STSCI 5080
Probability Models and Inference
Lecture 13: Moment Generating Function

October 11, 2018

Conditional expectation

Definition

(a) If (X, Y) is discrete with joint pmf $p(x, y)$, then the conditional expectation of X given Y is defined by

$$E(X \mid Y = y) = \sum_x x p_{X|Y}(x \mid y) \quad \text{for any } y,$$

provided that $\sum_x |x| p_X(x) < \infty$.

(b) If (X, Y) is continuous with joint pdf $f(x, y)$, then the conditional expectation of X given Y is defined by

$$E(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx \quad \text{for any } y,$$

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

The conditional expectation $E(X \mid Y = y)$ is a function of y.

Suppose that $\sum_x |x| p_X(x) < \infty$. Then

$$\sum_y \sum_x |x| p_{X|Y}(x \mid y) p_Y(y) = \sum_y \sum_x |x| p(x,y)$$
$$= \sum_x |x| \sum_y p(x,y) = \sum_x |x| p_X(x) < \infty,$$

so that as long as $p_Y(y) > 0$, we have

$$\sum_x |x| p_{X|Y}(x \mid y) < \infty.$$

If $p_Y(y) = 0$, then $p_{X|Y}(x \mid y) = 0$ for any x, and so

$$\sum_x |x| p_{X|Y}(x \mid y) = 0.$$

Moment generating function

Definition

Suppose that $E(e^{\theta X}) < \infty$ for all $|\theta| < a$ for some $a > 0$. Then the function

$$\psi(\theta) = E(e^{\theta X}),\ |\theta| < a$$

is called the moment generating function (mgf) of X.

The mgf is determined by the pmf/pdf or equivalently the cdf.

Theorem

If the mgf exists for a random variable X, then $E(|X|^k) < \infty$ for any positive integer k.

For example, a Cauchy random variable does not have mgf.

Proof

Suppose that $E(e^{\theta X}) < \infty$ for all $|\theta| < a$. By Taylor's expansion,

$$e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^k}{k!} + \cdots ,$$

so that

$$e^{|x|} \geq \frac{|x|^k}{k!}, \text{ i.e., } k!e^{|x|} \geq |x|^k.$$

Since $e^{|x|} \leq e^x + e^{-x}$, for sufficiently small $\theta \neq 0$, we have

$$|\theta X|^k \leq k!e^{|\theta X|}, \text{ i.e., } |X|^k \leq k!|\theta|^{-k}(e^{\theta X} + e^{-\theta X}).$$

Hence, we have

$$E(|X|^k) \leq k!|\theta|^{-k}\{\psi(\theta) + \psi(-\theta)\} < \infty.$$

Calculating moments using mgf

Theorem

Let X have mgf $\psi(\theta)$. Then for any $k = 1, 2, \ldots$,

$$\psi^{(k)}(0) = \frac{d^k}{d\theta^k}\psi(\theta)\Big|_{\theta=0} = E(X^k).$$

For example, $\psi'(0) = E(X)$ and $\psi''(0) = E(X^2)$, so that

$$\mathrm{Var}(X) = \psi''(0) - \{\psi'(0)\}^2.$$

Proof

We note that

$$\frac{\partial}{\partial \theta} e^{\theta X} = X e^{\theta X},$$

so that

$$\psi'(\theta) = \frac{d}{d\theta} E(e^{\theta X}) = E\left(\frac{\partial}{\partial \theta} e^{\theta X}\right) = E(X e^{\theta X})$$

Hence, we have

$$\psi'(0) = E(X).$$

Likewise, we have $\psi^{(k)}(0) = E(X^k)$.

Example 13.1

Example

Find the mgf of $Po(\lambda)$ and then calculate the variance of $Po(\lambda)$.

Example 13.1

Example

Find the mgf of $Po(\lambda)$ and then calculate the variance of $Po(\lambda)$.

