Review B: Expectations

Professor Ji-Woong Chung
Korea University

This lecture note is based on Thomas Wiemann's.

Recap

In Part A of the probability theory review, we discussed probability distributions:

- CDFs and pdfs (or pmfs) fully characterize a random variable.
- Joint CDFs and joint pdfs (or pmfs) fully characterize relationships between random variables.

But we may not always require a full characterization. Often, we are content with knowing about key features of a random variable that partly characterize it or its relation to other random variables.

- Recall the returns to education example where we were interested in, e.g.,

$$\tau_{ATT} = E[Y_i(1) - Y_i(0) | D = 1],$$

  and not the conditional distribution of $Y_i(1) - Y_i(0)$ given $D = 1$.

The key concept we will cover in this lecture is expectations.

Outline

Features of Probability Distributions
- Expectation
- Variance
- Covariance
- Correlation

Features of Conditional Probability Distributions
- Conditional Expectation
- Conditional Variance

Mean Independence

Outline

Outline

Expectation

Definition (Expected Value)

The expected value of a random variable X is defined as

$$E_X[X] = \begin{cases} \sum_{x \in \text{supp } X} x f_X(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f_X(x) dx, & \text{if } X \text{ is continuous.} \end{cases}$$

The expected value is a one-number summary of a random variable.

- X is a random variable but $E_X[X]$ is a number.
- Considered a measure of central tendency.

We say that the expectation of X exists if $E[|X|] < \infty$.

- In this course, we always (implicitly) assume that expectations exist.

Note: You may encounter various other names for the expectation, including "mean" or "first moment," as well as alternative notations. For example, we may also express the expectation as $E_X[X] = \int x dF(x)$.

Expectation (Contd.)

Example: Consider tossing a fair coin twice. Let X be the number of heads. Then

$$f_X(x) = \begin{cases} \frac{1}{4}, & \text{if } x = 0, \\ \frac{1}{2}, & \text{if } x = 1, \\ \frac{1}{4}, & \text{if } x = 2, \\ 0, & \text{otherwise}, \end{cases}$$

and the expected number of heads is

$$E_X[X] = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1.$$

Expectation (Contd.)

Example Consider $X \sim U(a, b)$. Then

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{for } x \in [a, b], \\ 0, & \text{otherwise.} \end{cases}$$

and we have

$$E_X[X] = \int_a^b x \frac{1}{b-a} dx = \frac{x^2}{2(b-a)} \Big|_a^b = \frac{a+b}{2}.$$

Law of the Unconscious Statistician

The next result is crucial when working with economic models involving random variables.

Theorem (Law of the Unconscious Statistician)

Let X be a random variable and define $Y = h(X)$ for some function h. Then

$$E_Y[Y] = E_X[h(X)] = \begin{cases} \sum_{x \in supp\ X} h(x) f_X(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} h(x) f_X(x) dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Without LOTUS, we would first find $f_Y(y)$, then
$E_Y[Y] = \sum_{y \in \text{supp}\ Y} y f_Y(y)$
With LOTUS, we do not need to go through the trouble of deriving its distribution. Instead, we may work with the distribution of X.

Law of the Unconscious Statistician (Contd.)

Example Let X be a continuous random variable. Consider $Y = h(X)$ where $h(x) = \mathbb{1}\{x \in \mathcal{A}\}$ for some set $\mathcal{A} \subset \mathbb{R}$. By the theorem, we have

$$E_Y[Y] = E_X[h(X)] = \int_{-\infty}^{\infty} \mathbb{1}\{x \in \mathcal{A}\} f_X(x) dx = \int_{\mathcal{A}} f_X(x) dx = P(X \in \mathcal{A}).$$

More generally, for any random variable X and set $\mathcal{A} \subset \mathbb{R}$, it holds that

$$E_X[\mathbb{1}\{X \in \mathcal{A}\}] = P(X \in \mathcal{A}).$$

Expectations (Contd.)

Expectations are defined as sums and integrals and thus inherit their useful properties:

Theorem

Let X be a random variable. Then

$$E_X[a + bX] = a + bE_X[X],$$

$\forall\ a, b \in \mathbb{R}$.

Expectations (Contd.)

Theorem

Let $X_1, \ldots, X_n$ be random variables. Then

$$E_{X_1,\ldots,X_n}\left[\sum_{i=1}^{n} b_i X_i\right] = \sum_{i=1}^{n} b_i E_{X_i}[X_i],$$

$\forall\ b_1, \ldots, b_n \in \mathbb{R}$.

Expectations (Contd.)

Theorem

Let $X_1, \ldots, X_n$ be independent random variables. Then

$$E_{X_1,\ldots,X_n}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} E_{X_i}[X_i].$$

Outline

Variance

Definition (Variance & Standard Deviation)

The variance of a random variable X with $\mu_X = E_X[X]$ is defined as

$$\text{Var}(X) = E_X[(X - \mu_X)^2].$$

The standard deviation of a random variable X is defined as

$$\text{sd}(X) = \sqrt{\text{Var}(X)}.$$

The variance (and standard deviation) are measures of dispersion.

