Statistics I – Supplements for Chapters 5 and 6 Moment Generating Functions

Ling-Chieh Kung

Department of Information Management
National Taiwan University

October 31, 2012

Introduction

- Today we will study an important **mathematical tool** for Probability and Statistics: The moment generating function.
- It is useful in deriving means and variances.
- It is useful in finding the **distribution** of a random variable.
- It is required to understand materials in Chapters 7 to 9.
  - To memorize them, you do not need it.
  - To know **why they are true**, you need it.
- But it may be hard...

Road map

- **Moment generating functions (MGF)**.
- MGF for distributions.
- MGF for independent sums.

Moments

- For a random variable, we typically use its mean and variance to describe it.
- In general, we may use **moments**:

Definition 1 (Moments)

The k^{th} moment of a random variable X is defined as

$$\mu'_k \equiv \mathbb{E}\left[X^k\right].$$

Moments: an example

- Consider the uniform distribution $\mathrm{Uni}(0,1)$:
  - $f(x) = 1$.
  - $\mu_1' = \mathbb{E}[X] = \frac{1}{2}$.
  - $\mu_2' = \mathbb{E}[X^2] = \int_0^1 x^2 dx = \frac{1}{3}$.
  - $\mu_3' = \mathbb{E}[X^3] = \int_0^1 x^3 dx = \frac{1}{4}$.
  - In general, $\mu_k' = \frac{1}{k+1}$.

Moments: the general case

- The **first moment**:
  - $\mu'_1 \equiv \mathbb{E}[X^1] = E[X] = \mu$.
- The **second moment**:
  - $\mu'_2 \equiv \mathbb{E}[X^2]$.
  - Moreover, $\sigma^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mu'_2 - \mu^2$.
- For most practical random variables, there are infinitely many moments.

Moments and distributions

- When we use moments to describe distributions:
  - When two RV have the same mean and variance (and thus the same second moment), they may follow different distributions.
  - When their first, second, and third moments are all the same, it is more likely that they are the same.
  - When their first four moments are all the same...
- In **all moments** are the same:

Proposition 1 (Moments and distributions)

If two random variables have all their moments identical, they must follow exactly the same distribution.

Proof. Beyond the scope of this course. □

Moment generating functions

- The proposition is attractive but hard to use.
- It will be a nightmare to calculate all the (infinitely many) moments of a random variable.
- Fortunately, statisticians have found an easier way through **moment generating functions** (MGF).

Definition 2

The moment-generating function $m(t)$ for a random variable X is defined as

$$m(t) \equiv \mathbb{E}\left[e^{tX}\right].$$

Moment generating functions

- $m(t) \equiv \mathbb{E}[e^{tX}]$ is called the moment generating function because it **generates moments**. Why?
- Recall that you may do a Taylor expansion on e^{tx} as

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots .$$

Moment generating functions

- With this, the MGF (assuming X is discrete) satisfies

$$
\begin{aligned}
\mathbb{E}\left[e^{tX}\right] &= \sum_{x \in S} e^{tx} \operatorname{Pr}(x) \\
&= \sum_{x \in S} \left[1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots\right] \operatorname{Pr}(x) \\
&= \sum_{x \in S} \operatorname{Pr}(x) + t \sum_{x \in S} x \operatorname{Pr}(x) + \frac{t^2}{2!} \sum_{x \in S} x^2 \operatorname{Pr}(x) + \frac{t^3}{3!} \sum_{x \in S} x^3 \operatorname{Pr}(x) + \cdots \\
&= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots .
\end{aligned}
$$

Moment generating functions

- Now consider the **first-order derivative** of $m(t)$:

$$\frac{d}{dt}m(t) = \mu_1' + \frac{t}{1!}\mu_2' + \frac{t^2}{2!}\mu_3' + \cdots .$$

- If we plug in $t = 0$ into the above equation, we get

$$\left.\frac{d}{dt}m(t)\right|_{t=0} = \mu_1',$$

which is the first moment.