The pmf of $X \sim Po(\lambda)$ is

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \; x = 0, 1, 2, \ldots$$

and so

$$E(e^{\theta X}) = \sum_{x=0}^{\infty} e^{\theta x} p(x) = e^{-\lambda} \underbrace{\sum_{x=0}^{\infty} \frac{(e^{\theta}\lambda)^x}{x!}}_{=\exp(\lambda e^{\theta})} = \exp\{\lambda(e^{\theta} - 1)\}.$$

The mgf of X is

$$\psi(\theta) = \exp\{\lambda(e^{\theta} - 1)\}, \; -\infty < \theta < \infty.$$

The variance of X is $\mathrm{Var}(X) = \psi''(0) - \{\psi'(0)\}^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Example 13.2

Example

Find the mgf of $N(\mu, \sigma^2)$.

Example 13.2

Example

Find the mgf of $N(\mu, \sigma^2)$.

We first calculate the mgf of $X \sim N(0,1)$. We note that

$$E(e^{\theta X}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\theta x - x^2/2} dx,$$

where

$$\theta x - \frac{x^2}{2} = -\frac{1}{2}(x^2 - \theta x) = -\frac{1}{2}(x-\theta)^2 + \frac{\theta^2}{2}.$$

Hence, we have

$$E(e^{\theta X}) = e^{\theta^2/2} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-\theta)^2/2} dx = e^{\theta^2/2},$$

so that the mgf is $\psi_X(\theta) = e^{\theta^2/2}$ for $-\infty < \theta < \infty$.

Next, if $Y \sim N(\mu, \sigma^2)$, then

$$Y = \mu + \sigma X$$

for some $X \sim N(0, 1)$. Hence,

$$E(e^{\theta Y}) = E(e^{\theta(\mu+\sigma X)}) = e^{\mu\theta}\psi_X(\sigma\theta) = e^{\mu\theta+\sigma^2\theta^2/2}.$$

The mgf of Y is

$$\psi_Y(\theta) = e^{\mu\theta+\sigma^2\theta^2/2},\ -\infty < \theta < \infty.$$

Uniqueness theorem of mgf

Theorem

Suppose that $X \sim F$ and $Y \sim G$ have mgfs $\psi_F(\theta)$ and $\psi_G(\theta)$ in an open interval I containing the origin, respectively. If

$$\psi_F(\theta) = \psi_G(\theta) \text{ for all } \theta \in I,$$

then $F \equiv G$.

See the supplementary material for a proof.

Corollary

If X has mgf that is identical to the mgf of a known cdf F, then $X \sim F$.

Finding distribution of $X + Y$

Definition

If $X \sim F$ and $Y \sim G$ are independent, then the cdf of $Z = X + Y$ is called the convolution of F and G and denoted by $F * G$, i.e., $Z \sim F * G$.

Finding distribution of $X + Y$

Definition

If $X \sim F$ and $Y \sim G$ are independent, then the cdf of $Z = X + Y$ is called the convolution of F and G and denoted by $F * G$, i.e., $Z \sim F * G$.

- If X and Y have pmfs p_X and p_Y, then $Z = X + Y$ has pmf

$$p_Z(z) = \sum_x p_X(x) p_Y(z - x).$$

  This is called the convolution of p_X and p_Y, and denoted by $p_X * p_Y$.

- If X and Y have pdfs f_X and f_Y, then $Z = X + Y$ has pdf

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx.$$

  This is called the convolution of f_X and f_Y, and denoted by $f_X * f_Y$.

Theorem

If X and Y are independent and have mgfs $\psi_X(\theta)$ and $\psi_Y(\theta)$ for $|\theta| < a$ for some $a > 0$, then $Z = X + Y$ has mgf $\psi_Z(\theta) = \psi_X(\theta)\psi_Y(\theta)$ for $|\theta| < a$.

Proof.

