Proof Practice

Emanuel Mejía

Week 5

A - Best Linear Predictor of a Constrained Outcome Space

Suppose that discrete random variables X and Y have joint probability mass function given by:

$$f(x, y) = \begin{cases} \frac{1}{2} & (x, y) \in \{(0,0), (2,1)\} \\ 0 & otherwise \end{cases}$$

(This means that there is equal probability that the points $(0, 0)$ and $(2, 1)$ are drawn; there is zero probability that any other point is drawn.)

Let $g(x) = \beta_0 + \beta_1 x$ be a predictor for y that is a function of x, and define the error, ϵ, to be the difference between the true value of y and the prediction $g(x)$, $\epsilon = Y - g(X)$.

1. If you impose the moment condition (that is, you require that), $E[\epsilon] = 0$, what one point in the plane must the predictor pass through? (In some places, this point is referred to as the grand mean.)
2. Because we have defined $\epsilon = Y - g(X)$, we can ask the question, "What is the covariance between X and ϵ?"

Because how we have defined ϵ, we can know that the answer probably starts with a substitution:

$$Cov[X, \epsilon] = Cov[X, Y - g(x)]$$

Assume (or you might say, "require") that the expected value of ϵ is zero, $E[\epsilon] = 0$. Then, prove that $Cov[X, \epsilon]$ has the form $a + b\beta_1$.

Given the constraints of the pdf, $f(x, y)$, you have been provided, what is the specific value of b?

3. How is the sign of $Cov[X, \epsilon]$ is related to the angle of the line.
4. Compute the BLP in this way:

a. Assume (or you might say require) that $E[\epsilon] = 0$.

b. Then, set $Cov[X, \epsilon] = 0$ and solve for β_1.

What is the value of β_1?

Solution:

0. Calculations

Before answering the questions, we'll be calculating the following values based on the information provided: First, we'll present the following joint PMF for X and Y, and the marginals for each one of them in the edge:

$\boldsymbol{f(x,y)}$	$Y=0$	$Y=1$	$\boldsymbol{f_X(x)}$
$X=0$	1/2	0	**1/2**
$X=2$	0	1/2	**1/2**
$\boldsymbol{f_Y(y)}$	**1/2**	**1/2**	**1**

We'll also calculate the following:

- $E[X] = \sum_x x f_X(x) = 0 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = 1$
- $E[Y] = \sum_y y f_Y(x) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$
- $E[X^2] = \sum_x x^2 f_X(x) = (0)^2 \cdot \frac{1}{2} + (2)^2 \cdot \frac{1}{2} = 2$
- $E[XY] = \sum_x \sum_y xyf(x,y) = (0)(0) \cdot \frac{1}{2} + (2)(1) \cdot \frac{1}{2} = 1$
- $V[X] = E[X^2] - E[X]^2 = 2 - (1)^2 = 1$
- $Cov[X,Y] = E[XY] - E[X]E[Y] = 1 - (1)(\frac{1}{2}) = \frac{1}{2}$

1. The Grand Mean

We'll impose that $E[\epsilon] = 0$, Then:

$$\begin{aligned} 0 &= E\left[Y - g(X)\right] \\ &= E\left[Y\right] - E\left[g(X)\right] \\ &= E\left[Y\right] - E\left[\beta_0 + \beta_1 x\right] \\ &= E\left[Y\right] - E\left[\beta_0\right] - E\left[\beta_1 x\right] \\ &= E\left[Y\right] - E\left[\beta_0\right] - \beta_1 E\left[x\right] \\ &= E\left[Y\right] - \beta_0 - \beta_1 E\left[x\right] \end{aligned}$$

$$\begin{aligned} \Rightarrow E[Y] &= \beta_0 + \beta_1 E\left[x\right] \\ &= g(E[X]) \end{aligned}$$

$\therefore$ The point (x,y) where the predictor must pass is $(E[X], E[Y])$

2. Covariance $Cov[X,\epsilon]$

We know that $Cov[X,\epsilon] = E[X\epsilon] - E[X]E[\epsilon]$, but we've also defined $E[\epsilon] = 0$. Therefore:

$$
\begin{aligned}
Cov[X,\epsilon] &= E[X\epsilon] - E[X]E[\epsilon] \\
&= E[X\epsilon] \\
&= E[X(Y - g(X))] \\
&= E[XY - Xg(X)] \\
&= E[XY] - E[Xg(X)] \\
&= E[XY] - E[X(\beta_0 + \beta_1 X)] \\
&= E[XY] - E[\beta_0 X + \beta_1 X^2] \\
&= E[XY] - E[\beta_0 X] - E[\beta_1 X^2] \\
&= E[XY] - \beta_0 E[X] - \beta_1 E[X^2] \\
&= 1 - \beta_0 \cdot 1 - \beta_1 \cdot 2 && \text{substituting values} \\
&= 1 - \beta_0 - 2\beta_1
\end{aligned}
$$

