2. Random Variables, Distributions and Densities

1. Random Variables and Distributions

Let a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ be given. We define

Definition 1 *A random variable X is a measurable function from Ω to $\mathbf{R}$.*

Remarks

(a) A random variable X is a *function* from Ω to $\mathbf{R}$, i.e., it assigns a number to each outcome.

(b) It is *measurable*, i.e.,

$$X^{-1}(A) = \{\omega |\, X(\omega) \in A\} \in \mathcal{F}$$

for any $A \in \mathcal{B}(\mathbf{R})$.

Example For the random experiment of tossing a coin, we may define a random variable X by $X(H) = 1$ and $X(T) = 0$, where H and T denote outcomes of the coin landing on its head and tail, respectively.

Definition 2 *The distribution P_X of a random variable X is the probability measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ induced by X. It is thus defined by*

$$\mathrm{P}_X(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}\{\omega |\, X(\omega) \in A\} \quad \textit{for} \;\; A \in \mathcal{B}(\mathbf{R})$$

We often write $\mathrm{P}_X = \mathbf{P} \circ X^{-1}$. When there is no ambiguity about the underlying random variable, we usually write P in the place of P_X for simplicity.

Remarks We may easily show that P is indeed a probability measure, i.e., it satisfies three axioms of probability. Notice that

(a) $\mathrm{P}(A) = \mathbf{P}(X^{-1}(A)) \geq 0$ for any $A \in \mathcal{B}(\mathbf{R})$

(b) $\mathrm{P}(\mathbf{R}) = \mathbf{P}(\Omega) = 1$

(c) If $A_1, A_2, \ldots \in \mathcal{B}(\mathbf{R})$ are disjoint, then

$$\mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \;=\; \mathbf{P}\left(X^{-1} \bigcup_{n=1}^{\infty} A_n\right)$$

$$= \mathbf{P}\left(\bigcup_{n=1}^{\infty} X^{-1}(A_n)\right)$$
$$= \sum_{n=1}^{\infty} \mathbf{P}(X^{-1}(A_n))$$
$$= \sum_{n=1}^{\infty} \mathrm{P}(A_n)$$

The properties of probability introduced earlier thus apply to P_X.

Example Consider the random variable X introduced in the above example, and let $\mathbf{P}\{H\} = 1/3$ and $\mathbf{P}\{T\} = 2/3$. The distribution P_X of X is then given by

$$\mathrm{P}_X\{0\} = \frac{2}{3} \quad \text{and} \quad \mathrm{P}_X\{1\} = \frac{1}{3}$$

More precisely, $\mathrm{P}_X(A) = 1/3, 2/3, 1$ depending whether A contains only 1, only 0 or both. If A contains neither 0 or 1, then $\mathrm{P}_X(A) = 0$.

Definition 3 *The distribution function F_X of a random variable X is defined by*

$$F_X(x) = \mathrm{P}_X(-\infty, x] \quad \textit{for} \;\; x \in \mathbf{R}$$

As in the case of distribution, we often omit the subscript X of F_X and write F when there is no danger of confusion.

Remark Note that the distribution P is uniquely determined by the distribution function F, since $\mathcal{P} = \{(-\infty, x]\}$ is a π-system that generates $\mathcal{B}(\mathbf{R})$. See Theorem 6 of the previous section.

Properties of Distribution Function

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

(b) $F(x) \le F(y)$ if $x \le y$

(c) F is right continuous

Proof

(a) Let $x_n \downarrow -\infty$. Since $(-\infty, x_n] \downarrow \emptyset$, we have

$$F(x_n) = \mathrm{P}(-\infty, x_n] \to \mathrm{P}(\emptyset) = 0$$

as is to be shown. The proof for the case $x_n \uparrow \infty$ is similar.

(b) Let $x \leq y$. Then $(-\infty, x] \subset (-\infty, y]$ and it follows that

$$F(x) = \mathrm{P}(-\infty, x] \leq \mathrm{P}(-\infty, y] = F(y)$$

by the monotonicity of P.

