Probability and Statistics Homework #9

Ari Feiglin

Question 9.1:

A fair coin and fair die are both flipped and rolled, respectively. We define X to indicate the result of the coin toss, and Y to indicate whether or not the result of the die is even. X and Y are independent. We define $Z = X + Y$. Compute $\mathrm{Var}\left(Z\right)$.

Answer:

Since X and Y are independent:

$$\mathrm{Var}\left(Z\right) = \mathrm{Var}\left(X + Y\right) = \mathrm{Var}\left(X\right) + \mathrm{Var}\left(Y\right)$$

Furthermore, we know $X, Y \sim \mathrm{Ber}\left(\frac{1}{2}\right)$, so $\mathrm{Var}\left(X\right) = \mathrm{Var}\left(Y\right) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$.
Therefore:

$$\mathrm{Var}\left(Z\right) = \frac{1}{4} + \frac{1}{4} = \boxed{\frac{1}{2}}$$

$\blacksquare$

Question 9.2:

X and Y are random variables with a joint mass probability distribution of:

$$P_{X,Y}(x,y) = \begin{cases} k(2x+y) & x,y=1,2 \\ 0 & \text{else} \end{cases}$$

(1) Find the value of k.

(2) Compute $\text{Var}(X+Y)$.

(3) Are X and Y dependent?

Answer:

(1) We know that:

$$P_{X,Y}(1,1) + P_{X,Y}(1,2) + P_{X,Y}(2,1) + P_{X,Y}(2,2) = 1$$

Which means:

$$k(2+1+2+2+4+1+4+2) = 1 \implies 18k = 1 \implies \boxed{k = \frac{1}{18}}$$

(2) First, we must find the mass probability functions of X and Y.
We know:

$$\mathbb{P}(X=x) = \mathbb{P}(X=x, Y=1) + \mathbb{P}(X=x, Y=2) = k(2x+1+2x+2) = k(4x+3)$$

And:

$$\mathbb{P}(Y=y) = \mathbb{P}(X=1, Y=y) + \mathbb{P}(X=2, Y=y) = k(2+y+4+y) = k(6+2y)$$

Notice that:

$$\mathbb{P}(X-1=1) = \mathbb{P}(X=2) = 11k$$

And:

$$\mathbb{P}(Y-1=1) = \mathbb{P}(Y=2) = 10k$$

Which means that $X \sim \text{Ber}(11k)$ and $Y \sim \text{Ber}(10k)$ since they have a support of $\{0,1\}$.
Now notice that:

$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[(X-1)^2 + 2x - 2 + 1\right] = \mathbb{E}\left[(X-1)^2\right] + 2\mathbb{E}[X-1] + 1 = 3\mathbb{E}[X-1] + 1 = 33k + 1$$

And similarly $\mathbb{E}\left[Y^2\right] = 30k + 1$.
Furthermore:

$$\mathbb{E}[XY] = \mathbb{P}(X=1, Y=1) + 2\left(\mathbb{P}(X=1, Y=2) + \mathbb{P}(X=2, Y=1)\right) + 4\mathbb{P}(X=2, Y=2) =$$

$$= k(2+1+2(2+2+4+1)+4(4+2)) = k(3+18+24) = 45k$$

This means that:

$$\mathbb{E}\left[(X+Y)^2\right] = \mathbb{E}\left[X^2 + 2XY + Y^2\right] = \mathbb{E}\left[X^2\right] + 2\mathbb{E}[XY] + \mathbb{E}\left[Y^2\right] = 33k + 1 + 90k + 30k + 1 = 153k + 2$$

And:

$$\mathbb{E}[X+Y] = \mathbb{E}[X-1] + \mathbb{E}[Y-1] + 2 = 10k + 11k + 2 = 21k + 2$$

Therefore:

$$\text{Var}(X+Y) = \mathbb{E}\left[(X+Y)^2\right] - \mathbb{E}[X+Y]^2 = 153k + 2 - (21k+2)^2 = -441k^2 + 69k - 2$$

Plugging in $k = \frac{1}{18}$, we get:

$$\boxed{\text{Var}(X+Y) = \frac{17}{36}}$$

(3) Yes they are:

$$P_{X,Y}(1,1) = k(2+1) = \frac{1}{6}$$

But:

$$P_X(1) = \frac{7}{18}$$

And

$$P_Y(1) = \frac{8}{18}$$

Which don't multiply to $\frac{1}{6}$.

