Conditional Expectation

Let X be a random variable and A be an event.

$$E[X] = \sum_x x\, P[X = x]$$

$$E[X \mid A] = \sum_x x\, P[X = x \mid A]$$

Often, the event A will use some other random variable Y.

$$E[X \mid Y = y] = \sum_x x\, P[X = x \mid Y = y]$$

Let $f(y) = E[X \mid Y = y]$ be a function of y.

The function f is the minimum mean square estimate (MMSE) of X.

$E[X \mid Y] = f(Y)$ is a random variable (because it is a function of y).

Law of Total/Iterated Expectation

$$E[X] = E[E[X \mid Y]]$$

$$E[X] = \sum_y E[X \mid Y = y]\, P[Y = y]$$

Linear Least Squares Estimate (LLSE)

$$L(Y \mid X) = E[Y] + \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}\left(X - E[X]\right)$$

1 LLSE

We have two bags of balls. The fractions of red balls and blue balls in bag A are $2/3$ and $1/3$ respectively. The fractions of red balls and blue balls in bag B are $1/2$ and $1/2$ respectively. Someone gives you one of the bags (unmarked) uniformly at random. You then draw 6 balls from that same bag with replacement. Let X_i be the indicator random variable that ball i is red. Now, let us define $X = \sum_{1 \le i \le 3} X_i$ and $Y = \sum_{4 \le i \le 6} X_i$.

(a) Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(b) Compute $\text{Var}(X)$.

(c) Compute $\text{cov}(X, Y)$. (*Hint*: Recall that covariance is bilinear.)

(d) Now, we are going to try and predict Y from a value of X. Compute $L(Y \mid X)$, the best linear estimator of Y given X. (*Hint*: Recall that

$$L(Y \mid X) = \mathbb{E}[Y] + \frac{\text{cov}(X,Y)}{\text{Var}(X)}\big(X - \mathbb{E}[X]\big).$$

)

a) $E[X] = E[X_1 + X_2 + X_3] = E[X_1] + E[X_2] + E[X_3]$

$E[X_i] = \frac{1}{2}\cdot\frac{2}{3} + \frac{1}{2}\cdot\frac{1}{2} = \frac{7}{12}$

$E[X] = \frac{7}{12} + \frac{7}{12} + \frac{7}{12} = \frac{7}{4}$

$E[Y] = \frac{7}{4}$

b) $Var(X) = E[X^2] - E[X]^2 = E[(X_1+X_2+X_3)^2] - \left(\frac{7}{4}\right)^2$

$E[X^2] = E[X_1^2 + X_2^2 + X_3^2 + 2X_1X_2 + 2X_1X_3 + 2X_2X_3]$

$E[X_1X_2] = P[X_1 = 1, X_2 = 1] = \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{2}{3} + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{25}{72}$

$E[X^2] = E[X_1^2] + E[X_2^2] + E[X_3^2] + 2E[X_1X_2] + 2E[X_1X_3] + 2E[X_2X_3]$

$= \frac{7}{12} + \frac{7}{12} + \frac{7}{12} + 2\left(\frac{25}{72}\right) + 2\left(\frac{25}{72}\right) + 2\left(\frac{25}{72}\right) = \frac{37}{48}$

c) $Cov(X,Y) = E[XY] - E[X]E[Y] = E[(X_1+X_2+X_3)(X_4+X_5+X_6)] - \frac{7}{4}\cdot\frac{7}{4}$

$= 9\left(\frac{25}{72}\right) - \frac{49}{16} = \frac{1}{16}$

d) $L(Y|X) = \frac{7}{4} + \frac{3}{37}\left(X - \frac{7}{4}\right) = \frac{3}{37}X + \frac{119}{74}$

2 Number Game

Sinho and Vrettos are playing a game where they each choose an integer uniformly at random from $[0, 100]$, then whoever has the larger number wins (in the event of a tie, they replay). However, Vrettos doesn't like losing, so he's rigged his random number generator such that it instead picks randomly from the integers between Sinho's number and 100. Let S be Sinho's number and V be Vrettos' number.

(a) What is $\mathbb{E}[S]$?

$$E[S] = \frac{0+100}{2} = 50$$

(b) What is $\mathbb{E}[V|S=s]$, where s is any constant such that $0 \le s \le 100$?

$$E[V|S=s] = \frac{s+100}{2}$$

(c) What is $\mathbb{E}[V]$?

$$E[V] = \sum_{s=0}^{100} P[S=s]\, E[V|S=s]$$

$$= \sum_{s=0}^{100} \frac{1}{101} \frac{s+100}{2}$$

$$= \frac{1}{202} \sum_{s=0}^{100} (s+100)$$

$$= \frac{101 \cdot 50 + 101 \cdot 100}{202}$$

$$= 75$$

3 Number of Ones

In this problem, we will revisit dice-rolling, except with conditional expectation.

(a) If we roll a die until we see a 6, how many ones should we expect to see?

(b) If we roll a die until we see a number greater than 3, how many ones should we expect to see?

(Hint: for both of the above subparts, the Law of Total Expectation may be helpful)

X = number of total rolls

Y = number of 1's

$X \sim \text{Geometric}(\frac{1}{6})$

$Y|X \sim \text{Binomial}(X-1, \frac{1}{5})$

a) $E[Y] = E[E[Y|X]]$

$= E[\frac{1}{5}(X-1)]$

$= \frac{1}{5}(E[X]-1)$

$= 1$

b) $E[Y] = E[E[Y|X]]$

$= E[\frac{1}{3}(X-1)]$

$= \frac{1}{3}(E[X]-1)$

$= \frac{1}{3}(2-1) = \frac{1}{3}$

4 Marbles in a Bag

We have r red marbles, b blue marbles, and g green marbles in the same bag. If we sample marbles with replacement until we get 3 red marbles (not necessarily consecutively), how many blue marbles should we expect to see? (*Hint*: It might be useful to use Law of Total Expectation, $E(Y) = E(E(Y|X))$.)

X = number of samples until we get 3 red marbles

Y = number of blue marbles

$X = X_1 + X_2 + X_3 \qquad X_i \sim \text{Geometric}\left(\frac{r}{r+g+b}\right)$

$E[X] = E[X_1] + E[X_2] + E[X_3] = 3\left(\frac{r+g+b}{r}\right)$

$E[Y|X] = \frac{b}{b+g}(X-3)$

$E[Y] = E[E[Y|X]]$

$= E[\frac{b}{b+g}(X-3)]$

$= \frac{b}{b+g}(E[X]-3)$

$= \frac{b}{b+g}\left(\frac{3(r+g+b)}{r} - 3\right) = \frac{3b}{r}$