CHAPTER 2

2. RANDOM VARIABLES & PROBABILITY DISTRIBUTIONS

DEFINITION

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 $\mathbb{R}$.

$$X: \quad \begin{aligned} \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto x \end{aligned}$$

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

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

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

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

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

CONSTRUCTING 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 $b \in \mathbb{Z}$

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

2.1. DISCRETE RANDOM VARIABLES

DEFINITION

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**,

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

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

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

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

NOTE: For completeness, we define

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

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

$$f_X(x) > 0$$

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

$$\text{(i) } f_X(x_i) \geq 0 \qquad \text{(ii) } \sum f_X(x_i) = 1$$

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

DEFINITION

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

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

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.

PROPERTIES OF F_X

(i) In the limiting cases, $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$.

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

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

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

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

(iii) $\mathrm{P}[a < X \leq b] = F_X(b) - F_X(a)$

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 $\mathbb{X}$.

2.2. CONTINUOUS RANDOM VARIABLES

DEFINITION

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

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

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

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

$$F_X(x) = \mathrm{P}[X \leq x] \quad \text{for} \quad x \in \mathbb{R}.$$

Note : for the cdf of a continuous random variable,

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

DEFINITION

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

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

NOTES:

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

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

(iii) If X is continuous, $f_X(x) \neq \mathrm{P}[X = x] = \lim_{h \to 0}[F_X(x+h) - F_X(x)] = 0$.

(iv) $\mathrm{P}[a < X \leq b] = \mathrm{P}[a \leq X < b] = \mathrm{P}[a \leq X \leq b] = \mathrm{P}[a < X < b] = F_X(b) - F_X(a)$.

THEOREM

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

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

Proof

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

EXAMPLE 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 $(\{H\}, \{TH\}, \{TTH\}, \{TTTH\} ...)$ with associated probabilities 1/2, 1/4, 1/8, 1/16, ...

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

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

and zero otherwise. For $x \in \mathbb{R}$, let $k(x)$ be the largest integer not greater than x, then

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

Graphs of the probability mass function (left) and cumulative distribution function (right) are shown in Figure 1. Note that the mass function is only non-zero at points that are elements of $\mathbb{X}$, 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: 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 $\mathbb{R}^+$, and if A_x is the event that the component functions for longer than $x > 0$ time units, suppose that $\mathrm{P}(A_x) = \exp\{-x^2\}$.

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

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

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

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

and zero otherwise.

Graphs of the probability density function (left) and cumulative distribution function (right) are shown in Figure 2. Note that both the pdf and cdf are defined for all real values of x, and that both are continous functions.

Note that here

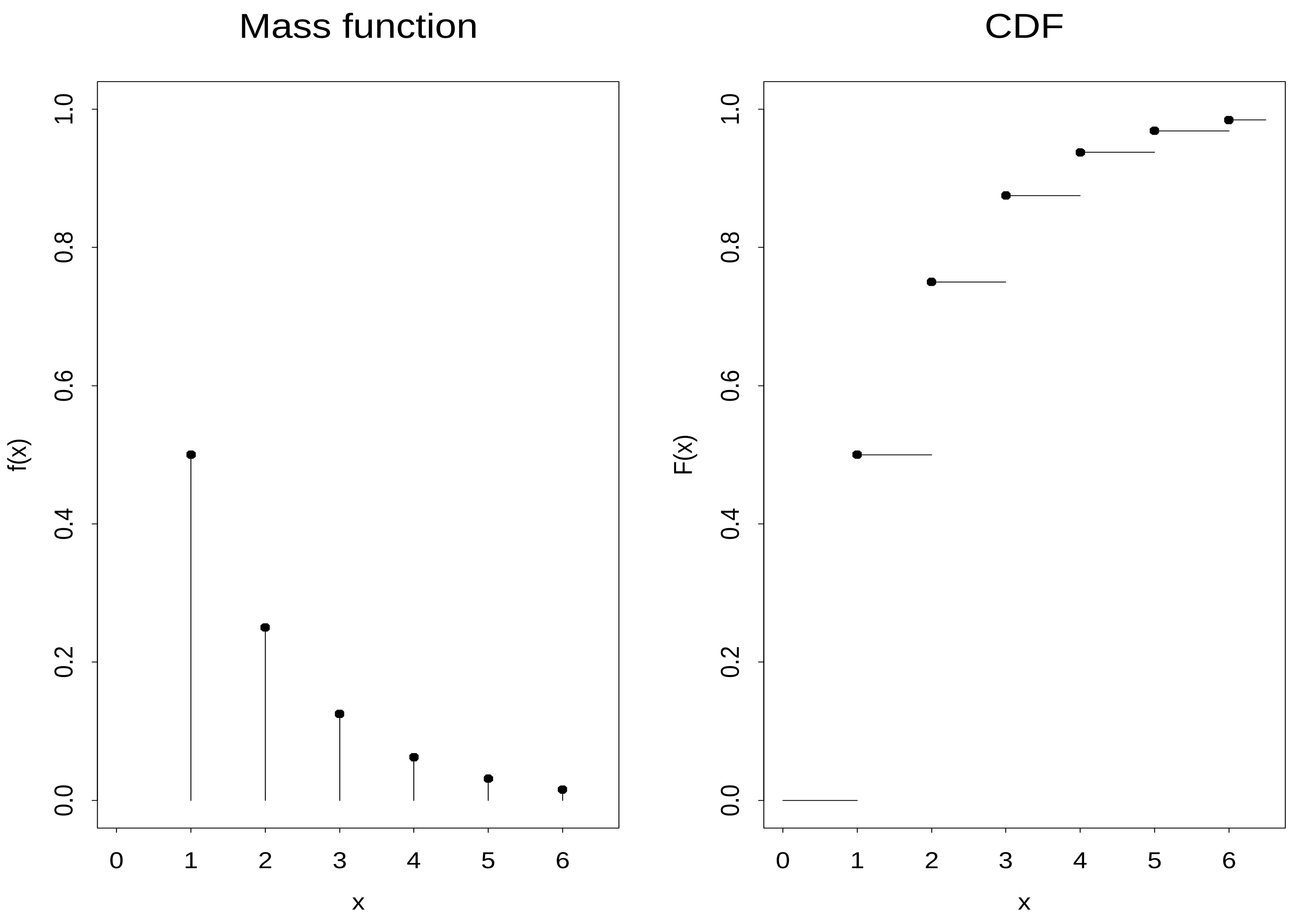
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\; dt = \int_0^x f_X(t)\; dt$$

