Review of Probability Theory
Part B: Expectations

Thomas Wiemann
University of Chicago

Econometrics
Econ 21020

Updated: April 6, 2022

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

$$E_U[\tau(U)|W = 1] = E_U[g(1, U) - g(0, U)|W = 1], \tag{1}$$

and not the conditional distribution of $\tau(U)$ given $W = 1$.

The key concept we will cover in this lecture are *expectations*.

Outline

1. Features of Probability Distributions
   - ▷ Expectation
   - ▷ Variance
   - ▷ Covariance
   - ▷ Correlation
2. Features of Conditional Probability Distributions
   - ▷ Conditional Expectation
   - ▷ Conditional Variance
3. Mean Independence

These notes benefit greatly from the exposition in Wasserman (2003).

Outline

1. **Features of Probability Distributions**
   - ▷ **Expectation**
   - ▷ Variance
   - ▷ Covariance
   - ▷ Correlation

2. Features of Conditional Probability Distributions
   - ▷ Conditional Expectation
   - ▷ Conditional Variance

3. Mean Independence

These notes benefit greatly from the exposition in Wasserman (2003).

Definition 1 (Expected Value)

The *expected value* of a random variable X is defined as

$$E_X[X] = \begin{cases} \sum_{x \in \operatorname{supp} X} x f_X(x), & \text{if } X \text{ is discrete,} \\ \int_{\mathbb{R}} x f_X(x) dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{2}$$

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 Equation* (37) *as a Riemann–Stieltjes integral:* $E_X[X] = \int x dF(x)$.

Expectation (Contd.)

Example 1

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

$$f_X(x) = \begin{cases} 1/4 & \text{if } x = 0 \\ 1/2 & \text{if } x = 1 \\ 1/4 & \text{if } x = 2 \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

and the expected number of heads is

$$E_X[X] = \sum_{x \in \operatorname{supp} X} x f(x)$$

$$= 0 f(0) + 1 f(1) + 2 f(2) \tag{4}$$

$$= 1 \tfrac{1}{2} + 2 \tfrac{1}{4}$$

$$= 1$$

Example 2

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

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \forall x \in [a, b], \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

and we have

$$\begin{aligned} E_X[X] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{-\infty}^{\infty} x \, \mathbb{1}\{x \in [a,b]\} \frac{1}{b-a} dx \\ &= \int_a^b x \frac{1}{b-a} dx = \left[\frac{x^2}{2(b-a)} \right]_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{(b+a)\cancel{(b-a)}}{2\cancel{(b-a)}} = \frac{b+a}{2} \end{aligned} \tag{6}$$

Law of the Unconscious Statistician

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

Theorem 1 (Law of the Unconscious Statistician)

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

$$E_Y[Y] = E_X[h(X)] = \begin{cases} \sum_{x \in \operatorname{supp} X} h(x) f_X(x), & \text{if } X \text{ is discrete,} \\ \int_{\mathbb{R}} h(x) f_X(x) dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{7}$$

Proof.

See Exercise 4 in Problem Set 1 for the discrete case. □

The theorem is remarkable because $h(X)$ defines a new random variable, yet, we do not need to go through the trouble of deriving its distribution. Instead, we may work with the distribution of X.

Note*: The result gets its name from the fact that Equation* (7) *is often stated w/o the realization that it requires a proof and does not immediately follow from Definition 1.*

Law of the Unconscious Statistician (Contd.)

Example 3

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

$$
\begin{aligned}
E_Y[Y] = \; & E_X[h(X)] \\
& = \int_{-\infty}^{\infty} \mathbb{1}\{x \in \mathcal{A}\} f_X(x)\,dx \\
& = \int_{\mathcal{A}} f_X(x)\,dx \\
& = P(X \in \mathcal{A})
\end{aligned}
\tag{8}
$$

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

$$E_X\left[\mathbb{1}\{X \in \mathcal{A}\}\right] = P\left(X \in \mathcal{A}\right). \tag{9}$$

Expectations (Contd.)

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

Theorem 2

Let X be a random variable. Then

$$E_X[a + bX] = a + bE_X[X], \tag{10}$$

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

Proof.

