ISEG - Lisbon School of Economics and Management
Statistics I
(Date of this version: 6/11/2017)
Handout 3 (Part 1) – Expected values and parameters

Issues covered:

3.1 Expected values of a random variables

In order to obtain a measure of the centre of a probability distribution, we introduce the notion of the expected value (or mean or expectation) of a random variable.

Example: Consider the following example: A review of textbooks in a segment of the business area found that 81% of all pages of texts were error free, 17% of all pages contained one error, and the remaining 2% contained two errors. We use the random variable X to denote the number of errors on a page chosen at random from one of these books, with possible values of 0, 1, and 2, and the probability distribution function $f_X(0) = 0.81$, $f_X(1) = 0.17$; $f_X(2) = 0.02$. We could consider using the simple average of the values as the central location of a random variable. In this example the possible numbers of errors on a page are 0, 1, and 2. Their average is, then, $(0 + 1 + 2)/3 = 1$ error. However, a moment of reflection reveals that that this is an absurd measure of central location. In calculating this average, we paid no attention to the fact that 81% of all pages contain no errors, 17% contain one error, while only 2% contain two errors. In order to obtain a sensible measure of central location, we should weight the various possible outcomes by the probabilities of their occurrence which yields

$$0 \times 0.81 + 1 \times 0.17 + 2 \times 0.02 = 0.21.$$

Expected values (or mean or expectation) of a discrete random variables

Let X be a a discrete random variable and let D_X be the set of discontinuity points of the cumulative distribution function X. For generality let us assume that the number of elements of D_X is countably infinite, that this $D_X = \{x_1, x_2, ...\}$. The probability function of X is given by

$$f_X(x) = \begin{cases} P(X = x) & , x \in D_X \\ 0 & , x \notin D_X \end{cases}$$

Expected Value of a discrete random variable: The expected value of a random variable, denoted as $E(X)$ or μ_X, also known as its population mean, is the weighted average of its possible values, the weights being the probabilities attached to the values

$$\mu_X = E(X) = \sum_{x \in D_X} x \times f_X(x) = \sum_{i=1}^{\infty} x_i \times f_X(x_i).$$

provided that $\sum_{x \in D_X} |x| \times f_X(x) = \sum_{i=1}^{\infty} |x_i| \times f_X(x_i) < +\infty$.

Remarks:

1. If the number of elements of D_x is finite that this $D_X = \{x_1, x_2, ..., x_k\}$ where k is a finite integer, then $\sum_{x \in D_X} |x| \times f_X(x) = \sum_{i=1}^{k} |x_i| \times f_X(x_i)$ and the condition $\sum_{i=1}^{k} |x_i| \times f_X(x_i) < +\infty$ is always satisfied.

2. Note that μ_X can take values that are not in D_X.

Exercise: Let X be Bernoulli random variable with $P(X = 1) = p$, where $p \in (0, 1)$. Compute $E(X)$.

Exercise: Let X be a discrete random variable with probability function given by $f_X(x) = 1/3$, $x = -1$, 0, 1. Compute $E(X)$.
Exercise: Suppose that X is a random variable that takes values $x = 1, 2, 3, ...$ and that its probability function is given by

$$f_X(x) = \frac{6}{\pi^2 x^2}, x = 1, 2, ...$$

where $\pi = 3.14159....$ Show that $f_X(x)$ is a probability function, but $E(X)$ does not exist (**Hints:**.$\sum_{x=1}^{\infty} x^{-2} = \pi^2/6$, $\sum_{x=1}^{\infty} x^{-1} = +\infty$.)

Expected values (or mean or expectation) of a continuous random variable: If X is a continuous random variable and $f_X(x)$ is its probability density function at x, the expected value of X is

$$\mu_X = E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

provided that $\int_{-\infty}^{+\infty} |x| f_X(x) dx < \infty$.

Remark: Thus, the mean can be thought of as the centre of the distribution and, as such, it describes its location. Consequently, the mean is considered as a *measure of location*.

