Multiple Random Variables

Definition 4.1
An n-dimensional random vector is a function from a sample space S into $\mathbb{R}^n$.

* Each of the components of a random vector are random variables and so each can be continuous or discrete.

* For convenience of notation we will primarily deal with the situation where all components are either discrete or they are all continuous.

* When some components are discrete and some are continuous we use integration for the continuous parts and summations for the discrete parts as appropriate.

Bivariate Random Vectors

* For simplicity we shall consider $n = 2$ at first so our random vector is the ordered pair (X_1, X_2).

Definition 4.2

The joint cumulative distribution function of the bivariate random vector (X_1, X_2) *is defined as*

$$\begin{aligned} F_{X_1,X_2}(x_1, x_2) &= P(X_1 \leqslant x_1, X_2 \leqslant x_2) \\ &= P(X_1 \leqslant x_1 \bigcap X_2 \leqslant x_2) \quad \textit{for all } (x_1, x_2) \in \mathbb{R}^2 \end{aligned}$$

Joint Probability Mass Function

Definition 4.3

Let (X_1, X_2) be a discrete bivariate random vector. The joint probability mass function is defined as

$$f_{X_1,X_2}(x_1, x_2) \;=\; P(X_1 = x_1, X_2 = x_2) \qquad \textit{for all } (x_1, x_2) \in \mathbb{R}^2$$

* The joint pmf $f(x_1, x_2)$ satisfies

  1. $f(x_1, x_2) \geqslant 0$ for every $(x_1, x_2) \in \mathbb{R}^2$.

  2. $\displaystyle\sum_{(x_1,x_2)\in\mathbb{R}^2} f(x_1, x_2) = 1.$

Discrete Bivariate Probabilities and Expectations

* For any set $A \subset \mathbb{R}^2$ we have

$$P\Big((X_1, X_2) \in A\Big) \;=\; \sum_{(x_1,x_2)\in A} f(x_1, x_2)$$

* Expectations of scalar functions $g(x_2, y_2)$ are defined as

$$\mathsf{E}\left(g(X_1, X_2)\right) \;=\; \sum_{(x_1,x_2)\in \mathbb{R}^2} g(x_1, x_2) f(x_1, x_2).$$

Joint Probability Density Function

Definition 4.4

A non-negative function $f(x_1, x_2)$ *mapping* $\mathbb{R}^2$ *to* $\mathbb{R}$ *is called the joint probability density function of a continuous bivariate random vector* (X_1, X_2) *if*

$$P\Big((X_1, X_2) \in A\Big) \;=\; \iint_A f(x_1, x_2)\, dx\, dy \quad \textit{for every } A \subset \mathbb{R}^2.$$

* The joint pdf $f(x_1, x_2)$ satisfies

  1. $f(x_1, x_2) \geqslant 0$ for every $(x_1, x_2) \in \mathbb{R}^2$.

  2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1\, dx_2 \;=\; 1.$

* Expectations of scalar functions $g(x_1, x_2)$ are defined as

$$\mathsf{E}\left(g(X_1, X_2)\right) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2) f(x_1, x_2)\, dx_1\, dx_2.$$

Marginal Distributions

* The joint distribution describes the behaviour of the random vector.

* In many situations we wish to extract the distribution of just one component.

* We use the term marginal distribution to describe the distribution of one component of the random vector.

* Note that these are simply univariate distributions as we saw previously.

* Note that, although we can derive the marginal distributions from the joint the reverse is generally not true.

Marginal Probability Mass Functions

Theorem 4.1

Let (X_1, X_2) *be a discrete bivariate random vector with joint pmf* f_{X_1,X_2}. *Then the marginal pmfs of* X_1 *and* X_2 *are given by*

$$f_{X_1}(x_1) = P(X_1 = x_1) = \sum_{x_2 \in \mathbb{R}} f_{X_1,X_2}(x_1, x_2)$$

$$f_{x_2}(x_2) = P(X_2 = x_2) = \sum_{x_1 \in \mathbb{R}} f_{X_1,X_2}(x_1, x_2)$$

Marginal Probability Density Functions

Theorem 4.2

Let (X_1, X_2) be a continuous bivariate random vector with joint pdf f_{X_1,X_2}. Then the marginal pdfs of X_1 and X_2 are given by

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_2$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_1$$

Properties of the Joint CDF

* $F_{X_1,X_2}(x_1, x_2)$ is right continuous in both of its arguments.