We prove the result for continuous X.

$$E_X[a + bX] = \int_{-\infty}^{\infty}(a+bx)f(x)dx = \int_{-\infty}^{\infty} a f(x)dx + \int_{-\infty}^{\infty} bx f(x)dx$$

$$= a\underbrace{\int_{-\infty}^{\infty} f(x)dx}_{=1} + b\underbrace{\int_{-\infty}^{\infty} x f(x)dx}_{=E[X]} \tag{11}$$

$$= a + bE[X]$$

$\square$

Expectations (Contd.)

Theorem 3

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

$$E_{X_1,\ldots,X_n}\left[\sum_{i=1}^{n} b_i X_i\right] = \sum_{i=1}^{n} b_i E_{X_i}[X_i], \tag{12}$$

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

Proof.

Left as a self-study exercise. (Hint: Prove this for continuous random variables by using linearity of integrals, as in the proof of Theorem 2, and the definition of marginal pdfs.) □

Expectations (Contd.)

Theorem 4

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

$$E_{X_1,\ldots,X_n}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} E_{X_i}\left[X_i\right]. \tag{13}$$

Proof.

We prove the result for continuous X.

$$\begin{aligned} E_{X_1,\ldots,X_n}\left[\prod_{i=1}^{n} X_i\right] &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\prod_{i=1}^{n} x_i\right) f(x_1,\ldots,x_n)\,dx_1\cdots dx_n \\ &\overset{\text{II}}{=} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\prod x_i\right) f(x_1)\cdots f(x_n)\,dx_1\cdots dx_n \\ &= \int_{-\infty}^{\infty} x_1 f(x_1)\,dx_1 \cdots \int_{-\infty}^{\infty} x_n f(x_n)\,dx \\ &= \prod_{i=1}^{n} E_{X_i}[X_i] \end{aligned} \tag{14}$$

$\square$

Outline

1. **Features of Probability Distributions**
   - ▷ Expectation
   - ▷ **Variance**
   - ▷ Covariance
   - ▷ Correlation

2. Features of Conditional Probability Distributions
   - ▷ Conditional Expectation
   - ▷ Conditional Variance

3. Mean Independence

Variance

Definition 2 (Variance & Standard Deviation)

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

$$Var(X) = E_X\left[(X - \mu_X)^2\right]. \tag{15}$$

The *standard deviation* of a random variable X is defined as

$$sd(X) = \sqrt{Var(X)}. \tag{16}$$

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

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

From Equation (15), it follows that

$$Var(X) = E[X^2 - 2X\mu + \mu^2] = E[X^2] - 2\mu E[X] + E[\mu^2] \tag{17}$$

$$= E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2$$

Example 4

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

$$
\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= (0-1)^2 f(0) + (1-1)^2 f(1) + (2-1)^2 f(2) \\
&= 1(f(0) + f(2)) \\
&= \frac{1}{2}
\end{aligned}
\tag{18}
$$

OR: $Var(X) = E[X^2] - E[X]^2 = 1 + \frac{1}{2} - 1$

where $E[X^2] = 0^2 f(0) + 1^2 f(1) + 2^2 f(2) = \frac{1}{2} + \frac{4}{4} = 1 + \frac{1}{2}$

Variance (Contd.)

Corollary 1

Let X be a random variable. Then

$$Var(a + bX) = b^2 Var(X), \tag{19}$$

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

Proof.
We have

$$\begin{aligned} Var(a + bX) &= E[(a + bX - E[a + bX])^2] = E[(a + bX - a - bE[X])^2] \\ &= E[(b(X - E[X]))^2] \\ &= b^2 E[(X - E[X])^2] \\ &= b^2 Var(X) \end{aligned} \tag{20}$$

□

Variance (Contd.)

