Lösung

Diskrete Wahrscheinlichkeitstheorie – Hausaufgabe 11

Abgabe bis zum 11.7. bis 8:30.

Alle Antworten sind unter Angabe des Rechenwegs zu begründen, soweit nicht anders gefordert! Fragen gerne im infler-Forum posten :).

Aufgabe 11.1 Abzugeben 2P

Wir schätzen den Mittelwert einer $\mathcal{N}(\mu, 3)$-verteilten ZV indem wir n identisch verteilte Stichproben $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 3)$ betrachten und ihren Mittelwert bilden $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$.

Sei $n = 3000$. Bestimmen Sie ein möglichst kleines (symmetrisches) Konfidenzintervall für den Parameter μ zum Konfidenzniveau 0.99.

Lösung: Beachte zunächst, dass $\overline{X} \sim \mathcal{N}(\mu, \frac{3}{n})$, also ist $\frac{\sqrt{n}(\overline{X}-\mu)}{\sqrt{3}}$ standardnormalverteilt.

Wir wollen ein minimales $\delta > 0$ bestimmen, so dass:

$$\Pr\left[|\overline{X} - \mu| > \delta\right] \leq 0.01 \Leftrightarrow \Pr\left[|\frac{\sqrt{n}(\overline{X} - \mu)}{\sqrt{3}}| > \frac{\delta\sqrt{n}}{\sqrt{3}}\right] \leq 0.01$$

$$\Leftrightarrow 2\Phi(-\frac{\delta\sqrt{n}}{\sqrt{3}}) \leq 0.01 \Leftrightarrow \Phi(-\frac{\delta\sqrt{n}}{\sqrt{3}}) \leq 0.005$$

$$\Leftrightarrow \frac{\delta\sqrt{n}}{\sqrt{3}} \geq 2.5758 \Leftrightarrow \delta \geq \frac{\sqrt{3} \cdot 2.5758}{\sqrt{n}} = 0.08145\ldots$$

Hinreichend (aber noch möglichst klein) ist also z.B. $\delta = 0.0815$, damit ist das Konfidenzintervall für μ zu der gegebenen Stichprobe $K = [\overline{X} - \delta, \overline{X} + \delta]$. Für eine gegebene Stichprobe liegt der wahre Wert von μ also mit mindestens 99%-iger Wkeit im Intervall K.

Aufgabe 11.2 Abzugeben 3P

Sei X eine ZV mit folgender Dichte für unbekannten Parameter $\lambda > 0$:

$$f(t) = 2\lambda t \cdot e^{-\lambda t^2} I_{[0,\infty)}(t).$$

Sei $\vec{x} = (x_1, \ldots, x_n)$ eine Stichprobe von X.

Bestimmen Sie den Maximum-Likelihood-Schätzer für den Parameter λ.

Hinweis: Maximieren Sie den Logarithmus der Likelihood-Funktion $\ln(L(\vec{x}; \lambda))$.

Lösung: Wir betrachten die Log-Likelihood-Funktion $f(\lambda) := \ln(L(\vec{x};\lambda))$. Diese ist maximal für λ genau dann, wenn die Likelihood-Funktion maximal ist, da ln monoton ist.

$$f(\lambda) = \ln(\prod_{i=1}^{n} 2\lambda x_i e^{-\lambda x_i^2}) = n\ln(2\lambda) + \sum_{i=1}^{n}(\ln x_i - \lambda x_i^2)$$

Um das Maximum dieser Funktion zu finden, differenzieren wir sie nach λ und setzen sie 0.

$$\frac{\partial f}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i^2 \stackrel{!}{=} 0$$

$$\Leftrightarrow \lambda = \frac{n}{\sum_{i=1}^{n} x_i^2}$$

(wie man sofort sieht, ist dies tatsächlich ein Maximum, da z.B. die zweite Ableitung von f überall negativ ist).

Aufgabe 11.3 **Abzugeben** **3P+1P+1P+1P**

Michel und seine kleine Schwester Maxi streiten sich fast täglich. In letzter Zeit hatte Michel öfters blaue Flecken am Arm. Maxi behauptet, dass diese daher kommen, dass ihr Bruder sehr tollpatschig sei und öfter gegen Einrichtungsgegenstände laufe.

Xaver, der Vater der beiden, entschließt sich dazu Michels Tollpatschigkeit zu testen. Dazu lässt er den Kleinen über mehrere Wochen hinweg insgesamt 8 mal durch (verschiedene) Parcours im Spielplatz turnen und zählt, wie oft er dabei blaue Flecken bekommt.

Sei X_i eine binäre Zufallsvariable, die signalisiert, ob sich Michel beim i-ten Parcours blaue Flecken geholt hat. Wir nehmen an, dass $X_i \sim \text{Bin}(1,p)$ für unbekanntes p (ein Maß für die Ungeschicklichkeit Michels) und alle X_i unabhängig sind.

