Technische Universität München
Fakultät für Informatik
Lehrstuhl Informatik XIV
Prof. Dr. Susanne Albers
Dr. Werner Meixner

Sommersemester 2014
Lösungsblatt 9
27. Juni 2014

Diskrete Wahrscheinlichkeitstheorie

Hausaufgabe 1 (5 Punkte)

Sie werfen eine kreisförmige Münze mit Radius 0.5 auf ein Quadrat mit Seitenlänge 4, wobei der Mittelpunkt der Münze stets innerhalb des Quadrats zu liegen kommt.

Wie hoch ist die Wahrscheinlichkeit, dass die Münze über den Rand des Quadrats hinausragt, wenn wir eine Gleichverteilung des Mittelpunkts der Münze auf der Quadratfläche annehmen?

Lösung

Die Münze ragt genau dann nicht über den Rand des Quadrats Q hinaus, wenn der Mittelpunkt der Münze in einem abgeschlossenen Quadrat B mit Seitenlänge 3 liegt, das zentriert innerhalb von Q liegt.

Damit ist klar, dass die gesuchte Wahrscheinlichkeit Pr durch das Verhältnis der Flächen von $Q \setminus B$ zu Q gegeben ist.

Ergebnis: $\Pr = \frac{7}{16}$.

Bemerkung: Die Aufgabe kann auch durch Rechnung gelöst werden, wie folgt:
Wir bezeichnen $[a, b]_{\mathbb{R}} := \{x \in \mathbb{R}\,;\, a \leq x \leq b\}$.
Seien $Q = \{(x, y)\,;\, x, y \in [0, 4]_{\mathbb{R}}\}$ und $Q = \{(x, y)\,;\, x, y \in [0.5, 3.5]_{\mathbb{R}}\}$.
Man definiert zwei kontinuierliche Zufallsvariablen X und Y über einem Kolmogorov'schen Wahrscheinlichkeitsraum $\langle \Omega, \mathcal{A}, \Pr \rangle$ durch Angabe der 2-dimensionalen Dichte

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{16} : (x, y) \in Q, \\ 0 \;\; : \text{sonst.} \end{cases}$$

Dann gilt

$$\Pr[(x, y) \notin B] \; = \; 1 - \int_B f_{X,Y}(x, y) \, \mathrm{d}x \, \mathrm{d}y \, .$$

(5P)

Hausaufgabe 2 (5 Punkte)

Die Lebensdauer T eines Rechners habe die folgende mit $\lambda > 0$ und a parametrisierte Dichte:

$$f_T(t) = \begin{cases} a\lambda^2 t e^{-\lambda t} & : t \geq 0 \\ 0 & : t < 0 \end{cases}$$

1. Welchen Wert muss a besitzen, so dass für alle $\lambda > 0$ die Funktion f_T tatsächlich eine Dichte ist, d. h., dass $\int_{-\infty}^{\infty} f_T(t) \, dt = 1$ gilt. Beweis!
2. Berechnen Sie $\Pr[T \leq \frac{2}{\lambda}]$.
3. Berechnen Sie $\mathbb{E}[T]$ in Abhängigkeit von λ .

Lösung

1. Wir wenden partielle Integration an wie folgt.

$$\begin{aligned}
\int_{-\infty}^{\infty} f_T(t)\,\mathrm{d}t &= a\lambda^2 \int_0^{\infty} te^{-\lambda t}\,\mathrm{d}t \\
&= a\lambda^2 \cdot \left(\left[-\frac{1}{\lambda} te^{-\lambda t} \right]_0^{\infty} - \int_0^{\infty} -\frac{1}{\lambda} e^{-\lambda t}\,\mathrm{d}t \right) \\
&= a\lambda^2 \cdot \left(0 + \left[-\frac{1}{\lambda^2} e^{-\lambda t} \right]_0^{\infty} \right) \\
&= a\,.
\end{aligned}$$

   Daraus folgt $a = 1$. (2P)

