Exercises on probability and statistics

Machine Learning 2020-2021 - UMONS
Lab 1
Souhaib Ben Taieb

1 Exercise 1

Consider a sample space Ω comprising three possible outcomes:

$$\Omega = \{\omega_1, \omega_2, \omega_3\}.$$

Suppose the three possible outcomes are assigned the following probabilities:

$$P(\omega_1) = \frac{1}{5} \quad P(\omega_2) = \frac{2}{5} \quad P(\omega_3) = \frac{2}{5}$$

Define the events

$$A_1 = \{\omega_1, \omega_2\}, \quad A_2 = \{\omega_1, \omega_3\},$$

and denote by A_1^c the complement of A_1.

Compute $P(A_2|A_1^c)$, the conditional probability of A_2 given A_1^c.

2 Exercise 2

An economics consulting firm has created a model to predict recessions. The model predicts a recession with probability 80% when a recession is indeed coming and with probability 10% when no recession is coming. The unconditional probability of falling into a recession is 20%. If the model predicts a recession, what is the probability that a recession will indeed come?

3 Exercise 3

1. Let X be a discrete random variable taking values in $R_X = \{0, 1, 2, 3, 4\}$ with uniform distribution. Compute $P(1 \leq X < 4)$.

2. Let X be a continuous random variable taking values in $R_X = [0, 1]$. Let its probability density function $f_X(x)$ be

$$f_X(x) = \begin{cases} 2x & \text{if } x \in R_X; \\ 0 & \text{if } x \notin R_X \end{cases}$$

   Compute $P(\frac{1}{4} \leq X < \frac{1}{2})$.

4 Exercise 4

1. Let X be a continous random variable with uniform distribution on the interval $[1,3]$. Compute the **expected value** and **variance** of X.

2. Let X be a discrete random variable taking values in $R_X = \{1,2,3\}$. Let its probability mass function $p_X(x)$ be

$$p_X(x) = \begin{cases} x/6 & \text{if } x \in R_X; \\ 0 & \text{if } x \notin R_X \end{cases}$$

   Compute the **expected value** and **variance** of X.

5

For the following joint distributions between random variables Y and X, find both marginal distributions and the conditional distribution requested. Also, are the two random variables independent?

5.1

Find the marginal distributions and the distribution of Y conditional on $X = 0$.

	$X = 0$	$X = 1$
$Y = 0$	0.14	0.26
$Y = 1$	0.21	0.39

5.2

Find the marginal distributions and the distribution of X conditional on $Y = 1$.

	$X = 0$	$X = 1$
$Y = 1$	0.45	0.25
$Y = 3$	0.05	0.25

5.3

Find the marginal distributions and the distribution of Y conditional on $X = 1$.

	$X = 0$	$X = 1$	$X = 2$
$Y = 1$	0.1	0.2	0.3
$Y = 2$	0.05	0.15	0.2

5.4

Find the marginal distributions and the distribution of Y conditional on $X = 2$.

	$X = 0$	$X = 1$	$X = 2$
$Y = 1$	0.05	0.04	0.01
$Y = 2$	0.1	0.08	0.02
$Y = 3$	0.35	0.28	0.07

6

We have a population of people, 47% of whom were men and the remaining 53% were women. Suppose that the average height of the men was 70 inches, and the women was 71 inches. What is the average height of the entire population? [Hint: Use the law of total expectation]

7

Let $X_1, X_2, \dots, X_n$ be a collection of n random variables, and $a_1, a_2, \dots, a_n$, a set of constants, we have

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \,\mathrm{Cov}(X_i, X_j).$$

Prove the above fact. You can use the fact that, for a set of numbers $e_1, e_2, \dots, e_n$,

$$\left(\sum_{i=1}^{n} e_i\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} e_i e_j.$$

8

Let $X_1, X_2, \ldots, X_n$ be **i.i.d.** random variables with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$ for $i = 1, 2, \ldots, n$. If

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

show that

1. $E[\bar{X}_n] = \mu$
2. $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$

9 Exercise 6

Let F_1 be a normal distribution $\mathcal{N}(\mu, \sigma^2)$, and F_2, a Bernoulli distribution, $\mathrm{Bern}(p)$, where $\mu \in \mathbb{R}$, $\sigma > 0$ and $p \in [0,1]$. Consider the four scenarios where $F = F_1$ or F_2 and $n = 10$ or $n = 1000$. For each scenario,

1. repeat the following procedure $10,000$ times:
   (a) Generate n i.i.d. realizations $X_1, X_2, \ldots, X_n$ where $X_i \sim F$.
   (b) Compute $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
2. compute the average and variance of the $10,000$ values computed in 1(b)
3. plot a histogram of these $10,000$ values, and add vertical lines at the true mean and the average.

Discuss the results using Exercise 5. Experiment with different values of μ and σ for F_1, and different values of p for F_2.