Probability and Statistics

Omid Safarzadeh

December 20, 2021

Table of contents

*__Acknowledgement__: This slide is prepared based on Casella and Berger, 2002

Distribution of Functions of a Random Variable

- If X is a random variable with cdf F_X, then $Y = g(X)$ is also a random variable.
- Importantly, since Y is a function of X, we can determine its random behaviour in terms of the behaviour of X.
- Then, for any set A,

$$P(Y \in A) = P(g(X) \in A).$$

  This clearly shows that the distribution of Y depends on the function $g(.)$ and the cdf F_X.
- Formally,

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

  where $\mathcal{X}$ and $\mathcal{Y}$ are the sample spaces of X and Y, respectively.

Distribution of Functions of a Random Variable

- Notice that the mapping $g(.)$ is associated with the inverse mapping $g^{-1}(.)$, a mapping from the subsets of $\mathcal{Y}$ to those $\mathcal{X}$:

$$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\}. \tag{1}$$

- Therefor, the mapping $g^{-1}(.)$ takes sets into sets, that is, $g^{-1}(A)$ is the set of points in $\mathcal{X}$that $g(x)$ takes into the set A.
- If $A = \{y\}$, a point set, then

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

- Now, if $Y = g(X)$, then for all $A \in \mathcal{Y}$,

$$\begin{aligned} P(Y \in A) &= P(g(X) \in A) \\ &= P(\{x \in \mathcal{X} : g(x) \in A\}) \\ &= P(X \in g^{-1}(A)), \end{aligned} \tag{2}$$

where the last line follows from (1). This defines the probability distribution of Y.

Example 1.1

A discrete random variable X has binomial distribution if its pmf is of the form

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n,$$

where n is a positive integer and $0 \leq p \leq 1$. Values such as n and p are called parameters of a distribution. Different parameter values imply different distributions.

Distribution of Functions of a Random Variable

- The cdf of $Y = g(X)$ is

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

Distribution of Functions of a Random Variable

Theorem 1.1

Let X have cdf $F_X(x)$, let $Y = g(X)$ and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as

$$\mathcal{X} = \{x : f_X(x) > 0\} \quad and \quad \mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}. \tag{3}$$

1. *If g is an increasing function on $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.*
2. *If g is a decreasing function on $\mathcal{X}$ and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.*

Proof: *Exercise!*

Distribution of Functions of a Random Variable

Theorem 1.2

Let X have pdf $f_X(x)$ and let $Y = g(X)$,where g is a monotone function, Let $\mathcal{X}$ and $\mathcal{Y}$ be defined as in (3). Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. The the pdf of Y is given by

$$f_Y(y) = \begin{cases} f_x(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| & y \in \mathcal{Y} \\ \\ 0 & \text{otherwise} \end{cases}$$

Proof: *Exercise!*

Distribution of Functions of a Random Variable

Example 1.2

Suppose X $f_X(x) = 1$ for $0 < x < 1$ and 0 otherwise, which is the *uniform*$(0, 1)$ distribution. Observe that $F_X(x) = x$, $0 < x < 1$. We now make the transformation $Y = g(X) = -\log X$. Then,

$$g'(x) = \frac{d}{dx}(-\log x) = -\frac{1}{x} < 0 \quad \text{for } 0 < x < 1;$$

hence, $g(x)$ is monotone and has e continuous derivative on $0 < x < 1$. Also, $\mathcal{Y} = (0, \infty)$. Observe that $g^{-1}(y) = e^{-y}$. Then, using Theorem (1.2),

$$\begin{aligned} f_Y(y) &= 1 * |-e^{-y}| \quad \text{if } 0 < y < \infty \\ &= e^{-y} \quad \text{if } 0 < y < \infty. \end{aligned}$$

Distribution of Functions of a Random Variable

Example 1.3

Let X have the standard normal distribution,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \qquad -\infty < X < \infty.$$

- Consider $Y = X^2$. The function $g(x) = x^2$ is monotone on $(-\infty, 0)$ and on $(0, \infty)$. The set $\mathcal{Y} = (0, \infty)$. Applying Theorem (1.2), we take

$$A_0 = \{0\};$$

$$A_1 = (-\infty, 0), \qquad g_1(x) = x^2, \qquad g_1^{-1} = -\sqrt{y};$$
$$A_2 = (0, \infty), \qquad g_2(x) = x^2, \qquad g_2^{-1} = -\sqrt{y}.$$

Example 1.3 cont.

