Exercise Class- Probability Review

Instructor: Irene Iodice
Email: irene.iodice@univ-paris1.fr

Ex.1: Introduction to the concept of a RV
Recall that a RV is a a variable whose possible values are numerical outcomes of a random phenomenon. Let's introduce a random phenomenon...

Suppose you flip a fair coin three times independently.

1. *What is the sample space?*

   The sample space, denoted as Ω, is the set of all possible outcomes in the experiment. A coin has two sides, heads denoted H and tails denoted T. Each coin flip will turn up either H or T, and 3 independent coin flips yield a sequence of length 3 containing H and/or T. There are a total of 2^3 possible outcomes, thus $|\Omega| = 8$[1]. The sample space is given by:

$$\Omega = \begin{pmatrix} \{H,H,H\} & \{H,H,T\} & \{H,T,H\} & \{T,H,H\} \\ \{H,T,T\} & \{T,H,T\} & \{T,T,H\} & \{T,T,T\} \end{pmatrix} \tag{1}$$

2. *What is the set identified by the event that the number of heads is exactly 2? What is its probability?*

   We group all the sequences in Ω which contain 2 H into a set that we call S:

$$S = (\{H,H,T\} \quad \{H,T,H\} \quad \{T,H,H\}) \tag{2}$$

   Note that there are 3 such cases, i.e. $|S| = 3$. Then, the probability that we get 2 heads on 3 independent coin flips is

$$\frac{|S|}{|\Omega|} = \frac{3}{8} \tag{3}$$

   To solve this problem we have employed a mechanism called 'enumeration', i.e. we have identified all possible outcomes of the random experiment and counted those of interest (which usually are denoted with 'success'). This mechanism become very time-expensive when the number of possible outcomes is large (for example if the experiment required to flip a coin 5 times you would have to enumerate 2^5 possible outcomes).

   There is another way one can tackle this problem. Note, that if we get 2 heads, this implies we also get 1 tails. What is the probability of this case? Recall from basic probability that with independent events, y and z, their joint probability is $Pr(Y = y, Z = z) = Pr(Y = y)Pr(Z = z)$. We know that the probability of a head on one flip is $pr = 1/2$, and of a tail $1 - pr = 1/2$, then the probability of observing 2 heads and one tail is:

$$pr^2(1 - pr) = \frac{1}{8} \tag{4}$$

   . Still, note that the order of each sequence of flips matters, i.e. our sample space contains both $\{H,H,T\}$ and $\{H,T,H\}$. Recall that the number of different ways you can order two heads in set of three elements is given by the binomial coefficient $\binom{3}{2} = 3$. Then the probability we can get two heads and 1 tails in an ordered sequence is given by:

$$\binom{3}{2} pr^2(1 - pr) = \frac{3}{8} = 0.375 \tag{5}$$

[1]This notation is used to denote the cardinality of a set, i.e. the number of different elements it contains.

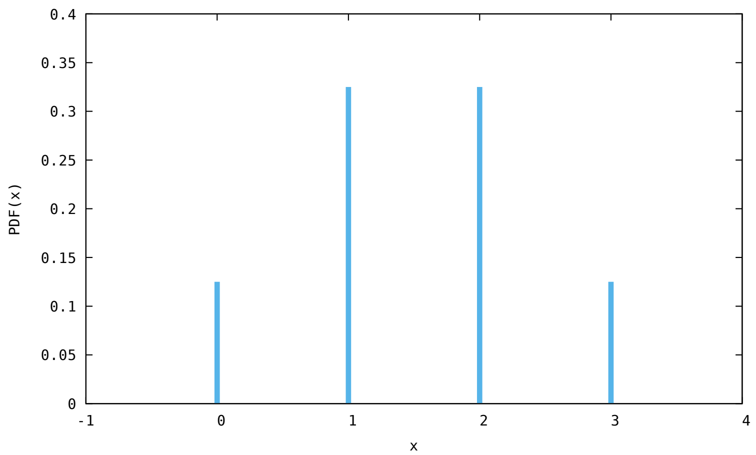
.

3. *Identify with X a RV that represents the number of heads in the sequence. Graph the PDF and CDF.*

   Recall that the probability distribution function of a discrete RV, is the list of all possible values of the variable with the probability that each value will occur.

