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 $\lambda(A) = |3 - 7| = 4$.

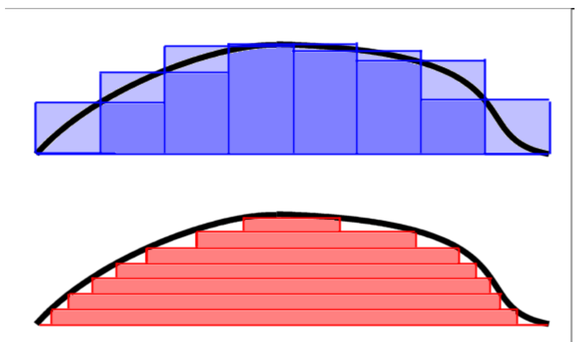
• The Lebesgue integral of f on $[a,b]$ is defined in terms of $\Sigma_i\, y_i\, \lambda(A_i)$, where $0 = y_1 \leq y_2 \leq ... \leq y_n$, $A_i = \{x : y_i \leq f(x) < y_{i+1}\}$, and $\lambda(A_i)$ is the Lebesgue measure of the set A_i.

• 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 $\{y_n\}$ and break the x-axis into the corresponding grid $\{A_n\}$ 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.



Taking expectations: Riemann vs Lebesgue

• 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 (Ω, Σ, μ) to $(\bar{R}, \bar{B})$, where $\bar{R} = R \cup \{-\infty, \infty\}$, $\bar{B} = \sigma(B \cup \{\{\infty\}, \{-\infty\}\})$.

Example: If $\Omega = R$ and μ 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 μ=P, in statistics, $\int X\, dP = EX = E[X]$ is called the *expectation* or *expected value* of X.

Expected Value

Consider our probability space (Ω, Σ, P) . Take an event (a set *A* of ω є Ω) and *X*, a RV, that assigns real numbers to each ω є *A*.

• If we take an observation from *A* without knowing which ω є *A* will be drawn, we may want to know what value of *X*(ω) we should *expect* to see.

• Each of the ω є *A* has been assigned a probability measure P[ω], which induces P[*x*]. Then, we use this to weight the values *X*(ω).

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

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

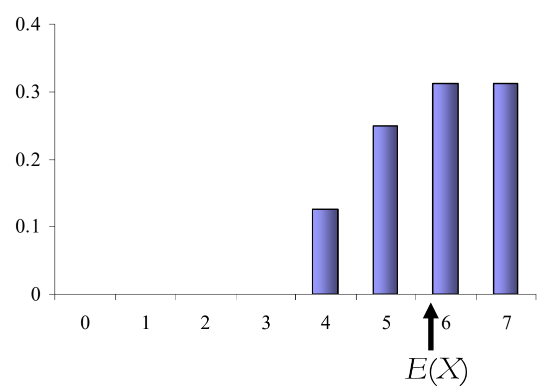
and if X is *continuous* with probability density function $f(x)$

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

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 = the number of successes in n independent repetitions of a Bernoulli trial. Find the expected value of X, $E(X)$.

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

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

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

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

Example: Solution

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

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

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

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

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

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

Example: Solution

$$= np\left[\frac{(n-1)!}{0!(n-1)!}p^0(1-p)^{n-1} + \frac{(n-1)!}{1!(n-2)!}p^1(1-p)^{n-2} + \right.$$

$$\left.\ldots + \frac{(n-1)!}{(n-2)!1!}p^{n-2}(1-p) + \frac{(n-1)!}{(n-1)!0!}p^{n-1}\right]$$

$$= np\left[\binom{n-1}{0}p^0(1-p)^{n-1} + \binom{n-1}{1}p^1(1-p)^{n-2} + \right.$$

$$\left.\ldots. + \binom{n-1}{n-2}p^{n-2}(1-p) + \binom{n-1}{n-1}p^{n-1}\right]$$

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

Example: Exponential Distribution

Let X have an exponential distribution with parameter λ. 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:

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{0}^{\infty} x\lambda e^{-\lambda x}dx$$

