Probability

Jan-April 2014

Chennai Mathematical Institute

Professor Nandini Kannan

Chapter 5

Moments and Generating Functions

Definition: The r-th **moment** of a random variable X is

$$\mu_r^{'} = E[X^r]. \tag{5.1}$$

The r-th **central moment** is

$$\mu_r = E[(X-\mu)^r], \tag{5.2}$$

where $\mu = \mu_1^{'} = E(X)$ is the expected value or mean of X.

Definition: The **variance** of a random variable is its second central moment.

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = E(X^2) - \mu^2.$$

The positive square root of the variance is called the **standard deviation**.

Remarks:

- The variance and the standard deviation provide a measure of spread of the distribution about its mean.
- $Var(X) = 0$ iff X is a degenerate random variable, i.e. X is constant with probability 1. This implies no variation in X.

Definition: For any random variable X,

$$\alpha_3 = \frac{\mu_3}{(\mu_2)^{3/2}} \tag{5.3}$$

is called the **coefficient of skewness** and is used as a measure of asymmetry.

Definition: A pdf is said to be **symmetric** about the point a if for all $\epsilon > 0$,

$$f(a+\epsilon) = f(a-\epsilon). \tag{5.4}$$

Definition: A rv X is **symmetric** about a point a if

$$P(X \geq a+\epsilon) = P(X \leq a-\epsilon) \qquad \forall \ \ \epsilon$$
$$F(a-\epsilon) = 1 - F(a+\epsilon) + P(X = a+\epsilon).$$

Theorem 5.0.1. *If a pdf is symmetric about the point a, then $\alpha_3 = 0$.*

Proof: We have $\mu_3 = E(X-\mu)^3$. The mean is given by

$$\begin{aligned} \mu &= E(X) = \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_{-\infty}^{a} xf(x)dx + \int_{a}^{\infty} xf(x)dx. \end{aligned}$$

Let $y = x-a$. Then $dy = dx$ and

$$\begin{aligned} E(X) &= \int_{-\infty}^{0} (y+a)f(y+a)dy + \int_{0}^{\infty} (y+a)f(y+a)dy \\ &= \int_{-\infty}^{0} (y+a)f(y+a)dy + \int_{0}^{\infty} (y+a)f(a-y)dy \\ &= a\int_{-\infty}^{\infty} f(a+y)dy = a. \end{aligned}$$

$$\begin{aligned} \mu_3 &= E(X-a)^3 = \int_{-\infty}^{\infty} (x-a)^3 f(x)dx \\ &= 0. \end{aligned}$$

This implies $\alpha_3 = 0$.
■

Example: Let $f(x) = e^{-x}, x > 0$. We have

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

$$E(X^2) = \int_{0}^{\infty} x^2e^{-x}dx = \Gamma(3) = 2.$$
$$E(X^3) = \Gamma(4) = 6.$$

Therefore

$$\mu_3 = E[(X-\mu)^3] = 2, \qquad \mu_2 = 1.$$

Substituting, we have

$$\alpha_3 = 2 > 0,$$

This is a distribution that is skewed right.

■

Definition: For any random variable X,

$$\alpha_4 = \frac{\mu_4}{(\mu_2)^2} \tag{5.5}$$

is called the **coefficient of kurtosis** and measures the peakededness or flatness of the pdf.

Example: Let $X \sim N(0,1)$. We have

$$E(X^4) = 3.$$
$$E(X^2) = 1.$$

Substituting, we have

$$\alpha_4 = 3.$$

This is called a **mesokurtic** curve.
■

Example: Let

$$f_X(x) = \frac{1}{2}, \qquad -1 < x < 1.$$

We have

$$E(X) = 0.$$
$$E(X^2) = \int_{-1}^{1} \frac{x^2}{2} dx = \left.\frac{x^3}{6}\right|_{-1}^{1} = \frac{2}{6}.$$
$$E(X^4) = \frac{2}{10}.$$