Moment generating functions

- Now consider the **second-order derivative** of $m(t)$:

$$\frac{d^2}{dt^2}m(t) = \mu_2' + \frac{t}{1!}\mu_3' + \cdots$$

- If we plug in $t = 0$ into the above equation, we get

$$\left.\frac{d^2}{dt^2}m(t)\right|_{t=0} = \mu_2',$$

  which is the second moment.
- The **k^{th}-order derivative** generates the k^{th} moment:

$$\left.\frac{d^k}{dt^k}m(t)\right|_{t=0} = \mu_k'.$$

MGF of the Poisson distribution

- As our first example, we derive the MGF of a Poisson RV:

Proposition 2 (MGF of the Poisson distribution)

The moment generating function for $X \sim \mathrm{Poi}(\lambda)$ *is*

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

Proof. First, we have

$$m(t) = \mathbb{E}\left[e^{tX}\right] = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^t\right)^x}{x!}.$$

MGF of the Poisson distribution

Proof (cont'd). Now, note that the summation is another Taylor expansion:

$$e^{\lambda e^t} = \sum_{x=0}^{\infty} \frac{\left(\lambda e^t\right)^x}{x!}.$$

Therefore, we have

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

and the proof is complete. □

MGF of the Poisson distribution

- Let's apply the MGF of the Poisson distribution:

Proposition 3

Let $X \sim \mathrm{Poi}(\lambda)$, *then*

$$\mathbb{E}[X] = \mathrm{Var}(X) = \lambda.$$

Proof. We have

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

and thus $m'(0) = \mathbb{E}[X] = \lambda$.

MGF of the Poisson distribution

Proof (cont'd). Moreover, we have

$$\begin{aligned} m''(t) &= \frac{d}{dt}\left[\lambda e^t \cdot e^{\lambda(e^t-1)}\right] \\ &= \lambda e^t \cdot e^{\lambda(e^t-1)} + \left(\lambda e^t\right)^2 \cdot e^{\lambda(e^t-1)} \\ &= \lambda e^t \cdot e^{\lambda(e^t-1)}\left(1 + \lambda e^t\right) \end{aligned}$$

and thus $m''(0) = \mathbb{E}[X^2] = \lambda(1+\lambda) = \lambda + \lambda^2$. It then follows that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda$. $\square$

MGF of the Bernoulli distribution

- So with the MGF, it can (sometimes) be much **easier** to find the mean and variance of a given random variable.
- As another example, let's consider the Bernoulli distribution.

Proposition 4

Let $X \sim \mathrm{Ber}(p)$, *then* $\mathbb{E}[X] = p$ *and* $\mathrm{Var}(X) = p(1-p)$.

Proof. The MGF $m(t) = \mathbb{E}[e^{tX}] = p \cdot e^t + (1-p) \cdot 1$. Then we have $m'(t) = pe^t$ and $m'(0) = \mathbb{E}[X] = p$. Moreover, we have $m''(t) = pe^t$ and $m''(0) = \mathbb{E}[X^2] = p$. Then $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1-p)$. $\square$

Summary

- You may treat the MGF as a pure mathematical tool.
- It is an expectation and thus not a random variable.
- It generates moments through differentiation.
- It can be used to find means and variances.

Road map

- Moment generating functions (MGF).
- **MGF for distributions**.
- MGF for independent sums.

Two properties of MGFs

- There are two very important properties of MGFs:

Proposition 5 (Uniqueness of MGF)

For any random variable, its MGF is unique.

Proof. Beyond the scope of this course. □

Proposition 6 (MGF and distributions)

If two random variables have the same MGF, then they follow the same distribution.

Proof. Having identical MGF means having all moments identical, which mean the distributions are identical. □

MGFs for distributions

- How may we apply the above proposition to derive the distribution of a random variable?
  - As an example, suppose for a random variable X we find its MGF is $e^{4(e^t-1)}$.
  - Also we know the MGF of $\text{Poi}(\lambda)$ is $e^{\lambda(e^t-1)}$.
  - Then we may conclude that $X \sim \text{Poi}(4)$.
- In other words, we need to first find the MGF or those well-known distributions (binomial, Poisson, exponential, normal, etc.) before we use this method.

MGFs for distributions

Distribution	MGF $m(t)$	Distribution	MGF $m(t)$
$\mathrm{Ber}(p)$	$pe^t + (1-p)$	$\mathrm{Uni}(a, b)$	?
$\mathrm{Bi}(n, p)$	?	$\mathrm{Exp}(\lambda)$	?
$\mathrm{HG}(N, A, n)$	?	$\mathrm{ND}(\mu, \sigma)$	?
$\mathrm{Poi}(\lambda)$	$e^{\lambda(e^t-1)}$	$\mathrm{Gamma}(\alpha, \beta)$	?
		$\chi^2(n)$	?

MGF of the exponential distribution

- Let's try the exponential distribution.

