
Diskrete Wahrscheinlichkeitstheorie

Hausaufgabe 1 (5 Punkte)

Sei X eine Zufallsvariable mit unbekanntem Erwartungswert μ und einer Standardabweichung $\sigma_X = 2$. Wir verwenden $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ als Schätzvariable für μ , wobei die X_i unabhängige Wiederholungen von X seien. Wir setzen voraus, dass \bar{X} für $n \geq 1500$ normalverteilt ist.

1. Leiten Sie eine möglichst kleine untere Schranke n_0 für n her, so dass für alle $n \geq n_0$ gilt

$$\Pr[|\bar{X} - \mu| < 0,1] \geq 0,9.$$

Für das 0,95-Quantil der Standardnormalverteilung ist dabei $z_{0,95} \approx 1,65$ zu verwenden.

2. Nun sei $n = 2500$. Geben Sie ein Konfidenzintervall für μ zum Konfidenzniveau 0,9 an!

Lösung

1. Es gilt

$$\begin{aligned} \text{Var}[\bar{X}] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] \\ &= \frac{1}{n^2} \cdot n \cdot \sigma_X^2 = \frac{4}{n}. \end{aligned}$$

Wir setzen $\bar{\sigma}^2 = \text{Var}[\bar{X}]$. Dann gilt $\bar{\sigma} = \frac{2}{\sqrt{n}}$.

$$\begin{aligned} \Pr[|\bar{X} - \mu| < 0,1] &= \Pr[-0,1 < \bar{X} - \mu < 0,1] \\ &\approx \Phi\left(\frac{0,1}{\bar{\sigma}}\right) - \Phi\left(-\frac{0,1}{\bar{\sigma}}\right) \\ &= 2 \cdot \Phi\left(\frac{0,1}{\bar{\sigma}}\right) - 1. \end{aligned}$$

Es gilt $2 \cdot \Phi\left(\frac{0,1}{\bar{\sigma}}\right) - 1 \geq 0,9$ genau dann, wenn $\Phi\left(\frac{0,1}{\bar{\sigma}}\right) \geq 0,95$.

Mit dem Quantil $z_{0,95} \approx 1,65$ erhalten wir für n die Ungleichung

$$\frac{0,1}{\bar{\sigma}} = \frac{\sqrt{n}}{20} \geq 1,65,$$

d. h.
$$n \geq 20^2 \cdot 1,65^2 = 1089.$$

Da die Normalverteilung $n \geq 1500$ voraussetzt, folgt

$$n \geq \max\{1500, 1089\} = 1500 =: n_0. \quad (3P)$$

2. Für $n = 2500$ und einen zu bestimmenden Parameter c betrachten wir $\Pr[|\bar{X} - \mu| < c]$ und erhalten die Gleichung

$$\frac{c}{\sigma} = \frac{c\sqrt{2500}}{2} = 1,65$$

mit der Lösung $c = \frac{3,3}{50} = 0,066$.

Das gesuchte Konfidenzintervall für μ ist demnach

$$\bar{X} - 0,066 < \mu < \bar{X} + 0,066. \quad (2P)$$

Hausaufgabe 2 (5 Punkte)

Zeigen oder widerlegen Sie:

1. Wenn zwei unabhängige Zufallsvariablen X und Y normalverteilt sind, dann ist auch ihre Differenz $X - Y$ normalverteilt.
2. Jede erwartungstreue Schätzvariable für einen Parameter δ schätzt den Erwartungswert von δ .
3. Die Anwendung des Maximum-Likelihood-Prinzips setzt die Verfügbarkeit einer Stichprobe voraus.
4. Bei echten Alternativtests ist die Summe der Wahrscheinlichkeit eines Fehlers 1. Art und der Wahrscheinlichkeit eines Fehlers 2. Art stets gleich 1.
5. Wir betrachten eine erwartungstreue, diskrete Schätzvariable X für einen Parameter α . Für die Verteilungsdichte f_X gelte $f_X(i) = \frac{3^i e^{-3}}{i!}$ für $i \in \mathbb{N}_0$. Dann ist 3 der durch X für α geschätzte Wert.

Lösung

Pro Teilaufgabe 1P.

