Stochastik

Serie 5

Aufgabe 5.1

Verwenden Sie R, um folgende Grössen zu berechnen.

Es sei $X \sim \text{Poisson}(200)$ die Zufallsvariable, die die Anzahl Unfälle in einem Jahr beschreibt.

a) Wie gross ist die Wahrscheinlichkeit, dass in einem Jahr genau 200 Unfälle passieren?

b) Wie gross ist die Wahrscheinlichkeit, dass in einem Jahr höchstens 210 Unfälle passieren?

c) Wie gross ist die Wahrscheinlichkeit, dass in einem Jahr zwischen 190 und 210 Unfälle passieren (beide Grenzen eingeschlossen)?

Aufgabe 5.2

Die Zufallsvariable, die die Anzahl eingehender Telefonanrufe in einer Telefonzentrale innerhalb von 10 Minuten beschreibt, nennen wir X. Sie folgt einer Poissonverteilung mit Erwartungswert $\lambda = 2$, d. h. $X \sim \text{Poisson}(\lambda)$.

a) Wie gross ist die Wahrscheinlichkeit, dass es in einer bestimmten 10-Minuten-Periode keinen einzigen Anruf gibt?

b) Wie gross ist die Wahrscheinlichkeit, dass es nicht mehr als drei Telefonanrufe in einer bestimmten 10-Minuten-Periode gibt?

c) Wie gross ist die Wahrscheinlichkeit, dass es mehr als drei Telefonanrufe in einer bestimmten 10-Minuten-Periode gibt?

d) Angenommen, die Anzahl Anrufe in einer 10-Minuten-Periode ist von der Anzahl Anrufe in einer anderen 10-Minuten-Periode unabhängig. Die Zufallsvariable, die die Anzahl Anrufe in einer Stunde beschreibt bezeichnen wir mit Y. Welcher Verteilung folgt Y?

Aufgabe 5.3

Der Verkehr auf einer Hauptstrasse kann als Poissonprozess modelliert werden. Die Intensität $\lambda = 5$ gibt dabei an, wie viele Fahrzeuge im Mittel pro Minute an einem bestimmten Punkt vorbeikommen. Wie hoch ist die Wahrscheinlichkeit, dass während 2 Minuten kein Auto vorbeifährt? Wie hoch ist die Wahrscheinlichkeit, dass während 30 Sekunden kein Auto vorbeifährt?

Aufgabe 5.4

Angenommen Sie gehen über den Jahrmarkt und kaufen bei einer Losbude 30 Lose. Unter den 30 Losen sind 2 Gewinne. Am nächsten Tag erzählt Ihnen Ihr Studienkollege, dass er am Vorabend bei der gleichen Losbude 50 Lose gekauft hat und darunter 4 Gewinne hatte. Wie kombinieren Sie die beiden Ergebnisse, um mit der Maximum-Likelihood-Methode eine möglichst gute Schätzung der Erfolgswahrscheinlichkeit zu erhalten?

a) Sei X_1 die Zufallsvariable, die die Anzahl Gewinne unter 30 Losen beschreibt („Ihre" Gewinne). Wenn wir annehmen, dass jedes Los unabhängig von jedem anderen Los ein Gewinn oder eine Niete ist, dann folgt X einer Binomialverteilung mit $n_1 = 30$ und unbekanntem Erfolgsparameter π. Abgekürzt schreiben wir: $X_1 \sim \text{Bin}(n_1, \pi)$. Analog sei X_2 die Zufallsvariable, die die Gewinne Ihres Kollegen beschreibt: $X_2 \sim \text{Bin}(n_2, \pi)$ mit $n_2 = 50$ und dem gleichen Wert für die Erfolgswahrscheinlichkeit wie bei X_1. Angenommen, die Anzahl Gewinne, die Sie gezogen haben, ist unabhängig von der Anzahl Gewinne, die Ihr Kollege gezogen hat. Wie lässt sich dann $P(X_1 = x_1 \cap X_2 = x_2)$ schreiben?

