Econ 508B: Lecture 5

Expectation, MGF and CGF

Hongyi Liu

Washington University in St. Louis

July 31, 2017

Outline

Outline

1. Expected Values
2. Moment Generating Functions
3. Cumulative Generating Functions

Motivation: *Probability v.s. Expectation*

- To start with, people probably have a better understanding for an expected value than for probability.
- Like optimization and approximation problems, they are phrased in terms of expectations.
- Expectations are indeed seen as special cases and are treated with uniformity and economy.

Definition 1.1

Let X be a random variable on $(\Omega, \mathcal{F}, P)$.The **expected value** of X, EX, is defined as

$$EX = \int_\Omega X dP,$$

given the integral is well-defined, i.e., at least one of the two quantities $\int X^+ dP$ and $\int X^- dP$ is finite.

Proposition 1.1 (Change of variable formula)

If X is a random variable on $(\Omega, \mathcal{F}, P)$ and $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable and $Y = g(X)$ is also a random variable on $(\Omega, \mathcal{F}, P)$.

- $\int_\Omega |Y| dP = \int_\mathbb{R} |g(x)| P_X(dx) = \int_\mathbb{R} |y| P_Y(dy)$.
- If $\int_\Omega |Y| dP < \infty$, then

$$\int_\Omega Y dP = \int_\mathbb{R} h(x) P_X(dx) = \int_\mathbb{R} y P_Y(dy).$$

Moment

Definition 1.2

For any positive integer n, the n^{th} **moment** μ_n and the n^{th} **central moment** μ'_n of a random variable X is defined by

$$\mu_n \equiv EX^n, \mu'_n \equiv E(X - EX)^n$$

provided the expectation is well-defined.

In particular, the variance of a random variable X is the 2^{th} central moment, namely $Var(X) = E(X - EX)^2$, provided $EX^2 < \infty$.

Outline

MGF

Definition 2.1

The **moment generating function (MGF)** of a random variable X is

$$M_X(t) \equiv E(e^{tX}), \text{for all } t \in \mathbb{R}$$

- e^{tX} is always non-negative, therefore, $E(e^{tX})$ is well-defined but could be infinity (Why?).

Definition 2.1

The **moment generating function (MGF)** of a random variable X is

$$M_X(t) \equiv E(e^{tX}), \text{for all } t \in \mathbb{R}$$

- e^{tX} is always non-negative, therefore, $E(e^{tX})$ is well-defined but could be infinity (Why?).
- The payoff of **MGF** is that it gives the direct connection between **MGF** and the moments of a random variable X as follows.

non-negative case

Proposition 2.1

Let X be a non-negative random variable $t > 0$. Then

$$M_X(t) \equiv E(e^{tX}) = \sum_{n=0}^{\infty} \frac{t^n \mu_n}{n!}$$

Proof: By *Taylor expansion*, $e^{tX} = \sum_{n=0}^{\infty} \frac{t^n X^n}{n!}$ and X is non-negative, this comes from M.C.T.

bounded case

Proposition 2.2

Let X be a random variable and let $M_X(t)$ be finite for all $|t| < \epsilon$, for some $\epsilon > 0$, then

(1) $E|X|^n < \infty$ for all $n \geq 1$,

(2) $M_X(t) = \sum_{n=0}^{\infty} t^n \frac{\mu_n}{n!}$ for all $|t| < \epsilon$,

(3) $M_X(\cdot)$is infinitely differentiable on $(-\epsilon, +\epsilon)$ and for $r \in \mathbb{N}$, the r^{th} derivative of $M_X(\cdot)$ is

$$M_X^{(r)}(t) = \sum_{n=0}^{\infty} \mu_{n+r} \frac{t^n}{n!} = E(e^{tX} X^r) \text{for } |t| < \epsilon.$$

In particular,

$$M_X^{(r)}(0) = \mu_r = EX^r$$

Proof

(1) : According to $M_X(t)$ is finite and the fact that $\frac{|t|^n|X|^n}{n!} \leq e^{|tX|}$ for all $n \in \mathbb{N}$, then

$$E(e^{|tX|}) \leq E(e^{tX}) + E(e^{-tX}) < \infty \text{ for } |t| < \epsilon$$

Therefore, choosing a $t \in (-\epsilon, +\epsilon)$ leads to the outcome of (1).

Proof

(1) : According to $M_X(t)$ is finite and the fact that $\frac{|t|^n|X|^n}{n!} \leq e^{|tX|}$ for all $n \in \mathbb{N}$, then

$$E(e^{|tX|}) \leq E(e^{tX}) + E(e^{-tX}) < \infty \text{ for } |t| < \epsilon$$

Therefore, choosing a $t \in (-\epsilon, +\epsilon)$ leads to the outcome of (1).

