DATA 1010
In-class exercises
Samuel S. Watson
12 October 2018

Problem 1

Suppose you're 90% sure that your package was delivered today and 75% sure that if it was delivered it would be on your door step rather than tucked away in your mailbox. When you arrive at home and do not see your package right away, what is the conditional probability—given the observed information—that you'll find it in your mailbox?

Solution

Let's use the sample space

$$\Omega = \{\text{delivered and visible}, \text{delivered and invisible}, \text{undelivered and visible}, \text{undelivered and invisible}\}$$

The probability masses assigned to these four outcomes are

$$\mathbb{P}(\text{visible} \,|\, \text{delivered})\mathbb{P}(\text{delivered}) = (0.75)(0.9) = 0.675,$$

$$\mathbb{P}(\text{invisible} \,|\, \text{delivered})\mathbb{P}(\text{delivered}) = (0.25)(0.9) = 0.225,$$

$$\mathbb{P}(\text{undelivered and visible}) = 0,$$

$$\mathbb{P}(\text{invisible} \,|\, \text{undelivered})\mathbb{P}(\text{undelivered}) = (1)(0.1) = 0.1.$$

The desired conditional probability is $\mathbb{P}(\text{delivered} \,|\, \text{invisible})$, which is equal to

$$\mathbb{P}(\text{delivered} \,|\, \text{invisible}) = \frac{\mathbb{P}(\text{delivered and invisible})}{\mathbb{P}(\text{invisible})}$$

The numerator is equal to $(0.25)(0.9) = 0.225$, while the denominator is equal to

$$\mathbb{P}(\text{invisible} \,|\, \text{delivered})\mathbb{P}(\text{delivered}) + \mathbb{P}(\text{invisible} \,|\, \text{undelivered})\mathbb{P}(\text{undelivered}) = (0.25)(0.9) + (1)(0.1) = 0.325.$$

Dividing these two quantities, we get a conditional probability of approximately $\boxed{69.2\%}$ that the package is in the mailbox.

The equation

$$\mathbb{P}(A \,|\, E) = \frac{\mathbb{P}(E \,|\, A)\mathbb{P}(A)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E \,|\, A)\mathbb{P}(A)}{\mathbb{P}(E \,|\, A)\mathbb{P}(A) + \mathbb{P}(E \,|\, A^c)\mathbb{P}(A^c)}$$

relating $\mathbb{P}(A \,|\, E)$ and $\mathbb{P}(E \,|\, A)$ is called **Bayes' theorem**. The factor $\mathbb{P}(A)/\mathbb{P}(E)$ can have some unintuitive consequences: $\mathbb{P}(A \,|\, E)$ can be much larger or smaller than $\mathbb{P}(E \,|\, A)$ if $\mathbb{P}(A)/\mathbb{P}(E)$ is very large or small. For example, if A is the event that a patient has a disease and E is the event that the patient tests positive for the disease, then $\mathbb{P}(E \,|\, A)$ can be small even though $\mathbb{P}(A \,|\, E)$ is high. In other words, the patient probably doesn't have the disease even though the test is very accurate.

Problem 2

Does $\text{Cov}(X, Y) = 0$ imply that X and Y are independent?

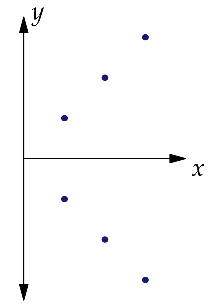
Consider a random variable X which is uniformly distributed on $\{1, 2, 3\}$ and an independent random variable Z which is uniformly distributed on $\{-1, 1\}$. Set $Y = ZX$. Consider the pair (X, Y).

Solution

The suggested random variables X and Y have zero covariance, because

$$\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X^2 Z] - \mathbb{E}[X]\mathbb{E}[ZX] = \mathbb{E}[X^2]\mathbb{E}[Z] - \mathbb{E}[X]^2\mathbb{E}[Z] = \mathbb{E}[X^2]\cdot 0 - \mathbb{E}[X]^2 \cdot 0 = 0,$$

but they are not independent. For example, $\mathbb{P}(X = 2 \text{ and } Y = 1) = 0$ even though $\mathbb{P}(X = 2)$ and $\mathbb{P}(Y = 1)$ are both positive. The joint distribution of X and Y is shown in the figure.



Problem 3

Suppose that $X_1, \ldots, X_n$ are independent random variables with the same distribution. Find the mean and variance of

$$\frac{X_1 + \cdots + X_n}{n}$$

Solution

By linearity of expectation, we have

$$\begin{aligned}
\mathbb{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] &= \frac{\mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]}{n} \\
&= \mathbb{E}[X_1] \qquad \text{(Identical distribution).}
\end{aligned}$$

Then

$$\begin{aligned}
\operatorname{Var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) &= \sum_{k=1}^{n} \operatorname{Var}\left(\frac{X_k}{n}\right) + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \operatorname{Cov}\left(\frac{X_j}{n}, \frac{X_k}{n}\right) \\
&= \sum_{k=1}^{n} \operatorname{Var}\left(\frac{X_k}{n}\right) \qquad \text{(Independence)} \\
&= \sum_{k=1}^{n} \frac{1}{n^2} \operatorname{Var}(X_k) \qquad \text{(Variance property)} \\
&= \frac{\operatorname{Var}(X_1)}{n}. \qquad \text{(Identical distribution)}
\end{aligned}$$

This hints at the law of large numbers: the average of n independent random variables with the same distribution is a new random variable whose mean is the mean of the common distribution and whose variance decreases to 0 as $n \to \infty$. This means that if we sample from a distribution many times, the average of those samples is likely to be close to the distribution's mean.

Problem 4

Consider the probability space with $\Omega = [0, 1]$ and probability measure given by the density $f(x) = 2x$ for $x \in [0, 1]$. Find $\mathbb{P}([\frac{1}{2}, 1])$.

Solution

We calculate $\mathbb{P}([\frac{1}{2}, 1]) = \int_{\frac{1}{2}}^{1} 2x \, \mathrm{d}x = \frac{3}{4}$.

Problem 5

Find the expectation of a random variable whose density is the function $f(x) = 2x$ on $[0, 1]$.

Solution

We have $\mathbb{E}[X] = \int_0^1 x(2x)\,\mathrm{d}x = \frac{2}{3}$.