Advanced Statistics

2. Random Variables and their Probability Distributions

Christian Aßmann

Chair of Survey Statistics and Data Analysis – Otto-Friedrich-Universität Bamberg

2.1 Univariate Random Variables

In many experiments it is easier to deal with a summary variable than with the original probability structure.

For example, consider the experiment of tossing a coin 50 times.

- Typically, we are not interested in knowing which of the 2^{50} possible 50-tuples in sample space S has occurred.

  Rather we would like to know the number of heads in 50 tosses, which can be defined as a variable X.

- Note that the sample space of X is the set $\{0, 1, 2, \ldots, 50\}$ which is much easier to deal with than the original sample space S.

- By defining the variable X, we have defined a function from the original sample space S to a new sample space and, hence, we have defined a random variable.

DEFINITION (UNIVARIATE RANDOM VARIABLE): Let $\{S, Y, P\}$ be a probability space. If $X : S \to \mathbb{R}$ (or simply, X) is a real-valued function having as its domain the elements of S, then $X : S \to \mathbb{R}$ (or X) is called a random variable.

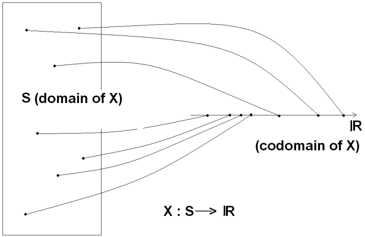


Fig. 5.

EXAMPLE: In some experiments random variables are implicitly used; Examples are:

Experiment	Random variable
Toss two dice	$X =$ sum of the numbers
Toss a coin 50 times	$X =$ number of heads in 50 tosses
Toss a coin 50 times	$X =$ squared number of heads in 50 tosses $\|\|$

Remark on notation:

$X(w)$	:	denotes the image of $w \in S$ generated by the random variable $X : S \to \mathbb{R}$.
$x = X(w)$	:	(realized) value of the function X

Uppercase letters (X) will be used to denote random variables and corresponding lowercase letters (x) will denote the realized values. $\diamondsuit$

Range of a random variable

Note that by defining a random variable, we have also defined a new sample space, namely, the range of the random variable.
This range, denoted by $R(X)$, obtains as the set of all x-values which can be generated on the sample space S using the function X:

$$R(X) = \{x : x = X(w), w \in S\}.$$

This raises the following important questions:

How can we embed the new sample space $R(X)$ within a probability space that can be used for assigning probabilities to events in terms of random-variable outcomes?

Hence, what is the probability function on $R(X)$, say P_X?

Induced probability function

- Suppose we have a discrete sample space

$$S = \{w_1, ..., w_n\} \quad \text{with a probability function } P \ .$$

- Now define a random variable

$$X(w) \quad \text{with range } R(X) = \{x_1, ..., x_m\}.$$

Assume that we observe $X = x_i$ iff the experiment's outcome is w_j such that

$$x_i = X(w_j).$$

- Since the elementary event $w_j \in S$ is equivalent to the event $x_j \in R(X)$, both events should have the same probability. Thus

$$P_X(X = x_i) = P(\{w_j : x_i = X(w_j), w_j \in S\}).$$

Note that the function P_X on the left-hand side is an induced probability set function on $R(X)$ defined in terms of the original function P.

EXAMPLE: Consider the experiment of tossing a fair coin two times.

- Define the random variable X to be the number of heads in the two tosses. Thus

Experiment's outcome $w \in S$	(H,H)	(H,T)	(T,H)	(T,T)
Variable's Realization $x = X(w)$	2	1	1	0

- The random variable's range is $R(X) = \{0, 1, 2\}$
- Since, for example, $P_X(X = 1) = P(\{H, T\}) + P(\{T, H\})$, the induced probability function on $R(X)$ obtains as

x	0	1	2
$P_X(X = x)$	1/4	1/2	1/4

||

Discrete Probability Density Function

REMARK: In practice, it is useful to have a representation of the induced probability set function, P_X, in a compact closed-form formula. This leads us to the definition of a probability density function. $\diamondsuit$

With every random variable, we associate a probability density function. Random variables can be either discrete or continuous. In the following, we start to consider discrete random variables and their probability density function.

> DEFINITION (DISCRETE RANDOM VARIABLE): A random variable X is called discrete iff its range $R(X)$ is countable.

Definition (Discrete probability density function): The discrete probability density function (pdf) of a discrete random variable X, denoted by f, is defined by

$$f : \mathbb{R} \to [0,1] \quad \text{such that} \quad f(x) = \begin{cases} P_X(X = x) & \text{if} \quad x \in R(X) \\ 0 & \text{else.} \end{cases}$$

Remark: Even though the range $R(X)$ of a discrete random variable consists of a countable number of elements, the domain of the pdf is the entire uncountable real line $\mathbb{R}$.
However, the value of $f(x)$ is set to zero at all points $x \notin R(X)$. This definition is adopted for the sake of convenience – it standardizes the domain for all random variables to be $\mathbb{R}$. $\diamondsuit$

Remark: The pdf allows us to obtain the probability for an event in $R(X)$.

- Consider the event $A \subset R(X)$, written as a union of elementary events $A = \cup_{x \in A}\{x\}$.
- Since elementary events are disjoint, we know from Axiom 1.3 that

$$P_X(A) = P_X(\cup_{x \in A}\{x\}) \overset{(Ax.3)}{=} \sum_{x \in A} P_X(x) = \sum_{x \in A} f(x).$$

- Thus, we can use the pdf to calculate probabilities for events on $R(X)$ by summing the probabilities of the elementary events given by the pdf. $\diamondsuit$

EXAMPLE: Consider the experiment of tossing two fair dice and observing the number of dots facing up.

- The sample space is $S = \{(i,j) : i = 1,\ldots,6; j = 1,\ldots,6\}$, where i,j are the number of dots. S consists of 36 elementary events.
- Define a random variable X to be the sum of the dots, such that $x = X((i,j)) = i + j$.

