Lecture 04: Moments and Moment Generating Functions

Mathematical Statistics I, MATH 60061/70061

Thursday September 9, 2021

Reference: Casella & Berger, 2.3

Expectation of $g(X)$

To find $E(g(X))$:

- Find the distribution of the random variable $g(X)$
- Use the definition of expectation

The **law of the unconscious statistician** (LOTUS) is a powerful alternative. If X is a random variable and g is a function from $\mathbb{R}$ to $\mathbb{R}$, then

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

provided that the integral or sum exists.

Variance and standard deviation

One important application of LOTUS is for finding the variance of a random variable, a summary for the *spread* of the distribution.

The **variance** of a random variable X is

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

The square root of the variance is called the **standard deviation** (SD):

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}.$$

Properties of variance

For any random variable X with finite variance and any constant c,

$$\begin{aligned} \mathrm{Var}(X+c) &= \mathrm{Var}(X), \\ \mathrm{Var}(cX) &= c^2\mathrm{Var}(X). \end{aligned}$$

Variance is the expectation of the nonnegative random variable $(X - EX)^2$, so $\mathrm{Var}(X) \geq 0$, with equality if and only if $P(X = a) = 1$ for some constant a.

Equivalent expression for variance

For any random variable X,

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

Let $\mu = EX$. Expanding $(X-\mu)^2$ and using linearity, the variance of X is

$$\begin{aligned} E(X-\mu)^2 &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2. \end{aligned}$$

$E(X^2)$ is called the second moment of X.

Moments

Let X be a random variable with mean μ and variance σ^2. For any positive integer n, the nth **moment** of X is $E(X^n)$, the nth **central moment** is $E((X-\mu)^n)$, and the nth **standardized moment** is $E\left(\left(\frac{X-\mu}{\sigma}\right)^n\right)$, where "if it exists" is left implicit.

- Mean: the first moment
- Variance: the second central moment

Bernoulli variance

Let $X \sim \mathrm{Bern}(p)$. What is the variance of X?

$$E(X^2) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p.$$

So, the variance of X is given by

$$\mathrm{Var}(X) = E(X^2) - (EX)^2 = p - p^2 = p(1 - p).$$

Binomial variance

Let $X \sim \mathrm{Bin}(n, p)$. What is the variance of X?

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} \\
&= n \sum_{x=1}^{n} x \binom{n-1}{x-1} p^x (1-p)^{n-x} \quad [x^2 \binom{n}{x} = xn \binom{n-1}{x-1}] \\
&= n \sum_{y=0}^{n-1} (y+1) \binom{n-1}{y} p^{y+1} (1-p)^{n-1-y} \\
&= np \sum_{y=0}^{n-1} y \binom{n-1}{y} p^y (1-p)^{n-1-y} + np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y}.
\end{aligned}
$$

The first sum is equal to $(n-1)p$ (since it is the mean of a $\mathrm{Bin}(n-1, p)$), and the second sum is equal to 1. Hence,

$$E(X^2) = n(n-1)p^2 + np,$$

and

$$\mathrm{Var}(X) = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

Poisson random variable

A random variable X has the Poisson distribution with parameter λ, where $\lambda > 0$, if the PMF of X is

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!},$$

for $x = 0, 1, 2, \dots$.

This is a valid PMF because of the Taylor series $\sum_{x=0}^{\infty} \lambda^x / x! = e^{\lambda}$.

Poisson expectation

Let $X \sim \mathrm{Pois}(\lambda)$. The expected value of X is

$$\begin{aligned} E(X) &= e^{-\lambda} \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} \\ &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda. \end{aligned}$$

Poisson variance

Let $X \sim \text{Pois}(\lambda)$. By LOTUS,

$$E(X^2) = \sum_{x=0}^{\infty} x^2 P(X = x) = e^{-\lambda} \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!}.$$