- Then, the pdf of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-(-\sqrt{y})^2} \left| -\frac{1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{(-\sqrt{y})^2} \left| \frac{1}{2\sqrt{y}} \right|$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{y/2}, \qquad 0 < y < \infty.$$

- This is the pdf of a chi squared random variables with 1 degree of freedom. You will use this distribution frequently in your later econometrics courses.

Theorem 1.3

Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then, Y is uniformly distributed on (0,1), that is

$$P(Y \leq y) = y, \quad 0 < y < 1.$$

- **Proof**: *Exercise!*

Distribution of Functions of a Random Variable

- Why is this result useful?
- This result connects any random variable with some cdf $F_X(x)$ with a uniformly distributed random variable. Hence, if we want to simulate random numbers from some distribution $F_X(x)$, all we have to do is to generate uniformly distributed random variables, Y, and then solve for $F_X(x) = y$. As long as we can compute $F_X^{-1}(y)$, we can generate random numbers from the distribution $F_X(x)$.

Expected Value

- In this section, we will introduce one of the most widely used concepts in econometrics, the expected value.
- As we will see in more detail, this is one of the momentsthat a random variable can possess.
- This concepts is akin to the concept of "average". The standard "average" is an arithmetic average where all available observations are weighted equally.
- The expected value, on the other hand, is the average of all possible values a random variable can take, weighted by the probability distribution.
- The question is,which value would we expect the random variable to take on, on average.

Expected Value

Definition 2.1

The expected value or mean of a random variable $g(X)$, denoted by $E[g(X)]$, is

$$E[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) = \sum_{x \in \mathcal{X}} g(x) P(X = x) & \text{if } X \text{ is discrete} \end{cases}$$

provided that the integral or sum exists. If $E[g(X)] = \infty$, we say that $E[g(X)]$ does not exist.

- In both cases, the idea is that we are taking the average of $g(x)$ over all of its possible values ($x \in \mathcal{X}$), where these values are weighted by the respective value of the pdf, $f_X(x)$.

Expected Value

Example 2.1

Suppose X has an exponential(λ) distribution, that is, it has pdf given by

$$f_X(x) = \frac{1}{\lambda}e^{-x/\lambda}, \quad 0 \leq x < \infty \quad \lambda > 0.$$

Then,

$$E[X] = \int_0^\infty \frac{1}{\lambda} x e^{-x/\lambda} dx = -xe^{-x/\lambda}|_0^\infty + \int_0^\infty e^{-x/\lambda} dx \tag{4}$$

$$= \int_0^\infty e^{-2/\lambda} dx = \lambda. \tag{5}$$

- To obtain this result, we use a method called integration by parts. This is based on

$$\int udv = uv - \int vdu.$$

Example 2.1 cont.

- Then, taking

$$u = x, \quad du = 1dx,$$
$$v = -e^{-x/\lambda}, \quad dv = \lambda^{-1}e^{-x/\lambda}dx,$$

gives (4).

- To obtain (5), notice that, by L'Hôpital's Rule,

$$\lim_{x\to\infty} \frac{x}{e^{x/\lambda}} = \lim_{x\to\infty} \frac{\frac{d}{dx}x}{\frac{d}{dx}e^{x/\lambda}} = \lim_{x\to\infty} \frac{1}{\lambda^{-1}e^{x/\lambda}} = 0.$$

- Finally,

$$\int_0^\infty e^{x/\lambda|}dx = -\lambda^{-x/\lambda}|_0^\infty = \lambda.$$

Expected Value

- A very useful property of the expectation operator is that it is a linear operator.
- For example, consider some X such that $E[X] = \mu$.
- Then, for twoconstants$a$ and b,

$$E[a + Xb] = a + E[Xb] = a + be[x] = a + b\mu.$$

- Notice that, clearly, the expectation of a constant is equal to itself.

