

Stochastik

Musterlösung 11

1. Diese Aufgabe behandelt verschiedene Themenbereiche aus dem gesamten bisherigen Vorlesungsmaterial. Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig.

a) Es sei die diskrete Zufallsvariable X gegeben durch $P(X = \frac{2}{3}) = \frac{1}{6}$, $P(X = 1) = \frac{1}{2}$, $P(X = \frac{3}{2}) = \frac{1}{3}$. Wie gross ist $F_X(5/4)$?

1. $\frac{1}{2}$.
2. $\frac{2}{3}$.
3. 1.

b) Welche Aussage ist korrekt?

1. $\text{Cov}(X + Y, X) = \text{Var}(X)$.
2. $\text{Cov}(X + Y, X) = \text{Var}(X) + \text{Cov}(X, Y)$.
3. $\text{Cov}(X + Y, X) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$.

c) Betrachte eine Reihe von unabhängigen Wiederholungen eines Experiments mit Erfolgswahrscheinlichkeit p . Welche ist die korrekte Verteilung für die Anzahl Wiederholungen, bis der erste Erfolg eintritt?

1. $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x \in \{1, 2, \dots\}$.
2. $p(x) = (1-p)p^{x-1}$, $x \in \{1, 2, \dots\}$.
3. $p(x) = p(1-p)^{x-1}$, $x \in \{1, 2, \dots\}$.

d) Es seien A und B Ereignisse mit $P(A) > 0$, $P(B) > 0$. Wir nehmen an, dass $P(A|B) > P(A)$. Dann gilt:

1. $P(B|A) > P(B)$.
2. $P(A|B) = P(B|A)$.
3. $P(A) = P(B)$.

e) Es sei die Zufallsvariable X gegeben mit der Dichte

$$f(x) = \begin{cases} c+x & -c \leq x \leq 0 \\ c-x & 0 \leq x \leq c \end{cases}$$

Bestimme die Konstante c .

1. $c = 2$.
2. $c = 1$.
3. $c = \frac{1}{2}$.

- f) Gegeben sind i.i.d. $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariablen $X_i, i = 1, 2, \dots, n$. Definiere $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$. Wie gross ist der Erwartungswert von

$$\frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2?$$

1. σ^2 .
 2. μ^2 .
 3. $\frac{n-1}{n}\sigma^2$.
- g) Wenn das Signifikanzniveau α eines Tests kleiner wird, dann
1. wird der Verwerfungsbereich für die Nullhypothese H_0 grösser.
 2. wird das Vertrauensintervall grösser.
 3. wird der P-Wert grösser.
- h) Gegeben sind n Beobachtungen x_1, \dots, x_n von Zufallsvariablen X_1, \dots, X_n . Wir nehmen an dass die X_i unabhängig und normalverteilt sind mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2 . Welcher Test ist geeignet um die Nullhypothese $H_0 : \mu = 5$ gegen die Alternative $H_A : \mu \neq 5$ zu testen?
1. t-Test.
 2. z-Test.
 3. Binomialtest.

Es wird ein Test mit Nullhypothese $H_0 : \mu = 0$ und der Alternative $H_A : \mu \neq 0$ auf dem Signifikanzniveau α geführt. Betrachte die Funktion $g(\mu) = P(H_0 \text{ ablehnen} | \mu)$. Die Funktion g gibt die Wahrscheinlichkeit an, H_0 zu verwerfen, für verschiedene Werte des unbekanntes wahren Parameters μ .

- i) Wenn $\mu = 0$ ist, dann gilt
1. $g(\mu) > \alpha$.
 2. $g(\mu) < 1 - \alpha$.
 3. $g(\mu) \leq \alpha$.
- j) Wenn $\mu \neq 0$, dann ist welche Aussage falsch?
1. $1 - g(\mu)$ ist die Wahrscheinlichkeit für den Fehler 1. Art.
 2. $1 - g(\mu)$ ist die Wahrscheinlichkeit für den Fehler 2. Art.
 3. $g(\mu)$ ist die Macht eines Tests.

Lösung:

a) 2 b) 2 c) 3 d) 1 e) 2 f) 3 g) 2 h) 1 i) 3 j) 1

2. Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig.

a) Sei die Funktion f explizit gegeben durch

$$f(x) = \begin{cases} cx^3, & 0 \leq x \leq 2 \\ c, & -2 \leq x \leq 0. \end{cases}$$

Für welchen Wert c stellt f eine Dichtefunktion dar?