We will determine $\int x\lambda e^{-\lambda x}dx$

using integration by parts $\int u\,dv = uv - \int v\,du$

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 $\int x\lambda e^{-\lambda x}dx = -xe^{-\lambda x} + \int e^{-\lambda x}dx = -xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x}$

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x}dx = -xe^{-\lambda x}\Big|_0^\infty - \frac{1}{\lambda}e^{-\lambda x}\Big|_0^\infty$$

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

Summary: If X has an exponential distribution with parameter λ, then:

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

Example: The Uniform Distribution

Suppose X has a uniform distribution from a to b.
Then:

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

The *expected value* of X is:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_a^b x\frac{1}{b-a}dx$$

$$= \left[\frac{1}{b-a}\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

Example: The Normal Distribution

Suppose X has a Normal distribution with parameters μ and σ.
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:

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

Example: The Normal Distribution

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

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

Now
$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1 \text{ and } \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz = 0$$

The second integral is an example of an odd function. Recall that an odd function gives:

$f(-x) = -f(x)$. Then, $\int_{-a}^{a} f(x) dx = 0$.

Thus $E(X) = \mu$

Example: The Gamma Distribution

Suppose X has a Gamma distribution with parameters α and λ.

Then:

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

Note:

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

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

Example: The Gamma Distribution

The *expected value* of X is:

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

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

This is now equal to 1.

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

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

Example: The Gamma Distribution

Thus, if X has a Gamma (α, λ) distribution, the **expected value** of X is:

$$E(X) = \alpha / \lambda$$

Special Cases: (α, λ) distribution then the **expected value** of X is:

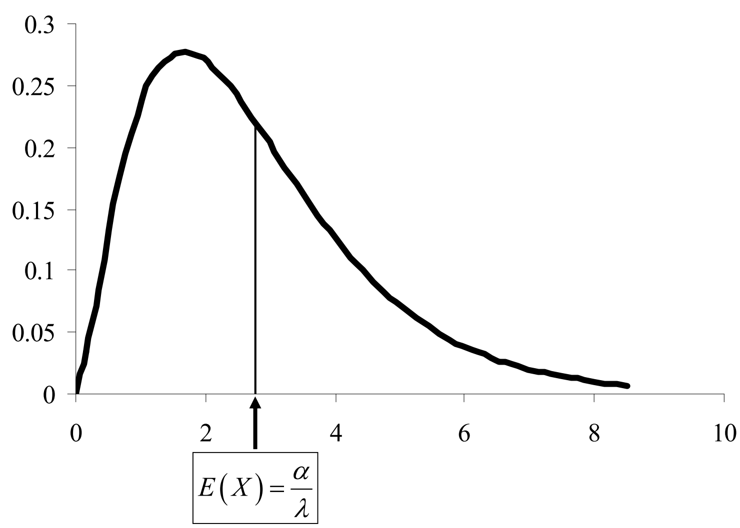
1. **Exponential** (λ) **distribution:** $\alpha = 1$, λ arbitrary

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

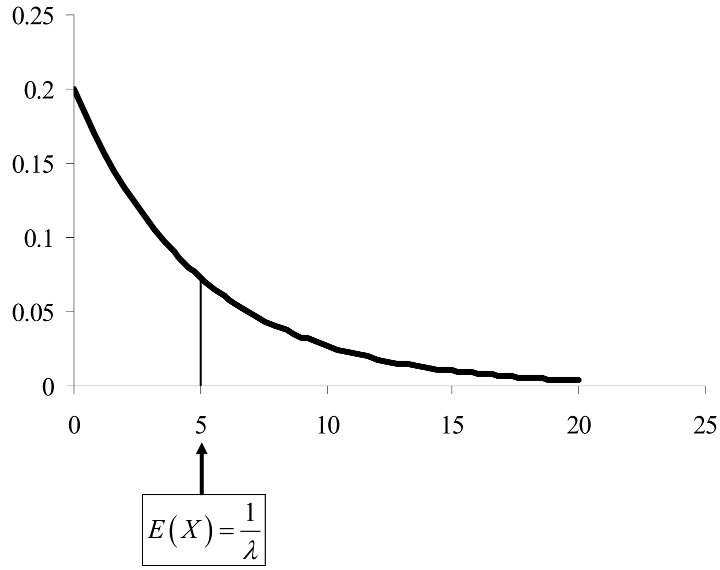
2. **Chi-square** (ν) **distribution:** $\alpha = \nu/2$, $\lambda = 1/2$.