Example 5

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

$$E_X[X] = 0f(0) + 1f(1) = p \tag{21}$$

and

$$Var(X) = E[X^2] - E[X]^2 = E[X] - E[X]^2 = p - p^2 = p(1-p) \tag{22}$$

$X = 1\!\!1\{X=1\}$
$X^2 = 1\!\!1\{X=1\}^2 = 1\!\!1\{X=1\} = X$

Example 6

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

Outline

1. **Features of Probability Distributions**
   - ▷ Expectation
   - ▷ Variance
   - ▷ **Covariance**
   - ▷ Correlation
2. Features of Conditional Probability Distributions
   - ▷ Conditional Expectation
   - ▷ Conditional Variance
3. Mean Independence

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, and begin with a measure of joint dispersion: the covariance.

Definition 3 (Covariance)

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

$$Cov(X, Y) = E_{X,Y}\left[(X - \mu_X)(Y - \mu_Y)\right]. \tag{23}$$

Note: $Cov(X, X) = Var(X)$

From Equation (23) it follows that

$$Cov(X, Y) = E[XY - \mu_x Y - X\mu_y + \mu_x\mu_y] = E[XY] - \mu_x E[Y] - E[X]\mu_y + \mu_x\mu_y \tag{24}$$

$$= E[XY] - 2E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y]$$

Example 7

Consider random variables X and Y with joint pmf given by

	$Y = 0$	$Y = 1$	Total
$X = 0$	$1/5$	$1/10$	$3/10$
$X = 1$	$3/10$	$2/5$	$7/10$
Total	$5/10$	$5/10$	1

We have $E_X[X] = 7/10$ and $E_Y[Y] = 1/2$, and

$$Cov(X, Y) = E[XY] - E[X]E[Y] \tag{25}$$

$$= 1 \cdot f(1,1) - \frac{7}{10}\frac{1}{2}$$

$$= \frac{2}{5} - \frac{7}{20} = \frac{1}{20}$$

Corollary 2

Let X and Y be random variables. Then

$$X \perp\!\!\!\perp Y \Rightarrow Cov(X, Y) = 0. \tag{26}$$

The converse does not hold in general.

Proof.

We have

$$Cov(X, Y) = E[(X-\mu_x)(Y-\mu_y)] \overset{\perp\!\!\!\perp}{=} E[(X-\mu_x)]\underbrace{E[Y-\mu_y]} = 0 \tag{27}$$

$$= E[Y - E[Y]]$$

$$= E[Y] - E[E[Y]] = E[Y] - E[Y] = 0$$

See Exercise 7c) in Problem Set 1 for a counterexample of the converse. □

Covariance (Contd.)

Corollary 3

Let X and Y be random variables. Then

$$Cov(a + bX, Y) = bCov(X, Y), \tag{28}$$

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

Proof.

We have

$$\begin{aligned} Cov(a + bX, Y) &= E\Big[\big(\overbrace{a + bX - E[a + bX]}^{a + bX - (a + bE[X])}\big)(Y - E[Y])\Big] \\ &= E[b(X - E[X])(Y - E[Y])] \\ &= bE[(X - E[X])(Y - E[Y])] \\ &= bCov(X, Y) \end{aligned} \tag{29}$$

$\square$

Corollary 4

Let X and Y be random variables. Then

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y). \tag{30}$$

Proof.

We have

$$\begin{aligned} Var(X+Y) &= E\left[(X+Y-E(X+Y))^2\right] = E\left[\left((X-E[X])+(Y-E[Y])\right)^2\right] \\ &= E\left[(X-E[X])^2 + 2(X-E[X])(Y-E[Y]) + (Y-E[Y])^2\right] \\ &= E\left[(X-E[X])^2\right] + E\left[(Y-E[Y])^2\right] + 2E\left[(X-E[X])(Y-E[Y])\right] \\ &= Var(X) + Var(Y) + 2Cov(X,Y) \end{aligned} \tag{31}$$

□

Covariance (Contd.)

Corollary 5

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

$$Var\left(\sum_{i}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i). \tag{32}$$

Proof.

