BIOS 660/BIOS 672 (3 Credits):
Probability and Statistical Inference I

Jianwen Cai

https://sakai.unc.edu/portal/site/bios660-bios672-3-credits

Notes 9

Moment Generating Function

(C-B 2.3, Gut III.3)

The *moment generating function (mgf)* of the rv X is defined to be

$$M_X(t) = \mathsf{E}(e^{tX})$$

provided that the expectation exists in a neighbourhood $(-h, h)$ of $t = 0$.

Theorem: Suppose the mgf $M_X(t)$ of X exists for $t \in (-h, h)$ for some $h > 0$. Then for any positive integer n,

$$\mathsf{E}(X^n) = M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\bigg|_{t=0}$$

Notice that $M_X(0) = 1$ always.

cont.

Proof: Assuming that we can interchange expectation and differentiation,

$$\begin{aligned} & \frac{d}{dt} M_X(t) = \frac{d}{dt}\mathsf{E}(e^{tX}) = \mathsf{E}\left(\frac{d}{dt}e^{tX}\right) = \mathsf{E}(Xe^{tX}) \\ \Rightarrow \quad & \frac{d}{dt} M_X(t)\bigg|_{t=0} = \mathsf{E}(X) \\ & \frac{d^2}{dt^2} M_X(t)\bigg|_{t=0} = \mathsf{E}(X^2 e^{tX})\bigg|_{t=0} = \mathsf{E}(X^2) \end{aligned}$$

Another way to see this is

$$M_X(t) = \mathsf{E}(e^{tX}) = \mathsf{E}\left(\sum_{n=0}^{\infty} \frac{t^n}{n!} X^n\right) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathsf{E}X^n$$

so the moments can be obtained from a Taylor expansion of $M_X(t)$ around $t = 0$.

Example: Continuous

Mgf of an exponential rv: Let $f_X(x) = \lambda e^{-\lambda x}\, 1(x > 0)$. Then

$$M_X(t) = \mathsf{E}(e^{tX}) = \int_0^\infty e^{tx} \cdot \lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx$$
$$= \lambda \frac{-1}{\lambda - t} e^{-(\lambda - t)x}\Big|_0^\infty = \begin{cases} \frac{\lambda}{\lambda - t} & \text{if } t < \lambda \\ \infty & \text{otherwise} \end{cases}$$

This is fine as we only need the mgf to be defined near zero. To obtain the moments, assume $|t| < \lambda$:

$$M_X(t) = \frac{1}{1 - t/\lambda} = \sum_{n=0}^\infty \frac{t^n}{\lambda^n} = \sum_{n=0}^\infty \frac{t^n}{n!}\mathsf{E}(X^n) \quad \Rightarrow \quad \mathsf{E}(X^n) = \frac{n!}{\lambda^n}$$

In particular,

$$\mathsf{E}X = 1/\lambda$$
$$\mathsf{Var}X = \mathsf{E}X^2 - \mathsf{E}^2X = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$$

Example: Discrete

Mgf of a geometric rv: Let $f_X(x) = pq^{x-1}$, $x = 1, 2, \dots$ ($q = 1 - p$).

$$M_X(t) = \mathsf{E}(e^{tX}) = \sum_{x=1}^\infty e^{tx} \cdot pq^{x-1} = \frac{p}{q} \sum_{x=1}^\infty (qe^t)^x$$
$$= \frac{p}{q}\left(\frac{1}{1 - qe^t} - 1\right) = \frac{pe^t}{1 - qe^t}$$

The sum converges if $e^t q < 1$, that is, $t < \log(1/q)$.

The moments can be obtained by differentiation:

$$\mathsf{E}X = \frac{d}{dt}\left[\frac{p}{e^{-t} - q}\right]_{t=0} = \frac{pe^{-t}}{(e^{-t} - q)^2}\Big|_{t=0} = \frac{1}{p}$$

Linear transformations

For any constants a, b, the mgf of the rv $g(X) = aX + b$ is

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Proof:

$$\begin{aligned} M_{aX+b}(t) &= \mathsf{E}(e^{t(aX+b)}) = \mathsf{E}(e^{taX} e^{bt}) \\ &= e^{bt} \mathsf{E}(e^{(at)X}) = e^{bt} M_X(at) \end{aligned}$$

Assign as HW

Example: Continuous

Mgf of a Gaussian: Let $X \sim N(0,1)$. Then

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx = e^{t^2/2} \int_{-\infty}^{\infty} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}}\, dx = e^{t^2/2}$$

Also,

$$e^{t^2/2} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{t^2}{2}\right)^n = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!} = \sum_{m:\ \text{even}} \frac{t^m}{m!} \frac{m!}{2^{m/2}(m/2)!}$$

Matching coefficients of $t^m/m!$ we get that all the odd moments are zero and that the even moments are $E(X^m) = m!/(2^{m/2}(m/2)!)$.

