MODULE 3

FUNCTION OF A RANDOM VARIABLE AND ITS DISTRIBUTION

LECTURE 16

Topics

3.5.3 Jensen Inequality

Theorem 5.2

Let $I \subseteq \mathbb{R}$ be an interval and let $\phi: I \to \mathbb{R}$ be a twice differentiable function such that its second order derivative $\phi''(\cdot)$ is continuous on I and $\phi''(x) \geq 0, \forall x \in \mathbb{R}$. Let X be a random variable with support $S_X \subseteq I$ and finite expectation. Then

$$E\big(\phi(X)\big) \geq \phi\big(E(X)\big).$$

If $\phi''(x) > 0, \quad \forall x \in I$, then the inequality above is strict unless X is a degenerate random variable.

Proof. Let $\mu = E(X)$. On expanding $\phi(x)$ into a Taylor series about μ we get

$$\phi(x) = \phi(\mu) + (x-\mu)\phi'(\mu) + \frac{(x-\mu)^2}{2!}\phi''(\xi), \forall x \in I,$$

for some ξ between μ and x. Since $\phi''(x) \geq 0, \forall x \in I$, it follows that

$$\phi(x) \geq \phi(\mu) + (x-\mu)\phi'(\mu), \forall x \in I \qquad (5.2)$$

$$\Rightarrow \phi(X) \geq \phi(\mu) + (X-\mu)\phi'(\mu), \quad \forall X \in S_X$$

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

$$= \phi(\mu)$$

$$= \phi\big(E(X)\big).$$

Clearly the inequality in (5.2) is strict unless $E((X-\mu)^2) = 0$, i.e., $P(\{X = \mu\}) = 1$ (using Corollary 3.1 (ii)). ∎

Example 5.3

Let X be a random variable with support S_X, an interval in $\mathbb{R}$. Then

(i) $E(X^2) \geq \big(E(X)\big)^2$ (taking $\phi(x) = x^2, x \in \mathbb{R}$, in Theorem 5.2);

(ii) $E(\ln X) \leq \ln E(X)$, provided $S_X \subseteq (0,\infty)$ (taking $\phi(x) = -\ln x, x \in \mathbb{R}$, in Theorem 5.2);

(iii) $E(e^{-X}) \geq e^{-E(X)}$ (taking $\phi(x) = e^{-x}, x \in \mathbb{R}$, in Theorem 5.2);

(iv) $E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}$, if $S_X \subseteq (0,\infty)$ $\left(\text{taking } \phi(x) = \frac{1}{x}, x \in \mathbb{R} \text{ in Theorem 5.2}\right)$;

provided the involved expectations exist. ∎

Definition 5.2

Let X be a random variable with support $S_X \subseteq (0,\infty)$. Then, provided they are finite, $E(X)$ is called the arithmetic mean (AM) of X, $e^{E(\ln X)}$ is called the geometric mean (GM) of X, and $\frac{1}{E\left(\frac{1}{X}\right)}$ is called harmonic mean (HM) of X. ∎

3.5.4 AM-GM-HM inequality

Example 5.4

(i) Let X be a random variable with support $S_X \subseteq (0,\infty)$. Then

$$E(X) \geq e^{E(\ln X)} \geq \frac{1}{E\left(\frac{1}{X}\right)},$$

provided the expectations are finite.

(ii) Let $a_1, \cdots, a_n$ be positive real constants and let $p_1, \cdots, p_n$ be another set of positive real constants such that $\sum_{i=1}^{n} p_i = 1$. Then

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

Solution.

(i) From Example 5.3 (ii) we have

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

$$\Longrightarrow E(X) \geq e^{E(\ln X)} \tag{5.3}$$

Using (5.3) on $\frac{1}{X}$,we get

$$E\left(\frac{1}{X}\right) \geq e^{E\left(\ln \frac{1}{X}\right)} = e^{-E(\ln X)}$$

$$\Longrightarrow e^{E(\ln X)} \geq \frac{1}{E\left(\frac{1}{X}\right)}. \tag{5.4}$$

The assertion now follows on combining (5.3) and (5.4).

(ii) Let X be a discrete type random variable with support $S_X = \{a_1, a_2 \cdots\}$ and $P(\{X = a_i\}) = p_i, i = 1, \cdots, n.$ Clearly,
$P(\{X = x\}) > 0 \; \forall x \in S_X$ and $\sum_{x \in S_X} P(\{X = x\}) = \sum_{i=1}^{n} p_i = 1.$
On using (i), we get

$$E(X) \geq e^{E(\ln X)} \geq \frac{1}{E\left(\frac{1}{X}\right)}$$

$$\Longrightarrow \sum_{i=1}^{n} a_i p_i \geq e^{\sum_{i=1}^{n} (\ln a_i) p_i} \geq \frac{1}{\sum_{i=1}^{n} \frac{p_i}{a_i}}$$

$$\Longrightarrow \sum_{i=1}^{n} a_i p_i \geq e^{\ln\left(\prod_{i=1}^{n} {a_i}^{p_i}\right)} \geq \frac{1}{\sum_{i=1}^{n} \frac{p_i}{a_i}}$$

$$\Rightarrow \sum_{i=1}^{n} a_i p_i \geq \left(\prod_{i=1}^{n} a_i^{p_i}\right) \geq \frac{1}{\sum_{i=1}^{n} \frac{p_i}{a_i}}.$$

3.6 DESCRIPTIVE MEASURES OF PROBABILITY DISTRIBUTIONS

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$, associated with a random experiment $\mathcal{E}$. Let F_X and f_X denote, respectively, the distribution function and the p.d.f./p.m.f. of X. The probability distribution (i.e., the distribution function/p.d.f./p.m.f.) of X describes the manner in which the random variable X takes values in different Borel sets. It may be desirable to have a set of numerical measures that provide a summary of the prominent features of the probability distribution of X. We call these measures as *descriptive measures.* Four prominently used descriptive measures are: (i) *Measures of central tendency* (or location), also referred to as averages; (ii) *measures of dispersion;* (iii) *measures of skewness*, and (iv) *measures of kurtosis*.