2. Wir wenden partielle Integration an wie folgt.

$$\begin{aligned}
\Pr\left[T \leq \frac{2}{\lambda}\right] &= \int_0^{2/\lambda} f_T(t)\,\mathrm{d}t \\
&= \lambda^2 \int_0^{2/\lambda} te^{-\lambda t}\,\mathrm{d}t \\
&= \lambda^2 \cdot \left(\left[-\frac{1}{\lambda} te^{-\lambda t} \right]_0^{2/\lambda} + \left[-\frac{1}{\lambda^2} e^{-\lambda t} \right]_0^{2/\lambda} \right) \\
&= \lambda^2 \cdot \left(-\frac{1}{\lambda} \cdot \frac{2}{\lambda} e^{-2} - \frac{1}{\lambda^2} e^{-2} + \frac{1}{\lambda^2} \right) \\
&\approx 0{,}59399415\,.
\end{aligned}$$

   (2P)

3. Teilweise wie vorausgehend rechnen wir

$$\begin{aligned}
\mathbb{E}[T] &= \int_0^{\infty} t \cdot f_T(t)\,\mathrm{d}t \\
&= \int_0^{\infty} \lambda^2 t^2 e^{-\lambda t}\,\mathrm{d}t \\
&= \left[-\lambda t^2 e^{-\lambda t} \right]_0^{\infty} - \int_0^{\infty} -2\lambda te^{-\lambda t}\,\mathrm{d}t \\
&= 2\lambda \int_0^{\infty} te^{-\lambda t}\,\mathrm{d}t \\
&= 2\lambda \cdot \frac{1}{\lambda^2} \quad \text{(siehe oben)} \\
&= \frac{2}{\lambda}\,.
\end{aligned}$$

   (1P)

Hausaufgabe 3 (5 Punkte)

Wahr oder falsch? Begründen Sie Ihre Antwort!

1. Falls $X \sim \mathcal{N}(0,1)$ und $Y \sim \mathcal{N}(0,2)$ normalverteilt sind, dann folgt $\mathrm{Var}[X+Y] = 3$.
2. Seien X und Y standardnormalverteilt. Dann gilt $\Pr[X \leq 0] = \Pr[Y \geq 0]$.
3. Bei kontinuierlichen Zufallsvariablen existiert stets der Erwartungswert.
4. Seien $X \sim \mathrm{Po}(1)$ und $Y \sim \mathrm{Bin}(n, \frac{1}{n})$. Dann gilt $|\Pr[X=2n] - \Pr[Y=2n]| < 2^{-n}$.
5. Sei X exponentialverteilt. Dann gilt $\Pr[X > 2 \mid X > 1] + \Pr[X \leq 1] = 1$.

Lösung

Für die richtige Antwort und für die richtige Begründung gibt es jeweils einen $\frac{1}{2}$ Punkt.

1. Falsch! Begründung: Das ist allgemein nur richtig, wenn X und Y unabhängig sind.
2. Wahr! Es gilt $\Pr[X \leq 0] = \frac{1}{2} = \Pr[Y \leq 0] = 1 - \Pr[Y \leq 0] = \Pr[Y \geq 0]$.
3. Falsch! Begründung: Durch Vertauschung von Abschnitten der Dichtefunktion können den kleinen Dichtewerten beliebig hohe Werte der Variablen zugeordnet werden.
4. Wahr! Begründung: $\Pr[Y = 2n] = 0$ und $\Pr[X = 2n] = \frac{e^{-1}}{(2n)!} < 2^{-n}$.
5. Wahr! Begründung: Da die Exponentialverteilung gedächtnislos ist, gilt $\Pr[X > 2 \mid X > 1] = \Pr[X > 1] = 1 - \Pr[X \leq 1]$.

Hausaufgabe 4 (5 Punkte)

Sei $a > 0$, und seien X, Y kontinuierliche Zufallsvariablen mit gemeinsamer Dichtefunktion

$$f_{X,Y}(x,y) = \begin{cases} a\cdot(1 - x\cdot y) & : \quad 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & : \quad \text{sonst} \end{cases}$$

1. Berechnen Sie die Randdichten $f_X(x)$ und $f_Y(y)$.
2. Bestimmen Sie a.
3. Sind die Variablen X und Y unabhängig? Beweisen Sie Ihre Antwort.

