Conditional expectation & conditional variance

(part 1)

Lecture 8a (STAT 24400 F24)

Conditional expectation and conditional distribution

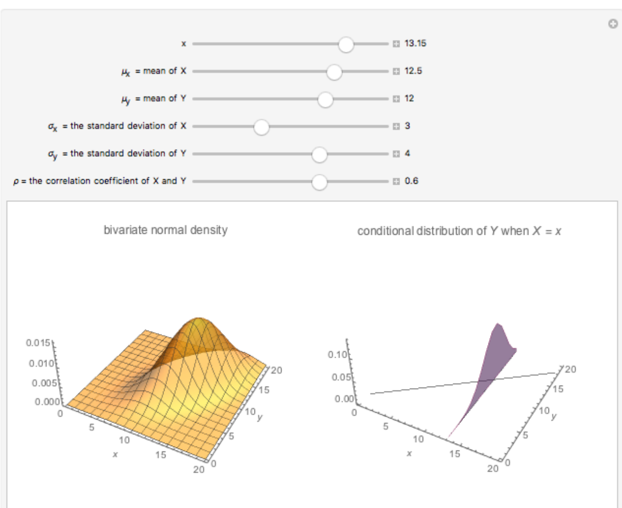
The meaning of **conditional expectation** $\mathbb{E}(X|Y)$:

Expected value of a r.v. X conditional on observing the value of another r.v. Y, i.e. expected value from a conditional distribution.

For a joint distribution on (X, Y), we can ask about the distribution of X conditional on observing $Y = y$.

As for any distribution, we can calculate its expected value:
e.g. from conditional PMF $p_{X|Y}(x|y)$ or conditional density $f_{X|Y}(x|y)$.

Conditional distribution (demo)



Conditional expectation (frequentist interpretation)

Intuitively, we can still think of $\mathbb{E}(X \mid Y = y)$ as a long-run average:

- Imagine drawing (X, Y) from its joint distribution many times
- Now throw out all trials except those for which we got $Y = y$
- Among those trials, what is the average X value?
  These X are from the conditional distribution given $Y = y$.

(This intuition works for discrete Y, where $\mathbb{P}(Y = y) > 0$.
For the continuous case, we can imagine taking limits as it's done before.)

Conditional expectation (discrete case)

Discrete case:

$$\mathbb{E}(X \mid Y = y) = \sum_x x \cdot p_{X|Y}(x \mid y)$$

Similarly,

$$\mathbb{E}(Y \mid X = x) = \sum_y y \cdot p_{Y|X}(y \mid x)$$

More generally

$$\mathbb{E}(g(X) \mid Y = y) = \sum_x g(x) \cdot p_{X|Y}(x \mid y)$$

Recall $p_{X|Y}(x \mid y) = \mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}$, etc.

Conditional expectation (continuous case)

Continuous case:

$$\mathbb{E}(X \mid Y = y) = \int_x x \cdot f_{X|Y}(x \mid y)\, \mathrm{d}x$$

More generally

$$\mathbb{E}(g(X) \mid Y = y) = \int_x g(x) \cdot f_{X|Y}(x \mid y)\, \mathrm{d}x$$

Recall $f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

Conditional variance

We can also define conditional variance,

$$\mathrm{Var}(X \mid Y = y) = \mathbb{E}[(X - \mathbb{E}[X|Y = y])^2 \mid Y = y]$$

or equivalently,

$$\mathrm{Var}(X \mid Y = y) = \mathbb{E}(X^2 \mid Y = y) - \mathbb{E}(X \mid Y = y)^2$$

Intuition: Among all trials of (X, Y), consider the onex with $Y = y$, what is the variability among the corresponding X values?

Note: $\mathbb{E}(X \mid Y = y)^2 = (\mathbb{E}(X \mid Y = y))^2$.

Cnditional expectation and conditional variance as r.v.'s

- $\mathbb{E}(Y \mid X = x)$ is a real function of $x \in \mathbb{R}$.
- When $\mathbb{E}(Y \mid X = x)$ is well defined for all x in the support of X, $\mathbb{E}(Y \mid X)$ is a random variable — a function of random variable X.
- $Var(Y \mid X = x) \geq 0$ is a real function of $x \in \mathbb{R}$.
- When $Var(Y \mid X = x)$ is well defined for all x in the support of X, $Var(Y \mid X) \geq 0$ is a random variable — a function of random variable X.

Tower law for expectations (Law of Total Expectation)

For jointly distributed (X, Y),

$$\mathbb{E}(Y) = \mathbb{E}\left[\mathbb{E}(Y \mid X)\right]$$

What do we mean by these successive expectations?

We can write this more explicitly as

$$\underbrace{\mathbb{E}(Y) = \mathbb{E}_X\left[\overbrace{\mathbb{E}_{Y|X}(Y \mid X)}^{\text{taking expectation over conditional distrib. of } Y \text{ given } X \text{ (the answer is a function of } X\text{)}}\right]}_{\text{taking expectation over marginal distrib. of } X \text{ (i.e., expectation of } \mathbb{E}(Y|X)\text{, which is a function of } X\text{)}}$$

Similarly, we may write $E(X) = \mathbb{E}[\mathbb{E}(X \mid Y)]$.

Examples (Tower law for expectations)

1. Let's play the following game: at each round, you toss a coin.
   - If it's Heads, you roll a die and win \$1 if you rolled a 6
   - If it's Tails, the game ends

   What is the expected amount of money you win?