$\blacksquare$

Question 9.3:

We are given $n+1$ urns numbered from 0 to n. Each urn has n balls. In the kth urn there are k white balls, the rest are black. We remove a single ball from each urn. What is the expected number and variance of the number of white balls we remove? What is the expected number and variance of the number of black balls we remove?

Answer:

Suppose instead each urn has $m \geq n$ balls.
Let X be the number of white balls we removed in total, and let X_k indicate if we removed a white ball from the kth urn. Then:

$$X = \sum_{k=0}^{n} X_k$$

And we know that in the kth urn there are k white balls and m balls total. Therefore

$$X_k \sim \text{Ber}\left(\frac{k}{m}\right)$$

This means:

$$\mathbb{E}\left[X\right] = \sum_{k=0}^{n} \mathbb{E}\left[X_k\right] = \sum_{k=0}^{n} \frac{k}{m} = \frac{n(n+1)}{2m}$$

And since the X_ks are independent:

$$\text{Var}\left(X\right) = \sum_{k=0}^{n} \text{Var}\left(X_k\right) = \sum_{k=0}^{n} \frac{k}{m} - \frac{k^2}{m^2} = \frac{n(n+1)}{2m} - \frac{n(n+1)(2n+1)}{6m^2} = \frac{n(n+1)}{2m}\left(1 - \frac{2n+1}{3m}\right)$$

let Y be the number of black balls we removed. And we know that in total we removed $n+1$ balls, so $Y = n+1-X$. Therefore:

$$\mathbb{E}\left[Y\right] = n+1-\mathbb{E}\left[X\right] = n+1-\frac{n(n+1)}{2m} = \frac{(n+1)(2m-n)}{2m}$$

And:

$$\text{Var}\left(Y\right) = \text{Var}\left(n+1-X\right) = \text{Var}\left(-X\right) = \text{Var}\left(X\right)$$

Plugging in $m = n$ gives:

$$\boxed{\begin{array}{rl} \mathbb{E}\left[Y\right] = \ \mathbb{E}\left[X\right] & = \frac{n+1}{2} \\ \text{Var}\left(X\right) = \text{Var}\left(Y\right) & = \frac{n+1}{2}\left(1 - \frac{2n+1}{3n}\right) \end{array}}$$

■

Question 9.4:

There are 200 sweets: 100 of which are candies and 100 which are chocolates. The sweets are divided randomly uniformly among 100 children, each of which receive 2. Let X be the number of children who received 2 candies and Y be the number of children who received one candy and one chocolate.

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

(2) Compute $\operatorname{Cov}(X, Y)$.

(3) Compute $\operatorname{Var}(X)$ and $\operatorname{Var}(Y)$.

Answer:

Let X_i indicate whether the ith child received 2 candies, and let Y_i indicate whether the ith child received one candy and one chocolate.
We know that

$$\mathbb{P}(X_i = 1) = \frac{n}{2n} \cdot \frac{n-1}{2n-1} = \frac{n-1}{4n-2}$$

Since we miust first choose one of the n candies out of $2n$ sweets, and then $n-1$ out of $2n-1$. Therefore:

$$X_i \sim \text{Ber}\left(\frac{n-1}{4n-2}\right)$$

And:

$$X = \sum_{i=1}^{n} X_i$$

Since the sum counts the number of children who received two candies.
And:

$$\mathbb{P}(Y_i = 1) = 2 \cdot \frac{n}{2n} \cdot \frac{n}{2n-1} = \frac{n}{2n-1}$$