Lösung

1. $f_X(x) = a \cdot (1 - \frac{x}{2})$. Berechnung:

$$\begin{aligned} f_X(x) &= \int_0^1 a\cdot(1 - x\cdot y)\,\mathrm{d}y \\ &= a \cdot \left[y - \frac{xy^2}{2} \right]_{y=0}^{y=1} = a \cdot (1 - \frac{x}{2}). \end{aligned}$$

   Entsprechend gilt $f_Y(y) = a \cdot (1 - \frac{y}{2})$. (1P)

2. Aus der Form des Gebiets, in dem die Dichte ungleich Null ist, ergibt sich die Gleichung

$$
\begin{aligned}
1 \;=\; F_{X,Y}(1,1) \;&=\; \int_0^1 f_X(x)\,\mathrm{d}x \;=\; \int_0^1 a\cdot(1-\frac{x}{2})\,\mathrm{d}x \\
&=\; a\cdot\left[x-\frac{x^2}{4}\right]_{y=0}^{y=1} \;=\; \frac{3}{4}\cdot a
\end{aligned}
$$

Mithin $a=\frac{4}{3}$. (2P)

3. Nein! I.A. gilt

$f_{X,Y}(x,y)=a(1-xy)\neq a(1-\frac{x}{2})\cdot a(1-\frac{y}{2})=f_X(x)\cdot f_Y(y)$. (2P)

Hinweis: *Die Vorbereitungsaufgaben bereiten die Tutoraufgaben vor und werden in der Zentralübung unterstützt. Tutoraufgaben werden in den Übungsgruppen bearbeitet. Hausaufgaben sollen selbstständig bearbeitet und zur Korrektur und Bewertung abgegeben werden.*

Vorbereitung 1

Seien $X_1, X_2, \ldots, X_{100}$ unabhängige Bernoulli-verteilte Zufallsvariable mit gleicher Erfolgswahrscheinlichkeit $\frac{1}{50}$. Sei $S_{100}=\sum_{i=1}^{100} X_i$.

1. Berechnen Sie Zahlen $a, b\in\mathbb{R}$ mit $a>0$, so dass $\mathbb{E}[Y]=0$ und $\mathrm{Var}[Y]=1$ für $Y=a\cdot S_{100}+b$ gelten.

2. Wenden Sie den zentralen Grenzwertsatz an zur approximativen Berechnung eines Intervalls $[d_1, d_2]$, so dass

$$\Pr[d_1 \le S_{100} \le d_2] \;\approx\; 1-\alpha = 1-0.05\,.$$

Benutzen Sie dabei das Quantil $c=x_{1-\frac{\alpha}{2}}\approx 1.96$ der Standardnormalverteilung.

Lösung

1. $\mathbb{E}[S_{100}]=100\cdot\frac{1}{50}=2$.
   $\mathrm{Var}[S_{100}]=2\cdot(1-\frac{1}{50})=\frac{49}{25}$.
   Aus $\mathbb{E}[Y]=a\cdot\mathbb{E}[S_{100}]+b$ folgt $0=2a+b$.
   Aus $\mathrm{Var}[Y]=a^2\mathrm{Var}[S_{100}]$ folgt $1=a^2\cdot\frac{49}{25}$.
   Mithin $a=\frac{5}{7}$ und $b=-\frac{10}{7}$.

2. Ansatz: $\Pr[d_1\le S_{100}\le d_2]=\Pr[-c\le Y\le c]=1-\alpha$.
   Mit $Y=a\cdot S_{100}+b$ ergibt sich $d_1=\frac{-c-b}{a}=\ldots$ und $d_2=\frac{c-b}{a}=\ldots$.

Vorbereitung 2

Bei einem Einwahlserver für $n = 10^3$ Teilnehmer nehmen wir an, dass zu einem festen Zeitpunkt jeder Teilnehmer mit Wahrscheinlichkeit $p = 0,05$ Zugriff auf den Server wünscht.

