Chapter 3
Moments of a Distribution

Expectation

We develop the expectation operator in terms of the Lebesgue integral.

• Recall that the Lebesgue measure $\lambda(A)$ for some set $A$ gives the length/area/volume of the set $A$. If $A = (3; 7)$, then $\lambda(A) = |3 - 7| = 4$.

• The Lebesgue integral of $f$ on $[a, b]$ is defined in terms of $\sum y_i \lambda(A_i)$, where $0 = y_1 \leq y_2 \leq ... \leq y_n$, $A_i = \{x: y_i \leq f(x) < y_{i+1}\}$, and $\lambda(A)$ is the Lebesgue measure of the set $A$.

• The value of the Lebesgue integral is the limit as the $y_i$'s are pushed closer together. That is, we break the $y$-axis into a grid using $\{y_i\}$ and break the $x$-axis into the corresponding grid $\{A_i\}$ where

$$A_i = \{x: f(x) \in [y_i, y_{i+1})\}.$$
Taking expectations: Riemann vs Lebesgue

- Riemann's approach
  Partition the base. Measure the height of the function at the center of each interval. Calculate the area of each interval. Add all intervals.
- Lebesgue approach
  Divide the range of the function. Measure the length of each horizontal interval. Calculate the area of each interval. Add all intervals.

A Borel function (RV) \( f \) is integrable if and only if \( |f| \) is integrable.

For convenience, we define the integral of a measurable function \( f \) from \((\Omega, \Sigma, \mu)\) to \((\overline{\mathbb{R}}, \overline{\mathcal{B}})\), where \( \overline{\mathbb{R}} = \mathbb{R} \cup \{ -\infty, \infty \} \), \( \overline{\mathcal{B}} = \sigma(\mathcal{B} \cup \{\infty\}, \{-\infty\}) \).

**Example:** If \( \Omega = \mathbb{R} \) and \( \mu \) is the Lebesgue measure, then the Lebesgue integral of \( f \) over an interval \([a, b]\) is written as \( \int_{[a,b]} f(x) \, dx = \int_a^b f(x) \, dx \), which agrees with the Riemann integral when the latter is well defined.

However, there are functions for which the Lebesgue integrals are defined but not the Riemann integrals.

- If \( \mu = P \), in statistics, \( \int X \, dP = EX = E[X] \) is called the expectation or expected value of \( X \).
Expected Value

Consider our probability space \((\Omega, \Sigma, P)\). Take an event (a set \(A\) of \(\omega \in \Omega\)) and \(X\), a RV, that assigns real numbers to each \(\omega \in A\).

- If we take an observation from \(A\) without knowing which \(\omega \in A\) will be drawn, we may want to know what value of \(X(\omega)\) we should expect to see.

- Each of the \(\omega \in A\) has been assigned a probability measure \(P[\omega]\), which induces \(P[\alpha]\). Then, we use this to weight the values \(X(\omega)\).

- \(P\) is a probability measure: The weights sum to 1. The weighted sum provides us with a weighted average of \(X(\omega)\). If \(P\) gives the "correct" likelihood of \(\omega\) being chosen, the weighted average of \(X(\omega)\) – \(E[X]\) – tells us what values of \(X(\omega)\) are expected.

### Expected Value

- Now with the concept of the Lebesgue integral, we take the possible values \(\{x_i\}\) and construct a grid on the \(y\)-axis, which gives a corresponding grid on the \(x\)-axis in \(A\), where 
  \[ A_i = \{\omega \in A: X(\omega) \in [x_i, x_{i+1})\} \].

Let the elements in the \(x\)-axis grid be \(A_i\). The weighted average is

\[
\sum_{i=1}^{n} x_i P[A_i] = \sum_{i=1}^{n} x_i P_X [X = x_i] = \sum_{i=1}^{n} x_i f_X (x_i)
\]

- As we shrink the grid towards 0, \(A\), becomes infinitesimal. Let \(d\omega\) be the infinitesimal set \(A\). The Lebesgue integral becomes:

\[
\lim_{n \to \infty} \sum_{i=1}^{n} x_i P[A_i] = \int_{-\infty}^{\infty} x P[d\omega] = \int_{-\infty}^{\infty} x P_X [X = x_i] = \int_{-\infty}^{\infty} x f_X (x_i) dx
\]
The Expectation of $X$: $E(X)$

The expectation operator defines the mean (or population average) of a random variable or expression.

**Definition**

Let $X$ denote a discrete RV with probability function $p(x)$ (probability density function $f(x)$ if $X$ is continuous) then the expected value of $X$, $E(X)$ is defined to be:

$$E(X) = \sum_{x} xp(x) = \sum_{i} x_{i} p(x_{i})$$

and if $X$ is continuous with probability density function $f(x)$

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

Sometimes we use $E[.]$ as $E_{X}[.]$ to indicate that the expectation is being taken over $f_{X}(x) \, dx$.

**Interpretation of $E(X)$**

1. The expected value of $X$, $E(X)$, is the center of gravity of the probability distribution of $X$.
2. The expected value of $X$, $E(X)$, is the long-run average value of $X$. (To be discussed later: Law of Large Numbers)
Example: The Binomial distribution

Let $X$ be a discrete random variable having the Binomial distribution -- i.e., $X$ is the number of successes in $n$ independent repetitions of a Bernoulli trial. Find the expected value of $X$, $E(X)$.

