Board questions

1. Prove: if $X \sim \text{Bernoulli}(p)$ then $\text{Var}(X) = p(1-p)$.

2. Prove: if $X \sim \text{bin}(n, p)$ then $\text{Var}(X) = n\,p(1-p)$.

3. Suppose $X_1, X_2, \ldots, X_n$ are independent and all have the same standard deviation $\sigma = 2$. Let $\overline{X}$ be the average of $X_1, \ldots, X_n$.

What is the standard deviation of $\overline{X}$?

Solution

1. For $X \sim \text{Bernoulli}(p)$ we use a table. (We know $E(X) = p$.)

X	0	1
$p(x)$	$1-p$	p
$(X-\mu)^2$	p^2	$(1-p)^2$

$$\text{Var}(X) = E((X-\mu)^2) = (1-p)p^2 + p(1-p)^2 = p(1-p)$$

2. $X \sim \text{bin}(n, p)$ means X is the sum of n *independent* Bernoulli(p) random variables $X_1, X_2, \ldots, X_n$. For independent variables, the variances add. Since $\text{Var}(X_j) = p(1-p)$ we have

$$\text{Var}(X) = \text{Var}(X_1) + \text{Var}(X_2) + \ldots + \text{Var}(X_n) = np(p-1).$$

continued on next slide

Solution continued

3. Since the variables are independent, we have

$$\mathrm{Var}(X_1 + \ldots + X_n) = 4n.$$

$\overline{X}$ is the sum scaled by $1/n$ and the rule for scaling is $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$, so

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n^2}\mathrm{Var}(X_1 + \ldots + X_n) = \frac{4}{n}.$$

This implies $\sigma_{\overline{X}} = \dfrac{2}{\sqrt{n}}$.

Note: this says that the average of n independent measurements varies less than the individual measurements.

Board question continued

6. (New scenario) From the following table compute $F(3.5, 4)$.

$X \backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

answer: See next slide

Solution 6

6. $F(3.5, 4) = P(X \leq 3.5, Y \leq 4)$.

$X \backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

Add the probability in the shaded squares: $F(3.5, 4) = 12/36 = 1/3$.

Board question: computing covariance

Flip a fair coin 3 times.

Let $X =$ number of heads in the first 2 flips

Let $Y =$ number of heads on the last 2 flips.

Compute $\text{Cov}(X, Y)$,

Solution

$X\backslash Y$	0	1	2	$p(x_i)$
0	1/8	1/8	0	1/4
1	1/8	2/8	1/8	1/2
2	0	1/8	1/8	1/4
$p(y_j)$	1/4	1/2	1/4	1

From the marginals compute $E(X) = 1 = E(Y)$. By the table compute

$$E(XY) = 1 \cdot \frac{2}{8} + 2\frac{1}{8} + 2\frac{1}{8} + 4\frac{1}{8} = \frac{5}{4}.$$

So $\mathrm{Cov}(X, Y) = \frac{5}{4} - 1 = \boxed{\frac{1}{4}}$.

A more conceptual solution is on the next slide.

Alternative Solution

Use the properties of covariance.
X_i = the number of heads on the i^{th} flip. (So $X_i \sim$ Bernoulli(.5).)

$$X = X_1 + X_2 \quad \text{and} \quad Y = X_2 + X_3.$$

Know $E(X_i) = 1/2$ and $\text{Var}(X_i) = 1/4$. Therefore $\mu_X = 1 = \mu_Y$.
Use Property 2 (linearity) of covariance

$$\begin{aligned}\text{Cov}(X, Y) &= \text{Cov}(X_1 + X_2, X_2 + X_3) \\ &= \text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3) + \text{Cov}(X_2, X_2) + \text{Cov}(X_2, X_3).\end{aligned}$$

Since the different tosses are independent we know

$$\text{Cov}(X_1, X_2) = \text{Cov}(X_1, X_3) = \text{Cov}(X_2, X_3) = 0.$$

Looking at the expression for $\text{Cov}(X, Y)$ there is only one non-zero term

$$\text{Cov}(X, Y) = \text{Cov}(X_2, X_2) = \text{Var}(X_2) = \boxed{\frac{1}{4}}.$$

Concept question

Toss a fair coin $2n+1$ times. Let X be the number of heads on the first $n+1$ tosses and Y the number on the last $n+1$ tosses.

If $n = 1000$ then $\text{Cov}(X, Y)$ is:

(a) 0 (b) $1/4$ (c) $1/2$ (d) 1

(e) More than 1 (f) tiny but not 0

Board question

Toss a fair coin $2n+1$ times. Let X be the number of heads on the first $n+1$ tosses and Y the number on the last $n+1$ tosses.

Compute $\mathrm{Cov}(X, Y)$ and $\mathrm{Cor}(X, Y)$.

Board question

Toss a fair coin $2n+1$ times. Let X be the number of heads on the first $n+1$ tosses and Y the number on the last $n+1$ tosses.

Compute $\text{Cov}(X, Y)$ and $\text{Cor}(X, Y)$.

As usual let X_i = the number of heads on the i^{th} flip, i.e. 0 or 1. Then

$$X = \sum_{1}^{n+1} X_i, \qquad Y = \sum_{n+1}^{2n+1} X_i$$

X is the sum of $n+1$ independent Bernoulli(1/2) random variables, so

$$\mu_X = E(X) = \frac{n+1}{2}, \quad \text{and} \quad \text{Var}(X) = \frac{n+1}{4}.$$

Likewise, $\mu_Y = E(Y) = \dfrac{n+1}{2}$, and $\text{Var}(Y) = \dfrac{n+1}{4}$.

Continued on next slide.

Solution continued

Now,

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}\left(\sum_{1}^{n+1} X_i \sum_{n+1}^{2n+1} X_j\right) = \sum_{i=1}^{n+1}\sum_{j=n+1}^{2n+1} \mathrm{Cov}(X_i X_j).$$

Because the X_i are independent the only non-zero term in the above sum is $\mathrm{Cov}(X_{n+1}X_{n+1}) = \mathrm{Var}(X_{n+1}) = \frac{1}{4}$ Therefore,

$$\mathrm{Cov}(X,Y) = \frac{1}{4}.$$

We get the correlation by dividing by the standard deviations.

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{1/4}{(n+1)/4} = \frac{1}{n+1}.$$

This makes sense: as n increases the correlation should decrease since the contribution of the one flip they have in common becomes less important.