Berechnen Sie eine Näherung der Wahrscheinlichkeit, mit der gleichzeitig mehr als 55 Verbindungswünsche auftreten? Approximieren Sie dabei die Binomialverteilung durch die entsprechende Normalverteilung und benutzen Sie ggf. geeignete Tabellen für die Werte der Standardnormalverteilung.

Lösung

Sei X die Anzahl der Verbindungswünsche. X ist näherungsweise normalverteilt. Wir nehmen also an $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu = np = 50$ und $\sigma^2 = np(1-p) = 47,5$. Damit folgt

$$\begin{aligned}
\Pr[X > 55] &= 1 - \Pr[X \leq 55] \\
&= 1 - \Pr\left[\frac{X-\mu}{\sigma} \leq \frac{55-50}{\sqrt{47,5}}\right] \\
&= 1 - \Phi\left(\frac{55-50}{\sqrt{47,5}}\right) \\
&\approx 0,2343\,.
\end{aligned}$$

Vorbereitung 3

Sei X eine binomialverteilte Zufallsvariable mit $X \sim \mathrm{Bin}(2000, 0.05)$. Wir nehmen an, dass die Voraussetzungen sowohl für eine Approximation von entsprechenden Verteilungen mit Poisson-Verteilung als auch mit Normalverteilung vorliegen.
Berechnen Sie approximativ

1. $\Pr[X = 110]\,,$
2. $\Pr[X > 110]\,.$

Begründen Sie jeweils die Wahl einer der Approximationen.

Lösung

1. Für die Wahrscheinlichkeit des Auftretens eines bestimmten Wertes einer diskreten Zufallsvariablen, wie hier $X = 110$, ist es naheliegend die approximierende diskrete Verteilung zu verwenden. In diesem Fall wählen wir deshalb die Poisson-Verteilung.

   Es gilt zunächst $E[X] = 2000 \cdot 0.05 = 100$. Mit $\lambda = n \cdot p_n = 100$ gilt

   $$\lim_{n\to\infty} b(k; n, p_n) = \lim_{n\to\infty} \binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

   und wir erhalten mit $k = 110$

   $$\Pr[X = 110] \;=\; b(110; 2000, 0.05) \;\approx\; e^{-100} \cdot \frac{100^{110}}{110!}\;.$$

   Die Berechnung erfordert nun die Anwendung der Stirling-Formel auf die Fakultätsfunktion.

   $$k! \;\approx\; \sqrt{2\pi k} \cdot \left(\frac{k}{e}\right)^k\;.$$

Damit rechnen wir

$$
\begin{aligned}
e^{-100} \cdot \frac{100^{110}}{110!} &= e^{-100} \cdot \frac{100^{110}}{\sqrt{2\pi \cdot 110}} \cdot \left(\frac{e}{110}\right)^{110} \\
&= \frac{e^{10}}{\sqrt{2\pi \cdot 110}} \cdot \left(\frac{100}{110}\right)^{110} \\
&\approx 0.0234\ldots
\end{aligned}
$$

Ergebnis: $\Pr[X=110] \approx 0.0234$.

Bemerkung: In praktischen Anwendungen müsste nun eine Abschätzung der Genauigkeit der Berechnungen folgen.

2. Es giltt $\Pr[X > 110] = 1 - \Pr[X \leq 110]$. D.h., wir müssen lediglich den Wert der Verteilungsfunktion der binomialen Dichtefunktion an der Stelle 110 berechnen.

   Allerdings haben wir die Verteilungsfunktion der Poisson-Dichte noch nicht zur Verfügung. Schon aus diesem Grund bietet sich die Normalverteilung für eine Approximation an. In diesem Fall steht uns die Verteilungsfunktion Φ für die Standardnormalverteilung zur Verfügung.