(c) Fix an arbitrary x. For the right continuity of F at x, it suffices to show that $F(x_n) \to F(x)$ for any sequence $\{x_n\}$ such that $x_n \downarrow x$. However, this is obvious since

$$F(x_n) = \mathrm{P}(-\infty, x_n] \to \mathrm{P}(-\infty, x] = F(x)$$

due to the monotone convergence. ■

Remark F is not necessarily left-continuous. For $x_n \uparrow x$, we have $(-\infty, x_n] \uparrow (-\infty, x)$. It thus follows that

$$F(x_n) = \mathrm{P}(-\infty, x_n] \to \mathrm{P}(-\infty, x) \neq F(x)$$

We indeed have

$$F(x) = F(x^-) + \mathrm{P}\{x\}$$

where $F(x^-) = \mathrm{P}(-\infty, x)$, and therefore F is continuous at x if and only if P does not have any point probability mass at x.

2. Random Vectors and Joint Distributions

Let $X_1, \ldots, X_n$ be random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We now define a *random vector* X by

$$X(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix}$$

Therefore, an n-dimensional random vector is just an n-tuple of random variables. We may of course view X as a measurable function from Ω to $\mathbf{R}^n$.

Examples

(a) Consider the random experiment of flipping a coin. Denote by X_1 a random variable given by $X_1(H) = 1$ and $X_1(T) = 0$. Also, let X_2 be another random variable given by $X_2(H) = 0$ and $X_2(T) = 1$. Define a random vector X by $X = (X_1, X_2)'$. We may view X as a function from from $\Omega = \{H, T\}$ to $\mathbf{R}^2$, since

$$X(H) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad X(T) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

(b) Look at the random experiment of tossing two coins. Let X_1 be a random variable taking value 1 if the first coin is flipped down with head and 0 otherwise. Similarly, denote by X_2 the random variable assigning value 1 if the second coin is flipped down with head and 0 otherwise. The random vector $X = (X_1, X_2)'$ is then a function from $\Omega = \{HH, HT, TH, TT\}$ to $\mathbf{R}^2$, and is given by

$$X(HH) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \; X(HT) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; X(TH) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \; X(TT) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The distribution of a random vector is defined similarly as for the case of a random variable. The distribution P_X of an n-dimensional random vector $X = (X_1, \ldots, X_n)'$ is a probability measure on $\mathbf{R}^n$, which is given by

$$\mathrm{P}_X(A) = \mathbf{P}\{\omega | \, X(\omega) \in A\}$$

for $A \in \mathcal{B}(\mathbf{R}^n)$. The distribution of a random vector X is often called the *joint distribution* of random variables $X_1, \ldots, X_n$. This is to make it explicit that more than one random variables are involved. The distribution of a subvector of X is called the *marginal* distribution, which we can derive from the joint distribution. For example consider a random vector $X = (X_1, X_2)$ with two random variables, and denote by $\mathrm{P}_{X_1}(A_1)$ the marginal distribution of the sub-vector X_1 for $A_1 \in \mathbf{R}$. We may consider the following cylinder set

$$A_1 \times \mathbf{R} = \{(x_1, x_2) | \, x_1 \in A_1, x_2 \in \mathbf{R}\} \in \mathcal{B}(R^2)$$

for any $A_1 \in \mathcal{B}(\mathbf{R})$, then we have

$$\mathrm{P}_X(A_1 \times \mathbf{R}) = \mathbf{P}(X^{-1}(A_1 \times \mathbf{R})) = \mathbf{P}\left\{\omega | X_1(\omega) \in A_1\right\} = \mathrm{P}_{X_1}(A_1)$$

Similarly, the marginal distribution of X_2 can be obtained from the joint distribution.

For an n-dimensional random vector $X = (X_1, \cdots, X_n)$, the distribution function F_X is defined by

$$F_X(x_1, \cdots, x_n) = \mathbf{P}\{\omega |\, X_1(\omega) \le x_1, \ldots, X_n(\omega) \le x_n\}$$

It is thus a real valued funtion on $\mathbf{R}^n$. The distribution function of a random vector X is often called the *joint distribution function* of the component random variables $X_1, \ldots, X_n$.

