2.7 JOINT PROBABILITY DISTRIBUTIONS

Consider a collection of random variables $X_1, ..., X_k$, with X_i a function from sample space Ω_i to $\mathbb{X}_i \subseteq \mathbb{R}$. Then the vector of random variables $X = (X_1, ..., X_k)$ is a vector function from the Cartesian product of the sample spaces to $\mathbb{X}^{(k)}$

$$\Omega_1 \times ... \times \Omega_k \longrightarrow \mathbb{X}^{(k)} \subseteq \mathbb{X}_1 \times ... \times \mathbb{X}_k \subseteq \mathbb{R}^k$$

The vector random variable $X = (X_1, ..., X_k)$ could correspond to the outcomes of k different experiments carried out once, or a single experiment carried out k times.

Definition 2.7.1 JOINT MASS FUNCTION

The **joint probability mass function** of the k dimensional discrete random variable, that is, $X = (X_1, ..., X_k)$ is denoted $f_{X_1,...,X_k}$ and is defined by

$$f_{X_1,...,X_k}(x_1, ..., x_k) = P[(X_1 = x_1) \cap ... \cap (X_k = x_k)] = P[X_1 = x_1, ..., X_k = x_k]$$

for all possible values of the real vector, $x = (x_1, ..., x_k) \in \mathbb{X}^{(k)}$.

Note: As for the single variable case, the joint mass function must be non-negative for all $x \in \mathbb{X}^{(k)}$, and the sum of the joint mass function evaluated for all $x \in X^{(k)}$ must be 1.

Definition 2.7.2 JOINT DISCRETE CDF

The **joint cumulative distribution function** of k dimensional discrete random variable $X = (X_1, ..., X_k)$ is denoted $F_{X_1,...,X_k}$ and is defined by for any real vector $(x_1, ..., x_k)$ by

$$F_{X_1,...,X_k}(x_1, ..., x_k) = P[X_1 \leq x_1, ..., X_k \leq x_k]$$

Note: The joint cdf must satisfy certain properties (i.e. behaviour for limiting values of $x_1, ..., x_k$ individually and jointly) analogous to the univariate case.

Definition 2.7.3 MARGINAL MASS FUNCTION

The **marginal mass function** of random variable X_i, denoted f_{X_i}, is defined in terms of the joint mass function $f_{X_1,...,X_k}$ for $x_i \in \mathbb{X}_i$ by

$$f_{X_i}(x_i) = \sum_{\mathbb{X}_1} ... \sum_{\mathbb{X}_{i-1}} \sum_{\mathbb{X}_{i+1}} ... \sum_{\mathbb{X}_k} f_{X_1,...,X_k}(x_1, ..., x_k)$$

that is, the summation of joint mass function evaluated at $(x_1, ..., x_k)$ for all values of $x_j \in \mathbb{X}_j$ for $j \neq i$.

Definition 2.7.4 JOINT CONTINUOUS CDF

The **joint cumulative distribution function** of k dimensional continuous random variable, that is, $X = (X_1, ..., X_k)$ is denoted $F_{X_1,...,X_k}$ and is defined for any real vector $(x_1, ..., x_k)$ by

$$F_{X_1,...,X_k}(x_1, ..., x_k) = P[X_1 \leq x_1, ..., X_k \leq x_k]$$

Note: The joint cdf must satisfy certain properties (i.e. behaviour for limiting values of $x_1, ..., x_k$ individually and jointly) analogous to the univariate case.

Definition 2.7.5 JOINT PDF

The **joint probability density function** of k dimensional continuous random variable $(X_1, ..., X_k)$ is denoted $f_{X_1,...,X_k}$, and is defined in terms of the joint cdf $F_{X_1,...,X_k}$ for vector $(x_1, ..., x_k) \in \mathbb{X}^{(k)}$ by

$$F_{X_1,...,X_k}(x_1, ..., x_k) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_k} f_{X_1,...,X_k}(t_1, ..., t_k) dt_1...dt_k$$

so that

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \frac{\partial^k}{\partial t_1...\partial t_k} \{\{F_{X_1,...,X_k}(t_1, ..., t_k)\}\}_{t_1=x_1,...,t_k=x_k}$$

Note : As for the single variable case, the joint pdf need not exist, but if it does exist, the joint pdf must be take non-negative values for all $(x_1, ..., x_k) \in \mathbb{X}^{(k)}$, and

$$\int_{\mathbb{X}_1} ... \int_{\mathbb{X}_k} f_{X_1,...,X_k}(x_1, ..., x_k) dx_1...dx_k = 1$$

Definition 2.7.6 MARGINAL PDF

The **marginal probability density function** of random variable X_i, denoted f_{X_i}, is defined in terms of the joint pdf $f_{X_1,...,X_k}$ for $x_i \in \mathbb{X}_i$ by

$$f_{X_i}(x_i) = \int_{\mathbb{X}_1} ... \int_{\mathbb{X}_{i-1}} \int_{\mathbb{X}_{i+1}} ... \int_{\mathbb{X}_k} f_{X_1,...,X_k}(x_1, ..., x_k) dx_1...dx_{i-1}dx_{i+1}...dx_k$$

that is, the joint pdf integrated out over the ranges of the remaining $k-1$ variables $X_j, j \neq i$.

