# Review of probability and statistics

Machine Learning I (2021-2022)
UMONS

## 1

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?

### 1.1 Solution

Let $R$ be a Bernouilli random variable with support $\mathscr{X} \in \{0,1\}$ indicating whether we fell into a recession ($R=1$ means we fell into a recession, $R=0$ means we did not). Let $M$ also be a Bernouilli random variable with support $\mathscr{M} \in \{0,1\}$ indicating the outcome of the prediction model ($M=1$ means that the model predicted that a recession was coming, $M=0$ means that it did not).

We know that $\mathbb{P}(R=1)=0.2$, $\mathbb{P}(M=1|R=1)=0.8$ and $\mathbb{P}(M=1|R=0)=0.1$. We are interested in finding the probability that a recession will come, conditioned on the fact that the model predicted it, i.e. we are looking for $\mathbb{P}(R=1|M=1)$. By the definition of conditional probability, we have:

$$
\begin{aligned}
\mathbb{P}(R=1|M=1) &= \frac{\mathbb{P}(R=1,M=1)}{\mathbb{P}(M=1)} \\
&= \frac{\mathbb{P}(M=1|R=1)\mathbb{P}(R=1)}{\mathbb{P}(M=1)} \\
&= \frac{\mathbb{P}(M=1|R=1)\mathbb{P}(R=1)}{\mathbb{P}(M=1|R=1)\mathbb{P}(R=1)+\mathbb{P}(M=1|R=0)\mathbb{P}(R=0)} \qquad \text{Law of total probability} \\
&= \frac{\mathbb{P}(M=1|R=1)\mathbb{P}(R=1)}{\mathbb{P}(M=1|R=1)\mathbb{P}(R=1)+(1-\mathbb{P}(M=0|R=0))(1-\mathbb{P}(R=1))} \\
&= \frac{0.8*0.2}{0.8*0.2+0.1*0.8} \\
&= \frac{2}{3}
\end{aligned}
$$

# 2

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?

## 2.1

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 |

### 2.1.1 Solution

$p_X(0) = p_{XY}(0,0) + p_{XY}(0,1) = 0.21 + 0.14 = 0.35$
$p_X(1) = p_{XY}(1,0) + p_{XY}(1,1) = 0.26 + 0.39 = 0.65$
$p_Y(0) = p_{XY}(0,0) + p_{XY}(1,0) = 0.14 + 0.26 = 0.4$
$p_Y(1) = p_{XY}(0,1) + p_{XY}(1,1) = 0.21 + 0.39 = 0.6$

$p_{Y|X}(1|0) = \frac{p_{XY}(0,1)}{p_X(0)} = \frac{0.21}{0.35} = 0.6$
$p_{Y|X}(0|0) = \frac{p_{XY}(0,0)}{p_X(0)} = \frac{0.14}{0.35} = 0.4$

**Independence ?**

Two discrete random variables $X$ and $Y$ are independent iff :

$$p_{XY}(x,y) = p_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)p_X(x) = p_X(x)p_Y(y)$$

We must check that the equality holds for all realizations of the random variables $X$ and $Y$ :

$p_{XY}(0,0) = 0.14 = p_X(0)p_Y(0)$
$p_{XY}(0,1) = 0.21 = p_X(0)p_Y(1)$
$p_{XY}(1,0) = 0.26 = p_X(1)p_Y(0)$
$p_{XY}(1,1) = 0.39 = p_X(1)p_Y(1)$

The random variables $X$ and $Y$ are independent.

## 2.2

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 |

### 2.2.1 Solution

$p_X(0) = 0.5$
$p_X(1) = 0.5$
$p_Y(1) = 0.7$
$p_Y(3) = 0.3$

$p_{X|Y}(0|1) = 0.64$

$p_{X|Y}(1|1) = 0.36$

**Independence ?**

$p_{XY}(0,1) = 0.45 \neq 0.35 = p_X(0)p_Y(1) \rightarrow$ Not independent.

### 2.3

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 |

#### 2.3.1 Solution

$p_X(0) = 0.15$
$p_X(1) = 0.35$
$p_X(2) = 0.5$
$p_Y(1) = 0.6$
$p_Y(2) = 0.4$

$p_{Y|X}(1|1) = 0.57$
$p_{Y|X}(2|1) = 0.43$

**Independence ?**

$p_{XY}(0,1) = 0.1 \neq 0.09 = p_X(0)p_Y(1) \rightarrow$ Not independent.