1. Wahr! Beweis: Linearkombinationen von unabhängigen und normalverteilten Zufallsvariablen sind wieder normalverteilt (Satz der Vorlesung).
2. Falsch! Da δ keine Zufallsvariable ist, kann auch der Erwartungswert von δ nicht definiert werden.
3. Wahr! Es wird die Wahrscheinlichkeit $L(\vec{x}; p)$ maximiert, wobei \vec{x} eine Stichprobe ist.
4. Falsch! Echte Alternativtests werden mit dem Ziel konstruiert, beiden Fehlerarten eine kleine Wahrscheinlichkeit zu geben.
5. Wahr! Beweis: X ist Poisson-verteilt mit $\lambda = 3$. Damit gilt $\mathbb{E}[X] = 3$. Da X den Wert α erwartungstreu abschätzt gilt $\alpha = \mathbb{E}[X] = 3$.

Hausaufgabe 3 (5 Punkte)

Seien $\lambda > 0$ und X_1, X_2, \dots, X_n unabhängige und identisch verteilte kontinuierliche Zufallsvariable jeweils mit der gleichen Dichte

$$f(t) := \begin{cases} \frac{1}{2} \lambda^3 t^2 e^{-\lambda t} & : t \geq 0, \\ 0 & : \text{sonst.} \end{cases}$$

1. Sei $\vec{k} = (k_1, \dots, k_n)$ ein Stichprobenvektor. Stellen Sie die Likelihood-Funktion $L(\vec{k}; \gamma)$ für den Parameter $\gamma = \frac{1}{\lambda}$ auf.
2. Bestimmen Sie den ML-Schätzer für γ durch Maximieren von $L(\vec{k}; \gamma)$.

Lösung

1.

$$\begin{aligned} L(\vec{k}; \gamma) &= \prod_{i=1}^n \left(\frac{1}{2\gamma^3} k_i^2 e^{-\frac{1}{\gamma} k_i} \right) \\ &= \frac{1}{2^n} \frac{1}{\gamma^{3n}} \left(\prod_{i=1}^n k_i^2 \right) e^{-\frac{1}{\gamma} \sum_{i=1}^n k_i}. \end{aligned}$$

Zwecks besserer Lesbarkeit kann man Bezeichnungen $C := \frac{1}{2^n} \left(\prod_{i=1}^n k_i^2 \right)$ und $K := \sum_{i=1}^n k_i$ einführen und kann dann wie folgt schreiben.

$$L(\vec{k}; \gamma) = \frac{C}{\gamma^{3n}} e^{-\frac{K}{\gamma}}. \quad (2P)$$

2. Wir bestimmen bei gegebener Stichprobe \vec{k} das Maximum der Funktion $L(\vec{k}; \gamma)$ von γ .

Nullstellen der Ableitung von L nach γ :

$$\frac{d}{dx} L(\vec{k}; \gamma) = -C \cdot 3n \gamma^{-3n-1} e^{-\frac{K}{\gamma}} + C \gamma^{-3n} K \gamma^{-2} e^{-\frac{K}{\gamma}}.$$

Es gilt $\frac{d}{dx} L(\vec{k}; \gamma) = 0$ genau dann, wenn

$$\gamma = \frac{K}{3n} = \frac{\sum_{i=1}^n k_i}{3n}.$$

Damit gibt es genau eine Nullstelle. Da $L(\vec{k}; \gamma) > 0$ für alle γ und $L(\vec{k}; \gamma) \rightarrow 0$ sowohl für $\gamma \rightarrow 0$ als auch für $\gamma \rightarrow +\infty$ gilt, folgt, dass die Nullstelle ein Maximum ist.

Damit ist

$$U = \frac{\sum_{i=1}^n k_i}{3n}$$

der gesuchte ML Schätzer für γ . (3P)

Hausaufgabe 4 (5 Punkte)

Seien X_1, \dots, X_n unabhängige Stichproben einer Zufallsvariablen X und sei \bar{X} das arithmetische Mittel der X_i . Wir verwenden die Zufallsvariable

$$V = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

als Schätzer für die Varianz von X .

Berechnen Sie den Bias von V ! Welche Aussage gilt für $n \rightarrow \infty$?