Which has the form $a + b\beta_1$ with $a = 1 - \beta_0$ and $b = -2$

3. Sign for $Cov[X,\epsilon]$ related to angle for g(X)

We know that the angle of the line $g(X)$ is given by β_1 which has a positive sign in $g(X)$ and a negative sign in $Cov[X,\epsilon]$ so, this seems like an inverse relationship, so if we have a fixed Y-intercept (β_0):

- As the angle gets "higher" (more positive and steeper), $Cov[X,\epsilon]$ decreases.

And viceversa:

- As the angle gets "lower" (more negative and also steeper), $Cov[X,\epsilon]$ increases.

We can go deeper in this question by trying to understand what happens when $Cov[X,\epsilon]$ is strictly negative or positive.

3.1 $Cov[X,\epsilon] < 0$ For the case where the mentioned covariance is strictly negative we have the following:

$$
\begin{aligned}
& 1 - \beta_0 - 2\beta_1 < 0 && \text{since } Cov[X,\epsilon] = 1 - \beta_0 - 2\beta_1 \\
& \Rightarrow 1 - \beta_0 < 2\beta_1 \\
& \Rightarrow \frac{1 - \beta_0}{2} < \beta_1 && \text{So if } \beta_0 < 1 \Rightarrow 0 < \beta_1
\end{aligned}
$$

Meaning that if $Cov[X,\epsilon] < 0$ and $\beta_0 < 1 \Rightarrow \beta_1$ will be strictly positive (the slope for $g(X)$ will be positive).

The other way around isn't always true, but we can also prove that if $0 < \beta_1$ and $1 < \beta_0 \Rightarrow Cov[X,\epsilon] > 0$.

3.2 $Cov[X,\epsilon] > 0$ On the other hand, when the mentioned covariance is strictly positive we get the following:

$$
\begin{aligned}
& 1 - \beta_0 - 2\beta_1 > 0 && \text{since } Cov[X,\epsilon] = 1 - \beta_0 - 2\beta_1 \\
& \Rightarrow 1 - \beta_0 > 2\beta_1 \\
& \Rightarrow \frac{1 - \beta_0}{2} > \beta_1 && \text{So if } 1 < \beta_0 \Rightarrow \beta_1 < 0
\end{aligned}
$$

Meaning that if $Cov[X, \epsilon] > 0$ and $1 < \beta_0 \Rightarrow \beta_1$ must be strictly negative (slope for $g(X)$ will be negative). The other way around isn't always true, but we can prove that if $\beta_1 < 0$ and $\beta_0 < 1 \Rightarrow Cov[X, \epsilon] < 0$.

4. Computing the BLP

What about the case $Cov[X, \epsilon] = 0$? We'll use it to compute the BLP as follows:

We know that $E[\epsilon] = 0$

$$
\begin{aligned}
&\Rightarrow E[Y] = \beta_0 + \beta_1 E[x] && \text{from exercise 1.}\\
&\Rightarrow \frac{1}{2} = \beta_0 + \beta_1 \cdot 1 && \text{substituting values}\\
&\Rightarrow \beta_0 + \beta_1 = \frac{1}{2}
\end{aligned}
$$

On the other hand we'll set $Cov[X, \epsilon] = 0$

$$
\begin{aligned}
&\Rightarrow 0 = 1 - \beta_0 - 2\beta_1 && \text{from exercise 2.}\\
&\Rightarrow \beta_0 = 1 - 2\beta_1
\end{aligned}
$$

By substituting this in our first result we have:

$$
\begin{aligned}
&(1 - 2\beta_1) + \beta_1 = \frac{1}{2}\\
&\Rightarrow 1 - \beta_1 = \frac{1}{2}\\
&\Rightarrow \beta_1 = \frac{1}{2}
\end{aligned}
$$

And since we knew that $\beta_0 = 1 - 2\beta_1 \Rightarrow \beta_0 = 0$.

