STSCI 5080
Probability Models and Inference

Lecture 11: Expected Values

October 2, 2018

Review of Definition

Definition

The expected value/expectation/mean of a random variable X is defined by

$$E(X) = \begin{cases} \sum_x xp(x) & \text{if } X \text{ is discrete with pmf } p(x) \\ \int_{-\infty}^{\infty} xf(x)dx & \text{if } X \text{ is continuous with pdf } f(x) \end{cases}$$

provided that the sum/integral is absolutely convergent.

Expectation of a function of a random variable

- Find the expectation of $Y = g(X)$.
- If X is discrete with pmf $p_X(x)$, Y is also discrete.
- The definition of $E(Y)$ is:

$$E(Y) = \sum_{y} y p_Y(y),$$

  where $p_Y(y)$ is the pmf of Y.
- Finding the pmf of Y is often not convenient!

Theorem

Let $Y = g(X)$*.*

(a) *If* X *is discrete with pmf* $p_X(x)$*, then*

$$E(Y) = \sum_x g(x)p_X(x)$$

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

(b) *If* X *is continuous with pdf* $f_X(x)$*, then*

$$E(Y) = \int_{-\infty}^{\infty} xf_X(x)dx$$

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

Proof of Case (a)

The definition of $E(Y)$ is

$$E(Y) = \sum_y y p_Y(y).$$

The pmf of Y is by definition

$$p_Y(y) = P(Y = y) = P(g(X) = y) = \sum_{x:g(x)=y} p_X(x).$$

Define the indicator function

$$I(x,y) = \begin{cases} 1 & \text{if } g(x) = y \\ 0 & \text{otherwise} \end{cases}.$$

Then

$$p_Y(y) = \sum_x p_X(x) I(x,y).$$

Now,

$$E(Y) = \sum_y y p_Y(y) = \sum_y y \sum_x g(x) I(x,y) = \sum_y \sum_x y p_X(x) I(x,y).$$

Because $I(x,y) = 0$ unless $g(x) = y$,

$$\begin{aligned}\sum_y \sum_x y p_X(x) I(x,y) &= \sum_y \sum_x g(x) p_X(x) I(x,y) \\ &= \sum_x \sum_y g(x) p_X(x) I(x,y) = \sum_x g(x) p_X(x) \sum_y I(x,y).\end{aligned}$$

For a fixed x, $I(x,y) = 1$ only if $y = g(x)$, and so $\sum_y I(x,y) = 1$. In conclusion, we have

$$E(Y) = \sum_x g(x) p_X(x).$$

$\square$

- The previous theorem extends to the case where g is non-negative without the condition that $\sum_x g(x)p_X(x) < \infty$ or $\int_{-\infty}^{\infty} g(x)f_X(x)dx < \infty$, namely, if X is discrete and $g \geq 0$,

$$E(Y) = \sum_x g(x)p_X(x)$$

  even when $\sum_x g(x)p_X(x) = \infty$.

- In particular, we always have

$$E(|X|) = \sum_x |x|p_X(x)$$

  including the case where $\sum_x |x|p_X(x) = \infty$.

Linear function

Theorem

$E(aX + b) = aE(X) + b.$

Linear function

Theorem

$E(aX + b) = aE(X) + b.$

Proof.

$$E(aX+b) = \int_{-\infty}^{\infty} (ax+b) f_X(x)dx = a \underbrace{\int_{-\infty}^{\infty} x f_X(x)dx}_{=E(X)} + b \underbrace{\int_{-\infty}^{\infty} f_X(x)dx}_{=1}$$

$$= aE(X) + b.$$

$\square$

Example 11.1

Example

If $Y \sim N(\mu, \sigma^2)$, then $E(Y) = \mu$.

Example 11.1

Example

If $Y \sim N(\mu, \sigma^2)$, then $E(Y) = \mu$.

By definition, $Y = \mu + \sigma X$ for some $X \sim N(0, 1)$, and so

$$E(Y) = \mu + \sigma E(X).$$

However,

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{xe^{-x^2/2}}_{\text{odd function}} \, dx = 0,$$

so that $E(Y) = \mu$.

Example 11.2

In general, $E(g(X)) \neq g(E(X))$.

Example

If $X \sim U[1,2]$, then compare $1/E(X)$ and $E(1/X)$.

