CHAPTER 2

RANDOM VARIABLES & PROBABILITY DISTRIBUTIONS

This chapter contains the introduction of random variables as a technical device to enable the general specification of probability distributions in one and many dimensions to be made. The key topics and techniques introduced in this chapter include the following:

- **NORMALIZATION**
- **EXPECTATION**
- **TRANSFORMATION**
- **STANDARDIZATION**
- **GENERATING FUNCTIONS**
- **JOINT MODELLING**
- **MARGINALIZATION**
- **MULTIVARIATE TRANSFORMATION**
- **MULTIVARIATE EXPECTATION & COVARIANCE**
- **SUMS OF VARIABLES**
- **ORDER STATISTICS**

2.1 CONSTRUCTING RANDOM VARIABLES & PROBABILITY MODELS

Definition 2.1.1 A **random variable** X is a function defined on a sample space Ω that associates a *real number* $X(\omega) = x$ with each possible outcome $\omega \in \Omega$.

Formally, we regard X as a (possibly many-to-one) mapping from Ω to $\mathbb{R}$

$$X : \begin{array}{l} \Omega \longrightarrow \mathbb{R} \\ \omega \longmapsto x \end{array}$$

Implication : we can associate any sample space Ω (for any experiment) with a sample space that is a set of real numbers, in which the events are subsets.

For example, we could regard set $B \subseteq \mathbb{R}$ as an event associated with event $A \subseteq \Omega$ if

$$A = \{\omega | X(\omega) = x \text{ for some } x \in B\}$$

A and B are events in **different** sample spaces but are termed **equivalent**, and

$$P[X \in B] = P(A)$$

so that, after defining the random variable X as a function on the experimental sample space, attention switches to assigning the probability P$[X \in B]$ for a set $B \subseteq .\mathbb{R}$

Note : Strictly, when referring to random variables, we should make explicit the connection to original sample space Ω, and write

$$P[X \in B] = P[\{\omega : X(\omega) \in B\}]$$

but, generally, we will suppress this and merely refer to X.

EVENTS IN $\mathbb{R}$

We will assign probability to subsets B of $\mathbb{R}$ that are equivalent to events (subsets) in Ω that form the basis of a σ-algebra of subsets of Ω.

If Ω is *countable*, $\Omega = \{\omega_1, \omega_2, ...\}$, then the events of interest will be of the form $[X = b]$, or equivalently of the form $[X \leq b]$ for $b \in \mathbb{R}$

If Ω is *uncountable*, then the events of interest will be of the form $[X \leq b]$ for $b \in$.

2.2 DISCRETE RANDOM VARIABLES

Definition 2.2.1 A random variable X is **discrete** if the set of all possible values of X (that is, the *range* of the function represented by X), denoted $\mathbb{X}$, is **countable**, that is

$$\mathbb{X} = \{x_1, x_2, ..., x_n\} \qquad \text{[FINITE]} \qquad \text{or} \qquad \mathbb{X} = \{x_1, x_2, ...\} \qquad \text{[INFINITE]}$$

Definition 2.2.2 PROBABILITY MASS FUNCTION

The function f_X, defined on $\mathbb{X}$ by

$$f_X(x) = P[X = x] x \in \mathbb{X}$$

that assigns probability to each $x \in \mathbb{X}$ is the (discrete) **probability mass function**.

NOTE: For completeness, we define

$$f_X(x) = 0 \qquad x \notin \mathbb{X}$$

so that f_X is defined for all $x \in .\mathbb{R}$ Furthermore we will regard $\mathbb{X}$ as the *support* of random variable X, that is, the set of $x \in \mathbb{R}$ such that $f_X(x) > 0$

2.2.1 PROPERTIES OF MASS FUNCTION f_X

THEOREM

A function f_X is a probability mass function for discrete random variable X with range $\mathbb{X}$ of the form $\{x_1, x_2, ...\}$ if and only if

(i) $f_X(x_i) \geq 0$

(ii) $\sum f_X(x_i) = 1$

PROOF

Events $[X = x_1], [X = x_2]$ etc. are **equivalent** to events that partition Ω, that is

$$[X = x_i] \text{ is equivalent to event } A_i = \{\omega_i\}\,.$$

hence P$[X = x_i]$ =P(A_i), and the two parts of the theorem follow immediately.