And the BLP is: $g(x) = \frac{1}{2}x$

Note: These values for β_0 and β_1 are consistent with the definition of the BLP:

$$g(x) = \alpha + \beta X$$

With:

$$
\begin{aligned}
\beta_0 = \alpha &= E[Y] - \frac{Cov[X, Y]}{V[X]} E[X]\\
&= \frac{1}{2} - \frac{\frac{1}{2}}{1} \cdot 1\\
&= 0\\
\beta_1 = \beta &= \frac{Cov[X, Y]}{V[X]}\\
&= \frac{\frac{1}{2}}{1}\\
&= \frac{1}{2}
\end{aligned}
$$

B - Think of a Friendly Type of Function

Let $T_{(i)}$ be a sequence of discrete random variables for $i \in \{1, 2, 3, ...\}$. Suppose that $T_{(i)}$ has the pmf,

$$f_i(t) = \begin{cases} \frac{1}{2} & t = \frac{1}{i} \\ \frac{1}{2} & t = -\frac{1}{i} \\ 0 & otherwise \end{cases}$$

Define $g : \mathbb{R} \to \mathbb{R}$ by $g(t) = t^2 + e^t$

a. Prove that $T_{(n)} \xrightarrow{p} 0$

b. Prove that $g(T_{(n)}) \xrightarrow{p} 1$, without computing the distribution of $g(T_{(n)})$.

Solution:

a. Using the definition for Congergence in Probability

We know that $T_{(n)} \xrightarrow{p} 0$ means that $\forall \varepsilon > 0$

$$\lim_{n\to\infty} P\left[T_{(n)} \in (c - \varepsilon, c + \varepsilon)\right] = 1$$

Now, let's define $\varepsilon_n = \frac{2}{n}$, and $c = 0$ so for any $n \in \mathbb{N}$:

$$\begin{aligned} P\left[T_{(n)} \in (c - \varepsilon_n, c + \varepsilon_n)\right] &= P\left[T_{(n)} \in (0 - \frac{2}{n}, 0 + \frac{2}{n})\right] \\ &= P\left[T_{(n)} \in (-\frac{2}{n}, \frac{2}{n})\right] \\ &= 1 && \text{because of the way } T_{(n)} \text{is defined} \end{aligned}$$

This also means that:

$$\lim_{n\to\infty} P\left[T_{(n)} \in (-\varepsilon_n, \varepsilon_n)\right] = \lim_{n\to\infty} 1 = 1$$

This ε_n has the property that as $n \to \infty$, $\varepsilon_n \to 0^+$ and $-\varepsilon_n \to 0^-$.

Therefore, $\forall \varepsilon > 0$ we'll always be able to find a big enough value for n such that $(-\varepsilon_n, \varepsilon_n) \subseteq (-\varepsilon, \varepsilon)$ which would also mean that:

$$P\left[T_{(n)} \in (-\varepsilon_n, \varepsilon_n)\right] \leq P\left[T_{(n)} \in (-\varepsilon, \varepsilon)\right]$$

So:

$$\lim_{n\to\infty} P\left[T_{(n)} \in (-\varepsilon_n, \varepsilon_n)\right] \leq \lim_{n\to\infty} P\left[T_{(n)} \in (-\varepsilon, \varepsilon)\right]$$

But since P is a probability measure $\Rightarrow 0 \leq P\left[T_{(n)} \in (-\varepsilon, \varepsilon)\right] \leq 1$

Then, the previous statements: $1 = \lim_{n\to\infty} P\left[T_{(n)} \in (-\varepsilon_n, \varepsilon_n)\right] \leq \lim_{n\to\infty} P\left[T_{(n)} \in (-\varepsilon, \varepsilon)\right]$ will hold:

$$\iff \lim_{n\to\infty} P\left[T_{(n)} \in (-\varepsilon, \varepsilon)\right] = 1$$

Which is the definition for Convergence in Probability for $c = 0$:

$$\therefore T_{(n)} \xrightarrow{p} 0$$

b. Using the Continuous Mapping Theorem

Using the CMT we know that if $T_{(n)} \xrightarrow{p} a$, then $g(T_{(n)}) \xrightarrow{p} g(a)$.

From the previous result we know that $T_{(n)} \xrightarrow{p} 0$, so evaluating:

$$\begin{aligned} g(0) &= (0)^2 + e^0 \\ &= 1 \end{aligned}$$

$$\therefore g(T_{(n)}) \xrightarrow{p} 1$$