.$$

Are X and Y are independent?

Theorem: If X and Y are independent, then

(i) $E(Y|X) = E(Y)$.

(ii) The events $\{X \in A\}$ and $\{Y \in B\}$ are independent.

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \quad \forall A \subset R, B \subset R.$$

(iii)

$$E(g(X)h(Y)) = E(g(X))E(h(Y).$$

In particular, $E(XY) = E(X)E(Y)$.

(iv) In addition, we have $M_{X,Y}(t,s) = E(e^{tX+sY}) = M_X(t)M_Y(s)$. And

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = M_X(t)M_Y(t).$$

If it is easy to identify the right-hand side as the MGF of some standard distribution, then the sum of two independent variables is easy to find.

Example 1. $X \sim \text{Bin}(n_1, p)$, $Y \sim \text{Bin}(n_2, p)$, and they are independent.

Example 2. $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$, and they are independent.

Example 3. $X \sim \mathrm{NB}(r_1, p)$, $Y \sim \mathrm{NB}(r_2, p)$, and they are independent.

Example 4. $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$, and they are independent.

Example 5. $X \sim \mathrm{Gamma}(\alpha_1, \beta)$, $Y \sim \mathrm{Gamma}(\alpha_2, \beta)$, and they are independent.

4 Bivariate transformation

In this section, we only consider continuous bivariate random vector (X, Y). Consider the following bivariate transformation of (X, Y):

$$U = g_1(X, Y), \quad V = g_2(X, Y).$$

4.1 Transformation for Discrete Random Variables

Assume that (X, Y) is a discrete bivariate random vector with the support $\mathcal{A}$, i.e. $P(X = x, Y = y) > 0$ on $\mathcal{A}$. Consider the bivariate transformation

$$U = g_1(X, Y), \quad V = g_2(X, Y).$$

The the support of (U, V) is

$$\mathcal{B} = \{(u, v) : u = g_1(x, y), v = g_2(x, y) \text{ for some } (x, y) \in \mathcal{A}\}.$$

For any $(u, v) \in \mathcal{B}$, define $A_{(u,v)} = \{(x, y) \in \mathcal{A} : g_1(x, y) = u, g_2(x, y) = v\}$. Then the joint pmf of (U, V) is given by

$$f_{U,V}(u, v) = P(U = u, V = v) = P((X, Y) \in A_{(u,v)}) = \sum_{(x,y)\in A_{(u,v)}} f_{X,Y}(x, y).$$

Example 1: Assume $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\theta)$, and they are independent. Find the joint pmf of $(X + Y, Y)$ and the marginal pmf of U.

4.2 One-to-One Transformation for Continuous Random Variables

Assume that g_1 and g_2 are continuous, differentiable, and one-to-one. Therefore, we can define their inverse transformations as

$$X = h_1(U, V), \quad Y = h_2(U, V).$$

Def: Jacobian matrix and determinant

$$J = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{pmatrix}.$$

is the Jacobian matrix and $\det(J)$ is the Jacobian determinant, or simply the Jacobian.

Example 1. Linear transform.

$$U = X + Y, \quad V = X - Y.$$

Example 2. Polar transform. Assume $(X, Y) \in R^2$. Consider

$$x = r\cos\theta, \quad y = r\sin\theta,$$

where $r \in (0, \infty)$ and $\theta \in (0, 2\pi)$. How to express (r, θ) in terms of (x, y)?

Theorem. If $f_{X,Y}(x,y)$ is the joint density of (X,Y), then

$$f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v))|\det(J)|.$$

Proof follows from change of variable rules for integration — omitted.

Example. Assume $X, Y \sim N(0,1)$ and they are independent. Let $U = X+Y, V = X-Y$. Find the joint and marginal distributions of (U,V).

Example. Polar transform of independent normals.

EXAMPLE. Assume $X \sim \text{Gamma}(\alpha_1, \beta)$ and $Y \sim \text{Gamma}(\alpha_2, \beta)$, and they are independent. Let $U = X + Y, V = \frac{X}{X+Y}$. Find the joint and marginal distributions of (U, V)

EXAMPLE. Assume $X, Y \sim N(0, 1)$ and they are independent. Let $U = X/Y$. Find the distribution of U.

EXAMPLE. Assume $X \sim \text{Beta}(\alpha, \beta)$ and $Y \sim \text{Beta}(\alpha + \beta, \gamma)$ and they are independent. Find the distribution of XY.