3.6.1 Measures of Central Tendency

A measure of central tendency or location (also called an average) gives us the idea about the central value of the probability distribution around which values of the random variable are clustered. Three commonly used measures of central tendency are *mean, median* and *mode.*

3.6.1.1 Mean.

Recall (Definition 3.2 (i)) that the mean (of probability distribution) of a random variable X is given by $\mu_1' = E(X)$. We have seen that the mean of a probability distribution gives us idea about the average observed value of X in the long run (i.e., the average of observed values of $X$when the random experiment is repeated a large number of times). Mean seems to be the best suited average if the distribution is symmetric about a point μ (i.e., $X-\mu \overset{d}{=} \mu - X$, in which case $\mu = E(X)$ provided it is finite), values in the neighborhood of μ occur with high probabilities, and as we move away from μ in either direction $f_X(\cdot)$ decreases. Because of its simplicity mean is the most commonly used average (especially for symmetric or nearly symmetric distributions). Some of the demerits of this measure are that in some situations this may not be defined (Examples 3.2 and 3.4) and that it is very sensitive to presence of a few extreme values of X which are different from other values of X (even though they may occur with small positive probabilities). So this measure should be used with caution if probability distribution assigns positive probabilities to a few Borel sets having some extreme values.

3.6.1.2 Median.

A real number m satisfying in $F_X(m-) \leq \frac{1}{2} \leq F_X(m)$, i.e., $P(\{X < m\}) \leq \frac{1}{2} \leq P(\{X \leq m\})$, is called the *median* (of the probability distribution) of X. Clearly if m is the median of a probability distribution then, in the long run (i.e., when the random experiment $\mathcal{E}$ is repeated a large number of times), the values of X on either side of m in S_X are observed with the same frequency. Thus the median of a probability distribution, in some sense, divides S_X into two equal parts each having the same probability of occurrence. It is evident that if $X$is of continuous type then the median m is given by $F_X(m) = 1/2$. For some distributions (especially for distributions of discrete type random variable) it may happen that $F_X(a-) < \frac{1}{2}$ and $\{x \in \mathbb{R}: F_X(x) = 1/2\} = [a, b)$, for some $-\infty < a < b < \infty$, so that the median is not unique. In that case $P(\{X = x\}) = 0, \forall \mathrm{x} \in (\mathrm{a}, \mathrm{b})$ and thus we take the median to be $m = a = \inf\{x \in \mathbb{R}: F_X(x) \geq 1/2\}$. For random variables having a symmetric probability distribution it is easy to verify that the mean and the median coincide (see Problem 33). Unlike the mean, the median of a probability distribution is always defined. Moreover the median is not affected by a few extreme values as it takes into account only the probabilities with which different values occur and not their numerical values. As a measure of central tendency the median is preferred over the mean if the distribution is asymmetric and a few extreme observations are assigned positive probabilities. However the fact that the median does not at all take into account the numerical values of X is one of its demerits. Another disadvantage with median is that for many probability distributions it is not easy to evaluate (especially for distributions whose distribution functions $F_X(\cdot)$ do not have a closed form).

3.6.1.3 Mode.

Roughly speaking the *mode* m_0 of a probability distribution is the value that occurs with highest probability and is defined by $f_X(m_0) = \sup\{f_X(x): x \in S_X\}$. Clearly if m_0 is the mode of a probability distribution of X then, in the long run, either m_0 or a value in the neighborhood of m_0 is observed with maximum frequency. Mode is easy to understand and easy to calculate. Normally, it can be found by just inspection. Note that a probability distribution may have more than one mode which may be far apart. Moreover mode does not take into account the numerical values of X and it also does not take into account the probabilities associated with all the values of X. These are crucial deficiencies of mode which make it less preferable over mean and median. A probability distribution with one (two/three) mode(s) is called an *unimodal (bimodal/trimodal)* distribution. A distribution with multiple modes is called a *multimodal* distribution.

3.6.2 Measures of Dispersion

Measures of central tendency give us the idea about the location of only central part of the distribution. Other measures are often needed to describe a probability distribution. The values assumed by a random variable X usually differ from each other. The usefulness of mean or median as an average is very much dependent on the variability (or dispersion) of values of X around mean or median. A probability distribution (or the corresponding random variable X) is said to have a high dispersion if its support contains many values that are significantly higher or lower than the mean or median value. Some of the commonly used measures of dispersion are standard deviation, quartile deviation (or semi-inter-quartile range) and coefficient of variation.

3.6.2.1 Standard Deviation.

Recall (Definition 3.2 (v)) that the variance (of the probability distribution) of a random variable X is defined by $\sigma^2 = E((X-\mu)^2)$, where $\mu = E(X)$ is the mean (of the probability distribution) of X. The *standard deviation* (of the probability distribution) of X is defined by $\sigma = \sqrt{\mu_2} = \sqrt{E((X-\mu)^2)}$. Clearly the variance and the standard deviation give us the idea about the average spread of values of Xaround the mean μ. However, unlike the variance, the unit of measurement of standard deviation is the same as that of X. Because of its simplicity and intuitive appeal, standard deviation is the most widely used measure of dispersion. Some of the demerits of standard deviation are that in many situations it may not be defined (distribution for which second moment is not finite) and that it is sensitive to presence of a few extreme values of X which are different from other values. A justification for having the mean μ in place of median or any other average in the definition of $\sigma = \sqrt{E((X-\mu)^2)}$ is that $\sqrt{E((X-\mu)^2)} \leq \sqrt{E((X-c)^2)}, \forall c \in \mathbb{R}$ (Problem 32).