Note In both discrete and continuous cases, the concept of marginalization can be extended from consideration of the marginal probability distribution of a single variable. For example, consider the pair of variables (X_i, X_j) for $i \neq j$; the joint marginal mass function/pdf can be obtained by summation/integration over the remaining $k-2$ variables. This can be further extended to consideration of more than two variables.

Definition 2.7.7 CONDITIONAL MASS/DENSITY FUNCTION

The **conditional probability mass/density function** of random variable X_i, **given** that

$$X_1 = x_1, , ..., X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, ..., X_k = x_k$$

is denoted

$$f_{X_i|X_1,...,X_{i-1},X_{i+1},...,X_k},$$

is defined by

$$f_{X_i|X_1,...,X_{i-1},X_{i+1},...,X_k}(x_i|x_1, ..., x_{i-1}, x_{i+1}, ..., x_k) = \frac{f_{X_1,...,X_k}(x_1, ..., x_k)}{f_{X_1,...,X_{i-1},X_{i+1},...,X_k}(x_1, ..., x_{i-1}, x_{i+1}, ..., x_k)}$$

if the denominator is strictly positive, that is, the ratio of the (full) joint mass function/pdf for the k variables $X_1, ..., X_k$ to the (marginal) joint mass function/pdf for the $k-1$ variables $X_1, ..., X_{i-1}, X_{i+1}, ..., X_k$. The form of the denominator is obtained by summing/integrating the (full) joint mass function/pdf out over the variable X_i.

Note : This definition is directly related to the definition of conditional probability given earlier, and can be extended to the definition of conditional distribution of two variables, given the values of the remaining $k-2$, and so on.

2.7.1 JOINT DISTRIBUTION SPECIAL CASE: $k = 2$

We consider in detail the case when $k = 2$; the extension to higher order multivariate distributions is straightforward. Suppose that X and Y are random variables with ranges $\mathbb{X}$ and $\mathbb{Y}$ respectively, so that the vector (X, Y) is a vector random variable with range (contained in)

$$\mathbb{X} \times \mathbb{Y} = \{(x, y) : x \in \mathbb{X} \text{ and } y \in \mathbb{Y}\} \subseteq \mathbb{R}^2$$

DISCRETE CASE

If X and Y are **DISCRETE** the joint probability mass function of (X, Y), denoted $f_{X,Y}$, is

$$f_{X,Y}(x, y) = P[(X = x) \cap (Y = y)] = P[X = x, Y = y]$$

for all possible values of the vector $(x, y) \in \mathbb{X} \times \mathbb{Y}$. The joint cumulative distribution function of (X, Y), denoted $F_{X,Y}$, is

$$F_{X,Y}(x, y) = P[X \leq x, Y \leq y]$$

for **any** real vector (x, y).

The joint mass function $f_{X,Y}(x, y)$ essentially specifies a two-way table of probabilities. For example, suppose that

$$\mathbb{X} = \{1, 2, 3, 4, 5, 6\} \qquad \mathbb{Y} = \{1, 2, 3, 4\}$$

and let $p_{xy} = f_{X,Y}(x, y)$ Then we have the following table

Y	4	p_{14}	p_{24}	p_{35}	p_{44}	p_{54}	p_{64}
	3	p_{13}	p_{23}	p_{33}	p_{43}	p_{53}	p_{63}
	2	p_{12}	p_{22}	p_{32}	p_{42}	p_{52}	p_{62}
	1	p_{11}	p_{21}	p_{31}	p_{41}	p_{51}	p_{61}
		1	2	3	4	5	6
							X

The relationship between $f_{X,Y}$ and $F_{X,Y}$ is given by

$$F_{X,Y}(x, y) = \sum_{t=-\infty}^{[x]} \sum_{s=-\infty}^{[y]} f_{X,Y}(t, s)$$

where $[x]$ is the largest value in $\mathbb{X}$ not greater than x etc.

PROPERTIES

The axioms of probability automatically require that for a valid probability model, we must have

PMF $$0 \leq f_{X,Y}(x, y) \leq 1$$

$$\sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} f_{X,Y}(x, y) = 1$$

CDF $$0 \leq F_{X,Y}(x, y) \leq 1$$

$$\lim_{x \longrightarrow -\infty} F_{X,Y}(x, y) = 0$$

$$\lim_{y \longrightarrow -\infty} F_{X,Y}(x, y) = 0$$

$$\lim_{x \longrightarrow \infty, y \longrightarrow \infty} F_{X,Y}(x, y) = 1$$

$F_{X,Y}(x, y)$ is *non-decreasing* in both x and y

CONTINUOUS CASE

If X and Y are **CONTINUOUS** the **joint cumulative distribution function** (joint cdf) of (X, Y), denoted $F_{X,Y}$, is

$$F_{X,Y}(x, y) = P[X \leq x, Y \leq y]$$

for **any** real vector (x, y). The **joint probability distribution function** (joint pdf) of (X, Y), denoted $f_{X,Y}$, is

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t, s)\, ds\, dt \qquad \Leftrightarrow \qquad f_{X,Y}(x, y) = \frac{\partial^2}{\partial t \partial s} \left\{ F_{X,Y}(t, s) \right\}_{t=x, s=y}$$

for any real vector (x, y). The joint cdf and joint pdf are merely two real-valued functions of **two** variables.

