2.4. FUNCTIONS OF RANDOM VARIABLES

Consider a discrete/continuous random variable X with range $\mathbb{X}$ and probability distribution described by mass/pdf f_X, or cdf F_X. Suppose g is a real-valued function whose domain includes $\mathbb{X}$, and suppose that

$$g: \begin{array}{l} \mathbb{X} \longrightarrow \mathbb{Y} \\ x \longmapsto y \end{array}$$

Then $Y = g(X)$ is also a random variable as Y is a function from Ω to $\mathbb{R}$.

For $A \subseteq \mathbb{R}$, the event $[\ Y \in A\ ]$ is an event in terms of the transformed variable Y. If f_Y is the mass/density function for Y, then

$$\mathrm{P}[\ Y \in A\ ] = \begin{cases} \displaystyle\sum_{y \in A} f_Y(y) & Y \text{ discrete} \\ \\ \displaystyle\int_A f_Y(y)\ dy & Y \text{ continuous} \end{cases}$$

We wish to derive the probability distribution of random variable Y; in order to do this, we first consdier the inverse transformation g^{-1} from $\mathbb{Y}$ to $\mathbb{X}$ defined for set $A \subseteq \mathbb{Y}$ (and for $y \in \mathbb{Y}$) by

$$g^{-1}(A) = \{\ x \in \mathbb{X}\ :\ g(x) \in A\} \qquad g^{-1}(y) = \{\ x \in \mathbb{X}\ :\ g(x) = y\}$$

that is, $g^{-1}(A)$ is the set of points in $\mathbb{X}$ that map into A, and $g^{-1}(y)$ is the set of points in $\mathbb{X}$ that map to y, under transformation g. By construction, we have

$$\mathrm{P}[\ Y \in A\ ] = \mathrm{P}[\ X \in g^{-1}(A)\ ]$$

and hence $[\ Y \in A\ ]$ and $[\ X \in g^{-1}(A)\ ]$ are **equivalent** events.

Consider first the cdf of Y, F_Y, evaluated at a point $y \in \mathbb{R}$. We have

$$F_Y(y) = \mathrm{P}[\ Y \leq y\ ] = \ \mathrm{P}[\ g(X) \leq y\ ] = \begin{cases} \displaystyle\sum_{x \in A_y} f_X(x) & \text{if } X \text{ is discrete} \\ \\ \displaystyle\int_{A_y} f_X(x)\ dx & \text{if } X \text{ is continuous} \end{cases}$$

where $A_y = \{\ x \in \mathbb{X}\ :\ g(x) \leq y\}$. Often, the set is A_y is easy to identify for a given y, that is, $F_Y(y) = \mathrm{P}[\ g(X) \leq y\ ] = \mathrm{P}[\ x_1 \leq X \leq\ x_2]$, where x_1 and x_2 depend on y and g or g^{-1}. The main objective is therefore to identify the set A_y.

1-1 Transformations

The mapping $g(X)$ is a function of X from $\mathbb{X}$ which is *1-1 and onto* $\mathbb{Y}$ if,

(i) for each $x \in \mathbb{X}$, there exists one and only one y such that $y = g(x)$, and

(ii) for each $y \in \mathbb{Y}$, there exists an $x \in \mathbb{X}$ such that $g(x) = y$.

The following theorem gives the distribution for random variable $Y = g(X)$ when g is 1-1.

THEOREM

Let X be a random variable with mass/density function f_X and support $\mathbb{X}$. Let g be a 1-1 function from $\mathbb{X}$ onto $\mathbb{Y}$ with inverse g^{-1}. Then $Y = g(X)$ is a random variable with support $\mathbb{Y}$ and

Discrete Case : The mass function of random variable Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \qquad y \in \mathbb{Y} = \{\ y \mid f_Y(y) > 0\ \}.$$

where x is the unique solution of $y = g(x)$ (so that $x = g^{-1}(y)$).

Continuous Case : The pdf of random variable Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} \right| \qquad y \in \mathbb{Y} = \{\ y \mid f_Y(y) > 0\ \}$$

where $y = g(x)$, provided that the derivative $\dfrac{d}{dt}\left\{g^{-1}(t)\right\}$ is *continuous* and *non-zero* on $\mathbb{Y}$.

PROOF

Discrete case: by direct calculation, $f_Y(y) = \mathrm{P}[\ Y = y\ ] = \mathrm{P}[\ g(X) = y\ ] = \mathrm{P}[\ X = g^{-1}(y)\ ] = f_X(x)$. where $x = g^{-1}(y)$, and hence $f_Y(y) > 0 \iff f_X(x) > 0$.

Continuous case: function g is either (I) a monotonic increasing, or (II) a monotonic decreasing function.