   Wir berechnen die Wahrscheinlichkeit $\Pr[X \leq 110]$ mit $\mathbb{E}[X] = 2000 \cdot 0.05 = 100$ und $\sigma = \sqrt{np(1-p)} = \sqrt{100 \cdot 0.95} = \sqrt{95}$ wie folgt.

$$
\begin{aligned}
\Pr[X \leq 110] &= F_X(110) \\
&\approx \Phi\left(\frac{k - np}{\sqrt{np(1-p)}}\right) \\
&= \Phi\left(\frac{110 - 100}{\sqrt{95}}\right) \\
&\approx \Phi(1.025978) \approx 0.847
\end{aligned}
$$

   Ergebnis: $\Pr[X > 110] \approx 1 - 0.847 = 0.153$.

Vorbereitung 4

Eine Werbeagentur möchte am letzten Tag der Fußballweltmeisterschaft mit einer Blitzumfrage schätzen, welcher Anteil ϑ der per Bahn anreisenden Fußballfans einen Platz im Stadion hat. Jeder der 12 Mitarbeiter befragt so lange zufällig ausgewählte Fans, bis er einen Fan gefunden hat, der eine Karte für das Stadion besitzt. Die Anzahl der vom Mitarbeiter i befragten Fans sei X_i.

Wir nehmen an, dass alle X_i die gleiche geometrische Verteilung besitzen mit

$$
\Pr_\vartheta[X_i = k] = (1-\vartheta)^{k-1} \cdot \vartheta\,, \ \forall k \in \mathbb{N}\,.
$$

1. Man bestimme auf der Basis der ermittelten Stichprobenwerte

$$
3,\ 1,\ 2,\ 3,\ 2,\ 1,\ 2,\ 1,\ 5,\ 3,\ 2,\ 3
$$

   einen Maximum-Likelihood-Schätzwert für ϑ.

2. Man gebe mit Hilfe des Zentralen Grenzwertsatzes ein konkretes 95%-Konfidenzintervall für ϑ an.

Lösung

1. Für die Likelihood-Funktion $L(\vec{x};\vartheta)$ der Stichprobe $\vec{x} = (x_1, \ldots, x_{12})$ gilt mit $n = 12$

$$\begin{aligned} L(\vec{x};\vartheta) &= \prod_{i=1}^{n} \Pr_\vartheta[X_i = x_i] \\ &= \prod_{i=1}^{n} \left[(1-\vartheta)^{x_i-1} \cdot \vartheta\right]. \end{aligned}$$

Wir schreiben $\sum_{i=1}^{n} x_i - n = n\bar{x} - n$ und erhalten

$$L(\vec{x};\vartheta) = (1-\vartheta)^{n\bar{x}-n} \cdot \vartheta^n .$$

Gesucht ist der Wert $\vartheta = \bar{\vartheta}$ mit $0 \le \vartheta \le 1$, für den $L(\vec{x};\vartheta)$ das Maximum in dem $[0,1]$-Intervall annimmt.

Falls $\bar{x} = 1$, dann folgt sofort $\vartheta = 1$. Dies bedeutet, dass mit maximaler Wahrscheinlichkeit (die der Test erlaubt) jeder Fan eine Stadionkarte besitzt.

Falls $\bar{x} \neq 1$, dann bestimmen wir das Maximum von ϑ zwischen 0 und 1 im ersten Schritt durch Bestimmung einer Nullstelle der Ableitung von L nach ϑ wie folgt.

$$\begin{aligned} \frac{\mathrm{d}\, L(\vec{x};\vartheta)}{\mathrm{d}\,\vartheta} &= (n\bar{x}-n)(1-\vartheta)^{n\bar{x}-n-1}(-1)\vartheta^n + (1-\vartheta)^{n\bar{x}-n} n\vartheta^{n-1} \\ &= (1-\vartheta)^{n\bar{x}-n-1}\vartheta^{n-1}[(n\bar{x}-n)(-\vartheta) + n(1-\vartheta)] \\ &= (1-\vartheta)^{n\bar{x}-n-1}\vartheta^{n-1}[n - n\bar{x}\vartheta] \end{aligned}$$