Exercise: Suppose $X \sim U(a, b)$, that is, X is an uniform random variable in the set (a, b), where $a < b$. Compute $E(X)$.
Exercise: Suppose that the random variable X has probability density function given by

$$f_X(x) = \begin{cases} 3x^{-4} & , \text{ for } x > 1 \\ 0 & , \text{ otherwise} \end{cases}.$$

Compute the expected value of X.
Exercise: Suppose that the random variable X has probability density function given by

$$f_X(x) = \begin{cases} x & , \text{ for } x \in (0, 1) \\ 1/2 & , \text{ for } x \in (1, 2) \\ 0 & , \text{ otherwise} \end{cases}.$$

Compute the expected value of X.

3.2 Expected values of functions of random variables

Expected value of a function of a discrete random variable: If X is a discrete random variable and $f_X(x)$ is the value of its probability function at x, the expected value of $g(X)$ is

$$E[g(X)] = \sum_{x \in D_X} g(x) \times f_X(x) = \sum_{i=1}^{\infty} g(x_i) \times f_X(x_i).$$

provided that $\sum_{x \in D_X} |g(x)| \times f_X(x) = \sum_{i=1}^{\infty} |g(x_i)| \times f_X(x_i) < +\infty$.

Expected value of a function of a continuous random variable: If X is a continuous random variable and $f_X(x)$ is its probability density function at x, the expected value of $g(X)$ is

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

provided that $\int_{-\infty}^{+\infty} |g(x)| f_X(x) dx < \infty$.

Remarks:

- The existence of $E(X)$ does not imply the existence of $E(g(X))$ and the inverse is also true.
- $E(g(X))$ can be calculated using the above definition or finding the distribution of $Y = g(X)$ and computing directly $E(Y)$.

Example: Let X be a discrete random variable with probability function given by $f_X(x) = 1/3$, $x = -1$, 0, 1 and $Y = g(X) = X^2$. We can compute $E(X^2)$ if the following two ways:

1. $E(X^2) = (-1)^2 f_X(-1) + (0)^2 f_X(0) + (1)^2 f_X(1) = 1 \times 1/3 + 0 \times 1/3 + 1 \times 1/3 = 2/3.$
2. Note that $f_Y(0) = P(Y = 0) = P(X = 0) = 1/3$. $f_Y(1) = P(Y = 1) = P(X = -1$ or $X = 1) = P(X = -1) + P(X = 1) = 1/3 + 1/3 = 2/3$. Thus $E(X^2) = E(Y) = 0 f_Y(0) + 1 f_Y(1) = 0 \times 1/3 + 2 \times 1/3 = 2/3$.

Example: Suppose that the random variable X has probability density function given by

$$f_X(x) = \begin{cases} 3x^{-4} & , \text{ for } x > 1 \\ 0 & , \text{ otherwise} \end{cases}.$$

and cumulative distribution function

$$F_X(x) = \begin{cases} 1 - x^{-3} & , \text{ for } x > 1 \\ 0 & , \text{ otherwise} \end{cases}$$

Compute $E(X^2)$.
Note that

$$\begin{aligned} E\left[X^2\right] &= \int_1^{+\infty} x^2 f_X(x)\, dx \\ &= \int_1^{+\infty} 3x^{-2} dx = 3 \end{aligned}$$

Alternatively

$$\begin{aligned} F_Y(y) &= P\left(X^2 \le y\right) \\ &= P\left(-\sqrt{y} \le X \le \sqrt{y}\right) \\ &= F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right) \\ &= F_X\left(\sqrt{y}\right) \end{aligned}$$

as $F_X\left(-\sqrt{y}\right) = 0$ Therefore

$$\begin{aligned} F_Y(y) &= \begin{cases} 1 - y^{-3/2} & , \text{ for } \sqrt{y} > 1 \\ 0 & , \text{ otherwise} \end{cases} \\ &= \begin{cases} 1 - y^{-3/2} & , \text{ for } y > 1 \\ 0 & , \text{ otherwise} \end{cases} \end{aligned}$$

Hence

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} \frac{3}{2} y^{-5/2} & , \text{ for } y > 1 \\ 0 & , \text{ otherwise} \end{cases}.$$

Consequently

$$\begin{aligned} E\left[X^2\right] &= E[Y] = \int_1^{+\infty} y f_Y(y)\, dy \\ &= \int_1^{+\infty} \frac{3}{2} y^{-3/2} dy \\ &= 3 \end{aligned}$$

3.3 Properties of expected values

The expected values satisfy the following properties:

1. $E(a + bX) = a + bE(X)$, where a and b are constants.
2. $E(X - \mu_X) = E(X) - \mu_X = 0$.
3. If a is a constant, $E(a) = a$.
4. If b is a constant, $E(bg(X)) = bE(g(X))$.
5. Given n functions $u_i(X)$ $i = 1, ..., n$ and , $E\left[\sum_{i=1}^n u_i(X)\right] = \sum_{i=1}^n E\left[u_i(X)\right]$.