Proposition 7 (MGF of an exponential RV)

The moment generating function for $X \sim \mathrm{Exp}(\lambda)$ is

$$m(t) = \frac{\lambda}{\lambda - t} \quad \text{or} \quad \frac{1}{1 - \frac{t}{\lambda}} \quad \forall t < \lambda.$$

Proof. For all $t < \lambda$, we have

$$m(t) = \mathbb{E}\left[e^{tX}\right] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{(t-\lambda)x}\Big|_0^\infty = \frac{\lambda}{\lambda - t},$$

which is equivalent to the second expression. $\square$

MGF of the exponential distribution

- The mean and variance may then be derived:

Proposition 8

Let $X \sim \mathrm{Exp}(\lambda)$, *then*

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \textit{and} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Proof. We have $m'(t) = \frac{\lambda}{(\lambda - t)^2}$ and $m'(0) = \mathbb{E}[X] = \frac{1}{\lambda}$.
Moreover, we have $m''(t) = \frac{\lambda}{(\lambda - t)^3}$ and $m''(0) = \mathbb{E}[X^2] = \frac{2}{\lambda^2}$.
It then follows that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}$. □

MGF of the uniform distribution

- Let's try the uniform distribution.

Proposition 9 (MGF of the uniform distribution)

The moment generating function $m(t)$ for $X \sim \mathrm{Uni}(a, b)$ is

$$m(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

Proof. Homework! □

MGF of the normal distribution

- Let's try the normal distribution.

Proposition 10 (MGF of the normal distribution)

The moment generating function $m(t)$ for $X \sim \mathrm{ND}(\mu, \sigma)$ is

$$m(t) = e^{\mu t + \frac{\sigma^2}{2} t^2} = \exp\left(\mu t + \frac{\sigma^2}{2} t^2\right).$$

- Suppose this is true, would you verify that the mean and standard deviation are indeed μ and σ?

MGF of the normal distribution

Proof. By definition, we have

$$
\begin{aligned}
m(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[tx - \frac{1}{2}\Big(\frac{x^2-2\mu x+\mu^2}{\sigma^2}\Big)\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\Big[x^2-2(\mu+t\sigma^2)x+\mu^2\Big]\right\} dx.
\end{aligned}
$$

MGF of the normal distribution

Proof (cont'd). Now, let's try to complete the square for the exponent by adding and subtracting a term:

$$
\begin{aligned}
& x^2 - 2\big(\mu + t\sigma^2\big) + \mu^2 \\
= {} & x^2 - 2\big(\mu + t\sigma^2\big) + \big(\mu + t\sigma^2\big)^2 - 2\mu\sigma^2 t - \sigma^4 t^2 \\
= {} & \big[x - \big(\mu + t\sigma^2\big)\big]^2 - \big(2\mu\sigma^2 t + \sigma^4 t^2\big).
\end{aligned}
$$

In the original derivation, this means multiplying and dividing $e^{\mu t + \frac{\sigma^2}{2} t^2} = \exp\left(\mu t + \frac{\sigma^2}{2} t^2\right)$.

MGF of the normal distribution

Proof (cont'd). We thus have

$$\begin{aligned} m(t) &= e^{\mu t + \frac{\sigma^2}{2} t^2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left[\frac{x-(\mu+t\sigma^2)}{\sigma}\right]^2 \right\} dx \\ &= e^{\mu t + \frac{\sigma^2}{2} t^2}, \end{aligned}$$

where the last equality follows because the integral is the pdf of $\text{ND}(\mu + t\sigma^2, \sigma)$. $\square$

MGF of the normal distribution

- Now we can show that the mean and variance of a normal random variable are indeed μ and σ^2.

Proposition 11

Let $X \sim \mathrm{ND}(\mu, \sigma)$, *then*

$$\mathbb{E}[X] = \mu \quad \textit{and} \quad \mathrm{Var}(X) = \sigma^2.$$

Proof. We have $m'(t) = (\mu + \sigma^2 t)e^{\mu t + \frac{\sigma^2}{2}t^2}$ and $m'(0) = \mu$. Moreover, we have $m''(t) = e^{\mu t + \frac{\sigma^2}{2}t}[\sigma^2 + (\mu + \sigma^2 t)^2]$ and $m''(0) = \sigma^2 + \mu^2$. It then follows that $\mathrm{Var}(X) = \sigma^2$.