For a similar reason (and there are two ways to choose a candy and a chocolate, depending on the order). Therefore:

$$Y_i \sim \text{Ber}\left(\frac{n}{2n-1}\right)$$

And:

$$Y = \sum_{i=1}^{n} Y_i$$

Since the sum counts the number of children who received a candy and a chocolate.
Furthermore, let Z be the number of children who received 2 chocolates. By symmetry we know that:

$$X \overset{d}{==} Z$$

Since there is an equal amount of candies as chocolates.
Furthermore, since each child receives either 2 candies (X), a candy and a chocolate (Y), or 2 chocolates (Z), and since there are n children, we know that:

$$X + Y + Z = n$$

(1) By the linearity of expected value:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{n-1}{4n-2} = \frac{n(n-1)}{4n-2} = \frac{n^2-n}{4n-2}$$

And similarly:

$$\mathbb{E}[Y] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = \sum_{i=1}^{n} \frac{n}{2n-1} = \frac{n^2}{2n-1}$$

Plugging in $n = 100$ yields:

$$\boxed{\mathbb{E}[X] = \frac{4950}{190} \quad \mathbb{E}[Y] = \frac{10{,}000}{199}}$$

(2) We know that:

$$\text{Cov}(X, Y) = \text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, Y_j)$$

And by the definition of covariance:

$$\text{Cov}(X_i, Y_j) = \mathbb{E}[X_i Y_j] - \mathbb{E}[X_i]\,\mathbb{E}[Y_j]$$

Since X_i and Y_j are indicators, so is $X_i Y_j$. And:

$$\mathbb{E}[X_i Y_j] = \mathbb{P}(X_i = 1, Y_j = 1) = \frac{n}{2n} \cdot \frac{n-1}{2n-1} \cdot 2 \cdot \frac{n-2}{2n-2} \cdot \frac{n}{2n-3} = \frac{n(n-2)}{2(2n-1)(2n-3)}$$

If $i \neq j$. If $i = j$ then the same child can't both get two candies and a candy and a chocolate, so the probability in this case is 0.
So if $i \neq j$ then:

$$\text{Cov}(X_i, Y_j) = \frac{n(n-2)}{2(2n-1)(2n-3)} - \frac{n-1}{4n-2} \cdot \frac{n}{2n-1}$$

And if $i = j$ then:

$$\text{Cov}(X_i, Y_j) = -\frac{n-1}{4n-2} \cdot \frac{n}{2n-1}$$

Thus:

$$\text{Cov}(X, Y) = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{n(n-2)}{2(2n-1)(2n-3)} - \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{n(n-1)}{2(2n-1)^2} =$$

$$\frac{n^2(n-1)(n-2)}{2(2n-1)(2n-3)} - \frac{n^3(n-1)}{2(2n-1)^2}$$

Upon simplifying, we get that:

$$\text{Cov}(X, Y) = -\frac{n^2(n-1)^2}{(2n-1)^2(2n-3)}$$

Plugging in $n = 100$ yields:

$$\boxed{\text{Cov}(X, Y) \approx -12.563}$$

(3) First notice that:

$$\text{Cov}(X, Y) = \text{Cov}(n - Y - Z, Y) = \text{Cov}(n, Y) - \text{Cov}(Y, Y) - \text{Cov}(Z, Y) = -\text{Var}(Y) - \text{Cov}(Z, Y)$$

By symmetry, $\text{Cov}(Z, Y)$ since there is no real difference between candies and chocolates, and Y is independent of the choice of candy or chocolate in a way. So:

$$2\text{Cov}(X, Y) = -\text{Var}(Y) \implies \text{Var}(Y) = -2\text{Cov}(X, Y) = \frac{2n^2(n-1)^2}{(2n-1)^2(2n-3)}$$

And we also know:

$$\text{Cov}(X, Z) = \text{Cov}(X, n - X - Y) = -\text{Var}(X) - \text{Cov}(X, Y)$$

$$\implies \text{Var}(X) = -\text{Cov}(X, Y) - \text{Cov}(X, Z)$$

So let's compute $\text{Cov}(X, Z)$. By linearity:

$$\text{Cov}(X, Z) = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, Z_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[X_i Z_j] - \mathbb{E}[X_i]\,\mathbb{E}[Z_j]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{P}\left(X_i = 1, Z_j = 1\right) - n^2 \cdot \frac{(n-1)^2}{4(2n-1)^2}$$