Expected Value

Theorem 2.1

Let X be a random variable and let a, b and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

- $E[ag_1(X) + bg_2(X) + c] = aE[g_1(X)] + bE[g_2(X)] + c$.
- *If* $g_1(x) \geq 0$ *for all* x*, then* $E[g_1(X)] \geq 0$.
- *If* $g_1(x) \geq g_2(x)$ *for all* x*, then* $E[g_1(X)] \geq E[g_2(X)]$.
- *If* $a \leq g_1(x) \leq b$ *for all* x*, then* $a \leq E[g(X)] \leq b$.

Proof: *Exercise!*

Expected Value

Example 2.2

Let X have a uniform distribution, such that

$$f_X(x) = \begin{cases} 1 & \text{if} \quad 0 \leq x \leq 1 \\ \\ 0 & \text{if} \quad \mathrm{otherwise} \end{cases}$$

Define $g(X) = -\log X$. Then,

$$E[g(X)] = E[-\log X] = \int_0^1 -\log x dx = (-x \log x + x)|_0^1 = 1,$$

where we use integration by parts. We can also use $f_Y(y)$ to calculate $E[Y]$ directly.

Example 2.2 cont.

- In Example (1.2), it was shown that, for the case at hand

  $$f_Y(y) = e^{-y} \quad \text{if } 0 \le y \le \infty.$$

  Remember from Example (2.1) that Y $f_Y(y) = \lambda^{-1}e^{-y/\lambda}$, where $0 \le y < \infty$ and $\lambda > 0$, we have $E[Y] = \lambda$. Notice that this comes down to the same pdf if we pick $\lambda = 1$. Hence, $E[Y] = 1$.
- Note that the textbook is a bit vague on the set of possible values for X. In Example (1.2) we have $0 < x < 1$ while in this Example, the textbook actually considers $0 \le x \le 1$.

Moments

- Another widely used moment is the varianceof a random variable. Obviously, this moment measures the variation/dispersion/spread of the random variable (around expectation).
- While the expectation is usually denoted by μ, σ^2 is generally used for variance.
- Variance is a second-order moment. If available, higher order moments of a random variable can be calculated, as well.
- For example, the third and fourth moments are concerned with how symmetric and fat-tailed the underlying distribution is. We will talk more about these.

Moments

Definition 2.2

For each of integer n, the n^{th} moment of X is

$$\mu_n' = E[X^n].$$

The n^{th}central momentof X, μ_n, is

$$\mu_n = E[(X - \mu)^n],$$

where $\mu = \mu_1' = E[X]$.

Moments

Definition 2.3

The variance of a random variable X is its second central moment,

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

while $\sqrt{Var(X)}$ is known as the standard deviation of X.

- Importantly,

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

Moments

- Let us digress for a moment and briefly mention another important concept: covariance. When it exists, the covariance of two random variables X and Y is defined as
$$Cov(X, Y) = E(\{X - E[X]\}\{Y - E[Y]\}).$$

Discrete Uniform Distribution

- A random variable X has a discrete *uniform*(1,N) distribution if

$$P(X = x|N) = \frac{1}{N}, \qquad x = 1, 2, ..., N,$$

where N is a specified integer. This distribution puts equal mass on each of the outcomes $1, 2, ..., N$.

Binomial Distribution

- This is based on a Bernoulli trial (after James Bernoulli which is an experiment with two, and only, two, possible outcomes.
- A random variable X has Bernoulli(p) distribution if

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases} \qquad 0 \leq p \leq 1.$$

- $X = 1$ is often termed as "success" and p is, accordingly, the probability of success. Similarly, $X = 0$ is termed a "failure".
- Now,

$$E[X] = 1 * p + 0 * (1-p) = p,$$
$$\text{and} \quad Var(X) = (1-p)^2 p + (0-p)^2 (1-p) = p(1-p).$$

- $E[X] = np$ (**Proof**: Exercise!)
- $Var(X) = np(1-p)$ (**Proof**: Exercise!)
- Examples:
  1. Tossing a coin (p =probability of a head, $X = 1$ if heads)
  2. Roulette ($X = 1$ if red occurs, p =probability of red)
  3. Election polls ($X = 1$ if candidate A gets vote)
  4. Incidence of disease (p =probability that a random person gets infected)

