Multivariate RVs and their distributions

Probability calculus / Adv Stat I

Prof. Dr. Matei Demetrescu

Increasing the numbers

Random variables allow us to include stochastic components in mathematical models.

The scalar case is often just the beginning,

... as we may need more than just one stochastic component.

So we have to discuss ways of dealing with several random variables

at the same time.

Today's outline

Multivariate RVs and their distributions

Outline

What if mapping to $\mathbb{R}^K$?

Definition (Multivariate random variable)

Let $\{\mathcal{S}, \mathcal{Y}, \mathrm{P}\}$ be a probability space. If $X : \mathcal{S} \to \mathbb{R}^n$ (or simply, X) is a real-valued vector function having as its domain the elements of $\mathcal{S}$, then $X : \mathcal{S} \to \mathbb{R}^n$ (or X) is called a multivariate (n-variate) random variable.

The realized value of the multivariate random variable is

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix} = \boldsymbol{X}(\omega) \quad \text{for} \quad \omega \in \mathcal{S},$$

and its range is

$$\mathrm{R}(\boldsymbol{X}) = \{(x_1, ..., x_n) : x_i = X_i(\omega), i = 1, ..., n, \omega \in \mathcal{S}\}.$$

More complicated pdfs

Definition (Discrete multivariate pdf)

A multivariate random variable $\boldsymbol{X} = (X_1, ..., X_n)$ is called discrete iff its range $\mathrm{R}(\boldsymbol{X})$ is countable. The discrete joint pdf of a discrete random vector $\boldsymbol{X}$, denoted by f, is defined by

$$f : \mathbb{R}^n \to [0, 1] \quad \text{such that}$$

$$f(x_1, ..., x_n) = \begin{cases} \mathrm{P}_{\boldsymbol{X}}(X_1 = x_1, ..., X_n = x_n) & \text{if} \quad (x_1, ..., x_n) \in \mathrm{R}(\boldsymbol{X}) \\ 0 & \text{else.} \end{cases}$$

And the continuous case

Definition (Continuous multivariate pdf)

A multivariate random variable $\boldsymbol{X} = (X_1, ..., X_n)$ is called continuous iff

- its range $\mathrm{R}(\boldsymbol{X})$ is uncountably infinite and
- there exists a function

$$f : \mathbb{R}^n \to [0, \infty) \quad \text{such that for any event A,}$$

$$\mathrm{P}_{\boldsymbol{X}}(A) = \int \cdots \int_{(x_1,...,x_n)\in A} f(x_1, ..., x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

and

$$f(x_1, ..., x_n) = 0 \;\forall\; (x_1, ..., x_n) \notin \mathrm{R}(\boldsymbol{X}).$$

The function f is called a continuous joint pdf.

Requirements

Definition (Class of discrete joint pdfs)

The function $f : \mathbb{R}^n \to \mathbb{R}$ is a member of the class of discrete joint pdfs iff

(i$_a$) the set $C = \{(x_1, ..., x_n) : f(x_1, ..., x_n) > 0, (x_1, ..., x_n) \in \mathbb{R}^n\}$ is countable;

(ii$_a$) $f(x_1, ..., x_n) = 0 \;\forall\; (x_1, ..., x_n) \in \bar{C}$;

(iii$_a$) $\sum \cdots \sum_{(x_1,...,x_n)\in C} f(x_1, ..., x_n) = 1$.

Definition (Class of continuous joint pdfs)

The function $f : \mathbb{R}^n \to \mathbb{R}$ is a member of the class of continuous joint pdfs iff

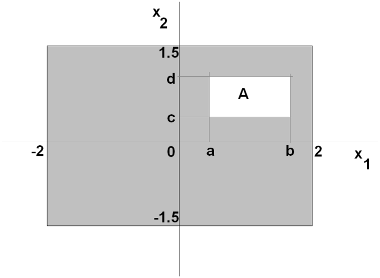
(i$_b$) $f(x_1, ..., x_n) \geq 0 \;\forall\; (x_1, ..., x_n) \in \mathbb{R}^n$;

(ii$_b$) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, ..., x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n = 1$.

An example from outer space I

Consider that the NASA announces that a small meteorite will hit a rectangular area of 12km^2=4km$\times$3km.