Definition 2.2.3 DISCRETE CUMULATIVE DISTRIBUTION FUNCTION

The **cumulative distribution function**, or **cdf**, F_X of a discrete random variable X is defined by

$$F_X(x) = P[X \leq x] \qquad x \in \mathbb{R}.$$

2.2.2 CONNECTION BETWEEN F_X AND f_X

THEOREM

Let X be a discrete random variable with range $\mathbb{X} = \{x_1, x_2, ...\}$, where $x_1 < x_2 <$ ..., and probability mass function f_X and cdf F_X. Then for any real value x, if $x < x_1$, then $F_X(x) = 0$, and for $x \geq x_1$,

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i)$$

and hence $f_X(x_1) = F_X(x_1)$ and

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \qquad i = 2, 3, ...$$

PROOF

Events of the form $[X \leq x_i]$ can be represented as countable unions of the events $A_i = \{\omega_i\}$. The first result therefore follows from Probability Axiom 3. The second result follows immediately.

2.2.3 PROPERTIES OF DISCRETE CDF F_X

(i) In the limiting cases,

$$\lim_{x \to -\infty} F_X(x) = 0 \qquad \lim_{x \to \infty} F_X(x) = 1.$$

(ii) F_X is **continuous from the right** (but not continuous) on $\mathbb{R}$ that is, for $x \in \mathbb{R}$,

$$\lim_{h \to 0^+} F_X(x+h) = F_X(x)$$

(iii) F_X is **non-decreasing**, that is

$$a < b \Longrightarrow F_X(a) \leq F_X(b)$$

(iv) For $a < b$,

$$P[a < X \leq b] = F_X(b) - F_X(a)$$

The functions f_X and/or F_X can be used to describe the **probability distribution** of random variable X. A graph of the function f_X is non-zero only at the elements of $\mathbb{X}$. A graph of the function F_X is a **step-function** which takes the value zero at minus infinity, the value one at infinity, and is non-decreasing with points of discontinuity at the elements of $\mathbb{X}$.

2.3 CONTINUOUS RANDOM VARIABLES

Definition 2.3.1 A random variable X is **continuous** if the range of X, $\mathbb{X}$, is **uncountable**, and the function F_X defined on $\mathbb{R}$ by

$$F_X(x) = P[X \leq x]$$

for $x \in \mathbb{R}$ is a **continuous** function on $\mathbb{R}$, that is, for $x \in \mathbb{R}$,

$$\lim_{h \to 0} F_X(x+h) = F_X(x).$$

Definition 2.3.2 CONTINUOUS CUMULATIVE DISTRIBUTION FUNCTION

The **cumulative distribution function**, or cdf, F_X of a continuous random variable X is defined by

$$F_X(x) = P[X \leq x] \qquad x \in \mathbb{R}.$$

Definition 2.3.3 PROBABILITY DENSITY FUNCTION

The **probability density function**, or pdf, f_X of a continuous random variable X is defined in terms of F_X by

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

2.3.1 PROPERTIES OF CONTINUOUS F_X AND f_X

(i) Such a function f_X need not exist but continuous random variables where f_X cannot be defined in this way will be ignored. The function f_X can be defined piecewise on intervals of $\mathbb{R}$.

(ii) For the cdf of a continuous random variable,

$$\lim_{x \to -\infty} F_X(x) = 0 \qquad \lim_{x \to \infty} F_X(x) = 1$$

(iii) Directly from the definition, at values of x where F_X is differentiable x,

$$f_X(x) = \frac{d}{dt}\left\{F_X(t)\right\}_{t=x}$$

(iv) If X is continuous,

$$f_X(x) \neq P[X = x] = \lim_{h \to 0}[F_X(x+h) - F_X(x)] = 0$$

(v) For $a < b$,

$$P[a < X \leq b] = P[a \leq X < b] = P[a \leq X \leq b] = P[a < X < b] = F_X(b) - F_X(a)$$

THEOREM

A function f_X is a pdf for a continuous random variable X if and only if

$$\text{(i)} f_X(x) \geq 0 \qquad \text{(ii)} \int_{-\infty}^{\infty} f_X(x)dx = 1$$

PROOF

Analogous to the discrete case, direct from definitions and properties of F_X.