We have

$$Var\left(\sum_{i}^{n} X_i\right) = E\left[\left(\sum X_i - \sum E[X_i]\right)^2\right] = E\left[\left(\sum (X_i - E[X_i])\right)^2\right]$$

$$= E\left[\sum_i \sum_j (X_i - E[X_i])(X_j - E[X_j])\right] \tag{33}$$

$$= \sum_i \sum_j E\left[(X_i - E[X_i])(X_j - E[X_j])\right] = \sum_i E\left[(X_i - E[X_i])^2\right] + \sum_i \sum_{j \neq i} E\left[(X_i - E[X_i])(X_j - E[X_j])\right]$$

$$= \sum_i Var(X_i) + \sum_i \sum_{j \neq i} E[(X_i - E[X_i])] E[(X_j - E[X_j])] = \sum_i Var(X_i)$$

$\square$

Theorem 5 (Cauchy-Schwarz Inequality)

Let X and Y be random variables. Then

$$Cov^2(X, Y) \leq Var(X)Var(Y). \tag{34}$$

$\iff \quad Cov(X, Y) \leq sd(X)\, sd(Y)$

Proof.

Take $a \in \mathbb{R}$.

$Var(aX + Y) = Var(aX) + Var(Y) + 2\,Cov(aX, Y)$

$= a^2 Var(X) + Var(Y) + 2a\,Cov(X, Y)$

FOC: $\quad 2a\,Var(X) + 2\,Cov(X, Y) = 0$

$\Rightarrow \quad a = \dfrac{-Cov(X, Y)}{Var(X)}$

Covariance (Contd.)

$$Var(aX+Y) = \frac{Cov^2(X,Y)}{Var(X)} + Var(Y) - 2\frac{Cov^2(X,Y)}{Var(X)}$$

$$= Var(Y) - \frac{Cov^2(X,Y)}{Var(X)}$$

$$\Rightarrow \frac{Cov^2(X,Y)}{Var(X)} = Var(Y) - Var(aX+Y)$$

$$\Rightarrow Cov^2(X,Y) = Var(X)Var(Y) - \underbrace{Var(X)}_{\geq 0}\underbrace{Var(aX+Y)}_{\geq 0}$$

$$\Rightarrow Cov^2(X,Y) \leq Var(X)Var(Y)$$

(Note: Holds w/ equality only if $\exists a, b \in \mathbb{R} : Y = a + bX$) $\square$

Outline

1. **Features of Probability Distributions**
   - ▷ Expectation
   - ▷ Variance
   - ▷ Covariance
   - ▷ **Correlation**
2. Features of Conditional Probability Distributions
   - ▷ Conditional Expectation
   - ▷ Conditional Variance
3. Mean Independence

Correlation

Notice the units of $Cov(X, Y)$ are the units of X times Y.

- ▷ Makes comparisons challenging to interpreted.
- ▷ Motivates normalization by the units of X times Y.

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

Definition 4 (Correlation)

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

$$corr(X, Y) = \frac{Cov(X, Y)}{sd(X)sd(Y)}. \tag{35}$$

***Note**: corr(X, Y) is considered a measure of linear dependence because*

$$corr(X, Y) \in \{-1, 1\} \quad \Leftrightarrow \quad \exists a, b \in \mathbb{R} : Y = a + bX.$$

We don't make use of this result in this course and thus state it here w/o proof.

Correlation (Contd.)

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

Corollary 6

Let X and Y be random variables. We have

$$-1 \leq corr(X, Y) \leq 1. \tag{36}$$

Proof.

$$-1 \leq corr(X,Y) \leq 1$$

$$\Leftrightarrow \quad corr^2(X,Y) \leq 1$$

$$\Leftrightarrow \quad \frac{Cov^2(X,Y)}{Var(X)\,Var(Y)} \leq 1$$

$$\Leftrightarrow \quad Cov^2(X,Y) \leq Var(X)\,Var(Y)$$ holds by C.S.

$\square$

Example 8

Reconsider the random variables X and Y of Example 7. We have

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

$$= \frac{1/20}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

$$= \frac{1/20}{\sqrt{\frac{7 \cdot 3}{100}}\sqrt{\frac{1}{4}}}$$

$$Var(X) = \frac{7}{10} \cdot \frac{3}{10}$$

$$Var(Y) = \frac{1}{2} \cdot \frac{1}{2}$$

Outline

1. Features of Probability Distributions
   - ▷ Expectation
   - ▷ Variance
   - ▷ Covariance
   - ▷ Correlation
2. **Features of Conditional Probability Distributions**
   - ▷ **Conditional Expectation**
   - ▷ Conditional Variance
3. Mean Independence

Conditional Expectation

We now introduce the concept of *conditional* expectations.