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

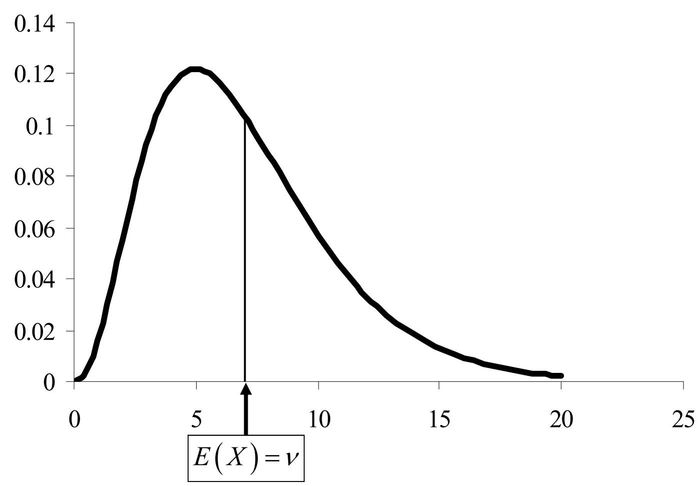
Example: The Gamma Distribution



Example: The Gamma Distribution - Exponential



Example: The Gamma Distribution - Chi-square



Expectation of a function of a RV

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

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

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

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

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

Expectation of a function of a RV

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

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

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

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

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

= 1/3 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 $\boldsymbol{X}$ is the unique number m that satisfies the following inequalities:

$$P(\boldsymbol{X} \leq m) \geq ½ \qquad \text{and} \qquad P(\boldsymbol{X} \geq m) \geq ½.$$

For a continuous distribution, we have that m solves:

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

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

$$E(|\mathbf{X} - c|)$$

is the median of the probability distribution of the random variable $\boldsymbol{X}$.

Example I: Median of the Exponential Distribution

Let X have an exponential distribution with parameter λ. 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:

$$\int_m^\infty f_X(x)dx = 1/2$$

$$\int_m^\infty \lambda\, e^{-\lambda x} dx = \lambda \int_m^\infty e^{-\lambda x} dx = \,-e^{-\lambda x}\,|_m^\infty = e^{-\lambda m} = 1/2$$

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

Example II: Median of the Pareto Distribution

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

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

The median m solves the following integral of X: $\int_m^\infty f_X(x)dx = 1/2$

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

$$= -x_s^\alpha\, x^{-\alpha} + C\,|_m^\infty = x_s^\alpha\, m^{-\alpha} = 1/2 \quad \Rightarrow\ m = x_s\, 2^{1/\alpha}$$

<u>Note</u>: 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:

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

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

• The central moments describe how the probability distribution is distributed about the center of gravity, μ.

Moments of a Random Variable – 1st and 2nd

The first central moments is given by:

$$\mu_1^0 = E\left[X - \mu\right]$$

The second central moment depends on the *spread* of the probability distribution of X about μ. It is called the variance of X and is denoted by the symbol var(X).

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

$$\sqrt{\mu_2^0} = \sqrt{E\left[(X - \mu)^2\right]}$$ is called the *standard deviation* of X and is denoted by the symbol σ.

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

Moments of a Random Variable – Skewness

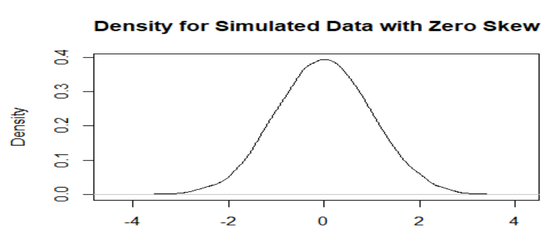
- The third central moment: $\mu_3^0 = E[(X - \mu)^3]$

μ_3^0 contains information about the *skewness* of a distribution.

