CS 70 Spring 2017 — Discrete Mathematics and Probability Theory — Rao — DIS 11a

1 Correlation and Independence

(a) What does it mean for two random variables to be uncorrelated?

(b) What does it mean for two random variables to be independent?

(c) Are all uncorrelated variables independent? Are all independent variables uncorrelated?

Solution:

(a) Recall that for two random variables X and Y,

$$\text{cov}(X,Y) = \mathbf{E}\left((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))\right) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y).$$

Two random variables are uncorrelated iff their covariance is equal to zero. If X and Y are uncorrelated, then there is no linear relationship between them.

(b) Recall that two random variables X and Y are independent if and only if the following criteria are met (the three criteria are equivalent and connected by Bayes rule):

$$\Pr(X = x \mid Y = y) = \Pr(X = x)$$

$$\Pr(Y = y \mid X = x) = \Pr(Y = y)$$

$$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$$

for all x, y.

If X and Y are independent, any information about one variable offers no information whatsoever about the other variable.

(c) Note that if two random variables are independent, they must have no relationship whatsoever, including linear relationships; therefore they must be uncorrelated. The converse, however, is not true: two uncorrelated variables may not be independent. Consider two variables X and Y that follow a uniform joint distribution over the points $(1,0),(0,1),(-1,0),(0,-1)$. Then

$$\text{cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = 0 - (0)(0) = 0$$

so there is no linear relationship, but X and Y are not independent (for example, $\Pr(Y = 0) = 1/2$ but $\Pr(Y = 0 \mid X = 1) = 1$).

2 Covariance

We have a bag of 5 red and 5 blue balls. We take two balls from the bag without replacement. Let X_1 and X_2 be indicator random variables for the first and second ball being red. What is $\text{cov}(X_1,X_2)$?

Solution:

We can use the formula $\text{cov}(X_1,X_2) = \mathbf{E}(X_1X_2) - \mathbf{E}(X_1)\mathbf{E}(X_2)$.

$$\begin{aligned} \mathbf{E}(X_1) &= \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2} \\ \mathbf{E}(X_2) &= \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2} \\ \mathbf{E}(X_1X_2) &= \frac{5}{10} \cdot \frac{4}{9} \times 1 + \left(1 - \frac{5}{10} \cdot \frac{4}{9}\right) \times 0 = \frac{2}{9} \end{aligned}$$

Therefore,

$$\mathbf{E}(X_1X_2) - \mathbf{E}(X_1)(X_2) = \frac{2}{9} - \frac{1}{2} \times \frac{1}{2} = -\frac{1}{36}.$$

3 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. Then we draw 6 balls from the 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 $Y = \sum_{4 \le i \le 6} X_i$. Find $L(Y \mid X)$. *Hint*: Recall that

$$L(Y \mid X) = \mathbf{E}(Y) + \frac{\text{cov}(X,Y)}{\text{var}(X)}(X - \mathbf{E}(X)).$$

Solution:

Note that although the indicator random variables are not independent, we can still apply linearity of expectation. By symmetry, we also know that each indicator follows the same distribution. Therefore:

$$\begin{aligned} \mathbf{E}(X) &= 3 \cdot \mathbf{E}(X_1) \\ &= 3 \cdot \Pr(X_1 = 1) \\ &= 3 \cdot \left(\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{2}\right) \\ &= \frac{7}{4}. \end{aligned}$$

$$\mathbf{E}(Y) = \mathbf{E}(X) = \frac{7}{4}.$$

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \operatorname{cov}\left(\sum_{1\le i\le 3} X_i, \sum_{4\le j\le 6} X_j\right) \\
&= 9\cdot \operatorname{cov}(X_1,X_4) \\
&= 9\cdot (\mathbf{E}(X_1X_4) - \mathbf{E}(X_1)\cdot \mathbf{E}(X_4)).
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}(X_1X_4) - \mathbf{E}(X_1)\mathbf{E}(X_4) &= \Pr(X_1 = 1, X_4 = 1) - \Pr(X_1 = 1)^2 \\
&= \left[\frac{1}{2}\cdot\left(\frac{2}{3}\right)^2 + \frac{1}{2}\cdot\left(\frac{1}{2}\right)^2\right] - \left[\frac{1}{2}\cdot\left(\frac{2}{3}\right) + \frac{1}{2}\cdot\left(\frac{1}{2}\right)\right]^2 \\
&= \frac{1}{144}.
\end{aligned}$$

$$\begin{aligned}
\operatorname{var}(X) &= \operatorname{cov}\left(\sum_{1\le i\le 3} X_i, \sum_{1\le j\le 3} X_j\right) \\
&= 3\cdot \operatorname{var}(X_1) + 6\cdot \operatorname{cov}(X_1,X_2) \\
&= 3(\mathbf{E}(X_1^2) - \mathbf{E}(X_1)^2) + 6\cdot\frac{1}{144} \\
&= 3\left(\frac{7}{12} - \left(\frac{7}{12}\right)^2\right) + 6\cdot\frac{1}{144} \\
&= \frac{111}{144}.
\end{aligned}$$

So,

$$L(Y \mid X) = \frac{7}{4} + \frac{9}{111}\left(X - \frac{7}{4}\right) = \frac{3}{37}X + \frac{119}{74}.$$