1.7.1 Moments and Moment Generating Functions

Definition 1.12. *The nth moment ($n \in \mathbb{N}$) of a random variable X is defined as*

$$\mu'_n = \mathrm{E}\, X^n$$

The nth central moment of X is defined as

$$\mu_n = \mathrm{E}(X - \mu)^n,$$

where $\mu = \mu'_1 = \mathrm{E}\, X$. □

Note, that the second central moment is the variance of a random variable X, usually denoted by σ^2.

Moments give an indication of the shape of the distribution of a random variable. Skewness and kurtosis are measured by the following functions of the third and fourth central moment respectively:

the coefficient of skewness is given by

$$\gamma_1 = \frac{\mathrm{E}(X-\mu)^3}{\sigma^3} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}};$$

the coefficient of kurtosis is given by

$$\gamma_2 = \frac{\mathrm{E}(X-\mu)^4}{\sigma^4} - 3 = \frac{\mu_4}{\mu_2^2} - 3.$$

Moments can be calculated from the definition or by using so called moment generating function.

Definition 1.13. *The **moment generating function (mgf)** of a random variable X is a function $M_X : \mathbb{R} \to [0, \infty)$ given by*

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

provided that the expectation exists for t in some neighborhood of zero. □

More explicitly, the mgf of X can be written as

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, \quad \text{if } X \text{ is continuous,}$$
$$M_X(t) = \sum_{x \in \mathcal{X}} e^{tx} P(X = x)dx, \quad \text{if } X \text{ is discrete.}$$

The method to generate moments is given in the following theorem.

Theorem 1.7. *If X has mgf $M_X(t)$, then*

$$\mathrm{E}(X^n) = M_X^{(n)}(0),$$

where

$$M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)|_0.$$

That is, the n-th moment is equal to the n-th derivative of the mgf evaluated at $t = 0$.

Proof. Assuming that we can differentiate under the integral sign we may write

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

Hence, evaluating the last expression at zero we obtain

$$\frac{d}{dt} M_X(t)|_0 = \mathrm{E}(X e^{tX})|_0 = \mathrm{E}(X).$$

For $n = 2$ we will get

$$\frac{d^2}{dt^2} M_X(t)|_0 = \mathrm{E}(X^2 e^{tX})|_0 = \mathrm{E}(X^2).$$

Analogously, it can be shown that for any $n \in \mathbb{N}$ we can write

$$\frac{d^n}{dt^n} M_X(t)|_0 = \mathrm{E}(X^n e^{tX})|_0 = \mathrm{E}(X^n).$$

□

Example 1.14. Find the mgf of $X \sim \mathrm{Exp}(\lambda)$ and use results of Theorem 1.7 to obtain the mean and variance of X.

By definition the mgf can be written as

$$M_X(t) = \mathrm{E}(e^t X) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx.$$

For the exponential distribution we have

$$f_X(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x),$$

where $\lambda \in \mathbb{R}_+$. Here we used the notation of the indicator function $I_{\mathcal{X}}(x)$ whose meaning is as follows:

$$I_{\mathcal{X}}(x) = \begin{cases} 1, & \text{if } x \in \mathcal{X}; \\ 0, & \text{otherwise.} \end{cases}$$

That is,

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \in (0,\infty); \\ 0, & \text{otherwise.} \end{cases}$$

Hence, integrating by the method of substitution, we get

$$M_X(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{(t-\lambda)x} dx = \frac{\lambda}{t-\lambda} \quad \text{provided that } |t| < \lambda.$$

Now, using Theorem 1.7 we obtain the first and the second moments, respectively:

$$\mathrm{E}(X) = M'_X(0) = \frac{\lambda}{(\lambda-t)^2}\Big|_{t=0} = \frac{1}{\lambda},$$

$$\mathrm{E}(X^2) = M_X^{(2)}(0) = \frac{2\lambda}{(\lambda-t)^3}\Big|_{t=0} = \frac{2}{\lambda^2}.$$

Hence, the variance of X is

$$\mathrm{var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

□

Exercise 1.10. Calculate mgf for Binomial and Poisson distributions.

Moment generating functions provide methods for comparing distributions or finding their limiting forms. The following two theorems give us the tools.

Theorem 1.8. *Let $F_X(x)$ and $F_Y(y)$ be two cdfs whose all moments exist. Then*

1. *If F_X and F_Y have bounded support, then $F_X(u) = F_Y(u)$ for all u iff $\mathrm{E}(X^n) = \mathrm{E}(Y^n)$ for all $n = 0, 1, 2, \ldots$.*

2. *If the mgfs of X and Y exist and are equal, i.e., $M_X(t) = M_Y(t)$ for all t in some neighborhood of zero, then $F_X(u) = F_Y(u)$ for all u.* □

Theorem 1.9. *Suppose that $\{X_1, X_2, \ldots\}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that*

$$\lim_{i\to\infty} M_{X_i}(t) = M_X(t), \quad \text{for all } t \text{ in a neighborhood of zero},$$

and $M_X(t)$ is an 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

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

□

This theorem means that the convergence of mgfs implies convergence of cdfs.