3.6.2.2 Mean Deviation.

Let A be a suitable average. The *mean deviation* (of probability distribution) of X around average A is defined by $\mathrm{MD}(A) = E(|X-A|)$. Among various mean deviations, the mean deviation about the median m is more preferable than the others. A reason for this preference is the fact that for any random variable $X, \mathrm{MD}(m) = E(|X-m|) \leq E(|X-c|) = \mathrm{MD}(c), \forall c \in \mathbb{R}$ (Problem 24). Since a natural distance between X and m is $|X-m|$, as a measure of dispersion, the mean deviation about median seems to be more appealing than the standard deviation. Although the mean deviation about median (or mean) has more intuitive appeal than the standard deviation, in most situations, it is not easy to evaluate. Some of the other demerits of mean deviation are that in many situations they may not be defined and that they are sensitive to presence of a few extreme values of X which are different from other values.

3.6.2.3 Quartile Deviation.

A common drawback with the standard deviation and mean deviations, as measures of dispersion, is that they are sensitive to presence of a few extreme values of X. *Quartile deviation* measures the spread in the middle half of the distribution and is therefore not influenced by extreme values. Let q_1 and q_3 be real numbers such that

$$F_X(q_1 -) \leq \frac{1}{4} \leq F_X(q_1) \text{ and } F_X(q_3 -) \leq \frac{3}{4} \leq F_X(q_3),$$

$\text{i.e.}, P(\{X < q_1\}) \leq \frac{1}{4} \leq P(\{X \leq q_1\})$ and $P(\{X < q_3\}) \leq \frac{3}{4} \leq P(\{X \leq q_3\})$.

The quantities q_1 and q_3 are called, respectively, the *lower and upper quartiles* of the probability distribution of random variable X. Clearly if q_1, m and q_3 are respectively the lower quartile, the median and the upper quartile of a probability distribution then they divide the probability distribution in four parts so that, in the long run (i.e., when the random experiment $\mathcal{E}$ is repeated a large number of times) twenty five percent of the observed values of X are expected to be less than q_1, fifty percent of the observed values of X are expected to be less than m and seventy five percent of the observed values of X are expected to be less than q_3. The quantity $\text{IQR} = q_3 - q_1$ is called the *inter quartile range* of the probability distribution of X and the quantity $\text{QD} = \frac{q_3 - q_1}{2}$ is called the *quartile deviation* or the *semi-inter-quartile range* of the probability distribution of X. It can be seen that if X is of absolutely continuous type then q_1 and q_3 are given by $F_X(q_1) = \frac{1}{4}$ and $F_X(q_3) = \frac{3}{4}$. For some distributions (especially for distributions of discrete type random variables) it may happen that $F_X(q_1 -) < \frac{1}{4}$ and/or $\{x \in \mathbb{R}: F_X(x) = 1/4\} = [a, b)$ $\left(F_X(q_3 -) < \frac{3}{4} \text{ and/or } \{x \in \mathbb{R}: F_X(x) = 3/4\} = [c, d)\right)$ for some $-\infty < a < b < \infty$ (for some $-\infty < c < d < \infty$), so that q_1 (q_3) is not uniquely defined. In that case $P(\{X = x\}) = 0, \forall x \in (a, b)$ ($x \in (c, d)$) and thus we take $q_1 = a = \inf\{x \in \mathbb{R}: F_X(x) \geq 1/4\}$ ($q_3 = c = \inf\{x \in \mathbb{R}: F_X(x) \geq 3/4\}$) . For random variables having a symmetric probability distribution it is easy to verify that $m = (q_1 + q_3)/2$ (Problem 33). Although, unlike the standard deviation and the mean deviation, quartile deviation is not sensitive to presence of some extreme values of X a major drawback with the quartile deviation is that it ignores the tails of the probability distribution (which constitute 50% of the probability distribution). Note that the quartile deviation depends on the units of measurement of random variable X and thus it may not be an appropriate measure for comparing dispersions of two different probability distributions. For comparing dispersions of two different probability distributions a normalized measure such as

$$\mathrm{CQD} = \frac{\frac{q_3 - q_1}{2}}{\frac{q_3 + q_1}{2}} = \frac{q_3 - q_1}{q_3 + q_1}$$

seems to be more appropriate. The quantity CQD is called the *coefficient of quartile deviation* of the probability distribution of X. Clearly the coefficient of quartile deviation is independent of unit and thus it can be used to be compare dispersion of two different probability distributions.

3.6.2.4 Coefficient of Variation.

Like quartile deviation, standard deviation σ also depends on the units of measurement of random variable X and thus it is not an appropriate measure for comparing dispersions of two different probability distributions. Let μ and σ, respectively, be the mean and the standard deviation of the distribution of X. Suppose that $\mu \neq 0$. The *coefficient of variation* of the probability distribution of X is defined by

$$CV = \frac{\sigma}{\mu}.$$

Clearly the coefficient of variation measures the variation per unit of mean and is independent of units. Therefore it seems to be an appropriate measure to compare dispersions of two different probability distributions. A disadvantage with the coefficient of variation is that when the mean μ is close to zero it is very sensitive to small changes in the mean.

3.6.3 Measures of Skewness

Skewness of a probability distribution is a measure of asymmetry (or lack of symmetry). Recall that the probability distribution of random variable X is said to be symmetric about point μ if $X - \mu \stackrel{d}{=} \mu - X$. In that case $\mu = E(X)$ (provided it exists) and $f_X(\mu + x) = f_X(\mu - x), \forall x \in \mathbb{R}$. Evidently, for symmetric distributions, the shape of the p.d.f./p.m.f. on the left of μ is the mirror image of that on the right side of μ. It can be shown that, for symmetric distribution the mean and the median coincide (Problem 33). We say that a probability distribution is *positively skewed* if the tail on the right side of the p.d.f./p.m.f. is longer than that on the left side of the p.d.f./p.m.f. and bulk of the values lie on the left side of the mean. Clearly a positively skewed distribution indicates presence of a few high values of X which pull up the value of the mean resulting in mean larger than the median and the mode. For unimodal positively skewed distribution we normally have $\mathrm{Mode} < \mathrm{Median} < \mathrm{Mean}$. Similarly we say that a probability distribution is *negatively skewed* if the tail on the left side of the p.d.f./p.m.f. is longer than that on the right side of the p.d.f./p.m.f. and bulk of the values lie on the right side of the mean. Clearly a negatively skewed distribution indicates presence of a few low values of X which pull down the value of the mean resulting in mean smaller than the median and the mode. For unimodal negatively skewed distribution we normally have Mean< Median <Mode. Let μ and σ, respectively, be the mean and the standard deviation of X and let $Z = (X - \mu)/\sigma$ be the standardized variable (independent of units). A measure of skewness of the probability distribution of X is defined by

$$\beta_1 = E(Z^3) = \frac{E((X-\mu)^3)}{\sigma^3} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}.$$