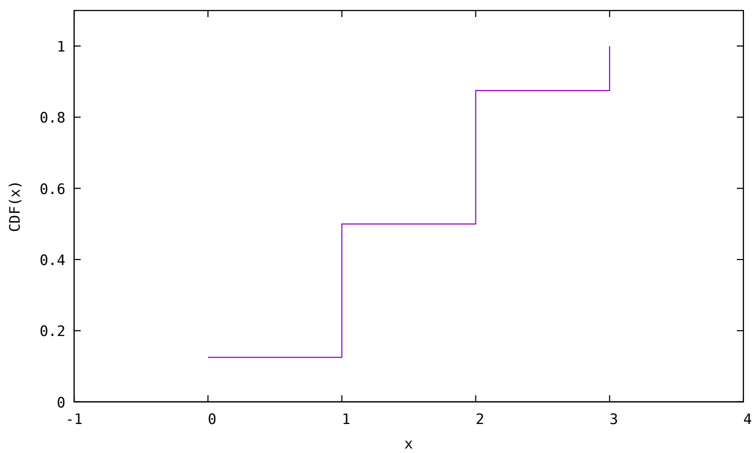
   Since we flip a coin three times, X can be takes values 0,1,2 or 3. We already computed $Pr(X = 2)$ above. For the remaining values we have

   - $Pr(X = 0) = \binom{3}{0} \times (1 - pr)^3 = 1 \times \frac{1}{8} = 0.125$
   - $Pr(X = 1) = \binom{3}{1} \times (pr)(1 - pr)^2 = 3 \times \frac{1}{8} = 0.375$
   - $Pr(X = 3) = \binom{3}{3} \times (pr)^3 = 1 \times \frac{1}{8} = 0.125$



   The cumulative distribution function (CDF) represents the probability that the random variable is less than or equal to a particular value. Thus, we have

   - $Pr(X \leq 0) = Pr(X = 0) = 0.125$
   - $Pr(X \leq 1) = Pr(X = 0) + Pr(X = 1) = 0.500$
   - $Pr(X \leq 2) = Pr(X = 0) + Pr(X = 1) + Pr(X = 2) = 0.875$
   - $Pr(X \leq 3) = Pr(X = 0) + Pr(X = 1) + Pr(X = 2) + Pr(X = 3) = 1$



4. *Compute $E(X)$, $Var(X)$.*
Recall that for a discrete RV we have:

$$E(X) = \mu_x = \sum_i x_i p_i \quad \text{and} \quad Var(X) = \sum_i (x_i - \mu_x)^2 p_i \tag{6}$$

where i denotes the possible values that the RV can take (in our case $i = 0, 1, 2, 3$) and $p_i = Pr(X = x_i)$. We obtain:

- $E(X) = 0 \times Pr(X = 0) + 1 \times Pr(X = 1) + 2 \times Pr(X = 2) + 3 \times Pr(X = 3) = 1.5$
- $Var(X) = (0 - 1.5)^2 \times Pr(X = 0) + (1 - 1.5)^2 \times Pr(X = 1) + (2 - 1.5)^2 \times Pr(X = 2) + (3 - 1.5)^2 \times Pr(X = 3) = 0.75$

5. *Do you expect the average number of heads ($\overline{X}$)in a sequence of three flips to be closer to $E(X)$ after you repeat the experiment 20 or 100 times? Why?*

We expect the average number of heads $\overline{X}$ to be closer to $E[X]$ after we repeat the experiment 100 times. Recall indeed that for the Law of Large numbers as the sample size grows ($n \to \infty$), the sequence of sample means, $\overline{X}_1, ..., \overline{X}_n$, converges in probability to the expected value, thus the larger n the more likely will be to observe a sample mean in an interval around the true mean of the distribution. However, note that this is the case only when we have finite variance ($Var(X) < \infty$), which is true in our exercise.

A practical example to give you an intuition behind the LLN: You are on Erasumus in a new town and you are looking for a room in a shared appartament but you have no clue on how much is a fair rent there, for example what students pay on average. If we randomly choose appartments and record the price, keeping a running average, then at the beginning we might see some larger fluctuations in our average. However, as we continue to call landlords and collect new information on prices (our sample increases in size, $n \uparrow$), we expect to see this running average settle and converge to the true mean price of a room in town.