Aus seinem Bekanntenkreis weiß Xaver, dass sich ein normales Kind etwa jedes dritte Mal auf dem Spielplatz blaue Flecken holt. Er möchte entscheiden, ob Michel ungeschickter ist, als ein normales Kind.

Bei seinem Test will er die Wkeit, dass er Michel zu Unrecht als ungeschickt bezeichnet auf höchstens 0.05 begrenzen.

(a) Formulieren Sie eine geeignete Nullhypothese und einen Test für diese Hypothese (d.h. definieren sie Testgröße und Ablehnungsbereich). Was ist das Signifikanzniveau Ihres Tests?

Hinweis: Der zentrale Grenzwertsatz ist hier nicht anwendbar (warum?)—rechnen Sie daher nicht approximativ!

(b) Berechnen Sie die Wkeit für einen Fehler 2. Art.

(c) Xaver bemerkt, dass sich Michel 3 mal auf dem Spielplatz gestoßen und blaue Flecken bekommen hat. Kann er, aufgrund dieser Beobachtung, die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$ ablehnen?

(d) Deuten die Ergebnisse darauf hin, dass Maxi wahrscheinlich ihren großen Bruder beim Streiten verprügelt? Warum?/Warum nicht?

Lösung:

(a) $H_0 : p \leq \frac{1}{3}$ In Worten: Michel ist höchstens so ungeschickt, wie ein normales Kind. $H_1 : p > 1/3$ (triviale Alternative). Testgröße $T := \sum_{i=1}^{8} X_i$. Ablehnungsbereich $K = \{k, \ldots, 8\}$ für zu bestimmendes k. Wir müssen nach Aufgabenstellung k so wählen, dass $\Pr_{p=1/3}[T \geq k] \leq 0.05$.

Man berechnet leicht die Zähldichte der $\text{Bin}(8, 1/3)$-Verteilung für k nahe bei 8:

k	$f_T(k)$ für $p = 1/3$
5	$\binom{8}{5}(\frac{1}{3})^5(\frac{2}{3})^3 = 0.06\ldots$
6	0.01707
7	0.002439
8	$1.524 \cdot 10^{-4}$

Also muss $k = 6$ gewählt werden. Das Signifikanzniveau ist dann $\alpha = \max_{p \leq 1/3} \Pr_p[T \geq 6] = \Pr_{p=1/3}[T \geq 6] \approx 0.0197$.

(b) Der Fehler 2. Art ist definiert als $\sup_{p>1/3} \Pr_p[T \leq 5] = 1 - \Pr_{p=1/3}[T \geq 6] = 1 - \alpha \approx 0.9803$. Wie man sieht, kann man bei einem Test mit trivialer Alternative nicht beide Fehler klein werden lassen!

(c) $3 \notin K$ also wird H_0 nicht abgelehnt.

(d) Nein... mögliche Gründe sind z.B.:

(i) Die Wkeit H_0 nicht abzulehnen obwohl Michel in Wahrheit eigentlich sehr ungeschickt ist, ist bei diesem Test ziemlich hoch. Z.B. sei $p = 2/3$, dann ist die Wkeit einen Fehler 2. Art zu begehen etwa 0.53.

(ii) Selbst wenn man annimmt, dass es sehr unwahrscheinlich ist dass sich Michel durch Zufall gestoßen hat, kann man noch nicht darauf schließen, dass Maxi ihren (großen!) Bruder verprügelt (was schon a priori sehr unwahrscheinlich ist!). Es gibt sicher wahrscheinlichere Gründe für Michels blaue Flecken. (siehe "Prosecutor's Fallacy")

Häufig missverstanden (v.a. im juristischen Kontext) wird, dass man bei einem Test nie die Wkeit berechnet, dass eine der Hypothesen "wahr" ist, sondern immer nur eine Art "bedingte Wahrscheinlichkeiten": **wenn** H_0 wahr ist, dann ist die Wkeit falsch zu entscheiden ≤ 0.05, was aber nichts über die Wkeit aussagt, **dass** H_0 wahr ist. (Vorsicht: das ist keine bedingte Wkeit im eigentlichen Sinn da "H_0 ist wahr" kein Ereignis ist! Es ist i.A. auch gar nicht möglich einen entsprechenden WRaum anzugeben in dem H_0 leben würde!).

Aufgabe 11.4 Abzugeben **3P+1P**

Wir wollen mit einem statistischen Test die Höhe zweier Berge vergleichen. Dazu vermessen wir beide wiederholt mit demselben Messgerät.

