Module 3

Functions of a Random Variable and Its Expectation

3.1. Probability distribution of a function of a discrete random variable

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X: \Omega \to \mathbb{R}$ be a r.v. with d.f. F, p.m.f. f, and support S. Let $h: \mathbb{R} \to \mathbb{R}$ be a given function. Define $Z: \Omega \to \mathbb{R}$ as

$$Z(\omega) = h(X(\omega)), \quad \omega \in \Omega.$$

Then Z is a r.v. and it is a function of r.v. X. Since we are only interested in values of r.v.s X and Z and not in the original probability space $(\Omega, \mathcal{F}, P)$, we simply write $X(\omega)$, $\omega \in \Omega$, as X and $Z(\omega)$, $\omega \in \Omega$, as Z. We have

$$F(x) = \Pr(X \le x), \quad x \in \mathbb{R}$$

$$f(x) = \Pr(X = x), \quad x \in \mathbb{R}, \quad \Pr(X \in S) = 1 \text{ and } \Pr(X = x) > 0, \ \forall x \in S.$$

Define $T = h(S) = \{h(x) : x \in S\}$. For any set $A \subseteq \mathbb{R}$, define $h^{-1}(A) = \{x \in S : h(x) \in A\}$. Then T is a countable set

$$\Pr(Z = z) > 0, \quad \forall z \in T \quad \left(\text{since } \Pr(X = x) > 0, \ \forall x \in S\right)$$

and

$$\Pr(Z \in T) = 1 \quad (\text{since } \Pr(X \in S) = 1)$$

It follows that Z is a discrete r.v. Moreover, for $z \in T$,

$$\begin{aligned}
\Pr(Z = z) &= \Pr(h(X) = z) \\
&= \sum_{\{x \in S : h(x) = z\}} \Pr(X = x) \\
&= \sum_{x \in h^{-1}(\{z\})} \Pr(X = x) \\
&= \sum_{x \in h^{-1}(z)} f(x)
\end{aligned}$$

and, for any $z \notin T$, $\Pr(Z = z) = 0$.

Thus we have the following theorem.

Theorem 3.1.1. Let X be a discrete r.v. with support S, d.f. F and p.m.f. f. Let $h: \mathbb{R} \to \mathbb{R}$ be a given function. Then $Z = h(X)$ is a discrete r.v., with support $T = \{h(x): x \in S\}$, p.m.f.

$$g(z) = \begin{cases} \sum_{x \in h^{-1}(\{z\})} f(x), & \text{if } z \in T \\ 0, & \text{otherwise} \end{cases}$$

and d.f.

$$G(z) = \Pr(Z \le z) = \sum_{\{t \in T: t \le z\}} g(t) = \sum_{\{x \in S: h(x) \le z\}} f(x)$$

$$= \sum_{x \in h^{-1}((-\infty, z]) \cap S} f(x).$$

In particular when $h: S \to \mathbb{R}$ is one-one, then

$$g(z) = \begin{cases} f(h^{-1}(z)), & \text{if } z \in T \\ 0, & \text{otherwise} \end{cases}$$

Example 3.1.1. Let X be a discrete r.v. with p.m.f

$$f(x) = \begin{cases} \frac{1}{7}, & \text{if } x \in \{-2, -1, 0, 1\} \\ \frac{3}{14}, & \text{if } x \in \{2, 3\} \\ 0, & \text{otherwise} \end{cases}.$$

Find the p.m.f. and d.f. of $Y = X^2$.

Solution Here the support of X is $S = \{-2, -1, 0, 1, 2, 3\}$. By Theorem 3.1.1 $Y = X^2$ is a discrete r.v. with support $T = \{0, 1, 4, 9\}$ and p.m.f.

$$g(z) = \Pr(X^2 = z) = \begin{cases} \Pr(X=0), & \text{if } z = 0 \\ \Pr(X=-1) + \Pr(X=1), & \text{if } z = 1 \\ \Pr(X=-2) + \Pr(X=2), & \text{if } z = 4 \\ \Pr(X=-3) + \Pr(X=3), & \text{if } z = 9 \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{1}{7}, & \text{if } z = 0 \\ \frac{2}{7}, & \text{if } z = 1 \\ \frac{5}{14}, & \text{if } z = 4 \\ \frac{3}{14}, & \text{if } z = 9 \\ 0, & \text{otherwise} \end{cases}$$

The d.f. of Y is

$$G(z) = \Pr(Y \le z) = \begin{cases} 0, & \text{if } z < 0 \\ \frac{1}{7}, & \text{if } 0 \le z < 1 \\ \frac{3}{7}, & \text{if } 1 \le z < 4 \\ \frac{11}{14}, & \text{if } 4 \le z < 9 \\ 1, & \text{if } z \ge 9 \end{cases}.$$

Example 3.1.2. In Example 3.1.1, directly find the d.f. of $Y = X^2$ (i.e. find d.f. of Y before finding the p.m.f. of Y). Hence find the p.m.f. of Y.

Solution By Theorem 3.1.1 Y is a discrete r.v. with support $T = \{0, 1, 4, 9\}$. Thus the d.f. of Y is

$$G(z) = Pr(Y \le z) = P(X^2 \le z) = \begin{cases} 0 & z<0 \\ Pr(X^2=0) & 0 \le z < 1 \\ Pr(X^2=0)+Pr(X^2=1), & 1 \le z < 4 \\ Pr(X^2=0)+Pr(X^2=1)+Pr(X^2=4) & 4 \le z < 9 \\ 1 & z \ge 9 \end{cases}$$

$$= \begin{cases} 0, & z<0 \\ \frac{1}{7}, & 0 \le z < 1 \\ \frac{1}{7}+\frac{1}{7}+\frac{1}{7}, & 1 \le z < 4 \\ \frac{1}{7}+\frac{1}{7}+\frac{1}{7}+\frac{1}{7}+\frac{3}{14}, & 4 \le z < 9 \\ 1 & z \ge 9 \end{cases}$$

$$= \begin{cases} 0, & z<0 \\ \frac{1}{7}, & 0 \le z < 1 \\ \frac{3}{7}, & 1 \le z < 4 \\ \frac{11}{14}, & 4 \le z < 9 \\ 1, & z \ge 9 \end{cases}$$

The p.m.f. of Y is

$$g(z) = \begin{cases} G(z)-G(z^-), & \text{if } z \in T \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{1}{7}, & \text{if } z=0 \\ \frac{2}{7}, & \text{if } z=1 \\ \frac{5}{14}, & \text{if } z=4 \\ \frac{3}{14}, & \text{if } z=9 \\ 0, & \text{otherwise.} \end{cases}$$

$\boxed{4/3}$

3.2. Probability distribution of a function of a continuous random variable

Let X be a continuous r.v. with d.f. F, p.d.f. $f(\cdot)$ and support

$$S = \{x \in \mathbb{R}: F(x+h) - F(x-h) = \int_{x-h}^{x+h} f(t)\,dt > 0, \ \forall h > 0\}.$$

For convenience assume that $S = [a, b]$ and $\{x \in \mathbb{R}: f(x) > 0\} = (a, b)$ for some $-\infty \le a < b \le \infty$ (with the convention that $[-\infty, b] \equiv (-\infty, b]$ $\forall b \in \mathbb{R}$, $[a, \infty] \equiv [a, \infty)$, $\forall a \in \mathbb{R}$ and $[-\infty, \infty] \equiv (-\infty, \infty)$).

Let $h: \mathbb{R} \to \mathbb{R}$ be a function such that h is strictly monotone and differentiable function on S. Then $Z = h(X)$ is a r.v. with d.f.

$$G(z) = \Pr(Z \le z)$$
$$= \Pr(h(X) \le z), \quad z \in \mathbb{R}.$$

For any sets $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$, define

$$h(A) = \{h(x): x \in A\}$$

and

$$h^{-1}(B) = \{x \in \mathbb{R}: h(x) \in B\}.$$

Clearly $\Pr(X \in (a, b)) = 1$ and therefore

$$\Pr(h(X) \in h((a, b)) = 1).$$

Consider the following cases.