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, 3, \ldots, n$$

$$E(X) = \sum_{x=0}^{n} xp(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} (1-p)^{n-x}$$

$$E(X) = p(1-p)^{n-1} + \frac{n!}{1!(n-2)!} p^2 (1-p)^{n-2} + \frac{n!}{2!(n-3)!} p^3 (1-p)^{n-3} \ldots + \frac{n!}{(n-1)!1!} p^n + \frac{n!}{n!0!} p^n$$

Example: Solution

$$E(X) = \sum_{x=0}^{n} xp(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{x!(n-x)!} (1-p)^{n-x}$$

$$E(X) = p(1-p)^{n-1} + \frac{n!}{1!(n-2)!} p^2 (1-p)^{n-2} + \frac{n!}{2!(n-3)!} p^3 (1-p)^{n-3} \ldots + \frac{n!}{(n-1)!1!} p^n + \frac{n!}{n!0!} p^n$$
Example: Solution

\[ np \left[ \frac{(n-1)!}{0!(n-1)!} p^0 (1-p)^{n-1} + \frac{(n-1)!}{1!(n-2)!} p^1 (1-p)^{n-2} + \right. \]
\[ \left. \cdots + \frac{(n-1)!}{(n-2)!} p^{n-2} (1-p) + \frac{(n-1)!}{(n-1)!} p^{n-1} \right] \]
\[ = np \left[ \begin{pmatrix} n-1 \\ 0 \end{pmatrix} p^0 (1-p)^{n-1} + \begin{pmatrix} n-1 \\ 1 \end{pmatrix} p^1 (1-p)^{n-2} + \right. \]
\[ \left. \cdots + \begin{pmatrix} n-1 \\ n-2 \end{pmatrix} p^{n-2} (1-p) + \begin{pmatrix} n-1 \\ n-1 \end{pmatrix} p^{n-1} \right] \]
\[ = np \left[ p + (1-p) \right]^{n-1} = np \left[ 1 \right]^{n-1} = np \]

Example: Exponential Distribution

Let \( X \) have an exponential distribution with parameter \( \lambda \). The probability density function of \( X \) is:

\[ f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \]

The expected value of \( X \) is:

\[ E(X) = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{\infty} x \lambda e^{-\lambda x} \, dx \]

We will determine \( \int x \lambda e^{-\lambda x} \, dx \)
\[ \int u \, dv = uv - \int v \, du \]
Example: Exponential Distribution

We will determine \( \int x \lambda e^{-\lambda x} \, dx \) using integration by parts.

In this case \( u = x \) and \( dv = \lambda e^{-\lambda x} \, dx \)

Hence \( du = dx \) and \( v = -e^{-\lambda x} \)

Thus \( \int x \lambda e^{-\lambda x} \, dx = -xe^{-\lambda x} + \int e^{-\lambda x} \, dx = -xe^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \)

\[ E(X) = \int_0^\infty x \lambda e^{-\lambda x} \, dx = -xe^{-\lambda x} \bigg|_0^\infty - \frac{1}{\lambda} e^{-\lambda x} \bigg|_0^\infty \]

\[ = (-0 + 0) - \left( 0 - \frac{1}{\lambda} \right) = \frac{1}{\lambda} \]

Summary: If \( X \) has an exponential distribution with parameter \( \lambda \), then:

\[ E(X) = \frac{1}{\lambda} \]

Example: The Uniform distribution

Suppose \( X \) has a uniform distribution from \( a \) to \( b \).

Then:

\[ f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x < a, x > b \end{cases} \]

The expected value of \( X \) is:

\[ E(X) = \int_{-\infty}^{\infty} x f(x) \, dx = \int_a^b x \frac{1}{b-a} \, dx \]

\[ = \left[ \frac{1}{b-a} x^2 \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a + b}{2} \]
Example: The Normal distribution

Suppose $X$ has a Normal distribution with parameters $\mu$ and $\sigma$.

Then:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The expected value of $X$ is:

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

Make the substitution:

$$z = \frac{x-\mu}{\sigma}$$

$$dz = \frac{1}{\sigma} \, dx \quad \text{and} \quad x = \mu + z\sigma$$

Hence

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} (\mu + z\sigma) e^{-\frac{z^2}{2}} \, dz$$

$$= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ze^{-\frac{z^2}{2}} \, dz$$

Now

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz = 1$$

and

$$\int_{-\infty}^{\infty} ze^{-\frac{z^2}{2}} \, dz = 0$$

Thus

$$E(X) = \mu$$
Example: The Gamma distribution

Suppose $X$ has a Gamma distribution with parameters $\alpha$ and $\lambda$. Then:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Note:

$$\int_{-\infty}^{\infty} f(x) \, dx = \int_{0}^{\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \, dx = 1 \quad \text{if} \quad \lambda > 0, \alpha \geq 0.$$

This is a very useful formula when working with the Gamma distribution.

\[\text{Example: The Gamma distribution}\]

The expected value of $X$ is:

$$E(X) = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{\infty} x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \, dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_{0}^{\infty} x^\alpha e^{-\lambda x} \, dx$$

$$= \frac{\lambda^\alpha \Gamma(\alpha + 1)}{\Gamma(\alpha) \lambda^{\alpha+1}} \int_{0}^{\infty} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha + 1)} x^{\alpha} e^{-\lambda x} \, dx$$

$$= \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha) \lambda} = \frac{\alpha \Gamma(\alpha)}{\Gamma(\alpha) \lambda} = \frac{\alpha}{\lambda}$$