Differentiating the familiar series

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$$

w.r.t. λ:

$$\sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{x!} = e^{\lambda},$$

$$\sum_{x=1}^{\infty} x \frac{\lambda^{x}}{x!} = \lambda e^{\lambda}.$$

Poisson variance, continued

Repeat:

$$\sum_{x=1}^{\infty} x^2 \frac{\lambda^{x-1}}{x!} = e^{\lambda} + \lambda e^{\lambda} = e^{\lambda}(1+\lambda),$$

$$\sum_{x=1}^{\infty} x^2 \frac{\lambda^{x}}{x!} = e^{\lambda}\lambda(1+\lambda).$$

Finally,

$$E(X^2) = e^{-\lambda} \sum_{x=0}^{\infty} x^2 \frac{\lambda^{x}}{x!} = e^{-\lambda} e^{\lambda}\lambda(1+\lambda) = \lambda(1+\lambda),$$

so

$$\mathrm{Var}(X) = E(X^2) - (EX)^2 = \lambda(1+\lambda) - \lambda^2 = \lambda.$$

Poisson variance, continued

Repeat:

$$\sum_{x=1}^{\infty} x^2 \frac{\lambda^{x-1}}{x!} = e^{\lambda} + \lambda e^{\lambda} = e^{\lambda}(1+\lambda),$$
$$\sum_{x=1}^{\infty} x^2 \frac{\lambda^{x}}{x!} = e^{\lambda}\lambda(1+\lambda).$$

Finally,

$$E(X^2) = e^{-\lambda} \sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda}\lambda(1+\lambda) = \lambda(1+\lambda),$$

so

$$\mathrm{Var}(X) = E(X^2) - (EX)^2 = \lambda(1+\lambda) - \lambda^2 = \lambda.$$

The mean and variance of a $\mathrm{Pois}(\lambda)$ random variable are both equal to λ.

Moment generating function

A moment generating function is a function that encodes the **moments** of a distribution.

The **moment generating function** (MGF) of a random variable X is $M_X(t) = E(e^{tX})$, as a function of t, if this is finite on some open interval $(-a, a)$ containing 0. Otherwise we say the MGF of X does not exist.

- If X is continuous

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

- If X is discrete

$$M_X(t) = \sum_x e^{tx} P(X = x).$$

Bernoulli and Binomial MGFs

- For $X \sim \mathrm{Bern}(p)$,

- For $X \sim \mathrm{Bin}(n, p)$,

Bernoulli and Binomial MGFs

- For $X \sim \mathrm{Bern}(p)$,

$$M_X(t) = E(e^{tX}) = e^t p + e^0(1-p) = pe^t + 1 - p.$$

- For $X \sim \mathrm{Bin}(n, p)$,

Bernoulli and Binomial MGFs

- For $X \sim \mathrm{Bern}(p)$,

$$M_X(t) = E(e^{tX}) = e^t p + e^0(1-p) = pe^t + 1 - p.$$

- For $X \sim \mathrm{Bin}(n,p)$,

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} \\ &= \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x(1-p)^{n-x} \\ &= [pe^t + (1-p)]^n, \end{aligned}$$

where the last equality follows from the binomial formula

$$\sum_{x=0}^{n}\binom{n}{x}u^x v^{n-x} = (u+v)^n.$$

MGF is important

1. The MGF encodes the **moments** of a random variable.
   - We could obtain the moments by taking derivatives of the MGF and evaluating at 0.
   - With LOTUS, it requires taking sums/integrals to compute moments.
2. The MGF of a random variable determines its **distribution**, like the CDF and PMF/PDF.
   - If two random variables have the same MGF, they must have the same distribution.

Moments via derivatives of the MGF

Given the MGF of X, we can get the nth moment of X by evaluating the nth derivative of the MGF at 0: $E(X^n) = M_X^{(n)}(0)$.

Moments via derivatives of the MGF

Given the MGF of X, we can get the nth moment of X by evaluating the nth derivative of the MGF at 0: $E(X^n) = M_X^{(n)}(0)$.
Taylor expansion of $M_X(t)$ about 0 is

$$M_X(t) = \sum_{n=0}^{\infty} M^{(n)}(0)\frac{t^n}{n!},$$

and by definition of MGF we also have

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

Under certain technical conditions being satisfied ($E(e^{tX})$ is finite in an interval around 0),

$$M_X(t) = \sum_{n=0}^{\infty} E(X^n)\frac{t^n}{n!}.$$

Matching the coefficients, we get $E(X^n) = M_X^{(n)}(0)$.

