Chapter 9

Multivariate Distributions

Outline

- 9.1 Joint distributions of $n > 2$ random variables
- 9.2 Order statistics
- 9.3 multinomial distributions

Section 9.1

Joint Distributions of $n > 2$ Random Variables

Definition of joint probability mass functions

- **Definition.** *Let* $X_1, X_2, \ldots, X_n$ *be discrete random variables defined on the same sample space, with sets of possible values* $A_1, A_2, \ldots, A_n$*, respectively. The function*

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

is called the joint probability mass function of $X_1, X_2, \ldots, X_n$.

- Note that
  1. $p(x_1, x_2, \ldots, x_n) \geq 0$.
  2. If for some i, $1 \leq i \leq n$, $x_i \notin A_i$, then $p(x_1, x_2, \ldots, x_n) = 0$.
  3. $\sum_{x_i \in A_i, i \leq i \leq n} p(x_1, x_2, \ldots, x_n) = 1$.

- The marginal probability mass function can be derived from the joint probability mass function

$$p_{X_i}(x_i) = P(X_i = x_i) = P(X_i = x_i, X_j \in A_j, 1 \leq j \leq n, j \neq i)$$
$$= \sum_{x_j \in A_j, j \neq i} p(x_1, x_2, \ldots, x_n).$$

- More generally, to find the joint probability mass function marginalized over a given set of k of these random variables, we sum up $p(x_1, x_2, \ldots, x_n)$ over all possible values of the remaining $n - k$ random variables.
- For example,

$$p_{X,Y}(x, y) = \sum_z p(x, y, z).$$

is the the joint probability mass function marginalized over X and Y.

Joint Probability Distribution Functions

- The joint probability distribution function of $X_1, X_2, \ldots, X_n$ is defined by

$$F(t_1, t_2, \ldots, t_n) = P(X_1 \le t_1, X_2 \le t_2, \ldots, X_n \le t_n)$$

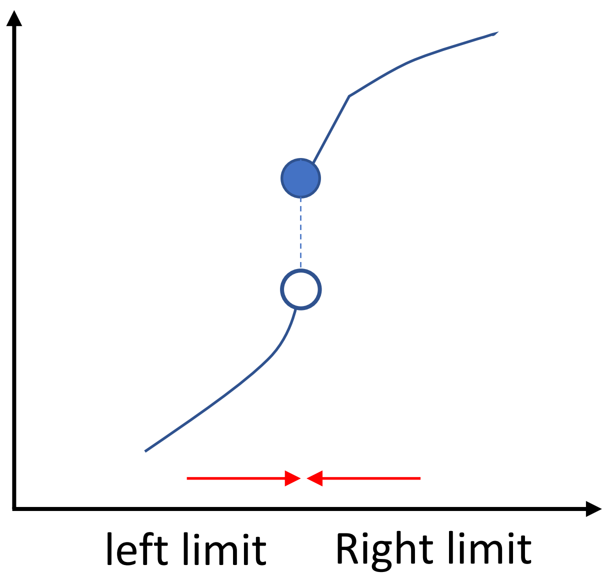
for all $-\infty < t_i < +\infty$, $i = 1, 2, \ldots, n$.

- The marginal probability distribution function of X_i, $1 \le i \le n$, can be found from F as follows:

$$\begin{aligned} F_{X_i}(t_i) &= P(X_i \le t_i) \\ &= P(X_1 < \infty, \ldots, X_{i-1} < \infty, X_i \le t_i, X_{i+1} < \infty, \ldots, X_n < \infty) \\ &= \lim_{\substack{t_j \to \infty \\ 1 \le j \le n, j \ne i}} F(t_1, t_2, \ldots, t_n). \end{aligned}$$

Properties of Joint Probability Distribution Functions

1. F is nondecreasing in each argument.
2. F is right continuous in each argument.
3. $F(t_1, t_2, \ldots, t_{i-1}, -\infty, t_{i+1}, \ldots, t_n) = 0$ for $i = 1, 2, \ldots, n$.
4. $F(\infty, \infty, \ldots, \infty) = 1$.



Marginal Probability Distribution Functions

- The marginal probability mass function can be derived from the joint probability mass function

$$p_{X_i}(x_i) = P(X_i = x_i) = P(X_i = x_i, X_j \in A_j, 1 \le j \le n, j \ne i)$$
$$= \sum_{x_j \in A_j, j \ne i} p(x_1, x_2, \ldots, x_n).$$

- More generally, to find the joint probability mass function marginalized over a given set of k of these random variables, we sum up $p(x_1, x_2, \ldots, x_n)$ over all possible values of the remaining $n - k$ random variables.
- For example,

$$p_{X,Y}(x, y) = \sum_z p(x, y, z).$$

is the the joint probability mass function marginalized over X and Y.