The quantity β_1 is simply called the *coefficient of skewness*. Clearly for symmetric distributions $\beta_1 = 0$ (Theorem 4.3 (vi)). However the converse may not be true, i.e., there are examples of skewed probability distributions for which $\beta_1 = 0$. A large positive value of β_1 indicates that the data is positively skewed and a small negative value of β_1 indicates that the data is negatively skewed. A measure of skewness can also be based on quartiles. Let q_1, m, q_3 and μ denote respectively the lower quartile, the median, the upper quartile and the mean of the probability distribution of X. We know that for random variables having a symmetric probability distribution $\mu = m = \frac{(q_1 + q_3)}{2}$, i.e., $q_3 - m = m - q_1$.

For positively (negatively) skewed distribution we will have $(q_3 - m) > (<)(m - q_1)$. Thus one may also define a measure of skewness based on $(q_3 - m) - (m - q_1) = q_3 - 2m + q_1$. To make this quantity independent of units one may consider

$$\beta_2 = \frac{q_3 - 2m + q_1}{q_3 - q_1}$$

as a measure of skewness. The quantity β_2 is called the *Yule coefficient of skewness.*

3.6.4 Measures of Kurtosis

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

$$f_{Y_{\mu,\sigma}}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0,$$

i.e., $Y_{\mu,\sigma} \sim N(\mu, \sigma^2)$ (see Example 4.2 and discussion following it). We have seen (Example 4.2 (iii)) that μ and σ^2 are respectively the mean and the variance of the distribution of $Y_{\mu,\sigma}$. We call the probability distribution corresponding to p.d.f. $f_{Y_{\mu,\sigma}}$ as the *normal distribution* with mean μ and variance σ^2 (denoted by $N(\mu, \sigma^2)$). We know that $N(\mu, \sigma^2)$ distribution is symmetric about μ (cf. Example 4.2 (ii)). Also it is easy to verify that $N(\mu, \sigma^2)$ distribution is unimodal with μ as the common value of mean, median and mode. Kurtosis of the probability distribution of X is a measure of peakedness and thickness of tail of p.d.f. of X relative to the peakedness and thickness of tails of the p.d.f. of normal distribution. A distribution is said to have higher (lower) kurtosis than the normal distribution if its p.d.f., in comparison with the p.d.f. of a normal distribution, has a sharper (rounded) peak and longer, fatter (shorter, thinner) tails. Let μ and σ respectively, be the mean and the standard deviation of distribution of X and let $Z = (X - \mu)/\sigma$ be the standardized variable. A measure of Kurtosis of the probability distribution of X is defined by

$$\gamma_1 = E(Z^4) = \frac{E((X-\mu)^4)}{\sigma^4} = \frac{\mu_4}{\mu_2^2}.$$

The quantity γ_1 is simply called the *kurtosis* of the probability distribution of X. it is easy to show that for any values of $\mu \in \mathbb{R}$ and $\sigma > 0$, the kurtosis of $N(\mu, \sigma^2)$ distribution is $\gamma_1 = 3$ (use Example 4.2 (iii)). The quantity

$$\gamma_2 = \gamma_1 - 3$$

is called the excess kurtosis of the distribution of X. It is clear that for normal distributions the excess kurtosis is zero. Distributions with zero excess Kurtosis are called mesokurtic. A distribution with positive (negative) excess Kurtosis is called $leptokurtic$ $(platykurtic)$. Clearly a leptokurtic (platykurtic) distribution has sharper (rounded) peak and longer, fatter (shorter, thinner) tails.