- Define $\boldsymbol{X} = (X_1, X_2)$ to be the coordinates of the point of strike, with a range $\mathrm{R}(\boldsymbol{X}) = \{(x_1, x_2) : x_1 \in [-2, 2], x_2 \in [-1.5, 1.5]\}$.
- Each point in that rectangle is equally likely to be struck.
- To obtain the continuous pdf of $\boldsymbol{X}$, consider a rectangle A in $\mathrm{R}(\boldsymbol{X})$:



An example from outer space II

Since all points are equally likely, we obtain

$$\mathrm{P}_{\boldsymbol{X}}(\boldsymbol{X} \in A) = \frac{\text{area of A}}{\text{area of } \mathrm{R}(\boldsymbol{X})} = \frac{(b-a)(d-c)}{12}.$$

According to the definition, the pdf f for $\boldsymbol{X}$ has to satisfy

$$\int_c^d \int_a^b f(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2 \stackrel{!}{=} \frac{(b-a)(d-c)}{12},$$

$\forall\ -2 \leq a \leq b \leq 2;\ -1.5 \leq c \leq d \leq 1.5$, with

$$\frac{\partial^2 \left[\int_c^d \int_a^b f(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2\right]}{\partial d \partial b} = f(b, d) \stackrel{!}{=} \frac{\partial^2 \left[(b-a)(d-c)/12\right]}{\partial d \partial b} = \frac{1}{12},$$

$\forall\ b \in [-2, 2],\ d \in [-1.5, 1.5]$.

An example from outer space III

Hence, the function

$$f(x_1, x_2) = \frac{1}{12}\mathbb{I}_{[-2,2]}(x_1)\mathbb{I}_{[-1.5,1.5]}(x_2)$$

can be used as a joint pdf for $\boldsymbol{X}$, and for any event $A \in \mathrm{R}(\boldsymbol{X})$ we obtain $\mathrm{P}_{\boldsymbol{X}}(A) = \int\int_{x \in A} \frac{1}{12}\mathrm{d}x_1\mathrm{d}x_2$.

Multivariate cdfs

Definition (Joint cdf)

The joint cdf of an n-dimensional random variable $\boldsymbol{X}$, denoted by F, is defined by

$$F : \mathbb{R}^n \to [0,1] \quad \text{with} \quad F(b_1, ..., b_n) = \mathrm{P}_{\boldsymbol{X}}(X_1 \leq b_1, ..., X_n \leq b_n),$$

$\forall (b_1, ..., b_n) \in \mathbb{R}^n$.

For a discrete random variable the joint cdf obtains as

$$F(b_1, ..., b_n) = \sum_{x_1 \leq b_1} \cdots \sum_{x_n \leq b_n} f(x_1, ..., x_n), \quad \forall (b_1, ..., b_n) \in \mathbb{R}^n,$$

and for a continuous random variable as

$$F(b_1, ..., b_n) = \int_{-\infty}^{b_n} \cdots \int_{-\infty}^{b_1} f(x_1, ..., x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n, \quad \forall (b_1, ..., b_n) \in \mathbb{R}^n.$$

Some properties

Theorem (2.4)

For any multivariate cdf F, we have that

(i) $\lim_{b_i \to -\infty} F(b_1, ..., b_n) = \mathrm{P}_{\boldsymbol{X}}(\emptyset) = 0,$ *for some* $i = 1, ..., n;$

(ii) $\lim_{b_i \to \infty, \forall i} F(b_1, ..., b_n) = \mathrm{P}_{\boldsymbol{X}}(\mathrm{R}(\boldsymbol{X})) = 1;$

(iii) *F is a non decreasing function on $(x_1, ..., x_n)$, that is, $F(\boldsymbol{a}) \leq F(\boldsymbol{b})$ for (the vector inequality)*

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} < \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \boldsymbol{b};$$

(iv) *Discrete joint cdfs have a countable number of jump discontinuities and joint cdfs for continuous random variables are continuous without jump discontinuities.*

Joint pdfs

Theorem (2.5)

