4. Multiple Random Variables

Definition 4.1

An n-dimensional random vector is a function from a sample space S into $\mathbb{R}^n$.

Bivariate Random Vectors

For simplicity we shall consider $n = 2$ at first so our random vector is the ordered pair (X, Y).

Definition 4.2

The joint cumulative distribution function of the bivariate random vector (X, Y) *is defined as*

$$F_{X,Y}(x, y) = P(X \leqslant x, Y \leqslant y) = P(X \leqslant x \bigcap Y \leqslant y) \quad \text{for all } (x, y) \in \mathbb{R}^2$$

- A bivariate random vector is said to be discrete if both of its component random variables are discrete.

- If both components are continuous then we describe the random vector as continuous.

- In many applied instances, however, one component may be discrete and the other continuous. Such random vectors, usually called mixed random vectors are relatively easy to deal with also but confuse the notation so I will usually just deal with discrete or continuous random vectors.

Definition 4.3

Let (X, Y) be a discrete bivariate random vector. The joint probability mass function is defined as

$$f_{X,Y}(x,y) \;=\; P(X=x, Y=y) \qquad \textit{for all } (x,y) \in \mathbb{R}^2$$

- The joint pmf $f(x,y)$ satisfies
  1. $f(x,y) \geqslant 0$ for every $(x,y) \in \mathbb{R}^2$.
  2. $\displaystyle\sum_{(x,y)\in\mathbb{R}^2} f(x,y) = 1$.
- For any set $A \subset \mathbb{R}^2$ we have

$$P\Big((X,Y) \in A\Big) \;=\; \sum_{(x,y)\in A} f(x,y)$$

- Expectations of scalar functions $g(x,y)$ are defined as

$$\mathrm{E}[g(X,Y)] \;=\; \sum_{(x,y)\in\mathbb{R}^2} g(x,y) f(x,y).$$

Theorem 4.1

Let (X, Y) be a discrete bivariate random vector with joint pmf $f_{X,Y}$. Then the marginal pmfs of X and Y are given by

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

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

- It is important to note that the joint pmf uniquely determines the marginals but the reverse is not true.

Definition 4.4

A non-negative function $f(x, y)$ mapping $\mathbb{R}^2$ to $\mathbb{R}$ is called the joint probability density function of a continuous bivariate random vector (X, Y) if

$$P\Big((X, Y) \in A\Big) \;=\; \iint_A f(x, y)\, dx\, dy \quad \text{for every } A \subset \mathbb{R}^2.$$

- The joint pdf $f(x, y)$ satisfies

  1. $f(x, y) \geqslant 0$ for every $(x, y) \in \mathbb{R}^2$.

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

- Expectations of scalar functions $g(x, y)$ are defined as

$$\mathsf{E}[g(X, Y)] \;=\; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy.$$

Theorem 4.2

Let (X, Y) be a continuous bivariate random vector with joint pdf $f_{X,Y}$. Then the marginal pdfs of X and Y are given by

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

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

- The joint cdf of a random vector is given by

$$
\begin{aligned}
F(x,y) &= \textstyle\sum_{u\leqslant x}\sum_{v\leqslant y} f(u,v) && \text{(discrete)}\\
F(x,y) &= \textstyle\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv && \text{(continuous)}
\end{aligned}
$$

- In the continuous case we also have

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

  at continuity points of $f(x,y)$.

- Taking limits of $F(x,y)$ as one of the components goes to infinity gives the marginal cdfs.

Conditional Distributions and Independence

Definition 4.5

Let (X, y) be a discrete bivariate random vector with joint pmf $f_{X,Y}(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$ the conditional probability mass function of Y given $X = x$ is

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

- It is not hard to see that this does define a valid probability mass function for the random variable Y.

Definition 4.6

Let (X, y) be a continuous bivariate random vector with joint pdf $f_{X,Y}(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$ the conditional probability density function of Y given $X = x$ is

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

- Note that we are conditioning on an event of probability 0.

- The definition does however define a valid pdf.

- Conditional pmfs and pdfs can be used in exactly the same way as other univariate pmfs and pdfs.

- In particular we can get the conditional expected value of $g(Y)$ given $X = x$ as

$$\mathsf{E}[g(Y) \mid X = x] = \begin{cases} \sum_y g(y) f_{Y|X}(y \mid x) & \text{(discrete)} \\ \int_{-\infty}^{\infty} g(y) f_{Y|X}(y \mid x)\, dy & \text{(continuous)} \end{cases}$$

- In particular we can get the conditional mean and variance of Y given $X = x$.