Substituting, we have

$$\alpha_4 = \frac{9}{5} < 3.$$

This is called a **platykurtic** curve.
■

Example: Let

$$f_X(x) = \frac{1}{2}e^{-|x|}, \qquad x \in \mathcal{R}.$$

We have

$$E(X^2) = 2.$$
$$E(X^4) = \Gamma(5) = 24.$$

Substituting, we have

$$\alpha_4 = 6 > 3.$$

This is called a **leptokurtic** curve.
■

Definition: Let X be a random variable with cdf $F_X(.)$. The **moment generating function** (mgf) of the random variable X, denoted by $M_X(t)$ is defined as

$$M_X(t) = E(e^{tX}), \tag{5.6}$$

provided the expectation exists in some neighbourhood of 0. We have

$$m_X(t) = \begin{cases} \sum_x e^{tx} p(x), & \text{if } X \text{ is discrete;} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Example Let

$$f_X(x) = \frac{1}{2} e^{-x/2}, \qquad x > 0.$$

We have

$$\begin{aligned} M_X(t) &= \frac{1}{2} \int_0^{\infty} e^{tx} e^{-x/2} dx \\ &= \frac{1}{2} \int_0^{\infty} e^{\left(t - \frac{1}{2}\right)x} dx \\ &= \frac{1}{1-2t} \qquad if \quad t < \frac{1}{2}. \end{aligned}$$

If $t > \frac{1}{2}$, the integral is infinite.
■

Theorem 5.0.2. *If the mgf $M_X(t)$ of X exists in a neighbourhood of 0, the derivatives of all orders exist at $t = 0$ and may be obtained by differentiating under the integral (or summation), i.e.*

$$M_X^n(0) = \frac{d^n}{dt^n} M_X(t)|_{t=0} = E(X^n).$$

Proof: We have

$$\begin{aligned} \frac{d}{dt} M_X(t) &= \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f(x) dx \\ &= \int_{-\infty}^{\infty} \frac{d}{dt}(e^{tx}) f(x) dx \\ &= E(X e^{tX}). \end{aligned}$$

$$\frac{d}{dt} M_X(t)|_{t=0} = E(X).$$

■

Remark: Since

$$\begin{aligned} M_X(t) &= E(e^{tX}) = E\left(1 + tX + \frac{t^2}{2!}X^2 + \ldots\right) \\ &= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \ldots \end{aligned}$$

$E(X^n)$ is the coefficient of $t^k/k!$ in the above expansion.

Example: Consider the Geometric rv with pmf

$$P(X = k) = p(1-p)^{k-1} \qquad k = 1, 2, \ldots.$$

We have

$$\begin{aligned} M_X(t) = \sum_k e^{tk} p(1-p)^{k-1} \\ &= \frac{p}{1-p}\sum_k [(1-p)e^t]^k \\ &= \frac{pe^t}{1-qe^t}. \end{aligned}$$

■

Theorem 5.0.3. *If $Z \sim N(0,1)$, we have*

$$M_Z(t) = e^{t^2/2}. \tag{5.7}$$

Proof: We have

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tz} e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}[z^2 - 2zt]\right] dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}[z^2 - 2zt + t^2]\right] \exp\left[t^2/2\right] dz \\ &= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(z-t)^2\right] dz \\ &= e^{t^2/2}, \end{aligned}$$

since the last integral is the $N(t,1)$ pdf.
■

Theorem 5.0.4. *If $X \sim N(\mu, \sigma^2)$, we have*

$$M_X(t) = e^{\mu t} e^{\frac{t^2\sigma^2}{2}}. \tag{5.8}$$

Proof: We have

$$\begin{aligned} M_X(t) &= E(e^{tX}) = E[e^{t(\sigma Z+\mu)}] \\ &= E[e^{\sigma t Z}]e^{\mu t} \\ &= \exp[\mu t]\exp[(\sigma t)^2/2]. \end{aligned}$$

■

Remarks:

- If the mgf exists, it characterizes an infinite set of moments.
- However. characterizing the infinite set of moments does not uniquely determine a distribution function, i.e. two random variables that are distinct may have the same moments.

Example: Let

$$f_1(x) = \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2}; \qquad x \geq 0$$

and

$$f_2(x) = f_1(x)[1 + \sin(2\pi \log x)]; \qquad x \geq 0.$$

We have

$$E(X_1^r) = e^{r^2/2} \qquad \text{all finite.}$$

We have

$$E(X_2^r) = E(X_1^r) + \int_0^\infty x^r \sin(2\pi \log x) f_1(x)dx.$$

The last integral is equal to zero. Therefore X_1 and X_2 have the same moments for $r = 0, 1, \ldots$ but distinct pdf's.
For this example, the mgf of X_1 does not exist.
■

Remark:

- Existence of moments **does not imply** existence of the mgf.
- If the cdf's have bounded support, i.e.

  $$\mathcal{X} = \{x : f_X(x) \geq 0\}$$

  is a bounded set, then the moments are unique. In this case the infinite sequence of moments will uniquely determine the distribution.
- When the mgf exists, the moment sequence determines the distribution uniquely.