Example 1.15. We know that the Binomial distribution can be approximated by a Poisson distribution when p is small and n is large. Using the above theorem we can confirm this fact.

The mgf of $X_n \sim \mathrm{Bin}(n,p)$ and of $Y \sim \mathrm{Poisson}(\lambda)$ are, respectively:

$$M_{X_n}(t) = [pe^t + (1-p)]^n, \quad M_Y(t) = e^{\lambda(e^t-1)}.$$

We will show that the mgf of X tends to the mgf of Y, where $\lambda = np$.

We will need the following useful result given in the lemma:

Lemma 1.1. *Let $a_1, a_2, \ldots$ be a sequence of numbers converging to a, that is, $\lim_{n\to\infty} a_n = a$. Then*

$$\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n = e^a.$$

Now, we can write

$$\begin{aligned}
M_{X_n}(t) &= \left(pe^t + (1-p)\right)^n \\
&= \left(1 + \frac{1}{n}np(e^t-1)\right)^n \\
&= \left(1 + \frac{\lambda(e^t-1)}{n}\right)^n \\
&\underset{n\to\infty}{\longrightarrow} e^{\lambda(e^t-1)} = M_Y(t).
\end{aligned}$$

Hence, by Theorem 1.9 the Binomial distribution converges to a Poisson distribution.

□

1.8 Functions of Random Variables

If X is a random variable with cdf $F_X(x)$, then any function of X, say $g(X) = Y$ is also a random variable. The question then is "what is the distribution of Y?"

The function $y = g(x)$ is a mapping from the induced sample space of the random variable X, $\mathcal{X}$, to a new sample space, $\mathcal{Y}$, of the random variable Y, that is

$$g(x) : \mathcal{X} \to \mathcal{Y}.$$

The inverse mapping g^{-1} acts from $\mathcal{Y}$ to $\mathcal{X}$ and we can write

$$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\} \quad \text{where } A \subset \mathcal{Y}.$$

Then, we have

$$\begin{aligned} P(Y \in A) &= P(g(X) \in A) \\ &= P\big(\{x \in \mathcal{X} : g(x) \in A\}\big) \\ &= P\big(X \in g^{-1}(A)\big). \end{aligned}$$

The following theorem relates the cumulative distribution functions of X and $Y = g(X)$.

Theorem 1.10. *Let X have cdf $F_X(x)$, $Y = g(X)$ and let domain and codomain of $g(X)$, respectively, be*

$$\mathcal{X} = \{x : f_X(x) > 0\}, \quad and\ \mathcal{Y} = \{y : y = g(x)\ for\ some\ x \in \mathcal{X}\}.$$

(a) *If g is an increasing function on $\mathcal{X}$ then $F_Y(y) = F_X\big(g^{-1}(y)\big)$ for $y \in \mathcal{Y}$.*

(b) *If g is a decreasing function on $\mathcal{X}$, then $F_Y(y) = 1 - F_X\big(g^{-1}(y)\big)$ for $y \in \mathcal{Y}$.*

Proof. The cdf of $Y = g(X)$ can be written as

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(g(X) \le y) \\ &= P\big(\{x \in \mathcal{X} : g(x) \le y\}\big) \\ &= \int_{\{x \in \mathcal{X} : g(x) \le y\}} f_X(x)dx. \end{aligned}$$

(a) If g is increasing, then

$$\{x \in \mathcal{X} : g(x) \le y\} = \{x \in \mathcal{X} : g^{-1}\big(g(x)\big) \le g^{-1}(y)\} = \{x \in \mathcal{X} : x \le g^{-1}(y)\}.$$

So, we can write

$$\begin{aligned}F_Y(y) &= \int_{\{x\in\mathcal{X}:g(x)\le y\}} f_X(x)dx \\ &= \int_{\{x\in\mathcal{X}:x\le g^{-1}(y)\}} f_X(x)dx \\ &= \int_{-\infty}^{g^{-1}(y)} f_X(x)dx \\ &= F_X\big(g^{-1}(y)\big).\end{aligned}$$

(b) Now, if g is decreasing, then

$$\{x \in \mathcal{X} : g(x) \le y\} = \{x \in \mathcal{X} : g^{-1}\big(g(x)\big) \ge g^{-1}(y)\} = \{x \in \mathcal{X} : x \ge g^{-1}(y)\}.$$

