MODULE 3

FUNCTION OF A RANDOM VARIABLE AND ITS DISTRIBUTION

LECTURE 15

Topics

3.4 PROPERTIES OF RANDOM VARIABLES HAVING THE SAME DISTRIBUTION

We begin this section with the following definition.

Definition 4.1

Two random variables X and Y, defined on the same probability space $(\Omega, \mathcal{F}, P)$, are said to have the same distribution (written as $X \overset{d}{=} Y$) if they have the same distribution function, i.e., if $F_X(x) = F_Y(x), \forall x \in \mathbb{R}$.

Theorem 4.1

(i) Let X and Y be random variables of discrete type with p.m.f.s f_X and f_Y respectively. Then $X \overset{d}{=} Y$ if, and only if, $f_X(x) = f_Y(x), \forall x \in \mathbb{R}$;

(ii) Let X and Y be random variables having distribution functions that are differentiable everywhere except, possibly, on some finite sets. Then both of them are of absolutely continuous type. Moreover, $X \overset{d}{=} Y$ if, and only if, there exist versions of p.d.f.s f_X and f_Y of X and Y, respectively, such that $f_X(x) = f_Y(x), \forall x \in \mathbb{R}$.

Proof.

(i) Suppose that $f_X(x) = f_Y(x), \forall x \in \mathbb{R}$. Then, clearly, $F_X(x) = F_Y(x),, \forall x \in \mathbb{R}$, and therefore $X \overset{d}{=} Y$. Conversely suppose that $X \overset{d}{=} Y$, i.e., $F_X(x) = F_Y(x) = G(x)$, say, $\forall x \in \mathbb{R}$. Then

$$\begin{aligned}\{x \in \mathbb{R}: F_X(x) - F_X(x-) > 0\} &= \{x \in \mathbb{R}: F_Y(x) - F_Y(x-) > 0\} \\ &= \{x \in \mathbb{R}: G(x) - G(x-) > 0\}\end{aligned}$$

$\Rightarrow S_X = S_Y = S$, say.

Moreover,

$$f_X(x) = f_Y(x) = \begin{cases} G(x) - G(x-), & \text{if } x \in S \\ 0, & \text{otherwise} \end{cases}.$$

(ii) Suppose that $f_X(x) = f_Y(x), \forall x \in \mathbb{R}$, for some versions of p.d.f.s of f_X and f_Y of X and Y respectively. Then, clearly, $F_X(x) = F_Y(x), \forall x \in \mathbb{R}$ and therefore $X \overset{d}{=} Y$. Conversely suppose that $X \overset{d}{=} Y$, i.e., suppose that $F_X(x) = F_Y(x) = G(x)$, say, $\forall x \in \mathbb{R}$. By the hypothesis, G is differentiable everywhere except possibly on a finite set C. Using Remark 4.2 (vii), Module 2, it follows that both X and Y are of absolutely continuous type with a common (version of) p.d.f.

$$g(x) = \begin{cases} G'(x), & \text{if } x \notin C \\ 0, & \text{otherwise} \end{cases}.$$

As a consequence of the above theorem we have the following corollary.

Theorem 4.2

Let X and Y be two random variables, of either discrete type or of absolutely continuous type, with $X \overset{d}{=} Y$. Then,

(i) for any Borel function $h, E(h(X)) = E(h(Y))$, provided the expectations are finite;

(ii) for any Borel function $\psi, \psi(X) \overset{d}{=} \psi(Y)$.

Proof.

(i) Since $X \overset{d}{=} Y$, we have $F_X(x) = F_Y(x) = G(x)$, say, $\forall x \in \mathbb{R}$

Case I. X is of discrete type.

Since $X \overset{d}{=} Y$, using Theorem 4.1 (i), it follows that $S_X = S_Y = S$, say, and $f_X(x) = f_Y(x) = g(x)$, say, $\forall x \in \mathbb{R}$. Therefore,

$$E\big(h(X)\big) = \sum_{x \in S} h\,(x) f_X(x)$$

$$= \sum_{x \in S} h\,(x) f_Y(x)$$

$$= E\big(h(Y)\big).$$

Case II. X is of absolutely continuous type.