And since if $i = j$, $\mathbb{P}\left(X_i = 1, Z_j = 1\right) = 0$, and if $i \neq j$:

$$\mathbb{P}\left(X_i = 1, Z_j = 1\right) = \frac{n}{2n} \cdot \frac{n-1}{2n-1} \cdot \frac{n-2}{2n-2} \cdot \frac{n-3}{2n-3} = \frac{(n-2)(n-3)}{4(2n-1)(2n-3)}$$

We get:

$$\mathrm{Cov}\left(X, Z\right) = \frac{n(n-1)(n-2)(n-3)}{4(2n-1)(2n-3)} - \frac{n^2(n-1)^2}{4(2n-1)^2}$$

So:

$$\mathrm{Var}\left(X\right) = \frac{n^2(n-1)^2}{(2n-1)^2(2n-3)} - \frac{n(n-1)(n-2)(n-3)}{4(2n-1)(2n-3)} + \frac{n^2(n-1)^2}{4(2n-1)^2}$$

Plugging in $n = 100$ yields:

$$\boxed{\mathrm{Var}\left(X\right) \approx 31.156 \quad \mathrm{Var}\left(Y\right) \approx 25.126}$$

$\blacksquare$

Question 9.5:

A mailman must distribute n letters to n people, and does so by putting each letter in a mailbox. Alas, the mailman cannot read and therefore does not knw to which mailbox to distribute the letters. Therefore he distributes the letters randomly.

(1) If the mailman can only put one letter in a mailbox, what is the expected value and variance of the number of letters which will arrive at their intended recipients?

(2) After the mailman puts each letter in a mailbox, he receives a hit on the head and forgets which mailboxes he already distributed letters to, and can therefore put multiple letters in a single mailbox. What is the expected value and variance of the number of letters which will arrive at their intended recipients?

Answer:

Let X be the number of letters which arrived at their intended recipients. Let X_i indicate whether the ith letter arrived at its intended recipient. Thus:

$$X = \sum_{i=1}^{n} X_i$$

(1) In this case, the probability that X_i equals 1 is $\frac{1}{n}$, as it is the number of permutations where $\sigma(i) = i$, which is equal to:

$$\frac{(n-1)!}{n!} = \frac{1}{n}$$

Thus:

$$X_i \in \mathrm{Ber}\left(\frac{1}{n}\right)$$

So:

$$\mathbb{E}\left[X\right] = \sum_{i=1}^{n} \frac{1}{n} = n \cdot \frac{1}{n} = 1$$

That is:

$$\boxed{\mathbb{E}\left[X\right] = 1}$$

Notice that if $i \neq j$, then:

$$\mathrm{Cov}\left(X_i, X_j\right) = \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right] = \mathbb{P}\left(X_i = 1, X_j = 1\right) - \frac{1}{n^2}$$

And $\mathbb{P}\left(X_i = 1, X_j = 1\right) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$, so:

$$\mathrm{Cov}\left(X_i, X_j\right) = \frac{1}{n(n-1)} = \frac{1}{n^2} = \frac{1}{n^2(n-1)}$$

So:

$$\mathrm{Var}\left(X\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) + \sum_{i \neq j} \mathrm{Cov}\left(X_i, X_j\right) = \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{1}{n^2}\right) + \sum_{i \neq j} \frac{1}{n^2(n-1)} = 1 - \frac{1}{n} + \frac{1}{n} = 1$$

