Solutions to problem set 2, 17 February, 2025

Definition 1 *A* **random variable** *is a function* $X : \Omega \to \mathbb{R}$ *such that for every* $t \in \mathbb{R}$

$$\{\omega \in \Omega : X(\omega) \leq t\} = \{X \leq t\}$$

is an event, that is, it is in $\mathcal{A}$*.*

A random variable is discrete, if its range is finite or countably infinite.

The **expectation** *of a discrete random variable is defined by*

$$\mathbb{E}(X) = \sum_{j=1}^{\infty} x_j \cdot \mathbb{P}(X = x_j), \qquad \textit{provided that } \sum_{j=1}^{\infty} |x_j| \cdot \mathbb{P}(X = x_j) < \infty,$$

where $x_1, x_2, \dots$ *are the possible values of* X*.*

The **variance** *of a random variable is defined by*

$$\mathrm{Var}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

provided that $\mathbb{E}(X^2)$ *is finite (there exist random variables with finite expectation but not well-defined variance).*

The square root of $\mathrm{Var}(X)$ *is called* **standard deviation***.*

A random variable has **binomial distribution** *with parameters* $n \geq 1$ *and* $0 < p < 1$*, if*

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \textit{for all } k = 0, 1, \dots, n.$$

Then the expectation of X *is* np*, and its variance is* $np(1-p)$*.*

A random variable has **Poisson distribution** *with parameter* $\lambda > 0$*, if*

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \textit{for all } k = 0, 1, 2, \dots.$$

Then $\mathbb{E}(X) = \mathrm{Var}(X) = \lambda$*.*

---

1. We organize a party, and we know in advance that the number of participants is eight with probability 1/4, nine with probability 1/3 and ten otherwise. Calculate the expected number and the variance of the number of participants.

   Let X be the number of guests. The expectation is as follows by definition:

   $$\mathbb{E}(X) = \frac{1}{4} \cdot 8 + \frac{1}{3} \cdot 9 + \frac{5}{12} \cdot 10 = 9.17.$$

   $$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{4} \cdot 8^2 + \frac{1}{3} \cdot 9^2 + \frac{5}{12} \cdot 10^2 - 9.17^2 = 84.67 - 9.17^2 = 0.58.$$

   Standard deviation: $\sqrt{0.58} = 0.76$

2. Peter is late from the university with probability 0.1 every day, independently of each other. We know that there are 21 days in March where he goes to the university.
   a) What is the probability that he is never late in March? What is the probability of exactly 1 occasion when he is late? Of 2? In general, what is the probability that he is late in March exactly k times?
   b) What is the distribution of the number of occasions when he is late?
   c) Randomize 100 samples from the distribution of the number of late arrivals of Peter in March. Make a histogram, and calculate the average in R.
   d) What is the expectation of number of occasions when he is late?
   e) What is the variance of the number of occasions when he is late?

   (a) possible values: $0, 1, \dots, 21$
   X: number of occasions when Peter is late

   $$\mathbb{P}(X = 0) = 0.9 \cdot 0.9 \cdot 0.9 \dots \cdot 0.9 = 0.9^{21}$$

   The days are independent of each other, hence we can multiply the probabilities.

$$\mathbb{P}(X=1) = 21 \cdot 0.1 \cdot 0.9^{20}$$

because there are 21 days on which he can be late, 0.1 is the probability of being late, and with probability 0.9^{20} he is not late on the other days. Again, the days are independent, hence we can multiply the probabilities.

Similarly, since we can choose the two "late" days $\binom{21}{2}$ many ways:

$$\mathbb{P}(X=2) = \binom{21}{2} \cdot 0.1^2 \cdot 0.9^{19}$$

In general, for every $k = 0, 1, \ldots, 21$ we have:

$$\mathbb{P}(X=k) = \binom{21}{k} \cdot 0.1^k \cdot 0.9^{21-k}$$

Hence the number of occasions that Peter is late is binomial with $n = 21$ and $p = 0.1$

In general: if we have n independent events, each of them occurs with probability p, and X is the number of events that occured, then X has binomial distribution. That is, for every $k = 0, 1, \ldots, n$ we have

$$\mathbb{P}(X=k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

(b)
```
> sample=rbinom(100, size=21, prob=0.1)
> hist(sample, col="orange", main="number of lates", xlab="possible values",
ylab="probabilities")
```

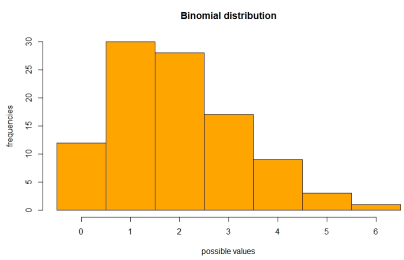

Figure 1: The distribution of the number of occasions when Peter is late: the possible values and the probabilities

```
> 21*20/2*0.1^2*0.9^19
[1] 0.2836789
> dbinom(2, size=21, prob=0.1)
[1] 0.2836789
> mean(sample)
[1] 2.23
> sd(sample)
[1] 1.398809
```

(c) $\mathbb{P}(\text{exactly two lates}) = \binom{21}{2} \cdot 0.1^2 \cdot 0.9^{19} = 0.284.$
```
> dbinom(2, size=21, prob=0.1)
[1] 0.2836789
```

(d) Expectation: $\mathbb{E}(X) = n \cdot p = 21 \cdot 0.1 = 2.1.$
Variance: $\text{Var}(X) = n \cdot p \cdot (1-p) = 21 \cdot 0.1 \cdot 0.9 = 1.89.$

3. Suppose that we have 10 servers in a system. On each day, each of them breaks down with probability 0.01, independently of each other. Let Z be the number of servers (among these 10) which break down tomorrow (that is, on a given day). Calculate the probability $\mathbb{P}(Z=2)$, and the expectation and variance of Z.