* Similar to the univariate case we have

$$\begin{aligned} \lim_{(x_1,x_2)\to(-\infty,-\infty)} F_{X_1,X_2}(x_1, x_2) &= 0 \\ \lim_{(x_1,x_2)\to(\infty,\infty)} F_{X_1,X_2}(x_1, x_2) &= 1 \end{aligned}$$

* When we take limits with respect to one component we have

$$\begin{aligned} \lim_{x_1\to-\infty} F_{X_1,X_2}(x_1, x_2) &= 0 \\ \lim_{x_1\to\infty} F_{X_1,X_2}(x_1, x_2) &= F_{X_2}(x_2) \end{aligned}$$

and similarly for x_2.

Conditional Distributions

* Conditional distributions specify the distribution of one component f we know the value of the other.

* In the discrete case we apply the definition of conditional probability to get

$$
\begin{aligned}
f_{X_1|X_2}(x_1 \mid x_2) &= \mathsf{P}(X_1 = x_1 \mid X_2 = x_2) \\
&= \frac{\mathsf{P}(X_1 = x_1, X_2 = x_2)}{\mathsf{P}(X_2 = x_2)} \\
&= \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}
\end{aligned}
$$

provided $f_{X_2}(x_2) = \mathsf{P}(X_2 = x_2) > 0$.

* Note that we have

$$
F_{X_1,X_2}(x_1, x_2) = \sum_{u \leqslant x_1} \sum_{v \leqslant x_2} f_{X_1|X_2}(u \mid v) f_{X_2}(v)
$$

Continuous Conditional Densities

* In the continuous case, we always have that $\mathrm{P}(X_2 = x_2) = 0$ so we cannot proceed in this way.

* However note that analogously to the discrete case we can write

$$F_{X_1,X_2}(x_1,x_2) \;=\; \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \frac{f_{X_1,X_2}(u,v)}{f_{X_2}(v)} f_{X_2}(v)\,dv\,du$$

* We can therefore define the conditional density function

$$f_{X_1|X_2}(x_1 \mid x_2) \;=\; \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)}$$

provided the marginal density $f_{X_2}(x_2) > 0$.

* It is easy to verify that this does indeed define a univariate density function.

Conditional Expectations

* 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 $g(X_2)$ given $X_1 = x_1$ as

$$\mathsf{E}\left(g(X_2) \mid X_1 = x_1\right) = \begin{cases} \displaystyle\sum_{x_2} g(x_2) f_{X_2|X_1}(x_2 \mid x_1) & \text{(discrete)} \\ \displaystyle\int_{-\infty}^{\infty} g(x_2) f_{X_2|X_1}(x_2 \mid x_1)\, dy & \text{(continuous)} \end{cases}$$

* In particular we can get the conditional mean and variance of X_2 given $X_1 = x_1$.

Conditional Mean and Variance

* If we do not specify an actual value of the conditioning variable the conditional moments become functions of the conditioning random variable.

* Hence $\mathsf{E}(X_2 \mid X_1)$ and $\mathsf{Var}(X_2 \mid X_1)$ are random variables.

Theorem 4.3

Let X_1 and X_2 be two random variables then

$$\begin{aligned}\mathsf{E}(X_1) &= \mathsf{E}\left(\mathsf{E}(X_1 \mid X_2)\right)\\ \mathsf{Var}(X_1) &= \mathsf{E}\left(\mathsf{Var}(X_1 \mid X_2)\right) + \mathsf{Var}\left(\mathsf{E}(X_1 \mid X_2)\right)\end{aligned}$$

* The inner moments are found using the conditional distribution of X_1 given X_2 and the outer moments are found using the marginal distribution of X_2.