PROPERTIES

The axioms of probability automatically require that for a valid probability model, we must have

PDF
$$0 \leq f_{X,Y}(x,y)$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

CDF
$$0 \leq F_{X,Y}(x,y) \leq 1$$
$$\lim_{x \longrightarrow -\infty} F_{X,Y}(x,y) = 0$$
$$\lim_{y \longrightarrow -\infty} F_{X,Y}(x,y) = 0$$
$$\lim_{x \longrightarrow \infty, y \longrightarrow \infty} F_{X,Y}(x,y) = 1$$

$F_{X,Y}(x,y)$ is *non-decreasing* in both x and y

In general, we can evaluate probabilities of events of interest by summing/integrating over appropriate regions of $\mathbb{R}^2$. For example, we may wish to evaluate

$$P[X+Y<6] \qquad P[a<XY<b] \qquad P[X<Y]$$

etc. which we can formulate generally as

$$P[g(X,Y) \in B]$$

for some function g and set B, and then evaluate as

$$P[g(X,Y) \in B] = \int_A \int f_{X,Y}(x,y)\,dx\,dy$$

where A is the region of $\mathbb{R}^2$ defined by

$$A = \{(x,y) : g(x,y) \in B,\, x \in \mathbb{X}\,,\, y \in \mathbb{Y}\}$$

For example,

$$P[X<Y] = P[X-Y<0] = \int_A \int f_{X,Y}(x,y)\,dx\,dy$$

where $A = \{(x,y) : x-y<0, x \in \mathbb{X}\,,\, y \in \mathbb{Y}\}$

MARGINAL DISTRIBUTIONS

The joint probability model expressed though $f_{X,Y}(x,y)$ or $F_{X,Y}(x,y)$ **automatically** specifies the probability model for each variable individually.

In the discrete case, the **marginal mass function** of random variable X, denoted f_X, is defined in terms of $f_{X,Y}$ for $x \in \mathbb{X}$ as

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

that is, the summation of joint mass function evaluated at (x,y) for **all** values of $y \in \mathbb{Y}$. Similarly, the **marginal mass function** for Y is

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

Essentially, the marginal distributions are obtained by summing out over the y in *column* x of the two-way table for $f_X(x)$, and summing out over x in *row* y for $f_Y(y)$. These results are a consequence of the Theorem of Total Probability; that is

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

Note that both f_X and f_Y are themselves probability mass functions, so must behave according to the rules specified in earlier sections.

In the continuous case, the **marginal probability density function** of random variable X, denoted f_X, is defined in terms of $f_{X,Y}$ for $x \in \mathbb{X}$ by

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

that is, the joint density function integrated out over y for a *fixed* value of x. Similarly, the **marginal probability density function** for Y is

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

Again, both f_X and f_Y are probability density functions, so must satisfy the required properties .

CONDITIONAL PROBABILITY DISTRIBUTIONS

The **conditional probability mass/density function** of random variable X given that $Y = y$ is denoted $f_{X|Y}(x|y)$ and is defined by

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

(that is, a function of argument x for **fixed** y) if the denominator is strictly positive, that is, the ratio of the joint mass function/pdf to the marginal mass/density function for Y. A similar definition gives the conditional of Y given $X = x$,

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

EXAMPLE In the continuous case, with $k = 3$, each of the following marginal and conditional density functions can be computed from the joint density function

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3)$$

MARGINALS:

$$f_{X_1}(x_1) = \int\limits_{\mathbb{X}_2}\int\limits_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3)dx_2dx_3 \qquad f_{X_2}(x_2) = \int\limits_{\mathbb{X}_1}\int\limits_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3)dx_1dx_3$$

$$f_{X_3}(x_3) = \int\limits_{\mathbb{X}_1}\int\limits_{\mathbb{X}_2} f_{X_1,X_2,X_3}(x_1, x_2, x_3)dx_1dx_2$$

$$f_{X_1,X_2}(x_1, x_2) = \int\limits_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3)dx_3 \qquad f_{X_1,X_3}(x_1, x_3) = \int\limits_{\mathbb{X}_2} f_{X_1,X_2,X_3}(x_1, x_2, x_3)dx_2$$

$$f_{X_2,X_3}(x_2, x_3) = \int\limits_{\mathbb{X}_1} f_{X_1,X_2,X_3}(x_1, x_2, x_3)dx_1$$