### 2.4

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 |

#### 2.4.1 Solution

$p_X(0) = 0.5$
$p_X(1) = 0.4$
$p_X(2) = 0.1$
$p_Y(1) = 0.1$
$p_Y(2) = 0.2$
$p_Y(3) = 0.7$

$p_{Y|X}(1|2) = 0.1$
$p_{Y|X}(2|2) = 0.2$
$p_{Y|X}(3|2) = 0.7$

**Independence ?**

$p_{XY}(0,1) = 0.05 = p_X(0)p_Y(1)$
$p_{XY}(0,2) = 0.1 = p_X(0)p_Y(2)$
$p_{XY}(0,3) = 0.35 = p_X(0)p_Y(3)$

$p_{XY}(1,1) = 0.04 = p_X(1)p_Y(1)$
$p_{XY}(1,2) = 0.08 = p_X(1)p_Y(2)$
$p_{XY}(1,3) = 0.28 = p_X(1)p_Y(3)$
$p_{XY}(2,1) = 0.01 = p_X(2)p_Y(1)$
$p_{XY}(2,2) = 0.02 = p_X(2)p_Y(2)$
$p_{XY}(2,3) = 0.07 = p_X(2)p_Y(3)$

The random variables $X$ and $Y$ are independent.

## 3

Alex and Bob each flips a fair coin twice. Denote "1" as head, and "0" as tail. Let $X$ be the maximum of the two numbers Alex gets, and let $Y$ be the minimum of the two numbers Bob gets.

a) Find the joint pmf $p_{X,Y}(x,y)$.

b) Find the marginal pmf $p_X(x)$ and $p_Y(y)$.

c) Find the conditional pmf $p_{X|Y}(x|y)$. Does $p_{X|Y}(x|y) = p_X(x)$ ? Why ?

### 3.1 Solution

For both Alex and Bob, the sample space for flipping a fair coin twice is $\Omega = \{00, 01, 10, 11\}$. if $X$ and $Y$ are the random variables respectively denoting the maximum of the two numbers Alex gets and the minimum of the two numbers Bob gets, then $\mathscr{X} \in \{0,1\}$ and $\mathscr{Y} = \{0,1\}$.

a) By definition of the joint pmf and as the random variables $X$ and $Y$ are independent, we have :

$p_{XY}(0,0) = p_X(0)p_Y(0) = \frac{1}{4} * \frac{3}{4} = \frac{3}{16}$

$p_{XY}(0,1) = p_X(0)p_Y(1) = \frac{1}{4} * \frac{1}{4} = \frac{1}{16}$

$p_{XY}(1,0) = p_X(1)p_Y(0) = \frac{3}{4} * \frac{3}{4} = \frac{9}{16}$

$p_{XY}(1,1) = p_X(0)p_Y(0) = \frac{3}{4} * \frac{1}{4} = \frac{3}{16}$

We can check that $\sum_{x\in\mathscr{X}, y\in\mathscr{Y}} p_{XY}(x,y) = \frac{3}{16} + \frac{1}{16} + \frac{9}{16} + \frac{3}{16} = 1$.

b) From the law of total probability, $p_X(x) = \sum_{y\in\mathscr{Y}} p_{XY}(x,y)$ and $p_Y(y) = \sum_{x\in\mathscr{X}} p_{XY}(x,y)$. Thus :

$p_X(0) = p_{XY}(0,1) + p_{XY}(0,0) = \frac{3}{16} + \frac{1}{16} = \frac{1}{4}$
$p_X(1) = 1 - p_X(0) = 1 - \frac{1}{4} = \frac{3}{4}$
$p_Y(0) = p_{XY}(0,0) + p_{XY}(1,0) = \frac{3}{16} + \frac{9}{16} = \frac{3}{4}$
$p_Y(1) = 1 - p_Y(0) = 1 - \frac{3}{4} = \frac{1}{4}$

c) From the definition of conditional density, we have that $p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}$. Thus :

$P_{X|Y}(0|0) = \frac{p_{XY}(0,0)}{p_Y(0)} = \frac{3}{16} * \frac{4}{3} = \frac{1}{4}$
$P_{X|Y}(0|1) = \frac{p_{XY}(0,1)}{p_Y(1)} = 4 * \frac{1}{16} = \frac{1}{4}$
$P_{X|Y}(1|0) = \frac{p_{XY}(1,0)}{p_Y(0)} = \frac{9}{16} * \frac{4}{3} = \frac{3}{4}$
$P_{X|Y}(1|1) = \frac{p_{XY}(1,1)}{p_Y(1)} = 4 * \frac{3}{16} = \frac{3}{4}$

As the random variables $X$ and $Y$ are independent, we have $p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)} = \frac{p_X(x)P_y(y)}{p_Y(y)} = p_X(x)$.