That is:

$$\boxed{\mathrm{Var}\left(X\right) = 1}$$

(2) In this case, the probability that X_i equals 1 is also $\frac{1}{n}$, as it is the number of functions where $f(i) = i$, which is equal to:

$$\frac{n^{n-1}}{n^n} = \frac{1}{n}$$

So in this case as well:

$$\boxed{\mathbb{E}\left[X\right] = 1}$$

And notice that for all $i \neq j$, X_i and X_j are independent. This is because the event $X_i = 1, X_j = 1$ is the number of functions where $f(i) = i$ and $f(j) = j$, whose probability is:

$$\frac{n^{n-2}}{n^n} = \frac{1}{n^2}$$

So:

$$\mathbb{P}\left(X_i = 1, X_j = 1\right) = \frac{1}{n^2} = \mathbb{P}\left(X_i = 1\right) \cdot \mathbb{P}\left(X_j = 1\right)$$

So:

$$\operatorname{Var}\left(X\right) = \sum_{i=1}^{n} \operatorname{Var}\left(X_i\right) = \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{1}{n^2}\right) = 1 - \frac{1}{n}$$

That is:

$$\boxed{\operatorname{Var}\left(X\right) = 1 - \frac{1}{n}}$$

■

Question 9.6:

In an urn there are 2 red balls, 3 white balls, and 4 blue balls. Three random balls are removed from the urn. Let X be the number of red balls removed, and Y be the number of white balls removed.

(1) What is the range of (X, Y)?

(2) What is the joint probability distribution of (X, Y)?

(3) What is the expected value and variance of each variable?

(4) What is the pearson correlation coefficient between X and Y?

Answer:

Let's generalize. Suppose instead there are n different colors of balls in the urn, and a_i balls of the ith color. Suppose we remove m balls from the urn, let X_i be the number of balls of color i removed.
Let:

$$A \coloneqq \sum_{i=1}^{n} a_i$$

Which is the total number of balls in the urn.
Thus:

$$X_i \sim \mathrm{HG}\left(A, a_i, m\right)$$

As it counts the number of balls of color i (of which there are a_i) removed among the m removed from the A total balls.

Lemma 9.6.1:

Suppose $X \sim \mathrm{HG}\left(N, D, n\right)$, then:

(1) $\mathbb{E}\left[X\right] = \frac{nD}{N}$

(2) $\mathrm{Var}\left(X\right) = \frac{nD(N-n)(N-D)}{N^2(N-1)}$

Proof:

Let X represent the number of special balls (of which there are D) removed among m removals of balls from an urn with N balls. Thus:

$$X \sim \mathrm{HG}\left(N, D, n\right)$$

Let X_i indicate whether a "special" ball was removed for the ith removal. The probability of this is $\frac{D}{N}$, thus:

$$X_i \sim \mathrm{Ber}\left(\frac{D}{N}\right)$$

And by definition:

$$X = \sum_{i=1}^{n} X_i$$

Thus:

$$\mathbb{E}\left[X\right] = \sum_{i=1}^{n} \mathbb{E}\left[X_i\right] = n \cdot \frac{D}{N} = \frac{nD}{N}$$

As required.
And by definition we know that if $i \neq j$:

$$\mathrm{Cov}\left(X_i, X_j\right) = \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right] = \mathbb{P}\left(X_i = 1, X_j = 1\right) - \frac{D^2}{N^2}$$

And we know that:

$$\mathbb{E}\left[X_i = 1, X_j = 1\right] = \frac{D}{N} \cdot \frac{D-1}{N-1}$$

So:

$$\mathrm{Cov}\left(X_i, X_j\right) = \frac{D(D-1)}{N(N-1)} - \frac{D^2}{N^2}$$

And so:

$$\mathrm{Var}\left(X\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) + \sum_{i \neq j} \mathrm{Cov}\left(X_i, X_j\right) = \sum_{i=1}^{n} \left(\frac{D}{N} - \frac{D^2}{N^2}\right) + \sum_{i \neq j} \left(\frac{D(D-1)}{N(N-1)} - \frac{D^2}{N^2}\right)$$

$$= \frac{nD}{N} - \frac{nD^2}{N^2} + \frac{n(n-1)D(D-1)}{N(N-1} - \frac{n(n-1)D^2}{N^2} = \frac{nD}{N}\left(1 - \frac{D}{N} + \frac{(n-1)(D-1)}{N-1} - \frac{(n-1)D}{N}\right)$$

$$= \frac{nD}{N}\left(1 + \frac{(n-1)(D-1)}{N-1} - \frac{nD}{N}\right)$$

Which upon simplifying, equals:

$$= \frac{nD}{N} \cdot \frac{(N-n)(N-D)}{N(N-1)} = \frac{nD(N-n)(N-D)}{N^2(N-1)}$$

As required.

