Gov 2001 Section 5, March 7, 2025

Conditional Expectation

1. Remember that conditional expectation $E[Y|\mathbf{X}]$ is a function of $\mathbf{X}$. When confused, remember it is in essence an expression $\mu(\mathbf{x}) = E[Y|\mathbf{X} = \mathbf{x}]$ for all possible $\mathbf{x}$, and then we replace $\mathbf{X} = \mathbf{x}$ with $\mathbf{X}$ (this is just for simplicity of notation) to get $\mu(\mathbf{X}) = E[Y|\mathbf{X}]$.

2. $E[Y|\mathbf{X}]$ is the projection of Y on $\mathbf{X}$, i.e., it minimizes mean squared error:

$$E[Y|\mathbf{X}] = \arg\min_{g(\mathbf{X})} E[(Y - g(\mathbf{X}))^2]$$

3. CEF error is orthogonal to every function of $\mathbf{X}$, i.e., for all functions h, we have

$$E[(Y - E[Y|\mathbf{X}])h(\mathbf{X})] = 0$$

   A quick proof using the Law of iterated expectation and factoring out $h(X)$:

$$\begin{aligned} E[(Y - E[Y|X])h(X)] &= E[h(X)Y] - E[E[Y|X]h(X)] \\ &= E[E[h(X)Y|X]] - E[E[Y|X]h(X)] \\ &= E[h(X)E[Y|X]] - E[E[Y|X]h(X)] = 0 \end{aligned}$$

4. Important properties of conditional expectation:

   (a) If $X \perp Y$, we have $E[Y|X] = E[Y]$, $E[XY] = E[X]E[Y]$.

   (b) $E[h(X)|X] = h(X)$. Notice that the result is still a function of X rather than an expected value. $E[h(X)Y|X] = h(X)E[Y|X]$ for all functions h.

   (c) Linearity: $E[cY|X] = cE[Y|X]$ for any constant c.

   (d) Law of iterated expectation (Adam's Law, law of total expectation): $E[Y] = E[E[Y|X]] = \int_x E[Y|X = x]f_X(x)\,dx = E[Y|B]P(B) + E[Y|B^c]P(B^c)$. With extra conditioning, we have $E[Y|Z] = E[E[Y|X,Z]|Z] = \int_x E[Y|X,Z]f_{X|Z}(x|Z)\,dx$.

5. Conditional variance $\text{Var}(Y|\mathbf{X})$ is also a function of $\mathbf{X}$, denote as $\sigma^2(\mathbf{X})$. We have

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

   which is repeatedly used in the proof of EVVE.

6. EVVE's law: $\text{Var}(Y|X) = E[\text{Var}(Y|X)] + \text{Var}(E[Y|X])$. We can interpret it as the sum of within-group variation and between-group variation.

Example

You are approached by a mysterious stranger, who allows you to bid on a mystery box containing a mystery prize! The value V of the prize is Uniform on $[0,1]$ (measured in millions of dollars). You can choose to bid any amount b (in millions of dollars). If $b \geq 2V/3$ then your bid is accepted (and your payoff is $V - b$, value minus bid). Otherwise, the bid is rejected and nothing happens. What is your optimal bid?

Solution: Let W be the payoff. If $b < 2/3$, we have by law of total expectation,

$$\begin{aligned} E[W] &= E[W|b \geq 2V/3]P(b \geq 2V/3) + E[W|b < 2V/3]P(b < 2V/3) \\ &= E[V - b|b \geq 2V/3]P(b \geq 2V/3) + 0 \\ &= (E[V|V \leq 3b/2] - b)P(V \leq 3b/2) \\ &= \left(\int_0^{3b/2} \frac{2v}{3b}\, dv - b\right)\left(\int_0^{3b/2} 1\, dv\right) \\ &= (3b/4 - b)(3b/2) = -3b^2/8 \leq 0 \end{aligned}$$

If $b \geq 2/3$, we always have $b \geq 2V/3$, and

$$E[W] = E[V - b] = \frac{1}{2} - b \leq -\frac{1}{6}$$

So the optimal bid is just 0.