Independent Random Variables

Definition 4.5

Two random variables X_1 and X_2 are said to be independent if, and only if, we can write

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \qquad \text{for every } (x_1,x_2) \in \mathbb{R}^2$$

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

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

Properties of Independent Random Variables

Theorem 4.4

Suppose that X_1 and X_2 are independent random variables then

1. *For any sets $A \subset \mathbb{R}$, $B \subset \mathbb{R}$, the events $\{X_1 \in A\}$ and $\{X_2 \in B\}$ are independent events; that is*

$$\mathsf{P}(X_1 \in A, X_2 \in B) \;=\; \mathsf{P}(X_1 \in A)\mathsf{P}(X_2 \in B)$$

2. *If g and h are univariate functions then*

$$\mathsf{E}\left(g(X_1)h(X_2)\right) \;=\; \mathsf{E}\left(g(X_1)\right)\mathsf{E}\left(h(X_2)\right)$$

Theorem 4.5

Let X_1 and X_2 be two independent random variables with moment generating functions $M_{X_1}(t)$ and $M_{X_2}(t)$ then the moment generating function of the random variable $Z = X_1 + X_2$ is

$$M_Z(t) \;=\; M_{X_1}(t)M_{X_2}(t)$$

Covariance

Definition 4.6

Suppose that X_1 and X_2 are two random variables then we define the covariance of the joint distribution to be

$$\text{Cov}(X_1, X_2) \;=\; \text{E}\Big((X_1 - \mu_1)(X_2 - \mu_2)\Big)$$

where $\text{E}(X_1) = \mu_1$ and $\text{E}(X_2) = \mu_2$, provided μ_1, μ_2 and $\text{E}(X_1 X_2)$ all exist.

* It is easy to see that $\text{Cov}(X_1, X_2) \;=\; \text{E}(X_1 X_2) - \text{E}(X_1)\,\text{E}(X_2)$.

* The sign of the covariance gives the direction of any linear relationship between X_1 and X_2.

* The magnitude, however, is very dependent on the measurement scales for the random variables.

  $\text{Cov}(aX_1, bX_2) \;=\; ab\,\text{Cov}(X_1, X_2)$ for any a, b not equal to 0.

The Correlation Coefficient

Definition 4.7

Suppose that X_1 and X_2 are two random variables then we define the correlation coefficient of the joint distribution to be

$$\rho_{X_1,X_2} = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}}$$

provided $E(X_1^2)$, $\mathrm{E}(X_2^2)$ and $\mathrm{E}(X_1X_2)$ all exist.

* The correlation coefficient clearly has the same sign as the covariance.

* It is a unit-free number, however, which does not depend on the scales of measurement of either random variable.

$$\rho_{aX_1,bX_2} = \rho_{X_1,X_2} \qquad \text{for any } a, b \text{ not equal to } 0.$$

Independence and Correlation

Theorem 4.6

If X_1 and X_2 are independent random variables then

$$\text{Cov}(X_1, X_2) = 0.$$

* Clearly the correlation coefficient is also 0 for independent random variables.

* The inverse is not true in general.

* This is because covariance and correlation only measure linear relationships. Two random variables can have strong non-linear relationships but have correlation equal to 0.

Properties of the Correlation

Theorem 4.7

Let X_1 and X_2 be any two random variables for which the correlation coefficient ρ exists. Then

1. $-1 \leqslant \rho \leqslant 1$.

2. $|\rho| = 1$ *if, and only if, there exist numbers* $a \neq 0$ *and* $b \in \mathbb{R}$ *with* $\mathrm{sign}(a) = \mathrm{sign}(\rho)$ *such that*

$$\mathsf{P}\left(X_2 = aX_1 + b\right) = 1$$

The Bivariate Normal Distribution

Definition 4.8

Let $\mu_1 \in \mathbb{R}$, $\mu_2 \in \mathbb{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$ *and* $-1 < \rho < 1$ *be five numbers. The joint probability denstiy function*

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right)\right\}$$

for any ordered pair $(x_1, x_2) \in \mathbb{R}^2$ *is called the Bivariate Normal joint pdf.*

Properties of the Bivariate Normal

Theorem 4.8

Suppose that the random vector (X_1, X_2) has the bivariate normal pdf given in Definition 4.8 then

1. *The correlation coefficient between X_1 and X_2 is equal to ρ.*

2. *The marginal distribution for X_1 is the normal(μ_1, σ^2) and similarly $X_2 \sim$ normal(μ_2, σ_2^2).*

3. *The conditional distribution of $X_1 \mid X_2 = x_2$ is a normal with mean and variance*

$$\begin{aligned}\mathrm{E}(X_1 \mid X_2 = x_2) &= \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2) \\ \mathrm{Var}(X_1 \mid X_2 = x_2) &= \sigma_2^2(1 - \rho^2)\end{aligned}$$

The Bivariate Normal and Independence

Theorem 4.9

Suppose that (X_1, X_2) *is a bivariate normal random vector with correlation coefficient* ρ *then* $\rho = 0$ *if, and only if,* X_1 *and* X_2 *are independent random variables.*

* Note that it is possible for two normal random variables to have correlation zero but not be independent.