This is now equal to 1.
Example: The Gamma distribution

Thus, if $X$ has a Gamma $(\alpha, \lambda)$ distribution, the expected value of $X$ is:

$$E(X) = \frac{\alpha}{\lambda}$$

**Special Cases:** $(\alpha, \lambda)$ distribution then the expected value of $X$ is:

1. **Exponential ($\lambda$) distribution:** $\alpha = 1$, $\lambda$ arbitrary

   $$E(X) = \frac{1}{\lambda}$$

2. **Chi-square ($\nu$) distribution:** $\alpha = \nu/2$, $\lambda = 1/2$. 

   $$E(X) = \frac{\nu}{2} \cdot \frac{1}{2} = \nu$$

![Example: The Gamma distribution](image-url)
The Exponential distribution

\[ E(X) = \frac{1}{\lambda} \]

The Chi-square (\( \chi^2 \)) distribution

\[ E(X) = \nu \]
Expectation of a function of a RV

• Let $X$ denote a discrete RV with probability function $p(x)$ (or pdf $f(x)$ if $X$ is continuous) then the expected value of $g(X)$, $E[g(X)]$, is defined to be:

$$E[g(X)] = \sum_x g(x) p(x) = \sum_i g(x_i) p(x_i)$$

and if $X$ is continuous with probability density function $f(x)$

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

• Examples:
  - $g(x) = (x - \mu)^2 \Rightarrow E[g(x)] = E[(x - \mu)^2]$
  - $g(x) = (x - \mu)^k \Rightarrow E[g(x)] = E[(x - \mu)^k]$

Expectation of a function of a RV

**Example:** Suppose $X$ has a uniform distribution from 0 to $b$. Then:

$$f(x) = \begin{cases} \frac{1}{b} & 0 \leq x \leq b \\ 0 & x < 0, x > b \end{cases}$$

Find the expected value of $A = X^2$.

If $X$ is the length of a side of a square (chosen at random from 0 to $b$) then $A$ is the area of the square

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) \, dx = \int_0^b x^2 \frac{1}{b} \, dx = \left[ \frac{1}{b} \cdot \frac{x^3}{3} \right]_0^b = \frac{b^3}{3} - \frac{0^3}{3(b)} = \frac{b^3}{3}$$

$$= 1/3 \text{ the maximum area of the square}$$
Median: An alternative central measure

• A median is described as the numeric value separating the higher half of a sample, a population, or a probability distribution, from the lower half.

Definition: Median
The median of a random variable $X$ is the unique number $m$ that satisfies the following inequalities:

$$P(X \leq m) \geq \frac{1}{2} \quad \text{and} \quad P(X \geq m) \geq \frac{1}{2}.$$ 

For a continuous distribution, we have that $m$ solves:

$$\int_{-\infty}^{m} f_X(x) dx = \int_{m}^{\infty} f_X(x) dx = \frac{1}{2}.$$

Median: An alternative central measure

• Calculation of medians is a popular technique in summary statistics and summarizing statistical data, since it is simple to understand and easy to calculate, while also giving a measure that is more robust in the presence of outlier values than is the mean.

An optimality property
A median is also a central point which minimizes the average of the absolute deviations. That is, a value of $c$ that minimizes

$$E(|X - c|)$$

is the median of the probability distribution of the random variable $X$. 
Example I: Median of the Exponential Distribution

Let $X$ have an exponential distribution with parameter $\lambda$. The probability density function of $X$ is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The median $m$ solves the following integral of $X$:

$$\int_{m}^{\infty} f_X(x) dx = \frac{1}{2}$$

$$\int_{m}^{\infty} \lambda e^{-\lambda x} dx = \lambda \int_{m}^{\infty} e^{-\lambda x} dx = -e^{-\lambda x} \bigg|_{m}^{\infty} = e^{-\lambda m} = \frac{1}{2}$$

That is, $m = \ln(2)/\lambda$.

Example II: Median of the Pareto Distribution

Let $X$ follow a Pareto distribution with parameters $\alpha$ (scale) and $x_s$ (shape, usually notated $x_m$). The pdf of $X$ is:

$$f(x) = \begin{cases} \frac{\alpha x_s^{\alpha}}{x^{\alpha+1}} & x \geq x_s > 0 \\ 0 & x < 0 \end{cases}$$

The median $m$ solves the following integral of $X$:

$$\int_{m}^{\infty} f_X(x) dx = \frac{1}{2}$$

$$\int_{m}^{\infty} \frac{\alpha x_s^{\alpha}}{x^{\alpha+1}} dx = \alpha x_s^{\alpha} \int_{m}^{\infty} x^{-(\alpha+1)} dx = \alpha x_s^{\alpha} \frac{x^{-(\alpha+1)+1}}{-(\alpha + 1) + 1} + C$$

$$= -x_s^{\alpha} x^{-\alpha} + C \bigg|_{m}^{\infty} = x_s^{\alpha} m^{-\alpha} = \frac{1}{2} \Rightarrow m = x_s 2^{1/\alpha}$$

Note: The Pareto distribution is used to describe the distribution of wealth.
Moments of a Random Variable

The moments of a random variable $X$ are used to describe the behavior of the RV (discrete or continuous).

**Definition: $K$th Moment**

Let $X$ be a RV (discrete or continuous), then the $k$th moment of $X$ is:

$$
\mu_k = E \left( X^k \right) = \begin{cases} 
\sum_{x} x^k p(x) & \text{if } X \text{ is discrete} \\
\int_{-\infty}^{\infty} x^k f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}
$$

- The first moment of $X$, $\mu = \mu_1 = E(X)$ is the center of gravity of the distribution of $X$.
- The higher moments give different information regarding the shape of the distribution of $X$.

