ETH Zürich FS 2013
D-MATH
Hans Rudolf Künsch

Koordinator
Blanka Horvath

Wahrscheinlichkeit & Statistik

Musterlösung Serie 11

1. **a)** Das interessierende Ereignis „vier Bauern" kommt in den $n = 60$ (idealerweise) unabhängigen Austeilungen jeweils mit gleicher Wahrscheinlichkeit $p = \frac{8}{935}$ vor. Die Anzahl erfolgreicher Austeilungen ist daher binomialverteilt mit Parametern n und p bzw., da p klein ist, genähert poissonverteilt mit Parameter $\frac{1}{2} = \lambda \approx np$.

„Exakt" gilt $P[X = 0] = \binom{60}{0}(\frac{8}{935})^0(\frac{927}{935})^{60} = 59.72\%$, und unter Verwendung der Poissonapproximation gilt $P[X = 0] \approx e^{-\frac{1}{2}} \frac{(\frac{1}{2})^0}{0!} = e^{-1/2} = 60.65\%$.

b) $P[X = 3] \approx e^{-\frac{1}{2}} \frac{(\frac{1}{2})^3}{3!} = 1.26\%$ und
$P[X \geq 3] = 1 - P[X \leq 2] \approx 1 - e^{-\frac{1}{2}} \frac{(\frac{1}{2})^0}{0!} - e^{-\frac{1}{2}} \frac{(\frac{1}{2})^1}{1!} - e^{-\frac{1}{2}} \frac{(\frac{1}{2})^2}{2!} = 1.44\%$.

Nur etwa jeden siebzigsten Jassabend ($100/1.44 \approx 70$), bei dem alles mit rechten Dingen zugeht, werden drei- oder mehrmal vier Bauern gewiesen. Das ist so selten, dass es wohl plausibler ist anzunehmen, jemand habe an diesem Abend mit gezinkten Karten gespielt.

2. Sei X die Anzahl markierter Fische im zweiten Fang. Beobachtet wurde $X = 3$. Die Wahrscheinlichkeit des Ereignisses, das wir beobachtet haben, hängt in folgender Weise von N ab: Es ist

$$P_N[X = 3] = \frac{\binom{N-5}{8}\binom{5}{3}}{\binom{N}{11}} = \frac{9900\,(N-5)!\,(N-11)!}{(N-13)!\,N!} =: g(N).$$

Der Maximum-Likelihood-Schätzer $\widehat{N}_{\text{ML}}$ für N ist so zu wählen, dass $g(N)$ für $N = \widehat{N}_{\text{ML}}$ maximal wird. Wir müssen also die Maximumstelle der Funktion g bestimmen. Wegen

$$\frac{g(N+1)}{g(N)} = \frac{(N-4)(N-10)}{(N-12)(N+1)} \begin{cases} > 1 & \text{falls } N \leq 17 \\ < 1 & \text{falls } N \geq 18 \end{cases}$$

ist $g(18)$ maximal. $\widehat{N}_{\text{ML}} = 18$ ist also der Maximum-Likelihood-Schätzer für N.

3. **a)**

- Da die X_i unabhängig und alle identisch $\mathcal{N}(\mu, \sigma^2)$-verteilt sind, gilt für das arithmetische Mittel:

$$\mathbb{P}\left(\frac{\sum_{i=1}^{2n+1} X_i - (2n+1)\mu}{\sigma\sqrt{2n+1}} \leq x \right) = \Phi(x),$$

wobei die Wahrscheinlichkeit auf der linken Seite umgeschrieben werden kann wie folgt

$$\begin{aligned} \mathbb{P}\left(\frac{\sum_{i=1}^{2n+1} X_i - (2n+1)\mu}{\sigma\sqrt{2n+1}} \leq x \right) &= \mathbb{P}\left(\frac{\frac{1}{2n+1}\sum_{i=1}^{2n+1} X_i - \mu}{\frac{\sigma}{\sqrt{2n+1}}} \leq x \right) \\ = \mathbb{P}\left(\tfrac{1}{2n+1} \sum_{i=1}^{2n+1} X_i - \mu \leq x \tfrac{\sigma}{\sqrt{2n+1}} \right) &= \mathbb{P}\left(T_{2n+1}^{(1)} - \mu \leq x \tfrac{\sigma}{\sqrt{2n+1}} \right). \end{aligned}$$