$\blacksquare$

(1) The range of (X_i, X_j) is all pairs (x_1, x_2) such that $0 \leq x_i \leq a_i$ and $x_1 + x_2 \leq m$, and $m - x_1 - x_2 \leq A - a_i - a_j \iff m - A + a_i + a_j \leq x_1 + x_2$.
So it is equal to the set:

$$\{(x_i, x_j) \in [0, a_i] \times [0, a_j] \mid m - A + a_i + a_j \leq x_i + x_j \leq m\}$$

In our case, $m = 3$, $a_i = 2$, $a_j = 3$, and $A = 9$. So we're requiring $0 \leq x_i + x_j \leq 3$. So the range is equal to the set:

$$\boxed{\left\{\begin{array}{l}(0,0),(0,1),(0,2),(0,3),\\(1,0),(1,1),(1,2),\\(2,0),(2,1)\end{array}\right\}}$$

(2) The number of ways to choose x_i balls of color i and x_j of color j is:

$$\binom{a_i}{x_i} \cdot \binom{a_j}{x_j} \cdot \binom{A - a_i - a_j}{m - x_i - x_j}$$

Since we must choose x_i and x_j balls of color i and j respectively, then $m - x_i - x_j$ balls of other colors. Since the probability of choosing m balls is independent of the balls chosen (ignoring color), the probability of choosing x_i balls of color i and x_j of color j is:

$$\frac{\binom{a_i}{x_i} \cdot \binom{a_j}{x_j} \cdot \binom{A-a_i-a_j}{m-x_i-x_j}}{\binom{A}{m}}$$

In our case, $m = 3$, a_i=2, $a_j = 3$, and $A = 9$, so:

$$\boxed{P_{X,Y}\left(x, y\right) = \frac{\binom{2}{x} \cdot \binom{3}{y} \cdot \binom{4}{3-x-y}}{84}}$$

(3) By lemma 9.6.1:

$$\mathbb{E}\left[X_i\right] = \frac{m \cdot a_i}{A}$$

So in our case:

$$\boxed{\mathbb{E}\left[X\right] = \frac{2}{3} \quad \mathbb{E}\left[Y\right] = 1}$$

And by lemma 9.6.1:

$$\mathrm{Var}\left(X_i\right) = \frac{ma_i(A-m)(A-a_i)}{A^2(A-1)}$$

So in our case:

$$\boxed{\text{Var}(X) = \frac{7}{18} \quad \text{Var}(Y) = \frac{1}{2}}$$

(4) Let X_k^i indicate if a ball of color i was removed for the kth ball removed. This means:

$$X_i = \sum_{k=1}^{m} X_k^i$$

And:

$$\text{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]$$

And:

$$\mathbb{E}[X_i X_j] = \sum_{k=1}^{m}\sum_{\ell=1}^{m} \mathbb{E}\left[X_k^i X_\ell^j\right]$$

Since $X_k^i X_\ell^j$ is an indicator:

$$\mathbb{E}\left[X_k^i X_\ell^j\right] = \mathbb{P}\left(X_k^i = 1, X_\ell^j = 1\right) = \frac{a_i}{A} \cdot \frac{a_j}{A-1}$$

If $k \neq \ell$, and if $k = \ell$ it is equal to 0 (since they are disjoint).
Thus:

$$\mathbb{E}[X_i X_j] = m(m-1) \cdot \frac{a_i a_j}{A(A-1)}$$

So:

$$\text{Cov}(X_i, X_j) = m(m-1) \cdot \frac{a_i a_j}{A(A-1)} - \frac{m^2 a_i a_j}{A^2} = \frac{m a_i a_j}{A}\left(\frac{m-1}{A-1} - \frac{m}{A}\right) = \frac{m a_i a_j (m-A)}{A^2(A-1)}$$