Generalization of Independence from Two Random Variables to n Random Variables

- Recall the definition of independence of two random variables in Chapter 8.
- **Definition.** *Two random variables X and Y are independent if, for arbitrary subsets A and B of real numbers, the events* $\{X \in A\}$ *and* $\{Y \in B\}$ *are independent, i.e.*

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

- **(Theorem 8.3)** The preceding definition implies that

$$F(t, u) = F_X(t)F_Y(u),$$

where F is the joint probability distribution function, and t and u are arbitrary real numbers.

- Also recall the definition of independence of multiple events in Chapter 3.
- **Definition.** *The set of events* $\{A_1, A_2, \ldots, A_n\}$ *is called independent if for every subset* $\{A_{i_1}, A_{i_2}, \ldots, A_{i_k}\}$, $k \geq 2$, of $\{A_1, A_2, \ldots, A_n\}$,

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k}).$$

- Suppose that $X_1, X_2, \ldots, X_n$ are random variables (discrete, continuous, or mixed) on a sample space. We say that they are independent if, for arbitrary subsets $A_1, A_2, \ldots, A_n$ of real numbers,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n)$$
$$= P(X_1 \in A_1)P(X_2 \in A_2) \cdots P(X_n \in A_n).$$

- Since the subsets are arbitrary, one may choose $A_{i_1} = A_{i_2} = \ldots = A_{i_k} = \mathbf{R}$ for $1 \le k \le n$ and has

$$
\begin{aligned}
&P(X_{i_j} \in A_{i_j}, k+1 \le j \le n) \\
&= P(X_{i_j} \in \mathbf{R}, 1 \le j \le k, X_{i_j} \in A_{i_j}, k+1 \le j \le n) \\
&= P(X_{i_1} \in \mathbf{R}) \cdots P(X_{i_k} \in \mathbf{R}) \cdot P(X_{i_{k+1}} \in A_{i_{k+1}}) \cdots P(X_{i_n} \in A_{i_n}) \\
&= P(X_{i_{k+1}} \in A_{i_{k+1}}) \cdots P(X_{i_n} \in A_{i_n})
\end{aligned}
$$

- If $\{X_1, X_2, \ldots, X_n\}$ is a sequence of independent random variables, its subsets are also independent sets of random variables.
- This observation motivates the following definition.
- **Definition** *A collection of random variables is called independent if all of its finite sub-collections are independent.*

- $X_1, X_2, \ldots, X_n$ are independent if and only if, for any $x_i \in \mathbf{R}$, $i = 1, 2, \ldots, n$,

$$P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$
$$= P(X_1 \le x_1)P(X_2 \le x_2) \cdots P(X_n \le x_n).$$

- That is, $X_1, X_2, \ldots, X_n$ are independent if and only if, for any $x_i \in \mathbf{R}$, $i = 1, 2, \ldots, n$,

$$F(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_n}(x_n).$$

- The preceding implies independence among a finite sub-collection, i.e.

$$F(x_1, x_2, \ldots, x_{n-1}) = \lim_{x_n \to \infty} F(x_1, x_2, \ldots, x_n)$$
$$= \lim_{x_n \to \infty} F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$
$$= F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_{n-1}}(x_{n-1}) \cdot 1$$

Generalization of Two Theorems

- **(Generalization of Theorem 8.5)** *If* $\{X_1, X_2, \ldots, X_n\}$ *is a sequence of independent random variables and for* $i = 1, 2, \ldots$, $g_i : \mathbf{R} \to \mathbf{R}$ *is a real-valued function, then the sequence* $\{g_1(X_1), g_2(X_2), \ldots\}$ *is also an independent sequence of random variables.*
- **Theorem 9.1 (Generalization of Theorem 8.4)** *Let* $X_1, X_2, \ldots, X_n$ *be jointly discrete random variables with the joint probability mass function* $p(x_1, x_2, \ldots, x_n)$. *Then* $X_1, X_2, \ldots, X_n$ *are independent if and only if* $p(x_1, x_2, \ldots, x_n)$ *is the product of their marginal densities* $p_{X_1}(x_1), p_{X_2}(x_2), \ldots, p_{X_n}(x_n)$.