We obtain the following correspondence between outcomes of X and events in S:

	$x = X((i,j))$	$B_x = \{(i,j) : x = i + j, (i,j) \in S)\}$	$P_X(x) = f(x) = P(B_x)$
$R(X)$	2	$\{(1,1)\}$	1/36
	3	$\{(1,2),(2,1)\}$	2/36
	4	$\{(1,3),(2,2),(3,1)\}$	3/36
		$\vdots$	
	12	$\{(6,6)\}$	1/36

- Consider the event $x \in A = \{3,4\}$. The probability obtains as $P_X(A) = \sum_{x \in A} f(x) = f(3) + f(4) = 5/36$.
- A compact algebraic form for the pdf f is $f(x) = \frac{6-|x-7|}{36} \mathbb{I}_{\{2,3,\ldots,12\}}(x)$. ||

Continuous Probability Density Function

Problem: If the the range $R(X)$ is continuous with events A defined as intervals on $R(X)$, the summation operation over the element in A (i.e. $\sum_{x \in A}$) is not defined. Thus, defining a probability set function on events in $R(X)$ as $P_X(A) = \sum_{x \in A} f(x)$ will not be possible!

Solution: Substitute the summation operation $\sum_{x \in A}$ by integration $\int_{x \in A}$. This leads us to the following definition of a continuous probability density function:

Definition (**Continuous probability density function**): A random variable X is called continuous iff

- its range $R(X)$ is uncountably infinite and
- there exists a function

$$f : \mathbb{R} \to [0, \infty) \quad \text{such that for any event A,} \quad P_X(A) = \int_{x \in A} f(x)dx$$

and

$$f(x) = 0 \; \forall \; x \notin R(X).$$

The function f is called a continuous probability density function.

EXAMPLE: Consider a Formula 1 circuit of 10 km. Suppose that accidents are equally likely to occur at each point of the circuit.

- Define the continuous random variable X to be the point of a potential accident with range $R(X) = [0, 10]$.
- In order to obtain the pdf for X, consider the event A of an accident between two points a and b, such that $A = [a, b]$.
- Since all points are equally likely, we obtain

$$P_X(A) = \frac{\text{length of } A}{R(X)} = \frac{b-a}{10}$$

- According to the definition, the pdf f for X has to satisfy

$$\int_{x \in A} f(x)dx = \int_a^b f(x)dx \overset{!}{=} P_X(A) = \frac{b-a}{10}, \quad \forall \quad 0 \leq a \leq b \leq 10,$$

with

$$\frac{\partial[\int_a^b f(x)dx]}{\partial b} = f(b) \overset{!}{=} \frac{\partial[\frac{b-a}{10}]}{\partial b} = \frac{1}{10}, \quad \forall \quad b \in [0, 10].$$

(CONTINUES)

EXAMPLE (CONTINUED):

- Hence, the function

$$f(x) = \frac{1}{10}\mathbb{I}_{[0,10]}(x)$$

can be used as a pdf for X, and for any event A on $R(X)$ we obtain $P_X(A) = \int_{x \in A} \frac{1}{10} dx$.

- For example, the probability for $A = [0,5]$ is $P_X(A) = \int_0^5 \frac{1}{10} dx = 1/2$. $||$

REMARK: The definition of the continuous pdf implies that the probability for an elementary event $A = \{a\}$ is zero, since

$$P_X(A) = \int_a^a f(x)dx = 0.$$

But this does not mean that the event A is impossible! Instead, we might interpret this to mean that A is 'relatively impossible', relative to all other outcomes that can occur in $R(X) - A$. $\diamondsuit$

REMARK: Consider the sets $[a, b]$, $(a, b]$, $[a, b)$, (a, b) and note that

$$[a, b] = (a, b] \cup \{a\} = [a, b) \cup \{b\} = (a, b) \cup \{a\} \cup \{b\}.$$

Since the sets are disjoint and since $P_X(\{a\}) = P_X(\{b\}) = 0$, probability Axiom 1.3 implies that

$$P_X([a, b]) = P_X((a, b]) = P_X([a, b)) = P_X((a, b)) = \int_a^b f(x)dx. \qquad \diamondsuit$$

REMARK: The interpretation of the function value of a continuous pdf $f(x)$ is fundamentally different from that of a discrete pdf:

- If f is discrete, $f(x) = P_X(x) =$ probability of the outcome x.
- If f is continuous, $f(x)$ is not the probability of outcome x, which is $P_X(x) = 0$. Note that if $f(x)$ was a probability, we would have $f(x) = 0\ \forall x$. $\diamondsuit$

Restrictions on the admissible choices of f as a pdf

An important task of statistical inference is the identification of an appropriate function f which can be used as a pdf representing the stochastic behavior of a random variable. Note that the selected f should ensure that the probabilities obtained from f adhere to the probability axioms.

The following definition identifies the restrictions on admissible choices of f:

Definition (Class of discrete pdfs): The function $f : \mathbb{R} \to \mathbb{R}$ is a member of the class of discrete pdfs iff

(i$_a$) the set $C = \{x : f(x) > 0, x \in \mathbb{R}\}$ is countable;

(ii$_a$) $f(x) = 0 \ \forall \ x \in \bar{C}$;

(iii$_a$) $\sum_{x \in C} f(x) = 1$.

Definition (Class of continuous pdfs): The function $f : \mathbb{R} \to \mathbb{R}$ is a member of the class of continuous pdfs iff

(i$_b$) $f(x) \geq 0 \ \forall \ x \in \mathbb{R}$;

(ii$_b$) $\int_{x \in \mathbb{R}} f(x) = 1$.