Binomial Distribution

- We can extend the scope to a collection of many independent trials.
- Define

$$A_i = \{X = 1 \text{ on the } i^{th} \text{ trial}\}, \qquad i = 1, 2, ..., n.$$

- Assuming that $A_1, ..., A_n$ are independent events, we can derive the distribution of the total number of successes in n trials.Define $Y =$"total number of successes in n trials".
- The event $\{Y = y\}$ means that out of n trials, y resulted as success. Therefore, $n - y$ trials have been unsuccessful.
- In other words, exactly y of $A_1, ..., A_n$ must have occurred.
- There are many possible orderings of the events that would lead to this outcome. Any particular such ordering has probability

$$p^y(1-p)^{n-y}.$$

- Since there are $\binom{n}{y}$ such sequences, we have

$$P(Y = y|n, p) = \binom{n}{y} p^y (1-p)^{n-y}, \qquad y = 0, 1, 2, ..., n,$$

and Y is called a *binomial(n,p)* random variable.

Binomial Distribution

Example 3.1

If you flip a fair coin 10 times, what is the probability of getting all tails?

- Let's first calculate the probability of getting tail on fair coin when you flip it one time.

$$P(1) = \frac{1}{2} = 50\%(\text{Because coin has two sides, H \& T})$$

- Since all the trails are independent, probability of getting head on *nth* turn is also $1/2$.
- Then,

$$\begin{aligned} P(10) &= \frac{1}{2} * \frac{1}{2} * \ldots * \frac{1}{2} \quad (10 \text{ times}) \\ &= (\frac{1}{2})^{10}. \end{aligned}$$

Poisson Distribution

- In modelling a phenomenon in which we are waiting for an occurrence (such as waiting for a bus), the number of occurrence in a given time interval can be modelled by the Poisson distribution.
- The basic assumption is as follows: for small time intervals, the probability of an arrival is proportional to the length of waiting time.
- If we are waiting for the bus, the probability that a bus will arrive within the next hour is higher than the probability that it will arrive within 5 minutes.
- Other possible applications are distribution of bomb hits in an area or distribution of fish in a lake.
- The only parameter is λ, also sometimes called the "intensity parameter."

Poisson Distribution

- $E[X] = \lambda$
- $Var(X) = \lambda$
- **Proof**: Exercise!

Poisson Distribution

Example 3.2

As an example of a waiting-for-occurrence application, consider a telephone operator who, on average, handles fie calls every 3 minutes. What is the probability that there will be no calls in the next minute? At leas two calls? If we let X =number of calls in a minute, then X has a Poisson distribution with $E[X] = \lambda = 5/3$. So,

$$P(\text{no calls in the next minute}) = P(X = 0)$$

$$= \frac{e^{-5/3}(5/3)^0}{0!} = e^{-5/3} = 0.189$$

and

$$P(\text{at least two calls in the next minute}) = P(X \geq 2)$$

$$= 1 - P(X = 0) - P(X = 1)$$

$$= 1 - 0.189 - \frac{e^{-5/3}(5/3)^1}{1!}$$

$$= 0.496.$$

Uniform Distribution

- The continuous uniform distribution is defined by spreading mass uniformly over an interval $[a, b]$. Its pdf is given by

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{if} \quad x \in [a,b] \\ \\ 0 & \text{if} \quad \text{otherwise} \end{cases}.$$

- One can easily show that

$$\int_a^b f(x)dx = 1,$$

$$E[X] = \frac{b+a}{2},$$

$$Var(X) = \frac{(b-a)^2}{12}.$$

- In many cases, when people say Uniform distribution, they implictly mean $(a, b) = (0, 1)$.

Exponential Distribution

- Now consider, $\alpha = 1$:

$$f(x|a,\beta) = f(x|1,\beta) = \frac{1}{\beta}e^{-e/\beta}, \quad 0 < x < \infty.$$

- Again, using our previous results, for X exponential (β) we have

$$E[X] = \beta \quad \text{and} \quad Var(X) = \beta^2$$

- A peculiar feature of this distribution is that it has no memory.