Theorem 9.2

- Theorem 9.2 is a generalization of Theorem 8.1.
- **Theorem 9.2** *Let* $p(x_1, x_2, \ldots, x_n)$ *be the joint probability mass function of discrete random variables* $X_1, X_2, \ldots, X_n$. *For* $1 \leq i \leq n$, *let* A_i *be the set of possible values of* X_i. *If* j *is a function of* n *variables from* $\mathbf{R}^n$ *to* $\mathbf{R}$, *then* $Y = h(X_1, X_2, \ldots, X_n)$ *is a discrete random variable with expected value given by*

$$E(Y) = \sum_{x_n \in A_n} \cdots \sum_{x_1 \in A_1} h(x_1, x_2, \ldots, x_n) p(x_1, x_2, \ldots, x_n),$$

provided that the sum is finite.

Expectation of the Product of Independent Random Variables

- The expected value of the product of several independent discrete random variables is equal to the product of their expected values.
- Assume that $X_1, X_2, \cdots, X_n$ are independent, then

$$E(X_1 X_2 \cdots X_n) = \prod_{i=1}^{n} E(X_i).$$

$$E(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} E(X_i)$$

always true

Definition of Joint PDFs of n Random Variables

- **Definition** *Let $X_1, X_2, \ldots, X_n$ be continuous random variables defined on the same sample space. We say that $X_1, X_2, \ldots, X_n$ have have a continuous joint distribution if there exists a nonnegative function of n variables, $f(x_1, x_2, \ldots, x_n)$, on $\mathbf{R}^n$ such that for any region R in $\mathbf{R}^n$ that can be formed from n-dimensional rectangles by a countable number of set operations,*

$$P((X_1, X_2, \ldots, X_n) \in R) = \int \cdots \int_R f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n.$$

The function $f(x_1, x_2, \ldots, x_n)$ is called the joint probability density function of $X_1, X_2, \ldots, X_n$.

- Let $R = \{(x_1, x_2, \cdots, x_n): x_i \in A_i, 1 \le i \le n\}$, where $A_i, 1 \le i \le n$, is any subset of real numbers that can be constructed from intervals by a countable number of set operations.
- Then,

$$P(X_1 \in A_1, X_2 \in A_2, \cdots, X_n \in A_n)$$
$$= \int_{A_n} \int_{A_{n-1}} \cdots \int_{A_1} f(x_1, x_2, \cdots x_n)\, dx_1 dx_2 \cdots dx_n.$$

Marginal Probability Density Function

- Let f_{X_i} be the marginal probability density function of X_i, $1 \leq i \leq n$.
- Then,

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_n.$$

- To find the joint probability density function marginalized over a given set of k of these random variables, we integrate $f(x_1, x_2, \ldots, x_n)$ over all possible values of the remaining $n - k$ random variables.
- For example,

$$f_{Y,T}(y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z, t)\, dx\, dz.$$

Theorem 9.3

- **Theorem 9.3** *Let* $X_1, X_2, \ldots, X_n$ *be jointly continuous random variables with the joint probability density function* $f(x_1, x_2, \ldots, x_n)$. *Then,* $X_1, X_2, \ldots, X_n$ *are independent if and only if* $f(x_1, x_2, \ldots, x_n)$ *is the product of their marginal densities* $f_{X_1}(x_1), f_{X_2}(x_2), \ldots, f_{X_n}(x_n)$.

Theorem 9.4

- **Theorem 9.4** *Let* $f(x_1, x_2, \ldots, x_n)$ *be the joint probability density function of random variables* $X_1, X_2, \ldots, X_n$. *If* h *is a function of* n *variables from* $\mathbf{R}^n$ *to* $\mathbf{R}$, *then* $Y = h(X_1, X_2, \ldots, X_n)$ *is a random variable with expected value given by*

$$E(Y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n.$$

provided that the integral is absolutely convergent.

Example 9.4

- A system has n components, the lifetime of each being an exponential random variable with parameter λ.
- Suppose that the lifetimes of the components are independent random variables, and the system fails as soon as any of its components fails.
- Find the probability density function of the time until the system fails.

- **Solution.**
- Let $X_1, X_2, \ldots, X_n$ be the lifetimes of the components.
- Then $X_1, X_2, \ldots, X_n$ independent random variables and for $i = 1, 2, \ldots, n$,

$$P(X_i \leq t) = 1 - e^{-\lambda t}.$$

- Letting X be the time until the system fails, we have

$$X = \min(X_1, X_2, \ldots, X_n).$$

- Therefore,

$$\begin{aligned} P(X > t) &= P(X = \min(X_1, X_2, \ldots, X_n) > t) \\ &= P(X_1 > t_1, X_2 > t_2, \ldots, X_n > t_n) \\ &= P(X_1 > t_1)P(X_2 > t_2)\ldots P(X_n > t_n) \\ &= (e^{-\lambda t})(e^{-\lambda t})\cdots(e^{-\lambda t}) = e^{-n\lambda t}. \end{aligned}$$