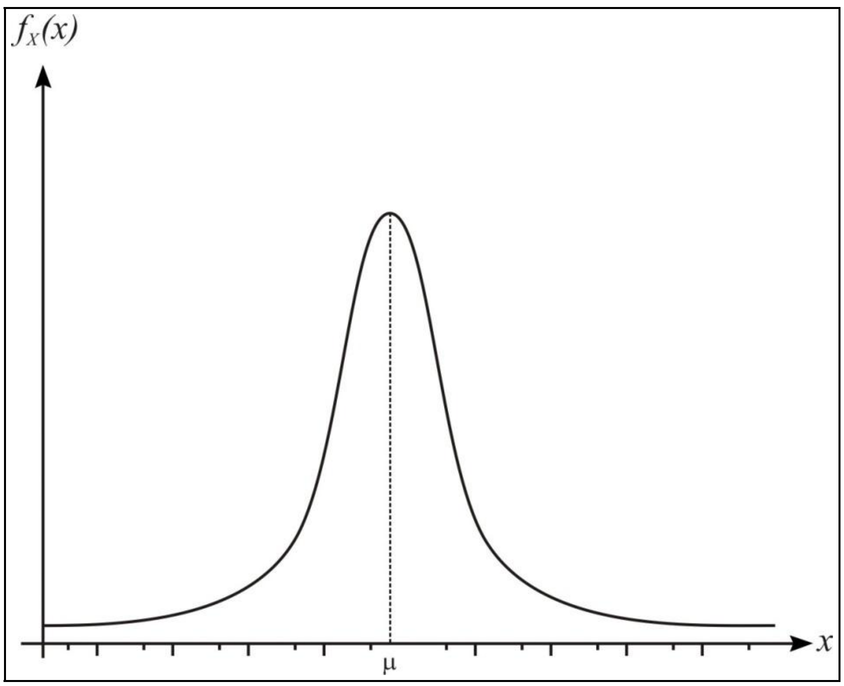

Figure 6.1. Distribution symmetric about mean μ

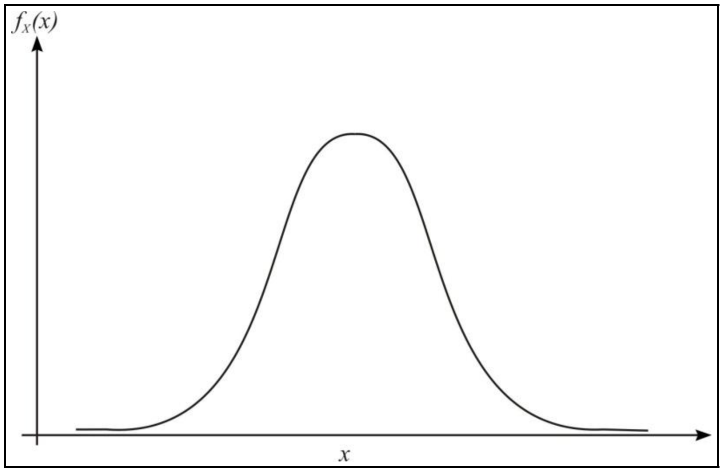

Figure 6.2 (a). Unimodal distribution

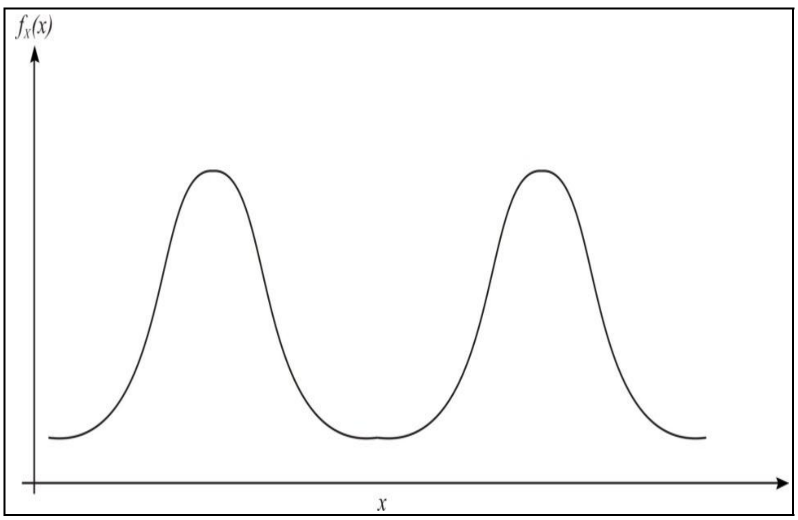

Figure 6.2 (b). Bimodal distribution

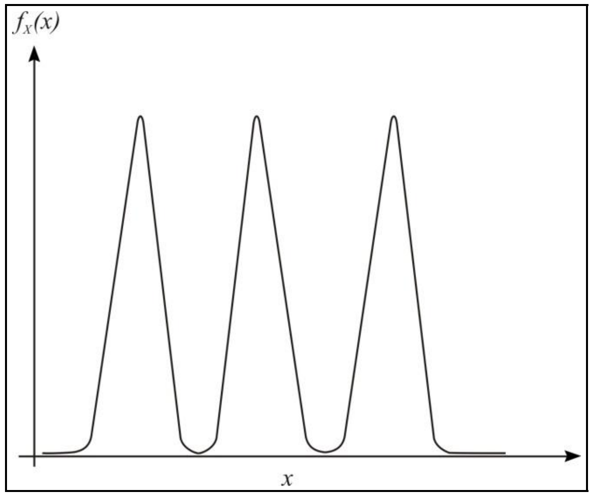

Figure 6.2 (c). Trimodal distribution

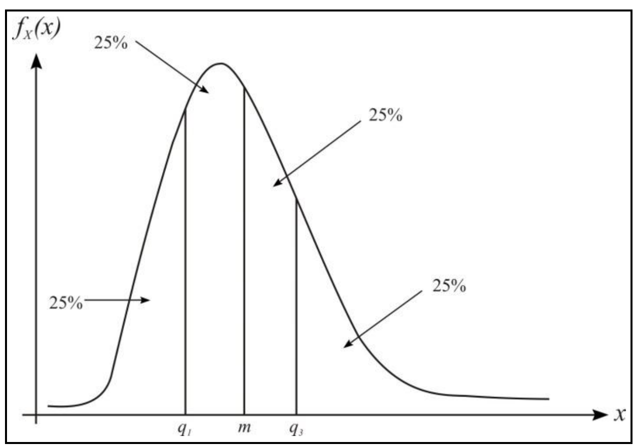

Figure 6.3. Lower quantile, median and upper quantile

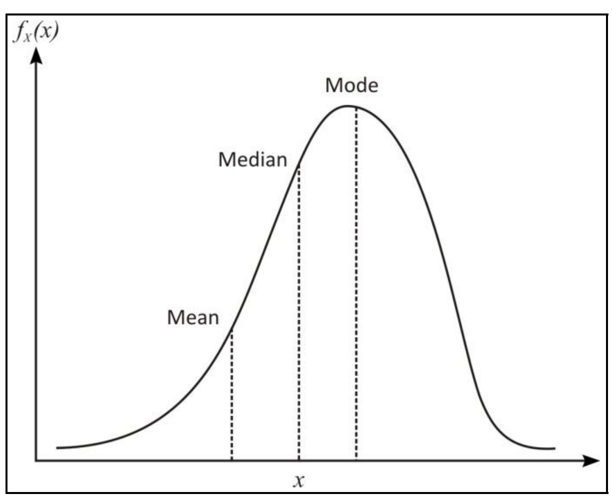

Figure 6.4 (a). Negatively skewed distribution

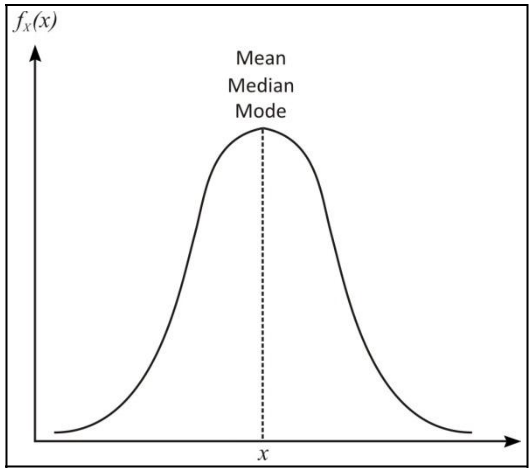

Figure 6.4 (b). Normal (no skew) distribution

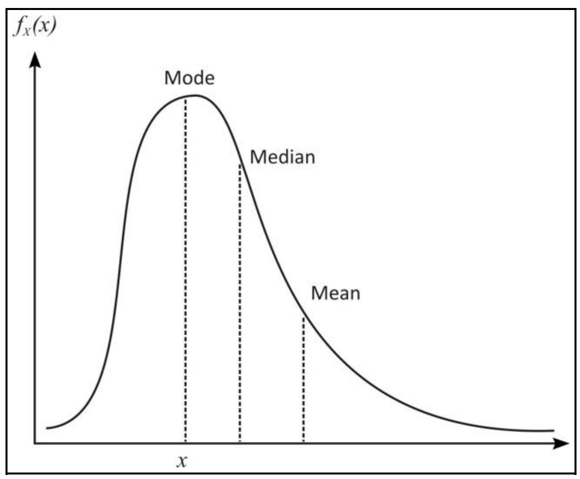

Figure 6.4 (c). Positively skewed distribution

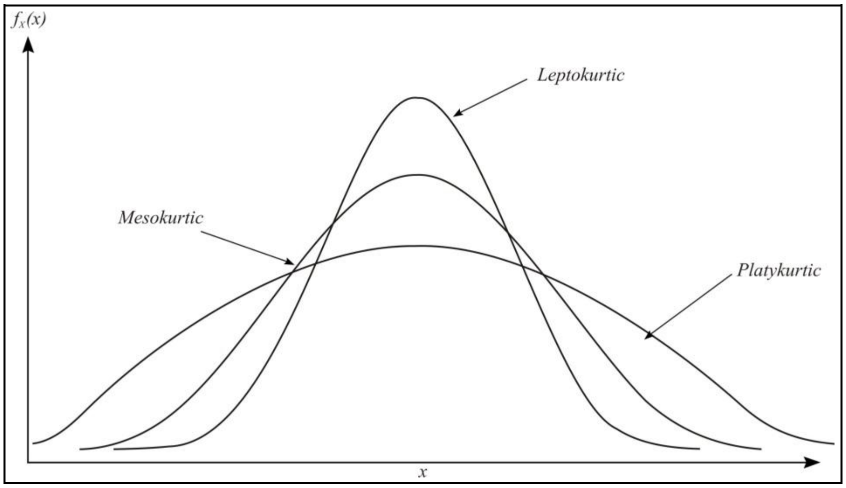

Figure 6.5. Kurtosis (Mesokurtic, Leptokurtic and Platykurtic distributions)

Example 6.1

For $\alpha \in [0,1]$, let X_α be a random variable having the p.d.f.

$$f_\alpha(x) = \begin{cases} \alpha e^{x}, & \text{if } x < 0 \\ (1-\alpha)e^{-x}, & \text{if } x \geq 0 \end{cases}.$$

(i) Show that, for a positive integer r,

$$\int_0^\infty e^{-x} x^{r-1}\, dx = (r-1)!.$$

Hence find $\mu_r'(\alpha) = E(X_\alpha^r), r \in \{1,2,\cdots\}$;