- Characterize the spread of the distribution of X around its mean.

From the definition, it follows that

$$\text{Var}(X) = E_X[X^2] - E[X]^2.$$

Variance (Contd.)

Example: Consider tossing a fair coin twice as in the earlier example. Let X be the number of heads and recall $E_X[X] = 1$. We have

$$\mathsf{Var}(X) = E_X[X^2] - 1^2 = \frac{1}{4} \times 0^2 + \frac{1}{2} \times 1^2 + \frac{1}{4} \times 2^2 - 1 = \frac{1}{2}.$$

Variance (Contd.)

Corollary

Let X be a random variable. Then

$$Var(a + bX) = b^2 Var(X),$$

$\forall\ a, b \in \mathbb{R}$.

Variance (Contd.)

Example Let $X \sim \text{Bernoulli}(p)$. Then

$$E_X[X] = 0f(0) + 1f(1) = p,$$

and

$$\text{Var}(X) = E[X^2] - E[X]^2 = p - p^2 == p(1-p).$$

Example Let $X \sim N(\mu, \sigma^2)$. Then $E_X[X] = \mu$ and $\text{Var}(X) = \sigma^2$.

Outline

Covariance

So far, we have discussed two important features of a random variable: its mean and its variance.

We now turn to features that characterize the joint distribution of random variables, beginning with a measure of joint dispersion: the covariance.

Definition (Covariance)

The covariance of two random variables X and Y with $\mu_X = E_X[X]$ and $\mu_Y = E_Y[Y]$ is defined as

$$\text{Cov}(X, Y) = E_{X,Y}[(X - \mu_X)(Y - \mu_Y)].$$

From the definition, it follows that

$$\text{Cov}(X, Y) = E_{X,Y}[XY] - E[X]E[Y].$$

Covariance (Contd.)

Example Consider random variables X and Y with joint pmf given by

	$Y = 0$	$Y = 1$	Total
$X = 0$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{3}{10}$
$X = 1$	$\frac{3}{10}$	$\frac{2}{5}$	$\frac{7}{10}$
Total	$\frac{1}{2}$	$\frac{1}{2}$	1

We have $E_X[X] = \frac{7}{10}$ and $E_Y[Y] = \frac{1}{2}$, and

$$\begin{aligned}\text{Cov}(X, Y) &= E_{X,Y}[XY] - E_X[X]E_Y[Y] \\ &= 1 \times 1 \times \frac{2}{5} - \frac{7}{10} \times \frac{1}{2} \\ &= \frac{1}{20}.\end{aligned}$$

Covariance (Contd.)

Corollary

Let X and Y be random variables. Then

$$X \perp\!\!\!\perp Y \quad \Rightarrow \quad \mathit{Cov}(X, Y) = 0.$$

The converse does not hold in general.

Covariance (Contd.)

Corollary

Let X and Y be random variables. Then

$$Cov(a + bX, Y) = b\,Cov(X, Y),$$

for all $a, b \in \mathbb{R}$.

Covariance (Contd.)

Corollary

Let X and Y be random variables. Then

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

Covariance (Contd.)

Corollary

Let $X_1, \ldots, X_n$ be a collection of independent random variables. Then

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

Covariance (Contd.)

Theorem (Cauchy-Schwarz Inequality)

Let X and Y be random variables. Then

$$\begin{aligned} & Cov^2(X, Y) \leq Var(X)Var(Y) \\ \iff & Cov(X, Y) \leq sd(X)sd(Y) \end{aligned}$$

Outline

Correlation

Notice the units of $\text{Cov}(X, Y)$ are the units of X times Y.

- This makes comparisons challenging to interpret.
- This motivates normalization by the units of X times Y.

This leads to a measure of linear dependence: the correlation.

Definition (Correlation)

The correlation of two random variables X and Y is defined as

$$\text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{sd}(X)\text{sd}(Y)}.$$

Note: $\text{corr}(X, Y)$ is considered a measure of linear dependence because $\text{corr}(X, Y) \in \{-1, 1\}$ if and only if there exist $a, b \in \mathbb{R}$ such that $Y = a + bX$.

Correlation (Contd.)

A consequence of the Cauchy-Schwarz inequality is the following result:

Corollary

Let X and Y be random variables. We have

$$-1 \leq \mathit{corr}(X, Y) \leq 1.$$

Correlation (Contd.)

Example

Reconsider the random variables X and Y from the earlier example.
We have

$$\text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{sd}(X)\text{sd}(Y)} = \frac{\frac{1}{20}}{\sqrt{\frac{7\times 3}{100} \times \frac{1}{4}}}.$$

$Var(X) = (7/10)(3/10)$ and $Var(Y) = (1/2)(1/2)$

Outline

Outline

Conditional Expectation

We now introduce the concept of conditional expectations.