We note that

$$\begin{aligned}
\psi_Z(\theta) &= E\{e^{\theta(X+Y)}\} \\
&= E(e^{\theta X} e^{\theta Y}) \\
&= E(e^{\theta X})E(e^{\theta Y}) \quad \text{(by independence)} \\
&= \psi_X(\theta)\psi_Y(\theta).
\end{aligned}$$

$\square$

Example 13.3

Example

If $X \sim Po(\lambda)$ and $Y \sim Po(\mu)$ are independent, then show that $X + Y \sim Po(\mu + \lambda)$. This can be written as

$$Po(\lambda) * Po(\mu) = Po(\lambda + \mu).$$

Example 13.3

Example

If $X \sim Po(\lambda)$ and $Y \sim Po(\mu)$ are independent, then show that $X + Y \sim Po(\mu + \lambda)$. This can be written as

$$Po(\lambda) * Po(\mu) = Po(\lambda + \mu).$$

The mgfs of X and Y are $\psi_X(\theta) = \exp\{\lambda(e^\theta - 1)\}$ and $\psi_Y(\theta) = \exp\{\mu(e^\theta - 1)\}$, and so the mgf of $Z = X + Y$ is

$$\psi_Z(\theta) = \exp\{(\lambda + \mu)(e^\theta - 1)\},$$

which is the mgf of $Po(\lambda + \mu)$. This implies that $X + Y \sim Po(\lambda + \mu)$.

Example 13.4

Example

If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then show that $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. This can be written as

$$N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Example 13.4

Example

If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then show that $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. This can be written as

$$N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

The mgfs of X and Y are $\psi_X(\theta) = e^{\mu_1\theta + \sigma_1^2\theta^2/2}$ and $\psi_Y(\theta) = e^{\mu_2\theta + \sigma_2^2\theta^2/2}$, and so the mgf of $Z = X + Y$ is

$$\psi_Z(\theta) = e^{(\mu_1+\mu_2)\theta + (\sigma_1^2+\sigma_2^2)\theta^2/2},$$

which is the mgf of $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. This implies that $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Gamma function and distribution

Definition

The function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx,\ \ \alpha > 0$$

is called the gamma function.

Theorem

(i) $\Gamma(1) = 1$.

(ii) $\Gamma(1/2) = \sqrt{\pi}$.

(iii) $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ *for any real* $\alpha > 0$.

(iv) $\Gamma(n + 1) = n!$ *for any positive integer* n.

Proof

(i)

$$\Gamma(1) = \int_0^\infty e^{-x}dx = [-e^{-x}]_0^\infty = 1.$$

(ii) Using the change of variables $y = \sqrt{2x}$ with $dx = ydy$, we have

$$\begin{aligned}\Gamma(1/2) &= \int_0^\infty x^{-1/2}e^{-x}dx = \frac{1}{\sqrt{2}}\int_0^\infty y^{-1}e^{-y^2/2}2ydy \\ &= \sqrt{2}\int_0^\infty e^{-y^2/2}dy = \sqrt{\pi} \times \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-y^2/2}dy = \sqrt{\pi}.\end{aligned}$$

(iii) By the integration by parts,

$$\Gamma(\alpha+1) = \int_0^\infty x^\alpha e^{-x}dx = [-x^\alpha e^{-x}]_0^\infty + \alpha\int_0^\infty x^{\alpha-1}e^{-x}dx = \alpha\Gamma(\alpha).$$

Definition

Let $\alpha > 0$ and $\beta > 0$. A random variable X follows the gamma distribution with shape parameter α and scale parameter β, $X \sim Ga(\alpha, \beta)$ in short, if X has pdf

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}.$$

If $\alpha = 1$ and $\beta = 1/\lambda$ for some $\lambda > 0$, then the pdf of $Ga(1, 1/\lambda)$ is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x > 0,$$

which is the pdf of $Ex(\lambda)$. Hence,

$$Ga(1, 1/\lambda) = Ex(\lambda).$$

Example 13.5

Example

Find the mgf of $Ga(\alpha, \beta)$ and then calculate its mean and variance.

Example 13.5

Example

Find the mgf of $Ga(\alpha, \beta)$ and then calculate its mean and variance.

If $X \sim Ga(\alpha, \beta)$ and $\theta < 1/\beta$, then

$$E(e^{\theta X}) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{-(1-\beta\theta)x/\beta} dx.$$

Changing the variables $y = (1 - \beta\theta)x$ with $dx = (1 - \beta\theta)^{-1} dy$, we have

$$= (1 - \beta\theta)^{-\alpha} \times \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{-x/\beta} dx = (1 - \beta\theta)^{-\alpha}.$$

Hence, the mgf of $Ga(\alpha, \beta)$ is

$$\psi(\theta) = (1 - \beta\theta)^{-\alpha}, \ \theta < 1/\beta.$$

In addition,

$$E(X) = \psi'(0) = \alpha\beta, \ E(X^2) = \psi''(0) = \alpha(\alpha+1)\beta^2$$

and

$$\mathrm{Var}(X) = E(X^2) - \{E(X)\}^2 = \alpha\beta^2.$$

What is the variance of $Ex(\lambda)$?