(ii) For $p \in (0,1)$, find $\xi_p \equiv \xi_p(\alpha)$ such that $F_\alpha(\xi_p) = p$, where F_α is the distribution function of X_α. The quantity ξ_p is called the p-th quantile of the distribution of X_α;

(iii) Find the lower quartile $q_1(\alpha)$, the median $m(\alpha)$ and the upper quartile $q_3(\alpha)$ of the distribution function of X_α;

(iv) Find the mode $m_0(\alpha)$ of the distribution of X_α;

(v) Find the standard deviation $\sigma(\alpha)$, the mean deviation about median $\mathrm{MD}(m(\alpha))$, the inter-quartile range) $\mathrm{IQR}(\alpha)$, the quartile deviation (or semi-inter-quartile range) $\mathrm{QD}(\alpha)$, the coefficient of quartile deviation $\mathrm{CQD}(\alpha)$ and the coefficient of variation $\mathrm{CV}(\alpha)$ of the distribution of X_α;

(vi) Find the coefficient of skewness $\beta_1(\alpha)$ and the Yule coefficient of skewness $\beta_2(\alpha)$ of the distribution of X_α. According to values of α, classify the distribution of X_α as symmetric, positively skewed and negatively skewed;

(vii) Find the excess kurtosis $\gamma_2(\alpha)$ of the distribution of X_α and hence comment on the kurtosis of the distribution of X_α.

Solution.

