Lecture 4: Expectations

The expected value, also called the expectation or mean, of a random variable is its average value weighted by its probability distribution.

Definition 2.2.1.

The expected value or mean of a random variable $g(X)$ is

$$E[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ has pdf } f_X \\ \sum_x g(x) f_X(x) & \text{if } X \text{ has pmf } f_X \end{cases}$$

provided that the integral or the sum exists (is finite); otherwise we say that the expected value of $g(X)$ does not exist.

- The expected value is a number that summarizes a typical, middle, or expected value of an observation of the random variable.
- If $g(X) \geq 0$, then $E[g(X)]$ is always defined except that it may be ∞.
- For any $g(X)$, its expected value exists iff $E|g(X)| < \infty$.
- The expectation is associated with the distribution of X, not with X.

Example

X has distribution

x	-2	-1	0	1	2	3
$f_X(x)$	0.1	0.2	0.1	0.2	0.3	0.1

$$E(X) = \sum x f_X(x) = -2 \times 0.1 - 0.2 + 0.2 + 2 \times 0.3 + 3 \times 0.1 = 0.7$$

$$E(X^2) = \sum x^2 f_X(x) = 4 \times 0.1 + 0.2 + 0.2 + 4 \times 0.3 + 9 \times 0.1 = 2.9$$

$Y = g(X) = X^2$ has distribution

y	0	1	4	9
$f_Y(y)$	0.1	0.4	0.4	0.1

$$E(Y) = E(X^2) = \sum y f_Y(y) = 0.4 + 4 \times 0.4 + 9 \times 0.1 = 2.9$$

Example 2.2.2.

Suppose $X \geq 0$ has pdf $f_X(x) = \lambda^{-1} e^{-x/\lambda}$, where $\lambda > 0$ is a constant. Then

$$E(X) = \int_0^\infty \lambda^{-1} x e^{-x/\lambda} dx = -x e^{-x/\lambda} \Big|_0^\infty + \int_0^\infty e^{-x/\lambda} dx = \lambda$$

In some cases, the calculation of the expectation requires some derivation.

Example 2.2.3.

Let X be a discrete random variable with pmf

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, ..., n,$$

where n is a fixed integer and $0 < p < 1$ is a fixed constant.
For any n and p,

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1.$$

By definition and the fact that $x\binom{n}{x} = n\binom{n-1}{x-1}$, we obtain

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^{n} n \binom{n-1}{x-1} p^x (1-p)^{n-x} = np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} \\ &= np \end{aligned}$$

Example 2.2.4 (Nonexistence of expectation)

Suppose that X has the following pdf:

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

Note that

$$\begin{aligned} E|X| &= \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx \\ &= \frac{2}{\pi} \lim_{M\to\infty} \int_0^M \frac{x}{1+x^2} dx = \frac{2}{\pi} \lim_{M\to\infty} \int_0^M \frac{1}{1+x^2} d(x^2/2) \\ &= \frac{2}{\pi} \lim_{M\to\infty} \frac{\log(1+x^2)}{2} \bigg|_0^M = \frac{2}{\pi} \lim_{M\to\infty} \frac{\log(1+M^2)}{2} = \infty \end{aligned}$$

Thus, neither X nor $|X|$ has an expectation.
We should not wrongly think

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} dx = \lim_{M\to\infty} \int_{-M}^{M} \frac{x}{\pi(1+x^2)} dx = \lim_{M\to\infty} 0 = 0$$

Linear operation

Taking expectation is a linear operation in the sense that
$E(c) = c$ for any constant c;
$E(aX) = aE(X)$ for any constant a and random variable X;
$E(X+Y) = E(X)+E(Y)$ for any random variables X and Y.

These are special cases of the following theorem.

Theorem 2.2.5

Let X and Y be random variables whose expectations exist, and let a, b, and c be constants.

a. $E(aX+bY+c) = aE(X)+bE(Y)+c$.
b. If $X \geq Y$, then $E(X) \geq E(Y)$.

Proof.

The proof of Theorem 2.2.5 is easy if $X = g_1(Z)$ and $Y = g_2(Z)$ for a random variable Z and some functions g_1 and g_2, and Z has a pdf or pmf.

In general, the proof of property a is not simple (a topic in stat 709).

Example 2.2.6 (Minimizing distance)

The mean of a random variable X is a good guess (predict) of X.
Suppose that we measure the distance between X and a constant b by $(X-b)^2$.
The closer b is to X, the smaller this quantity is.
We want to find a value b that minimizes the average $E(X-b)^2$.
If $\mu = E(X)$, then

$$E(X-\mu)^2 = \min_{b \in \mathscr{R}} E(X-b)^2$$

To show this, note that

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

There is an alternative proof using $E(X-b)^2 = E(X^2) - 2bE(X) + b^2$.