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

- Distribution according to skewness:

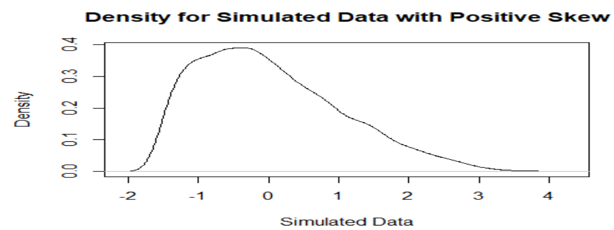
1) Symmetric distribution



$\mu_3^0 = 0, \gamma_1 = 0$

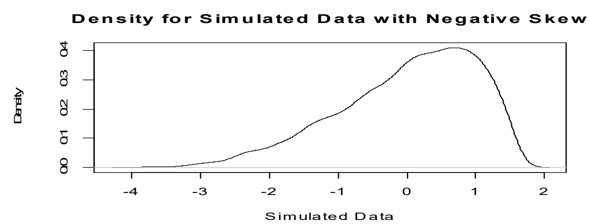
Moments of a Random Variable – Skewness

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 (positve skewness loved!).

Moments of a Random Variable – Kurtosis

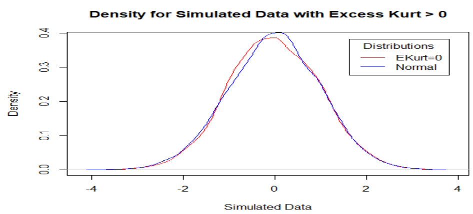
• The fourth central moment: $\mu_4^0 = E[(X-\mu)^4]$

μ_4^0 is a measure of the *shape* of a distribution. The property of shape measured by this moment is called *kurtosis,* usually estimated by $\kappa = \frac{\mu_4^0}{\sigma^4}$.

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

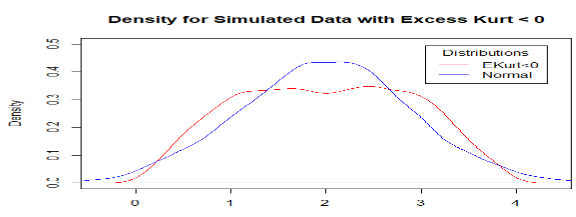
• Distributions:

1) Mesokurtic distribution ($\gamma_2 = 0$ or κ=3, like the normal distribution)

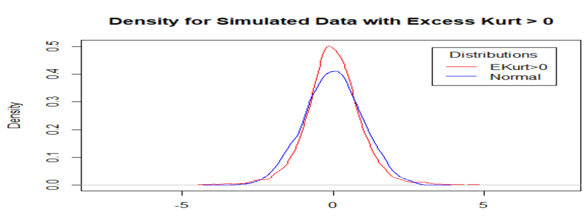


Moments of a Random Variable – Kurtosis

2) Platykurtic distribution ($\gamma_2 < 0$, μ_4^0 small in size)

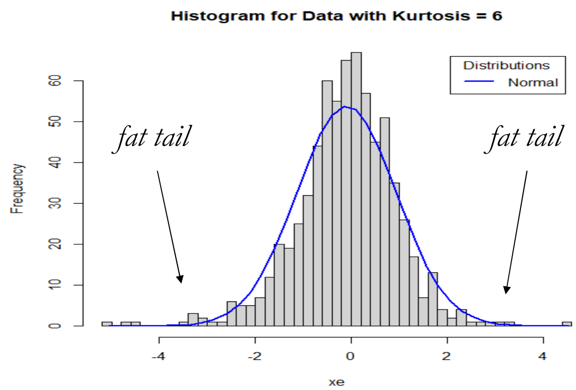


3) Leptokurtic distribution ($\gamma_2 > 0$, μ_4^0 large in size, usual shape)



Moments of a Random Variable – Kurtosis

• Typical financial returns series has $\gamma_2 > 0$. Below, I simulate a series with μ=0, σ=1, γ_1=0 & kurtosis = 6 (γ_2=3), overlaid with a standard normal distribution. Fat tails are seen on both sides of the distribution.