$\boxed{\text{Case I: } h(\cdot) \text{ is strictly increasing on } S}$

We have

$$\Pr(h(a) < Z < h(b)) = 1.$$

Therefore, for $z < h(a)$ $\Pr(Z \le z) = 0$ and, for $z \ge h(b)$, $\Pr(Z \le z) = 1$. For $h(a) < z < h(b)$

$$G(z) = \Pr(h(X) \le z)$$

$$= \Pr(X \le h^{-1}(z))$$

$$= \int_{-\infty}^{h^{-1}(z)} f(t)\,dt$$

$$= \int_{a}^{h^{-1}(z)} f(t)\,dt = \int_{h(a)}^{z} f(h^{-1}(y)) \left|\frac{d}{dy} h^{-1}(y)\right| dy$$

Thus

$$G(z) = \begin{cases} 0 & \text{if } z < h(a) \\ \int_{h(a)}^{z} f(h^{-1}(y)) \left|\frac{d}{dy} h^{-1}(y)\right| dy & \text{if } h(a) \le z < h(b) \\ 0, & \text{if } z \ge h(b) \end{cases}$$

Since f is continuous on (a, b) it follows that $G(z)$ is differentiable everywhere except possibly at $z = h(a)$ and $z = h(b)$.

Moreover

$$G'(z) = \begin{cases} f(h^{-1}(z)) \left|\frac{d}{dz} h^{-1}(z)\right|, & \text{if } h(a) < z < h(b) \\ 0, & \text{otherwise} \end{cases}$$

and

$$\int_{-\infty}^{\infty} G'(z)\,dz = \int_{h(a)}^{h(b)} f(h^{-1}(z)) \left|\frac{d}{dz} h^{-1}(z)\right| dz$$

$$= \int_{a}^{b} f(t)\,dt$$

$$= 1.$$

It follows that Z is a continuous r.v. with p.d.f.

$$g(z) = \begin{cases} f(h^{-1}(z)) \left|\frac{d}{dz} h^{-1}(z)\right|, & \text{if } h(a) < z < h(b) \\ 0, & \text{otherwise} \end{cases}$$

and support $S = [h(a), h(b)]$.

Case II. $h(\cdot)$ is strictly decreasing on S

here

$$\Pr\left(h(b) < h(X) < h(a)\right) = 1$$

and

$$G(z) = \Pr\left(h(X) \le z\right), \quad z \in \mathbb{R}.$$

Clearly, for $z \le h(b)$, $G(z) = 0$, and for $z \ge h(a)$, $G(z) = 1$.

For $h(b) < z < h(a)$

$$\begin{aligned} G(z) &= \Pr\left(X \ge h^{-1}(z)\right) \\ &= \int_{h^{-1}(z)}^{a} f(t)\,dt \\ &= \int_{h^{-1}(z)}^{b} f(t)\,dt \\ &= \int_{h(b)}^{z} f(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| dy \qquad \text{if } z \le h(b) \end{aligned}$$

Thus

$$\begin{cases} 0, & \\ \displaystyle\int_{h(b)}^{z} f(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y)\, dy \right|, & \text{if } h(b) \le z < h(a) \\ 1, & \text{if } z \ge h(a) \end{cases}$$

Since f is continuous on (a, b), it follows that $G(\cdot)$ is differentiable everywhere except possibly at $h(a)$ and $h(b)$. Moreover

$$G'(z) = \begin{cases} f(h^{-1}(z)) \left| \frac{d}{dz} h^{-1}(z) \right|, & \text{if } h(b) < z < h(a) \\ 0, & \text{otherwise} \end{cases}$$

and

$$\int_{-\infty}^{\infty} G'(z)\,dz = \int_{h(b)}^{h(a)} f(h^{-1}(z)) \left| \frac{d}{dz} h^{-1}(z) \right| dz = \int_{a}^{b} f(t)\,dt = 1.$$

Consequently Z is a continuous r.v. with p.d.f.

$$g(z) = \begin{cases} f(h^{-1}(z)) \left| \frac{d}{dz} h^{-1}(z) \right|, & \text{if } h(b) < z < h(a) \\ 0, & \text{otherwise} \end{cases}$$

and support $S = [h(b), h(a)]$.

Combining Case I and Case II, we get the following result.

$\boxed{\text{Theorem 3.2.1.}}$ Let X be a continuous r.v. with support $S = [a, b]$, for some $-\infty \le a < b \le \infty$. Suppose that $\{x \in \mathbb{R} : f(x) > 0\} = (a, b)$ and that f is continuous on (a, b). Let $h: \mathbb{R} \to \mathbb{R}$ be a function that is differentiable and strictly monotone on (a, b). Then $Z = h(X)$ is a continuous r.v. with p.d.f.

$$\begin{cases} f(h^{-1}(z)) \left| \frac{d}{dz} h^{-1}(z) \right|, & \text{if } z \in h((a, b)) \\ 0, & \text{otherwise} \end{cases}$$

and support $S = [\min\{h(a), h(b)\}, \max\{h(a), h(b)\}]$.

The following theorem is a generalization of the above result and can be proved on similar lines.

$\boxed{\text{Theorem 3.2.2.}}$ Let X be a continuous r.v. with support $S = \bigcup_{i \in \Lambda} [a_i, b_i]$, where Λ is a countable set and $[a_i, b_i]$, $i \in \Lambda$, are disjoint intervals. Suppose that $\{x \in \mathbb{R} : f(x) > 0\} = \bigcup_{i \in \Lambda} (a_i, b_i)$ and that f is continuous in each (a_i, b_i), $i \in \Lambda$. Let $h: \mathbb{R} \to \mathbb{R}$ be a function that is differentiable and strictly monotone in each (a_i, b_i), $i \in \Lambda$ (h may be monotonic increasing in some (a_i, b_i) and monotonic decreasing in some (a_i, b_i)). Let $h_i^{-1}(\cdot)$ be the inverse function of h_i on (a_i, b_i), $i \in \Lambda$. Then $Z = h(X)$ is a continuous r.v. with p.d.f.

$$g(z) = \sum_{j \in \Lambda} f(h_j^{-1}(z)) \left| \frac{d}{dz} h_j^{-1}(z) \right| I_{h((a_j, b_j))}(z)$$

Remark 3.2.1. Theorems 3.2.1 and 3.2.2 hold even in situations where the function h is differentiable everywhere except possibly at a finite number of points in S.

Example 3.2.1. Let X be a r.v. with p.d.f.

$$f(x) = \begin{cases} 3x^2, & \text{if } 0<x<1 \\ 0, & \text{otherwise} \end{cases}$$

Find the p.d.f. and d.f. of $Y = \frac{1}{X^2}$. What is the support of d.f. of Y.

Solution The support of F is $[0, 1]$ and $\{x \in \mathbb{R} : f(x) > 0\} = (0, 1)$. Moreover f is continuous on $(0, 1)$ and $h(x) = \frac{1}{x^2}$ is differentiable and strictly monotone on $(0, 1)$. $h((0, 1)) = (1, \infty)$, $h^{-1}(y) = \frac{1}{\sqrt{y}}$ and $\frac{d}{dy} h^{-1}(y) = -\frac{1}{2y\sqrt{y}}$, $y \in (1, \infty)$. Thus $Y = X^2$ is a continuous r.v. with p.d.f. of Y is

$$g(y) = f(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| I_{h((0,1))}(y)$$

$$= f(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| I_{(1,\infty)}(y)$$

$$= \begin{cases} \frac{3}{y} \times \frac{1}{2y\sqrt{y}}, & \text{if } y>1 \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{3}{2y^2\sqrt{y}}, & \text{if } y>1 \\ 0, & \text{otherwise} \end{cases}$$

The d.f. of Y is

$$G(y) = \int_{-\infty}^{y} g(t)\, dt$$

$$= \begin{cases} 0, & \text{if } y<1 \\ \int_1^y \frac{3}{2t^2\sqrt{t}}\, dt, & \text{if } y>1 \end{cases} = \begin{cases} 0, & \text{if } y<1 \\ 1 - \frac{1}{y^{3/2}}, & \text{if } y>1 \end{cases}$$

Clearly the support of G is $[1, \infty)$.

$\boxed{\text{Example 3.2.2}}$ Let X be r.v. with p.d.f.

$$f(x) = \begin{cases} \frac{|x|}{2}, & \text{if } -1<x<1 \\ \frac{x}{3}, & \text{if } 1\le x\le 2 \\ 0, & \text{otherwise} \end{cases}$$

and let $Y = X^2$.

(a) Find the p.d.f. of Y directly and hence find the d.f. of Y.

(b) Find the d.f. of Y and hence find the p.d.f. of Y.

(c) Find the support of d.f. of Y.