Als Nullstellen der Ableitung der Likelihood-Funktion erhalten wir $\vartheta = 0$, $\vartheta = 1$ und $\vartheta = \frac{1}{\bar{x}}$. Man beachte, dass $0 < \frac{1}{\bar{x}} < 1$ gilt. Da

$$L(\vec{x};0) = L(\vec{x};1) = 0 \quad \text{und} \quad L\left(\vec{x};\frac{1}{\bar{x}}\right) > 0$$

gilt, scheiden $\vartheta = 0$ und $\vartheta = 1$ als Maximumstellen aus. Es folgt

$$\bar{\vartheta} = \frac{1}{\bar{x}} .$$

Nun berechnen wir $\frac{1}{\bar{x}}$ aus der Stichprobe mit $\bar{x} = \frac{7}{3}$ und erhalten

$$\bar{\vartheta} = \frac{3}{7} .$$

Es haben also bei maximaler Testwahrscheinlichkeit $\frac{3}{7}$ der Fans Stadionkarten.

2. Die Stichprobe liefert den Wert $\bar{x} = \frac{7}{3}$ für die Zufallsvariable $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ mit $n = 12$. Für das Konfidenzintervall für ϑ zum Konfidenzniveau $1 - \alpha = 0.95$ setzen wir an

$$\Pr\left[-c \le \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \le c\right] \approx 0.95$$

mit dem Quantil $c = x_{1-\frac{\alpha}{2}} = x_{0.975} \approx 1.96$ und den Gleichungen $\mu_{\bar{X}} = \frac{1}{\vartheta}$ und $\sigma^2_{\bar{X}} = \frac{1-\vartheta}{n\vartheta^2}$ für $n = 12$. Wir bestimmen nun die Menge aller ϑ, für die die Ungleichung $-c \leq \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} \leq c$ gilt.

$$
\begin{aligned}
-c \leq \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \leq c \iff & (\bar{X} - \mu_{\bar{X}})^2 \leq c^2\sigma^2_{\bar{X}} \\
\iff & (\bar{X} - \frac{1}{\vartheta})^2 \leq c^2 \cdot \frac{1}{n} \cdot \frac{1-\vartheta}{\vartheta^2} \\
\iff & (\vartheta\bar{X} - 1)^2 \leq \frac{c^2}{n} \cdot (1-\vartheta) \\
\iff & \vartheta^2\bar{X}^2 + (\frac{c^2}{n} - 2\bar{X})\vartheta + (1 - \frac{c^2}{n}) \leq 0
\end{aligned}
$$

Nun setzen wir sämtliche Zahlenwerte ein und erhalten (wegen $c \approx 1.96$ näherungsweise)

$$
-c \leq \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \leq c \iff 49\vartheta^2 - 39.1188\vartheta + 6.1188 \leq 0
$$

Wenn wir nach bekannter Formel die Nullstellen der quadratischen Gleichung

$$
49\vartheta^2 - 39.1188\vartheta + 6.1188 = 0
$$

mit α_1, α_2 bezeichnen, wobei wir $\alpha_1 < \alpha_2$ annehmen können, dann ergibt sich als Lösung

$$
\alpha_1 \leq \vartheta \leq \alpha_2
$$

mit $\alpha_1 \approx 0.2135$ und $\alpha_2 \approx 0.5848$.

Tutoraufgabe 1

Sei $X = \sum_{i=1}^{2000} X_i$ die Summe der Augenzahlen, wenn man 2000-mal mit einem idealen Würfel würfelt.

1. Berechnen Sie näherungsweise $\Pr[7000 \leq X \leq 7100]$!
2. Wie groß muss man Δ wählen, damit $\Pr[7000 - \Delta \leq X \leq 7000 + \Delta] \approx \frac{1}{2}$ gilt?

Approximieren Sie die entsprechenden Wahrscheinlichkeiten mit Hilfe des Zentralen Grenzwertsatzes.