Now let $Y = \mu + \sigma X$ so that $Y \sim N(\mu, \sigma^2)$,

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

Existence of moments

Too hard. May eliminate. **Theorem**: Suppose the mgf $M_X(t)$ of X exists for $t \in (-h, h)$ for some $h > 0$. Then all moments exist: $|\mathsf{E}X^r| < \infty$ for all $r > 0$.

Proof: Fix $t \in (-h, h)$. There exists $C > 0$ such that

$$|x^r| \le Ce^{|tx|}, \qquad \forall x \in \mathbb{R}$$

(What is C?)
so

$$\begin{aligned}
|\mathsf{E}X^r| &\le \mathsf{E}|X^r| \\
&\le C\mathsf{E}e^{|tX|} \\
&\le C\mathsf{E}(e^{tX} + e^{-tX}) \\
&\le C\left[M_X(t) + M_X(-t)\right] < \infty
\end{aligned}$$

Can moments not exist?

Example: Cauchy distribution

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \qquad x \in \mathbb{R}$$

The mean is

$$\begin{aligned}
\mathsf{E}X &= \int_{-\infty}^{\infty} x\,\frac{1}{\pi}\frac{1}{1+x^2}\,dx \\
&= \int_{-\infty}^{0} \frac{1}{\pi}\frac{x}{1+x^2}\,dx + \int_{0}^{\infty} \frac{1}{\pi}\frac{x}{1+x^2}\,dx \\
&= \int_{0}^{\infty} \frac{1}{\pi}\frac{x}{1+x^2}\,dx - \int_{0}^{\infty} \frac{1}{\pi}\frac{x}{1+x^2}\,dx = \ ??
\end{aligned}$$

because

$$\int_0^\infty \frac{1}{\pi}\frac{x}{1+x^2}\,dx = \int_0^\infty \frac{1}{2\pi}\frac{dy}{1+y} = \frac{1}{2\pi}\log(1+y)\Big|_0^\infty = \infty$$

Cont.

One intuitive explanation for this is that the Cauchy distribution has infinite variance:

$$\mathsf{E}X^2 = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{x^2}{1+x^2}\, dx = \int_{-\infty}^{\infty} \frac{1}{\pi}\left(1 - \frac{1}{1+x^2}\right) dx$$
$$= \frac{1}{\pi}\int_{-\infty}^{\infty} dx - 1 = \infty$$

Can mgf's not exist?

Example: Cauchy distribution
If the 1st and 2nd moments do not exist, certainly the mgf does not either!
(How would you prove this statement is true?)

Example: Log-normal distribution
If $X \sim N(0,1)$ then $Y = e^X$ is called log-normal.

$$f_Y(y) = \frac{1}{y\sqrt{2\pi}} e^{-(\log y)^2/2}, \qquad y > 0$$

For $n = 0, 1, 2, \ldots$ the moments exist, but the mgf does not, i.e. the integral $\mathsf{E}(e^{tY})$ does not converge. (Homework).

Characterizing distributions

For rvs with unbounded support, the moments do not specify the distribution: there exist distributions with different pdfs and yet have all the same moments (see Example C-B 2.3.10).

However, moments uniquely identify distributions when the rvs have bounded support.

Also, mgfs uniquely identify distributions when the mgfs exist.

Theorem: Let $F_X(x)$ and $F_Y(y)$ be cdfs all of whose moments exist.

1. If X and Y have bounded support, then $F_X(u) = F_Y(u)$ for all u iff $\mathsf{E}X^n = \mathsf{E}Y^n$ for all $n = 0, 1, 2, \ldots$.
2. If the mgfs exist and $M_X(t) = M_Y(t)$ for all t in a neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u.

Convergence of mgfs

Convergence of mgfs implies convergence of cdfs.