Ex.2: Moments of a discrete RV

After some years of experience, the instructor of the course in Statistics ask to meet the chief of the Master to present him the following PDF of X, the number of students who miss his evening class on Fridays:

x	0	1	2	3	4	5	6	7	8	≥ 9
$f(x)$	0.35	0.15	0.1	0.00	0.00	0.05	0.05	0.1	0.2	0

1. *Compute the mean, the median, the mode and the standard deviation.*

- The mean is as before $E(X) = \mu_x = \sum_i x_i p_i = 0 \times 0.35 + 1 \times 0.15 + 2 \times 0.1 + 5 \times 0.05 + 6 \times 0.05 + 7 \times 0.1 + 8 \times 0.2 = 3.2$
- The median is the 'central observation', the one which has half of observation that display larger values and half smaller, formally $Pr(X \leq m_X) = 0.5 \quad \text{and} \quad Pr(X \geq m_X) = 0.5$. In our our exercise we are a bit unlucky: the central observation lies between 1 and 2, i.e. $m_x = 1.5$.
- The mode is the most frequent observation, which is 0.
- The standard deviation is the square root of the variance. $Var(X) = \sum_i (x_i - \mu_x)^2 p_i = (0 - 3.2)^2 \times 0.35 + (1 - 3.2)^2 \times 0.15 + (2 - 3.2)^2 \times 0.1 + (5 - 3.2)^2 \times 0.05 + (6 - 3.2)^2 \times 0.05 + (7 - 3.2)^2 \times 0.1 + (8 - 3.55)^2 \times 0.2 = 11.06$ and thus $\sigma_x = \sqrt{Var(X)} = 3.32$

2. *Is the distribution symmetric?*
   A symmetric distribution is one in which mean and median coincides. In our example the mean is greater than the median (the mean lies to the right wrt to the median). This is common for a distribution that is skewed to the right (that is, bunched up toward the left).

3. *What is your intuition on the kurtosis of X?*
   Note that the extreme observations (0 and 8) are the most frequent ones, thus we expect the kurtosis of $f(x)$ to be larger than 3 ('fat tails').

The instructor then shows the PDF of Y, the number of students who miss his evening class on Wensdays:

x	0	1	2	3	4	5	6	≥ 7
$f(x)$	0.05	0.10	0.20	0.25	0.20	0.15	0.05	0

4. *Compare the distribution with the previous one.*

   - The maximum value X takes up is smaller, there are at most 6 students missing the same class.
   - $E(Y) = 0 \times 0.05 + 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.25 + 4 \times 0.2 + 5 \times 0.15 + 6 \times 0.05 = 3.1$, which is slightly smaller than $E(X)$.
   - The median is 3, close to the average: the distribution is approximately symmetric. The mode is also 3.
   - $Var(Y) = (0 - 3.1)^2 \times 0.05 + (1 - 3.1)^2 \times 0.1 + (2 - 3.1)^2 \times 0.2 + (3 - 3.1)^2 \times 0.25 + (4 - 3.1)^2 \times 0.2 + (5 - 3.1)^2 \times 0.15 + (6 - 3.1)^2 \times 0.05 = 2.29$ and thus $\sigma_y = \sqrt{Var(Y)} = 1.51$. The standard deviation is much smaller than that of X.

5. *Based on this comparison, what might be the argument of the instructor?*

   The instructor might ask the chief to reschedule his class on Fridays evening. There are two reasons for this. First, on average the number of students that are absent is larger on Fridays. Second, the variability is much bigger.

Ex.3: Linear functions of a discrete RV
Let X be a discrete random variable with values $x = 0, 1, 2$ and probabilities $Pr(X = 0) = 0.25$, $Pr(X = 1) = 0.5$, and $Pr(X = 2) = 0.25$, respectively.

1. Find $E(X)$.
   E(X)=$\mu_x = \sum_i x_i p_i = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1$

2. Find $Var(X)$.
   $Var(X) = \sum_i (x_i - \mu_x)^2 p_i = (0-1)^2 \times 0.25 + (1-1)^2 \times 0.5 + (2-1)^2 \times 0.25 = 0.5$

3. Find the expected value and variance of Y = 5X + 2.
   - $E(Y) = E(5X+2) = 5E(X) + 2 = 7$. To see this note that the expected value of a constant is a constant, and that $\sum_i 5x_i p_i = 5\sum_i x_i p_i$; namely the expected value is a linear operator.
   - $Var(Y) = Var(5X+2) = 5^2 Var(X) = 25 \times 0.5 = 12.5$. To see this note that the variance of a constant is a zero, and that $Var(X) = \sum_i (5x_i - 5\mu_x)^2 p_i = 5^2 \sum_i (x_i - \mu_x)^2 p_i$.