1. $\frac{1}{2}$.
2. $\frac{1}{6}$.
3. 1.

b) Die diskreten Zufallsvariablen X und Y sind durch folgende gemeinsamen Verteilungen gegeben

$$\begin{aligned} P(X = 1, Y = -2) &= \frac{1}{8}, & P(X = 1, Y = 1) &= \frac{1}{4}, & P(X = 1, Y = 4) &= \frac{1}{8}, \\ P(X = 3, Y = -2) &= \frac{1}{8}, & P(X = 3, Y = 1) &= \frac{1}{8}, & P(X = 3, Y = 4) &= \frac{1}{4}. \end{aligned}$$

Wie gross sind die Erwartungswerte $\mathbb{E}(X)$ und $\mathbb{E}(Y)$?

1. $\mathbb{E}(X) = \frac{3}{2}$ und $\mathbb{E}(Y) = 2$.
2. $\mathbb{E}(X) = 1$ und $\mathbb{E}(Y) = \frac{5}{2}$.
3. $\mathbb{E}(X) = 2$ und $\mathbb{E}(Y) = \frac{11}{8}$.

c) Ist X eine binomialverteilte Zufallsvariable mit den Parametern n und p , dann konvergiert die Verteilungsfunktion $F_n(z)$ der standardisierten Zufallsvariablen

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

für $n \rightarrow \infty$ gegen die Verteilungsfunktion der

1. Exponentialverteilung.
2. Standardnormalverteilung.
3. Poissonverteilung.

d) Seien X und Y Zufallsvariablen mit Korrelation $\text{Corr}(X, Y) = -1$, Varianz $\text{Var}(X) = 1$ und Standardabweichung $\sqrt{\text{Var}(Y)} = 2$. Wie gross ist dann $\text{Var}(X + Y)$?

1. 1.
2. 3.
3. 5.

e) Seien X_1, \dots, X_n unabhängig und normalverteilt mit Erwartungswert μ und Varianz σ^2 und sei $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Dann ist für σ^2 der Schätzer

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

1. erwartungstreu.
2. nicht erwartungstreu.

3. Es kann keine Aussage gemacht werden.

f) Beim Testen einer Nullhypothese $H_0 : \mu = \mu_0$ wird bei gegebener Standardabweichung σ bei zweiseitiger Alternative $H_A : \mu \neq \mu_0$ die Nullhypothese H_0 abgelehnt, falls

1. $|\bar{X}_n - \mu_0| < \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2}),$

2. $\bar{X}_n - \mu_0 > \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2}),$

3. $|\bar{X}_n - \mu_0| > \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \frac{\alpha}{2}),$

wobei $\Phi(\cdot)$ die Verteilungsfunktion der Standardnormalverteilung ist.

g) Welche Aussage gilt für einen Test zum Niveau α ?

1. Je grösser das Niveau α , desto eher behalten wir die Nullhypothese.

2. Je kleiner das Niveau α , desto grösser die Wahrscheinlichkeit des Fehlers 2. Art.

3. α ist die Wahrscheinlichkeit des Fehlers 2. Art.

h) Seien X_1, \dots, X_n unabhängig und identisch verteilt mit der Dichtefunktion

$$f_{\vartheta}(x) = \begin{cases} (\vartheta - 1)x^{-\vartheta}, & x \geq 1 \\ 0, & \text{sonst,} \end{cases}$$

wobei $\vartheta \in (0, \infty)$ ein unbekannter Parameter ist. Der Maximum Likelihoodschätzer für ϑ lautet

1. $\frac{n}{\sum_{i=1}^n \log(X_i)}.$

2. $1 + \frac{n}{\sum_{i=1}^n \log(X_i)}.$

3. $\frac{n}{\vartheta-1} - \sum_{i=1}^n \log(X_i).$

i) Welche der folgenden Aussagen ist richtig?

1. Fehler 2. Art tritt auf, wenn die Nullhypothese H_0 akzeptiert wird, obwohl die Alternativhypothese H_A gilt.

2. Fehler 2. Art tritt auf, wenn die Nullhypothese H_0 abgelehnt wird, obwohl sie wahr ist.

3. Weder 1. noch 2. gilt.

j) Welche der nachfolgenden Interpretationen für den p -Wert eines Tests stimmt?