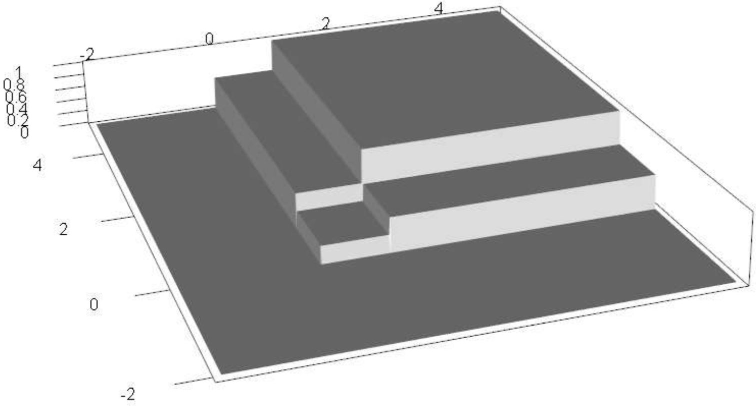
Let (X, Y) *be a discrete bivariate random variable with joint cdf* $F(x, y)$ *and range* $\mathrm{R}(X, Y) = \{x_1 < x_2 < x_3 < \cdots, y_1 < y_2 < y_3 < \cdots\}$. *Then the joint pdf obtains as*

$$\begin{aligned}
f(x_1, y_1) &= F(x_1, y_1), \\
f(x_1, y_j) &= F(x_1, y_j) - F(x_1, y_{j-1}), \quad j \geq 2, \\
f(x_i, y_1) &= F(x_i, y_1) - F(x_{i-1}, y_1), \quad i \geq 2. \\
\textit{For } i, j &\geq 2, \\
f(x_i, y_j) &= F(x_i, y_j) - F(x_i, y_{j-1}) - F(x_{i-1}, y_j) + F(x_{i-1}, y_{j-1}).
\end{aligned}$$

The result of the theorem for the bivariate case can be generalized to the n-variate case. (Cumbersome and omitted.)

Coin tosses

The cdf of two (0/1) fair coin tosses is as follows



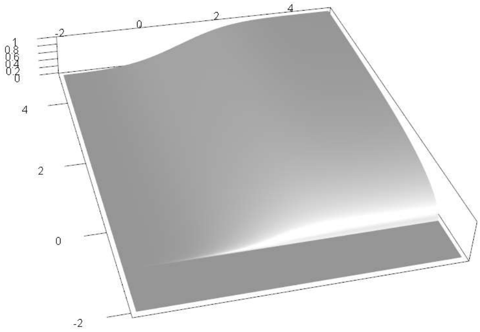
And the continuous case

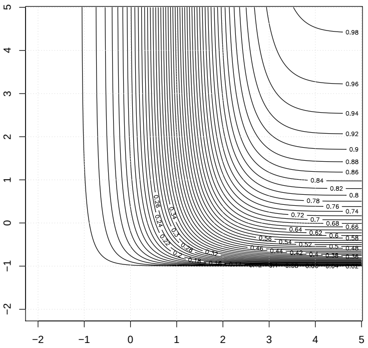
Theorem (2.6)

Let $f(x_1, ..., x_n)$ and $F(x_1, ..., x_n)$ denote the joint pdf and cdf for a continuous multivariate random variable $\boldsymbol{X} = (X_1, ..., X_n)$. Then the joint pdf for $\boldsymbol{X}$ obtains as

$$f(x_1, ..., x_n) = \begin{cases} \frac{\partial^n F(x_1,...,x_n)}{\partial x_1 \cdots \partial x_n}, & \textit{wherever } f(\cdot) \textit{ is continuous} \\ 0, & \textit{elsewhere.} \end{cases}$$

An anonymous continuous example





Left: 3d plot; right: contour plot.

Outline

One component alone

Take for instance the joint distribution of $(X_1, X_2)'$,

$X_1 \setminus X_2$	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$
$x_{1,1}$	0.2	0.2	0.05
$x_{1,2}$	0.1	0.2	0.25

Cells contain joint probabilities of $(X_1 = x_{1,i}, X_2 = x_{2,j})$.

For this distribution, what is the probability that $X_1 = x_{1,1}$ irrespective of X_2?

And how do we obtain the distribution of X_1 in general?

