Exercices 1

Björn Andersson

University of Oslo

October 24, 2021

L3 Task 1

Consider a random variable X which takes values 1, 2, 3 and 4 with corresponding probabilities 0.1, 0.2, 0.4 and 0.3.

a What is $E(X)$?

b What is $E(X^2)$?

c What is $\mathrm{Var}(X)$?

Hint: Recall that $\mathrm{Var}(X) = E[(X - E(X))^2]$.

L3 Task 2

X and Y are two random variables such that $E(X) = 10$, $E(X^2) = 150$, $E(Y) = 5$, $E(Y^2) = 75$ and $E(XY) = 20$.

a What is $\text{Cor}(X, Y)$?

b What is $\text{Cov}(5X, 10Y)$?

c What is $\text{Var}(5X + 10Y)$?

L3 Task 3

The following frequency table was observed from a two-item test where each item was scored $0/1$.

	Item $1 = 0$	Item $1 = 1$
Item $2 = 0$	42	20
Item $2 = 1$	22	16

a. Estimate the difficulty of each item.
b. Estimate the variance of the total score.

L3 Task 4

The following covariance matrix was observed from a three-item test.

$$\Sigma_{X_1,X_2,X_3} = \begin{bmatrix} 1.19 & 0.28 & 0.22 \\ 0.28 & 1.26 & 0.40 \\ 0.22 & 0.40 & 1.47 \end{bmatrix}.$$

a Calculate the sample variance of $X_1 + X_2 + X_3$.

b Calculate the estimated correlation between X_1 and X_3.

L3 Task 5

Show that the sample mean is an unbiased estimator of the expected value of a R.V. X. I.e. derive the expected value of

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

L3 Task 6

Assume that observations X_i are independent. Derive the variance of

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

L3 Task 7

Show that

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [\mathrm{Bias}(\hat{\theta})]^2.$$

L3 Task 8

Derive the expected value of

$$s^*_{xy} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{x}\bar{y}}{n-1},$$

where observations $k, l, k \neq l$ are independent.

L3 Task 1: solution

$$
\begin{aligned}
E(X) &= 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.4 + 4 \times 0.3 = 2.9 \\
E(X^2) &= 1^2 \times 0.1 + 2^2 \times 0.2 + 3^2 \times 0.4 + 4^2 \times 0.3 = 9.3 \\
\mathrm{Var}(X) &= E(X^2) - [E(X)]^2 = 9.3 - 2.9^2 = 0.89
\end{aligned}
$$

To see that $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$,

$$
\begin{aligned}
E[(X - E(X))^2] &= E[X^2 + E(X) \times E(X) - 2 \times X \times E(X)] \\
&= E(X^2) + E(X) \times E(X) - 2E(X) \times E(X) \\
&= E(X^2) - E(X) \times E(X) \\
&= E(X^2) - [E(X)]^2.
\end{aligned}
$$

L3 Task 2: solution

We can first note that

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= E(XY) - E(X)\times E(Y) = 20 - 50 = -30 \\
\mathrm{Var}(X) &= E(X^2) - (E(X))^2 = 150 - 100 = 50 \\
\mathrm{Var}(Y) &= E(Y^2) - (E(Y))^2 = 75 - 25 = 50
\end{aligned}
$$

We thus obtain

$$
\begin{aligned}
\mathrm{Cor}(X,Y) &= \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{-30}{\sqrt{50}\sqrt{50}} = -30/50 = -0.6 \\
\mathrm{Cov}(5X,10Y) &= 5\times 10\times \mathrm{Cov}(X,Y) = -1500 \\
\mathrm{Var}(5X+10Y) &= 5^2\mathrm{Var}(X) + 10^2\mathrm{Var}(Y) + 2\times 5\times 10\times \mathrm{Cov}(X,Y) \\
&= 25\times 50 + 100\times 50 + 100\times(-30) = 3250
\end{aligned}
$$

L3 Task 3: solution

The difficulty is, per the textbook, defined as the mean score of each item. Hence we obtain $\hat{p}_1 = (20 + 16)/100 = 0.36$ and $\hat{p}_2 = (22 + 16)/100 = 0.38$.