Moments of a Random Variable

**Definition: Central Moments**

Let $X$ be a RV (discrete or continuous). Then, the $k$th central moment of $X$ is defined to be:

$$
\mu_k^0 = E \left[ (X - \mu)^k \right] = \begin{cases} 
\sum_{x} (x - \mu)^k p(x) & \text{if } X \text{ is discrete} \\
\int_{-\infty}^{\infty} (x - \mu)^k f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}
$$

where $\mu = \mu_1 = E(X) = \text{the first moment of } X$.

- The central moments describe how the probability distribution is distributed about the center of gravity, $\mu$. 

Moments of a Random Variable – 1st and 2nd

The first central moments is given by:

\[ \mu_1^0 = E[X - \mu] \]

The second central moment depends on the spread of the probability distribution of \( X \) about \( \mu \). It is called the variance of \( X \) and is denoted by the symbol \( \text{var}(X) \).

\[ \mu_2^0 = E[(X - \mu)^2] = \text{2nd central moment.} \]

\[ \sqrt{\mu_2^0} = \sqrt{E[(X - \mu)^2]} \]

is called the standard deviation of \( X \) and is denoted by the symbol \( \sigma \).

\[ \text{var}(X) = \mu_2^0 = E[(X - \mu)^2] = \sigma^2 \]

Moments of a Random Variable – Skewness

• The third central moment \( \mu_3^0 = E[(X - \mu)^3] \)

contains information about the skewness of a distribution.

• A popular measure of skewness: \( \gamma_1 = \frac{\mu_3^0}{\sigma^3} = \frac{\mu_3^0}{(\mu_2^0)^{3/2}} \)

• Distribution according to skewness:

1) Symmetric distribution

\[ \mu_3^0 = 0, \gamma_1 = 0 \]
Moments of a Random Variable – Skewness

2) Positively skewed distribution

\[ \mu_3^0 > 0, \gamma_1 > 0 \]

3) Negatively skewed distribution

\[ \mu_3^0 < 0, \gamma_1 < 0 \]

Moments of a Random Variable – Skewness

- Skewness and Economics
  - Zero skew means symmetrical gains and losses.
  - Positive skew suggests many small losses and few rich returns.
  - Negative skew indicates lots of minor wins offset by rare major losses.

- In financial markets, stock returns at the firm level show positive skewness, but at stock returns at the aggregate (index) level show negative skewness.

- From horse race betting and from U.S. state lotteries there is evidence supporting the contention that gamblers are not necessarily risk-lovers but skewness-lovers: Long shots are overbet (positive skewness loved!).
Moments of a Random Variable – Kurtosis

• The fourth central moment \( \mu_4^0 = E\left[ (X - \mu)^4 \right] \)

It contains information about the *shape* of a distribution. The property of shape that is measured by this moment is called *kurtosis*.

• The measure of (excess) kurtosis: \( \gamma_2 = \frac{\mu_4^0}{\sigma^4} - 3 = \frac{\mu_4^0}{(\mu_2^0)^2} - 3 \)

• Distributions:
  1) Mesokurtic distribution

\[ \gamma_2 = 0, \mu_4^0 \text{ moderate in size} \]

Moments of a Random Variable – Kurtosis

2) Platykurtic distribution

\[ \gamma_2 < 0, \mu_4^0 \text{ small in size} \]

3) Leptokurtic distribution

\[ \gamma_2 > 0, \mu_4^0 \text{ large in size} \]
Moments of a Random Variable

Example: The uniform distribution from 0 to 1

\[ f(x) = \begin{cases} 
1 & 0 \leq x \leq 1 \\
0 & x < 0, x > 1 
\end{cases} \]

Finding the moments

\[ \mu_k = \int_{-\infty}^{\infty} x^k f(x) \, dx = \int_{0}^{1} x^k \, dx = \left[ \frac{x^{k+1}}{k+1} \right]_{0}^{1} = \frac{1}{k+1} \]

Finding the central moments:

\[ \mu_k^0 = \int_{-\infty}^{\infty} (x - \mu)^k f(x) \, dx = \int_{0}^{1} (x - \frac{1}{2})^k \, dx \]

Moments of a Random Variable

Finding the central moments (continuation):

\[ \mu_k^0 = \int_{-\infty}^{\infty} (x - \mu)^k f(x) \, dx = \int_{0}^{1} (x - \frac{1}{2})^k \, dx \]

making the substitution \( w = x - \frac{1}{2} \)

\[ \mu_k^0 = \int_{-\frac{1}{2}}^{\frac{1}{2}} w^k \, dw = \left[ \frac{w^{k+1}}{k+1} \right]_{-\frac{1}{2}}^{\frac{1}{2}} = \left( \frac{1}{2} \right)^{k+1} - \left( -\frac{1}{2} \right)^{k+1} \]

\[ = \frac{1 - (-1)^{k+1}}{2^{k+1} (k+1)} = \begin{cases} 
\frac{1}{2^k (k+1)} & \text{if } k \text{ even} \\
0 & \text{if } k \text{ odd} 
\end{cases} \]
Moments of a Random Variable

Hence  \( \mu_2^0 = \frac{1}{2^2} \left( \frac{3}{3} \right) = \frac{1}{12} \),  \( \mu_3^0 = 0 \),  \( \mu_4^0 = \frac{1}{2^4} \left( \frac{5}{5} \right) = \frac{1}{80} \)

Thus,  \( \text{var}(X) = \mu_2^0 = \frac{1}{12} \)

The standard deviation  \( \sigma = \sqrt{\text{var}(X)} = \sqrt{\mu_2^0} = \frac{1}{\sqrt{12}} \)