REMARK: The definition tells us whether a particular function f can be used as a pdf or not.
The conditions (i_a)-(iii_a) and (i_b),(ii_b), respectively, ensure that the corresponding *set functions* used to compute probabilities, namely,

$$P_X(A) = \sum_{x \in A} f(x) \quad \text{and} \quad P_X(A) = \int_{x \in A} f(x) dx$$

are in fact *probability set functions* which adhere to the probability axioms.

(CONTINUES)

REMARK (CONTINUED): As to the sufficiency of the conditions (i_a)-(iii_a) for the resulting P_X to satisfy the axioms for the discrete case:

- (i_a), (ii_a) imply that $f(x) \geq 0 \; \forall x$
  $\Rightarrow \quad P_X(A) = \sum_{x \in A} f(x) \geq 0 \; \forall$ events A (Ax.1).

- (i_a), (ii_a) imply that we can set $C = R(X)$ such that together with (iii_a)
  $\Rightarrow \quad P_X(R(X)) = \sum_{x \in R(X)} f(x) = 1$ (Ax.2).

- Let $\{A_i, i \in I\}$ be a collection of disjoint events. Then the set function used to compute the probability for $\cup_{i \in I} A_i$ is

$$P_X(\cup_{i \in I} A_i) = \sum_{x \in [\cup_{i \in I} A_i]} f(x) \overset{\text{(disjoint } A_i\text{s)}}{=} \sum_{i \in I} \left[\sum_{x \in A_i} f(x) \right] = \sum_{i \in I} P_X(A_i) \quad \text{(Ax.3)}.$$

(CONTINUES)

Remark (continued): As to the sufficiency of the conditions (i_b) and (ii_b) for the resulting P_X to satisfy the axioms for the continuous case:

- (i_b) says that $f(x) \geq 0 \; \forall x$
  $\Rightarrow \quad P_X(A) = \int_{x \in A} f(x)dx \geq 0 \; \forall$ events A (Ax.1).

- (ii_b) says that $\int_{\mathbb{R}} f(x) = 1$.
  $\Rightarrow \quad \exists$ at least one event $A \subset \mathbb{R}$ such that $\int_A f(x) = 1$
  $\Rightarrow \quad$ We can set $A = R(X)$
  $\Rightarrow \quad P_X(R(X)) = \int_{x \in R(X)} f(x)dx = 1$ (Ax.2).

- Let $\{A_i, i \in I\}$ be a collection of disjoint events. Then the set function used to compute the probability for $\cup_{i \in I} A_i$ is

$$P_X(\cup_{i\in I} A_i) = \underbrace{\int_{x \in [\cup_{i \in I} A_i]} f(x)dx = \sum_{i \in I} \left[\int_{x \in A_i} f(x)dx \right]}_{non-overlapping\ A_i s:\ addititivity\ prop.\ of\ Riemann\ integrals} = \sum_{i \in I} P_X(A_i) \quad \text{(Ax.3)}.$$

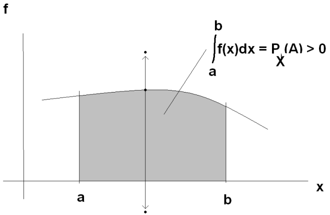
(continues)

REMARK (CONTINUED): For a discussion of the necessity of the conditions (i_a)-(iii_a) and (i_b),(ii_b) for the resulting P_X to satisfy the probability axioms, see Mittelhammer 2000, p. 57.

He shows that all conditions, except for the condition that $f(x) \geq 0 \ \forall \ x$ (i_b) in the continuous case, are necessary. In the continuous case, the property $f(x) \geq 0$ is not necessary for the following reason:

The function f could technically be negative for a finite number of x values, because the value of $\int_a^b f(x)dx$ is invariant to changes in $f(x)$ at a finite number of points having "measure zero". $\diamondsuit$

Fig. 6.



EXAMPLE:

1) Consider the function $f(x) = (0.3)^x\,(0.7)^{1-x}\mathbb{I}_{\{0,1\}}(x)$. Can this f serve as pdf?

Since (i) $f(x) > 0$ on the countable set $\{0, 1\}$, and (ii) $\sum_{x=0}^{1} f(x) = 1$, and (iii) $f(x) = 0\ \forall\ x \notin \{0, 1\}$, the function f can serve as a pdf.

2) Consider the function $f(x) = (x^2 + 1)\mathbb{I}_{[-1,1]}(x)$. Can this f serve as pdf?

While $f(x) \geq 0\ \forall\ x \in \mathbb{R}$, f does not integrate to 1:

$$\int_{\mathbb{R}} f(x)dx = \int_{-1}^{1}(x^2 + 1)dx = \frac{8}{3} \neq 1.$$

Thus, f can not serve as a pdf. (How do we get from f a function which can serve as a pdf ?) ||

Definition (Cumulative distribution function): The cumulative distribution function (cdf) of a random variable X, denoted by F, is defined by

$$F : \mathbb{R} \to [0,1] \quad \text{such that} \quad F(b) = P_X(X \leq b), \quad \forall b \in \mathbb{R}.$$

Remark: For a discrete random variable the cdf obtains as

$$F(b) = \sum_{x \leq b} f(x), \quad \forall b \in \mathbb{R},$$

and for a continuous random variable as

$$F(b) = \int_{-\infty}^{b} f(x)dx, \quad \forall b \in \mathbb{R}. \qquad \diamondsuit$$

EXAMPLE: Let the random variable X be the duration of a telephone call (in min), with range $R(X) = \{x : x > 0\}$.

- Let the pdf be: $f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \cdot \mathbb{I}_{(0,\infty)}(x)$, with $\lambda > 0$.
- The cdf obtains as: $F(b) = \int_0^b \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx = (1 - e^{-\frac{b}{\lambda}}) \cdot \mathbb{I}_{(0,\infty)}(b)$ – see Fig. 7.
- Assume that $\lambda = 100$ (average duration). Then the probability that the duration is less than 50 min is: $F(50) = 1 - e^{-\frac{50}{100}} = 0.39$. ||

EXAMPLE: Let the random variable X be the number of dots observed rolling a die, with range $R(X) = \{1, 2, \ldots, 6\}$.