3.4 Moments of a random variable (or of its distribution)

Objective: To characterize a random variable through a small number of indicators that describe the most significant aspects of its distribution.

Moments of a discrete random variable: The r^{th} moment of a discrete random variable (or its distribution), denoted as μ_r', is the expected value of X^r

$$\mu_r' = E(X^r) = \sum_{x \in D_X} x^r \times f_X(x) = \sum_{i=1}^{\infty} x_i^r \times f_X(x_i), \text{ for } r = 1, 2, ...$$

provided that $\sum_{x \in D_X} |x|^r \times f_X(x) = \sum_{i=1}^{\infty} |x_i|^r \times f_X(x_i) < +\infty$.

Moments of a continuous random variable: The r^{th} moment of a continuous random variable (or its distribution), denoted as μ'_r, is the expected value of X^r :

$$\mu'_r = E(X^r) = \int_{-\infty}^{+\infty} x^r f_X(x)dx$$

provided that $\int_{-\infty}^{+\infty} |x|^r f_X(x)dx < \infty$.

Most important moments:

1. The mean or expected value or expectation $\mu_X = \mu'_1 = E(X)$
2. The second moments $\mu'_2 = E(X^2)$.

3.5 Central moments of a random variable (or of its distribution)

The r^{th} central moment of the random variable (or its distribution) is also known as the r^{th} moment of a random variable about its mean (or its distribution).

Central moments of a discrete random variable: The r^{th} central moment of a discrete random variable (or its distribution), denoted as μ_r, is the expected value of $(X - \mu_X)^r$

$$\mu_r = E\left[(X - \mu_X)^r\right] = \sum_{x \in D_X} (x - \mu_X)^r \times f_X(x) = \sum_{i=1}^{\infty} (x_i - \mu_X)^r \times f_X(x_i), \text{ for } r = 1, 2, ...$$

provided that $\sum_{x \in D_X} |x - \mu_X|^r \times f_X(x) = \sum_{i=1}^{\infty} |x_i - \mu_X|^r \times f_X(x_i) < +\infty$.

Central moments of a continuous random variable: The r^{th} central moment of a continuous random variable (or its distribution), denoted as μ_r, is the expected value of $(X - \mu_X)^r$:

$$\mu_r = E\left[(X - \mu_X)^r\right] = \int_{-\infty}^{+\infty} (x - \mu_X)^r f_X(x)dx$$

provided that $\int_{-\infty}^{+\infty} |x - \mu_X|^r f_X(x)dx < \infty$.

Remarks:

1. μ_1 is of no interest because is it zero when it exists.
2. μ_2 is an important measure and is called variance.
3. μ_3 and μ_4 are also important.

3.6 The variance of a random variable

The second central moment about the mean of a random variable (μ_2), also called variance, is an indicator of the dispersion of the values of X about the mean.

The variance of a discrete random variable (or its distribution):

$$Var(X) = \sigma_X^2 = \mu_2 = E\left[(X - \mu_X)^2\right] = \sum_{x \in D_X} (x - \mu_X)^2 \times f_X(x) = \sum_{i=1}^{\infty} (x_i - \mu_X)^2 \times f_X(x_i),$$

provided that $Var(X) < +\infty$.

The variance of a continuous random variable (or its distribution):

$$Var(X) = \sigma_X^2 = \mu_2 = E\left[(X - \mu_X)^2\right] = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x)dx,$$

provided that $Var(X) < +\infty$.

Remark: We can show that if $\mu_2' = E(X^2)$ exists, then both μ_X and σ_X^2 exist.