The measure of skewness  \( \gamma_1 = \frac{\mu_3^0}{\sigma^3} = 0 \)

The measure of kurtosis  \( \gamma_1 = \frac{\mu_4^0}{\sigma^4} - 3 = \frac{1/80}{(1/12)^2} - 3 = -1.2 \)

Alternative measures of dispersion

When the median is used as a central measure for a distribution, there are several choices for a measure of variability:

- The range — the length of the smallest interval containing the data
- The interquartile range — the difference between the 3rd and 1st quartiles.
- The mean absolute deviation —  \( (1/n) \sum_i |x_i - \text{central measure}(X)| \)
- The median absolute deviation (MAD) —  \( \text{MAD} = m_i(|x_i - m(X)|) \)

These measures are more robust (to outliers) estimators of scale than the sample variance or standard deviation.

They especially behave better with distributions without a mean or variance, such as the Cauchy distribution.
Rules for Expectations

\[ E[g(X)] = \begin{cases} 
\sum_{x} g(x) p(x) & \text{if } X \text{ is discrete} \\
\int_{-\infty}^{\infty} g(x) f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases} \]

• Rules:

1. \( E[c] = c \) where \( c \) is a constant

Proof:

if \( g(X) \equiv c \) then \( E[g(X)] = E[c] = \int_{-\infty}^{\infty} cf(x) \, dx = c \int_{-\infty}^{\infty} f(x) \, dx = c \)

The proof for discrete random variables is similar.

Rules for Expectations

2. \( E[aX + b] = aE[X] + b \) where \( a, b \) are constants

Proof

if \( g(X) \equiv aX + b \) then \( E[aX + b] = \int_{-\infty}^{\infty} (ax + b) f(x) \, dx \)

\[ = a \int_{-\infty}^{\infty} xf(x) \, dx + b \int_{-\infty}^{\infty} f(x) \, dx \]

\[ = aE(X) + b \]

The proof for discrete random variables is similar.
Rules for Expectations

3. \( \text{var}(X) = \mu_2 - \mu_1^2 = E[(X - \mu)^2] = E(X^2) - [E(X)]^2 \)

Proof:

\[
\text{var}(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx \\
= \int_{-\infty}^{\infty} (x^2 - 2x\mu + \mu^2) f(x) \, dx \\
= \int_{-\infty}^{\infty} x^2 f(x) \, dx - 2\mu \int_{-\infty}^{\infty} xf(x) \, dx + \mu^2 \int_{-\infty}^{\infty} f(x) \, dx \\
= E(X^2) - 2\mu [E(X)] + \mu^2 = \mu_2 - \mu_1^2 
\]

The proof for discrete random variables is similar.

Rules for Expectations

4. \( \text{var}(aX + b) = a^2 \text{var}(X) \)

Proof:

\[
\mu_{aX+b} = E[aX + b] = aE[X] + b = a\mu + b \\
\text{var}(aX + b) = E\left((aX + b - \mu_{aX+b})^2\right) \\
= E\left((aX + b - [a\mu + b])^2\right) \\
= E\left[a^2 (X - \mu)^2\right] \\
= a^2 E\left((X - \mu)^2\right) = a^2 \text{var}(X)
\]
Definition: Moment Generating Function (MGF)

Let $X$ denote a random variable. Then, the moment generating function of $X$, $m_X(t)$, is defined by:

$$m_X(t) = E[e^{itX}] = \begin{cases} 
\sum_x e^{itx} p(x) & \text{if } X \text{ is discrete} \\
\int_{-\infty}^{\infty} e^{itx} f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}$$

Moment generating functions

The expectation of a function $g(X)$ is given by:

$$E[g(X)] = \begin{cases} 
\sum_x g(x) p(x) & \text{if } X \text{ is discrete} \\
\int_{-\infty}^{\infty} g(x) f(x) \, dx & \text{if } X \text{ is continuous}
\end{cases}$$
MGF: Examples

1. The Binomial distribution (parameters $p, n$)

$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, 2, K, n$$

The MGF of $X, m_X(t)$ is:

$$m_X(t) = E\left[e^{tX}\right] = \sum_x e^{tx} p(x)$$

$$= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1 - p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1 - p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$$

$$= (a + b)^n = (e^t p + 1 - p)^n$$

MGF: Examples

2. The Poisson distribution (parameter $\lambda$)

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x = 0, 1, 2, K$$

The MGF of $X, m_X(t)$ is:

$$m_X(t) = E\left[e^{tX}\right] = \sum_x e^{tx} p(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} \quad \text{using} \quad e^u = \sum_{x=0}^{\infty} \frac{u^x}{x!}$$

$$= e^{\lambda (e^t - 1)}$$
MGF: Examples

3. The Exponential distribution (parameter $\lambda$)

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The MGF of $X$, $m_X(t)$ is:

$$m_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) \, dx = \int_{0}^{\infty} e^{tx} \lambda e^{-\lambda x} \, dx$$

$$= \int_{0}^{\infty} \lambda e^{(t-\lambda)x} \, dx = \left[ \frac{\lambda e^{(t-\lambda)x}}{t-\lambda} \right]_{0}^{\infty}$$

$$= \begin{cases} \frac{\lambda}{\lambda - t} & t < \lambda \\ \text{undefined} & t \geq \lambda \end{cases}$$

MGF: Examples

4. The Standard Normal distribution ($\mu = 0, \sigma = 1$)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

The MGF of $X$, $m_X(t)$ is:

$$m_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) \, dx$$

$$= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{tx^2}{2}} \, dx$$
MGF: Examples