(i) For $r \in \{1,2,\cdots\}$, let

$$\mathrm{I}_r = \int_0^\infty e^{-x} x^{r-1}\, dx,$$

so that $I_1 = 1$. Performing integration by parts it is straightforward to see that $I_r = (r-1)I_{r-1},\ r \in \{2,3,\cdots\}$. On successively using this relationship we get

$$I_r = \int_0^\infty e^{-x} x^{r-1}\, dx = (r-1)!, \quad r \in \{1,2,\cdots\}.$$

Therefore, for a positive integer r,

$$\begin{aligned}\mu_r'(\alpha) &= E(X_\alpha^r)\\ &= \int_{-\infty}^0 \alpha x^r e^x\, dx + \int_0^\infty (1-\alpha) x^r e^{-x}\, dx\\ &= ((-1)^r\alpha + 1 - \alpha)\int_0^\infty x^r e^{-x}\, dx\\ &= ((-1)^r\alpha + 1 - \alpha) r!\\ &= \begin{cases}(1-2\alpha)r! & \text{if} \quad r \in \{1,3,5,\cdots\}\\ r!, & \text{if} \quad r \in \{2,4,6,\cdots\}\end{cases}.\end{aligned}$$

(ii) Let $p \in (0,1)$ and let ξ_p be such that $F_\alpha(\xi_p) = p$. Note that

$$F_\alpha(0) = \alpha \int_{-\infty}^0 e^x\, dx = \alpha.$$

Thus, for evaluation of ξ_p, the following two cases arise.

Case I. $0 \le \alpha < p$

We have $p = F_\alpha(\xi_p)$, i. e.,

$$\begin{aligned}p &= \int_{-\infty}^0 \alpha e^x\, dx + \int_0^{\xi_p} (1-\alpha) e^{-x}\, dx\\ &= 1 - (1-\alpha) e^{-\xi_p},\end{aligned}$$

i.e., $\xi_p = \ln\big((1-\alpha)/(1-p)\big)$.

Case II. $\alpha \geq p$

In this case we have

$$p = \int_{-\infty}^{\xi_p} \alpha e^x dx = \alpha e^{\xi_p},$$

i.e., $\xi_p = -\text{In}\left(\frac{\alpha}{p}\right)$. Combining the two cases we get

$$\xi_p = \begin{cases} \ln\left(\dfrac{1-\alpha}{1-p}\right), & \text{if } 0 \leq \alpha < p \\ -\ln\left(\dfrac{\alpha}{p}\right), & \text{if } p \leq \alpha \leq 1 \end{cases}.$$

(iii) We have

$$q_1(\alpha) = \xi_{\frac{1}{4}} = \begin{cases} \ln\left(\dfrac{4(1-\alpha)}{3}\right), & \text{if } 0 \leq \alpha < \dfrac{1}{4} \\ -\ln(4\alpha), & \text{if } \dfrac{1}{4} \leq \alpha \leq 1 \end{cases},$$

$$m(\alpha) = \xi_{\frac{1}{2}} = \begin{cases} \ln\big(2(1-\alpha)\big), & \text{if } 0 \leq \alpha < \dfrac{1}{2} \\ -\ln(2\alpha), & \text{if } \dfrac{1}{2} \leq \alpha \leq 1 \end{cases},$$

and $q_3(\alpha) = \xi_{\frac{3}{4}} = \begin{cases} \ln\big(4(1-\alpha)\big), & \text{if } 0 \leq \alpha < \frac{3}{4} \\ -\ln\left(\frac{4\alpha}{3}\right), & \text{if } \frac{3}{4} \leq \alpha \leq 1 \end{cases}.$

(iv) The mode $m_0 \equiv m_0(\alpha)$ of the distribution of X_α is such that $f_\alpha(m_0) = \sup\{f_\alpha(x): -\infty < x < \infty\}$. Clearly $m_0 = \max\{\alpha, 1-\alpha\}$.

(v) Using (i) we have $\mu_1'(\alpha) = E(X_\alpha) = 1 - 2\alpha$ and $\mu_2'(\alpha) = E(X_\alpha^2) = 2$. It follows that the standard deviation of the distribution of X_α is

$$\sigma(\alpha) = \sqrt{\text{Var}(X_\alpha)} = \sqrt{\mu_2'(\alpha) - \left(\mu_1'(\alpha)\right)^2} = \sqrt{1 + 4\alpha - 4\alpha^2}.$$

Note that, for $0 \leq \alpha < \frac{1}{2}, m(\alpha) = \ln\big(2(1-\alpha)\big) \geq 0$ and, for $\alpha > \frac{1}{2}, m(\alpha) = -\ln(2\alpha) < 0$. Thus, for the evaluation of the mean deviation about the median, the following cases arise:

Case I. $0 \leq \alpha < \frac{1}{2}$ (so that $m(\alpha) \geq 0$)

$$\begin{aligned}
\mathrm{MD}\big(m(\alpha)\big) &= E(|X-m(\alpha)|) \\
&= \alpha\int_{-\infty}^{0}(m(\alpha)-x)e^{x}dx + (1-\alpha)\int_{0}^{m(\alpha)}(m(\alpha)-x)\,e^{-x}dx \\
&\quad + (1-\alpha)\int_{m(\alpha)}^{\infty}\big(x-m(\alpha)\big)e^{-x}dx \\
&= m(\alpha)+2\alpha \\
&= \ln\big(2(1-\alpha)\big)+2\alpha.
\end{aligned}$$

Case II. $\frac{1}{2}\leq\alpha\leq 1$ (so that $m(\alpha)\leq 0$)

$$\begin{aligned}
\mathrm{MD}\big(m(\alpha)\big) &= E(|X-m(\alpha)|) \\
&= \alpha\int_{-\infty}^{m(\alpha)}(m(\alpha)-x)e^{x}dx + \alpha\int_{m(\alpha)}^{0}(x-m(\alpha))e^{x}dx \\
&\quad + (1-\alpha)\int_{0}^{\infty}\big(x-m(\alpha)\big)e^{-x}dx \\
&= 2(1-\alpha)-m(\alpha) \\
&= \ln(2\alpha)+2(1-\alpha).
\end{aligned}$$