For simplicity assume that G is differentiable everywhere except possibly on a finite set C. Using Remark 4.2 (vii), Module 2, we may take

$$f_X(x) = f_Y(x) = \begin{cases} G'(x), & \text{if } x \notin C \\ 0, & \text{if } x \in C \end{cases}.$$

Therefore,

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

$$= \int_{-\infty}^{\infty} h\,(x) f_Y(x) dx$$

$$= E\big(h(Y)\big).$$

(ii) Fix $x \in \mathbb{R}$. On taking

$$h(x) = I_{(-\infty,a]}\big(\psi(x)\big) = \begin{cases} 1, & \text{if } \psi(x) \le a \\ 0, & \text{if } \psi(x) > a \end{cases}$$

in (i), we get

$$E\left(I_{((-\infty,a])}\big(\psi(X)\big)\right) = E\left(I_{(-\infty,a]}\big(\psi(Y)\big)\right)$$

$$\Rightarrow P(\{\psi(X) \le a\}) = P(\{\psi(Y) \le a\}), \;\; \forall a \in \mathbb{R}$$

$$\Rightarrow \psi(X) \stackrel{d}{=} \psi(Y). \;\blacksquare$$

Example 4.1

(i) Let X be a random variable with p.m.f.

$$f_X(x) = \begin{cases} \binom{n}{x}\left(\frac{1}{2}\right)^n, & \text{if } x \in \{0, 1, \dots, n\} \\ 0, & \text{otherwise} \end{cases},$$

where n is a given positive integer. Let $Y = n - X$. Show that $Y \overset{d}{=} X$ and hence show that $E(X) = \frac{n}{2}$.

(ii) Let X be a random variable with p.d.f.

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

and let $Y = -X$. Show that $Y \overset{d}{=} X$ and hence show that $E(X^{2n+1}) = 0$, $n \in \{0, 1, \cdots\}$.

Solution.

(i) Clearly $E(X)$ is finite. Using Example 2.3 it follows that the p.m.f. of $Y = n - X$ is given by

$$f_Y(y) = P(\{Y = y\})$$

$$= \begin{cases} \binom{n}{y}\left(\frac{1}{2}\right)^y, & \text{if } y \in \{0, 1, \cdots, n\} \\ 0, & \text{otherwise} \end{cases}$$

$$= f_X(y), \qquad \forall y \in \mathbb{R},$$

i.e., $Y \overset{d}{=} X$. Hence, using Theorem 4.2 (i),

$$E(X) = E(Y) = E(n - X) = n - E(X)$$

$$\Longrightarrow E(X) = \frac{n}{2}.$$

(ii) Using Corollary 2.2, it can be shown that $Y = -X$ is a random variable of absolutely continuous type with p.d.f.

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

It follows that $Y \overset{d}{=} X$. For or $n \in \{0,1,2,\ldots\}$, it can be easily shown that $E(|X|^r)$is finite for every $r > -1$. Therefore

$$E(X^{2n+1}) = E(Y^{2n+1}) = -E(X^{2n+1})$$

$$\Longrightarrow E(X^{2n+1}) = 0. \blacksquare$$

Definition 4.2

A random variable X is said to have a *symmetric distribution* about a point $\mu \in \mathbb{R}$ if $X - \mu \stackrel{d}{=} \mu - X$. ■

Theorem 4.3

Let X be a random variable having p.d.f./p.m.f. f_X and distribution function F_X. Let $\mu \in \mathbb{R}$. Then

(i) the distribution of X is symmetric about μ if, and only if, $f_X(\mu - x) = f_X(\mu + x)$, $\forall x \in \mathbb{R}$;

(ii) the distribution of X is symmetric about μ if, and only if, $F_X(\mu + x) + F_X\big((\mu - x)-\big) = 1, \forall x \in \mathbb{R}$ $\big($i.e., if and only if, $P(\{X \leq \mu + x\}) = P(\{X \geq \mu - x\})\big)$;

(iii) the distribution of X is symmetric about μ if, and only if, the distribution of $Y = X - \mu$ is symmetric about 0;

(iv) if the distribution of X is symmetric about μ, then $F_X(\mu -) \leq \frac{1}{2} \leq F_X(\mu)$;

(v) if the distribution of X is symmetric about μ and the expected value of X is finite, then $E(X) = \mu$;

(vi) if the distribution of X is symmetric about 0, then $E(X^{2m+1}) = 0, m \in \{0, 1, 2, \cdots\}$, provided the expectations exist.

Proof. For simplicity we will assume that if X is of absolutely continuous type then its distribution function is differentiable everywhere expectations possibly on a finite set.