b) Der Maximum-Likelihood-Schätzer für π ist derjenige Zahlenwert, der, wenn man ihn anstelle von π einsetzt, den kleinstmöglichen Wert für $-\log(P(X_1 = x_1 \cap X_2 = x_2))$ liefert; diese Funktion wird negative Log-Likelihood-Funktion genannt. Benützen Sie das Resultat aus Teilaufgabe a) und die Eigenschaft des Logarithmus $\log(a \cdot b) = \log(a) + \log(b)$, um die Maximum-Likelihoodschätzung von π mit Hilfe der Software `R` zu verstehen.

```r
# Laden des package stats4 mit der Funktion mle(...)
require(stats4)
x <- c(2, 4)   # gezogene Gewinne
n.i <- c(30, 50)   # Anzahl gezogener Karten
# Definition der negativen Log-Likelihood-Funktion;
# log=TRUE fuehrt dazu, dass der Logarithmus der Wkeit
# berechnet wird
nLL <- function(pi, n) {
```

```
    -sum(dbinom(x.i, prob = pi, size = n.i, log = TRUE))
}
# Schaetzung des Parameters pi mit Startwert pi=0.010
# und festen Parametern n=30,50
fit0 <- mle(nLL, start = list(pi = 0.01), fixed = list(n = n.i),
    nobs = NROW(x))
fit0
```

c) Wie lässt sich $\log(P(X_1 = x_1 \cap X_2 = x_2))$ aufgrund der expliziten Ausdrücke für die Punktwahrscheinlichkeiten der Binomialverteilung schreiben? Versuchen Sie, diesen Term in eine Summe mit mehreren Termen umzuschreiben. Welche Terme hängen von π ab und welche nicht? Finden Sie durch Ableiten und gleich Null setzen den Wert von π in Abhängigkeit von n_1, n_2, x_1 und x_2, der $-\log(P(X_1 = x_1 \cap X_2 = x_2))$ minimiert.

Aufgabe 5.5

In einem Experiment wurde die Anzahl Zerfälle von Americium 241 in einem Intervall von 10 Sekunden gemessen. Das Experiment wurde 1207 Mal wiederholt, jedes Mal wurde die Anzahl Zerfälle in 10 Sekunden gemessen. In der Tabelle von Aufgabe 4 Serie 3 ist in der ersten Spalte die Anzahl Zerfälle aufgeführt, in der zweiten Spalte, wie oft diese Anzahl Zerfälle in den 1207 Experimenten beobachtet worden ist. So sind in 18 der 1207 Experimente 0, 1 oder 2 Alphateilchen detektiert worden. In 28 der 1207 Experimente wurden 3 Alphateilchen detektiert etc.

a) Welche Wahrscheinlichkeitsverteilung würden Sie zur Beschreibung der Anzahl Zerfälle in 10 Sekunden wählen? Berücksichtigen Sie dabei, dass die Gesamtzahl Atome extrem gross ist und ein Zerfall ein seltenes Ereignis darstellt.

b) Warum ist man überhaupt an einem (theoretischen) Wahrscheinlichkeitsmodell interessiert?

c) Bestimmen Sie die theoretische Wahrscheinlichkeitsverteilung, indem Sie die Parameter aus den Daten schätzen und zeichnen Sie diese als Stabdiagramm auf mit Mittelwert und Standardabweichung.

R-Hinweis:

```
plot(..., ..., type = "h",
     xlab = "Anzahl Zerfaelle in 10 Sekunden",
     ylab = "Wahrscheinlichkeit")
abline(v = ..., col = "red")
```

d) Sie machen ein weiteres Experiment und beobachten 20 Zerfälle in 10 Sekunden. Wenn Sie eine Million Experimente durchführen, wie oft werden Sie ein solches oder extremeres Ereignis beobachten? Wie nennt man diese Wahrscheinlichkeit?