Combining the two cases we get

$$\mathrm{MD}\big(m(\alpha)\big)=\begin{cases}\ln\big(2(1-\alpha)\big)+2\alpha, & \text{if}\quad 0\leq\alpha<\frac{1}{2}\\ \ln(2\alpha)+2(1-\alpha), & \text{if}\quad \frac{1}{2}\leq\alpha\leq 1\end{cases}.$$

Using (iii) the inter- quartile range of the distribution of X_α is

$$\begin{aligned}
\mathrm{IQR}\ (\alpha) &= q_3(\alpha)-q_1\,(\alpha) \\
&= \begin{cases}\ln 3, & \text{if } 0\leq\alpha<\dfrac{1}{4}\\ \ln\big(16\alpha(1-\alpha)\big), & \text{if } \dfrac{1}{4}\leq\alpha<\dfrac{3}{4}\\ \ln 3, & \text{if } \dfrac{3}{4}\leq\alpha\leq 1\end{cases}.
\end{aligned}$$

The quartile deviation of the distribution of X_α is

$$\mathrm{QD}(\alpha)=\frac{q_3(\alpha)-q_1(\alpha)}{2} = \begin{cases} \ln\sqrt{3}, & \text{if} \quad 0\le\alpha<\frac{1}{4} \\ \ln\left(4\sqrt{\alpha(1-\alpha)}\right), & \text{if} \quad \frac{1}{4}\le\alpha<\frac{3}{4} \\ \ln\sqrt{3}, & \text{if} \quad \frac{3}{4}\le\alpha\le 1 \end{cases}.$$

The coefficient quartile deviation of the distribution of X_α is

$$\mathrm{CQD}(\alpha)=\frac{q_3(\alpha)-q_1(\alpha)}{q_3(\alpha)+q_1(\alpha)}=\begin{cases} \dfrac{\ln 3}{\ln\left(\frac{16\,(1-\alpha)^2}{3}\right)}, & \text{if} \quad 0\le\alpha<\frac{1}{4} \\ \dfrac{\ln\big(16\alpha(1-\alpha)\big)}{\ln\left(\frac{1-\alpha}{\alpha}\right)}, & \text{if} \quad \frac{1}{4}\le\alpha<\frac{3}{4} \\ -\dfrac{\ln 3}{\ln\left(\frac{16\alpha^2}{3}\right)}, & \text{if} \quad \frac{3}{4}\le\alpha\le 1 \end{cases}.$$

For $\alpha\neq 1/2$, the coefficient of variation of the distribution of X_α is

$$\mathrm{CV}(\alpha)=\frac{\sigma(\alpha)}{\mu_1'(\alpha)}=\frac{\sqrt{1+4\alpha-4\alpha^2}}{1-2\alpha}.$$

(vi) We have

$$\mu_3(\alpha)=E\left(\left(X_\alpha-\mu_1'\right)^3\right)=\mu_3'(\alpha)-3\mu_1'(\alpha)\mu_2'(\alpha)+2\left(\mu_1'(\alpha)\right)^3=2(1-2\alpha)^3.$$

Therefore,

$$\beta_1(\alpha)=\frac{\mu_3(\alpha)}{\sigma(\alpha)}=\frac{2(1-2\alpha)^3}{\sqrt{1+4\alpha-4\alpha^2}}.$$

Using (iii), the Yule coefficient of skewness is

$$\beta_2(\alpha)=\frac{q_3(\alpha)-2m(\alpha)+q_1(\alpha)}{q_3(\alpha)-q_1(\alpha)}=\begin{cases} \dfrac{\ln\left(\frac{3}{4}\right)}{\ln 3}, & \text{if} \quad 0\le\alpha<\frac{1}{4} \\ -\dfrac{\ln\big(4\alpha(1-\alpha)\big)}{\ln\big(16\alpha(1-\alpha)\big)}, & \text{if} \quad \frac{1}{4}\le\alpha<\frac{1}{2} \\ \dfrac{\ln\big(4\alpha(1-\alpha)\big)}{\ln\big(16\alpha(1-\alpha)\big)}, & \text{if} \quad \frac{1}{2}\le\alpha<\frac{3}{4} \\ \dfrac{\ln\left(\frac{3}{4}\right)}{\ln 3}, & \text{if} \quad \frac{3}{4}\le\alpha\le 1 \end{cases}.$$

Clearly, for $0 \le \alpha < 1/2, \beta_i(\alpha) > 0, i = 1,2,$ and, for $\frac{1}{2} < \alpha \le 1, \beta_i(\alpha) < 0, i = 1,2.$ It follows that the probability distribution of X_α is positively skewed if $0 \le \alpha < 1/2$ and negatively skewed if $\frac{1}{2} < \alpha \le 1$. For $\alpha = \frac{1}{2}, f_\alpha(x) = f_\alpha(-x), \forall x \in \mathbb{R}.$ Thus, for $\alpha = 1/2$, the probability of X_α is symmetric about zero.

(vii) We have

$$\mu_4(\alpha) = E\left(\left(X_\alpha - \mu_1'\right)^4\right)$$

$$= \mu_4'(\alpha) - 4\mu_1'(\alpha)\mu_3'(\alpha) + 6\left(\mu_1'(\alpha)\right)^2 \mu_2'(\alpha) - 3\left(\mu_1'(\alpha)\right)^4$$

$$= 24 - 12(1-2\alpha)^2 - 3(1-2\alpha)^4.$$

Therefore,

$$\gamma_1(\alpha) = \frac{\mu_4(\alpha)}{\left(\mu_2(\alpha)\right)^2} = \frac{24 - 12(1-2\alpha)^2 - 3(1-2\alpha)^4}{(2-(1-2\alpha)^2)^2}$$

and

$$\gamma_2(\alpha) = \gamma_1(\alpha) - 3 = \frac{12 - 6(1-2\alpha)^4}{(2-(1-2\alpha)^2)^2}.$$

Clearly, for any $\alpha \in [0,1], \gamma_2(\alpha) > 0$. It follows that, for any value of $\alpha \in [0,1]$, the distribution of X_α is leptokurtic.