Lösung

Sei $S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Dann folgt

$$V = \frac{n-1}{n} \cdot S.$$

Berechnung des Bias von V :

$$\begin{aligned} \mathbb{E}[V - \text{Var}[X]] &= \mathbb{E}[V] - \text{Var}[X] \\ &= \frac{n-1}{n} \cdot \mathbb{E}[S] - \text{Var}[X] \\ &= \frac{n-1}{n} \cdot \text{Var}[X] - \text{Var}[X] \\ &= -\frac{\text{Var}[X]}{n}. \end{aligned}$$

Der Bias strebt gegen 0 für $n \rightarrow \infty$. (5P)

Zusatzaufgabe 4 (Wird nicht korrigiert)

Seien Φ bzw. Θ unabhängige ZV mit Φ gleichverteilt auf $[-\pi, \pi)$ bzw. $[0, 1]$. Dann ist durch

$$G(\Phi, \Theta) := \{(x, y) \in \mathbb{R}^2; x \cos \Phi + y \sin \Phi = \Theta\}$$

eine zufällige Gerade im \mathbb{R}^2 beschrieben, wobei Θ den Abstand der Geraden vom Ursprung angibt.

1. Für ein festes $r \in [0, \infty)$ sei $K_r = \{(x, y) \in \mathbb{R}^2; \sqrt{x^2 + y^2} = r\}$. Bestimmen Sie die Wahrscheinlichkeit

$$\Pr[G(\Phi, \Theta) \cap K_r \neq \emptyset].$$

2. Für ein festes $\rho \in [0, \pi]$ sei $L_\rho = \{(\cos \alpha, \sin \alpha) \in \mathbb{R}^2; \alpha \in [-\rho/2, \rho/2]\}$. Bestimmen Sie wiederum die Wahrscheinlichkeit

$$\Pr[G(\Phi, \Theta) \cap L_\rho \neq \emptyset].$$

Lösung

Wenn Φ und Θ unabhängige Zufallsvariable sind, dann müssen beide Abbildungen auf einer gemeinsamen Ergebnismenge Ω eines Wahrscheinlichkeitsraums $(\Omega, \mathcal{A}, \Pr)$ definiert sein, mithin die Funktionalität $\Omega \rightarrow \mathbb{R}$ haben.

Für Ω kommt $[-\pi, \pi) \times [0, 1]$ in Frage mit $\Phi((x, y)) = x$ und $\Theta((x, y)) = y$.

\Pr kann man aufgrund der genannten Unabhängigkeit aus den gegebenen Randdichten von Φ bzw. Θ herleiten.

Das führt zur Gleichverteilung auf der Fläche Ω , wobei $\Pr[A]$ durch das Flächenverhältnis von A zu Ω gegeben ist.

Man beachte, dass sowohl durch

$$G(\Phi, \Theta) \cap K_r \neq \emptyset$$

als auch durch

$$G(\Phi, \Theta) \cap L_\rho \neq \emptyset$$

Ereignisse für Punkte $(\phi, \vartheta) \in \Omega$ definiert werden.

1. Man bestimme zunächst die Menge $A = \{(\phi, \vartheta) ; G(\phi, \vartheta) \cap K_r \neq \emptyset\}$.

In welcher Beziehung steht die Fläche von A zu der gesuchten Wahrscheinlichkeit?

2. Der prinzipielle Zugang zur Lösung ist gleich wie vorhin wie folgt.

Wir bestimmen die Punktmenge E aller Punkte $(\phi, \vartheta) \in [-\pi, \pi) \times [0, 1]$, so dass $G(\phi, \vartheta) \cap L_\rho \neq \emptyset$.

E ist symmetrisch zur Θ -Achse : $(\phi, \vartheta) \in E \implies (-\phi, \vartheta)$.

Es reicht aus, die Menge $E^+ = E \cap [0, \pi) \times [0, 1]$ zu bestimmen.

Für $\vartheta = 1$ gilt:

$$(\phi, \vartheta) \in E^+ \iff 0 \leq \phi \leq \rho/2.$$

D.h., dass für $\vartheta = 1$ die Geraden aus $G(\phi, \vartheta)$ mit $0 \leq \phi$ das Kreissegment L_ρ gerade berühren, falls $0 \leq \phi \leq \rho/2$ gilt.