* What the above theorem says it is not possible for a bivariate normal random vector to have zero correlation and the components be independent.

* Marginal normality, however, does not imply bivariate normality.

Transformations of Discrete Bivariate Random Vectors

* Suppose that (X_1, X_2) is a bivariate random vector and we are interested in the random vector (Y_1, Y_2) given by

$$Y_1 = g_1(X_1, X_2) \qquad Y_2 = g_2(X_1, X_2)$$

for two functions g_1, g_2 which map $\mathbb{R}^2$ to $\mathbb{R}$.

* We can find the joint probability mass function in the discrete case similar to how we did for univariate random variables

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= \mathsf{P}\Big(g_1(X_1, X_2) = y_1, g_2(X_1, X_2) = y_2\Big) \\ &= \sum\sum_{A(y_1,y_2)} f_{X_1,X_2}(x_1, x_2) \end{aligned}$$

where $A(y_1, y_2) = \{(x_1, x_2) \in \mathcal{X} : g_1(x_1, x_2) = y_1, g_2(x_1, x_2) = y_2\}$.

Transformations of Discrete Bivariate Random Vectors

* In many instances we are actually interested in moving from a bivariate random vector (X_1, X_2) to a univariate random variable $Y = g(X_1, X_2)$.

* We can do this by first finding the joint pmf for the random vector defined by (Y, X_2) and then summing over all possible values of x_2 to get the marginal pmf for Y.

* In the discrete case, however, we can simplify that to

$$f_Y(y) = \sum_{A(y)}\sum f_{X_1,X_2}(x_1, x_2)$$

where $A(y) = \{(x_1, x_2) \in \mathcal{X} : g(x_1, x_2) = y\}$.

Transformations of Continuous Bivariate Random Vectors

* For continuous random vectors we shall assume that the transformation $(X_1, X_2) \to (Y_1, Y_2)$ is one-to-one.

* For such functions we can define the inverse transformation

$$X_1 = h_1(Y_1, Y_2) \qquad X_2 = h_2(Y_1, Y_2)$$

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

$$J = \begin{vmatrix} \dfrac{\partial h_1(y_1, y_2)}{\partial y_1} & \dfrac{\partial h_1(y_1, y_2)}{\partial y_2} \\ \dfrac{\partial h_2(y_1, y_2)}{\partial y_1} & \dfrac{\partial h_2(y_1, y_2)}{\partial y_2} \end{vmatrix} = \frac{\partial x_1}{\partial y_1} \times \frac{\partial x_2}{y_2} - \frac{\partial x_2}{\partial y_1} \times \frac{\partial x_1}{\partial y_2}$$

Transformations of Continuous Bivariate Random Vectors

Theorem 4.10

Let (X_1, X_2) *be a continuous bivariate random vector with joint pdf* $f_{X_1,X_2}(x_1, x_2)$ *on support* $\mathcal{A} = \{(x, y \in \mathbb{R} : f_{X_1,X_2}(x_1, x_2) > 0\}$. *Let the functions* $g_1(x_1, x_2)$ *and* $g_2(x_1, x_2)$ *define a one-to-one transformation of* $\mathcal{A}$ *to*

$\mathcal{B} = \{(y_1, y_2) : g_1(x_1, x_2) = y_1, g_2(x_1, x_2) = y_2$ *for some* $(x_1, x_2) \in \mathcal{A}\}$

and let the inverse transformation be given by $x_1 = h_1(y_1, y_2)$, $x_2 = h_2(y_1, y_2)$. *Then the pdf of the random vector* (Y_1, Y_2) *where* $Y_1 = g_1(X_1, X_2)$ *and* $Y_2 = g_2(X_1, X_2)$ *is given by*

$$f_{Y_1,Y_2}(y_1, y_2) \;=\; f_{X_1,X_2}(h_1(y_1, y_2), h_2(y_1, y_2))|J|$$

where J *is the Jacobian of the transformation given on the previous slide*

Piecewise Transformations

* 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 a common $\mathcal{B}$.