- The pdf is: $f(x) = \frac{1}{6} \cdot \mathbb{I}_{\{1,\ldots,6\}}(x)$.
- The cdf obtains as:

$$F(b) = \sum_{x \leq b} \frac{1}{6} \cdot \mathbb{I}_{\{1,\ldots,6\}}(x) = \frac{1}{6} \lfloor b \rfloor \cdot \mathbb{I}_{\{1,\ldots,6\}}(b) + \mathbb{I}_{(6,\infty)}(b)$$

($\lfloor b \rfloor$ denotes the integer part of the number b) – see Fig. 7. ||

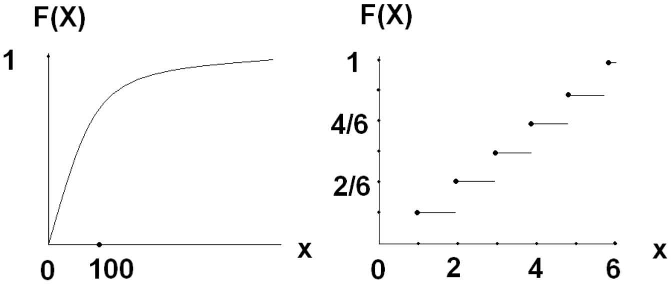


Fig. 7.

The definition of the cdf implies that a cdf $F(x)$ satisfies certain properties.

PROPERTIES OF A CDF:

(i) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$;

(ii) $F(x)$ is a non decreasing function on x; that is, $F(a) \leq F(b)$ for $a < b$;

(iii) $F(x)$ is right-continuous; that is, $\lim_{h \downarrow 0} F(x+h) = F(x)$.

REMARK:

- Property (i) follows from the fact that
  $\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} P_X(X \leq x) = P_X(\emptyset) = 0$, and
  $\lim_{x \to \infty} F(x) = \lim_{x \to \infty} P_X(X \leq x) = P_X(R(X)) = 1$

- Property (ii) follows from the fact that we accumulate (by integration / summation) non-negative values if we move from the left to the right.

- Property (iii) follows from the fact that
  $\lim_{h \downarrow 0} F(x+h) = \lim_{h \downarrow 0} P_X(X \leq x+h) = P_X(X \leq x) = F(x)$. $\diamondsuit$

The following theorems establish the relationship between a cdf and pdf.

THEOREM 2.1 *Let $x_1 < x_2 < x_3 < \cdots$ be the countable set of outcomes in the range of the discrete random variable X. Then the pdf for X obtains as*

$$f(x_i) = \begin{cases} F(x_i), & i = 1 \\ F(x_i) - F(x_{i-1}), & i = 2, 3, \dots \\ 0, & x \notin R(X). \end{cases}$$

PROOF: Since summation of f leads to F, differentiation of F leads to f. $\square$

Theorem 2.2 *Let $f(x)$ and $F(x)$ denote the pdf and cdf of a continuous random variable X. Then the pdf for X obtains as*

$$f(x) = \begin{cases} \frac{dF(x)}{dx}, & \text{wherever } f(x) \text{ is continuous} \\ 0, & \text{elsewhere.} \end{cases}$$

Proof: Wherever f is continuous we have

$$\frac{dF(x)}{dx} = \frac{d}{dx}\left[\int_{-\infty}^{x} f(u)du\right] = f(x) \qquad \text{(Fundamental Theorem of Calculus).}$$

At points where f is discontinuous (such that the derivative of F does not exist) we can set f to an arbitrary non-negative value (for example 0), since the value of $F(x) = \int_{-\infty}^{x} f(u)du$ is invariant to changes in $f(u)$ at a finite set of points having "measure zero". $\square$

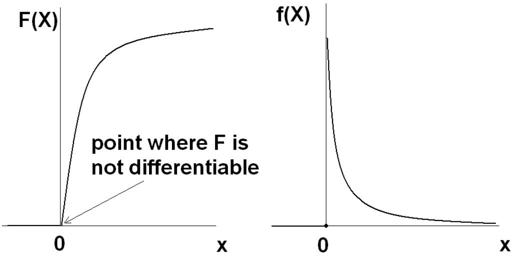
EXAMPLE: Recall the Ex., where X is the duration of a telephone call, with cdf

$$F(x) = (1 - e^{-\frac{x}{\lambda}}) \cdot \mathbb{I}_{(0,\infty)}(x).$$

A continuous pdf for X is given by

$$f(x) = \begin{cases} \frac{dF(x)}{dx} = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & \text{for} \quad x \in (0,\infty) \\ 0 \quad (\text{say}) & \text{for} \quad x = 0 \\ 0 & \text{for} \quad x \in (-\infty, 0). \end{cases} \qquad ||$$

Fig. 8.



So far, we have discussed univariate random variables, where only one real-valued function was defined on the sample space S. If we define concurrently two or more real-valued functions, we obtain multivariate random variables.

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

REMARK: The realized value of the multivariate random variable is

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

and its range is

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

The definitions of pdfs for multivariate discrete and continuous random variables are analogous to those in the univariate cases, and are as follows:

Definition (Discrete multivariate pdf): A multivariate random variable $X = (X_1, ..., X_n)$ is called discrete iff its range $R(X)$ is countable. The discrete joint pdf of a discrete random variable X, denoted by f, is defined by

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

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

DEFINITION (**CONTINUOUS MULTIVARIATE PDF**): A multivariate random variable $X = (X_1, ..., X_n)$ is called continuous iff

- its range $R(X)$ is uncountably infinite and
- there exists a function

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

$$P_X(A) = \int \cdots \int_{(x_1,...,x_n) \in A} f(x_1, ..., x_n) dx_1 \cdots dx_n$$

and

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

The function f is called a continuous joint pdf.