Case (I): If g is *increasing*, then for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$, we have that $g(x) \leq y \iff x \leq g^{-1}(y)$. Therefore, for $y \in \mathbb{Y}$,

$$F_Y(y) = \mathrm{P}[\ Y \leq y\ ] = \mathrm{P}[\ g(X) \leq y\ ] = \mathrm{P}[\ X \leq g^{-1}(y)\ ] = F_X(g^{-1}(y))$$

and, by differentiation, because g is monotonic increasing,

$$f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} = f_X(g^{-1}(y)) \left| \frac{d}{dy}\left\{g^{-1}(y)\right\}_{t=y} \right| \qquad \text{as } \frac{d}{dt}\left\{g^{-1}(t)\right\} > 0.$$

Case (II): If g is *decreasing*, then for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$ we have $g(x) \leq y \iff x \geq g^{-1}(y)$. Therefore, for $y \in \mathbb{Y}$,

$$F_Y(y) = \mathrm{P}[\ Y \leq y\ ] = \mathrm{P}[\ g(X) \leq y\ ] = \mathrm{P}[\ X \geq g^{-1}(y)\ ] = 1 - F_X(g^{-1}(y))$$

so

$$f_Y(y) = -f_X(g^{-1}(y)) \frac{d}{dt}\left\{g^{-1}(y)\right\} = f_X(g^{-1}(y)) \left| \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} \right| \qquad \text{as } \frac{d}{dt}\left\{g^{-1}(t)\right\} < 0.$$

DEFINITION

Suppose transformation $g : \mathbb{X} \longrightarrow \mathbb{Y}$ is 1-1, and is defined by $g(x) = y$ for $x \in \mathbb{X}$. Then the **Jacobian** of the transformation, denoted $J(y)$, is given by

$$J(y) = \left| \frac{d}{dt}\left\{g^{-1}(t)\right\}_{t=y} \right|$$

that is, the absolute value of first derivative of g^{-1} evaluated at $y = g(x)$. Note that the inverse transformation $g^{-1} : \mathbb{Y} \longrightarrow \mathbb{X}$ has Jacobian $1/J(x)$.

2.5. 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 $\mathbf{X} = (X_1, ..., X_k)$ 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** $\mathbf{X} = (X_1, ..., X_k)$ could correspond to the outcomes of k different experiments carried out once, or a single experiment carried out k times.

DEFINITION (JOINT MASS FUNCTION)

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

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

for all possible values of the random vector, $\mathbf{x} = (x_1, ..., x_k) \in \mathbb{X}^{(k)}$. As for the single variable case, the joint mass function must be **non-negative** for all $\mathbf{x} \in \mathbb{X}^{(k)}$, and the sum of the joint mass function evaluated for all $\mathbf{x} \in \mathbb{X}^{(k)}$ must be 1.

DEFINITION (JOINT DISCRETE CDF)

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

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

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 (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

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

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 (JOINT CONTINUOUS CDF)

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

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

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 (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 (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.

CONDITIONAL PROBABILITY DISTRIBUTIONS

DEFINITION

The **conditional probability mass/density function** of random variable X_i, given that $X_1 = x_1, , ..., X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, ..., X_k = x_k$ 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.

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:

$$f_{X_1}(x_1) = \int_{\mathbb{X}_2}\int_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_2 dx_3 \qquad f_{X_2}(x_2) = \int_{\mathbb{X}_1}\int_{\mathbb{X}_3} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_1 dx_3$$

$$f_{X_3}(x_3) = \int_{\mathbb{X}_1}\int_{\mathbb{X}_2} f_{X_1,X_2,X_3}(x_1, x_2, x_3)\ dx_1 dx_2$$

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

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

CONDITIONALS:

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

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

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

DEFINITION (INDEPENDENCE FOR RANDOM VARIABLES)

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

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

In fact, $X_1, ...X_k$ are independent **if and only if**

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

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 variables $X_1, ..., X_k$ to be independent are

(i) the **support** of the joint mass function/pdf of $\mathbf{X}^{(k)}$ is a Cartesian product, that is,

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

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

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,

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

so that, for example,

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

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.

CONDITIONAL EXPECTATION

DEFINITION

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

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

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

$$\mathrm{E}_{f_{X_1|X_2}}[X_1|X_2 = x_2] \quad = \quad \begin{cases} \displaystyle\sum_{x_1 \in \mathbb{X}_1} x_1 f_{X_1|X_2}(x_1|x_2) & X_1 \text{ discrete} \\ \\ \displaystyle\int_{\mathbb{X}_1} x_1 f_{X_1|X_2}(x_1|x_2)\ dx_1 & X_1 \text{ continuous} \end{cases}$$

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

THE LAW OF ITERATED EXPECTATION

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

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

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.

2.6. 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 case:

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

Continuous case:

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

where $\mathbf{x} = (x_1, ..., x_k)$ is the unique solution of $\mathbf{y} = \mathrm{g}(\mathbf{x})$, 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 modulus of 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, $\mathrm{P}[\ \mathbf{X} \in C\ ] = \mathrm{P}[\ \mathbf{Y} \in \mathbf{D}\ ]$. Now, $\mathrm{P}[\ \mathbf{X} \in C\ ]$ is the k dimensional integral of the joint density $f_{X_1,...,X_k}$ over the set C, and $\mathrm{P}[\ \mathbf{Y} \in D\ ]$ 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 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

$$\begin{array}{rcll} f_{X_1}(x_1) & = & 6x_1(1-x_1) & \qquad 0 \leq x_1 \leq 1 \\ f_{X_2}(x_2) & = & 3x_2^2 & \qquad 0 \leq x_2 \leq 1 \end{array}$$

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

$$Y_1 = X_1 X_2$$

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 $[0, 1]$, $X_1 \geq X_1 X_2$ so $Y_1 \leq Y_2$).

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

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

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

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

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

$$\begin{vmatrix} 0 & 1 \\ 1/y_2 & -y_1/y_2^2 \end{vmatrix}$$

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

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

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

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

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