Lösung

1. Zunächst gilt für jedes i

$$
\mu = \mathbb{E}[X] = 2000 \cdot \mathbb{E}[X_i] = 2000 \cdot \frac{7}{2} = 7000
$$

und

$$
\mathrm{Var}[X] = 2000 \cdot \mathrm{Var}[X_i] = 2000 \cdot \frac{35}{12} = \frac{17500}{3} .
$$

Mit

$$\sigma = \sqrt{\frac{17500}{3}} = 50 \cdot \sqrt{\frac{7}{3}}\,.$$

folgt

$$\begin{aligned}
\Pr[7000 \le X \le 7100] &= \Pr\left[\frac{7000-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{7100-\mu}{\sigma}\right] \\
&= \Pr\left[0 \le \frac{X-\mu}{\sigma} \le \sqrt{\frac{12}{7}}\right] \\
&\approx \Phi(1,309307341...) - \Phi(0) \\
&\approx 0,90478... - 0,5 \\
&= 0,40478... \quad .
\end{aligned}$$

Man beachte, dass $\frac{X-\mu}{\sigma}$ als standardnormalverteilt betrachtet und der Funktionswert $\Phi(1,309307341...)$ durch Interpolation aus den Tabellenwerten approximiert wurde. Außerdem wurde $0 < \frac{X-\mu}{\sigma}$ durch $0 \le \frac{X-\mu}{\sigma}$ ersetzt, weil Φ stetig ist.

2. Man beachte im Folgenden u. a. die Verwendung der Gleichung $\Phi(-z) = 1 - \Phi(z)$. Außerdem benützen wir wieder die Stetigkeit von Φ.

$$\begin{aligned}
\Pr[7000 - \Delta \le X \le 7000 + \Delta] &= \Pr\left[\frac{-\Delta}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{\Delta}{\sigma}\right] \\
&\approx \Phi\left(\frac{\Delta}{\sigma}\right) - \Phi\left(-\frac{\Delta}{\sigma}\right) \\
&= 2 \cdot \Phi\left(\frac{\Delta}{\sigma}\right) - 1\,.
\end{aligned}$$

Wir setzen nun $2 \cdot \Phi\left(\frac{\Delta}{\sigma}\right) - 1 = \frac{1}{2}$. Das bedeutet

$$\Phi\left(\frac{\Delta}{\sigma}\right) = 0,75\,.$$

Mit dem Quantil

$$z_{0,75} \approx 0,6745$$

erhalten wir

$$\Delta = z_{0,75} \cdot \sigma \approx 0,6745 \cdot 50 \cdot \sqrt{\frac{7}{3}} = 51,51788... \quad .$$

Bei der Bestimmung des Quantils aus der Tabelle von Φ haben wir eine Interpolation angewendet.

Tutoraufgabe 2

Ziel dieser Aufgabe ist es, einen ML-Schätzer für die Erfolgswahrscheinlichkeit $p \in (0,1)$ einer negativ binomialverteilten Zufallsvariablen zu bestimmen. Hierfür seien $X_1, \ldots, X_n$ unabhängige Zufallsvariable, wobei jedes X_i negativ binomialverteilt mit Erfolgswahrscheinlichkeit p bei m zu erzielenden Erfolgen sei, d. h., jedes X_i hat die Dichte

$$f_{X_i}(k) = \binom{k-1}{m-1} p^m (1-p)^{k-m} \quad (\text{mit } k \ge m)\,.$$

Der Parameter m sei bekannt. Zu schätzen ist p.

1. Es sei $\vec{k} = (k_1, \ldots, k_n)$ ein Stichprobenvector (mit $k \geq m$). Stellen Sie die Likelihood-Funktion $L(\vec{k}; p)$ auf.

2. Maximieren Sie $L(\vec{k}; p)$ und bestimmen Sie den entsprechenden ML-Schätzer für p.

3. Zeigen Sie, dass der hergeleitete ML-Schätzer i. A. nicht erwartungstreu ist.

   Hinweis: Verwenden Sie $\ln(1+x) = \sum_{i=1}^{\infty}(-1)^{i+1}\frac{x^i}{i}$ für $x \in (-1, 1]$.