CONDITIONALS:

$$f_{X_1,X_2|X_3}(x_1, x_2|x_3) = \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_3}(x_3)} \qquad f_{X_1,X_3|X_2}(x_1, x_3|x_2) = \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_2}(x_2)}$$

$$f_{X_2,X_3|X_1}(x_2, x_3|x_1) = \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_1}(x_1)}$$

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)} \qquad f_{X_1|X_3}(x_1|x_3) = \frac{f_{X_1,X_3}(x_1, x_3)}{f_{X_3}(x_3)}$$

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)} \qquad f_{X_2|X_3}(x_2|x_3) = \frac{f_{X_2,X_3}(x_2, x_3)}{f_{X_3}(x_3)}$$

$$f_{X_3|X_1}(x_3|x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)} \qquad f_{X_2|X_3}(x_2|x_3) = \frac{f_{X_2,X_3}(x_2, x_3)}{f_{X_3}(x_3)}$$

$$f_{X_3|X_1}(x_3|x_1) = \frac{f_{X_1,X_3}(x_1, x_3)}{f_{X_1}(x_1)} \qquad f_{X_3|X_2}(x_3|x_2) = \frac{f_{X_2,X_3}(x_2, x_3)}{f_{X_2}(x_2)}$$

$$f_{X_1|X_2,X_3}(x_1|x_2, x_3) = \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_2,X_3}(x_2, x_3)} \qquad f_{X_2|X_1,X_3}(x_2|x_1, x_3) = \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_1,X_3}(x_1, x_3)}$$

$$f_{X_3|X_1,X_2}(x_3|x_1, x_2) = \frac{f_{X_1,X_2,X_3}(x_1, x_2, x_3)}{f_{X_1,X_2}(x_1, x_2)}$$

2.7.2 INDEPENDENCE FOR RANDOM VARIABLES

Random variables $X_1, ...X_k$ are **independent** if, for $a_i < b_i$, $i = 1, ...k$,

$$P[a_1 < X_1 \leq b_1, ..., a_k < X_k \leq b_k] = \prod_{i=1}^{k} P[a_i < X_i \leq b_i]$$

More specifically, $X_1, ...X_k$ are independent **if and only if**

$$F_{X_1,...,X_k}(x_1, ..., x_k) = \prod_{i=1}^{k} F_{X_i}(x_i) \qquad \text{or equivalently, } f_{X_1,...,X_k}(x_1, ..., x_k) = \prod_{i=1}^{k} f_{X_i}(x_i)$$

that is, if and only if the joint cdf and the joint mass function/pdf factorizes into the k marginal cdfs and mass function/pdfs.

SUFFICIENT CONDITIONS FOR INDEPENDENCE

Sufficient conditions for variables $X_1, ..., X_k$ to be independent are

(i) the support of the joint mass function/pdf (the region on which the function is strictly positive) is a Cartesian product, that is,

$$\mathbb{X}^{(k)} = \mathbb{X}_1 \times ... \times \mathbb{X}_k = \{(x_1, ..., x_k) : x_1 \in \mathbb{X}_1, ..., x_k \in \mathbb{X}_k\}.$$

(ii) the joint mass function/pdf factorizes into a product of marginal functions.

2.7.3 THE CHAIN RULE FOR RANDOM VARIABLES

There is an explicit relationship between joint, marginal, and conditional mass/density functions. For example, consider three continuous random variables X_1, X_2, X_3, with joint pdf f_{X_1,X_2,X_3}. Then,

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) f_{X_3|X_1,X_2}(x_3|x_1, x_2)$$

so that, for example,

$$\begin{aligned} f_{X_1}(x_1) &= \int_{\mathbb{X}_2} \int_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3) dx_2 dx_3 \\ &= \int_{\mathbb{X}_2} \int_{\mathbb{X}_3} f_{X_1|X_2,X_3}(x_1|x_2, x_3) f_{X_2,X_3}(x_2, x_3) dx_2 dx_3 \\ &= \int_{\mathbb{X}_2} \int_{\mathbb{X}_3} f_{X_1|X_2,X_3}(x_1|x_2, x_3) f_{X_2|X_3}(x_2|x_3) f_{X_3}(x_3) dx_2 dx_3 \end{aligned}$$

Equivalent relationships hold in the discrete case, can be extended to determine the explicit relationship between joint, marginal, and conditional mass/density functions for any number of random variables.

NOTE: the discrete equivalent of this result is a DIRECT consequence of the Theorem of Total Probability; the event $[X_1 = x_1]$ is partitioned into sub-events $[(X_1 = x_1) \cap (X_2 = x_2) \cap (X_3 = x_3)]$ for all possible values of the pair (x_2, x_3).

2.7.4 CONDITIONAL EXPECTATION AND ITERATED EXPECTATION

Consider two discrete/continuous random variables X_1 and X_2 with joint mass function/pdf f_{X_1,X_2}, and the conditional mass function/pdf of X_1 given $X_2 = x_2$, defined in the usual way by

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

Then the **conditional expectation** of X_1 given $X_2 = x_2$ is defined by

$$E_{f_{X_1|X_2}}[X_1|X_2 = x_2] = \begin{cases} \displaystyle\sum_{x_1 \in \mathbb{X}_1} x_1 f_{X_1|X_2}(x_1|x_2) & X_1 \text{ DISCRETE} \\ \displaystyle\int_{\mathbb{X}_1} x_1 f_{X_1|X_2}(x_1|x_2)dx_1 & X_1 \text{ CONTINUOUS} \end{cases}$$

i.e. the expectation of X_1 *with respect to* the conditional density of X_1 given $X_2 = x_2$, (possibly giving a function of x_2).