37

Moments of a Random Variable

Example: The uniform distribution from 0 to 1

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

Finding the moments

$$\mu_k = \int_{-\infty}^{\infty} x^k f(x)\,dx = \int_0^1 x^k 1 dx = \left[\frac{x^{k+1}}{k+1}\right]_0^1 = \frac{1}{k+1}$$

Finding the central moments:

$$\mu_k^0 = \int_{-\infty}^{\infty} (x-\mu)^k f(x)\,dx = \int_0^1 \left(x - \tfrac{1}{2}\right)^k 1dx$$

Moments of a Random Variable

Example (continuation): Finding the central moments (continuation)

$$\mu_k^0 = \int_{-\infty}^{\infty} (x-\mu)^k f(x)\,dx = \int_0^1 \left(x-\tfrac{1}{2}\right)^k 1\,dx$$

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

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

$$= \frac{1-(-1)^{k+1}}{2^{k+1}(k+1)} = \begin{cases} \dfrac{1}{2^k(k+1)} & \text{if } k \text{ even} \\ 0 & \text{if } k \text{ odd} \end{cases}$$

Moments of a Random Variable

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

Thus, $\quad \operatorname{var}(X) = \mu_2^0 = \dfrac{1}{12}$

The standard deviation $\quad \sigma = \sqrt{\operatorname{var}(X)} = \sqrt{\mu_2^0} = \dfrac{1}{\sqrt{12}}$

The measure of skewness: $\quad \gamma_1 = \dfrac{\mu_3^0}{\sigma^3} = 0$

The measure of kurtosis: $\quad \gamma_2 = \dfrac{\mu_4^0}{\sigma^4} - 3 = \dfrac{1/80}{\left(\frac{1}{12}\right)^2} - 3 = -1.2$

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* – $(1/n) \Sigma_i \, |x_i - \textit{central measure}(X)|$
- The *median absolute deviation* (MAD) – 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.

Review – Rules for Expectations

• We will derive the rules for the continuous case, with X has a pdf $f(x)$. Proof are similar for the discrete case. That is, we define E[X] as

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

- **Rule 1**. $E[c] = c$, where c is a constant.

Proof: $g(x) = c$

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

- **Rule 2**. $E[c + dX] = c + d\, E[X]$, where c & d are constants.

Proof: $g(x) = c + dX$

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

$$= c \int_{-\infty}^{\infty} f(x) dx + d \int_{-\infty}^{\infty} x\, f(x) dx$$

$$= c + d\, E[X]$$

Review – Rules for Expectations

- **Rule 3**. Var[X] = $\mu_2^0 = E[(X-\mu)^2]$ = $E[X^2]$ - $[E(X)]^2$ = $\mu_2 - \mu_1^2$

Proof: $g(x) = (x-\mu)^2$

Var[X] = $E[(X-\mu)^2]$ = $\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$

$= \int_{-\infty}^{\infty}(x^2 - 2x\mu + \mu^2)f(x)dx$

$= \int_{-\infty}^{\infty} x^2 f(x)dx - \int_{-\infty}^{\infty} 2x\mu f(x)dx + \int_{-\infty}^{\infty} \mu^2 f(x)dx$

$= \int_{-\infty}^{\infty} x^2 f(x)dx - 2\mu \int_{-\infty}^{\infty} x f(x)dx + \mu^2 \int_{-\infty}^{\infty} f(x)dx$