3. The Lebesque Integral and Density

Let M be a set, and $\mathcal{M}$ be a σ-field of subsets of M on which a measure μ is defined. Consider a *simple function* $f : \mathcal{M} \to \mathbf{R}$, taking on only a finite m number of values. More explicitly, we let

$$f = \sum_{k=1}^{m} c_k \, \mathrm{I}(A_k)$$

with $A_k \cap A_j = \emptyset$ for $i \neq j$, and where $c_k \in \mathbf{R}$ and $A_k \in \mathcal{M}$. Define the *Lebesque* integral of f with respect to μ by

$$\int f \, d\mu = \sum_k c_k \mu(A_k)$$

For a general nonnegative function f, we define

$$\int f \, d\mu = \sup \int f_n \, d\mu$$

where sup is taken over all simple functions $f_n \le f$, i.e., all the simple functions below f. To get an explicit sequence of such simple functions, we let

$$\alpha_n(x) = \begin{cases} 0 & x = 0 \\ \dfrac{k-1}{2^n} & \text{if } \dfrac{k-1}{2^n} < x \le \dfrac{k}{2^n} \\ n & x > n \end{cases}$$

for $k = 1, \ldots, n2^n$. It is easy to see that $f_n = \alpha_n \circ f$ satisfies the desired properties, i.e., f_n is a sequence of simple functions such that $f_n \uparrow f$.

For any measurable function f, construct the following nonnegative functions

$$\begin{aligned} f^+(x) &= \max(f(x), 0) \\ f^-(x) &= \max(-f(x), 0) \end{aligned}$$

and write

$$f = f^+ - f^-$$

Define the Lebesque integral of f by

$$\int f\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu$$

If the integrals of f^+ and f^- are both finite, then f is said to be *integrable* with respect to μ.

Remarks

(a) The Lebesque integral generalizes the *Riemann-Stieltjes* integral. The Lebesque integral with respect to the Lebesque measure (the usual measure given by length, area, volume, etc) on $\mathbf{R}^n$ exists and agrees with the Riemann-Stieltjes integral whenever the latter exists.

(b) Let the set $M = \{x_1, x_2, \ldots\}$ be countable, and μ be the counting measure assigning unity to each x_i. Then f is integrable, provided that the sequence $\{f(x_i)\}$ is absolutely summable, i.e., $\sum_i |f(x_i)| < \infty$ and $\int f\,d\mu$ is given by $\sum f(x_i)$.

Let μ be a measure on a measure space $(M, \mathcal{M})$ as before, and $f : \mathcal{M} \to \mathbf{R}$ be a nonnegative function. We may construct another measure ν on $(M, \mathcal{M})$ by

$$\nu(A) = \int_A f\,d\mu$$

for $A \in \mathcal{M}$. For the measure ν thus constructed, it is obvious that

$$\nu(A) = 0 \quad \text{whenever} \quad \mu(A) = 0$$

i.e., the measure ν is *absolutely continuous* with respect to the measure μ.

We now introduce

Theorem 1 (Radon-Nikodym) *Let μ and ν be two nonnegative measures on a measure space $(M, \mathcal{M})$. If ν is absolutely continuous with respect to μ, then ν can be represented as*

$$\nu(A) = \int_A f \, d\mu \quad \textit{for} \;\; A \in \mathcal{M}$$

for a $\mathcal{M}$-measurable f.

In this case, the measure ν is said to have *density* f with respect to μ. The function f is also called the *Radon-Nikodym derivative* of ν with respect to μ.

We now go back and look at the distribution of a random variable. If the distribution is absolutely continuous with respect to the Lebesque measure, then by the Radon-Nikodym theorem, it has a density, with respect to the Lebesque measure. A density of a distribution is more often called a *probability density.* We call *continuous* the random variables that induce such absolutely continuous distributions. We may, for example, express the distribution P_X induced by a continuous random variable X as

$$\mathrm{P}_X(A) = \int_A p_X d\mu \quad \text{for} \;\; A \in \mathcal{B}(\mathbf{R})$$

where p_X is the probability density of P_X.

On the other hand, there are random variables which induce distributions that are absolutely continuous with respect to the counting measure. The densities are then given with respect to the counting measure. Such random variables are sometimes said to be of *discrete* type. We will mostly look at these two types of distributions.

4. Independence

Let X be a random variable. We define

Definition 4 *The σ-field $\sigma(X)$ generated by X is given by*

$$\sigma(X) = \{X^{-1}(A) \,|\, A \in \mathcal{B}(\mathbf{R})\}$$

Remarks

(a) It is easy to check that $\sigma(X)$ is indeed a σ-field.

(b) The σ-field $\sigma(X)$ is the smallest σ-field with respect to which X is measurable.