X = total \$ won, Y = # rounds played. Tower law: $\mathbb{E}(X) = \mathbb{E}\big[\mathbb{E}(X \mid Y)\big]$

$$X \mid Y = y \sim \text{Binomial}(y-1, \tfrac{1}{6}) \quad \Rightarrow \quad \mathbb{E}(X \mid Y = y) = \frac{y-1}{6}.$$

$$\Rightarrow \quad \mathbb{E}(X \mid Y) = \frac{Y-1}{6}.$$

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}\left(\frac{Y-1}{6}\right) = \frac{\mathbb{E}(Y)-1}{6} \underset{\text{since marginally } Y \sim \text{Geometric}(\frac{1}{2})}{=} \frac{2-1}{6} = \frac{1}{6}.$$

Examples (Tower law used for covariance)

We can also use the tower law to calculate $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$:

$$\mathbb{E}(XY \mid Y = y) = \mathbb{E}(X \mid Y = y) \cdot y = \frac{y(y-1)}{6}$$

$$\Rightarrow \quad \mathbb{E}(XY \mid Y) = \frac{Y(Y-1)}{6}$$

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(XY \mid Y)) = \mathbb{E}\left(\frac{Y(Y-1)}{6}\right) = \frac{\mathbb{E}(Y^2)}{6} - \frac{\mathbb{E}(Y)}{6} \underset{\text{calculate using } Y \sim \text{Geometric}(\frac{1}{2})}{=} \frac{6}{6} - \frac{2}{6} = \frac{2}{3}$$

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{2}{3} - \frac{1}{6} \cdot 2 = \frac{1}{3}.$$

Exercise: derive $\mathbb{E}[Y^2]$ when Y is geometric(p).

Examples (conditional density & conditional expectation)

2. Let (X, Y) be supported on the unit square $[0,1]^2$ with density

$$f(x, y) = x + y, \qquad x \in [0,1],\ y \in [0,1].$$

   What is the conditional expectation of $X \mid Y$?

Marginal density:

$$f_X(x) = \int_{y=0}^{1} (x + y)\, dy = x + 1/2\,, \qquad f_Y(y) = y + 1/2$$

Conditional density: for $(x, y) \in [0,1]^2$,

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{x + y}{y + 1/2}$$

Conditional expectation:

$$\mathbb{E}(X \mid Y = y) = \int_{x=0}^{1} x \cdot f_{X|Y}(x \mid y)\, \mathrm{d}x = \int_{x=0}^{1} x \cdot \frac{x+y}{y+1/2}\, \mathrm{d}x = \frac{y/2 + 1/3}{y + 1/2}$$

Tower law for probability

For any event A and any random variable X,

$$\mathbb{P}(A) = \mathbb{E}\big[\mathbb{P}(A \mid X)\big].$$

Proof:
Let $\mathbb{1}_A$ be the indicator variable for the event A.

$$\mathbb{1}_A \sim \text{Bernoulli}(\mathbb{P}(A)) \quad \Longrightarrow \quad \mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$$

Similarly, if we condition on X,

$$\mathbb{1}_A \mid X \sim \text{Bernoulli}(\mathbb{P}(A \mid X)) \quad \Longrightarrow \quad \mathbb{E}(\mathbb{1}_A \mid X) = \mathbb{P}(A \mid X)$$

By the tower law,

$$\mathbb{P}(A) = \mathbb{E}(\mathbb{1}_A) = \mathbb{E}\big[\mathbb{E}(\mathbb{1}_A \mid X)\big] = \mathbb{E}\big[\mathbb{P}(A \mid X)\big].$$

by tower law for expectations,

Examples (apply the tower law for probability)

1. You play the following game: at each round, you toss a coin.
   - If it's Heads, you roll a die and win $1 if you rolled a 6
   - If it's Tails, the game ends

   What is the probability that you win nothing?

X = total \$ won, Y = # rounds played.

$$\mathbb{P}(X=0) = \mathbb{E}\big(\mathbb{P}(X=0 \mid Y)\big) = \mathbb{E}\big((\tfrac{5}{6})^{Y-1}\big)$$

$$= \sum_{y=1}^{\infty} p_Y(y) \cdot (\tfrac{5}{6})^{y-1} = \sum_{y=1}^{\infty} (\tfrac{1}{2})^y (\tfrac{5}{6})^{y-1} = \frac{6}{7}.$$

Examples (apply the tower law for probability)

2. Suppose we build a toy tower of height H, where $H \sim \text{Uniform}[100, 200]$. However, a larger tower is more likely to break; a tower of height h breaks with probability $\frac{h}{300}$, in which case the final height is zero.

   What is the probability that the tower doesn't break?

Let A be the event that the tower doesn't break.

$$\mathbb{P}(A \mid H = h) = \left(1 - \frac{h}{300}\right)$$

$$\mathbb{P}(A) = \mathbb{E}(\mathbb{P}(A \mid H)) = \mathbb{E}\left(1 - \frac{H}{300}\right) = 1 - \frac{\mathbb{E}(H)}{300} = 1 - \frac{150}{300} = \frac{1}{2}$$

Examples (apply the tower law for probability)

What is the expected height of the tower?

Let A be the event that the tower doesn't break. Then the final height is $X = \mathbb{1}_A \cdot H$.

$$\mathbb{E}(X \mid H = h) = \mathbb{E}(\mathbb{1}_A \cdot H \mid H = h) = \mathbb{E}(\mathbb{1}_A \mid H = h) \cdot h = \left(1 - \frac{h}{300}\right) \cdot h$$

$$\mathbb{E}(X) = \mathbb{E}\big[\mathbb{E}(X \mid H)\big] = \mathbb{E}\left[\left(1 - \frac{H}{300}\right) \cdot H\right]$$

$$= \mathbb{E}(H) - \frac{\mathbb{E}(H^2)}{300} = 150 - \frac{23333.33}{300} = 72.22$$