Wir erweitern die Aussage für Geraden mit kleinerem Abstand ϑ , und zwar zunächst für $\cos \frac{\rho}{2} \leq \vartheta \leq 1$ wie folgt:

$$(\phi, \vartheta) \in E^+ \iff 0 \leq \phi \leq \frac{\rho}{2} + \arccos \vartheta.$$

Also ist die obere Begrenzung von ϕ für alle ϑ mit $\cos \frac{\rho}{2} \leq \vartheta \leq 1$ durch die Funktion $r(\vartheta) = \frac{\rho}{2} + \arccos \vartheta$ gegeben

und es gilt für alle ϑ mit $\cos \frac{\rho}{2} \leq \vartheta \leq 1$ die Bestimmung

$$(\phi, \vartheta) \in E^+ \iff 0 \leq \phi \leq r(\vartheta).$$

Man kann diesen Teil von E^+ als Fläche unterhalb der Funktion $r(\vartheta)$ im Abschnitt $\vartheta \in [\cos \frac{\rho}{2}, 1]$ auffassen.

Die nächste und letzte Erweiterung besteht in der Bestimmung der Menge E^+ im Bereich $0 \leq \vartheta \leq \cos \frac{\rho}{2}$.

Man sieht, dass eine Gerade mit Abstand $\vartheta < \cos \frac{\rho}{2}$ und Drehwinkel $\phi = 0$ den Sektor L_ρ nicht mehr schneidet.

Man erhält einen Schnitt mit L_ρ bei positiver Drehung α erstmalig für $\alpha = -\frac{\rho}{2} + \arccos \vartheta$.

Für alle ϑ mit $0 \leq \vartheta \leq \cos \frac{\rho}{2}$ gilt die Bestimmung

$$(\phi, \vartheta) \in E^+ \iff -\frac{\rho}{2} + \arccos \vartheta \leq \phi \leq \frac{\rho}{2} + \arccos \vartheta.$$

Da $\vartheta \in [\cos \frac{\rho}{2}, 1]$ genau dann gilt, wenn $-\frac{\rho}{2} + \arccos \vartheta \leq 0$ gilt, können wir vereinfachen und E^+ insgesamt wie folgt charakterisieren:

$$(\phi, \vartheta) \in E^+ \iff \vartheta \in [0, 1], \quad 0 \leq \phi, \quad \text{und} \\ -\frac{\rho}{2} + \arccos \vartheta \leq \phi \leq \frac{\rho}{2} + \arccos \vartheta.$$

Diese Charakterisierung kann nun bequem zur Berechnung der Fläche von E benutzt werden

Flächenberechnung F_{E^+} :

Beobachtung: Nach dem Prinzip von Cavallieri ist die Fläche im Streifen $\vartheta \in [0, 1]$ zwischen den beiden Randkurven $l(\vartheta) = -\frac{\rho}{2} + \arccos \vartheta$ und $r(\vartheta) = \frac{\rho}{2} + \arccos \vartheta$ gleich dem Rechteck mit Seitenlängen ρ und 1.

Allerdings muss nun die Bedingung $0 \leq \phi$ durch Abschneiden des Flächenstückes R für negative ϕ berücksichtigt werden:

$$R = \frac{\rho}{2} - \int_0^{\rho/2} \cos t \, dt = \frac{\rho}{2} - [\sin t]_0^{\rho/2} \\ = \frac{\rho}{2} - \sin \frac{\rho}{2}.$$

Ergebnis:

$$F_{E^+} = \rho - R \\ = \frac{\rho}{2} + \sin \frac{\rho}{2}, \\ F_E = 2 \cdot F_{E^+} \\ = \rho + 2 \sin \frac{\rho}{2}.$$

Für die gesuchte Wahrscheinlichkeit gilt:

$$\Pr[(\Phi, \Theta) \in E] = \Pr[G(\Phi, \Theta) \cap L_\rho \neq \emptyset] \\ = \frac{F_E}{2\pi} \\ = \frac{\rho + 2 \sin \frac{\rho}{2}}{2\pi}.$$

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

Auf zwei unabhängigen Servern stehe ein Web-Dienst zur Verfügung. Es soll festgestellt werden, welcher Server schnellere Antwortzeiten liefert. Dazu werden $n = 1000$ Anfragen an die Server geschickt und es wird festgestellt, von welchem Server die Antwort zuerst eintrifft. Dabei gehen wir davon aus, dass Pakete nicht gleichzeitig empfangen werden können. In 540 Fällen antwortet Server A vor Server B .

Wir wählen als Nullhypothese H_0 die Aussage, dass Server B im Mittel schneller ist als Server A .