(i) Let $Y_1 = X - \mu$ and $Y_2 = \mu - X$. Then the p.d.f.s/p.m.f.s of Y_1 and Y_2 are given by

$$f_{Y_1}(y) = f_X(\mu + y), \quad y \in \mathbb{R}$$

and

$$f_{Y_2}(y) = f_X(\mu - y), \quad y \in \mathbb{R},$$

respectively. Now, under the hypothesis,

$$\text{distribution of } X \text{ is symmetric about } \mu \Leftrightarrow Y_1 \stackrel{d}{=} Y_2$$

$$\Leftrightarrow f_{Y_1}(y) = f_{Y_2}(y), \quad \forall y \in \mathbb{R}$$

$$\Leftrightarrow f_X(\mu + y) = f_X(\mu - y), \forall y \in \mathbb{R}.$$

(ii) Let $Y_1 = X - \mu$ and $Y_2 = \mu - X$ so that the distribution functions of Y_1 and Y_2 are given by $F_{Y_1}(x) = F_X(\mu + x),\ x \in \mathbb{R}$, and $F_{Y_2}(x) = 1 - F_X\big((\mu - x)-\big),\ x \in \mathbb{R}$. Therefore, under the hypothesis,

$$Y_1 \stackrel{d}{=} Y_2 \Leftrightarrow F_{Y_1}(x) = F_{Y_2}(x), \quad \forall x \in \mathbb{R}$$

$$\Leftrightarrow F_X(\mu + x) + F_X\big((\mu - x)-\big) = 1, \forall x \in \mathbb{R}.$$

(iii) Clearly ,

distribution of X is symmetric about $\mu \Leftrightarrow X-\mu \stackrel{d}{=} \mu - X = -(X-\mu)$

$$\Leftrightarrow Y \stackrel{d}{=} -Y.$$

(iv) Using (ii), we have

distribution of X is symmetric about $\mu \Leftrightarrow F_X(\mu+x) + F_X\big((\mu-x)-\big) = 1,\ \forall x \in \mathbb{R}$

$$\Rightarrow F_X(\mu) + F_X(\mu-) = 1$$

$$\Rightarrow F_X(\mu-) \le \frac{1}{2} \le F_X(\mu),$$

since $F_X(\mu-) \le F_X(\mu)$.

(v) Suppose that $X-\mu \stackrel{d}{=} \mu - X$ and $E(|X|) < \infty$. Then $E(X-\mu) = E(\mu - X)$ (using Theorem 4.2 (i)) and therefore $E(X) = \mu$.

(vi) Suppose that $X \stackrel{d}{=} -X$. Then, using Theorem 4.2 (i),

$$E(X^{2m+1}) = E((-X)^{2m+1}), m \in \{0,1,2,\cdots\},$$

provided the expectations exist. Therefore,

$$E(X^{2m+1}) = E(-X^{2m+1}),\ \ m \in \{0,1,2,\cdots\}$$

$$\Rightarrow E(X^{2m+1}) = 0,\ \ m \in \{0,1,2,\cdots\}. \blacksquare$$

Theorem 4.4

Let X and Y be random variables having m.g.f.s M_X and M_Y respectively. Suppose that there exists a positive real number b such that $M_X(t) = M_Y(t),\ \forall t \in (-b,b)$. Then $X \stackrel{d}{=} Y$.

Proof. We will provide the proof for the special case where X and Y are of discrete type and $S_X = S_Y \subseteq \{0,1,2,\cdots\}$, as the proof for general X and Y is involved. We have

$$M_X(t) = M_Y(t),\ \ \forall t \in (-b,b)$$

$$\Rightarrow \sum_{k=0}^{\infty} e^{tk}\, P(\{X=k\}) = \sum_{k=0}^{\infty} e^{tk}\, P(\{Y=k\}), \forall t \in (-b,b)$$

$$\Rightarrow \sum_{k=0}^{\infty} s^{k}\, P(\{X=k\}) = \sum_{k=0}^{\infty} s^{k}\, P(\{Y=k\}), \forall s \in (e^{-b}, e^{b}).$$

We know that if two power series or polynomials match over an interval then they have the same coefficients. It follows that $P(\{X=k\}) = P(\{Y=k\}), k \in \{0,1,2,\cdots\}$, i.e., X and Y have the same p.m.f.. Now the result follows using Theorem 4.1 (i). $\blacksquare$