4.3 Piecewise One-to-One Transformation

Assume (X, Y) takes value from $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_k$, where $P((X, Y) \in \mathcal{A}_0) = 0$. Also $U = g_{1i}(X, Y), V = g_{2i}(X, Y)$ is one-to-one transformation from $\mathcal{A}_i$ to B, for $i = 1, \cdots, k$. Then

$$f_{U,V}(u, v) = \sum_{i=1}^{k} f_{X,Y}(h_{1i}(u, v), h_{2i}(u, v))|\det(J_i)|,$$

5 Hierarchical Mixtures.

Recall that

$$E(Y) = E(E(Y|X)), \quad \text{var}(Y) = E(\text{var}(Y|X)) + \text{var}(E(Y|X)).$$

Example. (binomial-Poisson). An insect lays a large number of eggs, each surviving with probability p. On the average, how many eggs will survive?

Let X=the number of eggs that survive

Let Y=the total number of eggs laid by the insect.

1. Describe their distributions.
2. Find the joint distribution of (X, Y).
3. Find the marginal distribution of X, $E(X)$ and $\text{Var}(X)$.

EXAMPLE: Assume $X|\Lambda \sim$ Poisson(Λ), $\Lambda \sim$ Gamma(α, β).

EXAMPLE: Assume $X|p \sim$ Binomial(n,p), $p \sim$ Beta(α, β).

EXAMPLE: binomial-Poisson-gamma (optional). Assume $X|Y \sim \text{Bin}(Y, p)$, $Y|\Lambda \sim \text{Poisson}(\Lambda)$, $\Lambda \sim \text{Gamma}(\alpha, \beta)$.

6 Covariance and Correlation.

Covariance: A measure of joint variation.

$$\text{Cov}(X, Y) = E[\{X - E(X)\}\{Y - E(Y)\}].$$

Note: the outer expectation is with respect to the joint distribution of (X, Y). And we have

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Correlation:

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

REMARK: If X and Y are independent, then $\text{cov}(X, Y) = 0$ and $\rho_{X,Y} = 0$. **But the converse is not true!**

EXAMPLE. $X \sim N(0,1)$, $Y = X^2$.

EXAMPLE. $X = X_1 + X_3$, $Y = X_2 + X_3$, where X_1, X_2, X_3 pairwise independent with common variance σ^2. Compute ρ.

EXAMPLE. $X \sim$Unif(0,1), $Z \sim$Unif$(0,\frac{1}{10})$ and they are independent. Let $Y = X + Z$. Compute ρ.

EXAMPLE. $Y = X^2 + Z$, where X, Z are independent, X symmetric about 0, Z any distribution. Compute $\text{Cov}(X, Y)$.

One Important Equation:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

One Important Inequality: Cauchy-Schwarz Inequality

$$|\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y$$

with equality iff X and Y are linearly related.

Corollary.

$$-1 \leq \rho_{X,Y} \leq 1.$$

And $|\rho_{X,Y}| = 1$ iff $Y = aX + b$ wp 1, where $a > 0$ iff $\rho_{X,Y} = 1$ and $a < 0$ iff $\rho_{X,Y} = -1$.
(Proofs can be found in the textbook and are omitted here.)

7 Bivariate Normal.

We say $(X, Y) \sim BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ if

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_1\sigma_2} \times \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}\right].$$

We will show that

$$X \sim N(\mu_1, \sigma_1^2), \quad Y \sim N(\mu_2, \sigma_2^2), \quad \rho_{X,Y} = \rho, \quad aX + bY \text{ is normal.}$$

Conditional distribution for Bivariate normal.

Suppose $(X,Y) \sim BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, then

$$X|Y=y \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\ \ \sigma_1^2(1-\rho)\right)$$
$$Y|X=x \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\ \ \sigma_2^2(1-\rho)\right)$$

8 Multivariate Distributions

Several variables $(X_1, \ldots, X_n)$.

8.1 Discrete Case

Joint pmf

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

satisfying $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \geq 0$ and $\sum_{x_1,\ldots,x_n} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = 1$.

For any subset A of R^n, we have

$$P\{(X_1,\ldots,X_n) \in A\} = \sum_{(x_1,\ldots,x_n)\in A} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n).$$

$$Eg(X_1,\ldots,X_n) = \sum \ldots \sum g(x_1,\ldots,x_n) f_{X_1,\ldots,X_n}(x_1,\ldots,x_n).$$

Marginal distribution of $(X_{i_1}, \ldots, X_{i_k})$:

$$f_{X_{i_1},\ldots,X_{i_k}}(x_{i_1},\ldots,x_{i_k}) = \sum_{\text{other indices}} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n).$$

One-dimensional marginals:

$$f_{X_i}(x_i) = \sum_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} f(x_1,\ldots,x_n).$$

Conditional distribution:

$$f_{X_{k+1},\ldots,X_n|X_1,\ldots,X_k}(x_{k+1},\ldots,x_n|x_1,\ldots,x_k) = \frac{f(x_1,\ldots,x_n)}{f_{X_1,\ldots,X_k}(x_1,\ldots,x_k)}.$$

Covariance: $\mathrm{Cov}(X_i, X_j)$ — based on pairwise distribution.