(c) The σ-field generated by a random vector $X = (X_1, \cdots, X_n)'$ is defined similarly as

$$\sigma(X) = \sigma(X_1, \ldots, X_n) = \{X^{-1}(A) \mid A \in \mathcal{B}(\mathbf{R}^n)\}$$

(d) Intuitively, $\sigma(X)$ is precisely the collection of events E such that, for a given ω, we can tell whether or not $\omega \in E$ solely on the basis of the value $X(\omega)$.

Example Let E be an event and define a random variable by $X = \mathrm{I}(E)$. The σ-field $\sigma(X)$ generated by X would then be given by

$$\sigma(X) = \{\emptyset, \Omega, E, E^c\}$$

Obviously, given the value $X(\omega)$, it can be unambiguously determined whether or not ω is in any of the sets in $\sigma(X)$.

Definition 5 (Independence) *Sub-σ-fields $\mathcal{F}_1, \mathcal{F}_2, \ldots$ of $\mathcal{F}$ are said to be independent if for any $E_i \in \mathcal{F}_i$*

$$\mathbf{P}\left(\bigcap_{k=1}^{n} E_{i_k}\right) = \prod_{k=1}^{n} \mathbf{P}(E_{i_k})$$

for any $n \geq 2$, where $(i_1, \ldots, i_n)$ is an arbitrary collection of n members in the index set $\{1, 2, 3, \ldots\}$. Random variables $X_1, X_2, \ldots$ are independent if the σ-fields $\sigma(X_1), \sigma(X_2), \ldots$ generated by them are independent.

By the abuse of notation, we use $p(x_{i_1}, \ldots, x_{i_n})$ to denote the joint density of the random variables $X_{i_1}, \ldots, X_{i_n}$. We may show that

Theorem 2 *The random variables $X_1, X_2, \ldots$ are independent if and only if*

$$p(x_{i_1}, \ldots, x_{i_n}) = \prod_{k=1}^{n} p(x_{i_k})$$

Proof We only show the necessity. Also, we only consider the case of two random variables X and Y, say. Let $Z = (X, Y)'$ be a two dimensional random vector. For any $A, B \in \mathcal{B}(\mathbf{R})$, we get

$$\begin{aligned} \mathrm{P}_Z(A \times B) &= \mathbf{P}(Z^{-1}(A \times B)) \\ &= \mathbf{P}(X^{-1}(A) \cap Y^{-1}(B)) \\ &= \mathbf{P}(X^{-1}(A))\,\mathbf{P}(Y^{-1}(B)) \\ &= \mathrm{P}_X(A)\mathrm{P}_Y(B) \end{aligned}$$

as required. The proof for the general case with more than two random variables is just a trivial extension.

5. Exercises

1. Let X be a random variable with distribution function F. Define $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$. Obtain the distribution functions of X^+ and X^-.

2. For a random variable X, define

$$F(x) = \mathbf{P}\left\{\omega | X(\omega) \geq x\right\}$$

State and prove the properties of F corresponding to the usual distribution function.

3. Let E and F be two events with probabilities $\mathbf{P}(E) = 1/2, \mathbf{P}(F) = 2/3$ and $\mathbf{P}(E \cap F) = 1/3$. Define random variables $X = \mathrm{I}(E)$ and $Y = \mathrm{I}(F)$. Find the joint distribution of X and Y. Also, obtain the conditional distribution of X given Y.

4. Let X and Y be two random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Is it possible that X and Y have the same distribution, and yet $\mathbf{P}\{\omega | X(\omega) \neq Y(\omega)\} = 1$?

5. For each of the probability density functions of X given by

$$p_1(x) = \frac{x^2}{18}\,\mathrm{I}\{-3 < x < 3\} \quad \text{and} \quad p_2(x) = \frac{x+2}{18}\,\mathrm{I}\{-2 < x < 4\}$$

compute $\mathbf{P}\{\omega | |X(\omega)| < 1\}$.

6. The joint probability density function of X and Y is given by

$$p(x,y) = 3(x+y)\,\mathrm{I}\{0 \le x+y \le 1, 0 \le x, y \le 1\}$$

(a) Find the marginal density of X.

(b) Find $\mathbf{P}\{\omega | X(\omega) + Y(\omega) < 1/2\}$.

7. Suppose X and Y are jointly distributed with the probability density function

$$p(x,y) = 2x\,\mathrm{I}\{0 \le x, y \le 1\}$$

Find $\mathbf{P}\{\omega | X^2(\omega) < Y(\omega) < X(\omega)\}$.