Kann man für einen entsprechenden statistischen Test auf einem Signifikanzniveau von $\alpha = 0,04$ die Nullhypothese annehmen?

Formulieren Sie hierzu den Test und weisen Sie Ihre Behauptung nach.

Lösung

Sei X eine Bernoulli-verteilte Zufallsvariable mit Wert 1 genau dann, wenn A schneller antwortet als B . Wir schreiben $p = \Pr[X = 1]$. Seien X_i für $i \in \{1, 2, \dots, n\}$ unabhängige Wiederholungen von X mit $n = 1000$. Dann ist die Testgröße $T = \sum_{i=1}^n X_i$ binomialverteilt.

Wir formulieren einen approximativen Binomialtest.

$$\text{Nullhypothese } H_0: p < \frac{1}{2}, \quad \text{Alternative } H_1: p \geq \frac{1}{2}.$$

Die Nullhypothese mit Signifikanzniveau $\alpha = 0,04$ anzunehmen, heißt aber, die triviale Alternative mit Signifikanzniveau $\alpha = 0,04$ abzulehnen. Um also das Testschema des approximativen Binomialtests nach Vorlesung anwenden zu können, bezeichnen wir die Alternative H_1 als H'_0 und H_0 als H'_1 .

Nun testen wir, ob die Hypothese H'_0 mit Signifikanzniveau $\alpha = 0,04$ abgelehnt werden kann. Wir setzen also

$$\text{Nullhypothese } H'_0: p \geq \frac{1}{2}, \quad \text{Alternative } H'_1: p < \frac{1}{2},$$

und wenden den approximativen Binomialtest wie folgt an.

Mit $p_0 = \frac{1}{2}$ und $T = h$ sei

$$Z = \frac{h - np_0}{\sqrt{np_0(1 - p_0)}} = \frac{540 - 500}{\sqrt{250}} \approx 2.5298 \dots$$

Das Ablehnungskriterium für H'_0 mit $\alpha = 0.04$ ist $Z < z_{0.04} \approx -0.9599$.

Dieses Kriterium ist wegen $Z = \frac{540-500}{\sqrt{250}} > z_{0.04}$ nicht erfüllt, d. h. wir können H_0 nicht mit Signifikanz $\alpha = 0,04$ annehmen.

Nun könnten wir sofort mit gleichem Test prüfen, ob H_0 mit Signifikanzniveau, d. h., maximaler Irrtumswahrscheinlichkeit $\alpha = 0.04$ ablehnen können. Und dies ist in der Tat so.

Es gilt nämlich, dass H_0 mit Signifikanzniveau $\alpha = 0.04$ abgelehnt werden kann. Dies folgt aus $Z > z_{1-\alpha} \approx 1.6667$.

Aber das war nicht die Frage. Gefragt war, ob wir H_0 mit Signifikanzniveau $\alpha = 0.04$ annehmen können.

Vorbereitung 2

An der dänischen Grenze werden Grenzkontrollen durchgeführt. Im Schnitt treffen alle 30 Sekunden an der Grenzstation Personen ein, die zu kontrollieren sind. Die Zeit zwischen zwei Kontrollen sei exponentialverteilt mit Parameter $\frac{1}{30}$. Wenn 2 Minuten lang kein Kontrollfall eingetroffen ist, dann machen die Grenzbeamten Ruhepause.

Seien T_1, T_2, \dots die Zeitspannen zwischen dem Eintreffen von zu kontrollierenden Personen und W die Wartezeit bis zur nächsten Ruhepause.

1. Berechnen Sie $\mathbb{E}[T_1 \mid T_1 \geq 120]$.
2. Berechnen Sie $\mathbb{E}[W]$.

Lösung

1. Sei $\lambda = \frac{1}{30}$. Dann gilt $\mathbb{E}[T_1] = 30$.

Es gilt $\mathbb{E}[T_1 \mid T_1 \geq 120] = 120 + \mathbb{E}[T_1] = 150$.

2. Wir wählen die Bezeichnung W' wie in TA 2 von Blatt 8.

Es gilt $W = 120 + \sum_{j=1}^{N-1} T_j$ oder $W = 120 + W'$ mit $W' = \sum_{j=1}^{N-1} T_j$.