Independence: $X_1, \cdots, X_n$ are called *mutually independent random variables* if their joint is the product of marginals:

$$f_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), \quad \forall (x_1, \cdots, x_n)$$

Remark: If $X_1, \cdots, X_n$ are mutually independent, then

(1) Any pair X_i and X_j are pairwise independent.

(2) Functions $g_1(X_1), \cdots, g_n(X_n)$ are independent, and

$$E(\prod_{i=1}^{n} g_i(X_i)) = \prod_{i=1}^{n} E(g_i(X_i)).$$

(3) MGF is the product of individual MGF's.

$$M_{X_1,\cdots,X_n}(t_1, \cdots, t_n) = \prod_{i=1}^{n} M_{X_i}(t_i).$$

(4) Let $Z = X_1 + \cdots + X_n$, then the mgf of Z is

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

In particular, if $X_1, \cdots, X_n$ all have the same distribution with mgf $M_X(t)$, then

$$M_Z(t) = [M_X(t)]^n.$$

Applications:

(i) Sum of independent normals is normal. Mean, variance add up.

(ii) Sum of independent gammas with the same scale parameter is gamma with the same scale and shape parameter added up. In particular, sum of independent exponentials is gamma.

(iii) Sum of independent Poisson is Poisson with parameters added up.

(iv) Sum of independent geometric is negative binomial.

Multinomial distribution. n categories, and each item can be from one and only one category. Sampling m times independently from the categories with probabilities $p_1, \ldots, p_n$, where $p_1 + \cdots p_n = 1$. Let X_i = the count of the ith category. Let $x_1, \ldots, x_n$ be non-negative integers adding up to m. Then

$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{m!}{x_1!x_2! \cdots x_n!} p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n}.$$

Prob. add up to one, as they are the terms in expansion of $(p_1 + \cdots + p_n)^m$.

(i) marginals are (lower order) multinomial. One dimensional $X_i \sim$Bin(m, p_i).

(ii) Conditionals:

(iii) Merging: $(X_1+X_2, X_3, \cdots, X_n) \sim$ Multinomial$(m; p1+p2, p_3, \cdots, p_n)$.

(iv) $Var(X_i) = mp_i(1 - p_i)$ and $Cov(X_i, X_j) = -mp_ip_j$ for all $i \neq j$.

8.2 Continuous Case

Joint pdf of $(X_1, \dots, X_n)$: If $f_{X_1,\dots,X_n}(x_1, \dots, x_n)$ satisfies $f_{X_1,\dots,X_n}(x_1, \dots, x_n) \geq 0$ and $\int f_{X_1,\dots,X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n = 1$.

Probabilities are obtained by

$$P\{(X_1, \dots, X_n) \in A\} = \int_{(x_1,\dots,x_n)\in A} f_{X_1,\dots,X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

$$Eg(X_1, \dots, X_n) = \int \dots \int g(x_1, \dots, x_n) f_{X_1,\dots,X_n}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Marginal of $(X_{i_1}, \dots, X_{i_k})$:

$$f_{X_{i_1},\dots,X_{i_k}}(x_{i_1}, \dots, x_{i_k}) = \int_{\text{other indices}} f_{X_1,\dots,X_n}(x_1, \dots, x_n).$$

One-dimensional marginals:

$$f_{X_i}(x_i) = \int f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Conditional:

$$f_{X_{k+1},\dots,X_n|X_1,\dots,X_k}(x_{k+1}, \dots, x_n | x_1, \dots, x_k) = \frac{f(x_1, \dots, x_n)}{f_{X_1,\dots,X_k}(x_1, \dots, x_k)}.$$

Covariance: $\mathrm{cov}(X_i, X_j)$ — based on pairwise distribution.

Independence: joint is the product of marginals. Equivalently, MGF is the product of individual MGF's.

EXAMPLE 1. Uniform over the ball.

$$f(x_1, x_2, x_3) = \frac{3}{4\pi} I\{x_1^2 + x_2^2 + x_3^2 < 1\}.$$

EXAMPLE 2. Dirichlet.

$$f(x_1,\ldots,x_{k-1}) = \frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)} x_1^{\alpha_1-1}\cdots x_{k-1}^{\alpha_{k-1}-1} x_k^{\alpha_k-1} I\{x_i>0, x_1+\cdots+x_k=1\}.$$

Properties:
(i) marginals are (lower order) Dirichlet. One dimensionals are beta.