- It implies that X is an exponential random variable with rate $n\lambda$.
- The density function of X is

$$f(t) = \frac{d}{dt}P(X \leq t) = n\lambda e^{-n\lambda t}.$$

Remark

- If $X_1, X_2, \cdots, X_n$ are n independent exponential random variables with parameters $\lambda_1, \lambda_2, \cdots, \lambda_n$, respectively, then $\min(X_1, X_2, \cdots, X_n)$ is an exponential random variable with parameter $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.
- Hence,

$$E[\min(X_1, X_2, \cdots, X_n)] = \frac{1}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}.$$

Example 9.5

Example 9.5

(a) Prove that the following is a joint probability density function.

$$f(x, y, z, t) = \begin{cases} \dfrac{1}{xyz} & \text{if } 0 < t \le z \le y \le x \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

(b) Suppose that f is the joint probability density function of random variables X, Y, Z, and T. Find $f_{Y,Z,T}(y, z, t)$, $f_{X,T}(x, t)$, and $f_Z(z)$.

Solution:

$$\boxed{0 < t \le z \le y \le x \le 1}$$

(a) Since $f(x, y, z, t) \ge 0$ and

$$\int_0^1 \int_0^x \int_0^y \int_0^z \frac{1}{xyz}\,dt\,dz\,dy\,dx = \int_0^1 \int_0^x \int_0^y \frac{1}{xy}\,dz\,dy\,dx$$

$$= \int_0^1 \int_0^x \frac{1}{x}\,dy\,dx = \int_0^1 dx = 1,$$

f is a joint probability density function.

(b) For $0 < t \le z \le y \le 1$,

$$f_{Y,Z,T}(y, z, t) = \int_y^1 \frac{1}{xyz}dx = \frac{1}{yz}\ln x\Big|_y^1 = -\frac{\ln y}{yz}.$$

Therefore,

$$f_{Y,Z,T}(y, z, t) = \begin{cases} -\dfrac{\ln y}{yz} & \text{if } 0 < t \le z \le y \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

To find $f_{X,T}(x,t)$, we have that for $0 < t \le x \le 1$,

$$f_{X,T}(x,t) = \int_t^x \int_t^y \frac{1}{xyz}\,dz\,dy = \int_t^x \Big[\frac{\ln z}{xy}\Big]_t^y dy$$

$$= \int_t^x \Big(\frac{\ln y}{xy} - \frac{\ln t}{xy}\Big)dy = \Big[\frac{1}{2x}(\ln y)^2 - \frac{\ln t}{x}\ln y\Big]_t^x$$

$$= \frac{1}{2x}(\ln x)^2 - \frac{1}{x}(\ln t)(\ln x) + \frac{1}{2x}(\ln t)^2 = \frac{1}{2x}(\ln x - \ln t)^2$$

$$= \frac{1}{2x}\ln^2\frac{x}{t}.$$

$$\boxed{0 < t \le z \le y \le x \le 1}$$

Therefore,

$$f_{X,T}(x,t) = \begin{cases} \dfrac{1}{2x}\ln^2\dfrac{x}{t} & \text{if } 0 < t \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

To find $f_Z(z)$, we have that for $0 < z \leq 1$,

$$f_Z(z) = \int_z^1 \int_z^x \int_0^z \frac{1}{xyz}\,dt\,dy\,dx = \int_z^1 \int_z^x \frac{1}{xy}\,dy\,dx$$

$$= \int_z^1 \Big[\frac{1}{x}\ln y\Big]_z^x dx = \int_z^1 \Big(\frac{1}{x}\ln x - \frac{1}{x}\ln z\Big)dx$$

$$= \Big[\frac{1}{2}(\ln x)^2 - (\ln x)(\ln z)\Big]_z^1 = \frac{1}{2}(\ln z)^2.$$

Thus

$$\boxed{0 < t \leq z \leq y \leq x \leq 1}$$

$$f_Z(z) = \begin{cases} \dfrac{1}{2}(\ln z)^2 & \text{if } 0 < z \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad \blacklozenge$$

Definition of Random Samples

- **Definition:** We say that n random variables $X_1, X_2, \ldots, X_n$ form a random sample of size n, from a (continuous or discrete) distribution function F, if they are independent and, for $1 \leq i \leq n$, the distribution function of X_i is F. Therefore, elements of a random sample are independent and identically distributed.

Abbreviated as i.i.d.

- The rest of Chapter 9 is skipped.