Wir nehmen an, dass die Messwerte jeweils normalverteilt sind mit bekannter Standardabweichung (Genauigkeit des Messgeräts) $\sigma = 15$.

Seien $X_1, \dots, X_m$ die ZVen, welche die Messungen der Höhe des ersten Berges beschreiben und $Y_1, \dots, Y_n$ die Messungen des zweiten Berges. Laut Annahme sind $X_i \sim \mathcal{N}(\mu_X, \sigma^2)$ und $Y_i \sim \mathcal{N}(\mu_Y, \sigma^2)$ mit unbekannten Erwartungswerten μ_X und μ_Y (aber bekannter Varianz $\sigma^2 = 225$).

Uns interessiert, ob beide Berge gleich hoch sind, also testen wir $H_0 : \mu_X = \mu_Y$ gegen die triviale Alternative $H_1 : \mu_X \neq \mu_Y$.

(a) Definieren Sie eine geeignete Testgröße und ein Ablehnungskriterium für die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$.

Hinweis: Die Testgröße sollte (unter der Nullhypothese) standardnormalverteilt sein.

(b) Gegeben seien die beiden Messreihen der Berghöhen

x_i	1202	1178	1191	1174	1211	1226	1199	1157	1208	1203
y_i	1224	1179	1182	1217	1247	1189	1221	1219	—	—

Können Sie, aufgrund dieser Daten, die Nullhypothese zum Signifikanzniveau $\alpha = 0.05$ ablehnen?

Lösung:

(a) $\overline{X} - \overline{Y} \sim \mathcal{N}(\mu_X - \mu_Y, \sigma^2(1/m + 1/n))$ also setzen wir $T := \frac{\overline{X} - \overline{Y}}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}}$. Dann ist T unter der Nullhypothese standardnormalverteilt.

Wir setzen als Ablehnungskriterium $|T| > z_{1-\alpha/2}$. Dann ist die Wkeit, dass die Nullhypothese H_0 abgelehnt wird, obwohl sie gilt: $\Pr\left[|T| > z_{1-\alpha/2}\right] = 2\Phi(-z_{1-\alpha/2}) = 2(1 - \Phi(z_{1-\alpha/2})) = 2 - (2 - \alpha) = \alpha$.

(b) Für die Daten ergibt sich $\overline{X} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1194.9$ und $\overline{Y} = \frac{1}{8}\sum_{i=1}^{8} y_i = 1209.75$.

Es ist $\sigma\sqrt{\frac{1}{m} + \frac{1}{n}} = 15 \cdot \frac{3}{2\sqrt{10}}$. Für die Testgröße T ergibt sich also $|T| = \frac{14.85\sqrt{10}}{15 \cdot \frac{3}{2}} = 2.087\dots > 1.96 = z_{0.975}$. Wir können die Nullhypothese also ablehnen (zum Signifikanzniveau 0.05).

Aufgabe 11.5 Abzugeben: ai), aiii), bi),bii) **1P+2P+1P+1P**

Die Dichte der χ^2-Verteilung mit n Freiheitsgraden ($n \in \mathbb{N}$) ist

$$\frac{2^{-n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-1/2x} I_{[0,\infty)}(x),$$

wobei $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(1) = 1$ und $\Gamma(r+1) = r\Gamma(r)$ gilt.

(a) Es seien X_1, X_2 unabhängige, normalverteilte ZV mit (unbekanntem) Erwartungswert μ und (unbekannter) Varianz σ^2. Es gelte $U := X_1 - X_2$ und $V := X_1 + X_2$.

(i) Bestimmen Sie die Verteilungen von U und V.

(ii) Zeigen Sie, dass U und V unabhängig sind.

(iii) Zeigen Sie nun, dass $\frac{1}{\sigma^2} S_X^2 = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2}{\sigma^2}$ gerade χ^2-verteilt ist mit einem Freiheitsgrad.

(b) Seien Y_1, Y_2 nun (unabhängig) normalverteilt mit bekanntem Erwartungswert 0 und unbekannter Varianz σ^2.

Da der Erwartungswert bekannt ist, können wir versuchem die Varianz direkt mittels $T := Y_1^2 + Y_2^2$ zu schätzen.

(i) Bestimmen Sie $c \in \mathbb{R}$ so, dass $c \cdot T$ ein erwartungstreuer Schätzer für σ^2 ist.

(ii) Zeigen Sie, dass $\frac{T}{\sigma^2}$ gerade χ^2-verteilt ist mit zwei Freiheitsgraden.

Hinweis: Verwenden Sie, dass für unabhängige ZVen Z_1, Z_2 mit Z_i χ^2-verteilt mit n_i Freiheitsgraden gilt: $Z_1 + Z_2$ ist χ^2-verteilt mit $n_1 + n_2$ Freiheitsgraden.