Da wir an $\mathbb{P}\left(\left| T_{2n+1}^{(1)} - \mu \right| \leq c_n^{(1)} \right)$ interessiert sind, erhalten wir

$$\begin{aligned} \mathbb{P}\left(\left| T_{2n+1}^{(1)} - \mu \right| \leq c_n^{(1)} \right) &= \mathbb{P}\left(\left| \frac{\sum_{i=1}^{2n+1} X_i - (2n+1)\mu}{\sigma\sqrt{2n+1}} \right| \leq x \right) \\ &= \Phi(x) - \Phi(-x), \end{aligned}$$

daher bekommt man unter Benutzung der Symmetrie von Φ, dass $c_n^{(1)} = x\frac{\sigma}{\sqrt{2n+1}}$, wobei man den Wert von x aus $95\% = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$ erhält:

$$x = \Phi^{-1}(97.5\%) = 1.96.$$

- Andererseits gilt für das Stichprobenmedian $T_{2n+1}^{(2)} = X_{(n+1)}$, mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$, dass der Median von $\widetilde{X}_i := X_i - \mu$ null ist, i.e. $F^{-1}\left(\frac{1}{2}\right) = 0$, wobei F die Verteilungsfunktion von $\widetilde{X}_i$ bezeichnet. Dann ist Beispiel 4.4 aus dem Skript für $\widetilde{X}_i$ anwendbar, und liefert

$$\mathbb{P}\left(\sqrt{(2n+1)}\widetilde{X}_{(n+1)} \leq x \right) \quad \overset{n\to\infty}{\longrightarrow} \quad \Phi(2F'(0)x),$$

wobei $F'(0) = \frac{1}{\sqrt{2\pi}\sigma}$. Die Wahrscheinlichkeit auf der linken Seite kann folgendermassen umgeschrieben werden

$$\mathbb{P}\left(\sqrt{2n+1}\widetilde{X}_{(n+1)} \leq x \right) = \mathbb{P}\left(\widetilde{X}_{(n+1)} \leq \tfrac{x}{\sqrt{2n+1}} \right).$$

Da wir an $\mathbb{P}\left(\left| T_{2n+1}^{(2)} - \mu \right| \leq c_n^{(2)} \right)$ interessiert sind, muss x von der Form $95\% = \Phi\left(\frac{\sqrt{2}}{\sqrt{\pi}\sigma} x \right) - \Phi\left(-\frac{\sqrt{2}}{\sqrt{\pi}\sigma} x \right)$ sein und somit analog wie oben

$$c_n^{(2)} = \frac{x}{\sqrt{(2n+1)}} = \frac{1.96\ \sigma\ \sqrt{\pi}}{\sqrt{2(2n+1)}}.$$

Siehe nächstes Blatt!

Man sagt, dass das Stichprobenmedian asymptotisch normalverteilt ist mit asymptotischer Varianz $\sigma^2 \frac{\pi}{2}$.

b) Wähle $q = \frac{\pi}{2}$. Dann ist

$$\frac{c_{qn}^{(2)}}{c_n^{(1)}} = \frac{\frac{1.645\ \sigma\ \sqrt{\pi}}{\sqrt{\left(2\frac{\pi}{2}n+1\right)\sqrt{2}}}}{\frac{1.645\ \sigma}{\sqrt{2n+1}}} = \frac{\sqrt{2\frac{\pi}{2}n + \frac{\pi}{2}}}{\sqrt{2\frac{\pi}{2}n + 1}} \xrightarrow{n\to\infty} 1.$$

Der Stichprobenmedian ist also bei der Normalverteilung weniger genau als das arithmetische Mittel. Man braucht für den Stichprobenmedian $q = \frac{\pi}{2}$-mal so viele Beobachtungen als für das arithmetische Mittel. Man sagt, dass er $\sqrt{\frac{\pi}{2}}$-mal so viel streut und nennt q die *relative effizienz.*

4. Die Aufgabe kann mit folgendem Code realisiert werden:

```
releff_mc<-function(N,n,k)
{
X<-matrix(nrow=N, ncol=n, data=rnorm(N*n, mean=0, sd=1))

T1<-rep(0, times=N)
for(i in 1:N)
{
T1[i]<-mean(X[i,])
}
T1_mcse<-mean(T1^2)

T2<-rep(0, times=N)
for(j in 1: N)
{
T2[j]<-mean(X[j,],k/n)
}
T2_mcse<-mean(T2^2)

return((T2_mcse)/(T1_mcse))
}
```

Man beachte dabei, dass die Funktion `mean` in R ein zweites, fakultatives Argument hat, nämlich `trim`. Dies bewirkt, dass an beiden Enden der Verteilung der angegebene Anteil der extremen Beobachtungen weggelassen wird.

Man erhält zum Beispiel für die Wahl $N = 10000$, $n = 1001$ und $k = 500$ Werte für `releff_mc(10000,1001,500)` um $1.57 \approx \frac{\pi}{2}$, d.h. Werte, die die relative effizienz aus Aufgabe 3/b approximieren.