As in the univariate case, a function f selected to serve as a joint pdf should ensure that the probabilities obtained from the selected function f adhere to the probability axioms. The following definition identifies the restrictions on admissible choices of f:

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$_b$) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, ..., x_n) dx_1 \cdots dx_n = 1$.

REMARK: The conditions (i_a)-(iii_a) and (i_b),(ii_b), respectively, ensure that the corresponding *set functions* used to compute joint probabilities, namely,

$$P_X(A) = \sum_{(x_1,\dots,x_n)\in A} \cdots \sum f(x_1, \dots, x_n)$$

and

$$P_X(A) = \int_{(x_1,\dots,x_n)\in A} \cdots \int f(x_1, \dots, x_n) dx_1 \cdots dx_n$$

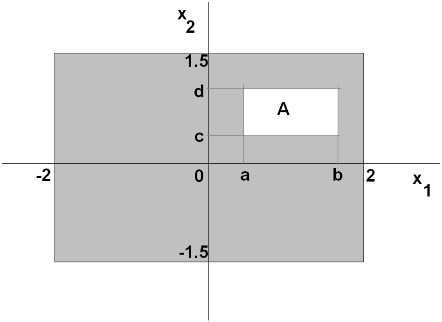
are in fact *probability set functions* which adhere to the probability axioms.

Sufficiency and necessity of the conditions can be shown by generalizing the arguments used in the univariate case. $\diamondsuit$

EXAMPLE: Consider that the NASA announces that a small meteorite will hit a rectangular area of 12km^2=$4\text{km}\times 3\text{km}$. Suppose that each point in that rectangle is equally likely to be struck.

- Define $X = (X_1, X_2)$ to be the coordinates of the point of strike, with a range $R(X) = \{(x_1, x_2) : x_1 \in [-2, 2], x_2 \in [-1.5, 1.5]\}$.
- In order to obtain the continuous pdf of X, consider a closed rectangle A in $R(X)$ (see Fig. 9).

Fig. 9.



(CONTINUES)

Example (continued):

- Since all points are equally likely, we obtain

$$P_x(x \in A) = \frac{\text{area of A}}{R(X)} = \frac{(b-a)(d-c)}{12}.$$

- According to the definition, the pdf f for X has to satisfy

$$\int_c^d \int_a^b f(x_1, x_2) dx_1 dx_2 \overset{!}{=} \frac{(b-a)(d-c)}{12},$$

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

$$\frac{\partial^2 \left[\int_c^d \int_a^b f(x_1, x_2) dx_1 dx_2 \right]}{\partial d \partial b} = f(b, d) \overset{!}{=} \frac{\partial^2 \left[(b-a)(d-c)/12 \right]}{\partial d \partial b} = \frac{1}{12},$$

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

- 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 X, and for any event $A \in R(X)$ we obtain $P_X(A) = \int \int_{x \in A} \frac{1}{12} dx_1 dx_2$. ||

Multivariate cdfs

DEFINITION (JOINT CDF): The joint cdf of an n-dimensional random variable X, denoted by F, is defined by

$$F : \mathbb{R}^n \to [0,1] \quad \text{such that} \quad F(b_1, ..., b_n) = P_X(X_1 \leq b_1, ..., X_n \leq b_n),$$

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

REMARK: 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

$$F(b_1, ..., b_n) = \int_{-\infty}^{b_n} \cdots \int_{-\infty}^{b_1} f(x_1, ..., x_n) dx_1 \cdots dx_n, \quad \forall (b_1, ..., b_n) \in \mathbb{R}^n. \qquad \diamondsuit$$

REMARK: Properties of joint cdfs are:

(i) $\lim_{b_i \to -\infty} F(b_1, ..., b_n) = P_X(\emptyset) = 0$, for any $i = 1, ..., n$;

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

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

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

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

Similar to the univariate case the joint cdf can be used to obtain the joint pdf.

THEOREM 2.3 *Let* (X, Y) *be a discrete bivariate random variable with joint cdf* $F(x, y)$ *and range* $R(X, Y) = \{x_1 < x_2 < x_3 < \cdots, y_1 < y_2 < y_3 < \cdots\}$. *Then the joint pdf obtains as*

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

PROOF: Since summation of f leads to F, differentiation of F leads to f. □

REMARK: The result of Theorem 2.3 for the bivariate case can be generalized to the n-variate case. However, this will require a somewhat cumbersome notation. ◇

THEOREM 2.4 *Let $f(x_1, ..., x_n)$ and $F(x_1, ..., x_n)$ denote the joint pdf and cdf for a continuous multivariate random variable $X = (X_1, ..., X_n)$. Then the joint pdf for X obtains as*

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

PROOF: Wherever f is continuous we have

$$\frac{\partial^n F(x_1, ..., x_n)}{\partial x_1 \cdots \partial x_n} = \underbrace{\frac{\partial^n \left[\int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(u_1, ..., u_n) du_1 \cdots du_n\right]}{\partial x_1 \cdots \partial x_n} = f(x_1, ..., x_n)}_{\text{(Fundamental Theorem of Calculus)}}.$$

At points where f is discontinuous (such that the derivative of F does not exist) we can set f to an arbitrary non-negative value (for example 0), since the value of $F(x_1, ..., x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(u_1, ..., u_n) du_1 \cdots du_n$ is invariant to changes in $f(\cdot)$ at a finite set of points having "measure zero". □