From the formula given, we have that the variance of the sum score is equal to the sum of the variances and covariances for all items/item pairs. We have

$$s_1^2 = \hat{p}_1 \times (1 - \hat{p}_1) = 0.36 \times 0.64 = 0.2304$$

$$s_2^2 = \hat{p}_2 \times (1 - \hat{p}_2) = 0.38 \times 0.62 = 0.2356$$

and

$$s_{12} = \hat{p}_{12} - \hat{p}_1 \times \hat{p}_2 = 0.16 - 0.36 \times 0.38 = 0.0232.$$

Hence we obtain

$$s_{1+2} = 0.2304 + 0.2356 + 2 \times 0.0232 = 0.5124$$

L3 Task 4: solution

$$\mathrm{Var}(X_1 + X_2 + X_3) = \sum_{i=1}^{3} \sum_{j=1}^{3} s_{ij} = 5.72.$$

$$\hat{\mathrm{Cor}}(X_1, X_3) = \frac{\hat{\mathrm{Cov}}(X_1, X_3)}{\hat{\sigma}_{X_1} \hat{\sigma}_{X_3}} = \frac{0.22}{\sqrt{1.19}\sqrt{1.47}} \approx 0.166.$$

L3 Task 5: solution

$$
\begin{aligned}
E(\bar{x}) &= E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) \\
&= \frac{\sum_{i=1}^{n} E(X_i)}{n} \\
&= \frac{nE(X)}{n} \\
&= E(X).
\end{aligned}
$$

We thus have that

$$E(\bar{x} - E(X)) = E(X) - E(X) = 0,$$

and $\bar{x}$ is an unbiased estimator of $E(X)$.

L3 Task 6: solution

$$
\begin{aligned}
\mathrm{Var}(\bar{x}) &= \mathrm{Var}\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) \\
&= \frac{\sum_{i=1}^{n} \mathrm{Var}(X_i)}{n^2} \\
&= \frac{n\mathrm{Var}(X)}{n^2} \\
&= \frac{\mathrm{Var}(X)}{n}.
\end{aligned}
$$

L3 Task 7: solution

$$
\begin{aligned}
\mathrm{MSE}(\hat{\theta}) &= E((\hat{\theta}-\theta)^2) \\
&= E(\hat{\theta}^2+\theta^2-2\hat{\theta}\theta) \\
&= E(\hat{\theta}^2)+\theta^2-2\theta E(\hat{\theta}) \\
&= E(\hat{\theta}^2)+(\theta-E(\hat{\theta}))^2-(E(\hat{\theta}))^2 \\
&= E(\hat{\theta}^2)-(E(\hat{\theta}))^2+(\theta-E(\hat{\theta}))^2 \\
&= \mathrm{Var}(\hat{\theta})+(\mathrm{Bias}(\hat{\theta}))^2.
\end{aligned}
$$

L3 Task 8: solution I

We have

$$E\left(\sum_{i=1}^{n} X_i X_i\right) = \sum_{i=1}^{n} E(X_i X_i) = nE(XY)$$

L3 Task 8: solution II

and

$$
\begin{aligned}
E\left[\sum_{i=1}^{n} \bar{x}\bar{y}\right] &= nE\left(\sum_{i=1}^{n} X_i/n \sum_{j=1}^{n} X_j/n\right) \\
&= (1/n)E\left(\sum_{i=1}^{n} X_i \sum_{j=1}^{n} X_j\right) \\
&= (1/n)\sum_{i=1}^{n} E\left(X_i \sum_{j=1}^{n} X_j\right) \\
&= (1/n)(n-1)nE(X)E(Y) + (1/n)nE(XY) \\
&= (n-1)E(X)E(Y) + E(XY).
\end{aligned}
$$

L3 Task 8: solution III

We thus obtain

$$E\left[\sum_{i=1}^{n} X_i X_i - n\bar{x}\bar{y}\right] = nE(XY) - (n-1)E(X)E(Y) - E(XY)$$
$$= (n-1)E(XY) - (n-1)E(X)E(Y).$$

Hence, we have that

$$E(s_{xy}) = \frac{n-1}{n}[E(XY) - E(X)E(Y)]$$
$$= \frac{n-1}{n}\sigma_{xy}.$$

Consequently, an unbiased estimator of σ_{xy} is

$$s^*_{xy} = \frac{\sum_{i=1}^{n}(X_i - \bar{x})(X_i - \bar{y})}{n-1}.$$