Nonlinear functions

When calculating expectations of nonlinear functions of X, we can proceed in one of two ways.

- Using the distribution of X, we can calculate

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \qquad \text{or} \qquad \sum_x g(x) f_X(x)$$

- We can find the pdf or pmf of $Y = g(X)$, f_Y, and then use

$$E[g(X)] = E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy \qquad \text{or} \qquad \sum_y y f_Y(y)$$

Example 2.2.7.

Suppose that X has a uniform pdf on (0,1),

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Consider $g(X) = -\log X$.

$$E[g(X)] = \int_0^1 -\log(x) dx = (x - x \log x)\Big|_0^1 = 1$$

On the other hand, we can use Theorem 2.1.5 to obtain the pdf of $Y = -\log X$:

$$f_Y(y) = f_X(e^{-y})|e^{-y}| = e^{-y}, \quad y > 0$$

Then

$$E[g(X)] = E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^{\infty} y e^{-y} dy = 1$$

Example

Suppose that X has the standard normal pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathscr{R}$$

Consider $g(X) = X^2$.
By Example 2.1.9, $Y = X^2$ has the chi-square pdf

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

Then

$$E(X^2) = E(Y) = \int_0^{\infty} y f_Y(y) dy = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \sqrt{y} e^{-y/2} dy = 1$$

Theorem

Let F be the cdf of a random variable X. Then

$$E|X| = \int_0^\infty [1 - F(x)]dx + \int_{-\infty}^0 F(x)dx$$

and $E|X| < \infty$ iff both integrals are finite. In the case where $E|X| < \infty$,

$$E(X) = \int_0^\infty [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx$$

Proof.

We only consider the case where X has a pdf or pmf, and give a proof when X has a pdf (the proof for the case where X has a pmf is similar).

When X has a pdf f,

$$\int_0^\infty [1 - F(x)]dx = \int_0^\infty \int_x^\infty f(t)dtdx = \int_0^\infty \int_0^t f(t)dxdt = \int_0^\infty tf(t)dt$$

$$\int_0^\infty F(x)dx = \int_{-\infty}^0 \int_{-\infty}^x f(t)dtdx = \int_{-\infty}^0 \int_t^0 f(t)dxdt = \int_{-\infty}^0 -tf(t)dt$$

These calculations are valid regardless of whether the integrals are finite or not.

By definition,

$$E|X| = \int_{-\infty}^{\infty} |t| f(t)dt = \int_0^{\infty} tf(t)dt + \int_{-\infty}^{0} -tf(t)dt$$
$$= \int_0^{\infty} [1 - F(x)]dx + \int_{-\infty}^{0} F(x)dx$$

Since both integrals ≥ 0, $E|X| < \infty$ iff both integrals are finite.
If $E|X| < \infty$, then both integrals are finite and

$$E(X) = \int_{-\infty}^{\infty} tf(t)dt = \int_0^{\infty} tf(t)dt + \int_{-\infty}^{0} tf(t)dt$$
$$= \int_0^{\infty} [1 - F(x)]dx - \int_{-\infty}^{0} F(x)dx$$

Property

For any random variable X,

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n).$$

Then $E|X| < \infty$ iff the sum is finite.

Using the previous theorem, we can prove this result regardless of whether X has a pdf (pmf) or not:

$$\begin{aligned}
E|X| &= \int_0^\infty [1-F(x)]dx + \int_{-\infty}^0 F(x)dx = \int_0^\infty [1-F(x)]dx + \int_0^\infty F(-x)dx \\
&= \int_0^\infty [1-F(x)+F(-x)]dx = \int_0^\infty [P(X>x)+P(X\le -x)]dx \\
&\le \int_0^\infty [P(X\ge x)+P(X\le -x)]dx = \int_0^\infty P(|X|\ge x)dx \\
&= \sum_{n=0}^\infty \int_n^{n+1} P(|X|\ge x)dx \le \sum_{n=0}^\infty \int_n^{n+1} P(|X|\ge n)dx \\
&= \sum_{n=0}^\infty P(|X|\ge n) = 1 + \sum_{n=1}^\infty P(|X|\ge n)
\end{aligned}$$

Similarly,

$$\begin{aligned}
E|X| &\ge \int_0^\infty [P(X>x)+P(X<-x)]dx = \int_0^\infty P(|X|>x)dx \\
&= \sum_{n=0}^\infty \int_n^{n+1} P(|X|>x)dx \ge \sum_{n=0}^\infty P(|X|\ge n+1) = \sum_{n=1}^\infty P(|X|\ge n)
\end{aligned}$$

Moments

The various moments of a random variable are an important class of expectations.

Definition 2.3.1.