(2) : Notice that $|\sum_{j=0}^{n} \frac{(tx)^j}{j!}| \leq e^{|tx|}$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$, then D.C.T. implies (2) holds.

Proof

(1) : According to $M_X(t)$ is finite and the fact that $\frac{|t|^n|X|^n}{n!} \leq e^{|tX|}$ for all $n \in \mathbb{N}$, then

$$E(e^{|tX|}) \leq E(e^{tX}) + E(e^{-tX}) < \infty \text{ for } |t| < \epsilon$$

Therefore, choosing a $t \in (-\epsilon, +\epsilon)$ leads to the outcome of (1).

(2) : Notice that $|\sum_{j=0}^{n} \frac{(tx)^j}{j!}| \leq e^{|tx|}$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$, then D.C.T. implies (2) holds.

(3) : The derivative of $M_X(\cdot)$ can be found by term-by-term differentiation of the power series. Hence,

$$\begin{aligned} M_X^{(r)}(t) &= \frac{d^r}{dt^r}\left(\sum_{n=0}^{\infty} t^n \frac{\mu_n}{n!}\right) = \sum_{n=0}^{\infty} \frac{d^r(t^n)}{dt^r} \frac{\mu_n}{n!} \\ &= \sum_{n=0}^{\infty} \mu_n \frac{t^{n-r}}{(n-r)!} = \sum_{n=0}^{\infty} \mu_{n+r} \frac{t^n}{n!} \end{aligned}$$

Example 2.1

Let $X \sim N(0,1)$, then for all $t \in \mathbb{R}$,

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{t^2/2} = \sum_{k=0}^{\infty} \frac{(t^2)^k}{k!} \frac{1}{2^k}.$$

Thus $\mu_n = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \frac{(2k)!}{k!2^k} & \text{if } n = 2k, k = 1, 2, ... \end{cases}$

Example 2.1

Let $X \sim N(0,1)$, then for all $t \in \mathbb{R}$,

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{t^2/2} = \sum_{k=0}^{\infty} \frac{(t^2)^k}{k!} \frac{1}{2^k}.$$

Thus $\mu_n = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \frac{(2k)!}{k! 2^k} & \text{if } n = 2k, k = 1, 2, ... \end{cases}$

Remark 2.1

- If $M_X(t)$ finite within a finite circle is fulfilled, then all the moments $\{\mu_n\}_{n \geq 1}$ of X are determined and its probability distribution as well.
- However, in general, probability distributions are not completely determined by their moments.

Intuitively speaking, if the sequence of moments does not grow so quickly, then the distribution is determined by its moments.

Example 2.2

A standard example of two distinct distributions with the same moment is based on the density of lognormal distribution (Billingsley, *Probability and Measure*, chapter 30.)

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

And its perturbed density:

$$f_a(x) = f(x)(1 + a \sin(2\pi \log x))$$

- They have the same moments and the n^{th} moment of each of them is $\exp(n^2/2)$.
- *Proof:* Homework!

Joint moment generating function

Definition 2.2

The *joint moment generating function* of a random vector $X = (X_1, ..., X_k)$ is defined by

$$M_{X_1,...,X_k}(t_1, ..., t_k) \equiv E(e^{t_1 X_1 + \cdots t_k X_k}),$$

for all $t_1, ..., t_k \in \mathbb{R}$. And the definition applied here for $M_{X_1,...,X_k}(\cdot)$ is similar to $M_X(t)$, namely the MGF of X 'exists' if $M_{X_1,...,X_k}(\cdot)$ is finite in a neighborhood of the origin of $\mathbb{R}^d$,$||t|| < t_0, t_0 > 0$.

$$M_X(t) = 1 + \sum_{i=1}^{k} \kappa^i t_i + \frac{1}{2} \sum_{i,j=1}^{r} \kappa^{ij} t_i t_j + \cdots$$

where $\kappa^{i_1 \cdots i_r} = E(Y^{i_1} \cdots Y^{i_r})$ for $i_1, ..., i_r = 1, ..., k$, which is referred to as the moment about the origin of order r of X, moments of order r form an array, symmetrical w.r.t permutations of indices.

Moreover,

$$\kappa^{i_1 \cdots i_r} = \frac{\partial^r M_X(t)}{\partial t_{i_1} \cdots \partial t_{i_r}}|_{t=0}$$

The relationship

$$M_X(t) = M_{X_1} \times \cdots \times M_{X_k}$$

holds *if and only if* the components of X are **independent**.