We will now use the fact that
\[ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi a}} e^{-\frac{(x-b)^2}{2a}} \, dx = 1 \text{ for all } a > 0, b \]

We have completed the square

\[ m_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx = e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} \, dx \]
\[ = e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx = e^{\frac{t^2}{2}} \]

This is 1

MGF: Examples

4. The Gamma distribution (parameters \( \alpha, \lambda \))

\[ f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \]

The MGF of \( X \), \( m_X(t) \) is:

\[ m_X(t) = E\left[ e^{itX} \right] = \int_{-\infty}^{\infty} e^{itx} f(x) \, dx \]
\[ = \int_{0}^{\infty} e^{itx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \, dx \]
\[ = \int_{0}^{\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} \, dx \]
**MGF: Examples**

We use the fact

\[ \int_0^\infty b^a x^{\alpha-1} e^{-bx} dx = 1 \text{ for all } a > 0, b > 0 \]

\[
m_X(t) = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx
\]

\[
= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \int_0^\infty \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx = \left( \frac{\lambda}{\lambda-t} \right)^\alpha
\]

Equal to 1

The Chi-square distribution with degrees of freedom \( v = \nu / \alpha, \lambda = \nu / 2 \):

\[ m_X(t) = (1 - 2t)^{\nu / 2} \]

**MGF: Properties**

1. \( m_{X(0)} = 1 \)

\[ m_X(t) = E(e^{itX}), \text{ hence } m_X(0) = E(e^{0X}) = E(1) = 1 \]

**Note:** The MGFs of the following distributions satisfy the property \( m_{X(0)} = 1 \)

i) Binomial Dist'n \( m_X(t) = (e^{t}p + 1-p)^n \)

ii) Poisson Dist'n \( m_X(t) = e^{\lambda(e^{t}-1)} \)

iii) Exponential Dist'n \( m_X(t) = \left( \frac{\lambda}{\lambda-t} \right) \)

iv) Std Normal Dist'n \( m_X(t) = e^{\frac{t^2}{2}} \)

v) Gamma Dist'n \( m_X(t) = \left( \frac{\lambda}{\lambda-t} \right)^\alpha \)
2. We use the expansion of the exponential function:

\[ e^u = 1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} + \cdots + \frac{u^k}{k!} + \cdots \]

\[ m_X(t) = E(e^{tx}) \]

\[ = E\{1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \cdots + \frac{t^kx^k}{k!} + \cdots \} \]

\[ = 1 + tE[X] + \frac{t^2E[X^2]}{2!} + \frac{t^3E[X^3]}{3!} + \cdots + \frac{t^kE[X^k]}{k!} + \cdots \]

\[ = 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \frac{\mu_3 t^3}{3!} + \cdots + \frac{\mu_k t^k}{k!} + \cdots \]

MGF: Properties

3. \[ m_X^{(k)}(0) = \frac{d^k}{dt^k} m_X(t) \bigg|_{t=0} = \mu_k \]

Now

\[ m_X(t) = 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \frac{\mu_3 t^3}{3!} + \cdots + \frac{\mu_k t^k}{k!} + \cdots \]

\[ m_X'(t) = \mu_1 + \frac{\mu_2}{2!} t + \frac{\mu_3}{3!} t^2 + \cdots + \frac{\mu_k}{k!} t^{k-1} + \cdots \]

\[ = \mu_1 + \frac{\mu_2}{2!} t + \frac{\mu_3}{3!} t^2 + \cdots + \frac{\mu_k}{(k-1)!} t^{k-1} + \cdots \]

and \( m_X'(0) = \mu_1 \)

\[ m_X''(t) = \mu_2 + \frac{\mu_3}{2!} t + \frac{\mu_4}{2!} t^2 + \cdots + \frac{\mu_k}{(k-2)!} t^{k-2} + \cdots \]

and \( m_X''(0) = \mu_2 \)

continuing we find \( m_X^{(k)}(0) = \mu_k \)
MGF: Applying Property 3 – Binomial

Property 3 is very useful in determining the moments of a RV $X$.

Examples:

i) Binomial Dist'n $m_X(t) = (e^t p + 1 - p)^n$

\[ m_X'(t) = n(e^t p + 1 - p)^{n-1}(pe^t) \]

\[ m_X(0) = n(e^0 p + 1 - p)^{n-1}(pe^0) = np = \mu_1 = \mu \]

\[ m_X''(t) = np \left[ (n-1)(e^t p + 1 - p)^{n-2}(e^t p) e^t + (e^t p + 1 - p)^{n-1} e^t \right] \]

\[ = np e^t (e^t p + 1 - p)^{n-2} \left[ (n-1)(e^t p) + (e^t p + 1 - p) \right] \]

\[ = np e^t (e^t p + 1 - p)^{n-2} \left[ ne^t p + 1 - p \right] \]

\[ m_X'''(0) = np[np + 1 - p] = np[np + q] = n^2 p^2 + npq = \mu_2 \]