Question 1

Suppose we want to model the relationship between legislation and politician quality. There are two types of politician quality: high and low. When a high quality politician proposes a bill, it has a probability p_1 to pass; conversely, when a low quality politician proposes a bill, it has a probability p_2 to pass, where $p_1 > p_2$. Unfortunately, we cannot directly observe politicians' quality, but instead rely on our prior that a politician is a high type with probability h and low type with probability $(1 - h)$, where $h \in (0,1)$. Let X be the number of passed bills after a randomly picked politician has made n proposals.

(a) Find the marginal distribution of X.

(b) Find the mean and variance of X.

Solution to Question 1

(a) Marginal distribution of X. Let H denote the event that the politician is of high quality, and L the event of low quality.

$$X \mid H \sim \text{Binomial}(n, p_1), \quad X \mid L \sim \text{Binomial}(n, p_2).$$

By the law of total probability and using the prior that the politician is high quality with probability h,

$$P(X = k) \;=\; P(X = k \mid H)\, P(H) \;+\; P(X = k \mid L)\, P(L).$$

Hence, the marginal distribution is a mixture of two Binomial distributions:

$$P(X = k) \;=\; h \binom{n}{k} p_1^k (1 - p_1)^{n-k} \;+\; (1 - h) \binom{n}{k} p_2^k (1 - p_2)^{n-k}, \quad k = 0, 1, \ldots, n.$$

(b) Mean and variance of X.

$$\mathbb{E}[X] \;=\; \mathbb{E}[\mathbb{E}[X \mid \text{type}]] \;=\; h\, \mathbb{E}[X \mid H] \;+\; (1 - h)\, \mathbb{E}[X \mid L].$$

Since $X \mid H \sim \text{Binomial}(n, p_1)$ and $X \mid L \sim \text{Binomial}(n, p_2)$,

$$\mathbb{E}[X \mid H] = n\, p_1, \quad \mathbb{E}[X \mid L] = n\, p_2.$$

Therefore,

$$\mathbb{E}[X] = h\, n\, p_1 + (1 - h)\, n\, p_2 \;=\; n\big[h\, p_1 + (1 - h)\, p_2\big].$$

For the variance, use the law of total variance:

$$\text{Var}(X) \;=\; \mathbb{E}[\text{Var}(X \mid \text{type})] \;+\; \text{Var}(\mathbb{E}[X \mid \text{type}]).$$

We know

$$\text{Var}(X \mid H) = n\, p_1 (1 - p_1), \quad \text{Var}(X \mid L) = n\, p_2 (1 - p_2).$$

Hence,

$$\mathbb{E}[\text{Var}(X \mid \text{type})] = h\, \big(n\, p_1 (1 - p_1)\big) + (1 - h)\, \big(n\, p_2 (1 - p_2)\big).$$

Also,

$$\mathbb{E}[X \mid H] = n\, p_1, \quad \mathbb{E}[X \mid L] = n\, p_2.$$

So

$$\text{Var}(\mathbb{E}[X \mid \text{type}]) = h\, (n\, p_1)^2 + (1 - h)\, (n\, p_2)^2 \;-\; \Big(n\, \big[h\, p_1 + (1 - h)\, p_2\big]\Big)^2.$$

Putting it all together gives the total variance:

$$\text{Var}(X) = h\, n\, p_1 (1 - p_1) + (1 - h)\, n\, p_2 (1 - p_2) + h\, (n\, p_1)^2 + (1 - h)\, (n\, p_2)^2 - \Big(n\big[h\, p_1 + (1 - h)\, p_2\big]\Big)^2.$$

You can leave it in this form or simplify it as needed.
!!!Notice: If you pick a new random politician each time for a new proposal, then we can do $X \sim \text{Bin}(n, p_1 h + p_2 (1 - h))$ and directly derive the mean and variance.