Ex.4: Describe two RVs, given their joint distribution

Two discrete RVs have the following joint probability distribution:

		Y		
		2	4	6
	1	1/8	1/4	1/8
X	3	1/24	1/4	1/24
	9	1/12	0	1/12

1. *Find the marginal probability distribution of X*
   Recall the formula $Pr(X = x) = \sum_i Pr(X = x, Y = y_i)$. Variable X takes values 1,3,9 while Y values 2,4, 6.
   - $Pr(X = 1) = Pr(X = 1, Y = 2) + Pr(X = 1, Y = 4) + Pr(X = 1, Y = 6) = \frac{1}{8} + \frac{1}{4} + \frac{1}{8} = \frac{1}{2}$
   - $Pr(X = 3) = \frac{1}{24} + \frac{1}{4} + \frac{1}{24} = \frac{1}{3}$
   - $Pr(X = 9) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$

2. *Find the conditional probability distribution of X given Y=2*

   Recall the formula $Pr(X = x|Y = y) = \frac{Pr(X=x,Y=y)}{Pr(Y=y)}$.
   We first compute the term at the denominator $Pr(Y = 2)$, which is the marginal probability that Y=2, as we did above: $Pr(Y = 2) = Pr(X = 1, Y = 2) + Pr(X = 3, Y = 2) + Pr(X = 9, Y = 2) = \frac{1}{8} + \frac{1}{24} + \frac{1}{12} = \frac{1}{4}$. We then have:
   - $Pr(X = 1|Y = 2) = \frac{Pr(X=1,Y=2)}{Pr(Y=2)} = \frac{1}{8} \times 4 = \frac{1}{2}$
   - $Pr(X = 3|Y = 2) = \frac{Pr(X=3,Y=2)}{Pr(Y=2)} = \frac{1}{24} \times 4 = \frac{1}{6}$
   - $Pr(X = 9|Y = 2) = \frac{Pr(X=9,Y=2)}{Pr(Y=2)} = \frac{1}{12} \times 4 = \frac{1}{3}$

   Check the answers to questions 2) and 3) with in mind the fact that all probability distributions must fulfill that $\sum_i p_i = 1$.

3. *Find the covariance of X and Y.*
Recall the formula $Cov(X,Y) = \sigma_{XY} = E\big[(X-\mu_x)(Y-\mu_y)\big] = \sum_i \sum_j (x_i-\mu_x)(y_j-\mu_y)Pr(X = x_i, Y = y_i)$.
We have:

   - $\mu_x = \frac{1}{2} \times 1 + \frac{1}{3} \times 3 + \frac{1}{6} \times 9 = 3$
   - $\mu_y = \frac{1}{4} \times 2 + \frac{1}{2} \times 4 + \frac{1}{4} \times 6 = 4$.
   - $Cov(X,Y) = (1-3)(2-4)\frac{1}{8} + (1-3)(6-4)\frac{1}{8} + (9-3)(2-4)\frac{1}{12} + (9-3)(6-4)\frac{1}{12} = 0$

4. *Are X and Y independent?*

   Recall that two variables are independent if knowing the value of one of the two provides no information about the other, which is $Pr(X|Y=y) = Pr(X)$ and vice versa. It follows that the joint distribtion function can be rewritten as the product of the marginal distribution functions, $Pr(X=x, Y=y) = Pr(X=x)Pr(Y=y)$.
In our case X and Y are dependent as, for example, $Pr(X=2)Pr(X=9) = \frac{1}{4} \times \frac{1}{6} \neq Pr(X = 9, Y=2) = \frac{1}{12}$. This shows that having zero covariance is a necessary but **NOT** sufficient condition for two variables to be independent.

Ex.5: Identify the distribution of two RV

Suppose you toss two tetrahedra (regular four-sided polyhedron) independently. Let X denote the number on the first tetrahedron and Y the larger between the numbers on the two tetrahedra.

1. *Detect the joint distributions of X and Y.*

   Both tetrahedra can take value 1,2,3,4 and thus we have $4^2 = 16$ possible outcomes for our toss. Also X and Y can take values 1,2,3,4, but not all $4^2 = 16$ combinations are possible, since we must have $Y \geq X$. The possible outcomes are just 10, they are $(X,Y) = (1,1),(1,2),(1,3),(1,4),(2,2),(2,3)(2,4),(3,3),(3,4),(4,4)$, however they are not all equally likely. Let's see why...
How many configurations of the toss of the two tetrahedra give as values (X=1,Y=1)? There is just one, the first thetraedron is 1 and the second is 1, thus $Pr(X=1,Y=1) = \frac{1}{4} \times \frac{1}{4}$ How many configurations of the toss of the two tetrahedra give as values (X=2,Y=2)? They are two: given the first is two (X=2) the second may be 1 or 2, in both cases $Y = 2$. Calling Z the number on the second thetrahedon we have: $Pr(X=2,Y=2) = Pr(X=2) \times \big(Pr(Z = 1) + Pr(Z=2)\big) = \frac{1}{4} \times (\frac{1}{4} + \frac{1}{4}) = \frac{2}{16}$.
Following this reasoning we have the following joint distribution:

