ETH Zürich HS 2014
D-INFK
Prof. Dr. P. Embrechts

Koordinator
Blanka Horvath

Wahrscheinlichkeit und Statistik
Lösungsskizze 10

10-1

$\lambda = n/N = 3.87$ ist die durchschnittliche Anzahl der Teilchen, die pro Zeitintervall registriert wurde. Für dieses λ vervollständigt sich die Tabelle wie folgt:

k	0	1	2	3	4	5	6	7	8	9	≥ 10
N_k	57	203	383	525	532	408	273	139	45	27	16
$N \cdot p_k$	54	210	407	525	508	394	254	141	68	29	17

Die "theoretischen Werte" scheinen recht gut zu den beobachteten Werten zu passen. Zwar sind insbesondere für $k = 2, 4, 8$ Abweichungen vorhanden, aber nicht extrem.

10-2

a) Die Likelihood-Funktion ist gegeben durch

$$L(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^{n} P_\lambda[X_i = x_i] = e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!},$$

und für die log-Likelihood-Funktion erhalten wir somit

$$\log L(x_1, \ldots, x_n; \lambda) = -n\lambda + \sum_{i=1}^{n} \left(x_i \log(\lambda) - \log(x_i!) \right).$$

Um über λ zu maximieren setzen wir die entsprechende Ableitung Null:

$$\frac{\partial}{\partial \lambda} \log L(x_1, \ldots, x_n; \lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0,$$

es folgt

$$\lambda = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}_n,$$

also ist der ML-Schätzer für λ

$$T = \bar{X}_n.$$

b) Mit $n = 3000$ erhalten wir den Schätzwert

$$\hat{\lambda} = \frac{1}{3000}(101 \cdot 0 + 317 \cdot 1 + 535 \cdot 2 + \cdots + 1 \cdot 11 + 1 \cdot 12) = \frac{10'569}{3000} = 3.523.$$

10-3

a) Für $i \geq 1$ sei X_i das Einfüllgewicht des i-ten Kartoffelsacks. Die X_i, $i \geq 1$, sind $i.i.d$ mit $X_i \sim \mathcal{U}(9.75, 10.75)$. Sei $S_n = \sum_{i=1}^n X_i$. Dann gilt:

$$\begin{aligned} E[X_i] &= 10.25, \qquad \text{also } E[S_{146}] = 146E[X_1] = 146 \cdot 10.25 = 1496.5, \\ \operatorname{var}[X_i] &= \frac{1}{12}, \qquad \text{also } \operatorname{var}[S_{146}] = \frac{146}{12} = 12.1\bar{6} \ \text{ und } \ \sigma(S_{146}) \approx 3.46. \end{aligned}$$

Mit dem Zentralen Grenzwertsatz folgt

$$Z = \frac{S_{146} - E[S_{146}]}{\sqrt{\operatorname{var}[S_{146}]}} \overset{\text{approx.}}{\sim} \mathcal{N}(0,1).$$

Also ist

$$\begin{aligned} P[S_{146} > 1500] &= P\left[\frac{S_{146} - E[S_{146}]}{\sqrt{\operatorname{var}[S_{146}]}} > \frac{1500 - 1496.5}{3.49}\right] \\ &\approx P[Z > 1.01] = 1 - P[Z \leq 1.01] = 1 - \Phi(1.01) \approx 0.156. \end{aligned}$$

b) Die Gleichung $P[S_n > 1500] = 0.01$ ist nach n aufzulösen, d.h.

$$0.01 \overset{!}{=} P[S_n > 1500] = P\left[\frac{S_n - E[S_n]}{\sigma(S_n)} > \frac{1500 - E[S_n]}{\sigma(S_n)}\right] \approx 1 - \Phi\left(\frac{1500 - 10.25n}{\sqrt{n/12}}\right);$$

also möchte man approximativ

$$\frac{1500 - 10.25n}{\sqrt{n/12}} \overset{!}{=} \Phi^{-1}(0.99) \approx 2.326.$$

Mit $x = \sqrt{n}$ erhalten wir somit die quadratische Gleichung

$$1500 - 10.25x^2 - \frac{2.326}{\sqrt{12}}x = 0$$

mit Lösungen $x_1 \approx -12.13$ und $x_2 \approx 12.06$ (wegen $x = \sqrt{n}$ ist nur die positive Lösung relevant). Man kann also höchstens $n = 145$ Säcke laden.
Bemerkung: Obwohl man im Vergleich zu **a)** nur einen Sack weniger lädt, reduziert sich die Wahrscheinlichkeit das zulässige Ladegewicht von 1500 kg zu überschreiten von 15.6% auf unter 1%. Dies ist plausibel, denn während der Erwartungswert $E[S_{146}] = 1496.5$ nur etwa eine Standardabweichung $\sigma(S_{146}) \approx 3.46$ unter dem zulässigen Ladegewicht liegt, ist das erwartete Ladegewicht von 145 Säcken $E[S_{145}] = 1486.25$ fast vier Standardabweichungen $\sigma(S_{145}) \approx 3.46$ unterhalb von 1500.

10-4

a) Für $i = 1, \ldots, n$ sei K_i das Ereignis, dass die Münze des i-ten Befragten Kopf zeigt und D_i das Ereignis, dass der i-te Befragte schon mindestens einmal einen Ladendiebstahl begangen hat. Wir definieren

$$X_i := I_{K_i^c} + I_{D_i^c \cap K_i};$$

d.h. $X_i = 1$ genau dann, wenn der i-te Befragte mit "Nein" antwortet und $X_i = 0$ genau dann, wenn der i-te Befragte mit "Ja" anwortet. Da die Ereignisse $(D_i, K_i)_{i=1,\dots,n}$ unabhängig sind, sind auch die Zufallsvariablen $X_1, \dots, X_n$ unabhängig. Ferner ist jedes X_i $Be(r)$-verteilt mit

$$r = P[X_i = 1] = 1 - P[X_i = 0] = 1 - P[D_i \cap K_i] = 1 - P[D_i]P[K_i] = 1 - \frac{p}{2}.$$

Die X_i, $i = 1, \dots, n$, sind also *i.i.d.* $\sim Be(r)$. Insbesondere ist die Wahrscheinlichkeit, dass eine beliebig ausgewählte Person mit "Nein" anwortet gleich $P[X_i = 1] = 1 - \frac{p}{2}$.

b) Gesucht ist $E[S_{1000}]$. Die Linearität des Erwartungswertes liefert

$$E[S_{1000}] = E\left[\sum_{i=1}^{1000} X_i\right] = \sum_{i=1}^{1000} E[X_i] = 1000\left(1 - \frac{1}{2}p\right).$$

c) Gesucht ist eine obere Schranke für $P[S_{1000} \geq 909]$. Da $X_1, \dots, X_{1000}$ unabhängige Bernoulli-Zufallsvariablen sind (vgl. **a)**), können wir Satz 5.7 aus der Vorlesung anwenden. Zunächst müssen wir $\delta > 0$ so bestimmen, dass $(1+\delta)E[S_{1000}] = 909$. Mit **b)** und $p = 0.2$ erhält man aus dieser Gleichung $\delta = 0.01$. Satz 5.7 gibt dann die Chernoff-Schranke

$$P[S_{1000} \geq 909] = P[S_{1000} \geq (1+\delta)E[S_{1000}]] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{E[S_{1000}]} \approx 0.956.$$

Die Wahrscheinlichkeit, dass mindestens 909 Personen mit "Nein" antworten ist also höchstens 95.6%.