Example 11.2

In general, $E(g(X)) \neq g(E(X))$.

Example

If $X \sim U[1,2]$, then compare $1/E(X)$ and $E(1/X)$.

The pdf of X is

$$f_X(x) = \begin{cases} 1 & \text{if } 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}.$$

On one hand,

$$E(X) = \int_1^2 x dx = \left[\frac{x^2}{2}\right]_1^2 = \frac{3}{2},$$

so that $1/E(X) = 2/3 \approx 0.67$. On the other hand,

$$E(1/X) = \int_1^2 \frac{1}{x} dx = [\log x]_1^2 = \log 2 - \log 1 = \log 2 \approx 0.69 > 1/E(X).$$

Moments

Definition

For each $k = 1, 2, \dots,$

$$E(X^k)$$

is called the k-th moment of X, provided that $E(|X|^k) < \infty$.

$E(X)$: 1st moment, $E(X^2)$: second moment, etc.

Theorem

If $E(|X|^k) < \infty$ *for some* $k = 2, 3, \ldots,$ *then* $E(|X|^j) < \infty$ *for any* $j = 1, \ldots, k-1.$

Proof.

The proof follows from the fact that

$$|x|^j \leq 1 + |x|^k,$$

which implies that $E(|X|^j) \leq 1 + E(|X|^k)$. $\square$

For example, if $E(X^2) < \infty$, then $E(|X|) < \infty$.

Functions of random vectors

Theorem

Let $(X_1, \ldots, X_n)$ *be a random vector and* $Y = g(X_1, \ldots, X_n)$.

(a) *If* $(X_1, \ldots, X_n)$ *is discrete with joint pmf* $p(x_1, \ldots, x_n)$,

$$E(Y) = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \ldots, x_n) p(x_1, \ldots, x_n)$$

provided that $\sum_{x_1} \cdots \sum_{x_n} |g(x_1, \ldots, x_n)| p(x_1, \ldots, x_n) < \infty$.

(b) *If* $(X_1, \ldots, X_n)$ *is continuous with joint pdf* $f(x_1, \ldots, x_n)$,

$$E(Y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \cdots dx_n$$

provided that $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(x_1, \ldots, x_n)| f(x_1, \ldots, x_n) dx_1 \cdots dx_n < \infty$.

Linear function

Theorem

$$E\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i E(X_i) + b.$$

In particular,

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i).$$

Example 11.3

Example

For $Y \sim Bin(n,p)$, find $E(Y)$.

Example 11.3

Example

For $Y \sim Bin(n,p)$, find $E(Y)$.

By definition, $Y = X_1 + \cdots + X_n$ for independent Bernoulli trials $X_1, \ldots, X_n$ with success probability p. Since

$$E(X_i) = p,$$

we have

$$E(Y) = E(X_1) + \cdots + E(X_n) = np.$$

Example 11.4

Example

If $X, Y \sim U[0, 1]$ i.i.d., then find $E(|X - Y|)$.

Example 11.4

Example

If $X, Y \sim U[0,1]$ i.i.d., then find $E(|X-Y|)$.

First of all, since X and Y are independent and continuous, the vector (X, Y) is continuous with joint pdf

$$f(x,y) = \begin{cases} 1 & \text{if } 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

$$\begin{aligned} E(|X-Y|) &= \int_0^1 \int_0^1 |x-y| dxdy \\ &= \int_0^1 \int_0^y (y-x) dxdy + \int_0^1 \int_y^1 (x-y) dxdy \\ &= \frac{1}{3}. \end{aligned}$$

Example 11.5

In some cases, it is easier to find the pmf/pdf of $Y = g(X_1, \ldots, X_n)$.

Example

If $X_1, \ldots, X_n \sim U[0, \theta]$ i.i.d., find $E(X_{(n)})$, where $X_{(n)} = \max_{1 \le i \le n} X_i$.

Example 11.5

In some cases, it is easier to find the pmf/pdf of $Y = g(X_1, \ldots, X_n)$.

Example

If $X_1, \ldots, X_n \sim U[0, \theta]$ i.i.d., find $E(X_{(n)})$, where $X_{(n)} = \max_{1 \leq i \leq n} X_i$.