Theorem 5.0.5. *Let $F_X(.)$ and $F_Y(.)$ be two distributions all of whose moments exist.*

(a) If $F_X(.)$ and $F_Y(.)$ have bounded support, then

$$F_X(u) = F_Y(u) \ \forall \ u \Leftrightarrow E(X^r) = E(Y^r), r = 1, 2, \ldots.$$

(b) If the mgf's exist and

$$M_X(t) = M_Y(t) \ \forall \ t \text{ in some neighbourhood of } 0,$$

then

$$F_X(u) = F_Y(u) \ \forall \ u.$$

These are **characterizations** of a distribution.
■

Theorem 5.0.6. ***Convergence of moment generating functions**. Suppose $\{X_i\}$ is a sequence of random variables each with mgf $M_{X_i}(t)$. Suppose*

$$\lim_{i\to\infty} M_{X_i}(t) = M_X(t) \ \forall \ t \in (-h, h),$$

and $M_X(t)$ is a mgf. Then there is a unique cdf $F_X(.)$ whose moments are determined by $M_X(t)$, and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i\to\infty} F_{X_i}(x) = F_X(x).$$

The proof relies on the uniqueness of the Laplace transforms.
■

Since the mgf may not exist for all random variables , we can define a different function that will always exist and has properties similar to the mgf.

Definition The characteristic function of a random variable X is given by

$$\phi_X(t) = E[e^{itX}], \tag{5.9}$$

which always exists and completely determines the distribution.
The **inversion theorem** helps us to compute the cdf from the mgf or the characteristic function.

Example: Let $X \sim Bin(n, p)$, the binomial distribution. We have

$$\begin{aligned} M_X(t) &= (pe^t + q)^n \\ &= (1 - p + pe^t)^n \\ &= \left[1 + \frac{1}{n}(e^t - 1)np\right]^n \\ &= \left[1 + \frac{1}{n}(e^t - 1)\lambda\right]^n, \qquad (np = \lambda). \end{aligned}$$

We have

$$\lim_{\substack{n \to \infty \quad np = \lambda \\ p \to 0}} M_X(t) = e^{\lambda(e^t - 1)},$$

which is the mgf of the Poisson distribution.
■

Example: Let $X \sim P(\lambda)$. Then

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

Let

$$Y = \frac{X - \lambda}{\sqrt{\lambda}}.$$

Then

$$M_Y(t) = e^{-t\sqrt{\lambda}} M\left(\frac{t}{\sqrt{\lambda}}\right).$$

Therefore

$$\begin{aligned}
\log M_Y(t) &= -t\sqrt{\lambda} + \lambda(e^{\frac{t}{\sqrt{\lambda}}} - 1) \\
&= \frac{-t}{(\sqrt{\lambda})^{-1}} + \lambda\left[\frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + \ldots +\right] \\
&= \frac{t^2}{2} + \frac{t^3}{3!\sqrt{\lambda}} + \ldots
\end{aligned}$$

As $\lambda \to \infty$, this converges to $t^2/2$, which implies

$$M_Y(t) \to e^{t^2/2},$$

which is the mgf of the standard normal.
■

Normal Approximation to the Binomial

Let $X \sim B(n,p)$. Then

$$Z = \frac{X - np}{\sqrt{npq}}$$

is approximately distributed as a standard normal random variable as $n \to \infty$.
Recall

$$M_X(t) = (pe^t + q)^n.$$

Let

$$Z = \frac{X - np}{\sqrt{npq}}.$$

Then

$$\begin{aligned}
M_Z(t) &= E(e^{tZ}) = \exp\left[\frac{-npt}{\sqrt{npq}}\right] E\left(\exp\left[\frac{tX}{\sqrt{npq}}\right]\right) \\
&= \exp\left[\frac{-npt}{\sqrt{npq}}\right]\left[p\exp\left(\frac{t}{\sqrt{npq}}\right)+q\right]^n \\
&= \left[p\exp\left(\frac{qt}{\sqrt{npq}}\right)+q\exp\left(\frac{-pt}{\sqrt{npq}}\right)\right]^n \\
&= \left[1+\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right]^n \\
&\to e^{t^2/2} \qquad \text{as } n \to \infty.
\end{aligned}$$

$\blacksquare$