- Conditional expectations characterize features of a random variable when there is information on another random variable.

Definition (Conditional Expectation)

The conditional expectation of X given $Y = y$ is defined as

$$E_{X|Y}[X|Y = y] = \begin{cases} \sum_{x \in \text{supp } X} x f_{X|Y}(x|y), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Notice that this is simply the definition of expectation where we have replaced the pdf (or pmf) of X with the conditional pdf (or pmf) of X given $Y = y$.

Note: $E_{X|Y}[X|Y = y]$ is a number, however, $E_{X|Y}[X|Y]$ is a random variable. In econometrics, $E_{X|Y}[X|Y]$ is often called the conditional expectation function (CEF).

Conditional Expectation (Contd.)

Example Suppose $X \sim U(0,1)$ and $Y|X \sim U(X,1)$. Then

$$E_{Y|X}[Y|X] = \int_X^1 y \frac{1}{1-X} dy = \frac{y^2}{2(1-X)}\bigg|_X^1$$
$$= \frac{1-X^2}{2(1-X)} = \frac{1+X}{2}$$

(Note: $f_{Y|X}(Y \mid X) = \mathbb{1}\{Y \in [X,1]\}\frac{1}{1-X}$) and

$$E_{Y|X}[Y|X=x] = \frac{1+x}{2}.$$

Notice that $E_{Y|X}[Y|X] \sim U\left(\frac{1}{2},1\right)$, but $E_{Y|X}[Y|X=x]$ is a number.

Conditional Expectation (Contd.)

Corollary

Let X *and* Y *be random variables. Then*

$$E_{Y|X}[X + XY|X] = X + XE_{Y|X}[Y|X].$$

Similarly, for all functions h_1, h_2, *and* g,

$$E_{Y|X}[h_1(X) + h_2(X)g(Y)|X] = h_1(X) + h_2(X)E_{Y|X}[g(Y)|X].$$

Law of Iterated Expectations

Theorem (Law of Iterated Expectations (LIE))

Let X and Y be random variables. Then

$$E_Y[Y] = E_X[E_{Y|X}[Y|X]].$$

Outline

Conditional Variance

Another useful feature of Y given X is its conditional variance.

- Measures dispersion of Y given X.

Definition (Conditional Variance)

The conditional variance of Y given X is defined as

$$\text{Var}(Y|X) = E_{Y|X}[(Y - \mu_{Y|X})^2|X] = E_{Y|X}[Y^2|X] - E_{Y|X}[Y|X]^2,$$

where $\mu_{Y|X} = E_{Y|X}[Y|X]$.

Example Consider the returns to education example from the previous lecture.

- $\text{Var}(Y|D = 1)$ is the variance of hourly wages of college graduates.
- $\text{Var}(Y|D = 0)$ is the variance of hourly wages of non-graduates.
- Intuitively, which do you think is greater? Why?

Law of Total Variance

Corollary (Law of Total Variance (LTV))

Let X and Y be random variables. Then

$$Var(Y) = E_X[Var(Y|X)] + Var(E_{Y|X}[Y|X]).$$

Proof:

$$\begin{aligned}
&E_X[Var(Y|X)] + Var(E_{Y|X}[Y|X]) \\
&= E[E[Y^2|X] - E[Y|X]^2] + E[(E[Y|X] - E[E[Y|X]])^2] \\
&= E[E[Y^2|X] - E[Y|X]^2] + E[E[Y|X]^2 - 2E[Y|X]E[Y] + E[Y]^2] \\
&= E[E[Y^2|X]] - 2E[Y]E[E[Y|X]] + E[Y]^2 \\
&= E[Y^2] - E[Y]^2 = Var(Y)
\end{aligned}$$

Outline

Mean Independence

Recall that independence of random variables places a strong restriction on their joint distribution.

We now turn to a weaker restriction: mean independence.

Definition (Mean Independence)

Y is said to be mean independent of X if

$$E_{Y|X}[Y|X] = E_Y[Y].$$

- Mean-independence of Y with respect to X implies that X has no predictive value for Y in terms of mean-squared error.
- Independence of Y and X implies that X has no predictive value for Y under any loss.

Mean Independence (Contd.)

The next result states that mean independence is a weaker restriction on the joint distribution than independence.

Corollary

Let X and Y be random variables. Then

$$X \perp\!\!\!\perp Y \quad \Rightarrow \quad E_{Y|X}[Y|X] = E_Y[Y].$$

The converse does not hold in general.

Summary

This conclude our review of probability theory!

- Part A discussed distributions of random variables.
- Part B discussed features of distributions of random variables.

But there is another distinct task in the analysis of causal questions.

- In the next lecture, we begin the review of estimation.