MGFs for distributions

Distribution	MGF $m(t)$	Distribution	MGF $m(t)$
$\text{Ber}(p)$	$pe^t + (1-p)$	$\text{Uni}(a,b)$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
$\text{Bi}(n,p)$	?	$\text{Exp}(\lambda)$	$\frac{\lambda}{\lambda-t}$
$\text{HG}(N,A,n)$	?	$\text{ND}(\mu,\sigma)$	$e^{\mu t+\frac{\sigma^2}{2}t^2}$
$\text{Poi}(\lambda)$	$e^{\lambda(e^t-1)}$	$\text{Gamma}(\alpha,\beta)$	?
		$\chi^2(n)$	?

Road map

- Moment generating functions (MGF).
- MGF for distributions.
- **MGF for independent sums**.

MGF for independent sums

- MGFs are particularly useful for deriving the distribution of a **sum** of **independent** random variables.

Proposition 12

Let X_1, X_2, ..., and X_n be independent random variables with MGFs $m_1(t)$, $m_2(t)$, ..., and $m_n(t)$, respectively. If $X = X_1 + \cdots + X_n$, then its MGF

$$m(t) = m_1(t) \times m_2(t) \times \cdots \times m_n(t).$$

MGF for independent sums

Proof. By definition, we have

$$\begin{aligned} m(t) &= E[e^{tX}] = E[e^{t(X_1+\cdots+X_n)}] \\ &= E[e^{tX_1}e^{tX_2}\cdots e^{tX_n}]. \end{aligned}$$

Because X_is are independent, we have

$$\begin{aligned} m(t) &= E[e^{tX_1}]E[e^{tX_2}]\cdots E[e^{tX_n}] \\ &= m_1(t) \times m_2(t) \times \cdots m_n(t), \end{aligned}$$

which completes the proof. □

Sum of independent Bernoulli RVs

- Let's apply the proposition on the binomial distribution.

Proposition 13

The moment generating function of $X \sim \mathrm{Bi}(n, p)$ *is*

$$m(t) = \left[pe^t + (1-p)\right]^n.$$

Proof. Let $X_i \sim \mathrm{Ber}(p)$, $i = 1, ..., n$, and X_is be independent. Then we know $X = \sum_{i=1}^n X_i \sim \mathrm{Bi}(n, p)$. Let the MGF of X_i be $m_i(t) = pe^t + (1-p)$ and that of X be $m(t)$. It then follows that $m(t) = \prod_{i=1}^n m_i(t) = [pe^t + (1-p)]^n$. □

Sum of independent exponential RVs

- Now let's try to prove that the sum of independent Exponential RVs is a gamma RV.

Proposition 14

Let $X_i \sim \mathrm{Exp}(\lambda)$, $i = 1, ..., n$, *and* $X_i s$ *be independent. Then* $X = \sum_{i=1}^{n} X_i \sim \mathrm{Gamma}(n, \frac{1}{\lambda})$.

Proof. We first find the MGF of the gamma distribution. Let $h = \frac{\beta}{1-\beta t}$, we have

$$\mathbb{E}\left[e^{tX}\right] = \int_0^\infty e^{tx}\left(\frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)}\right)dx = \frac{1}{\beta^\alpha \Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-\frac{x}{h}}dx.$$

Sum of independent exponential RVs

Proof (cont'd). Let's remove the integral by making the integrand a gamma pdf (if $h > 0$ or $t < \frac{1}{\beta}$):

$$\mathbb{E}\left[e^{tX}\right] = \frac{h^\alpha \Gamma(\alpha)}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \frac{x^{\alpha-1} e^{-\frac{x}{h}}}{h^\alpha \Gamma(\alpha)} dx = \left(\frac{h}{\beta}\right)^\alpha = \frac{1}{(1-\beta t)^\alpha}.$$

Now consider $X_i \sim \mathrm{Exp}(\lambda)$, $i = 1, ..., n$. Their MGFs are $\frac{\lambda}{\lambda - t} = \frac{1}{1-\frac{t}{\lambda}}$. As X is an independent sum of X_is, the MGF of X is

$$\left(\frac{1}{1-\frac{t}{\lambda}}\right)^n,$$

which is identical to the MGF of a gamma distribution with parameters $\alpha = n$ and $\beta = \frac{1}{\lambda}$. $\square$

Properties of normal RVs

- Now we are ready to derive some very important properties of the **normal distribution**.
  - The linear function of a normal RV is normal.
  - The linear combination of independent normal RVs is normal.
  - The standardization of a normal RV.
  - The distribution of a sample mean from a normal population.