MGF: Applying Property 3 – Poisson

ii) Poisson Dist'n $m_X(t) = e^{\lambda(e^t - 1)}$

\[ m_X'(t) = e^{\lambda(e^t - 1)} \left[ \lambda e^t \right] = \lambda e^{\lambda(e^t - 1) + t} \]

\[ m_X''(t) = \lambda e^{\lambda(e^t - 1) + t} \left[ \lambda e^t + 1 \right] = \lambda^2 e^{\lambda(e^t - 1) + 2t} + \lambda e^{\lambda(e^t - 1) + t} \]

\[ m_X'''(t) = \lambda^2 e^{\lambda(e^t - 1) + 2t} \left[ \lambda e^t + 2 \right] + \lambda e^{\lambda(e^t - 1) + t} \left[ \lambda e^t + 1 \right] \]

\[ = \lambda^2 e^{\lambda(e^t - 1) + 2t} \left[ \lambda e^t + 3 \right] + \lambda e^{\lambda(e^t - 1) + t} \]

\[ = \lambda^3 e^{\lambda(e^t - 1) + 3t} + 3\lambda^2 e^{\lambda(e^t - 1) + 2t} + \lambda e^{\lambda(e^t - 1) + t} \]
MGF: Applying Property 3 – Poisson

To find the moments we set \( t = 0 \).

\[
\mu_1 = m'_X(0) = \lambda e^{\frac{t}{\lambda}} = \lambda \\
\mu_2 = m''_X(0) = \lambda^2 e^{\frac{2t}{\lambda}} + \lambda e^{\frac{t}{\lambda}} = \lambda^2 + \lambda \\
\mu_3 = m'''_X(0) = \lambda^3 e^{0} + 3\lambda^2 e^{0t} + \lambda e^{0} = \lambda^3 + 3\lambda^2 + \lambda
\]

MGF: Applying Property 3 – Exponential

iii) Exponential Dist'n \( m_X(t) = \left( \frac{\lambda}{\lambda - t} \right) \)

\[
m'_X(t) = \frac{d}{dt} \left( \frac{\lambda}{\lambda - t} \right) = \lambda \frac{d}{dt} \left( \frac{\lambda - t^{-1}}{\lambda - t} \right) = \lambda (-1)(\lambda - t)^{-2} (-1) = \lambda (\lambda - t)^{-2}
\]

\[
m''_X(t) = \lambda (-2)(\lambda - t)^{-3} (-1) = 2\lambda (\lambda - t)^{-3}
\]

\[
m'''_X(t) = 2\lambda (-3)(\lambda - t)^{-4} (-1) = 2(3)\lambda (\lambda - t)^{-4}
\]

\[
m^{(4)}_X(t) = 2(3)\lambda (-4)(\lambda - t)^{-5} (-1) = (4!)\lambda (\lambda - t)^{-5}
\]

\[
m^{(k)}_X(t) = (k!)\lambda (\lambda - t)^{-k-1}
\]
MGF: Applying Property 3 – Exponential

Thus,
\[ \mu_1 = \mu = m'_X(0) = \lambda (\lambda)^{-2} = \frac{1}{\lambda} \]
\[ \mu_2 = m''_X(0) = 2\lambda (\lambda)^{-3} = \frac{2}{\lambda^2} \]
\[ \mu_k = m^{(k)}_X(0) = (k!)\lambda (\lambda)^{-k-1} = \frac{k!}{\lambda^k} \]

We can calculate the following popular descriptive statistics:
- \( \sigma^2 = \mu^2 - \mu^2 = (2/\lambda^2) - (1/\lambda)^2 = (1/\lambda)^2 \)
- \( \gamma_1 = \mu^3 / \sigma^3 = (2/\lambda^3) / [(1/\lambda)^2]/2 = 2 \)
- \( \gamma_2 = \mu^4 / \sigma^4 - 3 = (9/\lambda^4) / [(1/\lambda)^4] - 3 = 6 \)

Note: the moments for the exponential distribution can be calculated in an alternative way. This is done by expanding \( m_X(t) \) in powers of \( t \) and equating the coefficients of \( t^k \) to the coefficients in:

\[ m_X(t) = 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \frac{\mu_3 t^3}{3!} + \cdots + \frac{\mu_k t^k}{k!} + \cdots \]

Equating the coefficients of \( t^k \) we get:

\[ \frac{\mu_k}{k!} = \frac{1}{\lambda^k} \quad \text{or} \quad \mu_k = \frac{k!}{\lambda^k} \]
MGF: Applying Property 3 – Normal

iv) Standard normal distribution \( m_X(t) = \exp(t^2/2) \)

We use the expansion of \( e^u \).

\[
e^u = 1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} + \cdots + \frac{u^k}{k!} + \cdots
\]

\[
e^u = 1 + \left( \frac{t^2}{2} \right) + \frac{\left( \frac{t^2}{2} \right)^2}{2!} + \frac{\left( \frac{t^2}{2} \right)^3}{3!} + \cdots + \frac{\left( \frac{t^2}{2} \right)^k}{k!} + \cdots
\]

\[
e^u = 1 + \frac{1}{2} t^2 + \frac{1}{2^2 2!} t^4 + \frac{1}{2^3 3!} t^6 + \cdots + \frac{1}{2^k k!} t^{2k} + \cdots
\]

We now equate the coefficients \( \mu \) in:

\[
m_X(t) = 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \frac{\mu_3 t^3}{3!} + \cdots + \frac{\mu_k t^k}{k!} + \cdots + \frac{\mu_{2k} t^{2k}}{(2k)!} + \cdots
\]

MGF: Applying Property 3 – Normal

If \( k \) is odd: \( \mu_k = 0 \).

For even \( 2k \):

\[
\frac{\mu_{2k}}{(2k)!} = \frac{1}{2^k k!}
\]

or \( \mu_{2k} = \frac{(2k)!}{2^k k!} \)

Thus \( \mu_1 = 0, \mu_2 = \frac{2!}{2} = 1, \mu_3 = 0, \mu_4 = \frac{4!}{2^2 (2!)} = 3 \)
The log of Moment Generating Functions

Let \( l_X(t) = \ln m_X(t) \) = the log of the MGF.