- Note that the conditional variance is defined as

$$\begin{aligned} \mathsf{Var}(Y \mid X = x) &= \mathsf{E}\left[(Y - \mathsf{E}[Y \mid X = x])^2\right] \\ &= \mathsf{E}[Y^2 \mid X = x] - \left(\mathsf{E}[Y \mid X = x]\right)^2 \end{aligned}$$

- $\mathsf{E}[X \mid Y]$ and $\mathsf{Var}[X \mid Y]$ are random variables.

- They are a function of the random variable Y.

Theorem 4.3
Let X and Y be two random variables then

$$\begin{aligned} \mathsf{E}(X) &= \mathsf{E}[\mathsf{E}(X \mid Y)] \\ \mathsf{Var}(X) &= \mathsf{E}[\mathsf{Var}(X \mid Y)] + \mathsf{Var}[\mathsf{E}(X \mid Y)] \end{aligned}$$

- The inner moments are found using the conditional distribution of X given Y and the outer moments are found using the marginal distribution of Y.

Definition 4.7

Let (X, Y) be a bivariate random vector with joint pmf or pdf $f_{X,Y}(x, y)$ and marginal pmfs or pdfs $f_X(x)$ and $f_Y(y)$. Then X and Y are independent random variables if and only if

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \qquad \text{for every } x \in \mathbb{R} \text{ and } y \in \mathbb{R}.$$

Lemma 4.1

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

$$f_{X,Y}(x, y) = g(x) h(y) \qquad \text{for every } x \in \mathbb{R} \text{ and } y \in \mathbb{R}.$$

Theorem 4.4

Let X and Y be independent random variables.

a. *If $g(x)$ is a function of x only and $h(y)$ is a function of y only then*

$$\mathsf{E}[g(X)h(Y)] = \mathsf{E}[g(X)]\mathsf{E}[h(Y)]$$

b. *For any $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$ the events $(X \in A)$ and $(Y \in B)$ are independent events.*

Theorem 4.5

Let X and Y be two independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$ then the moment generating function of the random variable $Z = X + Y$ is

$$M_Z(t) = M_X(t)M_Y(t)$$

Bivariate Transformations

Theorem 4.6

Let (X, Y) be a discrete bivariate random vector with support $\mathcal{A}$ and define $U = g_1(X, Y)$ and $V = g_2(X, Y)$. Then (U, V) is a discrete bivariate random vector with support

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

and the pmf of (U, V) is given by

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

where $A_{u,v} = \{(x, y) \in \mathcal{A} : g_1(x, y) = u, g_2(x, y) = v\}$.

Theorem 4.7

Let (X, Y) be a continuous bivariate random vector with joint pdf $f_{X,Y}(x, y)$ on support $\mathcal{A} = \{(x, y \in \mathbb{R} : f_{X,Y}(x, y) > 0\}$. Let the functions $g_1(x, y)$ and $g_2(x, y)$ define a one-to-one transformation of $\mathcal{A}$ to

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

and let the inverse transformation be given by $x = h_1(u, v)$, $y = h_2(u, v)$. Then the pdf of the random vector (U, V) where $U = g_1(X, Y)$ and $V = g_2(X, Y)$ is given by

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

where J is the Jacobian of the transformation given by

$$J = \begin{vmatrix} \dfrac{\partial h_1(u, v)}{\partial u} & \dfrac{\partial h_1(u, v)}{\partial v} \\ \dfrac{\partial h_2(u, v)}{\partial u} & \dfrac{\partial h_2(u, v)}{\partial v} \end{vmatrix}$$

- As in the univariate case this can be extended to transformations which are not one-to-one.

- In that case we consider a partition $A_1, \ldots, A_k$ of $\mathcal{A}$ such that the transformation is one-to-one from each A_i to $\mathcal{B}$.

- We then apply the previous theorem to each set in the partition and sum the results to get the pdf of (U, V).

- The partition may also include a set A_0 such that $P((X, Y) \in A_0) = 0$ without changing the result.

Hierarchical Models and Mixture Distributions

- For many complicated random processes, it is easiest to model it using a sequence of conditional and marginal models.

- In the simplest hierarchy we have the conditional distribution of $X \mid Y$ and the marginal distribution of Y.

- The joint distribution is then given by

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

- The marginal distribution of X can then be found as

$$f_X(x) = \begin{cases} \sum_y f_{X|Y}(x \mid y) f_Y(y) & \text{if } Y \text{ is discrete} \\ \\ \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y)\, dy & \text{if } Y \text{ is continuous} \end{cases}$$

- The marginal distribution of X in this case is called a mixture distribution.