THE LAW OF ITERATED EXPECTATION
THEOREM

For two continuous random variables X_1 and X_2 with joint pdf f_{X_1,X_2},

$$E_{f_{X_1}}[X_1] = E_{f_{X_2}}\left[E_{f_{X_1|X_2}}[X_1|X_2 = x_2]\right]$$

PROOF

$$\begin{aligned} E_{f_{X_1}}[X_1] &= \int_{\mathbb{X}_1} x_1 f_{X_1}(x_1)dx_1 \\ &= \int_{\mathbb{X}_1} x_1 \left\{\int_{\mathbb{X}_2} f_{X_1,X_2}(x_1,x_2)dx_2\right\} dx_1 \\ &= \int_{\mathbb{X}_1} x_1 \left\{\int_{\mathbb{X}_2} f_{X_1|X_2}(x_1|x_2)f_{X_2}(x_2)dx_2\right\} \\ &= \int_{\mathbb{X}_1}\int_{\mathbb{X}_2} x_1 f_{X_1|X_2}(x_1|x_2)f_{X_2}(x_2)dx_2dx_1 \\ &= \int_{\mathbb{X}_2} \left\{\int_{\mathbb{X}_1} x_1 f_{X_1|X_2}(x_1|x_2)dx_1\right\} f_{X_2}(x_2)dx_2 \\ &= \int_{\mathbb{X}_2} \left\{E_{f_{X_1|X_2}}[X_1|X_2 = x_2]\right\} f_{X_2}(x_2)dx_2 = E_{f_{X_2}}\left[E_{f_{X_1|X_2}}[X_1|X_2 = x_2]\right] \end{aligned}$$

so the expectation of X_1 can be calculated by finding the conditional expectation of X_1 given $X_2 = x_2$, giving a function of x_2, and then taking the expectation of this function with respect to the marginal density for X_2. Note that this proof only works if the conditional expectation and the marginal expectation are finite. This results extends naturally to k variables.

2.8 MULTIVARIATE TRANSFORMATIONS

THEOREM

Let $\mathbf{X} = (X_1, ..., X_k)$ be a vector of random variables, with joint mass/density function $f_{X_1,...,X_k}$. Let $\mathbf{Y} = (Y_1, ..., Y_k)$ be a vector of random variables defined by $Y_i = g_i(X_1, ..., X_k)$ for some functions $g_i, i = 1, ..., k$, where the vector function $\mathbf{g}$ mapping $(X_1, ..., X_k)$ to $(Y_1, ..., Y_k)$ is a 1-1 transformation. Then the joint mass/density function of $(Y_1, ..., Y_k)$ is given by

DISCRETE $\quad f_{Y_1,...,Y_k}(y_1, ..., y_k) = f_{X_1,...,X_k}(x_1, ..., x_k)$

CONTINUOUS $\quad f_{Y_1,...,Y_k}(y_1, ..., y_k) = f_{X_1,...,X_k}(x_1, ..., x_k)\,|J(y_1, ..., y_k)|$

where $\mathbf{x} = (x_1, ..., x_k)$ is the unique solution of $\mathbf{y} = \mathbf{g}(\mathbf{x})$, so that $\mathbf{x} = \mathbf{g}^{-1}(\mathbf{y})$, and where $J(y_1, ..., y_k)$ is the **Jacobian**, of the transformation, that is, the determinant of the $k \times k$ matrix whose (i, j)th element is

$$\frac{\partial}{\partial t_j}\left\{{g_i}^{-1}(\mathbf{t})\right\}_{t_1=y_1,...,t_k=y_k}$$

where g_i^{-1} is the inverse function uniquely defined by $X_i = g_i^{-1}(Y_1, ..., Y_k)$.

PROOF

Discrete case proof follows univariate case precisely. For the continuous case, consider the **equivalent** events $[\mathbf{X} \in C]$ and $[\mathbf{Y} \in D]$, where D is the image of C under $\mathbf{g}$. Clearly, $P[\mathbf{X} \in C] = P[\mathbf{Y} \in D]$. Now, $P[\mathbf{X} \in C]$ is the k dimensional integral of the joint density $f_{X_1,...,X_k}$ over the set C, and $P[\mathbf{Y} \in D]$ is the k dimensional integral of the joint density $f_{Y_1,...,Y_k}$ over the set D. Result follows by changing variables in the first integral from $\mathbf{x}$ to $\mathbf{y} = \mathbf{g}(\mathbf{x})$, and equating the two integrands.

Note : As for single variable transformations, the ranges of the transformed variables must be considered carefully.