Properties of the Variance:

1. $Var(X) \geq 0$.
2. $\sigma_X^2 = Var(X) = E(X^2) - \mu_X^2$.
3. If c is a constant, $Var(c) = 0$.
4. If a and b are constants, $Var(a + bX) = b^2 Var(X)$.

Exercise: Let X be Bernoulli random variable with $P(X = 1) = p$, where $p \in (0, 1)$. Compute $Var(X)$.
Exercise: Let X be a discrete random variable with probability function given by $f_X(x) = 1/3$, $x = -1, 0, 1$. Compute $Var(X)$.
Exercise: Suppose $X \sim U(a, b)$, that is, X is an uniform random variable in the set (a, b), where $a < b$. Compute $Var(X)$.
Exercise: Suppose that the random variable X has probability density function given by

$$f_X(x) = \begin{cases} 3x^{-4} & , \text{ for } x > 1 \\ 0 & , \text{ otherwise} \end{cases}.$$

Compute the $Var(X)$.

The Standard deviation: The variance is not measured in the scale of the random variable as it is computed using the square function, in order to obtain a measure of dispersion about the mean which is measure in the same scale of the random variable we need to compute the standard deviation. The *Standard deviation* is given by:

$$\sigma_X = \sqrt{Var(X)}.$$

Coefficient of variation: If we are interested in a measure of dispersion which is independent of the scale of the random variable we should use the coefficient of variation. The *coefficient of variation* is given by

$$CV(X) = \frac{\sigma_X}{\mu_X}.$$

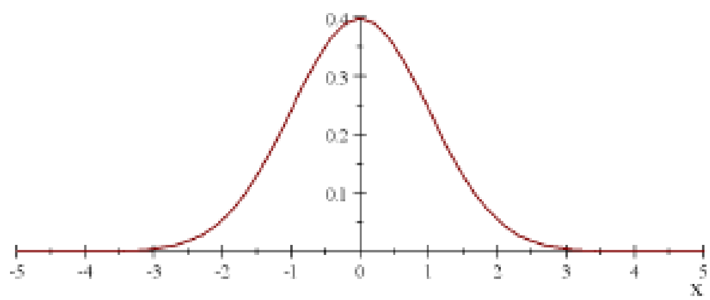
3.7 Skewness

Beyond the location and dispersion it is desirable to know the distribution behaviour about the mean. One parameter of interest is the coefficient of asymmetry also known as skewness. This parameter is a measure of asymmetry of a probability function/density about the mean of the random variable. It is given by

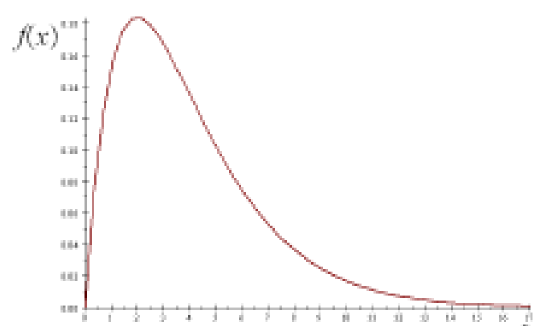
$$\gamma_1 = \frac{E\left[(X-\mu_X)^3\right]}{Var\,(X)^{3/2}} = \frac{\mu_3}{\sigma_X^3}$$

Remarks:

- For discrete random variables a probability function is symmetric if $f_X(\mu_x - \delta) = f_X(\mu_x + \delta)$ for all $\delta \in \mathbb{R}$.
- For continuous random variables the probability density function is symmetric if $f_X(\mu_x - \delta) = f_X(\mu_x + \delta)$ for all $\delta \in \mathbb{R}$



Symmetric density function



Asymmetric density function

Remark: The symmetry of probability function/density about the mean, implies that all central moments of odd order, which exist, are null. Hence if $\gamma_1 \neq 0$, the probability function/density is not symmetric about the mean.