Aufgabe 5.6

Wir betrachten die geometrische Verteilung. Dies ist eine diskrete Wahrscheinlichkeitsverteilung mit den Punktwahrscheinlichkeiten

$$P(X = x) = (1 - \pi)^{x-1}\pi, \quad x = 1, 2, 3, \ldots$$

wobei $0 < \pi < 1$ die Erfolgswahrscheinlichkeit ist. Eine geometrisch verteilte Zufallsvariable X beschreibt die Anzahl Versuche bis zum ersten Erfolg bei unabhängigen Bernoulliversuchen mit Erfolgswahrscheinlichkeit π. Der Erwartungswert von einer geometrisch verteilten Zufallsvariablen X lautet $\mathrm{E}(X) = \frac{1}{\pi}$.

Sie können die Punktwahrscheinlichkeit $P(X = x)$ der geometrischen Verteilung mit Gewinnwahrscheinlichkeit π mit Hilfe folgender R-Funktion berechnen

```
dgeom(x - 1, prob = pi)
```

Wir haben folgende Beobachtungen

$$x_1 = 8, x_2 = 7, x_3 = 12, x_4 = 5, x_5 = 10, x_6 = 9$$

und wir wollen den Parameter π aus dieser Stichprobe schätzen.

a) Bestimmen Sie die Likelihood-Funktion basierend auf n unabhängigen identisch verteilten Beobachtungen $x_1, \ldots, x_n$ einer geometrisch verteilten Zufallsvariablen.

b) Bestimmen Sie die negative Log-Likelihood-Funktion aufgrund der 6 unabhängigen Beobachtungen, wobei Sie die Punktwahrscheinlichkeiten nicht explizit ausschreiben müssen. Setzen Sie den Funktionsausdruck in den untenstehenden Code ein und schätzen Sie die Gewinnwahrscheinlichkeit von π.

```
require(stats4)
x <- c(8, 7, 12, 5, 10, 9)
nLL <- function(pi) {
    -sum(...)
}
fit0 <- mle(nLL, start = list(pi = 0.01), nobs = NROW(x))
fit0
```

c) Bestimmen Sie den zugehörigen Maximum-Likelihood-Schätzer für π aufgrund der 6 unabhängigen Stichproben, indem Sie die Log-Likelihood-Funktion mit den expliziten Ausdrücken für die Punktwahrscheinlichkeiten berechnen, ableiten, gleich Null setzen und nach π auflösen.

d) Bestimmen Sie den Momentenschätzer für π aufgrund der 6 unabhängigen Stichproben. Vergleichen Sie mit dem Maximum-Likelihood-Schätzer.

Kurzlösungen vereinzelter Aufgaben

A 5.1:

a) 0.0282
b) 0.773
c) 0.542

A 5.2:

a) 0.135
b) 0.857
c) 0.143

A 5.3:

a) $4.54 \cdot 10^{-5}$
b) 0.0821

A 5.4:

b)

$$-\log\left(P(X_1 = x_1; n_1, \pi) \cdot P(X_2 = x_2; n_2, \pi)\right) = -\log(P(X_1 = x_1; n_1, \pi) - \log(P(X_2 = x_2; n_2, \pi)$$

c) $\hat{\pi} = \frac{x_1+x_2}{n_1+n_2} = 0.075$

A 5.5:

c) $\widehat{\mathrm{E}(X)} = 8.37$ und $\widehat{\sigma(X)} = 2.89$
d) 0.00044 und 443

A 5.6:

a) $L(\pi; x_1, \ldots, x_n) = \prod_{i=1}^{n}(1-\pi)^{x_i-1}\pi$

b)
```
nLL <- function(pi) {
    -sum(dgeom(x - 1, prob = pi, log = TRUE))
}
```

$\hat{\pi}_{\mathrm{MLE}} = 0.118$

c) $\hat{\pi}_{\mathrm{MLE}} = \frac{n}{\sum_{i=1} x_i} = 0.118$

d) $\hat{\pi}_{\mathrm{MoM}} = 0.118$