(x,y)	(1,1)	(1,2)	(1,3)	(1,4)	(2,2)	(2,3)	(2,4)	(3,3)	(3,4)	(4,4)
$f(X=x,Y=y)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{4}{16}$

2. *Detect the marginal distributions of Y.*
The marginal distribution of Y is:

   - $Pr(Y=1) = Pr(X=1,Y=1) = \frac{1}{16}$
   - $Pr(Y=2) = Pr(X=1,Y=2) + Pr(X=2,Y=2) = \frac{3}{16}$
   - $Pr(Y=3) = Pr(X=1,Y=3) + Pr(X=2,Y=3) + Pr(X=3,Y=3) = \frac{5}{16}$
   - $Pr(Y=4) = Pr(X=1,Y=4) + Pr(X=2,Y=4) + Pr(X=3,Y=4) + Pr(X = 4, Y=4) = \frac{7}{16}$

3. *What do you expect about the sign of covariance? Why?*

 The covariance captures the extent to which the changes in one variable are associated with changes in a second variable. Note that $Y = min(X, Z)$, thus Y is a non-negative function of X. We therefore expect a positive co-variance between X and Y. Note also that the covariance measures the degree to which two variables are linearly associated, thus it will not be perfect in our case.

Ex.6: Sums of random variables

The firm 'Pippo' holds an investment portfolio consisting of two stocks A and B, with 80% of her capital invested in A and the remaining 20% in B. Stock A has an expected return of r_A = 10% and a standard deviation of σ_A=15%. Stock B has an expected return of $r_B = 15$ with a standard deviation of B = 25%.

1. *Compute the expected return on the portfolio.*

 Recall that $E(aX + bY) = aE(x) + bE(Y)$. The expected return of the overall portfolio is then: $E(r_p) = E\Big(\frac{4}{5} \times r_A + \frac{1}{5} \times r_B\Big) = \frac{4}{5} \times 10 + \frac{1}{5} \times 15 = 11$

2. *Compute the standard deviation of the returns on the portfolio assuming that the two stocks' returns are perfectly positively correlated, which is $Corr(A, B) = 1$.*

 Recall that $Corr(X, Y) = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$ and that $Var(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$.[2] If $Corr(r_a, r_b) = 1$, then

 - $Cov(r_a, r_b) = \sigma_X \times \sigma_Y = 15 \times 25 = 375$
 - $Var(r_p) = \left(\frac{4}{5}\right)^2 15^2 + \left(\frac{1}{5}\right)^2 25^2 + 2 \times \frac{4}{5} \times \frac{1}{5} \times 375 = 289$
 - $\sigma_{r_p} = \sqrt{Var(r_p)} = \sqrt{289} = 17\%$

3. Compute the standard deviation of the returns on the portfolio assuming that the two stocks returns have a correlation of 0.5.
 In this case we have:

 - $Cov(r_a, r_b) = 0.5\big(\sigma_X \times \sigma_Y\big) = 0.5(15 \times 25) = 187.5$
 - $Var(r_p) = \left(\frac{4}{5}\right)^2 15^2 + \left(\frac{1}{5}\right)^2 25^2 + 2 \times \frac{4}{5} \times \frac{1}{5} \times 187.5 = 229$
 - $\sigma_{r_p} = \sqrt{Var(r_p)} = \sqrt{229} = 15.13\%$

Ex. 7: Applying the Normal distribution

The length of life (in years) of a 'StatPhone' is approximately a normal distribution $N(3.1, 2.25)$.