Theorem 2.3.12 C-B: Let $X_1, X_2, \ldots$ be a sequence of rvs with corresponding mgfs $M_{X_1}(t), M_{X_2}(t), \ldots$ such that

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t), \qquad \forall t \in (-h, h), \;\; h > 0.$$

Then $\exists!$ a unique cdf $F_X(t)$ whose moments are given by $M_X(t)$ and for all x where $F_X(x)$ is continuous,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Application

Example: Normal approximation to Poisson.
Let $X \sim Poisson(\lambda)$, then $M_X(t) = \exp[\lambda(e^t - 1)]$
(Exercise C-B 2.33), with $\mathsf{E}X = \lambda$, Var$X = \lambda$.
Let $Y = (X - \lambda)/\sqrt{\lambda}$. Then

$$M_Y(t) = \mathsf{E}(e^{tY}) = \mathsf{E}(t(X-\lambda)/\sqrt{\lambda}) = e^{-\sqrt{\lambda}t} M_X(t/\sqrt{\lambda}).$$

Hence,

$$\begin{aligned} log(M_Y(t)) &= -t\sqrt{\lambda} + \lambda(e^{t/\sqrt{\lambda}} - 1) \\ \text{(when } \lambda \text{ is large)} &= -t\sqrt{\lambda} + \lambda(\frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + \frac{t^3}{3!\lambda^{3/2}} + \cdots) \\ &= \frac{t^2}{2} + \frac{t^3}{3!\lambda^{1/2}} + \cdots \end{aligned}$$

Therefore,

$$\lim_{\lambda \to \infty} M_Y(t) = e^{t^2/2},$$

which is the mgf of a $N(0,1)$ variable.

Relationship to other transforms

For continuous rvs:

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

is similar to the two-sided *Laplace transform* of the function $f_X(x)$.

The transform

$$S_X(t) = \log M_X(t) = \log \mathsf{E} e^{tX}$$

is called *cumulant generating function*. The derivatives at $t = 0$ are called *cumulants*. In particular, (Homework)

$$S_X(0) = 0, \qquad S_X^{(1)}(0) = \mathsf{E}X, \qquad S_X^{(2)}(0) = \text{Var}X$$

E.g. $X \sim N(\mu, \sigma^2)$,

$$S_X(t) = \log(e^{\mu t + \sigma^2 t^2/2}) = \mu t + \sigma^2 t^2/2$$

Characteristic Function

(Gut III.4)

The *characteristic function (cf)* of the rv X is defined as

$$\phi_X(t) = \mathsf{E}e^{itX} = \mathsf{E}[\cos(tX) + i\sin(tX)]$$

where $i^2 = -1$.

- The cf is complex-valued, $\phi_X(t) \in \mathbb{C}$.
- The cf always exists because
$$|\mathsf{E}e^{itX}| \leq \mathsf{E}|e^{itX}| = \mathsf{E}1 = 1$$
- For calculations, the cf can often be obtained from the mgf replacing the argument t by it (as long as the mgf exists!).

Examples

$X \sim Exp(\lambda)$:

$$\phi_X(t) = \frac{\lambda}{\lambda - it}$$

$X \sim N(\mu, \sigma^2)$:

$$\phi_X(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right)$$

$X \sim Geom(p)$:

$$\phi_X(t) = \frac{pe^{it}}{1 - qe^{it}}$$

And the range is $t \in \mathbb{R}$ in all cases.

Also, for X with the Cauchy distribution

$$f_X(x) = \frac{1}{\pi(1+x^2)} \qquad \Rightarrow \qquad \phi_X(t) = e^{-|t|}$$

Exists!

Properties

1. $|\phi_X(t)| \le \phi_X(0) = 1$
2. Complex conjugate: $\overline{\phi_X(t)} = \phi_X(-t)$ (Homework)
3. Linear transformations:
$$\phi_{aX+b}(t) = e^{ibt}\phi_X(at)$$
4. The distribution is symmetric about 0, $f_X(x) = f_X(-x)$, iff $\phi_X(t)$ is real.
5. Moment generation:
$$\phi_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\bigg|_{t=0} = i^n \mathsf{E}(X^n)$$

Characterizing distributions

Characteristic functions uniquely identify distributions, always.

Theorem: If X and Y are rvs with cfs $\phi_X(t) = \phi_Y(t)$, then $F_X(u) = F_Y(u)$ for all u.

Relationship to other transforms

For continuous rvs:

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dx$$

is similar to the *Fourier transform* of the function $f_X(x)$.

For discrete rvs:

$$\phi_X(t) = \sum_{x=-\infty}^{\infty} e^{itx} f_X(x)$$

is similar to the *discrete Fourier transform* of the sequence $f_X(x)$ (when x is sampled at equal intervals).

Thus Fourier transform tables can be helpful for finding cfs.