Question 2

We know from the definition of the variance that

$$\mathbb{E}\big[(Y - \mathbb{E}[Y])^2\big] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2.$$

Prove that this equality still holds when we condition on X, i.e.,

$$\mathbb{E}\big[(Y - \mathbb{E}[Y \mid X])^2 \mid X\big] = \mathbb{E}[Y^2 \mid X] \;-\; (\mathbb{E}[Y \mid X])^2.$$

Solution to Question 2

Recall that conditioning on X, we treat X as fixed while taking expectation with respect to Y. By definition:

$$\mathbb{E}\big[(Y - \mathbb{E}[Y \mid X])^2 \,\big|\, X\big] \;=\; \mathbb{E}\Big[Y^2 - 2\,Y\,\mathbb{E}[Y \mid X] + \big(\mathbb{E}[Y \mid X]\big)^2 \,\Big|\, X\Big].$$

Since $\mathbb{E}[Y \mid X]$ is treated as a constant when conditioning on X,

$$= \mathbb{E}[Y^2 \mid X] - 2\,\mathbb{E}[Y \mid X]\mathbb{E}[Y \mid X] + \big(\mathbb{E}[Y \mid X]\big)^2 \;=\; \mathbb{E}[Y^2 \mid X] - \big(\mathbb{E}[Y \mid X]\big)^2.$$

Hence,

$$\mathbb{E}[(Y - \mathbb{E}[Y \mid X])^2 \mid X] \;=\; \mathbb{E}[Y^2 \mid X] \;-\; (\mathbb{E}[Y \mid X])^2,$$

as required.

Question 3

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2, and $n \geq 2$. A bootstrap sample of $X_1, \ldots, X_n$ is a sample of n random variables $X_1^*, \ldots, X_n^*$ formed from the X_j by sampling *with replacement* with equal probabilities. Let $\overline{X}^*$ denote the sample mean of the bootstrap sample:

$$\overline{X}^* = \frac{1}{n}\big(X_1^* + \cdots + X_n^*\big).$$

(a) Find $\mathbb{E}[X_j^*]$ and $\mathrm{Var}(X_j^*)$ for each j. (Hint: What is the distribution of X_j^*?)

(b) Find $\mathbb{E}\big[\overline{X}^* \mid X_1, \ldots, X_n\big]$ and $\mathrm{Var}\big[\overline{X}^* \mid X_1, \ldots, X_n\big]$. (Hint: Conditional on $X_1, \ldots, X_n$, the X_j^* are independent, each putting probability $1/n$ at each of the points $X_1, \ldots, X_n$.)

(c) Find $\mathbb{E}\big[\overline{X}^*\big]$ and $\mathrm{Var}\big[\overline{X}^*\big]$. (Hint: Recall that the sample variance $\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2$ is an unbiased estimator of the population variance σ^2.)

Solution to Question 3

Let us denote the original sample as $X_1, \ldots, X_n$ (i.i.d. with mean μ and variance σ^2). A bootstrap sample $(X_1^*, \ldots, X_n^*)$ is drawn *with replacement* from $\{X_1, \ldots, X_n\}$.

(a) Distribution of each X_j^*. Conditioned on $X_1, \ldots, X_n$, the random variable X_j^* is equally likely to be any of $X_1, \ldots, X_n$, with probability $1/n$ each. Hence:

$$\mathbb{E}[X_j^* \mid X_1, \ldots, X_n] = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Unconditionally, this random variable still has expectation μ, but more precisely,

$$\mathbb{E}[X_j^*] = \mathbb{E}\big[\mathbb{E}[X_j^* \mid X_1, \ldots, X_n]\big] = \mathbb{E}\big[\tfrac{1}{n}\sum_{i=1}^{n} X_i\big] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mu.$$

For the variance, again conditioning on $X_1, \ldots, X_n$:

$$\operatorname{Var}(X_j^* \mid X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \big(\overline{X}\big)^2.$$

By EVVE's law, we have

$$\begin{aligned}
\operatorname{Var}(X_j^*) &= \mathbb{E}[\operatorname{Var}(X_j^* \mid X_1, \ldots, X_n)] + \operatorname{Var}(\mathbb{E}[X_j^* \mid X_1, \ldots, X_n]) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i^2] - \mathbb{E}[\bar{X}^2] + \operatorname{Var}(\bar{X}) \\
&= \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2 \\
&= \operatorname{Var}(X_1) = \sigma^2
\end{aligned}$$

(b) Mean and variance of $\overline{X}^*$ *conditional* on $X_1, \ldots, X_n$. We have

$$\overline{X}^* = \frac{1}{n}\sum_{j=1}^{n} X_j^*.$$

Since, given $X_1, \ldots, X_n$, each X_j^* is an i.i.d. draw from the empirical distribution that places mass $1/n$ on each X_i,

$$\mathbb{E}\big[\overline{X}^* \mid X_1, \ldots, X_n\big] = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[X_j^* \mid X_1, \ldots, X_n] = \frac{1}{n}\sum_{j=1}^{n}\overline{X} = \overline{X}.$$

Hence, conditional on the original sample, the mean of the bootstrap sample average is simply $\overline{X}$.