# 4

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]

## 4.1 Solution

Let $M$ be a Bernouilli random variable with support $\mathscr{M} \in \{0,1\}$ indicating whether an individual is either male or female ($M = 1$ means that the individual is male, $M = 0$ means that the individual is female). Let $H$ be a continuous random variable with support $\mathscr{H} \in \mathbb{R}^+$ indicating the height of an individual of the population.

We know that $p_M(1) = 0.47$ and that $p_M(0) = 0.53$. Moreover, $\mathbb{E}[H|M = 1] = 70$ and $\mathbb{E}[H|M = 0] = 71$. We are interested in finding the average height of the entire population, i.e $\mathbb{E}[H]$.

From the law of total expectation, we have:

$$\begin{aligned}\mathbb{E}[H] &= \mathbb{E}[H|M = 1]p_M(1) + \mathbb{E}[H|M = 0]p_M(0) \\ &= 70 * 0.47 + 71 * 0.53 \\ &= 70.53 \text{ inches}\end{aligned}$$

## 5

Let $X_1, X_2, \ldots, X_n$ be a collection of $n$ random variables, and $a_1, a_2, \ldots, a_n$, 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 $e_1, e_2, \ldots, e_n$,

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

### 5.1 Solution

By expanding the expression of the variance, and by successively applying the properties of the expectation, we get :

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) &= \mathbb{E}\Big[\Big(\sum_{i=1}^{n} a_i X_i - \mathbb{E}\Big[\Big(\sum_{i=1}^{n} a_i X_i\Big)\Big]\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\sum_{i=1}^{n} a_i X_i\Big)^2 - 2\mathbb{E}\Big[\sum_{i=1}^{n} a_i X_i\Big]\Big(\sum_{i=1}^{n} a_i X_i\Big) + \Big(\mathbb{E}\Big[\sum_{i=1}^{n} a_i X_i\Big]\Big)^2\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j X_i X_j - 2\Big(\sum_{i=1}^{n} a_i \mathbb{E}\Big[X_i\Big]\Big)\Big(\sum_{i=1}^{n} a_i X_i\Big) + \Big(\mathbb{E}\Big[\sum_{i=1}^{n} a_i X_i\Big]\Big)^2\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j X_i X_j\Big] - 2\Big(\sum_{i=1}^{n} a_i \mathbb{E}\Big[X_i\Big]\Big)\Big(\sum_{i=1}^{n} a_i \mathbb{E}[X_i]\Big) + \Big(\mathbb{E}\Big[\sum_{i=1}^{n} a_i X_i\Big]\Big)^2 \\
&= \mathbb{E}\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j X_i X_j\Big] - 2\Big(\sum_{i=1}^{n} a_i \mathbb{E}[X_i]\Big)^2 + \Big(\sum_{i=1}^{n} a_i \mathbb{E}[X_i]\Big)^2 \\
&= \mathbb{E}\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j X_i X_j\Big] - \Big(\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathbb{E}[X_i]\mathbb{E}[X_j]\Big) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \Big(\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]\Big) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \ \mathrm{Cov}(X_i, X_j)
\end{aligned}$$

## 6

Let $p_X$ be a normal distribution $\mathcal{N}(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$, and $\sigma > 0$. Consider the two scenarios where $n = 10$ or $n = 1000$. For each scenario,

1. repeat the following procedure 1000 times:
   (a) Generate $n$ i.i.d. realizations $X_1, X_2, \dots, X_n$ where $X_i \sim p_X$.
   (b) Compute $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.
2. compute the mean and variance of the 1000 values computed in 1(b)
3. plot a histogram of these 1000 values, and add vertical lines at the true mean and the computed mean.

Experiment with different values of $\mu$ and $\sigma$, and confirm that you obtain $E[\bar{X}_n] = \mu$ and $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$.