N ist eine geometrisch verteilte Zufallsvariable mit Erfolgswahrscheinlichkeit $p = \Pr[T \geq 120]$. Es gilt

$$p = \Pr[T \geq 120] = e^{-4}.$$

Berechnung von $\mathbb{E}[W']$:

Wir setzen $T = T_1$.

$$\begin{aligned} \mathbb{E}[W'] &= \sum_{n=1}^{\infty} \mathbb{E}[W' \mid N = n] \cdot p(1-p)^{n-1} \\ &= \sum_{n=1}^{\infty} \mathbb{E}[T \mid T \leq 120] \cdot (n-1) \cdot p(1-p)^{n-1} \\ &= \mathbb{E}[T \mid T \leq 120] \cdot \sum_{n=1}^{\infty} (n-1) \cdot p(1-p)^{n-1} \\ &= \mathbb{E}[T \mid T \leq 120] \cdot \mathbb{E}[N-1]. \end{aligned}$$

Es gilt

$$\mathbb{E}[N-1] = e^4 - 1.$$

$\mathbb{E}[T \mid T \leq 120]$ erhalten wir aus den Gleichungen

$$\begin{aligned} \mathbb{E}[T] &= \mathbb{E}[T \mid T \leq 120] \cdot \Pr[T \leq 120] + \mathbb{E}[T \mid T \geq 120] \cdot \Pr[T \geq 120] \\ &= \mathbb{E}[T \mid T \leq 120] \cdot (1 - e^{-4}) + 150 \cdot e^{-4} \\ &= 30. \end{aligned}$$

Es folgt

$$\mathbb{E}[T|T \leq 120] = \frac{30 - 150 \cdot e^{-4}}{1 - e^{-4}} = \frac{30 \cdot e^4 - 150}{e^4 - 1}.$$

Ergebnis:

$$\begin{aligned}\mathbb{E}[W'] &= 30 \cdot e^4 - 150, \\ \mathbb{E}[W] &= 30 \cdot e^4 - 150 + 120 \\ &\approx 26,8 \text{ (Minuten)}.\end{aligned}$$

Vorbereitung 3

Seien T_1, T_2, \dots unabhängige exponentialverteilte Zufallsvariablen mit gleichem Parameter λ . Seien $S_n = \sum_{i=1}^n T_i$ für alle $n \in \mathbb{N}_0$.

Man zeige für alle $n \in \mathbb{N}$ für die Verteilungsfunktion von S_n , dass für alle $t \geq 0$ gilt:

$$F_{S_{n+1}}(t) = -\frac{\lambda^n t^n}{n!} e^{-\lambda t} + F_{S_n}(t).$$

Lösung

Zunächst beweisen wir

$$F_{S_n}(t) - F_{S_{n+1}}(t) = \frac{1}{\lambda} f_{S_{n+1}}(t).$$

Beweis:

$$\begin{aligned}F_{S_{n+1}}(t) &= \int_{-\infty}^{\infty} f_T(T) \cdot \left(\int_{-\infty}^{t-T} f_{S_n}(S) \, dS \right) dT \\ &= \int_{-\infty}^{\infty} f_T(T) \cdot F_{S_n}(t-T) \, dT \\ &= \int_0^t f_T(T) \cdot F_{S_n}(t-T) \, dT \\ &= [F_T(T) F_{S_n}(t-T)]_{T=0}^{T=t} + \int_0^t F_T(T) f_{S_n}(t-T) \, dT \\ &= \int_0^t \left(1 - \frac{1}{\lambda} f_T(T)\right) f_{S_n}(t-T) \, dT \\ &= \int_0^t f_{S_n}(t-T) \, dT - \frac{1}{\lambda} \int_0^t f_T(T) \cdot f_{S_n}(t-T) \, dT \\ &= F_{S_n}(t) - \frac{1}{\lambda} f_{S_{n+1}}(t).\end{aligned}$$

Nun beweisen wir durch Induktion über $n \geq 1$

$$f_{S_n}(t) = \begin{cases} \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} & \text{für } t \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Beweis:

Für $n = 1$ gilt die Formel, da für $t \geq 0$ $f_T(t) = \lambda e^{-\lambda t}$ gilt.