Which means:

$$\rho(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i) \cdot \text{Var}(X_j)}} = -\sqrt{\frac{a_i a_j}{(A-a_i)(A-a_j)}}$$

Plugging in $a_i = 2$, $a_j = 3$, and $A = 9$, we get:

$$\boxed{\rho(X, Y) = -\frac{1}{\sqrt{7}}}$$

■

Question 9.7:

A fair coin is flipped n times. Let X be the number of heads and Y be the number of tails flipped. What is $\mathrm{Cov}(X, Y)$?

Answer:

We know that $X + Y = n$, so:

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, n - X) = \mathrm{Cov}(X, n) - \mathrm{Cov}(X, X) = -\mathrm{Var}(X)$$

And we know that X counts the number of heads we flipped out of n, so:

$$X \sim \mathrm{Bin}\left(n, \frac{1}{2}\right)$$

Which means that:

$$\mathrm{Var}(X) = n\left(\frac{1}{2} - \frac{1}{4}\right) = \frac{n}{4}$$

So:

$$\boxed{\mathrm{Cov}(X, Y) = -\frac{n}{4}}$$

■

Question 9.8:

Two coins are flipped three times. Coin A us a fair coin, but B is weighted so that the probability is lands on heads is $\frac{1}{4}$ and the probability it lands on tails is $\frac{3}{4}$. We define X to be the number of heads we flipped, and Y as the number of tails.

(1) What are the expected values of X and Y?

(2) What are the variances of X and Y?

(3) Are X and Y dependent?

Answer:

Let's generalize. Suppose coin A has a probability of flipping heads of p, and coin B of q. Furthermore, suppose we flip each coin n times.
Firstly, notice that $X + Y = 2n$ as there are $2n$ total flips.
Let X_A be the number of heads flipped on coin A and X_B be the number of heads flipped with coin B. Thus:

$$X = X_A + X_B$$

Furthermore, since these count the number of "successes" in n bernoulli trials:

$$X_A \sim \text{Bin}\left(n, p\right) \quad X_B \sim \text{Bin}\left(n, q\right)$$

And we know that the two coins are independent so $X_A \perp\!\!\!\perp X_B$.

(1) By the linearity of the expected value:

$$\mathbb{E}\left[X\right] = \mathbb{E}\left[X_A\right] + \mathbb{E}\left[X_B\right] = np + nq = n(p+q)$$

Since X_A and X_B are binomial. And since $Y = 2n - X$:

$$\mathbb{E}\left[Y\right] = 2n - \mathbb{E}\left[Y\right] = n(2 - p - q)$$

Plugging in $n = 3$, $p = \frac{1}{2}$ and $q = \frac{1}{4}$, we get:

$$\boxed{\mathbb{E}\left[X\right] = \frac{9}{4} \quad \mathbb{E}\left[Y\right] = \frac{15}{4}}$$

(2) Firstly, note that:

$$\text{Var}\left(Y\right) = \text{Var}\left(2n - X\right) = \text{Var}\left(X\right)$$

And since X_A and X_B are independent:

$$\text{Var}\left(Y\right) = \text{Var}\left(X\right) = \text{Var}\left(X_A + X_B\right) = \text{Var}\left(X_A\right) + \text{Var}\left(X_B\right) = n(p - p^2) + n(q - q^2) = n(p + q - p^2 - q^2)$$

Plugging in $n = 3$, $p = \frac{1}{2}$, and $q = \frac{1}{4}$, we get:

$$\boxed{\text{Var}\left(X\right) = \text{Var}\left(Y\right) = \frac{21}{16}}$$

(3) Yes, since:

$$\mathbb{P}\left(X = 2n, Y = 2n\right) = 0$$

Since if $X = 2n$ then $Y = 0$.

$\blacksquare$