$\boxed{\text{Solution}}$ (a) The support of F is $S=[-1, 2]$ and we may take $S=[-1, 0] \cup [0, 2]$, $\{x\in\mathbb{R}: f(x)>0\} = (-1, 0)\cup(0, 2)$. f is continuous on $(-1, 0)\cup(0, 2)$, $h(x)=x^2$ is differentiable on $(-1, 0)\cup(0, 2)$, $h(\cdot)$ is strictly $\downarrow$ on $(-1, 0)$ and strictly $\uparrow$ on $(0, 2)$.

$h(x)=x^2$ is strictly $\downarrow$ on $S_1=(-1, 0)$ with inverse function

$$h_1^{-1}(y) = -\sqrt{y}, \quad y\in(0, 1), \quad h(S_1)=(0, 1)$$

$h(x)=x^2$ is strictly $\downarrow$ on $S_2=(0, 1)$ with inverse function

$$h_2^{-1}(y)=\sqrt{y}, \quad y\in(0, 4), \quad h(S_2)=(0, 4)$$

Thus $Y=X^2$ is a continuous r.v. with p.d.f

$$g(y) = f(h_1^{-1}(y))\left|\frac{d}{dy}h_1^{-1}(y)\right| I_{(0,1)}(y) + f(h_2^{-1}(y))\left|\frac{d}{dy}h_2^{-1}(y)\right| I_{(0,4)}(y)$$

$$= f(-\sqrt{y})\left|\frac{-1}{2\sqrt{y}}\right| I_{(0,1)}(y) + f(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right| I_{(0,4)}(y)$$

$$= \frac{1}{2\sqrt{y}}\left[f(-\sqrt{y})\, I_{(0,1)}(y) + f(\sqrt{y})\, I_{(0,4)}(y)\right]$$

$$= \begin{cases} \frac{1}{2}, & \text{if } 0<y<1 \\ \frac{1}{6}, & \text{if } 1<y<4 \\ 0, & \text{otherwise} \end{cases}$$

~~The d.f. of Y is~~

$$\cancel{G(y) = \Pr(X^2 \le y) = \begin{cases} 0 & \text{if } y < \end{cases}}$$

$$G(y) = \Pr(X^2 \le y) = \int_{-\infty}^{y} g(t)\,dt = \begin{cases} 0, & \text{if } y<0 \\ \int_0^y \frac{1}{2}\,dt, & \text{if } 0 \le y < 1 \\ \int_0^1 \frac{dt}{2} + \int_1^y \frac{dt}{6}, & \text{if } 1 \le y < 4 \\ 1, & \text{if } y \ge 4 \end{cases}$$

$$= \begin{cases} 0, & \text{if } y<0 \\ \frac{y}{2}, & \text{if } 0 \le y < 1 \\ \frac{y+2}{6}, & \text{if } 1 \le y < 4 \\ 1 & \text{if } y \ge 1 \end{cases}.$$

(b) The d.f. of Y is

$$G(y) = \Pr(X^2 \le y) = \begin{cases} 0, & \text{if } y<0 \\ \Pr(-\sqrt{y} \le X \le \sqrt{y}), & \text{if } y \ge 0 \end{cases}.$$

For $0 \le y < 1$,

$$G(y) = \Pr(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{|x|}{2}\,dx = \frac{y}{2}$$

For $1 \le y < 4$ (note that $-2 < -\sqrt{y} \le -1$ and $1 \le \sqrt{y} \le 2$)

$$G(y) = \Pr(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-1}^{1} \frac{|x|}{2}\,dx + \int_1^{\sqrt{y}} \frac{x}{3}\,dx = \frac{y+2}{6}$$

For $y \geq 4$ $G(y) = 1$.

Therefore

$$G(y) = \begin{cases} 0, & \text{if } y < 0 \\ \frac{y}{2}, & \text{if } 0 \leq y < 1 \\ \frac{y+2}{6}, & \text{if } 1 \leq y < 4 \\ 1, & \text{if } y \geq 4 \end{cases}$$

Clearly G is differentiable everywhere except at finite number of points (0, 1 and 4) and we may take

$$G'(y) = \begin{cases} \frac{1}{2}, & \text{if } 0 < y < 1 \\ \frac{1}{6}, & \text{if } 1 < y < 4 \\ 0, & \text{otherwise} \end{cases}$$

Moreover

$$\int_{-\infty}^{\infty} G'(y)\, dy = \int_0^1 \frac{1}{2} dy + \int_1^4 \frac{1}{6} dy = 1$$

Thus Y is a continuous r.v. with p.d.f.

$$g(y) = \begin{cases} \frac{1}{2}, & \text{if } 0 < y < 1 \\ \frac{1}{6}, & \text{if } 1 < y < 4 \\ 0, & \text{otherwise} \end{cases}$$

3.3. Expectation (or Mean) of Random Variables

Let X be a discrete r.v. with p.m.f. $f(\cdot)$ and support S. For any $x \in S$ $f(x)$ gives an idea about proportion ~~of values x in the population~~ of times we will observe the event $\{X = x\}$ if the experiment is repeated a large number of times. Thus $\sum_{x \in S} x\, f(x)$ represents the mean (or expected) value of r.v. X if the experiment is repeated a large number of times.

Similarly if X is a continuous r.v. with p.d.f. $f(\cdot)$ then

$$\int_{-\infty}^{\infty} x f(x)\, dx \qquad \text{(provided the integral is finite)}$$

represents the mean (or expected) value of r.v. X.

Definition 3.3.1. (a) Let X be a discrete r.v. with p.m.f. $f(\cdot)$ and support S. We say that the expected value of X (or the mean of X, which we denote by $E(X)$) is finite and equals

$$E(X) = \sum_{x \in S} x f(x),$$

provided $\sum_{x \in S} |x| f(x) dx < \infty$.

(b) Let X be continuous r.v. with p.d.f. $f(\cdot)$ and support S. We say that the expected value of X (or the mean of X, which we denote by $E(X)$) is finite and equals

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx,$$

provided $\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$.

Example 3.3.1. (a) Let X be a discrete r.v. with p.m.f.

$$f(x) = \begin{cases} \left(\frac{1}{2}\right)^x, & \text{if } x \in \{1, 2, 3, \ldots\} \\ 0, & \text{otherwise} \end{cases}$$

Show that $E(X)$ is finite. Find $E(X)$.

(b) Let X be a r.v. with p.m.f.

$$f(x) = \begin{cases} \frac{3}{\pi^2 x^2}, & \text{if } x \in \{\pm 1, \pm 2, \ldots\} \\ 0, & \text{otherwise} \end{cases}$$

Show that $E(X)$ is not finite.

(c) Let X be a continuous r.v. with p.d.f.

$$f(x) = \frac{e^{-|x|}}{2}, \quad -\infty < x < \infty.$$

Show that $E(X)$ is finite. Find $E(X)$.

(d) Let X be a continuous r.v. with p.d.f.

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

Show that $E(X)$ is not finite.

$\boxed{\text{Solution}}$ (a) The support of the distribution is $S = \{1, 2, \ldots\}$.

Also

$$\sum_{x \in S} |x| f(x) = \sum_{n=1}^{\infty} \frac{n}{2^n} = \sum_{n \geq 1} a_n \text{ (say)},$$

where $a_n = \frac{n}{2^n} > 0$, $\forall n = 1, 2, \ldots$ and

$$\frac{a_{n+1}}{a_n} = \frac{n+1}{2n} \to \frac{1}{2} < 1, \text{ as } n \to \infty$$

Thus, by the ratio test

$$\sum_{x \in S} |x| f(x) = \sum_{n=1}^{\infty} \frac{n}{2^n} < \infty$$

It can be seen that $E(X) = 2$ (Exercise)

(b) Here the support of the distribution is $S = \{\pm 1, \pm 2, \pm 3, \ldots\}$

$$\sum_{x \in S_X} |x| f(x) = \frac{6}{\pi^2} \sum_{n \geq 1} \frac{1}{n} = \infty$$

Hence $E(X)$ is not finite.