EXAMPLE: Recall the meteorite example, where $X = (X_1, X_2)$ is the point of strike. The joint cdf obtains as

$$F(b_1, b_2) = \int_{-\infty}^{b_2} \int_{-\infty}^{b_1} \frac{1}{12} \mathbb{I}_{[-2,2]}(x_1) \mathbb{I}_{[-1.5,1.5]}(x_2) dx_1 dx_2,$$

with four different integration areas, such that

$$\begin{aligned} F(b_1, b_2) &= \left[\int_{-1.5}^{b_2} \int_{-2}^{b_1} \frac{1}{12} dx_1 dx_2 \right] \mathbb{I}_{[-2,2]}(b_1) \mathbb{I}_{[-1.5,1.5]}(b_2) \\ &+ \left[\int_{-1.5}^{b_2} \int_{-2}^{2} \frac{1}{12} dx_1 dx_2 \right] \mathbb{I}_{(2,\infty)}(b_1) \mathbb{I}_{[-1.5,1.5]}(b_2) \\ &+ \left[\int_{-1.5}^{1.5} \int_{-2}^{b_1} \frac{1}{12} dx_1 dx_2 \right] \mathbb{I}_{[-2,2]}(b_1) \mathbb{I}_{(1.5,\infty)}(b_2) \\ &+ \left[\int_{-1.5}^{1.5} \int_{-2}^{2} \frac{1}{12} dx_1 dx_2 \right] \mathbb{I}_{(2,\infty)}(b_1) \mathbb{I}_{(1.5,\infty)}(b_2) \\ &= \frac{(b_1 + 2)(b_2 + 1.5)}{12} \mathbb{I}_{[-2,2]}(b_1) \mathbb{I}_{[-1.5,1.5]}(b_2) \\ &+ \frac{4(b_2 + 1.5)}{12} \mathbb{I}_{(2,\infty)}(b_1) \mathbb{I}_{[-1.5,1.5]}(b_2) \\ &+ \frac{3(b_1 + 2)}{12} \mathbb{I}_{[-2,2]}(b_1) \mathbb{I}_{(1.5,\infty)}(b_2) + 1 \cdot \mathbb{I}_{(2,\infty)}(b_1) \mathbb{I}_{(1.5,\infty)}(b_2). \qquad || \end{aligned}$$

From the joint pdf $f(x_1, x_2)$ of a bivariate random variable (X_1, X_2) we can easily derive the marginal pdf of X_1 and X_2, denoted by $f_1(x_1)$ and $f_2(x_2)$, which can be used to assign (marginal) probabilities to the events $x_1 \in A_1$ and $x_2 \in A_2$, that is $P(x_1 \in A_1)$ and $P(x_2 \in A_2)$.

THEOREM 2.5 *Let $X = (X_1, X_2)$ be a discrete random variable with joint pdf $f(x_1, x_2)$ and a range $R(X) = R(X_1) \times R(X_2)$. The marginal pdfs are given by*

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

PROOF: For any $x_1 \in R(X_1)$, let

$$A = \{(x_1, x_2) : x_2 \in R(X_2)\}.$$

That is, A is a line in the plane $R(X)$ with first coordinate equal to x_1. Then for any $x_1 \in R(X_1)$,

$$\begin{aligned}
f_1(x_1) &= P(x_1) && \text{[by def.]} \\
&= P(x_1\ ,\ x_2 \in R(X_2)) && [P(x_2 \in R(X_2)) = 1] \\
&= P((x_1, x_2) \in A) && \text{[def. of } A] \\
&= \sum_{(x_1, x_2) \in A} f(x_1, x_2) \\
&= \sum_{x_2 \in R(X_2)} f(x_1, x_2)
\end{aligned}$$

The proof for $f_2(x_2)$ is analogous. $\square$

REMARK: In order to obtain the marginal pdf, we simply "sum out" the variables that are not of interest in the joint pdf. If the bivariate random variable is continuous, the marginal pdfs are obtained as in the discrete case with integrals replacing sums. $\diamondsuit$

THEOREM 2.6 *Let $X = (X_1, X_2)$ be a continuous random variable with joint pdf $f(x_1, x_2)$. The corresponding marginal pdfs are given by*

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2, \quad \text{and} \quad f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1.$$

PROOF: For any event $x_1 \in B$, let $A = \{(x_1, x_2) : x_1 \in B, x_2 \in R(X_2)\}$.
Then for any event $x_1 \in B$,

$$\begin{aligned} P(x_1 \in B) &= P(x_1 \in B; x_2 \in R(X_2)) = P((x_1, x_2) \in A) \\ &= \int\int_{(x_1,x_2)\in A} f(x_1, x_2) dx_2 dx_1 \\ &= \int_{x_1 \in B} \underbrace{\left[\int_{-\infty}^{\infty} f(x_1, x_2) dx_2\right]}_{\text{has to be the pdf of } X_1,\ f_1(x_1) \text{, in order to obtain } P(X_1 \in B)\ !} dx_1 = \int_{x_1 \in B} f_1(x_1) dx_1 \end{aligned}$$

The proof for $f_2(x_2)$ is analogous. $\square$

EXAMPLE: Consider the continuous random variable $X = (X_1, X_2)$ 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

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{\infty} f(x_1, x_2)dx_2 = \int_0^1 (x_1 + x_2)\mathbb{I}_{[0,1]}(x_1)dx_2 = \left[(x_1 x_2 + \frac{x_2^2}{2})\mathbb{I}_{[0,1]}(x_1)\right]_{x_2=0}^{x_2=1} \\ &= (x_1 + \frac{1}{2})\mathbb{I}_{[0,1]}(x_1). \qquad || \end{aligned}$$

REMARK: The concept of marginal pdfs can be straightforwardly generalized from the bivariate to the n-variate case. In this case marginal pdfs can be joint pdfs themselves. The n-variate generalization is presented in the following definition. $\diamondsuit$

DEFINITION (MARGINAL PDFS): Let $f(x_1, \ldots, x_n)$ be the joint pdf for the n-dimensional random variable $(X_1, \ldots, X_n)$. Let $J = \{j_1, j_2, \ldots, j_m\}, 1 \leq m < n$, 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

$$f_{j_1 \ldots j_m}\left(x_{j_1}, \ldots, x_{j_m}\right) = \begin{cases} \underset{\left(x_i \in R(X_i), i \in I - J\right)}{\sum \cdots \sum} f(x_1, \ldots, x_n) & \text{(discrete case).} \\ \displaystyle\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) \prod_{i \in I - J} dx_i & \text{(continuous case).} \end{cases}$$

From the joint pdf of a bivariate random variable (X_1, X_2) we can easily 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) $x_2 \in D$.