Example 2.8.1 Consider the case $k = 2$, and suppose that X_1 and X_2 are *independent* continuous random variables with ranges $\mathbb{X}_1 = \mathbb{X}_2 = [0, 1]$ and pdfs given respectively by

$$\begin{aligned} f_{X_1}(x_1) &= 6x_1(1-x_1) \quad 0 \le x_1 \le 1 \\ f_{X_2}(x_2) &= 3x_2^2 \quad\quad\quad\;\; 0 \le x_2 \le 1 \end{aligned}$$

and zero elsewhere. In order to calculate the pdf of random variable Y_1 defined

$$Y_1 = X_1X_2$$

using the transformation result, consider the additional random variable Y_2, where $Y_2 = X_1$ (note, as X_1 and X_2 take values on $[0, 1]$, $X_1 \ge X_1X_2$ so $Y_1 \le Y_2$).

The transformation $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ is then specified by the two functions

$$g_1(t_1, t_2) = t_1t_2 \qquad g_2(t_1, t_2) = t_1$$

and the inverse transformation $\mathbf{X} = \mathbf{g}^{-1}(\mathbf{Y})$ (i.e. $\mathbf{X}$ in terms of $\mathbf{Y}$) is

$$X_1 = Y_2 \qquad X_2 = Y_1/Y_2$$

giving

$$g_1^{-1}(t_1, t_2) = t_2 \qquad g_2^{-1}(t_1, t_2) = t_1/t_2.$$

Hence

$$\begin{array}{ll} \dfrac{\partial}{\partial t_1}\left\{g_1{}^{-1}(t_1, t_2)\right\} = 0 & \dfrac{\partial}{\partial t_2}\left\{g_1{}^{-1}(t_1, t_2)\right\} = 1 \\ \dfrac{\partial}{\partial t_1}\left\{g_2{}^{-1}(t_1, t_2)\right\} = 1/t_2 & \dfrac{\partial}{\partial t_2}\left\{g_2{}^{-1}(t_1, t_2)\right\} = -t_1/t_2^2 \end{array}$$

and so the Jacobian $J(y_1, y_2)$ of the transformation is given by the modulus of

$$\left| \begin{array}{cc} 0 & 1 \\ 1/y_2 & -y_1/y_2^2 \end{array} \right|$$

so that $J(y_1, y_2) = 1/y_2$. Hence, using the theorem

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}(y_2, y_1/y_2) \times |J(y_1, y_2)| \\ &= 6y_2(1-y_2) \times 3(y_1/y_2)^2 \times 1/y_2 \\ &= 18y_1^2(1-y_2)/y_2^2 \end{aligned}$$

on the set $\mathbb{Y}^{(2)} = \{(y_1, y_2) : 0 \le y_1 \le y_2 \le 1\}$, and zero otherwise. Hence

$$\begin{aligned} f_{Y_1}(y_1) &= \textstyle\int_{y_1}^1 18y_1^2(1-y_2)/y_2^2 dy_2 \\ &= 18y_1^2 \left[-1/y_2 - \log y_2\right]_{y_1}^1 \\ &= 18y_1^2(-1 + 1/y_1 + \log y_1) \\ &= 18y_1(1 - y_1 + y_1 \log y_1) \end{aligned}$$

for $0 \le y_1 \le 1$, and zero otherwise.

2.9 MULTIVARIATE EXPECTATIONS AND COVARIANCE

2.9.1 EXPECTATION WITH RESPECT TO JOINT DISTRIBUTIONS

Definition 2.9.1 For random variables $X_1, ..., X_k$ with range $\mathbb{X}^{(k)}$ with mass/density function $f_{X_1,...,X_k}$, the **expectation** of $g(X_1, ..., X_k)$ is defined in the discrete and continuous cases by

$$E_{f_{X_1,...,X_k}}[g(X_1, ..., X_k)] = \begin{cases} \sum_{x_1} ... \sum_{x_k} g(x_1, ..., x_k) f_{X_1,...,X_k}(x_1, ..., x_k) \\ \int_{\mathbb{X}_1} ... \int_{\mathbb{X}_k} g(x_1, ..., x_k) f_{X_1,...,X_k}(x_1, ..., x_k) dx_1 ... dx_k \end{cases}$$

PROPERTIES

(i) Let g and h be real-valued functions and let a and b be constants. Then, if $f_{\mathbf{X}} \equiv f_{X_1,...,X_k}$,

$$E_{f_{\mathbf{X}}}[ag(X_1, ..., X_k) + bh(X_1, ..., X_k)] = aE_{f_{\mathbf{X}}}[g(X_1, ..., X_k)] + bE_{f_{\mathbf{X}}}[h(X_1, ..., X_k)].$$

(ii) Let $X_1, ...X_k$ be **independent** random variables with mass functions/pdfs $f_{X_1}, ..., f_{X_k}$ respectively. Let $g_1, ..., g_k$ be scalar functions of $X_1, ..., X_k$ respectively (that is, g_i is a function of X_i *only* for $i = 1, ..., k$). If $g(X_1, ..., X_k) = g_1(X_1)...g_k(X_k)$, then

$$E_{f_{\mathbf{X}}}[g(X_1, ..., X_k)] = \prod_{i=1}^{k} E_{f_{X_i}}[g_i(X_i)]$$

where $E_{f_{X_i}}[g_i(X_i)]$ is the marginal expectation of $g_i(X_i)$ with respect to f_{X_i}.

