

Musterlösung zu Serie 5

1. Das Vertrauensintervall von λ hat die gleiche Form wie das Vertrauensintervall von $\tilde{\lambda}$ (um Faktor 5 gestreckt).

- a) Das approximative Vertrauensintervall zum Niveau 0.05 ist (vgl. Skript, Abschnitt 3.3.3):

$$I = I(x) \approx x \pm 1.96\sqrt{x}$$

Das heisst $\tilde{\lambda} \in 7 \pm 1.96\sqrt{7} = [1.81, 12.19]$. Für $\lambda = \tilde{\lambda}/5$ ist das zweiseitige Vertrauensintervall demnach $[0.36, 2.44]$.

- b) Um die genaue Form herauszufinden, müsste man für alle möglichen Werte von $\tilde{\lambda}_0$ einen Hypothesentest $H_0 : \tilde{\lambda} = \tilde{\lambda}_0$ gegen die *einseitige* Alternative $H_A : \tilde{\lambda} < \tilde{\lambda}_0$ für die Beobachtung $x = 7$ auf dem Signifikanzniveau α durchführen. Alle Werte von $\tilde{\lambda}_0$, für die H_0 nicht verworfen wird, liegen im *einseitigen* $1 - \alpha$ -Vertrauensintervall von $\tilde{\lambda}$.

Wir haben schon gesehen, dass der obige Test für $\tilde{\lambda}_0 = 15$ auf dem 5% Signifikanzniveau verworfen wird. Also liegt 15 nicht im 95%-Vertrauensintervall von $\tilde{\lambda}$. Für alle anderen Werte von $\tilde{\lambda}$ müsste man nun eigentlich eine ähnliche Rechnung durchführen. Das ist natürlich (ohne Computer) zu aufwändig und wir werden uns deshalb mit einer qualitativen Überlegung begnügen:

Für unseren Test ist der Annahmehereich immer von der Form $K = \{a + 1, a + 2, \dots\}$; dabei ist a so gewählt, dass $P_{H_0}[S \leq a] = \alpha$. Wenn $\tilde{\lambda}$ nun immer grösser wird, dann wird die ganze Wahrscheinlichkeitsverteilung von S nach rechts gestreckt, und daher wird auch a immer grösser. Irgendwann wird a grösser als x sein und dann wird H_0 verworfen, d.h., das entsprechende $\tilde{\lambda}$ gehört nicht mehr ins Vertrauensintervall. Wir haben also gesehen, dass sehr grosse Werte von $\tilde{\lambda}$ nicht ins Vertrauensintervall kommen.

Wenn umgekehrt $\tilde{\lambda}$ immer kleiner wird, so wird die Wahrscheinlichkeitsverteilung nach links gestaucht (wird aber nie negativ) und auch a wird immer kleiner. Somit ist für hinreichend kleine Werte von $\tilde{\lambda}$ sicher $x > a$. D.h., H_0 wird nicht abgelehnt und $\tilde{\lambda}$ kommt ins Vertrauensintervall. Wir haben also gesehen, dass sehr kleine Werte von $\tilde{\lambda}$ ins Vertrauensintervall kommen.

Das Vertrauensintervall (α beliebig) für $\tilde{\lambda}$ muss also von der Form $[0, c]$ sein. Zudem wissen wir, dass obiger Test für $\alpha = 0.05$ und $\lambda = 15$ verworfen wird. D.h., in dem 95% Vertrauensintervall ist $c < 15$.

Mit R kann man den kritischen Wert mit der Funktion `qpois()` bestimmen:

```
> ## Für lambda0tilde=15 wird H0 verworfen falls weniger als ...
> qpois(0.05,lambda=15)
[1] 9
> ## ... Fasern beobachtet werden. In unserem Falls ist die Beobachtung
> ## x=7 (siehe Serie 3, Aufgabe 1) und somit wird H0:lambda0tilde=15
> ## verworfen (--> 15 liegt nicht im 95% VI für lambdatilde)
>
> ## Weitere Versuche ermöglichen den Wert c fürs Vertrauensintervall
> ## anzunähern:
> qpois(0.05,lambda=14)
[1] 8
> qpois(0.05,lambda=13)
[1] 7
> ## Also: 14 liegt noch nicht im 95% VI für lambdatilde, 13 aber schon.
>
> ## Den exakten Wert für c kann man z.B. wie folgt berechnen:
> round(seq(0,23,by=0.001)[which(qpois(0.05,seq(0,23,by=0.001))==8)[1]],2)
[1] 13.15
```

Der Computer liefert also das 95% Vertrauensintervall $[0, 13.15]$ für $\tilde{\lambda}$, bzw. $[0, 2.63]$ für λ .

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.

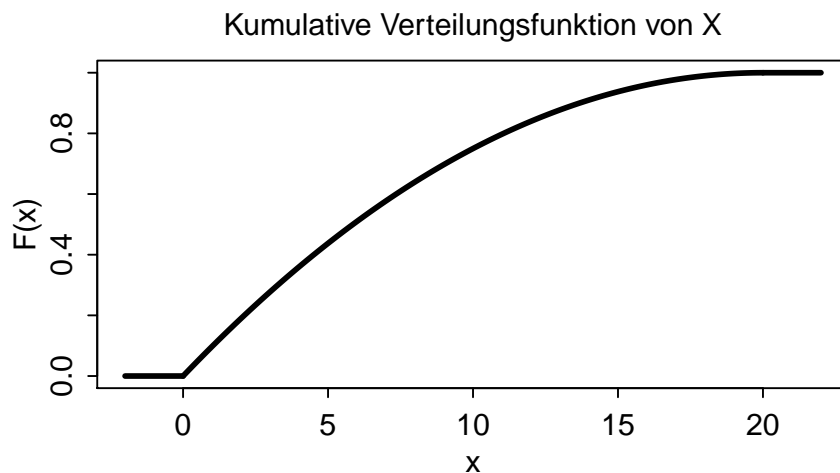
- b) Die kumulative Verteilungsfunktion von X lässt sich durch Integration der Dichtefunktion berechnen: Für $0 \leq x \leq 20$ gilt:

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

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

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

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

- 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

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

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.