Wahrscheinlichkeit und Statistik

Lösungen Serie 11

Lösung 11-1. (a) Die Likelihood-Funktion ist gegeben durch

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

und für die log-Likelihood-Funktion erhalten wir somit

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

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

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

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

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

Lösung 11-2. $\lambda=n/N=3.87$ 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.

Lösung 11-3. Seien $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim}$ UNIF(a,b)-verteilt und $X:=(X_1,\dots,X_n)$.

(a) Mit $\theta=(a,b)$ ist die Likelihood-Funktion gleich

$$L(x_1,\dots,x_n;\theta)=\frac{1}{(b-a)^n}\prod_{i=1}^n 1_{[a,b]}(x_i). \tag{1}$$

Nun muss $L(x_1,\dots,x_n;\theta)$ für feste $(x_1,\dots,x_n)$ bezüglich a und b maximiert werden. Seien $x_*:=\min_{1\leq i\leq n}x_i$ und $x^*:=\max_{1\leq i\leq n}x_i$. Falls x_* oder x^* ausserhalb $[a,b]$ liegen, dann verschwindet die rechte Seite von (1). Also muss die Maximumstelle $\hat{\theta}_{\mathrm{ML}}=(\hat{a}_{\mathrm{ML}},\hat{b}_{\mathrm{ML}})$

Siehe nächste Seite!

die Bedingungen $\hat{b}_{\mathrm{ML}} \geq x^* \geq x_* \geq \hat{a}_{\mathrm{ML}}$ erfüllen. Für $\hat{a}_{\mathrm{ML}} := x_* = \min_{1\leq i\leq n} x_i$ und $\hat{b}_{\mathrm{ML}} := x^* = \max_{1\leq i\leq n} x_i$ ist

$$L(x_1, \ldots, x_n; \theta) < L(x_1, \ldots, x_n; \hat{\theta}_{\mathrm{ML}})$$

f"ur alle $\theta = (a, b)$. Somit sind die Maximum-Likelihood Schätzer für a und b gegeben durch

$$T_a = \min_{1\leq i\leq n} X_i \quad \text{und} \quad T_b = \max_{1\leq i\leq n} X_i.$$

Für die Schätzwerte erhalten wir

$$\hat{a}_{\mathrm{ML}} = 9.2 \quad \text{und} \quad \hat{b}_{\mathrm{ML}} = 18.85.$$

(b) Es gilt:

$$E[X_i] = \frac{a+b}{2} \quad \text{und} \quad \mathrm{Var}(X_i) = \frac{(b-a)^2}{12}.$$

Die Schätzwerte mittels Momentenmethode erfüllen also

$$\bar{x} = \frac{\hat{a}+\hat{b}}{2} \quad \text{und} \quad \tilde{s} = \sqrt{\frac{(\hat{b}-\hat{a})^2}{12}}.$$

Es gilt $\bar{x} = 12.94$ und $\tilde{s}^2 = 8.5864$, also:

$$\hat{a}_{\mathrm{MM}} = 7.86 \quad \text{und} \quad \hat{b}_{\mathrm{MM}} = 18.02.$$

Lösung Challenge Serie 11. Wir leiten eine Gleichung für die Anzahl Schokostückchen N her in Abhängigkeit von der Anzahl Cookies (100) und der Wahrscheinlichkeit p, dass jeder Cookie mindestens 1 Schokostöckchen enthält.

Die Zahl $1-p$ ist dann die Wahscheinlichkeit, dass mindestens ein Cookie kein Schokostückchen enthält. Wir bezeichnen mit

$$A_k := \{\text{Cookie Nr. } k \text{ hat keine Schokostückchen}\} \quad k = 1, \ldots, 100.$$

Exakte Lösung: Dann gilt nach der Inklusion-Exklusionsformel Serie 2 Aufgabe 3 die Identität

$$1 - p = P\left[\bigcup_{k=1}^{100} A_k\right] = \sum_{k=1}^{100} (-1)^{k+1} \sum_{1\leq j_1 < \cdots < j_k \leq 100} P\left[\bigcap_{\ell=1}^{k} A_{j_\ell}\right].$$

Der k-te Summand in dieser Formel ist die Summe von Wahrscheinlichkeiten von der Form

$$P\left[\bigcap_{\ell=1}^{k} A_{j_\ell}\right] = P\left[A_{j_1} \cap \cdots \cap A_{j_k}\right]. \tag{2}$$

Davon gibt es genau $\binom{100}{k}$ Summanden. Wegen der Gleichverteilung (oder Symmetrie) sind die Wahrscheinlichkeiten (2) konstant in $(j_1, \ldots, j_k)$ und

$$P\left[\bigcap_{\ell=1}^{k} A_{j_\ell}\right] = \left(1 - \frac{\ell}{k}\right)^N.$$

Damit erhalten wir

$$1 - p = 1 - \sum_{k=1}^{100} (-1)^{k+1} \binom{100}{k} \left(1 - \frac{k}{100}\right)^N.$$

Siehe nächstes Blatt!

Setzen wir nun $p = 0.9$, dann erhalten wir mit Hilfe eines Computers $N \approx 683$.

Approximative Lösung (Union bound): Wir verwenden die allgemein gültige Formel

$$1 - p = P\left[\bigcup_{k=1}^{100} A_k\right] \leq \sum_{k=1}^{100} P[A_k].$$

Da $P[A_j] = \left(1 - \frac{1}{100}\right)^N$, erhalten wir

$$1 - p \leq 100\left(1 - \frac{1}{100}\right)^N \qquad \text{oder} \qquad p \geq 1 - 100\left(1 - \frac{1}{100}\right)^N.$$

Setzen wir $p = 0.9$ ein, so erhalten wir

$$0.9 \geq 1 - 100\left(1 - \frac{1}{100}\right)^N \qquad \text{also} \qquad N \leq \frac{\log(10^{-3})}{\log(0.99)} \approx 688.$$

Verglichen mit der exakten Lösung 683 ist die approximative Lösung 688 recht gut.