- ▷ Characterize features of a random variable when there is information on another random variable.

Definition 5 (Conditional Expectation)

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

$$E_{X|Y}[X|Y=y] = \begin{cases} \sum_{x \in \operatorname{supp} X} x f_{X|Y}(x|y), & \text{if } X \text{ is discrete,} \\ \int_{\mathbb{R}} x f_{X|Y}(x|y) dx, & \text{if } X \text{ is continuous.} \end{cases} \tag{37}$$

Notice that this is simply Definition 1 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 9

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

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

$$= \frac{1-X^2}{2(1-X)} = \frac{(1+X)(1-X)}{2(1-X)}$$

$$= \frac{1+X}{2} = \frac{1}{2}(1+X)$$

$$f_{Y|X}(y|x) = 1\!\!1\{y \in [X,1]\}\frac{1}{X-1}$$

and

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

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 7

Let X and Y be random variables. Then

$$E_{Y|X}[X + XY|X] = X + XE_{Y|X}[Y|X]. \tag{38}$$

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

$$E_{Y|X}[h_1(X) + h_2(X)g(Y)|X] = h_1(X) + h_2(X)E_{Y|X}[g(Y)|X]. \tag{39}$$

Proof.

We prove Equation (38) for continuous Y.

$$E_{Y|X}[X + XY|X] = \int_{-\infty}^{\infty} (x + xy) f_{Y|X}(y|x)\,dy$$

$$= x \underbrace{\int_{-\infty}^{\infty} f_{Y|X}(y|x)\,dy}_{=1} + x \underbrace{\int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,dy}_{=E[Y|X]} = X + X E[Y|X]$$

$\square$

Theorem 6 (Law of Iterated Expectations; LIE)

Let X and Y be random variables. Then

$$E_Y[Y] = E_X\left[E_{Y|X}[Y|X]\right]. \tag{40}$$

Proof.

We prove the result for continuous X and Y.

$$E_X\left[E_{Y|X}[Y|X]\right] = \int_{-\infty}^{\infty} E_{Y|X}[Y|X=x] f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\,dy\, f_X(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{Y|X}(y|x) f_X(x)\,dy\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{YX}(y,x)\,dy\,dx = \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{YX}(y,x)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} y\, f_Y(y)\,dy = E[Y]$$

□

Law of Iterated Expectations (Contd.)

Example 10 (A Real-Life Simpson's Paradox)

An actual example from my university studies: Let Y denote the final course score, X_g denote gender, and X_o country of origin. We may have

$$E_{Y|X_g}[Y|X_g = m] > E_{Y|X_g}[Y|X_g = f],$$

even though we also have

$$\begin{aligned} & E_{Y|X_g,X_o}[Y|X_g = m, X_o = a] < E_{Y|X_g,X_o}[Y|X_g = f, X_o = a], \\ \text{and} \quad & E_{Y|X_g,X_o}[Y|X_g = m, X_o = b] < E_{Y|X_g,X_o}[Y|X_g = f, X_o = b]. \end{aligned}$$

How is this possible? The LIE gives

$$E[Y|X_g=m] = E[E[Y|X_g=m, X_o]] = E[Y|X_g=m, X_o=a]\,P(X_o=a|X_g=m) + E[Y|X_g=m, X_o=b]\,P(X_o=b|X_g=m)$$

$$E[Y|X_g=f] = E[E[Y|X_g=f, X_o]] = E[Y|X_g=f, X_o=a]\,P(X_o=a|X_g=f) + E[Y|X_g=f, X_o=b]\,P(X_o=b|X_g=f)$$

Just need: $E[Y|X_g=m, X_o=a] \gg E[Y|X_g=f, X_o=b]$ & $P(X_o=a|X_g=m) \gg P(X_o=a|X_g=f)$

Outline

1. Features of Probability Distributions
    - ▷ Expectation
    - ▷ Variance
    - ▷ Covariance
    - ▷ Correlation
2. **Features of Conditional Probability Distributions**
    - ▷ Conditional Expectation
    - ▷ **Conditional Variance**
3. Mean Independence

Conditional Variance

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

- ▷ Measures dispersion of Y given X.