One component alone

Theorem (2.7)

Let $\boldsymbol{X} = (X_1, X_2)$ *be a discrete random variable with joint pdf* $f(x_1, x_2)$ *and a range* $\mathrm{R}(\boldsymbol{X}) = \mathrm{R}(X_1) \times \mathrm{R}(X_2)$. *The marginal pdfs are given by*

$$f_1(x_1) = \sum_{x_2 \in \mathrm{R}(X_2)} f(x_1, x_2), \quad \text{and} \quad f_2(x_2) = \sum_{x_1 \in \mathrm{R}(X_1)} f(x_1, x_2).$$

To obtain the marginal pdf, we simply "sum out" the variables that are not of interest.

The continuous case

Theorem (2.8)

Let $\boldsymbol{X} = (X_1, X_2)$ be a continuous random variable with joint pdf $f(x_1, x_2)$. The corresponding marginal pdfs are given by

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) \mathrm{d}x_2, \quad \textit{and} \quad f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) \mathrm{d}x_1.$$

The concept of marginal pdfs can be straightforwardly generalized from the bivariate to the n-variate case as follows.

Many components

Definition (Marginal pdfs)

Let $f\left(x_1, \ldots, x_n\right)$ be the joint pdf for the n-dimensional random variable $\left(X_1, \ldots, X_n\right)$. Let $J = \left\{j_1, j_2, \ldots, j_m\right\}, 1 \leq m < n$, be a set of indices selected from the index set $I = \{1, 2, \ldots, n\}$. Then the marginal density function for the m-dimensional random variable $\left(X_{j_1}, \ldots, X_{j_m}\right)$ is given by

$$f_{j_1 \ldots j_m}\left(x_{j_1}, \ldots, x_{j_m}\right) = \begin{cases} \displaystyle\sum_{\left(x_i \in \mathrm{R}(X_i), i \in I-J\right)} \cdots \sum f\left(x_1, \ldots, x_n\right) & \text{(discr.)}. \\ \displaystyle\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(x_1, \ldots, x_n\right) \prod_{i \in I-J} \mathrm{d}x_i & \text{(cont.)}. \end{cases}$$

An abstract example

Consider the continuous random variable $\boldsymbol{X} = (X_1, X_2)$ with a joint pdf

$$f(x_1, x_2) = (x_1 + x_2)\mathbb{I}_{[0,1]}(x_1)\mathbb{I}_{[0,1]}(x_2).$$

The corresponding marginal pdf of X_1 obtains as

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{\infty} f(x_1, x_2)\mathrm{d}x_2 = \int_0^1 (x_1 + x_2)\mathbb{I}_{[0,1]}(x_1)\mathrm{d}x_2 \\ &= \left[\left(x_1 x_2 + \frac{x_2^2}{2}\right)\mathbb{I}_{[0,1]}(x_1)\right]_{x_2=0}^{x_2=1} = \left(x_1 + \frac{1}{2}\right)\mathbb{I}_{[0,1]}(x_1). \end{aligned}$$

Outline

Assume that you know something...

Recall the joint distribution

	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$
$x_{1,1}$	0.2	0.2	0.05
$x_{1,2}$	0.1	0.2	0.25

E.g., the distribution of X_1 given $X_2 = x_2, 3$ is the 2-point distribution with

$$\begin{aligned} \mathrm{P}(X_1 = x_{1,1} | X_2 = x_{1,3}) &= 1/6 \\ \mathrm{P}(X_1 = x_{1,2} | X_2 = x_{2,3}) &= 5/6. \end{aligned}$$

So we can in principle derive the conditional pdf of X_1 given X_2, which can be used to assign the probability to the event $X_1 \in C$ given that (conditional on) $X_2 \in D$.

The discrete case

If (X_1, X_2) is a discrete random variable, the conditional pdf for X_1 given $X_2 \in D$ can be defined by

$$f(x_1|x_2 = d) = \frac{f(x_1, d)}{f_2(d)}$$

if D is a single point d, and, in general, by

$$f(x_1|x_2 \in D) = \frac{\sum_{x_2 \in D} f(x_1, x_2)}{\sum_{x_2 \in D} f_2(x_2)}.$$

From the conditional pdf we can straightforwardly derive the conditional cdf by using the conditional pdf in the general definition of a cdf. This holds in the continuous case as well, btw.