Lösung:

(a) (i) Nach Vorlesung gilt auf Grund der Unabhängigkeit:

$U \sim \mathcal{N}(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2) = \mathcal{N}(0, 2\sigma^2)$.

$V \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2) = \mathcal{N}(2\mu, 2\sigma^2)$.

(ii) Die gemeinsame Verteilung ergibt sich mittels der Transformation

$$\Theta\colon (x_1, x_2) \mapsto (u, v) := (x_1 - x_2, x_1 + x_2)$$

mit

$$J_\Theta = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad J_\Theta^{-1} = \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}, \quad \det J_\Theta = 1/2$$

zu

$$\begin{aligned}
\Pr[U \le s, V \le t] &= \int_{\{(x_1,x_2)\in\mathbb{R}^2 \mid x_1 - x_2 \le s, x_1 + x_2 \le t\}} \phi(x_1)\phi(x_2) dx_1 dx_2 \\
&= \int_{\{(u,v)\in\mathbb{R}^2 \mid u \le s, v \le t\}} \phi(\frac{u+v}{2})\phi(\frac{v-u}{2}) 1/2 dudv \\
&= \int_{u=-\infty}^{s} \int_{v=-\infty}^{t} \frac{1}{\sqrt{4\pi}\sigma} e^{-\frac{(u+v-\mu)^2}{8\sigma^2}} \frac{1}{\sqrt{4\pi}\sigma} e^{-\frac{(v-u-\mu)^2}{8\sigma^2}} dudv \\
&= \int_{u=-\infty}^{s} \int_{v=-\infty}^{t} \frac{1}{\sqrt{4\pi}\sigma} e^{-\frac{u^2+2u(v-\mu)+(v-\mu)^2}{8\sigma^2}} \frac{1}{\sqrt{4\pi}\sigma/2} e^{-\frac{u^2-2u(v-\mu)+(v-\mu)^2}{8\sigma^2}} dudv \\
&= \int_{u=-\infty}^{s} \int_{v=-\infty}^{t} \frac{1}{\sqrt{4\pi}\sigma} e^{-\frac{u^2}{4\sigma^2}} \frac{1}{\sqrt{4\pi}\sigma} e^{-\frac{(v-\mu)^2}{4\sigma^2}} dudv
\end{aligned}$$

Damit ist die gemeinsame Verteilung gerade das Produkt der Randverteilungen, und U, V somit unabhängig. Anschaulich entstehen U, V gerade durch eine Drechung um $\pi/2$ aus X_1, X_2.

(iii) Es gilt (siehe auch J_Θ^{-1}):

- $X_1 = \frac{1}{2}(U + V)$,
- $X_2 = \frac{1}{2}(-U + V)$ und
- $\overline{X} = \frac{1}{2}V$.

Damit ergibt sich

$$S_X^2 = (X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 = (\frac{1}{2}U + \frac{1}{2}V - \frac{1}{2}V)^2 + (-\frac{1}{2}U + \frac{1}{2}V - \frac{1}{2}V)^2 = \frac{1}{4}U^2 + \frac{1}{4}U^2 = \frac{1}{2}U^2.$$

Somit ist $\frac{1}{\sigma}S_X = \frac{U}{\sqrt{2\sigma^2}}$ standardnormalverteilt, und es folgt für $t \in [0, \infty)$:

$$\begin{aligned}
\frac{d}{dt}\Pr\left[\frac{U^2}{2\sigma^2} \le t\right] &= \frac{d}{dt}\Pr\left[\left|\frac{U}{\sqrt{2}\sigma}\right| \le \sqrt{t}\right] = \frac{d}{dt}(2\Phi(\sqrt{t}) - 1) \\
&= t^{-1/2}\varphi(\sqrt{t}) \\
&= \frac{1}{\sqrt{2\pi}} t^{-1/2} e^{-t/2} = \frac{2^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})} t^{\frac{1}{2}-1} e^{-\frac{1}{2}t},
\end{aligned}$$

was gerade die Dichte der χ^2-Verteilung mit einem Freiheitsgrad ist.

M.a.W.: Ist $X \sim \mathcal{N}(0,1)$, dann $X^2 \sim \chi_1^2$.

(b) (i) Es muss $c = 1/2$ gelten, da:

$$\mathbb{E}[cT] = c\mathbb{E}[T] = c2\mathbb{E}[Y_1^2] = c2\mathbb{E}[(Y_1 - \mathbb{E}[Y_1])^2] = 2c\mathrm{Var}[Y_1] = 2c\sigma^2.$$

(ii) Nach (aiii) gilt $Y_i^2/\sigma^2 \sim \chi_1^2$. Somit ist die Summe nach Hinweis χ^2-verteilt mit zwei Freiheitsgraden.