Example: Let X be a discrete random variable with probability function given by

$$f_X(x) = \begin{cases} 0.25 & \text{, for } x = -1 \\ 0.5 & \text{, for } x = 0 \\ 0.25 & \text{, for } x = 1 \end{cases} .$$

Note that $\mu_X = E(X) = (-1) \times 0.25 + (0) \times 0.5 + 1 \times 0.25 = 0$ and $f_X(x)$ is symmetric about $\mu_X = 0$. Note also that. $E(X^3) = (-1)^3 \times 0.25 + (0)^3 \times 0.5 + 1^3 \times 0.25. = 0$, and consequently $\gamma_1 = 0$.

Remark: Note however that we can have $\gamma_1 = 0$, and the probability function/density is not symmetric about the mean, that is $\gamma_1 = 0$ does not imply symmetry.

.

Example: Let X be a discrete random variable with probability function given by

$$f_X(x) = \begin{cases} 0.1 & , \text{ for } x = -3 \\ 0.5 & , \text{ for } x = -1 \\ 0.4 & , \text{for } x = 2 \end{cases}.$$

Note that $\mu_X = E(X) = (-3) \times 0.1 + (-1) \times 0.5 + 2 \times 0.4 = 0$. Since $f_X(-1) \neq f_X(1)$, this function is not symmetric around $\mu_X = 0$. Note however that. $E(X^3) = (-3)^3 \times 0.1 + (-1)^3 \times 0.5 + 2^3 \times 0.4. = 0$, and consequently $\gamma_1 = 0$.

3.8 Kurtosis

The kurtosis measures the “thickness” of the "tails" of the probability function/density or, equivalently, the “flattening” of the probability function/density in the central zone of the distribution.

$$\gamma_2 = \frac{E\left[(X - \mu_X)^4\right]}{Var(X)^2} = \frac{\mu_4}{\sigma_X^4}.$$

3.9 Quantiles

Other parameters of interest are the quantiles of a (cumulative) distribution or quantiles of a random variable. Quantiles have the advantage that they exist even for random variables that do not have moments.

Definition: Let be X be random variable and $\alpha \in (0, 1)$. The quantile of order α, q_α is the *smallest* value among all points x in $\mathbb{R}$ that satisfy the condition

$$F_X(x) \geq \alpha.$$

Remarks:

1. If X is a discrete random variable $q_\alpha \in D_X$.
2. The quantile 0.5 is called the *median* of a (cumulative) the distribution function. It can also be interpreted as a centre of the distribution and therefore it is also considered a measure of location.
3. When the probability function/ density is symmetric the $median = mean$.
4. The q_α are called quartiles if $\alpha = 0.25,\ 0.5,\ 0.75$. Therefore the first quartile is $q_{0.25}$, the second quartile is $q_{0.5}$ and the third quartile is $q_{0.75}$

5. The q_α are called deciles if $\alpha = 0.1,\ 0.2,...,0.9$. Therefore the first decile is $q_{0.1}$, the second decile is $q_{0.2}$, etc..

6. The q_α are called percentiles if $\alpha = 0.01,\ 0.02,\ldots,0.99$. Therefore the first percentile is $q_{0.01}$, the second percentile is $q_{0.02}$, etc.

7. The interquartile range $IQR = q_{0.75} - q_{0.25}$ is considered a measure of dispersion.

Example: Let X be the Bernoulli random variable with $P(X = 1) = p = 0.2$. It follows that

$$F_X(x) = \begin{cases} 0 & ,\text{ for } x < 0 \\ 0.8 & \text{for } 0 \leq x < 1 \\ 1 & ,\text{ for } x \geq 1 \end{cases}$$

Hence $q_{0.5}$ which is the <u>*smallest*</u> value among all points x in $\mathbb{R}$ that satisfy the condition

$$F_X(x) \geq 0.5.$$

is given by $q_{0.5} = 0$.

Note that if X is the Bernoulli random variable with $P(X = 1) = p = 0.6$, we have

$$F_X(x) = \begin{cases} 0 & ,\text{ for } x < 0 \\ 0.4 & \text{for } 0 \leq x < 1 \\ 1 & ,\text{ for } x \geq 1 \end{cases}$$

and consequently $q_{0.5} = 1$.

In general if X is the Bernoulli random variable with $p \leq 0.5$, $q_{0.5} = 0$ and if $p > 0.5$, $q_{0.5} = 1$.

