Solution Sheet 5

1. a) The binomial distribution (with $n = 50$) is suited to the description of these data. We assume that the melons contained in the shipment do not affect each others' quality (e.g. by passing on rot), i.e. that they are „independent“ of each other.

b) Assumption: The number of rotten melons follows a binomial distribution with parameters $n = 50$ and p. X denotes the number of overripe melons.
$H_0 : p = p_0 = 0.04$, $H_A : p > p_0$ (one-sided)
Trial and error gives us the following (for $p_0 = 0.04$):

$$\begin{aligned} \mathrm{P}[X \geq 4] &= 1 - \mathrm{P}[X \leq 3] = 1 - \sum_{k=0}^{3} \binom{n}{k} p_0^k (1-p_0)^{n-k} = 1 - 0.861 = 0.139 > 0.05\,, \\ \mathrm{P}[X \geq 5] &= 1 - \mathrm{P}[X \leq 4] = 1 - \sum_{k=0}^{4} \binom{n}{k} p_0^k (1-p_0)^{n-k} = 1 - 0.951 = 0.049 < 0.05\,. \end{aligned}$$

Thus the rejection set is $K = \{k;\ k \geq 5\}$. As merely 4 rotten melons have been found among the 50 investigated, we accept the null hypothesis; we cannot be sure that the trading company is lying.

c) The 50 melons investigated contain 4 rotten ones. Thus the p-value is

$$\mathrm{P}[X \geq 4] = 1 - \mathrm{P}[X \leq 3] = 1 - \sum_{k=0}^{3} \binom{50}{k} 0.04^k 0.96^{50-k} = 1 - 0.861 = 0.139\,.$$

d) The probability β of a Type II error is

$$\mathrm{P}[\text{accept } H_0 \text{ despite } H_A : p = 0.1 \text{ being true}] = \mathrm{P}[X < 5 \text{ and } p = 0.1] = 0.43\,.$$

This probability is very high, i.e. it is quite difficult to prove when the trader is lying. The power of the test is then

$$\mathrm{P}[H_0 \text{verwerfen, wenn } p = 0.1] = 1 - 0.43 = 0.57\,.$$

To improve the power of the test, the quality control officer would have to increase the sample size, e.g. by selecting 100 melons at random for investigation.

2. **a)** For $f(x)$ to be a density, the area under the triangle must be exactly 1. Thus we require that

$$\frac{c \cdot 20}{2} = 1\,,$$

from which $c = \frac{1}{10}$ follows. The density can thereby be written as

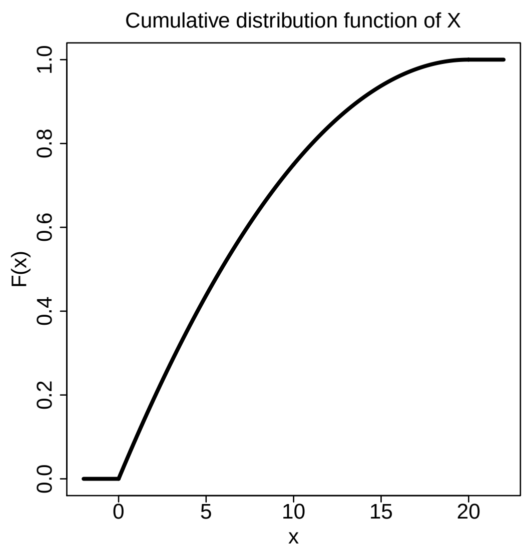
$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{10}(1 - \frac{x}{20}) & 0 \le x \le 20 \\ 0 & x > 20 \end{cases} .$$

b) We find the cumulative distribution function of X by integrating its density. For $0 \le x \le 20$, wer find that:

$$F(x) = \mathrm{P}\left[X \le x\right] = \int_0^x f(t)\,dt = \int_0^x (\frac{1}{10} - \frac{t}{200})\,dt = \frac{x}{10} - \frac{x^2}{400}$$

For $x \le 0$, we have $F(x) = 0$, and for $x \ge 20$, $F(x) = 1$.
In particular: $\mathrm{P}\left[X < 5\right] = F(5) = 0.4375$ and $\mathrm{P}\left[X < 10\right] = F(10) = 0.75$.

c) The cumulative distribution function $F(x)$ has already been computed in part b). It looks as follows:



d)

$$\begin{aligned}
\mathbf{E}\left[X\right] &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^{20} x \left[\frac{1}{10}\left(1 - \frac{x}{20}\right)\right] dx = \frac{1}{10}\left(\frac{x^2}{2} - \frac{x^3}{60}\right)\Bigg|_0^{20} = \frac{20}{3} \\
\mathrm{Var}\left(X\right) &= \mathbf{E}\left[X^2\right] - \left(\mathbf{E}\left[X\right]\right)^2 \\
\mathbf{E}\left[X^2\right] &= \int_0^{20} x^2 \frac{1}{10}\left(1 - \frac{x}{20}\right) dx = \frac{1}{10}\left(\frac{x^3}{3} - \frac{x^4}{80}\right)\Bigg|_0^{20} = \frac{200}{3} \\
\mathrm{Var}\left(X\right) &= \mathbf{E}\left[X^2\right] - \mathbf{E}\left[X\right]^2 = \frac{200}{3} - \left(\frac{20}{3}\right)^2 = \frac{200}{9}
\end{aligned}$$

So the standard deviation is $sd(X) = \sqrt{\mathrm{Var}\left(X\right)} = \sqrt{2} \cdot 10/3 \approx 4.71$.

The median $\tilde{m}$ can be computed by $F(\tilde{m}) \overset{!}{=} 0.5$. It certainly lies inside the interval [0,20], and thus

$$\frac{x}{10} - \frac{x^2}{400} \overset{!}{=} 0.5 \qquad \Longrightarrow \qquad \tilde{m} = 20 - 10\sqrt{2} \approx 5.858.$$

e)

$$\begin{aligned}
\mathrm{P}\left[K \leq 120'000\right] &= \mathrm{P}\left[40'000 \cdot \sqrt{X} \leq 120'000\right] = \mathrm{P}\left[\sqrt{X} \leq 3\right] \\
&= \mathrm{P}\left[X \leq 9\right] = F(9) = \frac{9}{10} - \frac{9^2}{400} = 0.6975
\end{aligned}$$

f) The density $g(x)$ of the exponential distribution is

$$g(x) = \begin{cases} 0 & x \leq 0 \\ \lambda \exp(-\lambda x) & x > 0 \end{cases} .$$

If X is exponentially distributed, its expectation is

$$E[X] = \frac{1}{\lambda}.$$

Choosing $\lambda = \frac{3}{20}$ thus gives us the same mean as for the previous distribution.

g) The cumulative distribution function $G(x)$ is

$$G(x) = P[X \leq x] = 1 - \exp(-\lambda x)$$

for $x > 0$. Thus

$$\begin{aligned}
\mathrm{P}\left[K \leq 120'000\right] &= \mathrm{P}\left[40'000 \cdot \sqrt{X} \leq 120'000\right] = \mathrm{P}\left[\sqrt{X} \leq 3\right] \\
&= \mathrm{P}\left[X \leq 9\right] = G(9) = 1 - \exp(-\frac{3}{20}9) = 1 - 0.259 = 0.741 \ .
\end{aligned}$$

If the duration of work is assumed to be exponentially distributed, the probability of costs below CHF 120'000.– is lower than if the previous distribution is used – despite these distributions having identical means.