So, we can write

$$\begin{aligned}F_Y(y) &= \int_{\{x\in\mathcal{X}:g(x)\le y\}} f_X(x)dx \\ &= \int_{\{x\in\mathcal{X}:x\ge g^{-1}(y)\}} f_X(x)dx \\ &= \int_{g^{-1}(y)}^{\infty} f_X(x)dx \\ &= 1 - F_X\big(g^{-1}(y)\big).\end{aligned}$$

□

Example 1.16. Find the distribution of $Y = g(X) = -\log X$, where $X \sim \mathcal{U}([0,1])$. The cdf of X is

$$F_X(x) = \begin{cases} 0, & \text{for } x < 0; \\ x, & \text{for } 0 \le x \le 1; \\ 1, & \text{for } x > 1. \end{cases}$$

For $x \in [0,1]$ the function $g(x) = -\log x$ is defined on $\mathcal{Y} = (0,\infty)$ and it is decreasing.

For $y > 0$, $y = -\log x$ implies that $x = e^{-y}$, i.e., $g^{-1}(y) = e^{-y}$ and

$$F_Y(y) = 1 - F_X\big(g^{-1}(y)\big) = 1 - F_X\big(e^{-y}\big) = 1 - e^{-y}.$$

Hence we may write

$$F_Y(y) = \big(1 - e^{-y}\big)I_{(0,\infty)}.$$

This is exponential distribution function for $\lambda = 1$.

□

For continuous rvs we have the following result.

Theorem 1.11. *Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Suppose that $f_X(x)$ is continuous on its support $\mathcal{X} = \{x : f_X(x) > 0\}$ and that $g^{-1}(y)$ has a continuous derivative on support $\mathcal{Y} = \{y : y = g(x)\ for\ some\ x \in \mathcal{X}\}$. Then the pdf of Y is given by*

$$f_Y(y) = f_X\big(g^{-1}(y)\big)\left|\frac{d}{dy}g^{-1}(y)\right|I_{\mathcal{Y}}.$$

Proof.

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy}F_Y(y) \\
&= \begin{cases} \frac{d}{dy}\big\{F_X\big(g^{-1}(y)\big)\big\}, & \text{if } g \text{ is increasing;} \\ \frac{d}{dy}\big\{1 - F_X\big(g^{-1}(y)\big)\big\}, & \text{if } g \text{ is decreasing.} \end{cases} \\
&= \begin{cases} f_X\big(g^{-1}(y)\big)\frac{d}{dy}g^{-1}(y), & \text{if } g \text{ is increasing;} \\ -f_X\big(g^{-1}(y)\big)\frac{d}{dy}g^{-1}(y), & \text{if } g \text{ is decreasing.} \end{cases}
\end{aligned}$$

which gives the thesis of the theorem.

□

Example 1.17. Suppose that $Z \sim \mathcal{N}(0, 1)$. What is the distribution of $Y = Z^2$?

For $Y > 0$, the cdf of $Y = Z^2$ is

$$\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(Z^2 \le y) \\
&= P(-\sqrt{y} \le Z \le \sqrt{y}) \\
&= F_Z(\sqrt{y}) - F_Z(-\sqrt{y}).
\end{aligned}$$

The pdf can now be obtained by differentiation:

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy}F_Y(y) \\
&= \frac{d}{dy}\big(F_Z(\sqrt{y}) - F_Z(-\sqrt{y})\big) \\
&= \frac{1}{2\sqrt{y}}f_Z(\sqrt{y}) + \frac{1}{2\sqrt{y}}f_Z(-\sqrt{y})
\end{aligned}$$

Now, for the standard normal distribution we have

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

This gives,

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi}}e^{(-\sqrt{y})^2/2} + \frac{1}{\sqrt{2\pi}}e^{(\sqrt{y})^2/2}\right] \\ &= \frac{1}{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-y/2}, \quad 0 < y < \infty. \end{aligned}$$

This is a well known pdf function, which we will use in statistical inference. It is called *chi squared random variable with one degree of freedom* and it is denoted by χ^2_1.

Note that $g(Z) = Z^2$ is not a monotone function, but the range of Z, $(-\infty, \infty)$, can be partitioned so that it is monotone on its sub-sets. □

Exercise 1.11. The pdf obtained in Example 1.17 is also pdf of a Gamma rv for some specific values of its parameters. What are these values?

Exercise 1.12. Suppose that $Z \sim \mathcal{N}(0, 1)$. Find the distribution of $Y = \mu + \sigma Z$ for constant μ and σ.

Exercise 1.13. Let X be a random variable with moment generating function M_X.

(i) Show that the moment generating function of $Y = a + bX$, where a and b are constants, is given by

$$M_Y(t) = e^{ta}M_X(tb).$$

(ii) Derive the moment generating function of $Y \sim \mathcal{N}(\mu, \sigma^2)$. Hint: First find $M_Z(t)$ for a standard normal rv Z.