(c) We have

$$\int_{-\infty}^{\infty} |x| f(x)\, dx = \int_{-\infty}^{\infty} |x| \frac{e^{-|x|}}{2}\, dx = \int_0^{\infty} x e^{-x}\, dx = 1 < \infty$$

$\Rightarrow E(X)$ is finite and

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-\infty}^{\infty} x \frac{e^{-|x|}}{2}\,dx = 0$$

(d) We have

$$\int_{-\infty}^{\infty} |x| f(x)\,dx = \int_{-\infty}^{\infty} |x| \cdot \frac{1}{\pi} \cdot \frac{1}{1+x^2}\,dx$$

$$= \frac{2}{\pi} \int_{0}^{\infty} \frac{x}{1+x^2}\,dx = \infty$$

$\Rightarrow$ $E(X)$ is not finite.

Example 3.3.2. **(St. Petersburg Paradox)** To make some money a gambler plays a sequence of fair games with the following strategy.

In the first bet bet Rs. 1 million. If the first bet is lost he doubles his bet in the second game. He keeps on doubling his bet until he wins a game.

If the gambler has not won by the nth trial he bets Rs. 2^n million in the $(n+1)$th game.

If he wins in kth game then:

Investment $= 1 + 2 + 4 + \cdots + 2^{k-1} = 2^k - 1$ million rupee

Win $= 2^k$ million rupee

Total earning if he wins on kth game $= 1$ million rupee

The above scheme seems to be fool proof for earning Rs. 1 million.

By this logic all gambler's should be billionaires!

X: the amount of money bet on the last game (the game he wins).

Then
$$P(X = 2^k) = \frac{1}{2^{k+1}}, \quad k = 0, 1, 2, \ldots$$

$$E(X) = \sum_{k=0}^{\infty} 2^k \times \frac{1}{2^{k+1}} = \infty \quad (E(X) \text{ is not finite})$$

$\rightarrow$ Enormous amount of money would be required.

Theorem 3.3.1. Let X be a discrete or continuous r.v. Then

$$E(X) = \int_0^\infty \Pr(X>y)\,dy - \int_{-\infty}^0 \Pr(X<y)\,dy,$$

provided $E(X)$ is finite.

Proof We will provide the proof for the case when X is a continuous r.v. with p.d.f., say f. We have

$$E(X) = \int_{-\infty}^\infty x f(x)\,dx$$

$$= \int_{-\infty}^0 x f(x)\,dx + \int_0^\infty x f(x)\,dx$$

$$= -\underbrace{\int_{-\infty}^0 \int_x^0}_{x<y<0} f(x)\,dy\,dx + \underbrace{\int_0^\infty \int_0^x}_{0<y<x} f(x)\,dy\,dx$$

$$= -\int_{-\infty}^0 \int_{-\infty}^y f(x)\,dx\,dy + \int_0^\infty \int_y^\infty f(x)\,dx\,dy$$

$$= -\int_{-\infty}^0 \Pr(X<y)\,dy + \int_0^\infty \Pr(X>y)\,dy.$$

Corollary 3.3.1. (a) Suppose that X is a discrete or continuous r.v. with $\Pr(X \geq 0)=1$. Then

$$E(X) = \int_0^\infty \Pr(X>y)\,dy$$

(b) Suppose that $\Pr(X \in \{0, \pm 1, \pm 2, \ldots\})=1$. Then

$$E(X) = \sum_{n\geq 1}^\infty P(X \geq n) - \sum_{n\geq 1}^\infty \Pr(X \leq -n).$$

(c) Suppose that $\Pr(X \in \{0, 1, 2, \ldots\})=1$. Then

$$E(X) = \sum_{n\geq 1}^\infty \Pr(X \geq n).$$

Proof. Exercise.

The following theorem suggests, that for any r.v. X and any function $h: \mathbb{R} \to \mathbb{R}$, $E(h(X))$ can be directly found using p.m.f./p.d.f. of X.

Theorem 3.3.2. (a) Let X be a discrete r.v. with support S and p.m.f. f. Let $h: \mathbb{R} \to \mathbb{R}$ be a given function, and let $Z = h(X)$. Then

$$E(Z) = \sum_{x \in S} h(x) f(x),$$

provided $\sum_{x \in S} |h(x)| f(x) < \infty$.

(b) Let X be a continuous r.v. with p.d.f. f and let $h: \mathbb{R} \to \mathbb{R}$ be a given function. If $Z = h(X)$, then

$$E(Z) = \int_{-\infty}^{\infty} h(x) f(x)\, dx,$$

provided $\int_{-\infty}^{\infty} |h(x)| f(x)\, dx < \infty$.

Proof We will provide the proof of (a) only. The proof of (b) follows on similar lines. The support of $Z = h(X)$ is $T = h(S)$. We have

$$
\begin{aligned}
E(T) &= \sum_{t \in T} t \Pr(T = t) \\
&= \sum_{t \in T} t \Pr(h(X) = t) \\
&= \sum_{t \in T} t \Big\{ \sum_{\{x \in S:\, h(x) = t\}} \Pr(X = x) \Big\} \\
&= \sum_{t \in T} \sum_{\{x \in S:\, h(x) = t\}} t \Pr(X = x) \\
&= \sum_{t \in T} \sum_{\{x \in S:\, h(x) = t\}} h(x) \Pr(X = x) \\
&= \sum_{t \in T} \sum_{\{x \in S:\, h(x) = t\}} h(x) \Pr(X = x) = \sum_{\bigcup_{t \in T} \{x \in S:\, h(x) = t\}} h(x) \Pr(X = x) \\
&= \sum_{x \in S} h(x) \Pr(X = x)
\end{aligned}
$$

~~16/3~~ 17/3

Example 3.3.2 (a) Let the r.v. X have the p.m.f.

$$f(x) = \begin{cases} \frac{1}{6}, & \text{if } x = -2, -1, 0, 1, 2, 3 \\ 0 & \text{otherwise.} \end{cases}$$

Fin $E(X^2)$.

(b) Let the r.v. X have the p.d.f.

$$f(x) = \begin{cases} 2x, & \text{if } 0<x<1 \\ 0, & \text{otherwise} \end{cases}$$

Find $E(X^3)$.

Solution (a)

$$E(X^2) = \sum_{x \in S} x^2 f(x) = 4 \times \frac{1}{6} + 1 \times \frac{1}{6} + 0 \times \frac{1}{6} + 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6}$$

$$= \frac{19}{6}$$

(b) $$E(X^3) = \int_{-\infty}^{\infty} x^3 f(x)\, dx$$

$$= 2 \int_0^1 x^4\, dx = \frac{2}{5}.$$

Theorem 3.3.3. Let X be a discrete or continuous r.v. with p.d.f./p.m.f. f and support S and let $h_i : \mathbb{R} \to \mathbb{R}$ $(i = 1, 2, \ldots, m)$ be given functions.

(a) Then, for real constants $c_1, \ldots, c_m$

$$E\left(\sum_{i=1}^{m} c_i h_i(X)\right) = \sum_{i=1}^{m} c_i\, E(h_i(X)),$$

provided involved expectations are finite.

(b) Let $h_1(x) \le h_2(x)$, $\forall x \in S$. Then

$$E(h_1(X)) \le E(h_2(X))$$

provided involved expectations are finite. In particular if $E(X)$ is finite and $\Pr(a \le X \le b) = 1$, for some real constants a and b $(a<b)$ then $a \le E(X) \le b$.

(c) If $\Pr(X \geq 0) = 1$ and $E(X) = 0$ then $\Pr(X=0) = 1$

(d) If $E(X)$ is finite then $|E(X)| \leq E(|X|)$

(e) For real constants a and b

$$E(aX+b) = aE(X) + b.$$

provided involved expectations are finite.

$\boxed{\text{Proof.}}$ The proofs for assertions (a), (b) and (e) are obvious

(c) We will provide the proof for the case when X is a continuous r.v. Then

$$\begin{aligned}
\Pr(X > 0) &= \Pr\left(\bigcup_{n \geq 1}^{\infty} \{X \geq \tfrac{1}{n}\}\right) \\
&= \lim_{n\to\infty} \Pr(X \geq \tfrac{1}{n}) \qquad (\{X \geq \tfrac{1}{n}\} \uparrow) \\
&= \lim_{n\to\infty} \int_{1/n}^{\infty} f(x)\,dx \\
&\leq \lim_{n\to\infty} \int_{1/n}^{\infty} nx\, f(x)\,dx \qquad (\because x \in [\tfrac{1}{n}, \infty) \Rightarrow nx \geq 1) \\
&\leq \lim_{n\to\infty} \left[n \int_{0}^{\infty} x f(x)\,dx\right] \\
&= \lim_{n\to\infty} [n E(X)] = 0
\end{aligned}$$

$$\Rightarrow \Pr(X > 0) = 0$$

$$\Rightarrow \Pr(X = 0) = 1$$

(d) We have

$$-|X| \leq X \leq |X|$$

$$\Rightarrow E(-|X|) \leq E(X) \leq E(|X|)$$

$$\Rightarrow |E(X)| \leq E(|X|).$$

Some Special Expectations

(i) $h(x)=x$: $E(X)=\mu_1'$ = mean of (distribution of) X

(ii) $h(x)=x^r$, $r\in\{1,2,\ldots\}$: $E(X^r)=\mu_r'$ = r-th moment of X about origin.