Linear function of a normal RV

- Consider a linear function of a normal RV:

Proposition 15

Let $X \sim \mathrm{ND}(\mu, \sigma)$*, then* $aX + b \sim \mathrm{ND}(a\mu + b, a\sigma)$*.*

Proof. We know the MGF of $\mathrm{ND}(\mu, \sigma)$ is $e^{\mu t} + \frac{\sigma^2}{2} t^2$. By definition, the MGF of $aX + b$ is

$$\begin{aligned} \mathbb{E}\left[e^{t(aX+b)}\right] &= \mathbb{E}\left[e^{taX} \cdot e^{tb}\right] = e^{tb}\mathbb{E}\left[e^{taX}\right] \\ &= e^{tb} \cdot e^{\mu(at) + \frac{\sigma^2}{2}(at)^2} = e^{(a\mu+b)t + \frac{(a\sigma)^2}{2} t^2}, \end{aligned}$$

which is the MGF of a normal RV with mean $a\mu + b$ and standard deviation $a\sigma$. □

Linear combination of indep. NDs

- Consider a linear combination of independent normal RVs:

Proposition 16

Let $X_i \sim \mathrm{ND}(\mu_i, \sigma_i)$ *and* X_i*s be independent, then*

$$X = \sum_{i=1}^{n} a_i X_i \sim \mathrm{ND}\left(\sum_{i=1}^{n} a_i \mu_i, \sqrt{\textstyle\sum_{i=1}^{n} a_i^2 \sigma_i^2} \right).$$

Linear combination of indep. NDs

Proof. First, note that $a_i X_i \sim \mathrm{ND}(a_i\mu_i, a_i\sigma_i)$ as this is a linear function of X_i. Now, we apply the result for independent sum and get

$$\mathbb{E}\left[e^{tX}\right] = \prod_{i=1}^{n}\left\{\exp\left[a_i\mu_i t + \frac{(a_i\sigma_i)^2}{2}t^2\right]\right\}$$

$$= \exp\left(a_1\mu_1 t + \frac{a_1^2\sigma_1^2}{2}t^2\right)\cdots\exp\left(a_n\mu_n t + \frac{a_n^2\sigma_n^2}{2}t^2\right)$$

$$= \exp\left[(a_1\mu_1 + \cdots + a_n\mu_n)t + \frac{1}{2}\left(a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2\right)t^2\right].$$

Compare this with the normal MGF and we are done. □

Standardization of a normal RV

- Consider the standardization of a normal RV:

Proposition 17

Let $X \sim \mathrm{ND}(\mu, \sigma)$, *then*

$$\frac{X-\mu}{\sigma} \sim \mathrm{ND}(0,1).$$

Proof. A direct application of the proposition for linear functions of normal random variables. □

The distribution of a sample mean

- The sample mean is one of the most important statistics.

Definition 3

Let $\{X_i\}_{i=1,\dots,n}$ be a sample from a (probably not normal) population , then

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

is the sample mean.

- A sample mean is also a random variable.
- We have computed its mean and variance. Suppose the population has mean μ and standard deviation σ:

$$\mathbb{E}[\overline{X}] = \mu \quad \text{and} \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

The distribution of a sample mean

- When the sample mean is draw from a **normal population**:

Proposition 18

Let $\{X_i\}_{i=1,\dots,n}$ *be a sample from a normal population with mean* μ *and standard deviation* σ*. Then*

$$\overline{X} \sim \mathrm{ND}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Proof. Homework! □

- The sample mean of a normal population is also normal.
- More about sample means and sampling distributions will be discussed in Chapter 7.

Summary of discrete distributions

Distribution	Mean	Variance	MGF $m(t)$
Ber(p)	p	$p(1-p)$	$pe^t + (1-p)$
Bi(n,p)	np	$np(1-p)$	$[pe^t + (1-p)]^n$
HG(N,A,n)	np $(p = \frac{A}{N})$	$np(1-p)\frac{N-n}{N-1}$	N/A
Poi(λ)	λ	λ	$e^{\lambda(e^t-1)}$

Summary of continuous distributions

Distribution	Mean	Variance	MGF $m(t)$
Uni(a, b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Exp(λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$
ND(μ, σ)	μ	σ	$e^{\mu t+\frac{\sigma^2}{2}t^2}$
Gamma(α, β)	$\alpha\beta$	$\alpha\beta^2$	$(\frac{1}{1-\beta t})^\alpha$
$\chi^2(n)$	n	$2n$	$(1-2t)^{-\frac{n}{2}}$