2.9.2 COVARIANCE AND CORRELATION

Definition 2.9.2 The <u>**covariance**</u> of two random variables X_1 and X_2 is denoted $Cov_{f_{X_1,X_2}}[X_1, X_2]$, and is defined by

$$Cov_{f_{X_1,X_2}}[X_1, X_2] = E_{f_{X_1,X_2}}[(X_1 - \mu_1)(X_2 - \mu_2)] = E_{f_{X_1,X_2}}[X_1 X_2] - \mu_1 \mu_2$$

where $\mu_i = E_{f_{X_i}}[X_i]$ is the marginal expectation of X_i, for $i = 1, 2$, and where

$$E_{f_{X_1,X_2}}[X_1 X_2] = \int \int g(x_1, x_2) f_{X_1,X_2}(x_1, x_2) dx_1 dx_2$$

that is, the expectation of function $g(x_1, x_2) = x_1 x_2$ with respect to the joint distribution f_{X_1,X_2}.

Definition 2.9.3 The **correlation** of X_1 and X_2 is denoted $Corr_{f_{X_1,X_2}}[X_1, X_2]$, and is defined by

$$Corr_{f_{X_1,X_2}}[X_1, X_2] = \frac{Cov_{f_{X_1,X_2}}[X_1, X_2]}{\sqrt{Var_{f_{X_1}}[X_1] Var_{f_{X_2}}[X_2]}}$$

If $Cov_{f_{X_1,X_2}}[X_1, X_2] = Corr_{f_{X_1,X_2}}[X_1, X_2] = 0$. then variables X_1 and X_2 are **uncorrelated**.

Note that if random variables X_1 and X_2 are independent then

$$Cov_{f_{X_1,X_2}}[X_1,X_2] = E_{f_{X_1,X_2}}[X_1X_2] - E_{f_{X_1}}[X_1]E_{f_{X_2}}[X_2]$$

$$= E_{f_{X_1}}[X_1]E_{f_{X_2}}[X_2] - E_{f_{X_1}}[X_1]E_{f_{X_2}}[X_2] = 0$$

and so X_1 and X_2 are also uncorrelated (the converse does not hold).

NOTES:
(i) For random variables X_1 and X_2, with (marginal) expectations μ_1 and μ_2 respectively, and (marginal) variances σ_1^2 and σ_2^2 respectively, if random variables Z_1 and Z_2 are defined

$$Z_1 = (X_1 - \mu_1)/\sigma_1 Z_2 = (X_2 - \mu_2)/\sigma_2$$

that is, Z_1 and Z_2 are *standardized* variables. Then

$$Corr_{f_{X_1,X_2}}[X_1,X_2] = Cov_{f_{Z_1,Z_2}}[Z_1,Z_2].$$

(ii) Extension to k variables: covariances can only be calculated for *pairs* of random variables, but if k variables have a joint probability structure it is possible to construct a $k \times k$ *matrix*, $\mathbf{C}$ say, of covariance values, whose (i,j)th element is

$$Cov_{f_{X_i,X_j}}[X_i,X_j]$$

for $i,j = 1,..,k$, that captures the complete covariance structure in the joint distribution. If $i \neq j$, then

$$Cov_{f_{X_j,X_i}}[X_j,X_i] = Cov_{f_{X_i,X_j}}[X_i,X_j]$$

so $\mathbf{C}$ is *symmetric*, and if $i = j$,

$$Cov_{f_{X_i,X_i}}[X_i,X_i] \equiv Var_{f_{X_i}}[X_i]$$

The matrix $\mathbf{C}$ is referred to as the **variance-covariance** matrix.

(iii) If random variable X is defined by

$$X = \sum_{i=1}^{k} a_i X_i$$

for random variables $X_1, ..., X_k$ and constants $a_1, ..., a_k$, then

$$E_{f_X}[X] = \sum_{i=1}^{k} a_i E_{f_{X_i}}[X_i]$$

$$Var_{f_X}[X] = \sum_{i=1}^{k} a_i^2 Var_{f_{X_i}}[X_i] + 2\sum_{i=1}^{k}\sum_{j=1}^{i-1} a_i a_j Cov_{f_{X_i,X_j}}[X_i,X_j]$$

(iv) Combining (i) and (iii) when $k = 2$, and defining standardized variables Z_1 and Z_2,

$$\begin{aligned} 0 \leq Var_{f_{Z_1,Z_2}}[Z_1 \pm Z_2] &= Var_{f_{Z_1}}[Z_1] + Var_{f_{Z_2}}[Z_2] \pm 2Cov_{f_{Z_1,Z_2}}[Z_1, Z_2] \\ &= 1 + 1 \pm 2Corr_{f_{X_1,X_2}}[X_1, X_2] \\ &= 2(1 \pm Corr_{f_{X_1,X_2}}[X_1, X_2]) \end{aligned}$$

and hence $-1 \leq Corr_{f_{X_1,X_2}}[X_1, X_2] \leq 1$.