- Mixture distributions often have rather formidable looking pmfs or pdfs.

- Theorem 3.3 gives us simple ways to find the moments of X using the hierarchical structure.

Covariance and Correlation

Definition 4.8

Suppose that X and Y are random variables with finite variances. Denote $\mathsf{E}[X] = \mu_X$, $\mathsf{E}[Y] = \mu_Y$, $\mathsf{Var}(X) = \sigma_X^2$ *and* $\mathsf{Var}(Y) = \sigma_Y^2$. *Then the covariance of X and Y is*

$$\mathsf{Cov}(X, Y) \;=\; \mathsf{E}[(X - \mu_X)(Y - \mu_Y)]$$

and the correlation coefficient of X and Y is

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

Theorem 4.8

For any random variable X and Y

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

Theorem 4.9

For any random variables X and Y,

a. $-1 \leqslant \rho_{X,Y} \leqslant 1$.

b. $|\rho_{X,Y}| = 1$ *if and only if there exist constants* $a \neq 0$ *and* b *such that* $P(Y = aX + b) = 1$. *In that case the signs of* a *and* $\rho_{X,Y}$ *are always the same.*

Theorem 4.10

If X and Y are two independent random variables then

$$\text{Cov}(X, Y) = \rho_{X,Y} = 0$$

- The reverse is not true since covariance and correlation only consider the linear relationship between X and Y.

Theorem 4.11

If X and Y are any two random variables and a and b are constants then

$$\begin{aligned} \text{E}[aX + bY] &= a\text{E}[X] + b\text{E}[Y] \\ \text{Var}(aX + bY) &= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y) \end{aligned}$$

General Multivariate Distributions

- A multivariate random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is a function mapping the sample space S to $\mathcal{X} \subset \mathbb{R}^n$.

- If $\mathcal{X}$ is countable then the joint probability mass function for any $\boldsymbol{x} = (x_1, \ldots, x_n)$ is

$$f(\boldsymbol{x}) = P(X_1 = x_1, \ldots, X_n = x_n)$$

  and for any $A \subset \mathbb{R}^n$ we have

$$P(\boldsymbol{X} \in A) \;=\; \sum_{\boldsymbol{x} \in A} f(\boldsymbol{x}).$$

- If $\boldsymbol{X}$ is continuous then the joint probability density function of $\boldsymbol{X}$ is the non-negative real function $f(x_1, \ldots, x_n)$ such that for any $A \subset \mathbb{R}^n$

$$P(\boldsymbol{X} \in A) = \int \cdots \int_A f(\boldsymbol{x})\, d\boldsymbol{x} = \int \cdots \int_A f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

- If $g(x_1, \ldots, x_n)$ is a real-valued function then

$$E[g(\boldsymbol{X})] = \begin{cases} \displaystyle\sum_{\boldsymbol{x} \in \mathbb{R}^n} g(\boldsymbol{x}) f(\boldsymbol{x}) & \text{(discrete case)} \\ \displaystyle\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{x}) f(\boldsymbol{x}) \, d\boldsymbol{x} & \text{(continuous case)} \end{cases}$$

- Marginal pmfs (pdfs) of some subset of the coordinates of $\boldsymbol{X}$ is found by summing (integrating) the pmf (pdf) over the remaining coordinates.

- Conditional pmfs (pdfs) of some subset of the coordinates given the rest is found by dividing the full joint pmf (pdf) by the marginal pmf (pdf) of the conditioning coordinates evaluated at their given values.

The multinomial distribution

- Generalizes the binomial to the case where there are more than two categories and we want to count the number in each category.

Definition 4.9

Suppose that m and n are positive integers and let $p_1, \ldots, p_n$ be constants such that $0 \leqslant p_i \leqslant 1$ for $i = 1, \ldots, n$ and $\sum_i p_i = 1$. Then the random vector $(X_1, \ldots, X_n)$ has a multinomial distribution with m trials and probabilities $p_1, \ldots, p_n$ if the joint pmf is

$$f(x_1, \ldots, x_n) = \frac{m!}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_n^{x_n}$$

for any $(x_1, \ldots, x_n)$ such that each x_i is a non-negative integer and $\sum_i x_i = m$.