1. *What is the share of phones that will die within the first year?*

 Recall that the area under the standard normal distribution Z can be interpreted as either a probability or as the proportion of the population with values within a certain range. In particular, the tables provided at the end of the exercises gives us the share of observations with a value less than a certain value z, $Pr(Z \leq z)$. Recall also that:

 - Z is obtained as the standardization of a normally distributed RV X, such that $Z = \frac{X - \mu_x}{\sigma_x}$. In our case X =length of life of a StatPhone, $\mu_x = 3.1$ and $\sigma_x = \sqrt{2.25} = 1.5$.
 - $\int_{-\infty}^{\infty} p(z)dz = 1$, thus, $Pr(Z \geq z) = 1 - Pr(Z \leq z)$
 - Z is symmetric: $P(Z \leq z) = 1 - P(Z \leq z)$

[2] To see this note that: $Var(aX + bY) = Cov(aX + bY, aX + bY)$

- Differently from the discrete case, $Pr(Z \leq z) = Pr(Z < z)$

We therefore have:

$$Pr(X \leq 1) = Pr\Big(Z \leq \frac{1-3.1}{1.5}\Big) = Pr\big(Z \leq -1.4\big) = 1 - \Pr(Z \leq 1.4) = 1 - 0.9192 = 0.0808$$

Around 8.1% of the phones will die within the first year.

2. *What is the share of phones that will survive 4 years or more?*

$$Pr(X \geq 4) = Pr\Big(Z \geq \frac{4-3.1}{1.5}\Big) = Pr\big(Z \geq 0.6\big) = 1 - \Pr(Z \leq 0.6) = 1 - 0.7257 = 0.2743$$

Around 27% of the phones will survive at least 4 years.

3. *What fraction of phones will last between 1 and 3.5 years?*

$$\begin{aligned} Pr(1 \leq X \leq 3) =& Pr\Big(\frac{1-3.1}{1.5} \leq Z \leq \frac{3.5-3.1}{1.5}\Big) = Pr\Big(-1.40 \leq Z \leq 0.27\Big) \\ =& Pr(z \leq 0.27) - Pr(Z \leq -1.4) \\ =& Pr(z \leq 0.27) - \Big(1 - Pr(Z \leq 1.4)\Big) \\ =& 0.6064 - 0.0808 = 0.5256 \end{aligned}$$

Around 52.5% of the phones will last between 1 and 3.5 years.

4. *If the manufacturer adopts a warranty policy in which only 10% of the phones have to be replaced, what will be the length of the warranty period?*

We use the inverse procedure we have done so far, starting from the share to find the length z.

- $Pr(Z \leq z) = 0.10$ is true for $z = -1.28$. Indeed $Pr(Z \leq -1.28) = 1 - Pr(Z \leq 1.28) = 1 - 0.8997 \approx 0.1$. We then have:
- $\frac{x-3.1}{1.5} = -1.28$ which is true for $x = 1.18$

Ex. 8: Random sampling

Suppose that X_1, X_2 and X_3 is a sample of observations from a $N(\mu_X, \sigma_X^2)$ population, with sample average $\overline{X}$. Suppose further that the three observations are not independent, in particular:

$$Cov(X_1, X_2) = Cov(X_2, X_3) = Cov(X_1, X_3) = 0.5\sigma^2 \tag{7}$$

!Notation!: you can rewrite $Cov(X_1, X_2) = \sigma_{X_1X_2}$

1. Find $E(\overline{X})$.
$E(\overline{X}) = E\Big(\frac{1}{3}\sum_{i=1}^{3} X_i\Big) = \frac{1}{3}\sum_{i=1}^{3} E(X_i) = \frac{1}{3}(3\mu_X) = \mu_X$

2. *Find $Var(\overline{X})$*

Recall that:$Var(X_1 + X_2) = Cov(X_1 + X_2, X_1 + X_2)$

- $Var(\overline{X}) = Var\Big(\frac{1}{3}\sum_{i=1}^{3} X_i\Big) = \frac{1}{9}\Big(Var(X_1 + X_2 + X_3)\Big)$
- $Var(X_1 + X_2 + X_3) = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \sigma_{X_3}^2 + 2\sigma_{X_1X_2} + 2\sigma_{X_1X_3} + 2\sigma_{X_2X_3}$
- $Var(\overline{X}) = \frac{1}{9}\Big(3\sigma_X^2 + 3 \times 2 \times \big(0.5\sigma_X^2\big)\Big) = \frac{2}{3}\sigma_X^2$

Note that, if the sample observations were independent, the sample variance would be smaller, $Var(\overline{X}) = \frac{1}{3}\sigma_X^2)$. Recall the intutition of this exercise once you will be asked to study the the different consequences on the OLS estimators on biasedness (ability to approximate the expected value of a RV) and efficiency (the variability of the estimator) when the errors that are not IID.