(iii) $h(x)=|x|^r$, $r\in\{1,2,\ldots\}$: $E(|X|^r)$ = r-th absolute moment of X about origin;

(IV) $h(x)=(x-\mu_1')^r$, $r\in\{1,2,\ldots\}$; $E((X-\mu_1')^r)=\mu_r$ = r-th moment of X about its mean or r-th central moment.

(V) $\mu_2=E((X-\mu_1')^2)=\sigma^2$ = Variance of X (also denoted by $\mathrm{Var}(X)$)

σ = standard deviation of X.

Remark 3.3.1.

(a)
$$\begin{aligned}\mathrm{Var}(X)=\sigma^2 &= E((X-\mu_1')^2)\\ &= E(X^2-2\mu_1'X+(\mu_1')^2)\\ &= E(X^2)-2\mu_1'E(X)+(\mu_1')^2\\ &= E(X^2)-2(\mu_1')^2+(\mu_1')^2\\ &= E(X^2)-(\mu_1')^2\\ &= E(X^2)-(E(X))^2.\end{aligned}$$

(b) Since $(X-\mu_1')^2\geq 0$, we have

$$\mathrm{Var}(X)=E((X-\mu_1')^2)\geq 0$$
$$\Rightarrow\quad E(X^2)\geq (E(X))^2.$$

(c) $\mathrm{Var}(X)=0 \iff E((X-\mu_1')^2)=0$

$$\iff \Pr(X=E(X))=1.$$

Theorem 3.3.4. Let X be a r.v. such that $E(|X|^s) < \infty$, for some $s > 0$. Then $E(|X|^r) < \infty$, $\forall\ 0 < r < s$.

Proof. Note that

$$|X|^r \le \max\{|X|^s, 1\}$$
$$\le |X|^s + 1$$
$$\Rightarrow E(|X|^r) \le E(|X|^s + 1) = E(|X|^s) + 1 < \infty$$

Thus the result follows.

3.4. Moment Generating Function

Let X be a r.v. with d.f. F and p.m.f./p.d.f. f

Definition 3.4.1. We say that the moment generating function (m.g.f.) of X (denoted by $M_X(t)$) exists and equals

$$M_X(t) = E(e^{tX}),$$

provided $E(e^{tX})$ is finite in $(-h, h)$ for some $h > 0$.

Remark 3.4.1. (i) $M_X(0) = 0$. Thus $A = \{t \in \mathbb{R} : E(e^{tX}) \text{ is finite}\} \neq \phi$.

(ii) $M_X(t) > 0$, $\forall\ t \in A = \{s \in \mathbb{R} : E(e^{sX}) \text{ is finite}\}$

(iii) Suppose that $M_X(t)$ exists and is finite on $(-h, h)$, for some $h > 0$. For real constants c and d, let $Y = cX + d$. Then the m.g.f. of Y also exists and is finite on $\left(-\frac{h}{|c|}, \frac{h}{|c|}\right)$ (with the convention that $\pm\frac{a}{0} = \pm\infty$ if $a > 0$). Moreover

$$M_Y(t) = E(e^{t(cX+d)})$$
$$= e^{td} M_X(ct), \quad t \in \left(-\frac{h}{|c|}, \frac{h}{|c|}\right).$$

(iv) The name m.g.f. to the transform M_X is motivated by the fact that M_X can be used to generate moments of any r.v., as illustrated in the following theorem.

Theorem 3.4.1. Let X be a r.v. with m.g.f. M_X that is finite on $(-h, h)$, $h>0$. Then

(a) for each $r \in \{1, 2, \dots\}$, $\mu_r' = E(X^r)$ is finite;

(b) for each $r \in \{1, 2, \dots\}$, $\mu_r' = E(X^r) = M_X^{(r)}(0)$, where $M_X^{(r)}(0) = \left[\frac{d^r}{dt^r} M_X(t)\right]_{t=0}$, the r-th derivative of M_X at the point 0;

(c) $M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{\underline{|r}} \mu_r'$, $t \in (-h, h)$ so that μ_r' is equal to coefficient of $\frac{t^r}{\underline{|r}}$ $(r = 1, 2, \dots)$ in the Maclaurin's series expansion of $M_X(t)$ around $t=0$.

Proof. (a) We have

$$E(e^{tX}) < \infty, \quad \forall\, t \in (-h, h)$$

$$\Rightarrow \int_{-\infty}^{0} e^{tx} f(x)\,dx < \infty, \ \forall\, t \in (-h, h), \text{ and, } \int_{0}^{\infty} e^{tx} f(x)\,dx < \infty \ \forall\, t \in (-h, h)$$

$$\Rightarrow \int_{-\infty}^{0} e^{-t|x|} f(x)\,dx < \infty \ \forall\, t \in (-h, h) \text{ and } \int_{0}^{\infty} e^{t|x|} f(x)\,dx < \infty \ \forall\, t \in (-h, h)$$

$$\Rightarrow \int_{-\infty}^{0} e^{|t||x|} f(x)\,dx < \infty \ \forall\, t \in (-h, h) \text{ and } \int_{0}^{\infty} e^{|t||x|} f(x)\,dx < \infty, \ \forall\, t \in (-h, h)$$

$$\Rightarrow \int_{-\infty}^{\infty} e^{|tx|} f(x) < \infty, \quad \forall\, t \in (-h, h);$$

here $f(\cdot)$ denotes the p.d.f. of r.v. X.

Fix $r \in \{1, 2, \dots\}$ and $t \in (-h, h) - \{0\}$. Then $\lim_{|x| \to \infty} \frac{|x|^r}{e^{|tx|}} = 0$ and therefore $\exists$ a positive real number $A_{r,t}$ such that $|x|^r < e^{|tx|}$, $\forall\, |x| > A_{r,t}$. Therefore

$$E(|X|^r) = \int_{-\infty}^{\infty} |x|^r f(x)\,dx$$

$$= \int_{|x| \le A_{r,t}} |x|^r f(x)\,dx + \int_{|x| > A_{r,t}} |x|^r f(x)\,dx$$

$$\le A_{r,t}^r \int_{|x| \le A_{r,t}} f(x)\,dx + \int_{|x| > A_{r,t}} e^{|tx|} f(x)\,dx$$

$$\le A_{r,t}^r + \int_{-\infty}^{\infty} e^{|tx|} f(x)\,dx < \infty, \quad \forall\, r = 1, 2, \ldots$$

(b) $$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

$$M_X^{(r)}(t) = \frac{d^r}{dt^r} \int_{-\infty}^{\infty} e^{tx} f(x)\,dx, \quad r = 1, 2, \ldots$$

Using the arguments of advanced calculus it can be shown that if $M_X(t) = E(e^{tX}) < \infty$, $\forall\, t \in (-h, h)$ then the derivatives can be passed through the integral sign. Therefore

$$M_X^{(r)}(t) = \int_{-\infty}^{\infty} \frac{d^r}{dt^r}(e^{tx} f(x))dx$$

$$= \int_{-\infty}^{\infty} x^r e^{tx} f(x)\,dx, \quad r = 1, 2, \ldots$$

$$M_X^{(r)}(0) = \int_{-\infty}^{\infty} x^r f(x)\,dx = E(X^r)$$

(c) $$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_{-\infty}^{\infty} \left(\sum_{r=0}^{\infty} \frac{t^r x^r}{\underline{|r}} \right) f(x)\,dx.$$

Under the assumption that $M_X(t) = E(e^{tX}) < \infty$, $\forall\, t \in (-h, h)$, using arguments of advanced calculus, it can be shown that the summation sign passed can be passed through the integral sign. Then

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{\underline{|r}} \int_{-\infty}^{\infty} x^r f(x)\,dx = \sum_{r=0}^{\infty} \frac{t^r}{\underline{|r}} E(X^r),$$

