5. Covariance of the multinomial

Problem 5. Covariance of the multinomial

1/3 points (graded)

Consider n independent rolls of a k-sided fair die with $k \geq 2$: the sides of the die are labelled $1, 2, \ldots, k$, and each side has probability $1/k$. Let the random variable X_i denote the number of rolls that result in side i. Thus, the random vector $(X_1, \ldots, X_k)$ has a multinomial distribution.

1. Which of the following statements is correct? Try to answer without doing any calculations.

- X_1 and X_2 are uncorrelated.
- X_1 and X_2 are positively correlated.
- X_1 and X_2 are negatively correlated. ✔

2. Find the covariance, $\textbf{cov}(X_1, X_2)$, of X_1 and X_2. Express your answer as a function of n and k using standard notation. *Hint:* Use indicator variables to encode the result of each roll.

$\textbf{cov}(X_1, X_2) =$ 0 ✖ **Answer:** -n/(k^2)

0

3. Suppose now that the die is biased, with a probability $p_i \neq 0$ that the result of any given die roll is i, for $i = 1, 2, \ldots, k$. We still consider n independent rolls of this biased die and define X_i to be the number of rolls that result in side i.

Generalize your answer to part 2: Find $\textbf{cov}(X_1, X_2)$ for this case of a biased die. Express your answer as a function of n, k, p_1, p_2 using standard notation. Write p_1 and p_2 as p_1 and p_2, respectively, for example, $2p_1p_2$ must be entered as 2*p_1*p_2.

$\textbf{cov}(X_1, X_2) =$ ✖ **Answer:** -n*(p_1)*(p_2)

STANDARD NOTATION

Solution:

1. The random variables X_1 and X_2 are negatively correlated. There is a fixed number, n, of rolls of the die. Intuitively, a large number of rolls that result in a 1 uses up many of the n total rolls, which leaves fewer remaining rolls that could result in a 2.

2. Let A_t (respectively, B_t) be a Bernoulli random variable that is equal to 1 if and only if the tth roll resulted in a 1 (respectively, 2). Note that $X_1 = \sum_{t=1}^{n} A_t$ and $X_2 = \sum_{t=1}^{n} B_t$, and so

$$\mathbf{E}[X_1] = \mathbf{E}[X_2] = \mathbf{E}\left[\sum_{t=1}^{n} A_t\right] = n\mathbf{E}[A_1] = \frac{n}{k}.$$

Since a single roll of the die cannot result in both a 1 and a 2, at least one of A_t and B_t must equal 0. Thus, $\mathbf{E}[A_t B_t] = 0$. Furthermore, since different rolls are independent, A_t and B_s are independent when $t \neq s$. Therefore,

$$\mathbf{E}[A_t B_s] = \mathbf{E}[A_t]\mathbf{E}[B_s] = \frac{1}{k} \cdot \frac{1}{k} = \frac{1}{k^2} \qquad \text{for} \quad t \neq s,$$

and so

$$\begin{aligned}
\mathbf{E}[X_1 X_2] &= \mathbf{E}\left[(A_1 + \cdots + A_n)(B_1 + \cdots + B_n)\right] \\
&= \mathbf{E}\left[\sum_{t=s} A_t B_t + \sum_{t \neq s} A_t B_s\right] \\
&= n \cdot 0 + n(n-1) \cdot \mathbf{E}[A_1 B_2] \\
&= n(n-1) \cdot \frac{1}{k^2}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{cov}(X_1, X_2) &= \mathbf{E}[X_1 X_2] - \mathbf{E}[X_1]\mathbf{E}[X_2] \\
&= n(n-1) \cdot \frac{1}{k^2} - \frac{n}{k} \cdot \frac{n}{k} \\
&= -\frac{n}{k^2}.
\end{aligned}$$

The covariance of X_1 and X_2 is negative as expected.

3. We follow the same reasoning as in part 2. Let A_t (respectively, B_t) be a Bernoulli random variable that is equal to 1 if and only if the tth roll resulted in a 1 (respectively, 2). As in part 2, a single roll of the die cannot result in both a 1 and a 2, so $\mathbf{E}[A_t B_t] = 0$. Different rolls of the die are independent, and so $\mathbf{E}[A_t B_s] = \mathbf{E}[A_t]\mathbf{E}[B_s] = p_1 \cdot p_2$, for $t \neq s$. Thus,

$$\begin{aligned}
\mathbf{E}[X_1X_2] &= \mathbf{E}\left[(A_1+\cdots+A_n)(B_1+\cdots+B_n)\right] \\
&= \mathbf{E}\left[\sum_{t=s} A_tB_t + \sum_{t\neq s} A_tB_s\right] \\
&= n\cdot 0 + n(n-1)\cdot \mathbf{E}[A_1B_2] \\
&= n(n-1)p_1p_2.
\end{aligned}$$

Note that $X_1 = \sum_{t=1}^{n} A_t$ and $X_2 = \sum_{t=1}^{n} B_t$, and so $\mathbf{E}[X_1] = \mathbf{E}\left[\sum_{t=1}^{n} A_t\right] = n\mathbf{E}[A_1] = np_1$. Similarly, $\mathbf{E}[X_2] = np_2$.

Therefore,

$$\begin{aligned}
\mathrm{cov}(X_1, X_2) &= \mathbf{E}[X_1X_2] - \mathbf{E}[X_1]E[X_2] \\
&= n(n-1)p_1p_2 - (np_1)(np_2) \\
&= -np_1p_2.
\end{aligned}$$

The covariance of X_1 and X_2 is again negative, even when the die is not fair.

提交 You have used 2 of 3 attempts

Answers are displayed within the problem

讨论

Topic: Unit 6 / Problem Set / 5. Covariance of the multinomial

显示讨论

© All Rights Reserved