$= E[X^2] - 2\,\mu\, E(X) + \mu^2 = \mu_2 - \mu_1^2$

Rules for Expectations

- **Rule 4**. $Var(aX+b) = a^2\, Var(X)$

Proof:

$$\mu_{aX+b} = E[aX+b] = aE[X] + b = a\mu + b$$

$$\operatorname{var}(aX+b) = E\left[\left(aX+b-\mu_{aX+b}\right)^2\right]$$

$$= E\left[\left(aX+b-[a\mu+b]\right)^2\right]$$

$$= E\left[a^2(X-\mu)^2\right]$$

$$= a^2 E\left[(X-\mu)^2\right] = a^2 \operatorname{var}(X)$$

Rules for Expectations for Vectors & Matrices

• Let $\boldsymbol{Z}$ be a random vector of k random variables: $Z_1, Z_2, \ldots, Z_k$. We have a similar definition for $\boldsymbol{W}$

- Expected value of $\boldsymbol{Z}$:

$$\mathrm{E}[\boldsymbol{Z}] = \begin{bmatrix} E[Z_1] \\ \vdots \\ E[Z_k] \end{bmatrix}$$

- Expected value of a linear function of random vectors. Let a & b be non-random scalars. Then:

$$\mathrm{E}[a\boldsymbol{Z} + b\boldsymbol{W}] = a\ \mathrm{E}[\boldsymbol{Z}] + b\ \mathrm{E}[\boldsymbol{W}]$$

- Variance of $\boldsymbol{Z}$: $\quad \mathrm{Var}[\boldsymbol{Z}] = \mathrm{E}[\boldsymbol{Z}\ \boldsymbol{Z}'] - \mathrm{E}[\boldsymbol{Z}]\ \mathrm{E}[\boldsymbol{Z}]' \qquad (k \times k)$

Rules for Expectations for Vectors & Matrices

- Variance of linear function of $\boldsymbol{Z}$:

$$\mathrm{Var}[a + b\boldsymbol{Z}] = b^2\ \mathrm{Var}[\boldsymbol{Z}]$$

- Variance of linear function of $\boldsymbol{Z}$, with a comformable non-random matrix **A**:

$$\mathrm{Var}[\mathbf{A}\ \boldsymbol{Z}] = \mathbf{A}\ \mathrm{Var}[\boldsymbol{Z}]\ \mathbf{A}'$$

- Expected value of a quadratic form $\boldsymbol{Z}'\ \mathbf{A}\ \boldsymbol{Z}$:

$$\mathrm{E}[\boldsymbol{Z}'\ \mathbf{A}\ \boldsymbol{Z}] = \mathrm{E}[\boldsymbol{Z}]'\ \mathbf{A}\ \mathrm{E}[\boldsymbol{Z}] - \mathrm{trace}(\mathbf{A}\ \mathrm{Var}[\boldsymbol{Z}]) \qquad (1 \times 1)$$

Derivation: Use properties of trace and expectations:

$$\begin{aligned}\mathrm{E}[\boldsymbol{Z}'\ \mathbf{A}\ \boldsymbol{Z}] &= \mathrm{E}[\mathrm{tr}(\mathbf{A}\boldsymbol{Z}\boldsymbol{Z}')] = \mathrm{tr}(\mathrm{E}[(\mathbf{A}\boldsymbol{Z}\boldsymbol{Z}')]) \\ &= \mathrm{tr}(\mathbf{A}\ \mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}']) = \mathrm{tr}(\mathbf{A}\ (\mathrm{Var}[\boldsymbol{Z}] + \mathrm{E}[\boldsymbol{Z}]\ \mathrm{E}[\boldsymbol{Z}]') \\ &= \mathrm{tr}(\mathbf{A}\ (\mathrm{Var}[\boldsymbol{Z}]) + \mathrm{tr}(\mathrm{E}[\boldsymbol{Z}]'\ \mathbf{A}\ \mathrm{E}[\boldsymbol{Z}]) \\ &= \mathrm{tr}(\mathbf{A}\ (\mathrm{Var}[\boldsymbol{Z}]) + \mathrm{E}[\boldsymbol{Z}]'\ \mathbf{A}\ \mathrm{E}[\boldsymbol{Z}]\end{aligned}$$