(ii) Conditionals:

(iii) Merging of categories:

(iv) Covariances:

EXAMPLE 3. Let $n = 4$ and the joint density of (X_1, X_2, X_3, X_4) is

$$f_{(X_1,X_2,X_3,X_4)}(x_1, x_2, x_3, x_4) = \frac{3}{4}(x_1^2+x_2^2+x_3^2+x_4^2), \quad \text{if} \quad 0 < x_i < 1, i = 1, 2, 3, 4;$$

and $= 0$ otherwise.

(i) Show that this is a valid pdf.

(ii) Compute $P(X_1 < \frac{1}{2}, X_2 < \frac{3}{4}, X_4 > \frac{1}{2})$

(iii) Obtain the marginal pdf of (X_1, X_2).

(iv) Find the conditional pdf of (X_3, X_4) given $X_1 = \frac{1}{3}$ and $X_2 = \frac{2}{3}$.

(v) Compute $E(X_1X_2)$.

8.3 Multivariate Transformation

Let $(X_1, \cdots, X_n)$ be a random vector with pdf $f_{X_1,\cdots,X_n}(x_1, \cdots, x_n)$. Let $\mathcal{A} = \{\mathbf{x} : f_{\mathbf{X}}(\mathbf{x}) > 0\}$. A new random vector $(U_1, \cdots, U_n)$ is defined by

$$\begin{aligned} U_1 &= g_1(X_1, \cdots, X_n), \\ U_2 &= g_2(X_1, \cdots, X_n), \\ \cdots & \quad \cdots \\ U_n &= g_n(X_1, \cdots, X_n). \end{aligned}$$

The transformation is one-to-one from $\mathcal{A}$ onto $\mathcal{B}$. The inverse of g_i's are

$$\begin{aligned} X_1 &= h_1(U_1, \cdots, U_n), \\ X_2 &= h_2(U_1, \cdots, U_n), \\ \cdots & \quad \cdots \\ X_n &= h_n(U_1, \cdots, U_n). \end{aligned}$$

Let J be the Jacobian from the inverse. The joint pdf of $U_1, \cdots, U_n$ is then

$$f_{U_1,\cdots,U_n}(u_1, \cdots, u_n) = f_{X_1,\cdots,X_n}(h_1(u_1, \cdots, u_n), \cdots, h_n(u_1, \cdots, u_n))|J|.$$

Example: Let (X_1, X_2, X_3, X_4) have the joint pdf

$$f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4) = 24e^{-x_1-x_2-x_3-x_4}, \quad 0 < x_1 < x_2 < x_3 < x_4 < \infty.$$

Consider the transformation

$$U_1 = X_1, \quad U_2 = X_2 - X_1, \quad U_3 = X_3 - X_2, \quad U_4 = X_4 - X_3.$$

Example 3. Multivariate normal.

Joint density of $Y = (Y_1, \cdots, Y_n)$:

$$f_Y(y_1, \ldots, y_n) = \frac{1}{(2\pi)^{n/2}(\det(\Sigma))^{1/2}} \exp\left(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right).$$

Generation of multivariate normal $N(\mu, \Sigma)$:

(1)Let $X_1, \cdots, X_n$ be iid $N(0, 1)$.

(2) Write $\mathbf{X} = (X_1, \ldots, X_n)'$ and let

$$Y = A\mathbf{X} + \mu,$$

where $\Sigma = AA^T$ (i.e. the Cholesky decomposition). Then Y is multivariate normal with the above density function, with $E(Y) = \mu$ and $\text{Var}(Y) = \Sigma$. Marginals and conditionals are also (multivariate) normal.

9 Some Useful Inequalities.

A. Cauchy-Schwarz

$$(E(XY))^2 \leq E(X^2)E(Y^2).$$

B. Hölder

$$|E(XY)| \leq (E(|X|^p))^{1/p}(E(|X|^q))^{1/q},$$

where $p^{-1} + q^{-1} = 1$.

C. Minkowski

$$(E(|X+Y|^p))^{1/p} \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p},$$

for $p \geq 1$.

D. Jensen

A function ψ is called convex if $\psi(at + (1-a)s) \leq a\psi(t) + (1-a)\psi(s)$. For any convex function ψ,

$$E(\psi(X)) \geq \psi(E(X)).$$