$r = 1, 2, \ldots$

Corollary 3.4.1 Under the notation and assumption of the above theorem, let $\psi_X(t) = \ln M_X(t)$, $t \in (-h, h)$. Then

$$\mu_1' = E(X) = \psi_X^{(1)}(0)$$

and $$\mu_2 = Var(X) = \psi_X^{(2)}(0).$$

Proof. For $t \in (-h, h)$

$$\psi_X^{(1)}(t) = \frac{M_X^{(1)}(t)}{M_X(t)} \text{ and } \psi_X^{(2)}(t) = \frac{M_X(t)\, M_X^{(2)}(t) - \left(M_X^{(1)}(t)\right)^2}{\left(M_X(t)\right)^2}$$

$$\Rightarrow \psi_X^{(1)}(0) = M_X^{(1)}(0) = E(X) \quad (M_X(0) = 1)$$

$$\text{and } \psi_X^{(2)}(0) = M_X^{(2)}(0) - \left(M_X^{(1)}(0)\right)^2$$
$$= E(X^2) - (E(X))^2 = Var(X).$$

Example 3.4.1 (a) Let X be a discrete r.v. with p.m.f.

$$f_X(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & \text{if } x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases},$$

where $\lambda > 0$. Show that the m.g.f of X exists and is finite on whole $\mathbb{R}$. Find $M_X(t)$, mean and variance of X and $E(X^3)$.

(b) Let X be a continuous r.v. with p.d.f

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases},$$

where $\lambda > 0$. Find the m.g.f. of X ~~(provided it exists)~~, mean and variance of X and $E(X^r)$, $r = 1, 2, \ldots$ (provided they exist).

(c) Let X be a continuous r.v. having the p.d.f. (called Cauchy p.d.f. and the corresponding probability distribution is called Cauchy distribution). Show that the m.g.f. of X does not exist.

Solution (a) We have

$$\sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^{x}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^{t})^{x}}{x!} = e^{-\lambda} e^{\lambda e^{t}} = e^{\lambda(e^{t}-1)}, \quad \forall\, t\in\mathbb{R}$$

Thus m.g.f. of X exists and is finite on whole of $\mathbb{R}$.

Moreover

$$M_X(t) = e^{\lambda(e^{t}-1)}, \quad t\in\mathbb{R}$$

$$\Psi_X(t) = \ln(M_X(t)) = \lambda(e^{t}-1)$$

$$\Psi_X^{(1)}(t) = \Psi_X^{(2)}(t) = \lambda e^{t}, \quad t\in\mathbb{R}$$

$$\Rightarrow E(X) = \Psi_X^{(1)}(0) = \lambda \quad \text{and} \quad Var(X) = \Psi_X^{(2)}(0) = \lambda.$$

$$M_X^{(1)}(t) = \lambda e^{t} e^{\lambda(e^{t}-1)} = \lambda e^{t} M_X(t)$$

$$M_X^{(2)}(t) = \lambda e^{t} M_X^{(1)}(t) + \lambda e^{t} M_X(t)$$

$$M_X^{(3)}(t) = \lambda e^{t} M_X^{(2)}(t) + 2\lambda e^{t} M_X^{(1)}(t) + \lambda e^{t} M_X(t)$$

$$\Rightarrow E(X) = M_X^{(1)}(0) = \lambda$$

$$E(X^2) = M_X^{(2)}(0) = \lambda M_X^{(1)}(0) + \lambda = \lambda^2 + \lambda$$

$$E(X^3) = M_X^{(3)}(0) = \lambda M_X^{(2)}(0) + 2\lambda M_X^{(1)}(0) + \lambda$$

$$= \lambda(\lambda^2+\lambda) + 2\lambda^2 + \lambda$$

$$= \lambda^3 + 3\lambda^2 + \lambda.$$

Alternatively, for $t\in\mathbb{R}$,

$$M_X(t) = e^{\lambda(e^{t}-1)}$$

$$= 1 + \lambda(e^{t}-1) + \frac{\lambda^2(e^{t}-1)^2}{2!} + \frac{\lambda^3(e^{t}-1)^3}{3!} + \frac{\lambda^4(e^{t}-1)^4}{4!} + \cdots$$

$$= 1 + \lambda\left(\sum_{j=1}^{\infty}\frac{t^j}{j!}\right) + \frac{\lambda^2}{2!}\left(\sum_{j=1}^{\infty}\frac{t^j}{j!}\right)^2 + \frac{\lambda^3}{3!}\left(\sum_{j=1}^{\infty}\frac{t^j}{j!}\right)^3 + \frac{\lambda^4}{4!}\left(\sum_{j=1}^{\infty}\frac{t^j}{j!}\right)^4 + \cdots$$

$$= 1 + \lambda t + t^2\left(\frac{\lambda}{2!} + \frac{\lambda^2}{2!}\right) + t^3\left(\frac{\lambda}{3!} + \frac{2\lambda^2}{(2!)^2} + \frac{\lambda^3}{3!}\right) + t^4(\cdots) + \cdots, \quad t\in\mathbb{R}$$

$E(X)$ = Coefficient of t in the expansion of $m_X(t) = \lambda$

$E(X^2)$ = Coefficient of $\frac{t^2}{\underline{|2}}$ in the expansion of $m_X(t) = \lambda + \lambda^2$

$E(X^3)$ = Coefficient of $\frac{t^3}{\underline{|3}}$ in the expansion of $m_X(t) = \lambda^3 + 3\lambda^2 + \lambda$.

(b) $$\int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx = \lambda \int_0^{\infty} e^{-\lambda(1-\frac{t}{\lambda})x}\,dx < \infty, \quad \text{if } t < \lambda$$

Thus the m.g.f of X exists and, for $t<\lambda$,

$$m_X(t) = \left(1-\frac{t}{\lambda}\right)^{-1}$$
$$= 1 + \frac{t}{\lambda} + \frac{t^2}{\lambda^2} + \cdots + \frac{t^r}{\lambda^r} + \cdots.$$

For $r = 1, 2, \ldots$

$\mu_r' = E(X^r)$ = Coefficient of $\frac{t^r}{\underline{|r}}$ in the expansion of $m_X(t)$

$$= \frac{\underline{|r}}{\lambda^r}, \quad r \in \{1, 2, \ldots\}.$$

Alternatively,

$$m_X^{(1)}(t) = \frac{1}{\lambda}\left(1-\frac{t}{\lambda}\right)^{-2}, \quad m_X^{(2)}(t) = \frac{2}{\lambda^2}\left(1-\frac{t}{\lambda}\right)^{-3} \text{ and}$$

$$m_X^{(r)}(t) = \frac{\underline{|r}}{\lambda^r}\left(1-\frac{t}{\lambda}\right)^{-(r+1)}, \quad t<\lambda.$$

$$\Rightarrow E(X^r) = m_X^{(r)}(0) = \frac{\underline{|r}}{\lambda^r}, \quad r = 1, 2, \ldots.$$

$$Var(X) = E(X^2) - (E(X))^2$$
$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

(c) Since $E(X)$ is not finite, the m.g.f. of X does not exist.

Definition 3.4.2 (Equality in Distribution) Let X and Y be two r.v.s with d.f.s F_X and F_Y, respectively. We say that X and Y have the same distribution (written as $X \stackrel{d}{=} Y$) if $F_X(x) = F_Y(x)$, $\forall x \in \mathbb{R}$.

Remark 3.4.2. (a) Let X and Y be discrete r.v.s with p.m.f.s f_X and f_Y, respectively. Then

$$X \stackrel{d}{=} Y \iff f_X(x) = f_Y(x), \quad \forall x \in \mathbb{R}.$$

(b) Let X and Y be continuous r.v.s. Then $X \stackrel{d}{=} Y$ iff there exist versions of p.d.f.s f_X and f_Y of X and Y, respectively, such that $f_X(x) = f_Y(x)$, $\forall x \in \mathbb{R}$.

(c) Suppose that $X \stackrel{d}{=} Y$. Then for any function $h: \mathbb{R} \to \mathbb{R}$ $h(X) \stackrel{d}{=} h(Y)$ and hence $E(h(X)) = E(h(Y))$.

Theorem 3.4.2. Let X and Y be r.v.s such that, for some $c>0$, $M_X(t) = M_Y(t)$, $\forall t \in (-c, c)$. Then $X \stackrel{d}{=} Y$.