Example 2.3.1 Consider a coin tossing experiment where a fair coin is tossed repeatedly under identical experimental conditions, with the sequence of tosses independent, until a Head is obtained. For this experiment, the sample space, Ω is then the set of sequences $(\{H\}, \{TH\}, \{TTH\}, \{TTTH\} \ldots)$ with associated probabilities 1/2, 1/4, 1/8, 1/16, ...

Define discrete random variable $X : \Omega \longrightarrow \mathbb{R}$, by $X(\omega) = x \Longleftrightarrow$ first H on toss x. Then

$$f_X(x) = P[X = x] = \left(\frac{1}{2}\right)^x \qquad x = 1, 2, 3, \ldots$$

and zero otherwise. For $x \geq 1$, let $k(x)$ be the largest integer not greater than x, then

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i) = \sum_{i=1}^{k(x)} f_X(i) = 1 - \left(\frac{1}{2}\right)^{k(x)}$$

and $F_X(x) = 0$ for $x < 1$.

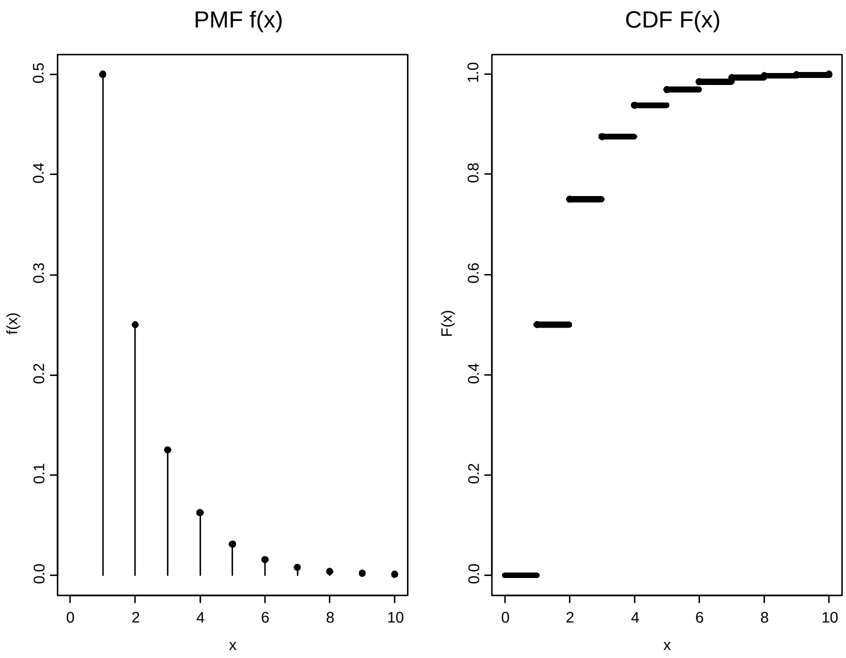

Figure 2.1: PMF $f_X(x) = \left(\frac{1}{2}\right)^x, x = 1, 2, 3, ...$ and CDF $F_X(x) = 1 - \left(\frac{1}{2}\right)^{k(x)}$

Graphs of the probability mass function (top) and cumulative distribution function (bottom) are shown in Figure 2.1. Note that the mass function is only non-zero at points that are elements of X, and that the cdf is defined for all real values of x, but is only continuous from the right. F_X is therefore a step-function.

Example 2.3.2 Consider an experiment to measure the length of time that an electrical component functions before failure. The sample space of outcomes of the experiment, Ω is $^+$, and if A_x is the event that the component functions for longer than $x > 0$ time units, suppose that $\mathrm{P}(A_x) = \exp\left\{-x^2\right\}$.