The continuous tentative

If (X_1, X_2) is a continuous random variable, we can substitute the summation operations by integrations, such that the conditional pdf for X_1 given $x_2 \in D$ is defined as

$$f(x_1|x_2 \in D) = \frac{\int_{x_2 \in D} f(x_1, x_2)\mathrm{d}x_2}{\int_{x_2 \in D} f_2(x_2)\mathrm{d}x_2}.$$

However, a problem arises when D is a single point d, such that

$$f(x_1|x_2 = d) = \frac{\int_d^d f(x_1, x_2)\mathrm{d}x_2}{\int_d^d f_2(x_2)\mathrm{d}x_2} = \frac{0}{0},$$

which is undefined!

This problem is circumvented by redefining the conditional probability in the continuous case in terms of a limit.

Concretely

Consider the definition

$$\mathrm{P}(X_1 \in A | X_2 = d) \equiv \lim_{\epsilon \downarrow 0} \mathrm{P}(X_1 \in A | d - \epsilon \leq X_2 \leq d + \epsilon).$$

After some algebra, we get

$$\mathrm{P}(X_1 \in A | X_2 = d) = \int_{x_1 \in A} \frac{f(x_1, d)}{f_2(d)} \mathrm{d}x_1.$$

Since pdfs are defined exactly this way, the desired conditional density must be the above integrand.

Hence the conditional pdf of X_1 given $X_2 = d$ in the continuous case can be defined as

$$f(x_1 | x_2 = d) = \frac{f(x_1, d)}{f_2(d)}.$$

Note that it has exactly the same form as in the discrete case.

Example

Consider the continuous random variable $\boldsymbol{X} = (X_1, X_2)$ with joint pdf

$$f(x_1, x_2) = (x_1 + x_2)\mathbb{I}_{[0,1]}(x_1)\mathbb{I}_{[0,1]}(x_2),$$

and marginal pdf (see above):

$$f_2(x_2) = \left(x_2 + \frac{1}{2}\right)\mathbb{I}_{[0,1]}(x_2).$$

Then the conditional pdf of X_1 given $X_2 \leq .5$ obtains as

$$\begin{aligned} f(x_1|x_2 \leq .5) \quad &\overset{(def.)}{=} \quad \frac{\int_{-\infty}^{.5} f(x_1, x_2)\mathrm{d}x_2}{\int_{-\infty}^{.5} f_2(x_2)\mathrm{d}x_2} = \frac{\int_{-\infty}^{.5}(x_1 + x_2)\mathbb{I}_{[0,1]}(x_1)\mathbb{I}_{[0,1]}(x_2)\mathrm{d}x_2}{\int_{-\infty}^{.5}\left(x_2 + \frac{1}{2}\right)\mathbb{I}_{[0,1]}(x_2)\mathrm{d}x_2} \\ &= \quad \left(\frac{4}{3}x_1 + \frac{1}{3}\right)\mathbb{I}_{[0,1]}(x_1). \end{aligned}$$

The conditional pdf of X_1 given $X_2 = .75$ is

$$f(x_1|x_2 = .75) \overset{(def.)}{=} \frac{f(x_1, .75)}{f_2(.75)} = \left(\frac{4}{5}x_1 + \frac{3}{5}\right)\mathbb{I}_{[0,1]}(x_1).$$

Multivariate case

Definition (Conditional pdfs)

Let $f(x_1, \ldots, x_n)$ be the joint pdf for the n-dimensional random variable $(X_1, \ldots, X_n)$. Let $J_1 = \{j_1, \ldots, j_m\}$ and $J_2 = \{j_{m+1}, \ldots, j_n\}$ be two mutually exclusive index sets whose union is equal to the index set $\{1, 2, \ldots, n\}$. Then the conditional pdf for the m-dimensional random variable $(X_{j_1}, \ldots, X_{j_m})$, given $\left(X_{j_{m+1}} = d_{m+1}, \ldots, X_{j_n} = d_n\right)$ is given by

$$f\left(x_{j_1}, \ldots, x_{j_m} \mid x_{j_i} = d_i, i = m+1, \ldots, n\right) = \frac{f(x_1, \ldots, x_n)}{f_{j_{m+1} \cdots j_n}(d_{m+1}, \ldots, d_n)}$$

where $x_{j_i} = d_i$ if $j_i \in J_2$, when the marginal density in the denominator is positive valued.