Proof (Special case) Suppose that X and Y are discrete r.v.s with supports $S_X = S_Y = \{1, 2, \ldots\}$,

$$p_k = \Pr(X = k) \quad \text{and} \quad q_k = \Pr(Y = k), \quad k = 1, 2, \ldots$$

Then

$$M_X(t) = M_Y(t), \quad \forall t \in (-c, c), \text{ for some } c>0$$

$$\Rightarrow \sum_{k=1}^{\infty} e^{kt} p_k = \sum_{k=1}^{\infty} e^{kt} q_k, \quad \forall t \in (-c, c)$$

$$\Rightarrow \sum_{k=1}^{\infty} s^k p_k = \sum_{k=1}^{\infty} s^k q_k, \quad \forall s \in (e^{-c}, e^{c})$$

$$\Rightarrow p_k = q_k, \quad \forall k = 1, 2, \ldots,$$

since if two power series are equal over an interval then their coefficients are the same. Thus $X \stackrel{d}{=} Y$.

Example 3.4.2. For any $p \in (0,1)$ and positive integer n let $X_{p,n}$ be discrete r.v. with p.m.f.

$$f_{p,n}(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x \in \{0,1,2,\dots,n\} \\ 0, & \text{otherwise} \end{cases}; \quad n \in \mathbb{N},\ p \in (0,1).$$

(such a r.v. probability (or distribution) is called binomial (distribution) r.v. with n trials and probability of success p). Define $Y_{p,n} = n - X_{p,n}$, $p \in (0,1)$, $n \in \mathbb{N}$. Using the m.g.f. of $X_{p,n}$, show that $Y_{p,n} \overset{d}{=} X_{1-p,n}$ and $E(X_{\frac{1}{2},n}) = \frac{n}{2}$.

Solution We have

$$\begin{aligned} M_{X_{p,n}}(t) &= E(e^{tX_{p,n}}) \\ &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x} \\ &= (1-p+pe^t)^n, \quad t \in \mathbb{R} \end{aligned}$$

$$\begin{aligned} M_{Y_{p,n}}(t) &= E(e^{tY_{p,n}}) \\ &= E(e^{t(n-X_{p,n})}) \\ &= e^{nt} M_{X_{p,n}}(-t) \\ &= e^{nt}(1-p+pe^{-t})^n \\ &= (p+(1-p)e^t)^n = (1-(1-p)+(1-p)e^t)^n \\ &= M_{X_{1-p,n}}(t), \quad \forall t \in \mathbb{R} \end{aligned}$$

Thus

$$Y_{p,n} \overset{d}{=} X_{1-p,n}$$

$$f_{Y_{p,n}}(y) = Pr(Y_{p,n} = y)$$

$$= Pr(X_{p,n} = n-y)$$

$$= \begin{cases} \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)}, & \text{if } n-y \in \{0, 1, 2, \dots, n\} \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \binom{n}{y} (1-p)^{y} (1-(1-p))^{n-y}, & \text{if } y \in \{0, 1, 2, \dots, n\} \\ 0, & \text{otherwise} \end{cases}$$

$$= f_{X_{1-p,n}}(y), \quad \forall y \in \mathbb{R}$$

$$\Rightarrow Y_{p,n} \stackrel{d}{=} X_{1-p,n}$$

For $p = \frac{1}{2}$

$$X_{\frac{1}{2},n} \stackrel{d}{=} n - X_{\frac{1}{2},n}$$

$$\Rightarrow E(X_{\frac{1}{2},n}) = E(n - X_{\frac{1}{2},n})$$

$$\Rightarrow E(X_{\frac{1}{2},n}) = \frac{n}{2}$$

$\boxed{\text{Example 3.4.3.}}$ Let X be a r.v. with p.d.f.

$$f_X(x) = \frac{e^{-|x|}}{2}, \quad -\infty < x < \infty,$$

and let $Y = -X$. Show that $Y \stackrel{d}{=} X$ and hence show that $E(X) = 0$.

$\boxed{\text{Solution}}$ We have

$$M_Y(t) = E(e^{tY}) = E(e^{-tX})$$

$$= \int_{-\infty}^{\infty} e^{-tx} \frac{e^{-|x|}}{2} \, dx$$

$$= \int_{-\infty}^{\infty} e^{+tx} e^{-\frac{|x|}{2}} \, dx$$

$$= M_X(t), \quad \forall t \in (-1,1)$$

$$\Big[M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{e^{-|x|}}{2} dx = \int_{-\infty}^{0} e^{tx} \frac{e^{+x}}{2} dx + \int_{0}^{\infty} e^{tx} \frac{e^{-x}}{2} dx$$

$$= \frac{1}{2}\left[\int_0^\infty e^{-(1+t)x} dx + \int_0^\infty e^{-(1-t)x} dx\right]$$

$$= \frac{1}{2}\left[\frac{1}{1+t} + \frac{1}{1-t}\right], \quad \forall t \in (-1,1)$$

$$= \frac{1}{1-t^2}, \quad \forall t \in (-1,1) \Big]$$

$$\Rightarrow \quad X \overset{d}{=} Y$$

Alternatively the p.d.f. of Y is

$$f_Y(y) = \frac{e^{-|y|}}{2}, \quad -\infty < y < \infty$$

$$= f_X(y), \quad \forall -\infty < y < \infty$$

$$\Rightarrow \quad X \overset{d}{=} Y.$$

Thus

$$E(Y) = E(X)$$

$$\Rightarrow E(-X) = E(X) \qquad \left(\text{since } \int_{-\infty}^{\infty} |x| f_X(x)\, dx < \infty\right).$$

$$\Rightarrow E(X) = 0.$$

3.5. Inequalities

Inequalities provide estimates of probabilities when they can not be evaluated precisely.

Theorem 3.5.1. Let X be a r.v. and let $g: \mathbb{R} \to \mathbb{R}$ be a non-negative function such that $E(g(X))$ is finite. Then for any $c > 0$,

$$\Pr(g(X) \geq c) \leq \frac{E(g(X))}{c}.$$

Proof. (For the case when X is a continuous r.v.).

Let $A = \{x \in \mathbb{R} : g(x) \geq c\}$ and let $f_X(\cdot)$ denote the p.d.f. of X. Then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

$$= \int_{-\infty}^{\infty} g(x)\, [I_A(x) + I_{A^c}(x)]\, f_X(x)\, dx$$

$$= \int_{-\infty}^{\infty} g(x)\, I_A(x) f_X(x)\, dx + \int_{-\infty}^{\infty} g(x)\, I_{A^c}(x)\, f_X(x)\, dx$$

$$\geq \int_{-\infty}^{\infty} g(x)\, I_A(x)\, f_X(x)\, dx$$

$$\geq c \int_{-\infty}^{\infty} I_A(x)\, f_X(x)\, dx = c \int_A f_X(x)\, dx$$

$$= c \Pr(g(X) \geq c)$$

$$\Rightarrow \quad \Pr(g(X) \geq c) \leq \frac{E(g(X))}{c}$$

$\boxed{\text{Corollary 3.5.1.}}$ (a) Let $g : [0, \infty) \to \mathbb{R}$ be a non-negative and strictly increasing function such that $E(g(X))$ is finite. Then, for any $c > 0$ such that $g(c) > 0$,

$$\Pr(|X| \geq c) \leq \frac{E(g(|X|))}{c}.$$

(b) Let $r > 0$ and $t > 0$. Then

$$\Pr(|X| \geq t) \leq \frac{E(|X|^r)}{t^r}, \quad \text{(Markov's Inequality)}$$

provided $E(|X|^r) < \infty$. In particular

$$\Pr(|X| \geq t) \leq \frac{E(|X|)}{t},$$

provided $E(|X|) < \infty$

$\boxed{\text{Proof.}}$ (a) $\Pr(|X| \geq c) = \Pr(g(|X|) \geq g(c))$ (since g is strictly $\uparrow$)

$$\leq \frac{E(g(|X|))}{g(c)} \quad \text{(Theorem 3.5.1)}$$

(b) Let $g : [0, \infty) \to \mathbb{R}$ be given by

$$g(x) = x^r, \quad x \geq 0, \; r > 0.$$

Then g is strictly increasing on $[0, \infty)$ and is non-negative. Using (a) we get

$$\Pr(|X| \geq t) \leq \frac{E(g(|X|))}{g(t)} = \frac{E(|X|^r)}{t^r}.$$

$\boxed{\text{Theorem 3.5.2.}}$ (Chebyshev Inequality) Let X be a r.v. with finite variance σ^2 and $E(X) = \mu$. Then, for any $\varepsilon > 0$,

$$\Pr(|X-\mu| \geq \varepsilon\sigma) \leq \frac{1}{\varepsilon^2}.$$

$\boxed{\text{Proof.}}$ Using the above Corollary

$$\Pr(|X-\mu| \geq \varepsilon\sigma) \leq \frac{E(|X-\mu|^2)}{\varepsilon^2\sigma^2}$$

$$= \frac{E((X-\mu)^2)}{\varepsilon^2\sigma^2} = \frac{1}{\varepsilon^2}.$$

$\boxed{\text{Example 3.5.1.}}$ (The above bounds are sharp).