Example 4.2

Let $\mu \in \mathbb{R}$ and $\sigma > 0$ be real constants and let $X_{\mu,\sigma}$ be a random variable having p.d.f.

$$f_{X_{\mu,\sigma}}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \cdot \qquad (4.1)$$

(i) Show that $f_{X_{\mu,\sigma}}$ is a p.d.f.;

(ii) Show that the probability distribution function of $X_{\mu,\sigma}$ is symmetric about μ. Hence find $E(X_{\mu,\sigma})$;

(iii) Find the m.g.f. of $X_{\mu,\sigma}$ and hence find the mean and variance of $X_{\mu,\sigma}$;

(iv) Let $Y_{\mu,\sigma} = aX_{\mu,\sigma} + b$, where $a \neq 0$ and b are real constants. Using the m.g.f. of $X_{\mu,\sigma}$, show that the p.d.f. of $Y_{\mu,\sigma}$ is

$$f_{Y_{\mu,\sigma}}(y) = \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{(y+(a\mu+b))^2}{2a^2\sigma^2}}, \quad -\infty < y < \infty.$$

Solution.

(i) Clearly $f_{X_{\mu,\sigma}}(x) \geq 0,\ \forall x \in \mathbb{R}$. Also,

$$\int_{-\infty}^{\infty} f_{X_{\mu,\sigma}}(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \qquad \text{(on making the transformation } \frac{x-\mu}{\sigma} = z)$$

$$= I, \text{say.}$$

Clearly $I \geq 0$ and

$$I^2 = \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy\right)\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz\right)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(y^2+z^2)}{2}} dydz.$$

On making the transformation $y = r\cos\theta, z = r\sin\theta, r > 0, 0 \leq \theta < 2\pi$ (so that the Jacobian of the transformation is r) we get

$$I^2 = \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} re^{-\frac{r^2}{2}} d\theta dr$$

$$= \int_0^{\infty} re^{-\frac{r^2}{2}} dr$$

$$= \int_0^{\infty} e^{-z} dz$$

$$= 1.$$

Since $I \geq 0$, it follows that $I = 1$ and thus $f_{X_{\mu,\sigma}}(x)$ is a p.d.f..

(ii) Clearly,

$$f_{X_{\mu,\sigma}}(\mu - x) = f_{X_{\mu,\sigma}}(\mu + x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2\sigma^2}}, \ \ \forall x \in \mathbb{R}.$$

Using Theorem 4.3 (i) and (v) it follows that the distribution of $X_{\mu,\sigma}$ is symmetric about μ and $E\left(X_{\mu,\sigma}\right) = \mu$.

(iii) For $t \in \mathbb{R}$

$$\begin{aligned} M_{X_{\mu,\sigma}}(t) &= E(e^{tX_{\mu,\sigma}}) \\ &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\mu+\sigma z)} e^{-\frac{z^2}{2}}\, dz \\ &= \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z-\sigma t)^2}{2}} dt \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}}, \end{aligned}$$

since, by (i),

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \sigma\sqrt{2\pi}, \ \ \forall \mu \in \mathbb{R} \text{ and } \sigma > 0.$$

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

$$\psi_{X_{\mu,\sigma}}(t) = \ln\left(M_{X_{\mu,\sigma}}(t)\right) = \mu t + \frac{\sigma^2 t^2}{2}$$

$$\Longrightarrow E(X) = \psi_{X_{\mu,\sigma}}^{(1)}(0) = \mu \text{ and } \mathrm{Var}(X) = \psi_{X_{\mu,\sigma}}^{(2)}(0) = \sigma^2.$$

(iv) From the discussion following Definition 3.3 we have, for $t \in \mathbb{R}$,

$$\begin{aligned} M_{Y_{\mu,\sigma}}(t) &= M_{aX_{\mu,\sigma}+b}(t) \\ &= e^{tb} M_{X_{\mu,\sigma}}(at) \\ &= e^{(a\mu+b)t + \frac{a^2\sigma^2 t^2}{2}} \end{aligned}$$

$$= M_{X_{a\mu+b,|a|\sigma}}(t)$$

$$\Longrightarrow Y_{\mu,\sigma} \stackrel{d}{=} X_{a\mu+b,|a|\sigma}.$$

Therefore the p.d.f. of $Y_{\mu,\sigma}$ is given by

$$f_{Y_{\mu,\sigma}}(y) = f_{X_{a\mu+b,|a|\sigma}}(y)$$

$$= \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}} y \in \mathbb{R}. \blacksquare$$

In the statistical literature, the probability distribution of the random variable $X_{\mu,\sigma}$ having a p.d.f. $f_{X_{\mu,\sigma}}(\cdot)$, defined by (4.1), is called the normal distribution (or Gaussian distribution) with mean μ and variance σ^2 (denoted by $X_{\mu,\sigma} \sim N(\mu, \sigma^2)$). Various properties of this distribution are further discussed in Module 5.

