

Musterlösung zur Serie 5

1. a) Die Binomialverteilung (mit $n = 50$) ist ein geeignetes Modell, um die Daten zu beschreiben. Wir setzen dabei voraus, dass sich in der Ladung benachbarte Melonen nicht gegenseitig mit Fäulnis „anstecken“, dass die Melonen also „unabhängig“ voneinander sind.
- b) Annahme: Die Anzahl fauler Melonen ist binomialverteilt mit $n = 50$ und Erfolgsparameter p . X beschreibt die Anzahl der faulen Melonen.
 $H_0 : p = p_0 = 0.04$, $H_A : p > p_0$ (einseitig)
 Durch Probieren findet man, dass (für $p_0 = 0.04$) gilt:

$$P[X \geq 4] = 1 - P[X \leq 3] = 1 - \sum_{k=0}^3 \binom{n}{k} p_0^k (1 - p_0)^{n-k} = 1 - 0.861 = 0.139 > 0.05,$$

$$P[X \geq 5] = 1 - P[X \leq 4] = 1 - \sum_{k=0}^4 \binom{n}{k} p_0^k (1 - p_0)^{n-k} = 1 - 0.951 = 0.049 < 0.05.$$

Also ist der Verwerfungsbereich $K = \{k; k \geq 5\}$. Da sich nur 4 faule Melonen unter den 50 untersuchten befinden, behalten wir die Nullhypothese bei; wir können nicht sicher sein, dass der Händler lügt.

- c) Unter den 50 untersuchten gibt es 4 faule Melonen. Somit ist der P-Wert

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

- d) Die Wahrscheinlichkeit β eines Fehlers 2. Art beträgt

$$P[H_0 \text{ annehmen, obwohl } H_A : p = 0.1 \text{ richtig}] = P[X < 5 \text{ und } p = 0.1] = 0.43.$$

Diese Wahrscheinlichkeit ist sehr hoch, d.h. wir können dem Händler nur schwer eine Lüge nachweisen. Die Macht des Test beträgt

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

Um die Situation zu verbessern, müsste der Grossverteiler die Stichprobengrösse erhöhen (z.B. 100 Melonen entnehmen).

2. a) Damit $f(x)$ eine Dichte ist, muss die Fläche des Dreiecks gleich 1 sein. Es muss also gelten

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

Daraus folgt $c = \frac{1}{10}$. Die Dichte lässt sich somit durch die Funktion

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

beschreiben.

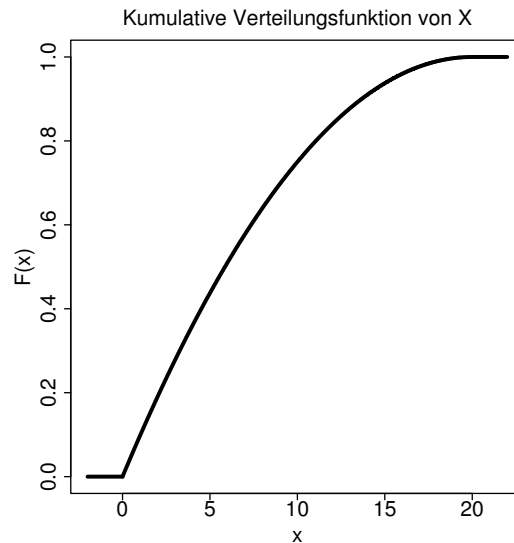
- b) Die kumulative Verteilungsfunktion von X lässt sich durch Integration der Dichtefunktion berechnen: Für $0 \leq x \leq 20$ gilt:

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

Für $x \leq 0$ ist $F(x) = 0$ und für $x \geq 20$ gilt $F(x) = 1$.

Insbesondere gilt: $P[X < 5] = F(5) = 0.4375$ und $P[X < 10] = F(10) = 0.75$.

- c) Die kumulative Verteilungsfunktion wurde bereits in b) berechnet.
Skizze von $F(x)$:



- d)

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{20} x \left[\frac{1}{10} \left(1 - \frac{x}{20}\right) \right] dx = \frac{1}{10} \left(\frac{x^2}{2} - \frac{x^3}{60} \right) \Bigg|_0^{20} = \frac{20}{3}$$

$$\text{Var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$$

$$\mathbf{E}[X^2] = \int_0^{20} x^2 \frac{1}{10} \left(1 - \frac{x}{20}\right) dx = \frac{1}{10} \left(\frac{x^3}{3} - \frac{x^4}{80} \right) \Bigg|_0^{20} = \frac{200}{3}$$

$$\text{Var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{200}{3} - \left(\frac{20}{3}\right)^2 = \frac{200}{9}$$

Also ist die Standardabweichung $sd(X) = \sqrt{\text{Var}(X)} = \sqrt{2} \cdot 10/3 \approx 4.71$.

Für den Median \tilde{m} muss gelten: $F(\tilde{m}) \stackrel{!}{=} 0.5$. Der Median liegt sicher im Intervall $[0, 20]$ und somit haben wir

$$\frac{x}{10} - \frac{x^2}{400} \stackrel{!}{=} 0.5 \quad \implies \quad \tilde{m} = 20 - 10\sqrt{2} \approx 5.858.$$

e)

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

f) Die Exponential-verteilung hat die Dichte $g(x)$:

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

Wenn X exponentialverteilt ist, dann ist der Erwartungswert

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

Für $\lambda = \frac{3}{20}$ erhalten wir somit denselben Erwartungswert wie in der bisherigen Verteilung.

g) Die kumulative Verteilungsfunktion $G(x)$ ist für $x > 0$

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

Daher

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

Wenn die Dauer der Baustellen als exponentialverteilt angenommen wird, ist die Wahrscheinlichkeit also grösser, dass die Kosten einer Baustelle unter 120'000 Fr. liegen, verglichen mit der ursprünglich angenommenen Verteilung, obwohl der Erwartungswert für die Dauer der Baustellen identisch ist für beide Verteilungen.