Lösung

1.

$$
\begin{aligned}
L(\vec{k}; p) &= \prod_{i=1}^{n} f_{X_i}(k_i) \\
&= \prod_{i=1}^{n} \left(\binom{k_i - 1}{m - 1} p^m (1-p)^{k_i - m} \right) \\
&= \left(\frac{p}{1-p} \right)^{mn} \cdot (1-p)^{\overbrace{\sum_{i=1}^{n} k_i}^{=:K}} \cdot \underbrace{\prod_{i=1}^{n} \binom{k_i - 1}{m - 1}}_{=:C} \\
&= p^{mn} \cdot (1-p)^{K - mn} \cdot C \ .
\end{aligned}
$$

2. Wir bestimmen das Maximum von L analog wie in der VA 1 durch Bildung der Ableitung und eine Begründung, dass die 2. Ableitung an der gefundenen Stelle negativ sein muss.

$$
\begin{aligned}
\frac{\mathrm{d}\, L(\vec{k}; p)}{\mathrm{d}\, p} &= Cmn \cdot p^{mn-1} \cdot (1-p)^{K-mn} - Cp^{mn} \cdot (K - mn) \cdot (1-p)^{K-mn-1} \\
&= Cp^{mn-1} \cdot (1-p)^{K-mn-1} \cdot [mn(1-p) - p(K - mn)] \\
&= Cp^{mn-1} \cdot (1-p)^{K-mn-1} \cdot [mn - pK] \, .
\end{aligned}
$$

Die Nullstellen sind $p = 0$, $p = 1$ und $p = \frac{mn}{K}$, wobei die mittlere Nullstelle das Maximum von L liefert. Der gesuchte Schätzer für die Erfolgswahrscheinlichkeit p ist also

$$
U = \frac{mn}{\sum_{i=1}^{n} X_i} \, .
$$

3. U ist erwartungstreu, falls $E[U] = p$ gilt. U ist aber nicht einmal für $m = n = 1$ erwartungstreu, wie die folgende Rechnung zeigt.

   Seien $m = n = 1$. Dann gilt $U = \frac{1}{X_1}$ und X_1 ist geometrisch verteilt:

$$
\mathbb{E}[U] = \mathbb{E}\left[\frac{1}{X_1}\right] = \sum_{i=1}^{\infty} \frac{1}{i} \cdot p \cdot (1-p)^{i-1} = \frac{p}{1-p} \sum_{i=1}^{\infty} \frac{(1-p)^i}{i} \, .
$$

Wir berechnen die unendliche Reihe (siehe Hinweis):

$$-\ln p = -\ln(1+(p-1)) = -\sum_{i=1}^{\infty}(-1)^{i+1}\frac{(p-1)^i}{i} = \sum_{i=1}^{\infty}\frac{(1-p)^i}{i}\,.$$

Damit folgt

$$\mathbb{E}[U] = p\cdot\frac{-\ln p}{1-p} \neq p\,.$$

Tutoraufgabe 3

Beim Testen von Hypothesen bezeichnet man die zu überprüfende Hypothese (Nullhypothese) generell mit H_0 und die Alternative mit H_1.

Ein Tierhändler erhält ein Paket mit 100 Frettchen. Er will testen, ob weniger als zehn (<10) dieser Frettchen aggressiv und bissig sind. Dazu hält er zehn Frettchen seinen Finger hin und nimmt das Paket nur an, wenn ihn keines davon beißt (wir nehmen an, dass ein aggressives Frettchen sofort zubeißen würde).

Wie lauten die Hypothesen des Händlers? Was ist das Signifikanzniveau des Tests?

Lösung

Die Zahl X der aggressiven Frettchen in der Stichprobe vom Umfang 10 ist hypergeometrisch verteilt mit $N = 100$ und $S=$ Anzahl der aggressiven Frettchen im Paket. Der Händler testet die Hypothese

$$H_0 : S \leq 9$$

gegen

$$H_1 : S \geq 10$$

und verwirft die Nullhypothese dabei im Fall $X \geq 1$. Für $S = 9$ ist

$$\Pr_S[X > 0] = 1 - \Pr_S[X = 0] = 1 - \frac{\binom{9}{0}\binom{91}{10}}{\binom{100}{10}} = 62.9\%.$$

Das ist auch das effektive Niveau des Tests, denn für $S < 9$ ist die Verwerfungswahrscheinlichkeit kleiner.