Example 4.3

Let $p \in (0,1)$ and let X_p be a random variable with p.m.f.

$$f_{X_p}(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & \text{if } x \in \{0, 1, \cdots, n\} \\ 0, & \text{otherwise} \end{cases}, \tag{4.2}$$

where n is a given positive integer and $q = 1 - p$.

(i) Find the m.g.f. of X_p and hence find the mean and variance of $X_p, p \in (0,1)$;
(ii) Let $Y_p = n - X_p$, p $\in (0,1)$. Using the m.g.f. of X_p show that the p.m.f. of Y_p is

$$f_{Y_p}(x) = \begin{cases} \binom{n}{y} q^y (1-q)^{n-y}, & \text{if } y \in \{0, 1, \cdots, n\} \\ 0, & \text{otherwise} \end{cases}.$$

Solution.

(i) From the solution of Example 3.6 (iii), it is clear that the m.g.f. of X_p is given by

$$M_{X_p}(t) = (1 - p + pe^t)^n, \quad t \in \mathbb{R}.$$

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

$$\psi_{X_p}(t) = \ln(M_{X_p}(t)) = n \ln(1 - p + pe^t), t \in \mathbb{R},$$

$$\psi_{X_p}^{(1)}(t) = \frac{npe^t}{1 - p + pe^t}, t \in \mathbb{R},$$

$$\psi_{X_p}^{(2)}(t) = np\frac{(1-p+pe^t)e^t - pe^{2t}}{(1-p+pe^t)^2} \quad ,t \in \mathbb{R}$$

$$\Rightarrow E(X) = \psi_{X_p}^{(1)}(0) = np \text{ and } \quad \mathrm{Var}(X) = \psi_{X_p}^{(2)}(0) = np(1-p).$$

(ii) For $t \in \mathbb{R}$

$$\begin{aligned} M_{Y_p}(t) &= E(e^{tY_p}) \\ &= e^{nt}M_{X_p}(-t) \\ &= e^{nt}(1-p+pe^{-t})^n \\ &= (p+(1-p)e^t)^n \\ &= M_{X_{1-p}}(t), \end{aligned}$$

i.e. , $Y_p \stackrel{d}{=} X_{1-p}$, p $\in (0,1)$. Therefore the p.m.f. of Y_p is given by

$$f_{Y_p}(y) = f_{X_{1-p}}(y)$$

$$= \begin{cases} \binom{n}{y} q^y(1-q)^{n-y}, & \text{if } y \in \{0,1,\cdots,n\} \\ 0, & \text{otherwise} \end{cases}. \blacksquare$$

The probability distribution of the random variable X_p, having p.m.f. $f_{X_p}(\cdot)$ defined by (4.2), is called a binomial distribution with n trials and success probability p (denoted by $X_p \sim \mathrm{Bin}(n,p)$). The binomial distribution and other related distributions are discussed in more detail in Module 5.

3.5 PROBABILITY AND MOMENT INEQUALITIES

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$, and let $B \in \mathcal{B}_1$ be a Borel set. In many situation $P(\{X \in B\})$ cannot be explicitly evaluated and therefore some estimate of this probability may be desired. For example if a random variable Z has the p.d.f.

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}, -\infty < z < \infty,$$

then

$$P(\{Z > 2\}) = \int_2^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \tag{5.1}$$

cannot be explicitly evaluated and, therefore, an estimate of this probability may be desired. To estimate this probability one has to either resort to numerical integration or use some other estimation procedure. Inequalities are popular estimation procedures and they play an important role in probability theory.

Theorem 5.1

Let X be a random variable and let $g: [0, \infty) \to \mathbb{R}$ be a non-negative and non decreasing function such that the expected value of $g(X)$ is finite. Then, for any $c > 0$ for which $g(c) > 0$,

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

Proof. We will provide the proof for the case when X is of absolutely continuous type. The proof for the discrete case follows in the similar fashion with integral signs replaced by summation signs.