- For each positive integer n, the nth moment of a random variable X (or F_X) is $E(X^n)$.
- For each positive integer n, the nth central moment of X is $E(X-\mu)^n$, where $\mu = E(X)$.
- For each constant $p > 0$, the pth absolute moment of X is $E(|X|^p)$.

Property.

For $p > 0$ and any random variable X,

$$E|X|^p < \infty \quad \text{iff} \quad \sum_{n=1}^{\infty} n^{p-1} P(|X| \geq n) < \infty$$

A consequence is, if the pth absolute moment (moment or central moment) exists and $q < p$, then the pth absolute moment (moment or central moment) also exists.

Proof.

From the discussion in the last lecture,

$$E|X|^p \le \int_0^\infty P(|X|^p \ge x)dx = p\int_0^\infty P(|X|^p \ge y^p)y^{p-1}dy$$

$$= p\sum_{n=0}^\infty \int_n^{n+1} P(|X|^p \ge y^p)y^{p-1}dy \le p\sum_{n=0}^\infty (n+1)^{p-1}P(|X| \ge n)$$

$$\le p + 2^{p-1}p\sum_{n=1}^\infty n^{p-1}P(|X| \ge n)$$

where the first equality is from changing variable $x = y^p$.

Similarly,

$$E|X|^p \ge \int_0^\infty P(|X|^p > x)dx = p\sum_{n=0}^\infty \int_n^{n+1} P(|X|^p > y^p)y^{p-1}dy$$

$$\ge p\sum_{n=0}^\infty n^{p-1}P(|X| \ge n+1) = p\sum_{n=2}^\infty (n-1)^{p-1}P(|X| \ge n)$$

$$\ge p2^{-p}\sum_{n=2}^\infty n^{p-1}P(|X| \ge n)$$

Definition 2.3.2.

The second central moment of a random variable X is called its variance and denoted by $\mathrm{Var}(X) = E(X-\mu)^2$.
The standard deviation of X is defined as $\sqrt{\mathrm{Var}(X)}$.

Sometime the following result is useful:

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

Measures of spread

- The variance gives a measure of the degree of spread of a distribution around its mean (center).
- A large variance means X is more variable.
- The unit on the variance is the square of the original unit.
- The standard deviation has the same qualitative interpretation as the variance, but its unit is the same as the original unit.
- If $\mathrm{Var}(X) = 0$, then $P(X = E(X)) = 1$.
  This is actually a special case of the following result:

If $g(X) \geq 0$ and $E[g(X)] = 0$, then $P(g(X) = 0) = 1$.

The result can be proved in general, but we only consider the case where X has a pdf or pmf.

Since $g(X) \geq 0$ and $E[g(X)] = 0$, X cannot have a pdf unless $g(x) = 0$ for all x.

If X is discrete and have positive probabilities p_j's to take two values x_j's, then

$$0 = E[g(X)] = \sum_k g(x_k)p_k \geq g(x_j)p_j$$

implies that $g(x_j) = 0$ for all j, i.e., $P(g(X) = 0) = 1$.

Theorem 2.3.4.

If X has a finite variance, then for any constants a and b,

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

Proof.

By definition,

$$\begin{aligned} \mathrm{Var}(aX + b) &= E[(aX + b) - E(aX + b)]^2 = E[aX - aE(X)]^2 \\ &= a^2 E[X - E(X)]^2 = a^2 \mathrm{Var}(X) \end{aligned}$$

Example 2.3.3 (the exponential distribution)

Suppose that X has the following pdf:

$$f_X(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where $\lambda > 0$ is a constant.

$$\mu = E(X) = \int_0^\infty \frac{1}{\lambda} x e^{-x/\lambda} dx = -xe^{-x/\lambda}\Big|_0^\infty + \int_0^\infty e^{-x/\lambda} dx = \lambda$$

$$\begin{aligned} \mathrm{Var}(X) &= E(X-\lambda)^2 = E(X^2) - \lambda^2 = \int_0^\infty \frac{x^2}{\lambda} e^{-x/\lambda} dx - \lambda^2 \\ &= -x^2 e^{-x/\lambda}\Big|_0^\infty + \int_0^\infty 2xe^{-x/\lambda} dx - \lambda^2 \\ &= 2\lambda \int_0^\infty \frac{1}{\lambda} x e^{-x/\lambda} dx - \lambda^2 \\ &= 2\lambda^2 - \lambda^2 = \lambda^2 \end{aligned}$$

Example 2.3.5 (binomial distribution)

Suppose that X is discrete with pmf

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n,$$

where n is a fixed positive integer and $0 < p < 1$ is a fixed constant.
We have obtained that $E(X) = np$.
Using the identity

$$x^2 \binom{n}{x} = x \frac{n!}{(x-1)!(n-x)!} = xn \binom{n-1}{x-1}$$

we obtain

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

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