This probability obtains as

$$P(x_1 \in C | x_2 \in D) = \begin{cases} \displaystyle\sum_{x_1 \in C} f(x_1 | x_2 \in D) & \text{(discrete case)} \\ \displaystyle\int_{x_1 \in C} f(x_1 | x_2 \in D) dx_1 & \text{(continuous case)}, \end{cases}$$

where $f(x_1 | x_2 \in D)$ denotes the conditional pdf of X_1 given that $x_2 \in D$.

The pdf $f(x_1 | x_2 \in D)$ can be derived as follows:

- Consider a discrete bivariate random variable (X_1, X_2), with joint pdf $f(x_1, x_2)$, and the following two pairs of equivalent events:

$$x_1 \in C \quad \Leftrightarrow \quad (x_1, x_2) \in A = \{(x_1, x_2) : x_1 \in C, x_2 \in R(X_2)\}$$

$$x_2 \in D \quad \Leftrightarrow \quad (x_1, x_2) \in B = \{(x_1, x_2) : x_1 \in R(X_1), x_2 \in D\}$$

- Then the conditional probability for $x_1 \in C$ given $x_2 \in D$ is given by:

$$P(x_1 \in C | x_2 \in D) \overset{\text{(equiv. of events)}}{=} P(A|B) \overset{\text{(by Def.)}}{=} \frac{P(A \cap B)}{P(B)} \quad \text{for} \quad P(B) > 0.$$

- Since the intersection of A and B is $A \cap B = \{(x_1, x_2) : x_1 \in C, x_2 \in D\}$, we get:

$$P(x_1 \in C | x_2 \in D) = \frac{\sum_{x_1 \in C} \sum_{x_2 \in D} f(x_1, x_2)}{\sum_{x_2 \in D} \underbrace{\sum_{x_1 \in R(X_1)} f(x_1, x_2)}_{\text{marginal } f_2(x_2)}} = \sum_{x_1 \in C} \underbrace{\left[\frac{\sum_{x_2 \in D} f(x_1, x_2)}{\sum_{x_2 \in D} f_2(x_2)} \right]}_{\text{has to be the pdf } f(x_1 | x_2 \in D)\ !} .$$

- Thus, 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

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

and, if D is a single point d, by

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

- 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

$$f(x_1|x_2 \in D) = \frac{\int_{x_2 \in D} f(x_1, x_2) dx_2}{\int_{x_2 \in D} f_2(x_2) dx_2}.$$

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

$$f(x_1|x_2 = d) = \frac{\int_d^d f(x_1, x_2) dx_2}{\int_d^d f_2(x_2) dx_2} = \frac{0}{0},$$

which is undefined!

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

- In particular, in the continuous case we define the conditional probability for $x_1 \in A$ given $x_2 = d$ as

$$\begin{aligned}
P(x_1 \in A | x_2 = d) &\equiv \lim_{\epsilon \downarrow 0} P(x_1 \in A | d - \epsilon \leq x_2 \leq d + \epsilon) \\
&= \lim_{\epsilon \downarrow 0} \int_{x_1 \in A} \left[\frac{\int_{d-\epsilon}^{d+\epsilon} f(x_1, x_2) dx_2}{\int_{d-\epsilon}^{d+\epsilon} f_2(x_2) dx_2} \right] dx_1 \\
&\qquad \text{(by def. of a conditional prob.)} \\
&= \lim_{\epsilon \downarrow 0} \int_{x_1 \in A} \left[\frac{2\epsilon f(x_1, x_2^{**})}{2\epsilon f_2(x_2^{*})} \right] dx_1, \quad (x_2^{*}, x_2^{**}) \in [d - \epsilon, d + \epsilon] \\
&\qquad \text{(by the Mean Value Theorem for integrals)} \\
&= \lim_{\epsilon \downarrow 0} \int_{x_1 \in A} \left[\frac{f(x_1, x_2^{**})}{f_2(x_2^{*})} \right] dx_1 .
\end{aligned}$$

As $\epsilon \downarrow 0$, the interval $[d - \epsilon, d + \epsilon]$ reduces to $[d, d] = d$, so that $x_2^{*} \to d$ and $x_2^{**} \to d$. Thus, we get:

$$P(x_1 \in A | x_2 = d) = \int_{x_1 \in A} \underbrace{\frac{f(x_1, d)}{f_2(d)}}_{\text{has to be the pdf } f(x_1 | x_2 = d)\,!} dx_2 .$$

- This implies that 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 this conditional pdf has exactly the same form as in the discrete case.