2.10 SUMS OF RANDOM VARIABLES - THE CONVOLUTION FORMULA

Suppose that X_1 and X_2 be continuous random variables taking values on $\mathbb{X}_1$ and $\mathbb{X}_2$ respectively with joint density function f_{X_1,X_2}. Then, if random variable Y is defined by $Y = X_1 + X_2$, the marginal density of Y is given by calculated as follows; consider the *multivariate transformation*

$$\mathbf{g} : (X_1, X_2) \longrightarrow (Y, Z) \qquad \text{so that} \qquad g(x_1, x_2) = (y, z) = (x_1 + x_2, x_1)$$

so that the inverse transformation is given by $(X_1, X_2) = (Z, Y - Z)$. Then by the multivariate transformation theorem,

$$f_{Y,Z}(y, z) = f_{X_1,X_2}(z, y - z)|J(y, z)| \text{ for } (y, z) \in B.$$

where

$$B = \{(y, z)|(y, z) = \mathbf{g}(x_1, x_2), x_1 \in \mathbb{X}_1, x_2 \in \mathbb{X}_2\}$$

$J(y, z)$ is the determinant of a 2×2 matrix of partial derivatives of the inverse functions

$$g_1{}^{-1}(t_1, t_2) = t_2 \qquad g_2{}^{-1}(t_1, t_2) = t_1 - t_2$$

which simply reduces to give $J(y, z) = -1$. Therefore

$$f_{Y,Z}(y, z) = f_{X_1,X_2}(z, y - z)$$

and so by the usual marginalization calculation

$$f_Y(y) = \int_{\mathbb{Z}} f_{Y,Z}(y, z)dz = \int_{\mathbb{X}_1} f_{X_1,X_2}(x_1, y - x_1)dx_1$$

If X_1 and X_2 are **independent**, then $f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ so if $Y = X_1 + X_2$, then Y has pdf given by

$$f_Y(y) = \int_{\mathbb{X}_1} f_{X_1}(x_1)f_{X_2}(y - x_1)dx_1$$

2.11 ORDER STATISTICS

For k random variables $X_1, ..., X_k$, the **order statistics**, $Y_1, ..., Y_k$, are defined by

$$Y_i = X_{(i)} - \text{“the i'th smallest value in } X_1, ..., X_k\text{”}$$

for $i = 1, ..., k$, so that $Y_1 = X_{(1)} = \text{Min}\{X_1, ..., X_k\}$, $Y_k = X_{(k)} = \text{Max}\{X_1, ..., X_k\}$.
For k independent, identically distributed random variables $X_1, ..., X_k$, with marginal density function f_X, there are two main results to consider;

RESULT 1 The joint density function of the order statistics $Y_1, ..., Y_k$ is given by

$$f_{Y_1,...,Y_k}(y_1, ..., y_k) = k! f_X(y_1)...f_X(y_k) y_1 < ... < y_k$$

RESULT 2 The marginal pdf of the jth order statistic Y_j for $j = 1, ..., k$ has the form

$$f_{Y_j}(y_j) = \frac{k!}{(j-1)!(k-j)!} \{F_X(y_j)\}^{j-1} \{1 - F_X(y_j)\}^{k-j} f_X(y_j)$$

Special Cases: MAXIMUM and MINIMUM

To derive the marginal pdf of Y_k, first consider the marginal cdf of Y_k;

$$\begin{aligned}
F_{Y_k}(y_k) &= P[Y_k \leq y_k] = P[\max\{X_1, ..., X_k\} \leq y_k] = P[X_1 \leq y_k, X_2 \leq y_k, ..., X_k \leq y_k] \\
&= \prod_{i=1}^{k} P[X_i \leq y_k] = \prod_{i=1}^{k} \{F_X(y_k)\} \\
&= \{F_X(y_k)\}^k \\
\Longrightarrow f_{Y_k}(y_k) &= k \{F_X(y_k)\}^{k-1} f_X(y_k)
\end{aligned}$$

By a similar calculation, we can find the marginal pdf/cdf for Y_1,

$$\begin{aligned}
F_{Y_1}(y_1) &= P[Y_1 \leq y_1] = 1 - P[Y_1 > y_1] = 1 - P[\min\{X_1, ..., X_k\} > y_1] \\
&= 1 - P[X_1 > y_1, X_2 > y_1, ..., X_k > y_1] \\
&= 1 - \prod_{i=1}^{k} P[X_i > y_1] = 1 - \prod_{i=1}^{k} \{1 - F_X(y_1)\} \\
&= 1 - \{1 - F_X(y_1)\}^k \\
\Longrightarrow f_{Y_1}(y_1) &= k \{1 - F_X(y_1)\}^{k-1} f_X(y_1)
\end{aligned}$$