## 7

You observe a sample of real values $y_1, y_2, \ldots, y_n$ where $y_i > 1$ for $i = 1, 2, \ldots, n$. Let us assume they are all i.i.d. observations of a random variable $Y$ with the following probability density function:

$$p(y;\alpha) = \begin{cases} \alpha e^{-\alpha y}, & \text{if } y \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

1. Write down the formula for the log-likelihood as a function of the observed data and the unknown parameter $\alpha$.

2. Compute the maximum likelihood estimate (MLE) of $\alpha$.

### 7.1 Solution

1)

$$\begin{aligned} L(\alpha) &= L(\alpha; y_1, y_2, ..., y_n) \\ &= p(y_1, y_2, ..., y_n; \alpha) \\ &= p(y_1;\alpha)p(y_2;\alpha)...p(y_n;\alpha) \qquad y_i,\ i = 1, ..., n \text{ are i.i.d. random variables.} \\ &= \prod_{i=1}^{n} p(y_i;\alpha) \\ &= \prod_{i=1}^{n} \alpha e^{-\alpha y_i} \end{aligned}$$

2)

$$\begin{aligned} \text{MLE} = \hat{\alpha} &= \underset{\alpha \in A}{\text{argmax}}\ L(\alpha) \\ &= \underset{\alpha \in A}{\text{argmax}}\ \ln L(\alpha) \\ &= \underset{\alpha \in A}{\text{argmax}}\ \ln\Big(\prod_{i=1}^{n} \alpha e^{-\alpha y_i}\Big) \\ &= \underset{\alpha \in A}{\text{argmax}}\ \sum_{i=1}^{n}\Big(ln\ \alpha - \alpha y_i\Big) \\ &= \underset{\alpha \in A}{\text{argmax}}\ n(ln\ \alpha - \alpha \bar{y}) \end{aligned}$$

with $A$ being the parameter space. Taking the derivative with respect to $\alpha$ and equalling to zero :

$$\Big(ln\ L\Big)'(\alpha) = n(\frac{1}{\alpha} - \bar{y})$$

$$\begin{aligned} &\Big(ln\ L\Big)'(\alpha) = 0 \\ \iff\ & n\Big(\frac{1}{\alpha} - \bar{y}\Big) = 0 \\ \iff\ & \alpha = \frac{1}{\bar{y}} \end{aligned}$$

Thus we have that the MLE $\hat{\alpha} = \frac{1}{\bar{y}}$. To check that $\hat{\alpha}$ is indeed a maximum, we can verify that the second derivative of the log-likelihood computed in $\hat{\alpha}$ is always negative :

$$\Big(ln\ L\Big)''(\hat{\alpha}) = -\frac{n}{\hat{\alpha}^2} < 0 \qquad \forall \hat{\alpha} \in A$$

# 8 Complementary exercise

Find the marginal pdf $f_X(x)$ if the joint pdf $f_{XY}(x,y)$ is defined as :

$$f_{XY}(x,y) = \frac{e^{-|y-x|-x^2/2}}{2\sqrt{2\pi}}$$

## 8.1 Solution

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy \\
&= \int_{-\infty}^{\infty} \frac{e^{-|y-x|-x^2/2}}{2\sqrt{2\pi}} dy \\
&= \frac{1}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-|y-x|} e^{-x^2/2} dy \\
&= \frac{e^{-x^2/2}}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-|y-x|} dy
\end{aligned}$$

We have that : $|y-x| = \begin{cases} y-x, & \text{if } y \geq x \\ x-y, & \text{if } y \leq x \end{cases}$, and thus :

$$\begin{aligned}
f_X(x) &= \frac{e^{-x^2/2}}{2\sqrt{2\pi}} \int_{-\infty}^{x} e^{y-x} dy + \int_{x}^{\infty} e^{x-y} dy \\
&= \frac{e^{-x^2/2}}{2\sqrt{2\pi}} \left( e^{-x} \left[ e^{y} \right]_{-\infty}^{x} + e^{x} \left[ -e^{-y} \right]_{x}^{\infty} \right) \\
&= \frac{e^{-x^2/2}}{2\sqrt{2\pi}} \left( e^{0} - e^{-\infty} - e^{-\infty} + e^{0} \right) \\
&= \frac{e^{-x^2/2}}{\sqrt{2\pi}}
\end{aligned}$$