Let X be a r.v. with p.m.f.

$$f(x) = \begin{cases} \frac{1}{8}, & \text{if } x = -1, 1 \\ \frac{3}{4}, & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$E(X^2) = \frac{1}{4}, \ \& \ \Pr(|X| \geq 1) = \frac{1}{4}.$$

Using the Markov inequality

$$\Pr(|X| \geq 1) \leq E(X^2) = \frac{1}{4}.$$

$\boxed{\text{Example 3.5.2.}}$ Let X be a r.v. with p.d.f.

$$f(x) = \begin{cases} \frac{1}{2\sqrt{3}}, & -\sqrt{3} < x < \sqrt{3} \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\mu = E(X) = \int_{-\sqrt{3}}^{\sqrt{3}} x \times \frac{1}{2\sqrt{3}}\, dx = 0$$

$$\sigma^2 = E(X^2) = \int_{-\sqrt{3}}^{\sqrt{3}} x^2 \times \frac{1}{2\sqrt{3}}\, dx = 1$$

and $\Pr(|X| \geq \frac{3}{2}) = 1 - \int_{-3/2}^{3/2} \frac{1}{2\sqrt{3}}\, dx = 1 - \frac{\sqrt{3}}{2} = 0.134$

Using the Markov inequality

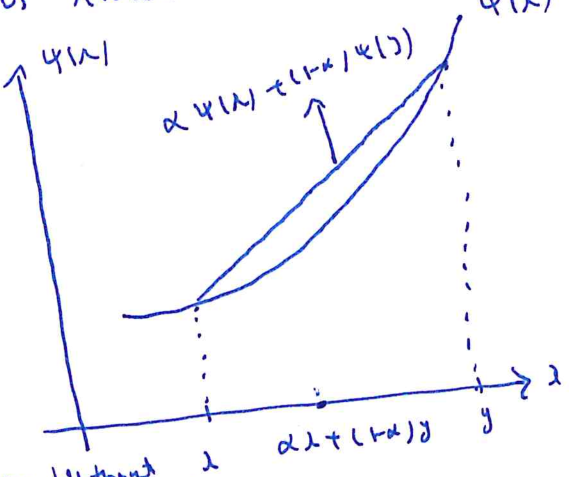
$$\Pr\left(|X| \geq \frac{3}{2}\right) \leq \frac{4}{9} E(X^2) = \frac{4}{9} = 0.444\ldots$$

$\downarrow$ Considerably Conservative.

$\boxed{\text{Definition 3.5.1.}}$ Let $-\infty \leq a < b \leq \infty$. A function $\psi : (a, b) \to \mathbb{R}$ is said to be a convex function if

$$\psi(\alpha x + (1-\alpha) y) \leq \alpha \psi(x) + (1-\alpha) \psi(y), \quad \forall x, y \in (a, b) \text{ and } \forall \alpha \in (0, 1).$$

The function $\psi(\cdot)$ is said to be strictly convex if the above inequality is strict.



Chord is above the curve.

We state the following theorem without providing its proof.

$\boxed{\text{Theorem 3.5.3.}}$ (a) Let $\psi : (a, b) \to \mathbb{R}$ be a convex function. Then ψ is continuous on (a, b) and is almost everywhere differentiable (i.e. if D is the set of points where ψ is not differentiable then D does not contain any interval)

(b) Let $\psi : (a, b) \to \mathbb{R}$ be a differentiable function. Then ψ is convex (strictly convex) on (a, b) iff ψ' is non-decreasing (strictly increasing) on (a, b).

(c) Let $\psi : (a, b) \to \mathbb{R}$ be a twice differentiable function. Then ψ is convex (strictly convex) on (a, b) iff $\psi''(x) \geq (>) \, 0 \quad \forall x \in (a, b)$.

Theorem 3.5.4. (Jensen's Inequality). Let $\psi : (a, b) \to \mathbb{R}$ be a convex function and let X be a r.v. with d.f. F having support $S \subseteq (a, b)$. Then

$$E(\psi(X)) \geq \psi(E(X)),$$

provided the expectations exist.

Proof. We provide the proove for the special case when ψ is twice differentiable on (a, b) so that $\psi''(x) \geq 0 \quad \forall x \in (a, b)$

Then, for $\mu = E(X)$,

$$\psi(x) = \psi(\mu) + (x-\mu)\psi'(\mu) + \frac{(x-\mu)^2}{L^2}\psi''(\xi), \quad \forall x \in (a, b),$$

for some ξ between μ and x. Clearly

$$\psi(x) \geq \psi(\mu) + (x-\mu)\psi'(\mu), \quad \forall x \in (a, b)$$

$$\Rightarrow E(\psi(X)) \geq E(\psi(\mu) + (X-\mu)\psi'(\mu))$$

$$= \psi(\mu)$$

$$= \psi(E(X)).$$

Example 3.5.3. (a) For any r.v. X

$$E(X^2) \geq (E(X))^2 \quad (\psi(x) = x^2, \; x \in \mathbb{R} \text{ is convex})$$

and $$E(|X|) \geq |E(X)| \quad (\psi(x) = |x|, \; x \in \mathbb{R} \text{ is convex})$$

(b) For any r.v. X with $\Pr(X > 0) = 1$

$$E(\ln X) \leq \ln E(X) \quad (\psi(x) = -\ln x, \text{ is convex on } (0, \infty))$$

(c) For any r.v. X

$$E(e^X) \geq e^{E(X)} \quad (\psi(x) = e^x, \; x \in \mathbb{R}, \text{ is convex})$$

(d) For any r.v. X with $\Pr(X > 0) = 1$

$$E(X)\, E\left(\tfrac{1}{X}\right) \geq 1 \quad \left(\psi(x) = \tfrac{1}{x}, \; x > 0, \text{ is convex}\right)$$

Example 3.5.4. Let $a_1 \cdots a_n$, $w_1 \cdots w_n$ be positive constants such that $\sum_{i=1}^{n} w_i = 1$. Prove the AM-GM-HM inequality

$$\sum_{i=1}^{n} a_i w_i \geq \prod_{i=1}^{n} a_i^{w_i} \geq \frac{1}{\sum_{i=1}^{n} \frac{w_i}{a_i}}.$$

$$(AM \geq GM \geq HM).$$

Solution Let X be a r.v. with p.m.f.

$$f(x) = \begin{cases} w_i, & \text{if } x = a_i, \quad i = 1, \cdots, n \\ 0, & \text{otherwise} \end{cases}$$

Then $\psi(x) = -\ln x$, $x > 0$, is a convex function. Therefore

$$E(\psi(X)) \geq \psi(E(X))$$

$$\Rightarrow E(-\ln X) \geq -\ln E(X)$$

$$\Rightarrow -\sum_{i=1}^{n} (\ln a_i) w_i \geq -\ln\left(\sum_{i=1}^{n} a_i w_i\right)$$

$$\Rightarrow \ln\left(\sum_{i=1}^{n} a_i w_i\right) \geq \ln\left(\prod_{i=1}^{n} a_i^{w_i}\right)$$

$$\Rightarrow \sum_{i=1}^{n} a_i w_i \geq \prod_{i=1}^{n} a_i^{w_i}$$

Replacing a_i's by $\frac{1}{a_i}$'s we get

$$\prod_{i=1}^{n} a_i^{w_i} \geq \frac{1}{\sum_{i=1}^{n} \frac{w_i}{a_i}}$$

Therefore

$$\sum_{i=1}^{n} a_i w_i \geq \prod_{i=1}^{n} a_i^{w_i} \geq \frac{1}{\sum_{i=1}^{n} \frac{w_i}{a_i}}.$$