Exponential Distribution

- If X exponential(β), then, for $s > t \geq 0$,

$$
\begin{aligned}
P(X > s | X > t) &= \frac{P(X > s, X > t)}{P(X > t)} = \frac{P(X > s)}{P(X > t)} \\
&= \frac{\int_s^\infty \frac{1}{\beta} e^{-x/\beta} dx}{\int_t^\infty \frac{1}{\beta} e^{-x/\beta} dx} = \frac{e^{-s/\beta}}{e^{-t/\beta}} \\
&= e^{-(s-t)/\beta} = P(X > s - t).
\end{aligned}
$$

- This is because,

$$\int_{s-t}^\infty \frac{1}{\beta} e^{-x/\beta} dx = -e^{-x/\beta}|_{x=s-t}^\infty = e^{-(s-t)/\beta}.$$

- What does this mean? When calculating $P(X > s|X > t)$, what matters is not whether X has passed a threshold or not. What matters is the distance between the threshold and the value to be reached.
- If Mr X has been fired more than 10 times, what is the probability that he will be fired more than 12 times? It is not different from the probability that a person, who has been fired once, will be fired more than two times. History does not matter.

Normal Distribution

- We now consider the normal distribution or the Gaussian distribution.
- Why is this distribution so popular?
  1. Analytical tractability
  2. Bell shaped or symmetric
  3. It is central to Central Limit Theorem; this type of results guarantee that, under (mild) conditions, the normal distribution can be used to approximate a large variety of distribution in large samples.
- The distribution has two parameters: mean and variance, denoted by μ and σ^2, respectively.
- The pdf is given by,

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp[-1/2\frac{(x-\mu)^2}{\sigma^2}].$$

Normal Distribution

- This distribution is usually denoted as $N(\mu, \sigma^2)$.
- A very useful result is that for $X\ N(\mu, \sigma^2)$,

$$Z = \frac{X - \mu}{\sigma}\ N(0, 1).$$

- $N(0, 1)$ is known as the standard normal distribution.
- To see this, consider the following:

$$P(Z \leq z) = P(\frac{(X - \mu)/\sigma \leq z}{)}$$

$$= P(X \leq z\sigma + \mu)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z\sigma+\mu} e^{-(x-\mu)^2} / 2\sigma^2 dx$$

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

where we substitute $t = (x - \mu)/\mu$. Notice that this implies that $dt/dx = 1/\sigma$. This shows that $P(Z \leq z)$ is the standard normal cdf.

Normal Distribution

- Then, we can do all calculations for the standard normal variable and then convert these results for whatever normal random variable we have in mind.
- Consider, for $Z\ N(0, 1)$, the following:

$$E[Z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ze^{-z^2/2} dz = -\frac{1}{\sqrt{2\pi}} e^{-z^2/2}|_{-\infty}^{\infty} = 0.$$

- Then, to find $E[X]$ for $X\ N(\mu, \sigma^2)$, we can use $X = \mu + Z\sigma$:

$$E[X] = E[\mu + Z\sigma] = \mu + \sigma E[Z] = \mu + \sigma * 0 = \mu.$$

- Similarly,

$$Var(X) = Var(\mu + Z\sigma) = \sigma^2 Var(Z) = \sigma^2.$$

- What about

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz \stackrel{?}{=} 1.$$

- Let X be a random variable such that

$$\log X \ N(\mu, \sigma^2).$$

  Then, X is said to have a lognormal distribution.

- By using a transformation argument (Theorem (1.2)), the pdf of X is given by,

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} exp[-\frac{(\log x - \mu)^2}{2\sigma^2}],$$

  where $0 < x < \infty$, $-\infty < \mu, \infty$, and $\sigma > 0$.

- How? Take $W = \log X$. We start from distribution of W and want to find the distribution of $X = expW$. Then, $g(W) = exp(W)$ and $g^{-1}(X) = log(X)$. The rest follows by using Theorem (1.2).