The pdf of $Y = \max_{1 \leq i \leq n} X_i$ is

$$f_Y(y) = nf(y)F(y)^{n-1}.$$

In this case, $f(x) = 1/\theta$ and $F(x) = x/\theta$ for $0 \leq x \leq \theta$, and so

$$f_Y(y) = \frac{ny^{n-1}}{\theta^n}$$

for $0 \leq y \leq \theta$, and $f_Y(y) = 0$ elsewhere. So

$$E(Y) = \frac{1}{\theta^n} \int_0^\theta ny^n dy = \frac{n}{n+1}\theta.$$

Expectation under independence

Theorem

If X and Y are independent random variables, then

$$E\{g(X)h(Y)\} = E\{g(X)\}E\{h(Y)\}$$

provided that $E\{|g(X)|\} < \infty$ and $E\{|h(Y)|\} < \infty$.

Proof?

Variance

- mean = weighted average of possible values.
- variance = a measure on how a random variable fluctuates around its mean.
- If the variance is high (low), then it is more likely that the variable is quite different from (close to) its mean.

Definition

Definition

The variance of a random variable X is defined by

$$\mathrm{Var}(X) = E[\{X - E(X)\}^2]$$

provided that $E(X^2) < \infty$.

If $X \equiv c$ (constant), then what is $\mathrm{Var}(X)$?

Theorem

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

Proof.

Because $\{X - E(X)\}^2 = X^2 - 2E(X)X + \{E(X)\}^2$, we have

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

□

Theorem

$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

Proof.

Let $Y = aX + b$. We have

$$E(Y) = aE(X) + b,$$

and so

$$Y - E(Y) = aX + b - \{aE(X) + b\} = a\{X - E(X)\}.$$

$\square$

Example 11.6

Example

Find the variances of the following random variables: (a) X is Bernoulli with success probability p. (b) $X \sim N(\mu, \sigma^2)$.

Example 11.6

Example

Find the variances of the following random variables: (a) X is Bernoulli with success probability p. (b) $X \sim N(\mu, \sigma^2)$.

(a) $\mathrm{Var}(X) = p(1-p)$. (b) $\mathrm{Var}(X) = \sigma^2$.

Some inequalities

Theorem (Markov's inequality)

For any random variable X,

$$P(|X| \geq t) \leq \frac{E(|X|)}{t}$$

for any $t > 0$.

Proof.

Pick any $t > 0$, and define a random variable

$$Y = \begin{cases} 1 & \text{if } |X| \geq t \\ 0 & \text{otherwise} \end{cases}.$$

Then $|X| \geq tY$, and so

$$E(|X|) \geq tE(Y) = tP(Y = 1) = tP(|X| \geq t).$$

$\square$

Theorem

If $E(|X|) = 0$, then $P(X = 0) = 1$.

Proof.

We have

$$P(|X| > 0) = P\left(\bigcup_{k=1}^{\infty}\{|X| \geq 1/k\}\right) \leq \sum_{k=1}^{\infty} P(|X| \geq 1/k).$$

But Markov's inequality shows that

$$P(|X| \geq 1/k) = kE(|X|) = 0,$$

so that $P(|X| > 0) = 0$, which implies that $P(|X| = 0) = 1$. $\square$

Theorem (Cauchy-Schwarz inequality)

For any random variables X and Y (that may not have finite second moments),

$$E(|XY|) \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

Proof

We may assume $0 < E(X^2) < \infty$ and $0 < E(Y^2) < \infty$ since otherwise there is nothing to prove. In addition, we may assume that $X \geq 0$ and $Y \geq 0$. Under this assumption, we also have $E(XY) < \infty$ since

$$XY \leq \frac{X^2}{2} + \frac{Y^2}{2}.$$

Now, for any $a \in \mathbb{R}$,

$$E\{(X - aY)^2\} \geq 0, \tag{*}$$

and the left hand side is

$$E(X^2) - 2aE(XY) + a^2E(Y^2),$$

which is minimized at

$$a = \frac{E(XY)}{E(Y^2)}.$$

Substituting this into (*), we have

$$E(X^2) - \frac{\{E(XY)\}^2}{E(Y^2)} \geq 0.$$

Definition

Definition

The covariance between two random variables X and Y is defined by

$$\mathrm{Cov}(X, Y) = E[\{X - E(X)\}\{Y - E(Y)\}]$$

provided that $E(X^2) < \infty$ and $E(Y^2) < \infty$.