From events to random variables

The independence of two events A and B means that $\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B)$.

(How) Does this extend to random variables?

Definition (Independence of Random Variables)

The random variables X_1 and X_2 are said to be independent iff

$$\mathrm{P}(X_1 \in A_1, X_2 \in A_2) = \mathrm{P}(X_1 \in A_1) \cdot \mathrm{P}(X_2 \in A_2)$$

for all $A_1 \subset \mathrm{R}(X_1)$ and $A_2 \subset \mathrm{R}(X_2)$.

Factorization

The definition is not immediately operational since the factorization has to hold for all pairs of events.

Theorem (2.9)

The random variables X_1 and X_2 with joint pdf $f(x_1, x_2)$ and marginal pdfs $f_1(x_1)$ and $f_2(x_2)$ are independent, iff

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2) \quad \forall (x_1, x_2),$$

(except possibly at points of discontinuity for a joint continuous pdf f).

Conditional and marginal distributions

An important implication of the independence of X_1 and X_2 is that the conditional pdfs are identical to the corresponding marginal pdfs, that is,

$$f(x_1|x_2 = d) \overset{(def.)}{=} \frac{f(x_1, d)}{f_2(d)} = \frac{f_1(x_1)f_2(d)}{f_2(d)} = f_1(x_1).$$

Thus the probability of event $X_1 \in A$ is unaffected by the occurrence or nonoccurrence of event $X_2 = d$.

Meteorite again

Recall the meteorite example, where $\boldsymbol{X} = (X_1, X_2)$ is the point of strike with joint pdf

$$f(x_1, x_2) = \frac{1}{12}\mathbb{I}_{[-2,2]}(x_1)\mathbb{I}_{[-1.5,1.5]}(x_2),$$

Are X_1 and X_2 independent? The marginal pdfs are

$$\begin{aligned} f_1(x_1) &= \frac{1}{12}\mathbb{I}_{[-2,2]}(x_1)\int_{-1.5}^{1,5} 1 \mathrm{d}x_2 = \frac{1}{4}\mathbb{I}_{[-2,2]}(x_1) \\ f_2(x_2) &= \frac{1}{12}\mathbb{I}_{[-1.5,1.5]}(x_2)\int_{-2}^{2} 1 \mathrm{d}x_1 = \frac{1}{3}\mathbb{I}_{[-1.5,1.5]}(x_2) \end{aligned}$$

Thus, $f(x_1, x_2) = f_1(x_1)f_2(x_2)$, and X_1 and X_2 are independent.

Implications of independence

If X_1 and X_2 are independent, then knowing the marginal pdfs f_1 and f_2 is sufficient to determine the joint pdf: $f(x_1, x_2) = f_1(x_1)f_2(x_2)$.

This is not true in general: consider for some $\alpha \in [-1, 1]$ the joint pdf

$$f(x_1, x_2; \alpha) = [1 + \alpha(2x_1 - 1)(2x_2 - 1)]\mathbb{I}_{[0,1]}(x_1)\mathbb{I}_{[0,1]}(x_2).$$

For any choice of $\alpha \in [-1, 1]$, the marginal pdfs are

$$f_1(x_1) = \mathbb{I}_{[0,1]}(x_1) \quad \text{and} \quad f_2(x_2) = \mathbb{I}_{[0,1]}(x_2).$$

Hence, for all suitable values of α in the joint pdf f, we obtain the very same marginal pdfs f_1 and f_2.

Thus, knowing f_1 and f_2 is insufficient to determine f and, in particular, the value of α.

The multivariate case

Definition (Independence in the n-variate case)

The random variables $X_1, \ldots, X_n$ are said to be independent iff

$$\mathrm{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} \mathrm{P}(X_i \in A_i), \quad \text{for all } A_i \subset \mathrm{R}(X_i).$$

Theorem (2.10)

The random variables $X_1, \ldots, X_n$ with joint pdf $f(x_1, \ldots, x_n)$ and marginal pdfs $f_i(x_i)$, $i = 1, \ldots, n$, are all independent of each other, iff

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i) \quad \forall (x_1, \ldots, x_n),$$

(except possibly at points of discontinuity for a joint continuous pdf f).

Outline

Coming up

Transformations of random variables