For the variance,

$$\operatorname{Var}\big(\overline{X}^* \mid X_1, \ldots, X_n\big) = \frac{1}{n^2}\sum_{j=1}^{n}\operatorname{Var}(X_j^* \mid X_1, \ldots, X_n),$$

because the X_j^* are independent given $X_1, \ldots, X_n$. Each term $\operatorname{Var}(X_j^* \mid X_1, \ldots, X_n)$ is the same, so

$$= \frac{1}{n^2}\, n \operatorname{Var}(X_1^* \mid X_1, \ldots, X_n) = \frac{1}{n}\operatorname{Var}(X_1^* \mid X_1, \ldots, X_n).$$

Using the fact that, conditionally, X_1^* takes the values $X_1, \ldots, X_n$ each with probability $1/n$:

$$\operatorname{Var}(X_1^* \mid X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \; - \; \big(\overline{X}\big)^2.$$

Thus

$$\operatorname{Var}\big(\overline{X}^* \mid X_1, \ldots, X_n\big) = \frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 \; - \; \big(\overline{X}\big)^2\right].$$

(c) Mean and variance of $\overline{X}^*$ *unconditional.* First, for the mean:

$$\mathbb{E}\big[\overline{X}^*\big] = \mathbb{E}\big[\mathbb{E}[\overline{X}^* \mid X_1, \ldots, X_n]\big] = \mathbb{E}[\overline{X}] = \mu.$$

Next, for the variance, use the law of total variance:

$$\mathrm{Var}\big(\overline{X}^*\big) = \mathbb{E}\big[\mathrm{Var}(\overline{X}^* \mid X_1, \ldots, X_n)\big] \;+\; \mathrm{Var}\big(\mathbb{E}[\overline{X}^* \mid X_1, \ldots, X_n]\big).$$

From part (b),

$$\mathbb{E}[\overline{X}^* \mid X_1, \ldots, X_n] \;=\; \overline{X}.$$

Hence

$$\mathrm{Var}\big(\mathbb{E}[\overline{X}^* \mid X_1, \ldots, X_n]\big) = \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

Also from part (b),

$$\mathrm{Var}(\overline{X}^* \mid X_1, \ldots, X_n) = \frac{1}{n}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2\Big].$$

Taking expectation over $(X_1, \ldots, X_n)$ yields

$$\mathbb{E}\big[\mathrm{Var}(\overline{X}^* \mid X_1, \ldots, X_n)\big] = \frac{1}{n}\,\mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\Big].$$

Observe that

$$\mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i^2\Big] = \mathbb{E}[X^2] = \sigma^2 + \mu^2,$$

and

$$\mathbb{E}[\overline{X}^2] = \mathrm{Var}(\overline{X}) + \big(\mathbb{E}[\overline{X}]\big)^2 = \frac{\sigma^2}{n} + \mu^2.$$

Hence

$$\mathbb{E}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\Big] = (\sigma^2 + \mu^2) - \Big(\frac{\sigma^2}{n} + \mu^2\Big) = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2\Big(1 - \frac{1}{n}\Big).$$

Thus

$$\mathbb{E}\big[\mathrm{Var}(\overline{X}^* \mid X_1, \ldots, X_n)\big] = \frac{1}{n}\,\sigma^2\left(1 - \frac{1}{n}\right) = \frac{\sigma^2}{n}\left(1 - \frac{1}{n}\right) = \sigma^2\,\frac{n-1}{n^2}.$$

Putting both pieces together for the total variance,

$$\mathrm{Var}\big(\overline{X}^*\big) = \sigma^2\,\frac{n-1}{n^2} + \frac{\sigma^2}{n} = \frac{\sigma^2(n-1)}{n^2} + \frac{\sigma^2}{n} = \frac{\sigma^2(n-1)}{n^2} + \frac{\sigma^2 n}{n^2} = \frac{\sigma^2\big(n-1+n\big)}{n^2} = \frac{\sigma^2(2n-1)}{n^2}.$$