Theorem 4.12

If $(x_1, \ldots, X_n)$ *is a multinomial random vector with* m *trials and probabilities* $p_1, \ldots, p_n$ *then the marginal distribution of* X_i *is the binomial distribution with parameters* m *and* p_i*.*

The conditional distribution of $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ *given* $X_i = x_i$ *is the multinomial distribution with* $m - x_i$ *trials and probabilities* $(p'_1, \ldots, p'_{i-1}, p'_{i+1}, \ldots, p'_n)$ *where*

$$p'_j = \frac{p_j}{1 - p_i} \qquad j = 1, \ldots, i-1, i+1, \ldots, n$$

Theorem 4.13

If $(x_1, \ldots, X_n)$ *is a multinomial random vector with* m *trials and probabilities* $p_1, \ldots, p_n$ *then for any* $i, j \in \{1, \ldots, n\}$ *with* $i \neq j$*,*

$$\text{Cov}(X_i, X_j) = -m p_i p_j$$

Definition 4.10

Let $X_1, \ldots, X_n$ *be random variables with joint pmf or pdf* $f(x_1, \ldots, x_n)$ *and let* $f_{X_i}(x_i)$ *denote the marginal pmf or pdf of* X_i. *Then* $X_1, \ldots, X_n$ *are called independent random variables if for every* $(x_1, \ldots, x_n) \in \mathbb{R}^n$

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

Theorem 4.14

Let $(X_1, \lambda, dots, X_n)$ *be a random vector with joint pdf or pmf* $f(x_1, \ldots, x_n)$. *Then the random variables* $X_1, \ldots, X_n$ *are independent if, and only if, there exist functions* $g_1, \ldots, g_n$ *such that for every* $(x_1, \ldots, x_n) \in \mathbb{R}^n$

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} g_1(x_i).$$

Theorem 4.15

Let $X_1, \ldots, X_n$ be independent random variables and let $g_1, \ldots, g_n$ be univariate real-valued functions. Then

$$\mathsf{E}\left[\prod_{i=1}^{n} g_i(X_i)\right] = \prod_{i=1}^{n} \mathsf{E}[g_i(X_i)]$$

Theorem 4.16

Let $X_1, \ldots, X_n$ be independent random variables with moment generating functions $M_{X_i}(t)$ $i = 1, \ldots, n$ and let $a_i, \ldots, a_n$ and b be real constants. Then the moment generating function of the random variable $Z = a_1 X_1 + \cdots + a_n X_n + b$ is

$$M_Z(t) = \mathsf{e}^{bt} \prod_{i=1}^{n} M_{X_i}(a_i t)$$

Definition 4.11

Suppose that there is a one-to-one transformation

$$G(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$$

with inverse

$$H(y_1, \ldots, y_n) = (x_1, \ldots, x_n).$$

The Jacobian of the transformation is defined as the determinant of the matrix of partial derivatives of the inverse functions.

$$J(\boldsymbol{x}, \boldsymbol{y}) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} \frac{\partial h_1(\boldsymbol{y})}{\partial y_1} & \frac{\partial h_1(\boldsymbol{y})}{\partial y_2} & \cdots & \frac{\partial h_1(\boldsymbol{y})}{\partial y_n} \\ \frac{\partial h_2(\boldsymbol{y})}{\partial y_1} & \frac{\partial h_2(\boldsymbol{y})}{\partial y_2} & \cdots & \frac{\partial h_2(\boldsymbol{y})}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n(\boldsymbol{y})}{\partial y_1} & \frac{\partial h_n(\boldsymbol{y})}{\partial y_2} & \cdots & \frac{\partial h_n(\boldsymbol{y})}{\partial y_n} \end{vmatrix}.$$

Theorem 4.17

Suppose that $\boldsymbol{X}$ is a continuous random vector pdf $f_{\boldsymbol{X}}$ and the transformation

$$Y_i = g_i(X_1, \ldots, X_n) \quad i = 1, \ldots, n$$

is a one-to-one transformation from the support, $\mathcal{X}$, of $\boldsymbol{X}$ to the support, $\mathcal{Y}$, of $\boldsymbol{Y}$ with inverse transformation

$$X_i = h_i(Y_1, \ldots, Y_n) \quad i = 1, \ldots, n.$$

Let $J(\boldsymbol{x}, \boldsymbol{y})$ be the Jacobian of the transformation and suppose that it is not identically equal to zero over $\mathcal{Y}$. Then the joint pdf of $\boldsymbol{Y}$ is

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_n) = f_{\boldsymbol{X}}(h_1(\boldsymbol{y}), \ldots, h_n(\boldsymbol{y})) \left| J(\boldsymbol{x}, \boldsymbol{y}) \right|.$$