Example 13.6

Example

If $X \sim Ga(\alpha_1, \beta)$ and $Y \sim Ga(\alpha_2, \beta)$ are independent, then show that $X + Y \sim Ga(\alpha_1 + \alpha_2, \beta)$. This can be written as

$$Ga(\alpha_1, \beta) * Ga(\alpha_2, \beta) = Ga(\alpha_1 + \alpha_2, \beta).$$

Example 13.6

Example

If $X \sim Ga(\alpha_1, \beta)$ and $Y \sim Ga(\alpha_2, \beta)$ are independent, then show that $X + Y \sim Ga(\alpha_1 + \alpha_2, \beta)$. This can be written as

$$Ga(\alpha_1, \beta) * Ga(\alpha_2, \beta) = Ga(\alpha_1 + \alpha_2, \beta).$$

The mgfs of X and Y are $\psi_X(\theta) = (1 - \beta\theta)^{-\alpha_1}$ and $\psi_Y(\theta) = (1 - \beta\theta)^{-\alpha_2}$, and so the mgf of $Z = X + Y$ is

$$\psi_Z(\theta) = (1 - \beta\theta)^{-(\alpha_1 + \alpha_2)},$$

which is the mgf of $Ga(\alpha_1 + \alpha_2, \beta)$. This implies that $X + Y \sim Ga(\alpha_1 + \alpha_2, \beta)$

Recap: regeneration property

Theorem

We have

$$
\begin{aligned}
&Bin(n,p) * Bin(m,p) = Bin(n+m,p),\\
&Po(\lambda) * Po(\mu) = Po(\lambda+\mu),\\
&N(\mu_1,\sigma_1^2) * N(\mu_2,\sigma_2^2) = N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2),\\
&Ga(\alpha_1,\beta) * Ga(\alpha_2,\beta) = Ga(\alpha_1+\alpha_2,\beta).
\end{aligned}
$$

Joint MGF

Definition

Let (X, Y) be a random vector such that $E(e^{\theta_1 X + \theta_2 Y}) < \infty$ for all $(\theta_1, \theta_2) \in A$ for some open set A of $\mathbb{R}^2$ containing the origin $(0, 0)$. Then the function

$$\psi(\theta_1, \theta_2) = E(e^{\theta_1 X + \theta_2 Y}), \ (\theta_1, \theta_2) \in A$$

is called the joint mgf of (X, Y).

By definition,

$$\psi(\theta_1, 0) = \underbrace{\psi_X(\theta_1)}_{\text{marginal mgf of } X} \quad \text{and} \quad \psi(0, \theta_2) = \underbrace{\psi_Y(\theta_2)}_{\text{marginal mgf of } Y} .$$

Some properties of joint mgf

- $$\frac{\partial^{j+k}}{\partial\theta_1^j \partial\theta_2^k}\psi(\theta_1,\theta_2)\Big|_{(\theta_1,\theta_2)=(0,0)} = E(X^j Y^k).$$
- The joint mgf uniquely determines the joint cdf.
- If X and Y are independent, then

  $$\psi(\theta_1,\theta_2) = E(e^{\theta_1 X+\theta_2 Y}) = E(e^{\theta_1 X})E(e^{\theta_2 Y}) = \psi_X(\theta_1)\psi_Y(\theta_2).$$

  The converse is also true.

Theorem

Suppose that (X, Y) *has joint mgf* $\psi(\theta_1, \theta_2)$*. Then* X *and* Y *are independent if and only if*

$$\psi(\theta_1, \theta_2) = \psi_X(\theta_1)\psi_Y(\theta_2)$$

on some open set of $\mathbb{R}^2$ *containing the origin.*