Notice that Z has binomial distribution: $n = 10$ independent events (a server breaking down), each occuring with probability $p = 0.01$, and Z is the number of events that occured.

$\mathbb{P}(Z=2) = \binom{10}{2} \cdot 0.99^8 \cdot 0.01^2 = 0.00415.$

```
> dbinom(2, size=10, prob=0.01)
[1] 0.004152351
```

Expectation of Z:

$\mathbb{E}(Z) = 10 \cdot 0.01 = 0.1.$

Variance: $\text{Var}(Z) = 10 \cdot 0.01 \cdot 0.99 = 0.096.$

4. Suppose that, according to our data from the last decades, the average number of earthquakes per year is 3.42 in a given city. Suppose that the number of earthquakes in a given year has Poisson distribution, and that its expectation is equal to the average that we observed.
a) Let us randomize a sample of size 200 from the distribution of the number of earthquakes in a given year, and make a histogram. What is the mean of the data? What is the proportion of 3 in the sample? What is the proportion of numbers which are at least 4?
b) What is the probability that there are exactly 3 earthquakes in a year?
c) What is the probability that there are at least 4 earthquakes in a year?

$\lambda = 3.42$, because for Poisson distribution the expectation and the parameter is the same, and the expectation has to be equal to the average that we observed.

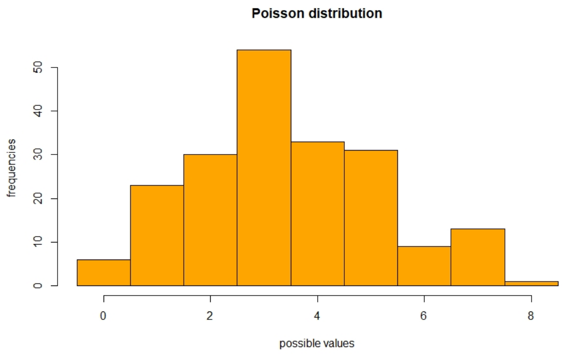

Figure 2: A sample of size 200 from a Poisson distribution with parameter $\lambda = 3.42$

```
> minta=rpois(200, lambda=3.42)
> hist(minta, col="orange", xlab="possible values", ylab="frequencies", main="Poisson distribution
breaks=c(-0.5, 0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5))
> mean(sample)
[1] 3.425
> sd(sample)
[1] 1.752063
#proportion of 3 in the sample:
> length(sample[sample==3])/length(sample)
[1] 0.27
# proportion of values at least 4 in the sample:
> length(sample[sample>=4])/length(sample)
[1] 0.435
```

b)

X: number of earthquakes

$$\mathbb{P}(X=3) = \frac{\lambda^k}{k!}e^{-\lambda} = \frac{3.42^3}{3!}e^{-3.42} = 0.218.$$

$$\mathbb{P}(X=0) = \frac{\lambda^0}{0!}e^{-\lambda} = e^{-\lambda}$$

c)

$$\begin{aligned}\mathbb{P}(X \geq 4) &= \mathbb{P}(X=4) + \mathbb{P}(X=5) + \ldots = \\ &= 1 - \mathbb{P}(X=0) - \mathbb{P}(X=1) - \mathbb{P}(X=2) - \mathbb{P}(X=3) = \\ &= 1 - \left(\frac{3.42^3}{3!} - \frac{3.42^2}{2!} - \frac{3.42^1}{1!} - \frac{3.42^0}{0!}\right)e^{-3.42} = \\ &= 1 - 0.218 - 0.1913 - 0.1119 - 0.0327 = 1 - 0.446 = 0.554.\end{aligned}$$

```
> dpois(3, lambda=3.42)
```

```
[1] 0.2180921
> 1-ppois(3, lambda=3.42, lower.tail=F)
[1] 0.5539899
```

5. Suppose that the number of downloads of a webpage within an hour has Poisson distribution, and the probability that there are 0 downloads is $1/e^2$. Suppose furthermore that the number of downloads independent for disjoint time intervals.

   a) What is the variance of the number of downloads within an hour?

   b) Given that the number of downloads within an hour is at most 1, what is the probability that there are 0 downloads within this hour?

   a) By the definition of the Poisson distribution, we have

$$\mathbb{P}(X = 0) = e^{-\lambda} \Rightarrow \lambda = 2$$

   By the properties of the Poisson distribution, we have $\mathrm{Var}(X) = \lambda = 2$.

   b) The question is the following conditional probability, which we can calculate by using the definition of this and the definition of the Poisson distribution:

$$\mathbb{P}(X = 0|X \leq 1) = \frac{\mathbb{P}(X = 0, X \leq 1)}{\mathbb{P}(X \leq 1)} = \frac{\mathbb{P}(X = 0)}{\mathbb{P}(X = 0) + \mathbb{P}(X = 1)} = \frac{e^{-2}}{e^{-2} + e^{-2} \cdot 2} = \frac{1}{3}.$$

Homework until 2 October2 0:00, 2 points. Suppose that 15 people attend a meeting. Everyone is late independently, with probability p. We also know that the expected number of people being late is 5.

a) What is the probability that exactly 2 people are late?

b) Given that at most two people are late, what is the conditional probability that no one is late?