Moment generating functions

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

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

Definition: Moment Generating Function (MGF)

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

$$m_X\left(t\right)=E\left[e^{tX}\right]=\begin{cases}\sum_{x} e^{tx}p\left(x\right) & \text{if } X \text{ is discrete}\\ \int_{-\infty}^{\infty} e^{tx}f\left(x\right)dx & \text{if } X \text{ is continuous}\end{cases}$$

MGF: Examples

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

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

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

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

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

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

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

MGF: Examples

2. The Poisson distribution (parameter λ)

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

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

$$m_X(t) = E\left[e^{tX}\right] = \sum_x e^{tx} p(x) \quad = \sum_{x=0}^{n} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda}$$

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

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

MGF: Examples

3. The Exponential distribution (parameter λ)

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

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

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

$$= \int_{0}^{\infty} \lambda e^{(t-\lambda)x}\,dx = \left[\lambda \frac{e^{(t-\lambda)x}}{t-\lambda}\right]_0^{\infty}$$

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

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:

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

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

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

MGF: Examples

We will now use the fact that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}a} e^{-\frac{(x-b)^2}{2a^2}} dx = 1 \quad \text{for all } a > 0, b$$

We have completed the square

$$m_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2-2tx}{2}} dx = e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2-2tx+t^2}{2}} dx$$

$$= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx = e^{\frac{t^2}{2}}$$

This is 1

MGF: Examples

4. The Gamma distribution (parameters α, λ)

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

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

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

$$= \int_0^{\infty} e^{tx} \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$$

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

MGF: Examples

We use the fact

$$\int_0^\infty \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} dx = 1 \quad \text{for all } a > 0, b > 0$$

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

$$= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \underbrace{\int_0^\infty \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx}_{\text{Equal to 1}} = \left(\frac{\lambda}{\lambda-t}\right)^\alpha$$

The Chi-square distribution with degrees of freedom ν ($\alpha = \nu/2$, $\lambda = 1/2$):

$$m_X(t) = (1-2t)^{-\frac{\nu}{2}}$$

MGF: Properties

1. $m_X(0) = 1$

$$m_X(t) = E\left(e^{tX}\right), \quad \text{hence } m_X(0) = E\left(e^{0 \cdot X}\right) = E(1) = 1$$

Note: The MGFs of the following distributions satisfy the property $m_X(0) = 1$

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

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

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

MGF: Properties

2. $$\mu_X(t) = 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \frac{\mu_3 t^3}{3!} + \cdots + \frac{\mu_k t^k}{k!} + \cdots$$

We use the expansion of the exponential function:

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

$$m_X(t) = E\left(e^{tX}\right)$$

$$= E\{1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots + \frac{t^kX^k}{k!} + \cdots\}$$

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

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

MGF: Properties

3. $$m_X^{(k)}(0) = \frac{d^k}{dt^k} m_X(t)\Big|_{t=0} = \mu_k$$

Now

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

$$m'_X(t) = \mu_1 + \frac{\mu_2}{2!}2t + \frac{\mu_3}{3!}3t^2 + \cdots + \frac{\mu_k t^k}{k!}(k-1)t^{k-1} + \cdots$$

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

and $m'_X(0) = \mu_1$

$$m''_X(t) = \mu_2 + \frac{\mu_3}{1!}t + \frac{\mu_4}{2!}t^2 + \cdots + \frac{\mu_k t^k}{(k-2)!}t^{k-2} + \cdots$$

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

$$m'_X(t) = n\left(e^t p + 1 - p\right)^{n-1}\left(pe^t\right)$$

$$m'_X(0) = n\left(e^0 p + 1 - p\right)^{n-1}\left(pe^0\right) = np = \mu_1 = \mu$$

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

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

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

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

MGF: Applying Property 3 – Poisson

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

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

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

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

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

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

MGF: Applying Property 3 – Poisson

To find the moments we set $t = 0$.