Define continuous random variable $X : \Omega \longrightarrow \mathbb{R}^+$, by $X(\omega) = x \Longleftrightarrow$ component fails at time x. Then, if $x > 0$,

$$F_X(x) = P[X \leq x] = 1 - P(A_x) = 1 - \exp\left\{-x^2\right\}$$

and $F_X(x) = 0$ if $x \leq 0$. Hence if $x > 0$,

$$f_X(x) = \frac{d}{dt}\left\{F_X(t)\right\}_{t=x} = 2x\exp\left\{-x^2\right\}$$

and zero otherwise.

Graphs of the probability density function (top) and cumulative distribution function (bottom)

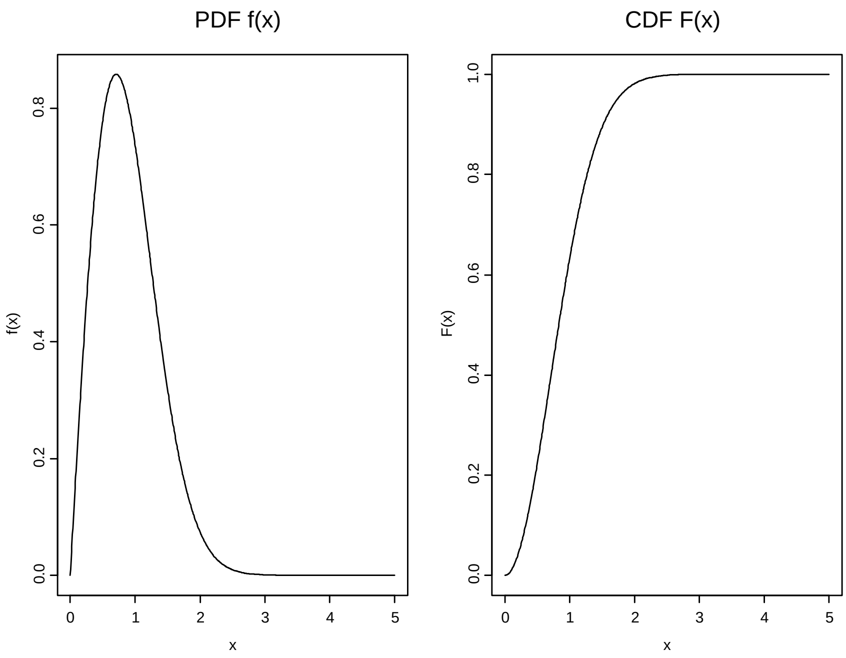

Figure 2.2: PDF $f_X(x) = 2x\exp\left\{-x^2\right\},\ x > 0$, and CDF $F_X(x) = 1 - \exp\left\{-x^2\right\}\ x > 0$

are shown in Figure 2.2. Note that both the pdf and cdf are defined for all real values of x, and that both are continuous functions. Note that here

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{0}^{x} f_X(t)dt$$

as $f_X(x) = 0$ for $x \leq 0$, and also that

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{0}^{\infty} f_X(x)dx = 1.$$

2.4 EXPECTATIONS

Definition 2.4.1 For a discrete random variable X with range X with probability mass function f_X, the **expectation** or **expected value** of X with respect to f_X is defined by

$$E_{f_X}[X] = \sum_{x=-\infty}^{\infty} x f_X(x) = \sum_{x \in \mathbb{X}} x f_X(x)$$

For a continuous random variable X with range X and pdf f_X, the **expectation** or **expected value** of X with respect to f_X is defined by

$$E_{f_X}[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{\mathbb{X}} x f_X(x)dx$$

Note : The sum/integral may not be convergent, and hence the expected value may be infinite. It is important always to check that the integral is finite: a sufficient condition is given by

$$\sum_x |x| f_X(x) < \infty \Longrightarrow \sum_x x f_X(x) = E_{f_X}[X] < \infty$$

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty \Longrightarrow \int_{-\infty}^{\infty} x f_X(x) dx = E_{f_X}[X] < \infty$$

Extension Let g be a real-valued function whose domain includes X. Then

$$E_{f_X}[g(X)] = \begin{cases} \displaystyle\sum_{x=-\infty}^{\infty} g(x) f_X(x) & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