Example: Suppose that X is a random variable that takes values $x = 1, 2, 3, ...$ and that its probability function is given by

$$f_X(x) = \frac{6}{\pi^2 x^2}, x = 1, 2, ...$$

where $\pi = 3.14159.....$ Compute $q_{0.5}$.

Solution: Note that $f_X(1) = P(X = 1) = 6/\pi^2 = 0.607927$. Hence $F_X(x) = 0$ for for $x < 1$ and $F_X(x) = 0.607927$ for $1 \leq x < 2$. Hence $q_{0.5}$ which is the <u>*smallest*</u> value among all points x in $\mathbb{R}$ that satisfy the condition

$$F_X(x) \geq 0.5.$$

is given by $q_{0.5} = 1$. (Recall that $E(X)$ does not exist in this case).

Remarks:

1. If X is a continuous random variable, the definition of quantile can be simplified. The quantile of order α, q_α is the <u>*smallest*</u> value among all points x in $\mathbb{R}$ that satisfy the condition

$$F_X(x) = \alpha.$$

2. If X is a continuous random variable and $F_X(x)$ is a strictly increasing function in x, then $q_\alpha = F_X^{-1}(\alpha)$, there $F_X^{-1}(\alpha)$ is the inverse function of $F_X(x)$ evaluated at α.

Example: Let X be a continuous random variable with density function

$$f_X(x) = \begin{cases} 0 & , \text{ for } x < 0 \\ e^{-x} & \text{for } x \geq 0 \end{cases}$$

Compute q_α, the median and the interquartile range.

Solution: Note that

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

Therefore $F_X(q_\alpha) = \alpha$ is equivalent to $1 - e^{-q_\alpha} = \alpha$. and consequently $1 - \alpha = e^{-q_\alpha}$, so $q_\alpha = -\log(1-\alpha)$. The $Median$=$q_{0.5} = -\log(0.5) = 0.69315$. The interquartile range $IQR = q_{0.75} - q_{0.25} = -\log(1-0.75) + \log(1-0.25) = 1.0986$.

3.10 The mode

Definition: The mode of a random variable or distribution is the value M that satisfies the condition $f_X(M) \geq f_X(x)$, for all $x \in \mathbb{R}$, where $f_X(x)$ is the probability function in the case of discrete random variables and it is the probability density function in the case of continuous random variables.

Remarks:

1. The mode can also be interpreted as a centre of the distribution and therefore it is also considered a measure of location.
2. In the case of discrete random variable the mode is the most frequent value.
3. The mode does not have to be unique.
4. If the variable probability distribution/density is symmetric and has only one mode, then the mode equals the median and the mean.

Example: Let X be the Bernoulli random variable with $P(X = 1) = p = 0.2$. It follows that $P(X = 0) = 0.8$, hence the mode is given by $M = 0$.

Note that if X is the Bernoulli random variable $P(X = 1) = p = 0.6$, then $P(X = 0) = 0.4$ and the mode is given by $M = 1$. However that if $P(X = 1) = p = 0.5$, we have $P(X = 0) = 0.5$, and therefore there are two modes $M = 0$ and $M = 1$.

In general if X is the Bernoulli random variable with $P(X = 1) = p < 0.5$, and $M = 0$. If $P(X = 1) = p > 0.5$, $M = 1$ and if $p = 0.5$ there are two modes $M = 0$ and $M = 1$.

Example: Suppose that X is a random variable that takes values $x = 1, 2, 3, ...$ and that its probability function is given by

$$f_X(x) = \frac{6}{\pi^2 x^2}, x = 1, 2, ...$$

where $\pi = 3.14159.....$ Compute the mode.

Solution: Since $f_X(x)$ is decreasing the x, the mode is given by $M = 1$.

Example: Let X be a continuous random variable with density function

$$f_X(x) = \begin{cases} 0 & , \text{ for } x < 0 \\ e^{-x} & \text{for } x \geq 0 \end{cases}$$

Compute the mode.
Solution: Since $f_X(x) = 0$ for $x < 0$ and it is a decreasing function for $x \geq 0$, the mode is given by $M = 0$.

3.11 The moment generating function

Definition: The moment generating function of a discrete random variable is given by

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x \in D_X} e^{tx} \times f_X(x) = \sum_{i=1}^{\infty} e^{tx_i} \times f_X(x_i),$$

provided that it is finite.