* We then apply the previous theorem to each set in the partition and sum the results to get the pdf of (Y_1, Y_2).

* The partition may also include a set A_0 such that $P((X_1, X_2) \in A_0) = 0$ without changing the result.

Hierarchical Models

* 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

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

Mixture Distributions

* The marginal distribution of X in this case is called a mixture distribution.
* Mixture distributions often have rather formidable looking pmfs or pdfs.
* Theorem 4.3 gives us simple ways to find the moments of X using the hierarchical structure.

Finite Mixture Distributions

Definition 4.9

A random variable X is said to have a finite mixture distribution if its pdf or pmf can be written as

$$f_X(x) \;=\; \sum_{j=1}^{k} p_i f_i(x)$$

where $0 < p_j$ $j = 1, \ldots, k$, $\sum_1^k p_j = 1$ *and each f_i is a pdf (or each f_i is a pmf).*

* Note that this can be thought of as a hierarchical model if Y is discrete with pmf

$$\mathsf{P}(Y = y) \;=\; p_y \qquad y = 1, \ldots, k$$

and the conditional distribution has pdf (pmf)

$$f_{X|Y}(x \mid y) \;=\; f_y(x) \qquad y = 1, \ldots, k$$

Countable Mixture Distributions

Definition 4.10

A random variable X is said to have a countable mixture distribution if there exists a discrete probability mass function $f_Y(y)$ and a sequence of conditional density or mass functions $f_{X|Y}(x \mid y)$ such that the marginal density (or mass) for X is given by

$$f_X(x) \;=\; \sum_y f_{X|Y}(x \mid y) f_Y(y).$$

* The pmf $f_Y(y)$ is usually called the mixing distribution.

* Most often, these mixture models arise when a "parameter" of the distribution of X is not fixed but is itself a random variable.

Uncountable Mixture Distributions

* We can generalize mixture distributions by allowing $f_Y(y)$ to be a probability density function.

* Then we get the mixture pdf (pmf) to be

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

* Some books call these compound distributions but they are really just mixture distributions with a continuous mixing distribution.

* They typically arise from a hierarchical structure in which the conditional distribution of X depends on a random “parameter” which is defined over an uncountable set (such as an interval).

General Multivariate Distributions

* A multivariate random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a function mapping a 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

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

  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 $f_{\boldsymbol{X}}(\boldsymbol{x})$ such that for any $A \subset \mathbb{R}^n$

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

Expectations, Marginal and Conditional Distributions

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

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

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

* Conditional pmfs (pdfs) of some subset of the components given the rest is found by dividing the full joint pmf (pdf) by the marginal pmf (pdf) of the conditioning components 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 4.11

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

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

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

Properties of the Multinomial Distribution

Theorem 4.11

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 4.12

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* $i \neq j$*,*

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

Multivariate Independence

Definition 4.12

Let $X_1, \ldots, X_n$ be random variables with joint pmf or pdf $f_{\boldsymbol{X}}(x_1, \ldots, x_n)$ 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 $(x_1, \ldots, x_n) \in \mathbb{R}^n$

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

Theorem 4.13

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

$$f_{\boldsymbol{X}}(x_1, \ldots, x_n) = \prod_{i=1}^{n} g_i(x_i).$$

Properties of Independent Random Variables

Theorem 4.14

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

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

Theorem 4.15

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 $Y = a_1 X_1 + \cdots + a_n X_n + b$ is

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

Multivariate Transformations

Definition 4.13

Suppose that there is a one-to-one transformation

$$G\,(x_1,\ldots,x_n) \;=\; \Big(g_1(\boldsymbol{x}),\ldots,g_n(\boldsymbol{x})\Big) \;=\; (y_1,\ldots,y_n)$$

with inverse

$$H(y_1,\ldots,y_n) \;=\; \Big(h_1(\boldsymbol{y}),\ldots,h_n(\boldsymbol{y})\Big) \;=\; \;=\; (x_1,\ldots,x_n).$$

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

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

Multivariate Transformations

Theorem 4.16

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

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