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

$$\int_{-\infty}^{\infty} f_X(x)\; dx = \int_0^{\infty} f_X(x)\; dx = 1$$

Note that in both examples we have that

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



Mass function
f(x)
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
x
CDF
F(x)
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
x

Figure 1:

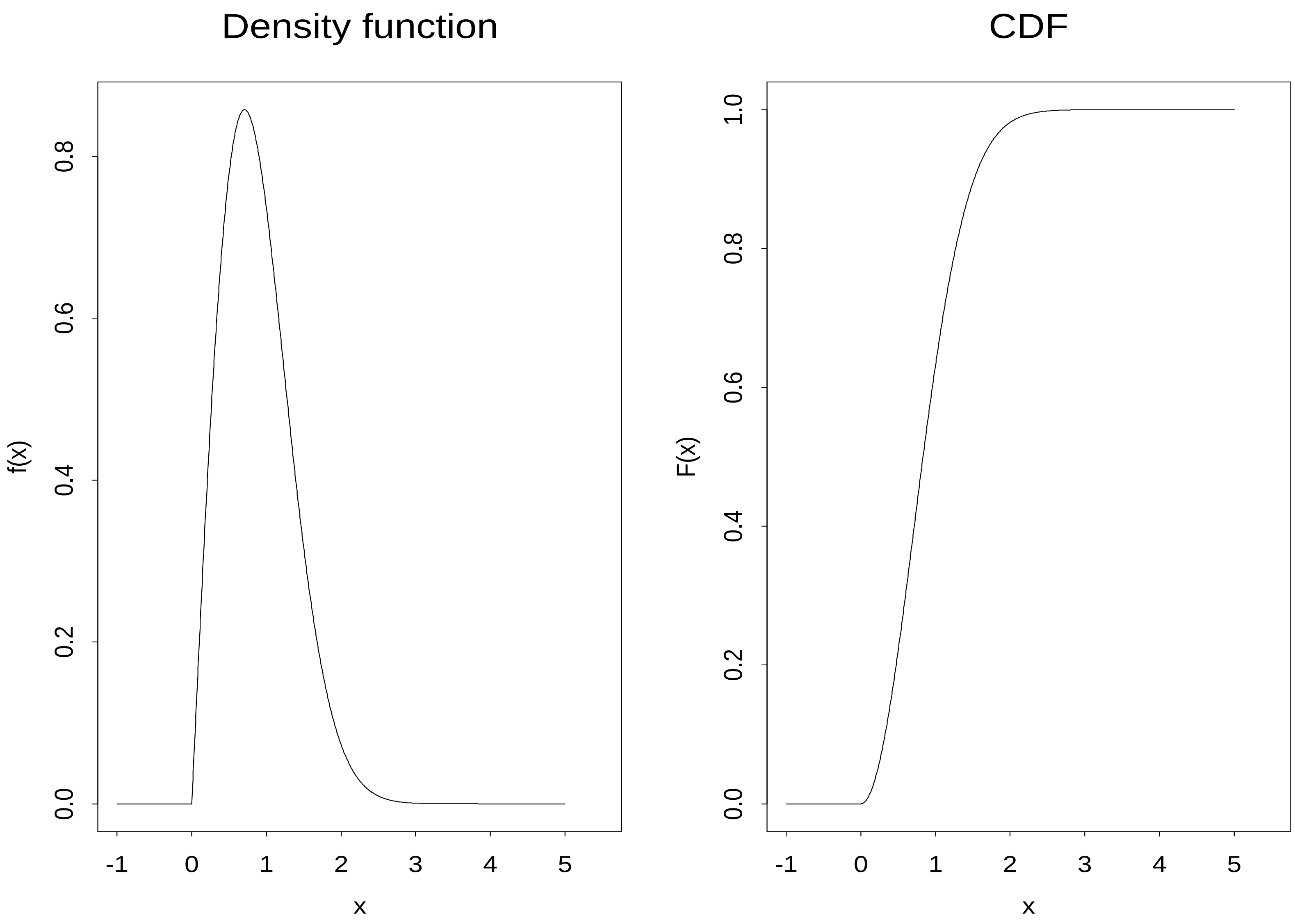

Density function
f(x)
0.0
0.2
0.4
0.6
0.8
-1
0
1
2
3
4
5
x
CDF
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
-1
0
1
2
3
4
5
x

Figure 2:

2.3. EXPECTATIONS

DEFINITION

For a **discrete** random variable X with range $\mathbb{X}$ with probability mass function f_X, the **expectation** or **expected value** of X with respect to f_X is defined by

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

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

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

NOTE: The expected value is a weighted average of the values of X, but the sum/integral may not be convergent, and hence the expected value may be infinite.

Extension Let g be a real-valued function whose domain includes $\mathbb{X}$. Then

$$\begin{aligned} \mathrm{E}_{f_X}[g(X)] &= \sum_{x=-\infty}^{\infty} g(x) f_X(x) \qquad \text{if X is discrete} \\ \mathrm{E}_{f_X}[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x)\ dx \quad \text{if X is continuous} \end{aligned}$$

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 $\mathbb{X}$, and let a and b be constants. Then

$$\mathrm{E}_{f_X}[ag(X) + bh(X)] = a\mathrm{E}_{f_X}[g(X)] + b\mathrm{E}_{f_X}[h(X)]$$

as (in the continuous case)

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

Special Cases :

(i) $\mathrm{E}_{f_X}[aX + b] = a\mathrm{E}_{f_X}[X] + b$