Definition: The moment generating function of a continuous random variable is given by

$$M_X(t) = E\left(e^{tX}\right) = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx..$$

provided that it is finite.

Remarks on the moment generating function (m.g.f.):

- The m.g.f. may not exist.
- If X is a discrete random variables and D_X is finite, then there is always a m.g.f.;
- The moment generating function is a function of t not X;
- If there is a m.g.f., then there are moments of every order. The reverse is not true.
- A distribution which has no moments – or has only the first k moments – does not have a m.g.f..
- The moment generating function is used to calculate the moments.
- The m.g.f. uniquely determines the distribution function. That is, if two random variables have the same m.g.f., then the cumulative distribution functions of the random variables coincide, except perhaps at a finite number of points.
- The moment generating function of a sum of independent random variables $S_n = \sum_{i=1}^{n} X_i$ equals the product of their m.g.f.(s).

$$M_{S_n}(t) = M_{X_1}(t) \times M_{X_2}(t) \times ... \times M_{X_n}(t)$$

Theorem:

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r' = E\left[X^r\right], \; r = 1, 2, 3, ...$$

Property: $M_{bX+a}(t) = E\left[e^{(bX+a)t}\right] = e^{at} M_X(bt)$

Example: Let X be the Bernoulli random variable with $P(X = 1) = p = 0.2$. Compute $M_X(t)$ and $E(X^r)$, $r = 1, 2, ...$

Solution:

$$M_X(t) = E\left(e^{tX}\right) = e^t p + (1-p)$$

Hence

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = p = E\left(X^r\right),$$

$r = 1, 2, 3, ...$

Example: Let X be a discrete random variable with probability function

$$f_X(x) = p(1-p)^{x-1}, x = 1, 2, ...$$

and $p \in (0,1)$. Compute $M_X(t)$ and $E(X)$

Solution: Note that

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) = \sum_{x=1}^{\infty} e^{tx} p (1-p)^{x-1} \\ &= \frac{p}{1-p} \sum_{x=1}^{\infty} \left(e^t(1-p)\right)^x \end{aligned}$$

Recall that $\sum_{i=1}^{\infty} c^i = \frac{c}{1-c}$ provided that $|c| < 1$.Hence provided that $e^t(1-p) < 1$ we have

$$M_X(t) = \frac{p}{1-p} \frac{e^t(1-p)}{1-e^t(1-p)} = \frac{e^t p}{1-e^t(1-p)}.$$

Note that $e^t(1-p) < 1$ holds provided that $t < -\log(1-p)$.

Hence

$$\begin{aligned} \frac{dM_X(t)}{dt} &= \frac{d\left(\frac{e^t p}{1-e^t(1-p)}\right)}{dt} \\ &= \frac{e^t p\left[1-e^t(1-p)\right] + e^t(1-p) e^t p}{\left[1-e^t(1-p)\right]^2} \end{aligned}$$

and consequently

$$E(X) = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \frac{p\left[1-(1-p)\right] + (1-p)p}{\left[1-(1-p)\right]^2} = \frac{p^2 + p - p^2}{p^2} = \frac{1}{p}.$$

Example: Let X be a continuous random variable with density function

$$f_X(x) = \begin{cases} 0 & , \text{ for } x < 0 \\ e^{-x} & \text{ for } x \geq 0 \end{cases}$$

Compute $M_X(t)$ and $E(X)$.

Solution:

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) = \int_0^{+\infty} e^{(t-1)x} dx \\ &= \lim_{z\to\infty} \int_0^z e^{(t-1)x} dx = \lim_{z\to\infty} \left[\frac{e^{(t-1)x}}{t-1}\right]_{x=0}^{x=z} \\ &= \lim_{z\to\infty} \left[\frac{e^{(t-1)z}}{t-1} - \frac{1}{t-1}\right] = -\frac{1}{t-1}, \end{aligned}$$

provided that $t < 1$. Now

$$\frac{dM_X(t)}{dt} = \frac{d\left(-\frac{1}{t-1}\right)}{dt} = \frac{1}{(t-1)^2}$$

and consequently

$$E(X) = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = 1.$$