EXAMPLE: Consider the continuous random variable $X = (X_1, X_2)$ 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):

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

- Then the conditional pdf of X_1 given $x_2 \leq .5$ obtains as

$$\begin{aligned} f(x_1|x_2 \leq .5) &\overset{(def.)}{=} \frac{\int_{-\infty}^{.5} f(x_1, x_2)dx_2}{\int_{-\infty}^{.5} f_2(x_2)dx_2} = \frac{\int_{-\infty}^{.5}(x_1 + x_2)\mathbb{I}_{[0,1]}(x_1)\mathbb{I}_{[0,1]}(x_2)dx_2}{\int_{-\infty}^{.5}(x_2 + \frac{1}{2})\mathbb{I}_{[0,1]}(x_2)dx_2} \\ &= (\frac{4}{3}x_1 + \frac{1}{3})\mathbb{I}_{[0,1]}(x_1). \end{aligned}$$

- The conditional pdf of X_1 given $x_2 = .75$ is

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

The concept of conditional pdfs can be straightforwardly generalized from the bivariate to the n-variate case. The n-variate generalization is presented in the following definition:

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 $\left(X_{j_1}, \ldots, X_{j_m}\right)$, given $\left(X_{j_{m+1}} = d_{m+1}, \ldots, X_{j_n} = d_n\right)$ is given by

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

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

REMARK: From the conditional pdf we can straightforwardly derive the conditional cdf by using the conditional pdf in the general definition of a cdf. $\diamondsuit$

The independence of two events A and B means that $P(A \cap B) = P(A) \cdot P(B)$ (see Chapter 1.5). This concept of independence can be straightforwardly applied to multivariate random variables.

DEFINITION (INDEPENDENCE OF RANDOM VARIABLES): The random variables X_1 and X_2 are said to be independent iff

$$P(x_1 \in A_1, x_2 \in A_2) = P(x_1 \in A_1) \cdot P(x_2 \in A_2), \quad \text{for all events } A_1,\ A_2.$$

REMARK: This definition is not immediately operational since the factorization has to hold for all pairs of events. Thus, the following result can be useful in practice: $\diamondsuit$

Theorem 2.7 *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).

Proof: (Continuous case) Let A_1, A_2 be any pair of events. Then, if the joint pdf can be factorized,

$$\begin{aligned}
P(x_1 \in A_1, x_2 \in A_2) &\overset{(def.)}{=} \int_{x_1 \in A_1} \int_{x_2 \in A_2} f(x_1, x_2) dx_2 dx_1 \\
&\overset{(by\ factorization)}{=} \int_{x_1 \in A_1} f_1(x_1) dx_1 \cdot \int_{x_2 \in A_2} f_2(x_2) dx_2 \\
&\overset{(def.)}{=} P(x_1 \in A_1) \cdot P(x_2 \in A_2).
\end{aligned}$$

Thus, the factorization is sufficient for independence.

(continues)

PROOF (CONTINUED): Now suppose that X_1, X_2 are independent and let $A_i = \{x_i : x_i < a_i\}$, $(i = 1, 2)$ for arbitrary a_i's. Then, by independence,

$$
\begin{aligned}
P(x_1 \in A_1, x_2 \in A_2) &\overset{(def.)}{=} \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} f(x_1, x_2) dx_2 dx_1 \\
&\overset{(by\ independence)}{=} P(x_1 \in A_1) \cdot P(x_2 \in A_2). \\
&\overset{(def.)}{=} \int_{-\infty}^{a_1} f_1(x_1) dx_1 \cdot \int_{-\infty}^{a_2} f_2(x_2) dx_2.
\end{aligned}
$$

Differentiating the integrals w.r.t. a_1 and a_2 yields $f(a_1, a_2) = f_1(a_1) \cdot f_2(a_2)$. Thus, the factorization is necessary for independence. The proof for the discrete case is analogous. $\square$

REMARK: 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 $x = b$. $\diamondsuit$

EXAMPLE: Recall the meteorite example, where $X = (X_1, X_2)$ 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

$$\begin{aligned} f_1(x_1) &= \frac{1}{12}\mathbb{I}_{[-2,2]}(x_1)\int_{-1.5}^{1,5} 1dx_2 = \frac{1}{4}\mathbb{I}_{[-2,2]}(x_1) \\ f_2(x_2) &= \frac{1}{12}\mathbb{I}_{[-1.5,1.5]}(x_2)\int_{-2}^{2} 1dx_1 = \frac{1}{3}\mathbb{I}_{[-1.5,1.5]}(x_2) \end{aligned}$$

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

REMARK: 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: $f(x_1, x_2) = f_1(x_1) f_2(x_2)$.
However, if X_1 and X_2 are dependent, then knowing the marginal pdfs f_1 and f_2 is not sufficient to determine the joint pdf f. $\diamondsuit$

EXAMPLE: Consider the joint pdf

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

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. $||$

So far, we considered the concept of independence for bivariate random variables. It can be extended to the n-variate case with the following definition:

DEFINITION (INDEPENDENCE IN THE n-VARIATE CASE): The random variables $X_1, \ldots, X_n$ are said to be independent iff

$$P(x_1 \in A_1, \ldots, x_n \in A_n) = \prod_{i=1}^{n} P(x_i \in A_i), \quad \text{for all events } A_1, \ldots, A_n.$$

The generalization of the joint pdf factorization theorem from the bivariate to the n-variate case is given in the following theorem :

THEOREM 2.8 *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).

PROOF: The proof is a direct extension of that for the bivariate case (Theorem 2.7). $\square$

The independence concept for random variables can be extended to the independence of random variables, which are defined as functions of other independent random variables:

THEOREM 2.9 *If X_1 and X_2 are independent random variables, and if Y_1 and Y_2 are defined as functions $y_1 = g_1(x_1)$ and $y_2 = g_2(x_2)$, then Y_1 and Y_2 are independent.*

PROOF: Define the equivalent events $y_i \in A_i$ and $x_i \in B_i$, this means,

$$B_i = \{x_i : g_i(x_i) \in A_i, x_i \in R(X_i)\}, \quad i = 1, 2.$$

The joint probability for $y_1 \in A_1$ and $y_2 \in A_2$ is

$$\begin{aligned} P(y_1 \in A_1, y_2 \in A_2) &\overset{(equiv.\ of\ events)}{=} P(x_1 \in B_1, x_2 \in B_2) \\ &\overset{(independence)}{=} P(x_1 \in B_1)P(x_2 \in B_2) \\ &\overset{(equiv.\ of\ events)}{=} P(y_1 \in A_1)P(y_2 \in A_2). \quad \square \end{aligned}$$