Induktionsschritt von n auf $n + 1$ für alle $n \geq 1$:

$$\begin{aligned} f_{S_{n+1}}(t) &= \int_0^t f_T(t-x) f_{S_n}(x) \, dx \\ &= \int_0^t \lambda e^{-\lambda(t-x)} \cdot \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} \, dx \\ &= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda t} \int_0^t x^{n-1} \, dx \\ &= \frac{\lambda^{n+1} t^n}{n!} e^{-\lambda t}. \end{aligned}$$

Tutoraufgabe 1

Sei $T = X_1 + X_2 + X_3$ eine Zufallsvariable, wobei X_1 , X_2 und X_3 unabhängige Bernoulli-verteilte Zufallsvariablen mit gleicher Erfolgswahrscheinlichkeit p seien. Wir betrachten T als Stichprobenvariable zum Test der Hypothese $H_0 : p \geq \frac{1}{3}$ mit Ablehnungsbereich $\tilde{K} = \{0, 1, 2\}$.

1. Berechnen Sie die Verteilungsfunktion F_T von T !
2. Berechnen Sie die Fehlerwahrscheinlichkeit 1. Art!
3. Berechnen Sie die Fehlerwahrscheinlichkeit 2. Art unter der Annahme, dass $H_1 : p \leq \frac{1}{4}$ eine echte Alternative zu H_0 sei!

Lösung

1. T ist binomialverteilt mit

$$f_T(i) = \binom{3}{i} p^i (1-p)^{3-i} \text{ und } W_T = \{0, 1, 2, 3\}.$$

Allgemein gilt

$$F_T(i) = \Pr[T \leq i].$$

$$F_T(0) = (1-p)^3,$$

$$\begin{aligned} F_T(1) &= (1-p)^3 + 3p(1-p)^2 \\ &= 1 - 3p^2 + 2p^3, \end{aligned}$$

$$F_T(2) = 1 - p^3,$$

$$F_T(3) = 1.$$

- 2.

$$\begin{aligned} \alpha_1 &= \max_{p \in H_0} \Pr[T \in \tilde{K}] \\ &= \max_{p \geq \frac{1}{3}} F_T(2) \\ &= \max_{p \geq \frac{1}{3}} (1 - p^3) \\ &= 1 - \left(\frac{1}{3}\right)^3 = \frac{26}{27}. \end{aligned}$$

- 3.

$$\begin{aligned} \alpha_2 &= \max_{p \in H_1} \Pr[T \notin \tilde{K}] \\ &= \max_{p \leq \frac{1}{4}} f_T(3) \\ &= \max_{p \leq \frac{1}{4}} p^3 \\ &= \left(\frac{1}{4}\right)^3 = \frac{1}{64}. \end{aligned}$$

Tutoraufgabe 2

Ein Hersteller von Nahrungsmittelkonserven gibt die Haltbarkeit eines bestimmten Produkts mit mindestens 50 Monaten an. Sie kaufen 20 derartige Konserven und messen ihre Haltbarkeit. Die Konserven haben im Schnitt 40 Monate gehalten mit einer Stichprobenstandardabweichung von $S = 30$ Monaten.

1. Zeigen Sie, dass die Angabe des Herstellers nicht abgelehnt werden kann (Signifikanzniveau 0.05).
2. Wie viele Konserven hätten Sie mindestens kaufen müssen, um die Angabe des Herstellers ablehnen zu können?

Lösung

Die Tabellenwerte für den t -Test wurden dem Buch von Schickinger/Steger entnommen.

1. Wir wenden einen t -Test an.

Seien $\mu_0 = 50$, $\bar{X} = 40$ und $S = 30$. Für die Testgröße T gilt:

$$T = \frac{\bar{X} - \mu_0}{S} \cdot \sqrt{n} = -\frac{1}{3}\sqrt{n}.$$

Die Herstellerangabe liefert die Nullhypothese $H_0 : \mu \geq 50 = \mu_0$.

Um H_0 ablehnen zu können, muss $T < t_{n-1,0.05}$ gelten. Für $n = 20$ ergibt sich

$$-\frac{1}{3}\sqrt{20} = -1.49 \not< -1.73 \stackrel{\text{Tabelle}}{\approx} t_{19,0.05}.$$

H_0 kann folglich nicht abgelehnt werden.