Laplace distribution

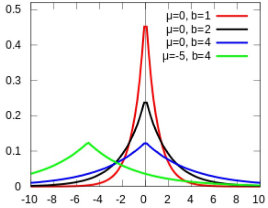
Laplace prior for β :

$$\beta_j \sim Lap(0, \frac{2\sigma^2}{\lambda} \Rightarrow p(\beta_j) = \frac{\lambda}{4\sigma^2} exp(-\frac{\lambda}{2\sigma^2}|\beta_j|)$$

where $\beta_j, j = 1, ..., p$ are i.i.d

- If $Z \sim Lap(\mu, b)$, then $E[Z] = \mu$, $Var(Z) = 2b^2$,

$$p(Z) = \frac{1}{2b} exp(-\frac{|x-\mu|}{b})$$



- **Likelihood**: Gaussian likelihood

Moments and Moment Generating Functions

- Now, let's get back on track.
- So far we have spoken mainly about the first two orders of moments.
- We now introduce a new function that is associated with a probability distribution, the moment generating function.
- This function can be used to obtain moments of a random variable.
- In practice, it is much easier in many cases to calculate moments directly than to use the moment generating function. However, the main use of the mgf is not to generate moments, but to help in characterising a distribution.

Moments and Moment Generating Functions

Definition 4.1

Let X be a random variable with cdf F_X.The moment generating function (mgf)of X (or F_X), denoted by $M_X(t)$, is

$$M_X(t) = E[e^{tX}],$$

provided that the expectation exists for t in some neighbourhood of 0. That is, there is an $h > 0$ such that, for all t in $-h < t < h$, $E[e^{tX}]$ exists.If the expectation does not exist in a neighbourhood of 0, we say that the mgf does not exist.

- We can write the mgf of X as

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

$$M_X(t) = \sigma_x e^{tx} P(X = x) \quad \text{if } X \text{ is discrete.}$$

- But why is this called a moment generating function?

Moments and Moment Generating Functions

Theorem 4.1

If X has mgf $M_X(t)$, then

$$E[X^n] = M_X^{(n)}(0),$$

where we define

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

That is, the n^{th} moment is equal to the n^{th} derivative of $M_X(t)$ evaluated at t=0.

- **Proof**: Assuming that we can differentiate under the integral sign,

$$\frac{d}{dt} M_X(t) = \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f_X(x)dx = \int_{-\infty}^{\infty} (\frac{d}{dt} e^{tx}) f_X(x)dx$$

$$= \int_{-\infty}^{\infty} x e^{tx} f_X(x)dx = E[Xe^{tx}].$$

- Hence,

$$\frac{d}{dt} M_X(t)|_{t=0} = E[Xe^{tx}]|_{t=0} = E[X].$$

Moments and Moment Generating Functions

- Similarly,

$$\frac{d^2}{dt^2}M_X(t) = \int_{-\infty}^{\infty}(\frac{d}{dt}e^{tx})f_X(x)dx = \int_{-\infty}^{\infty}(\frac{d^2}{dt^2}e^{tx})f_X(x)dx$$

$$= \int_{-\infty}^{\infty} x^2 e^{tx} f_X(x)dx = E[X^2 e^{tx}],$$

and

$$\frac{d^2}{dt^2}M_X(t)|_{t=0} = E[X^2 e^{tx}]|_{t=0} = E[X^2].$$

.

- Proceeding in the same manner, it can be shown that

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

Gamma mgf

Definition 4.2

The kernel of a function is the main part of the function, the part that remains when constants are disregarded.

Normal mgf

- Now consider the pdf for X $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2], \qquad -\infty < x < \infty.$$

- The mgf is given by

$$M_X(t) = E[e^{Xt}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx.$$

Normal mgf

- Then

$$M_X(t) = exp(\mu t + \frac{\sigma^2 t^2}{2}).$$

- Clearly,

$$E[X] = \frac{d}{dt} M_X(t)|_{t=0} = (\mu + \sigma^2 t) exp(\mu t + \frac{\sigma^2 t^2}{2})|_{t=0} = \mu,$$

$$E[X^2] = \frac{d^2}{dt^2} M_X(t)|_{t=0} = \sigma^2 exp(\mu t + \frac{\sigma^2 t^2}{2}|_{t=0}$$

$$+(\mu + \sigma^2 t)^2 exp(\mu t + \frac{\sigma^2 t^2}{2})^2|_{t=0}$$

$$= \sigma^2 + \mu^2,$$

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

References

Casella, G., & Berger, R. (2002). *Statistical inference*. Cengage Learning. https://books.google.fr/books?id=FAUVEAAAQBAJ