Definition 6 (Conditional Variance)

The *conditional variance* of Y given X is defined as

$$Var(Y|X) = E_{Y|X}\left[(Y - \mu_{Y|X})^2|X\right], \tag{41}$$

where $\mu_{Y|X} \equiv E_{Y|X}[Y|X]$. $= E_{Y|X}[Y^2|X] - E_{Y|X}[Y|X]^2$

Example 11

Consider the returns to education example from Lecture 1.

- ▷ $Var(Y|W = 1)$ is the variance of hourly wages of college graduates.
- ▷ $Var(Y|W = 0)$ is the variance of hourly wages of non-graduates.

Intuitively, which do you think is greater? Why?

Corollary 8 (Law of Total Variance; LTV)

Let X and Y be random variables. Then

$$Var(Y) = E_X\left[Var(Y|X)\right] + Var\left(E_{Y|X}[Y|X]\right). \tag{42}$$

Proof.
We have

$$E_X\left[Var(Y|X)\right] + Var\left(E_{Y|X}[Y|X]\right) =$$

$$= E\left[E[Y|X] - E[Y|X]^2\right] + E\left[\left(E[Y|X] - \overbrace{E[E[Y|X]]}^{\overset{LIE}{=} E[Y]}\right)^2\right]$$

$$= E\left[E[Y^2|X] - \cancel{E[Y|X]^2}\right] + E\left[\cancel{(E[Y|X]^2} - 2E[Y|X]E[Y] + E[Y]^2\right]$$

$$= \underbrace{E[E[Y^2|X]]}_{\overset{LIE}{=} E[Y^2]} - 2E[Y]\underbrace{E[E[Y|X]]}_{\overset{LIE}{=} E[Y]} + E[Y]^2 = E[Y^2] - E[Y]^2$$

$$= Var(Y)$$

$\square$

Outline

1. Features of Probability Distributions
   - ▷ Expectation
   - ▷ Variance
   - ▷ Covariance
   - ▷ Correlation
2. Features of Conditional Probability Distributions
   - ▷ Conditional Expectation
   - ▷ Conditional Variance
3. **Mean Independence**

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 7 (Mean Independence)

Y is said to be *mean independent* of X if

$$E_{Y|X}[Y|X] = E_Y[Y]. \tag{43}$$

Exercise 6 in Problem set 1 shows that we can interpret $E_{Y|X}[Y|X]$ as the best predictor of Y given X under the L^2-loss.

- ▷ Mean-independence of Y with respect to X implies that X has no predictive value for Y under the L^2-loss.
- ▷ Independence of Y and X implies that X has no predictive value for Y under *any* loss.

Mean Independence (Contd.)

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

Corollary 9

Let X and Y be random variables. Then

$$X \perp\!\!\!\perp Y \Rightarrow E_{Y|X}[Y|X] = E_Y[Y]. \tag{44}$$

The converse does not hold in general.

Proof.

See Exercise 7a) in Problem set 1 for a proof when X and Y are continuous. See Exercise 7c) for a counterexample of the converse. $\square$

Summary

This concludes our review of probability theory!

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

We are now fully equipped to revisit Task 1 (Definition) and Task 2 (Identification) from Lecture 1.

- ▷ Patience: We will do so in Lecture 6 & 7.

Even better: We are equipped for identification analysis under assumptions other than Random Assignment.

- ▷ Know everything to show identification under the Selection on Observables or the Instrumental Variables assumptions.
- ▷ Important because Random Assignment wasn't plausible in the returns to education example.

But there are *three* distinct tasks in the analysis of causal questions.

- ▷ In the next lecture, we begin the review of statistics.
- ▷ This is preparation for Task 3 (Estimation).

References

Wasserman, L. (2003). *All of statistics*. Springer.