Fix $c > 0$ such that $g(c) > 0$. Define $A = \mathbb{R} - (-c, c)$ so that, for $x \in A$, $|x| \geq c$. Then

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

$$\geq \int_{-\infty}^{\infty} g(|x|)\, I_A(x) f_X(x) dx \quad (\text{since } g(|x|) \geq g(|x|) I_A(x) \forall x \in \mathbb{R})$$

$$\geq g(c) \int_{-\infty}^{\infty} I_A(x) f_X(x) dx \quad (\text{since } g(|x|) I_A(x) \geq g(c) I_A(x) \forall x \in \mathbb{R}, \text{as } g \uparrow)$$

$$= g(c) P(\{X \in A\})$$

$$= g(c) P(\{|X| \geq c\})$$

$$\Rightarrow P(\{|X| \geq c\}) \leq \frac{\mathrm{E}\big(g(|X|)\big)}{g(c)}. \blacksquare$$

Corollary 5.1

Let X be a random variable.

3.5.1 Markov Inequality

Suppose that $E(|X|^r) < \infty$, for some $r > 0$. Then, for any $c > 0$,

$$P(\{|X| \geq c\}) \leq \frac{E(|X|^r)}{c^r}.$$

3.5.2 Chebyshev Inequality

Suppose that X has finite first two moments. If $\mu = E(X)$ and $\sigma^2 = \mathrm{Var}(X)$ $(\sigma \geq 0)$. Then for any $k > 0$,

$$P(\{|X-\mu|\geq k\})\leq \frac{\sigma^2}{k^2}.$$

Proof.

(i) Fix $c>0$ and $r>0$ and let $g(x)=x^r, x\geq 0$. Clearly g is a non-negative and non decreasing function. Using Theorem 5.1, we get

$$P(\{|X|\geq c\})\leq \frac{E(|X|^r)}{c^r}.$$

(ii) Using (i) on $Y=|X-\mu|$, for $r=2$, we get

$$P(\{|X-\mu|\geq k\})\leq \frac{E(|X-\mu|^2)}{k^2}=\frac{\sigma^2}{k^2}. \blacksquare$$

Example: 5.1

Let us revisit the problem of estimating $P(\{Z>2\})$, defined by (5.1). Using Example 4.2 (iii), we have $\mu=E(Z)=0$ and $\sigma^2=\mathrm{Var}(Z)=1$. Moreover, using Example 4.2 (ii), $Z\overset{d}{=}-Z$. Consequently $P(\{Z>2\})=P(\{-Z>2\})\,\big(=P(\{Z<-2\})\big)$, i.e.,

$$P(\{Z>2\})=\frac{P(\{|Z|>2\})}{2}=\frac{P(\{|Z|\geq 2\})}{2}.$$

Now, using the Chebyshev inequality we have

$$P(\{Z>2\})=\frac{P(\{|Z|\geq 2\})}{2}\leq\frac{E(|Z|^2)}{8}=\frac{1}{8}=0.125.$$

The exact value of $P(\{Z>2\})$, obtained using numerical integration, is 0.0228. $\blacksquare$

The following example illustrates that bounds provided in Theorem 5.1 and Corollary 5.1 are tight, i.e., the upper bounds provided there in may be attained.

Example 5.2

Let X be a random variable with p.m.f.

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

Clearly $E(X^2) = \frac{1}{4}$ and, therefore, using the Markov inequality we have

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

The exact probability is

$$P(\{|X| \geq 1\}) = P(\{X \in \{-1,1\}) = \frac{1}{4}. \blacksquare$$

Definition 5.1

A random variable X is said to be degenerate at a point $c \in \mathbb{R}$ if $P(\{X = c\}) = 1.$ $\blacksquare$

Suppose that a random variable X is degenerate at $c \in \mathbb{R}$. Then clearly X is of discrete type with support $S_X = \{c\}$, distribution function.

$$F_X(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases},$$

and p.m.f.

$$f_X(x) = \begin{cases} 1, & x = c \\ 0, & \text{otherwise} \end{cases}.$$

Note that a random variable X is degenerate at a point $c \in \mathbb{R}$ if, and only if, $E(X) = c$ and $\mathrm{Var}(X) = 0$.