(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

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

Write $\mathrm{Var}_{f_X}[X] = \mathrm{E}_{f_X}[X^2] - \{\mathrm{E}_{f_X}[X]\}^2$. Then $\mathrm{Var}_{f_X}[X]$ is the **variance** of the distribution.

Similarly, $\sqrt{\mathrm{Var}_{f_X}[X]}$ is the **standard deviation** of the distribution.

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

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

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

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

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

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

(v) Consider $g(x) = aX + b$. Then

$$\mathrm{Var}_{f_X}[aX+b] = a^2 \mathrm{Var}_{f_X}[X]$$

as

$$\begin{aligned}
\mathrm{Var}_{f_X}[g(X)] &= \mathrm{E}_{f_X}[(aX + b - \mathrm{E}_{f_X}[aX+b])^2] \\
&= \mathrm{E}_{f_X}[(aX + b - a\mathrm{E}_{f_X}[X] - b)^2] \\
&= \mathrm{E}_{f_X}[(a^2(X - \mathrm{E}_{f_X}[X])^2] \\
&= a^2 \mathrm{Var}_{f_X}[X]
\end{aligned}$$

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 $\mathbb{X}$ 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 $\mathbb{X}$. Then a Taylor approximation of g around μ is given for real-value x by,

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

where g' and g'' are the first and second derivatives of g respectively.

Using the Taylor approximation, and ignoring terms in $(x-\mu)^k$ for $k = 3, 4, \ldots$, the expectation of $g(X)$ with respect to f_X is given approximately by

$$\mathrm{E}_{f_X}[g(X)] \approx g(\mu) + \frac{1}{2}\sigma^2 g''(\mu).$$

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

$$\mathrm{Var}_{f_X}[g(X)] \approx \sigma^2 \left\{g'(\mu)\right\}^2$$