$$\mu_1 = m'_X(0) = \lambda e^{\lambda(e^0-1)+0} = \lambda$$

$$\mu_2 = m''_X(0) = \lambda^2 e^{\lambda(e^0-1)+0} + \lambda e^{\lambda(e^0-1)+0} = \lambda^2 + \lambda$$

$$\mu_3 = m'''_X(0) = \lambda^3 e^0 + 3\lambda^2 e^{0t} + \lambda e^0 = \lambda^3 + 3\lambda^2 + \lambda$$

MGF: Applying Property 3 – Exponential

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

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

$$= \lambda(-1)(\lambda - t)^{-2}(-1) = \lambda(\lambda - t)^{-2}$$

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

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

$$m_X^{(4)}(t) = 2(3)\lambda(-4)(\lambda - t)^{-5}(-1) = (4!)\lambda(\lambda - t)^{-5}$$

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

MGF: Applying Property 3 – Exponential

Thus,

$$\mu_1 = \mu = m'_X(0) = \lambda(\lambda)^{-2} = \frac{1}{\lambda}$$

$$\mu_2 = m''_X(0) = 2\lambda(\lambda)^{-3} = \frac{2}{\lambda^2}$$

$$\mu_k = m_X^{(k)}(0) = (k!)\lambda(\lambda)^{-k-1} = \frac{k!}{\lambda^k}$$

We can calculate the following popular descriptive statistics:

- $\sigma^2 = \mu^0_2 = \mu_2 - \mu^2 = (2/\lambda^2) - (1/\lambda)^2 = (1/\lambda)^2$
- $\gamma_1 = \mu^0_3/\sigma^3 = (2/\lambda^3) / [(1/\lambda)^2]^{3/2} = 2$
- $\gamma_2 = \mu^0_4/\sigma^4 - 3 = (9/\lambda^4) / [(1/\lambda)^4] - 3 = 6$

MGF: Applying Property 3 – Exponential

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:

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

$$m_X(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1 - t/\lambda} = \frac{1}{1-u} = 1 + u + u^2 + u^3 + \cdots$$

$$= 1 + \frac{t}{\lambda} + \frac{t^2}{\lambda^2} + \frac{t^3}{\lambda^3} + \cdots$$

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.

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

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

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

We now equate the coefficients t^k in:

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

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

$$l'_X(t) = \frac{1}{m_X(t)} m'_X(t) = \frac{m'_X(t)}{m_X(t)} \qquad l'_X(0) = \frac{m'_X(0)}{m_X(0)} = \mu_1 = \mu$$

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

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

The Log of Moment Generating Functions

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*)

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

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

$$l'_X(t) = n \frac{1}{e^t p + q} e^t p \qquad \mu = l'_X(0) = n \frac{1}{p+q} p = np$$

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

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

Log of MGF: Examples – Poisson

2. The Poisson distribution (parameter λ)

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

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

$$l'_X(t) = \lambda e^t \qquad \mu = l'_X(0) = \lambda$$

$$l''_X(t) = \lambda e^t \qquad \sigma^2 = l''_X(0) = \lambda$$

Log of MGF: Examples – Exponential

3. The Exponential distribution (parameter λ)

$$m_X(t) = \begin{cases} \dfrac{\lambda}{\lambda - t} & t < \lambda \\ \text{undefined} & t \geq \lambda \end{cases}$$

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

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

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

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

Log of MGF: Examples – Normal

4. The Standard Normal distribution (μ = 0, σ = 1)

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

$$l_X(t) = \ln m_X(t) = \tfrac{t^2}{2}$$

$$l'_X(t) = t, \quad l''_X(t) = 1$$

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

Log of MGF: Examples – Gamma

5. The Gamma distribution (parameters α, λ)

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

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

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

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

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 ν)

$$m_X(t) = (1 - 2t)^{-\frac{\nu}{2}}$$

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

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

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

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 $e^{itx} = \cos(xt) + i\sin(xt)$ and $\|e^{itx}\| \leq 1$, 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: $\varphi_X(t) = m_{iX}(t) = m_X(it)$

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