Nonunique moments

Consider the two PDFs given by

$$f_1(x) = \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2}, \quad 0 \le x < \infty$$
$$f_2(x) = f_1(x)[1 + \sin(2\pi \log x)], \quad 0 \le x < \infty.$$

If $X_1 \sim f_1(x)$, then the nth moment of X_1 is

$$\begin{aligned}
E(X_1^n) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{n-1} e^{-(\log x)^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{ny - y^2/2} dy \qquad\qquad [y = \log x] \\
&= \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(y-n)^2/2} dy \\
&= e^{n^2/2}.
\end{aligned}$$

Suppose that $X_2 \sim f_2(x)$, we have

$$\begin{aligned}
E(X_2^n) &= \int_0^\infty x^n f_1(x)[1+\sin(2\pi \log x)]dx \\
&= E(X_1^n) + \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{n-1} e^{-(\log x)^2/2} \sin(2\pi \log x) dx \\
&= E(X_1^n) + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{ny} e^{-y^2/2} \sin(2\pi y) dy \\
&= E(X_1^n) + \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(y-n)^2/2} \sin(2\pi y) dy \\
&= E(X_1^n) + \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-s^2/2} \sin(2\pi(s+n)) ds \\
&= E(X_1^n) + \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-s^2/2} \sin(2\pi s) ds \\
&= E(X_1^n)
\end{aligned}$$

since $e^{-s^2/2}\sin(2\pi s)$ is an odd function.

This shows that X_1 and X_2 have the same moments of order $n = 1, 2, \ldots$, but they have different distributions.

Determining a distribution

Let $F_X(x)$ and $F_Y(y)$ be two CDFs all of whose moments exist.

1. If X and Y have *bounded support*, then $F_X(u) = F_Y(u)$ for all u if and only if $E(X^n) = E(Y^n)$ for all integers $n = 0, 1, 2, \ldots$.
2. If the moment generating functions *exist* and $M_X(t) = M_Y(t)$ for all t in some neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u.

MGF of location-scale transformation

If X has MGF $M_X(t)$, then for any constants a and b, the MGF of the random variable $a + bX$ is given by

$$M_{a+bX}(t) = e^{at} M_X(bt).$$

MGF of location-scale transformation

If X has MGF $M_X(t)$, then for any constants a and b, the MGF of the random variable $a + bX$ is given by

$$M_{a+bX}(t) = e^{at} M_X(bt).$$

By definition,

$$\begin{aligned} M_{a+bX}(t) &= E\left(e^{(a+bX)t}\right) \\ &= E\left(e^{(bX)t} e^{at}\right) \\ &= e^{at} E\left(e^{(bt)X}\right) \\ &= e^{at} M_X(bt). \end{aligned}$$

Normal distribution

If Z is a standard Normal random variable $Z \sim \mathcal{N}(0, 1)$, then $X = \mu + \sigma Z$ is said to have the **Normal distribution** with mean μ and variance σ^2, for any real μ and σ^2 with $\sigma > 0$. We denote this by $X \sim \mathcal{N}(\mu, \sigma^2)$.

Expectation and variance of X:

$$E(\mu + \sigma Z) = E(\mu) + \sigma E(Z) = \mu,$$
$$\mathrm{Var}(\mu + \sigma Z) = \mathrm{Var}(\sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2.$$

The standardized version of X is

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

Normal MGF

The MGF of a standard Normal R.V. Z is

$$M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

After *completing the square*, we have

$$M_Z(t) = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} dz = e^{t^2/2},$$

since the $\mathcal{N}(t, 1)$ PDF integrates to 1.

Thus, the MGF of $X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$ is

$$M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{(\sigma t)^2/2} = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$