2. Durch Ausprobieren stellt man fest, dass $n = 27$ das kleinste n ist, für das die Hypothese abgelehnt werden kann. Es gilt:

$$-\frac{1}{3}\sqrt{26} = -1.700 \not< -1.708 \stackrel{\text{Tabelle}}{\approx} t_{25,0.05},$$

aber

$$-\frac{1}{3}\sqrt{27} = -1.732 < -1.706 \stackrel{\text{Tabelle}}{\approx} t_{26,0.05}.$$

Tutoraufgabe 3

Seien T_0, T_1, T_2, \dots unabhängige exponentialverteilte Zufallsvariablen mit gleichem Parameter λ . Beispiel: Grenzkontrollen in VA 2 mit $\lambda = \frac{1}{30}$.

Seien $S_n = \sum_{i=0}^n T_i$ für alle $n \in \mathbb{N}_0$. Wir betrachten den Stochastischen Prozess $(X(t))_{t>0}$, der gegeben ist durch

$$X(t) := \max\{n \in \mathbb{N}_0; S_n \leq t\}.$$

1. Zeigen Sie, dass für alle $t > 0$ und $n \in \mathbb{N}_0$ die folgende Gleichung gilt:

$$f_{X(t)}(n) = \Pr[S_n \leq t < S_{n+1}].$$

2. Für alle $t > 0$ sind die Variablen $X(t)$ Poisson-verteilt. Geben Sie eine explizite Darstellung der Dichtefunktion $f_{X(t)}$ an.
3. Beweisen Sie die vorgenannte explizite Darstellung der Dichtefunktion $f_{X(t)}$ durch direkte Berechnung und ohne Benutzung der Kenntnis, dass $X(t)$ Poisson-verteilt ist.

Lösung

1. Für $n = 0$ wird $S_n = 0$ gesetzt, wie üblich bei leeren Summen.

Nun gilt $m = \max\{n \in \mathbb{N}_0; S_n \leq t\}$ genau dann, wenn $S_m \leq t < S_{m+1}$. Damit folgt $X(t) = m \iff (S_m \leq t < S_{m+1})$, mithin

$$f_{X(t)}(m) = \Pr[X(t) = m] = \Pr[S_m \leq t < S_{m+1}].$$

Bemerkung: Der Wertebereich der Zufallsvariablen $X(t)$ ist $\mathbb{N}_0 \cup \{\infty\}$, denn sie ist erst dann vollständig definiert, wenn man ∞ bei der Maximumsbildung zulässt. Allerdings muss $\Pr[X(t) = \infty] = 0$ für alle $t > 0$ gesetzt werden, wie man durch Betrachtung der Verteilungsfunktion F_{S_n} und Grenzwertbildung für $n \rightarrow \infty$ zeigen kann.

2. Genau genommen gilt diese Aussage nur, wenn man die Betrachtung von ∞ ausschließt, denn eine Poisson-Verteilung sieht ∞ nicht vor. Die auf \mathbb{N}_0 eingeschränkte Dichtefunktion von $X(t)$ ist aber sehr wohl eine Poisson-Dichtefunktion.

Wir benutzen, die in der Vorlesung mitgeteilte Tatsache, dass für alle $t > 0$ die Variablen $X(t)$ Poisson-verteilt sind mit Parameter $\lambda \cdot t$. Es folgt für alle $n \in \mathbb{N}_0$

$$f_{X(t)}(n) = \frac{(\lambda t)^n}{n!} \cdot e^{-\lambda \cdot t}.$$

3. Wir benutzen die in Vorbereitungsaufgabe 3 bewiesene Rekursionsformel für die Verteilungsfunktion F_{S_n} .

Die Ereignisse $(t < S_n)$ und $(S_n \leq t < S_{n+1})$ sind disjunkt und es gilt $t < S_{n+1} \iff (t < S_n) \vee (S_n \leq t < S_{n+1})$, weil stets $S_n \leq S_{n+1}$ gilt. Daraus folgt

$$\Pr[t < S_{n+1}] = \Pr[t < S_n] + \Pr[S_n \leq t < S_{n+1}].$$

Wir rechnen

$$\begin{aligned} f_{X(t)}(n) &= \Pr[S_n \leq t < S_{n+1}] \\ &= \Pr[t < S_{n+1}] - \Pr[t < S_n] \\ &= (1 - F_{S_{n+1}}(t)) - (1 - F_{S_n}(t)) \\ &= F_{S_n}(t) - F_{S_{n+1}}(t) \\ &= \frac{\lambda^n t^n}{n!} e^{-\lambda t}. \end{aligned}$$