2.4.1 PROPERTIES OF EXPECTATIONS

Let X be a random variable with mass function/pdf f_X. Let g and h be real-valued functions whose domains include X, and let a and b be constants. Then

$$E_{f_X}[ag(X) + bh(X)] = aE_{f_X}[g(X)] + bE_{f_X}[h(X)]$$

as (in the continuous case)

$$\begin{aligned} E_{f_X}[ag(X) + bh(X)] &= \int [ag(x) + bh(x)] f_X(x) dx \\ &= a \int g(x) f_X(x) dx + b \int h(x) f_X(x) dx \\ &= aE_{f_X}[g(X)] + bE_{f_X}[h(X)] \end{aligned}$$

Special Cases :

(i) For a simple linear function

$$E_{f_X}[aX + b] = aE_{f_X}[X] + b$$

(ii) Consider $g(x) = (x - \mathrm{E}_{f_X}[X])^2$. Write $\mu = \mathrm{E}_{f_X}[X]$ (a constant that does not depend on x). Then, expanding the integrand

$$\begin{aligned} E_{f_X}[g(X)] &= \int (x-\mu)^2 f_X(x) dx = \int x^2 f_X(x) dx - 2\mu \int x f_X(x) dx + \mu^2 \int f_X(x) dx \\ &= \int x^2 f_X(x) dx - 2\mu^2 + \mu^2 = \int x^2 f_X(x) dx - \mu^2 \\ &= E_{f_X}[X^2] - \{E_{f_X}[X]\}^2 \end{aligned}$$

Then

$$Var_{f_X}[X] = E_{f_X}[X^2] - \{E_{f_X}[X]\}^2$$

is the **variance** of the distribution. Similarly, $\sqrt{Var_{f_X}[X]}$ is the **standard deviation** of the distribution.

(iii) Consider $g(x) = x^k$ for $k = 1, 2, ...$ Then in the continuous case

$$E_{f_X}[g(X)] = E_{f_X}[X^k] = \int x^k f_X(x)dx,$$

and $\mathrm{E}_{f_X}[X^k]$ is the kth **moment** of the distribution.

(iv) Consider $g(x) = (x - \mu)^k$ for $k = 1, 2, ...$ Then

$$E_{f_X}[g(X)] = E_{f_X}[(X - \mu)^k] = \int (x - \mu)^k f_X(x)dx,$$

and $\mathrm{E}_{f_X}[(X - \mu)^k]$ is the kth **central moment** of the distribution.

(v) Consider $g(x) = aX + b$. Then $Var_{f_X}[aX + b] = a^2 Var_{f_X}[X]$

$$\begin{aligned} Var_{f_X}[g(X)] &= E_{f_X}[(aX + b - E_{f_X}[aX + b])^2] \\ &= E_{f_X}[(aX + b - aE_{f_X}[X] - b)^2] \\ &= E_{f_X}[(a^2(X - E_{f_X}[X])^2] \\ &= a^2 Var_{f_X}[X] \end{aligned}$$

2.4.2 APPROXIMATIONS TO MEAN AND VARIANCE

A Taylor series expansion method can be used to obtain approximations to expectations of functions of a random variable. Let X be a continuous random variable, with range X and pdf f_X. Suppose that the expectation and variance of X with respect to f_X are denoted μ and σ^2 respectively, and let g be a real-valued function whose domain includes X. Then a Taylor approximation of g around μ is given for real-value x by,

$$g(x) \approx g(\mu) + (x - \mu)g\prime(\mu) + \frac{1}{2}(x - \mu)^2 g\prime\prime(\mu)$$

where $g\prime$ and $g\prime\prime$ are the first and second derivatives of g respectively. Using the Taylor approximation, and ignoring terms in $(x - \mu)^k$ for $k = 3, 4, ...$, the expectation of $g(X)$ with respect to f_X is given approximately by

$$E_{f_X}[g(X)] \approx g(\mu) + \frac{1}{2}\sigma^2 g\prime\prime(\mu).$$

Ignoring terms in $(x-\mu)^2$ and higher, the variance of $g(X)$ with respect to f_X is given approximately by

$$Var_{f_X}[g(X)] \approx \sigma^2 \{g\prime(\mu)\}^2$$