Example

Suppose $M_X(t) = \frac{1}{8}e^{-5t} + \frac{1}{4}e^{t} + \frac{5}{8}e^{7t}$. $E(X^n)$?

Answer:

$$M_X^{(n)}(t) = \frac{1}{8}(-5)^n e^{-5t} + \frac{1}{4}e^{t} + \frac{5}{8}7^n e^{7t}$$

$$E[X^n] = M_X^{(n)}(0) = \frac{1}{8}(-5)^n + \frac{1}{4} + \frac{5}{8}7^n$$

Example

Suppose $M_X(t) = \frac{1}{8}e^{-5t} + \frac{1}{4}e^t + \frac{5}{8}e^{7t}$. $E(X^n)$?
Answer:

$$M_X^{(n)}(t) = \frac{1}{8}(-5)^n e^{-5t} + \frac{1}{4}e^t + \frac{5}{8}7^n e^{7t}$$
$$E[X^n] = M_X^{(n)}(0) = \frac{1}{8}(-5)^n + \frac{1}{4} + \frac{5}{8}7^n$$

Alternatively, by the definition of expectation and MGF, random variable X, occurs -5 with probability $1/8$, occurs 1 with probability $1/4$, and occurs 7 with probability $5/8$. Thus its $E(X^n)$ is trivially $E[X^n] = M_X^{(n)}(0) = \frac{1}{8}(-5)^n + \frac{1}{4} + \frac{5}{8}7^n$.

Outline

Cumulant Generating Function

Definition 3.1

Let $M_X(t)$ be finite for $|t| < t_0$. The **cumulant generating function** of X is defined as

$$K_X(t) = \log M_X(t)$$

The CGF also completely determines the distribution of X and it can be expanded in a power series with same radius of convergence $R \geq t_0$ as follows

$$K_X(t) = \kappa_1 t + \kappa_2 \frac{t^2}{2!} + \kappa_3 \frac{t^3}{3!} + \cdots.$$

The coefficient κ_r of $t^r/r!$ is referred to as the **cumulant of order** r of X,

$$\kappa_r = \kappa_r(X) = \frac{d^r}{dt^r} K_X(t)|_{t=0}$$

Multivariate Cumulative generating function

When $X = (X_1, ..., X_k)$ is a vector, the CGF is defined as

$$K_X(t) = log M_X(t)$$

If $M_X(t)$ exists, then the CGF admits a multivariate Taylor series expansion in a neighborhood of the origin, with the coefficients corresponding to cumulants of X.

Definition 3.2

The **joint cumulant** of order r is

$$\kappa^{i_1, i_2, \cdots, i_r} = \frac{\partial^r K_X(t)}{\partial t_{i_1} \cdots \partial t_{i_r}} \Big|_{t=0}.$$

Sums of I.I.D. random variables

Let $S_n = \sum_{i=1}^{n} X_i$ and M_{X_i} exists, then

$$M_{S_n}(t) = (M_X(t))^n,\ K_{S_n}(t) = nK_X(t),$$

Also,

$$\kappa_r(S_n) = n\kappa_r(X) = n\kappa_r.$$

In a word, when working with sums of i.i.d random variables, its cumulants are simply times n by each random variable's cumulants.

Example 3.1

Let $X \sim N(\mu, \sigma^2)$ and then

$$M_X(t) = e^{\mu t + \sigma^2 \frac{t^2}{2}}, K_X(t) = \mu t + \sigma^2 \frac{t^2}{2}$$

Therefore, $\kappa_1 = \mu, \kappa_2 = \sigma^2, \kappa_r = 0$ for $r = 3, 4, ...$

Cumulants of order larger than 2 are all zero if and only if X has a normal distribution.

Location Shifts

Shifting from X to $X + a$ induce the corresponding transformation of $M_X(\cdot)$ and $K_X(\cdot)$, respectively

$$M_{X+a}(t) = E(e^{t(X+a)}) = e^{at} M_X(t),$$

and

$$K_{X+a}(t) = at + K_X(t).$$

Only the first cumulant is affected, i.e., $\kappa_1(X + a) = a + \kappa_1$.

Scale Changes

Scaling change of X by $b, b > 0$ obtains that X/b. It follows that

$$M_{X/b}(t) = E(e^{tX/b}) = M_X(t/b),$$
$$K_{X/b}(t) = K_X(t/b),$$
$$\kappa_r(X/b) = \kappa_r(X)/b^r = \kappa_r/b^r.$$

All cumulants are affected by a scale change unless $b = 1$.