Then \( l_X(0) = \ln m_X(0) = \ln 1 = 0 \)

\[
l'_X(t) = \frac{1}{m_X(t)} m'_X(t) = \frac{m'_X(t)}{m_X(t)} \quad l'_X(0) = \frac{m'_X(0)}{m_X(0)} = \mu_1 = \mu
\]

\[
l''_X(t) = \frac{m''_X(t)m_X(t) - [m'_X(t)]^2}{[m_X(t)]^2}
\]

\[
l''_X(0) = \frac{m''_X(0)m_X(0) - [m'_X(0)]^2}{[m_X(0)]^2} = \mu_2 - [\mu_1]^2 = \sigma^2
\]

Thus \( l_X(t) = \ln m_X(t) \) is very useful for calculating the mean and variance of a random variable

1. \( l'_X(0) = \mu \)
2. \( l''_X(0) = \sigma^2 \)
Log of MGF: Examples – Binomial

1. The Binomial distribution (parameters $p$, $n$)

$$m_X(t) = \left(e^t p + 1 - p\right)^n = \left(e^t p + q\right)^n$$

$$l_X(t) = \ln m_X(t) = n \ln \left(e^t p + q\right)$$

$$l'_X(t) = n \frac{1}{e^t p + q} e^t p$$

$$\mu = l'_X(0) = n \frac{1}{p + q} p = np$$

$$l''_X(t) = n \frac{e^t p (e^t p + q) - e^t p (e^t p)}{(e^t p + q)^2}$$

$$\sigma^2 = l''_X(0) = n \frac{p (p + q) - p (p)}{(p + q)^2} = npq$$

Log of MGF: Examples – Poisson

2. The Poisson distribution (parameter $\lambda$)

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

$$l_X(t) = \ln m_X(t) = \lambda (e^t - 1)$$

$$l'_X(t) = \lambda e^t$$

$$\mu = l'_X(0) = \lambda$$

$$l''_X(t) = \lambda e^t$$

$$\sigma^2 = l''_X(0) = \lambda$$
Log of MGF: Examples – Exponential

3. The Exponential distribution (parameter $\lambda$)

\[
m_X(t) = \begin{cases} 
\frac{\lambda}{\lambda - t} & t < \lambda \\
\text{undefined} & t \geq \lambda 
\end{cases}
\]

\[
l_X(t) = \ln m_X(t) = \ln \lambda - \ln(\lambda - t) \quad \text{if} \quad t < \lambda
\]

\[
l'_X(t) = \frac{1}{\lambda - t} = (\lambda - t)^{-1}
\]

\[
l''_X(t) = -1(\lambda - t)^{-2}(-1) = \frac{1}{(\lambda - t)^2}
\]

Thus $\mu = l'_X(0) = \frac{1}{\lambda}$ and $\sigma^2 = l''_X(0) = \frac{1}{\lambda^2}$

Log of MGF: Examples – Normal

4. The Standard Normal distribution ($\mu = 0, \sigma = 1$)

\[
m_X(t) = e^{\frac{t^2}{2}}
\]

\[
l_X(t) = \ln m_X(t) = \frac{t^2}{2}
\]

\[
l'_X(t) = t, \quad l''_X(t) = 1
\]

Thus $\mu = l'_X(0) = 0$ and $\sigma^2 = l''_X(0) = 1$
Log of MGF: Examples – Gamma

5. The Gamma distribution (parameters $\alpha$, $\lambda$)

\[
m_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha
\]

\[
l_X(t) = \ln m_X(t) = \alpha \left[ \ln \lambda - \ln (\lambda - t) \right]
\]

\[
l'_X(t) = \alpha \left[ \frac{1}{\lambda - t} \right] = \frac{\alpha}{\lambda - t}
\]

\[
l''_X(t) = \alpha (-1)(\lambda - t)^{-2}(-1) = \frac{\alpha}{(\lambda - t)^2}
\]

Hence $\mu = l'_X(0) = \frac{\alpha}{\lambda}$ and $\sigma^2 = l''_X(0) = \frac{\alpha}{\lambda^2}$

Log of MGF: Examples – Chi-squared

6. The Chi-square distribution (degrees of freedom $\nu$)

\[
m_X(t) = (1-2t)^{-\nu/2}
\]

\[
l_X(t) = \ln m_X(t) = -\frac{\nu}{2} \ln (1-2t)
\]

\[
l'_X(t) = -\frac{\nu}{2} \frac{1}{1-2t}(-2) = \frac{\nu}{1-2t}
\]

\[
l''_X(t) = \nu(-1)(1-2t)^{-2}(-2) = \frac{2\nu}{(1-2t)^2}
\]

Hence $\mu = l'_X(0) = \nu$ and $\sigma^2 = l''_X(0) = 2\nu$
Characteristic functions

Definition: Characteristic Function
Let \( X \) denote a random variable. Then, the characteristic function of \( X \), \( \varphi_X(t) \) is defined by:

\[
\varphi_X(t) = E(e^{itx})
\]

Since \( e^{ix} = \cos(xt) + i\sin(xt) \) and \( \| e^{ix} \| \leq 1 \), then \( \varphi_X(t) \) is defined for all \( t \). Thus, the characteristic function always exists, but the MGF need not exist.

Relation to the MGF: \( \varphi_X(t) = m_X(t) = m_X(it) \)

Calculation of moments: 
\[
\frac{\partial^k \varphi_X(t)}{\partial t} \bigg|_{t=0} = i^k \mu_k
\]