1. Der p -Wert ist die Wahrscheinlichkeit, dass die Nullhypothese H_0 stimmt.

2. Der p -Wert ist eine Obergrenze für die Wahrscheinlichkeit eines Fehlers 1. Art.

3. Weder 1. noch 2. sind richtig.

Lösung:

a) 2. b) 3. (Erklärung: $\mu_X = 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = 2$ und $\mu_Y = -2 \cdot \frac{1}{4} + 1 \cdot \frac{3}{8} + 4 \cdot \frac{3}{8} = \frac{11}{8}$) c) 2. d) 1. e) 1. f) 3. g) 2. h) 2. i) 1. j) 3.

3. Der Fernsehsender SkiTV möchte ein Skirennen mit 81 Skifahrern live übertragen. Der i -te Skifahrer ($i = 1, \dots, 81$) erreicht nach einer Zeit T_i (in Minuten) das Ziel. Sobald ein Skifahrer im Ziel ist,

kann der nächste ohne Zeitverzögerung direkt starten.
Die Zufallsvariablen T_i seien iid mit Dichte f_T , wobei

$$f_T(x) = \begin{cases} \lambda e^{-\lambda(x-1)}, & x \geq 1, \\ 0, & \text{sonst,} \end{cases}$$

und $\lambda = 3$.

- a) Bestimme den Erwartungswert $E[T_i]$ und die Varianz $\text{Var}(T_i)$.
Tip: Falls $S \sim \text{Exp}(\lambda)$, so ist $E[S] = 1/\lambda$ und $\text{Var}(S) = 1/\lambda^2$.
- b) Wie gross ist die Wahrscheinlichkeit, dass ein Skifahrer mehr als 70 Sekunden die Piste besetzt?
- c) Aufgrund langjähriger Erfahrung hat der TV-Sender genauere Informationen, nämlich dass die erwartete Dauer bis ein Skifahrer im Ziel ist $\mu = 13/9$ beträgt, mit einer Varianz von $\sigma^2 = 1/9$. Da der TV-Sender das ganze Rennen live übertragen will, fragt er sich wie lange das Rennen dauern wird. Man möchte das Rennen mit einer Wahrscheinlichkeit von 97.7% vollständig übertragen können. Wie viele Stunden/Minuten muss der Sender in seinem Sendeplan für das Skirennen reservieren? Benutze den zentralen Grenzwertsatz um dieses Problem zu lösen.

Lösung:

Beachte, dass $T_i - 1 \sim \text{Exp}(3)$.

- a) $E[T_i] = 1 + 1/\lambda = 4/3$ und
 $\text{Var}(T_i) = 1/\lambda^2 = 1/9$.
- b) $P(T > 7/6) = P(T - 1 > 1/6) = e^{-\lambda/6} = e^{-1/2} = 60.7\%$.
- c) Aus $S = \sum T_i$ folgt $E[S] = 81 \cdot 13/9 = 117$ und $\text{var}(S) = 81 \cdot 1/9 = 9$. Somit gilt mit dem ZGS

$$\frac{S - 117}{3} = 2,$$

und damit $S = 123$ Minuten oder 2 Stunden und 3 Minuten.

4. Die Firma Pharma AG vertreibt Schmerzmittel in Pulverform und Tablettenform. Bei der Einnahme des Pulvers klagen 40% der Patienten über Nebenwirkungen. Sie möchten nun testen: Wenn Tabletten statt Pulver eingenommen werden, ändert sich dann der Anteil der Patienten mit Nebenwirkungen signifikant?

Dazu werden zufällig 50 Patienten ausgewählt und es sei X die Anzahl Patienten unter diesen 50, welche bei der Einnahme von Tabletten *keine* Nebenwirkungen haben. Wir nehmen an, dass $X \sim \text{Bin}(50, p)$, wobei p die Erfolgswahrscheinlichkeit im Einzelfall ist ("Erfolg" = keine Nebenwirkungen).

- a) Obige Studie ergab das Folgende: Unter den 50 Patienten befanden sich 35, bei welchen durch die Tabletten keine Nebenwirkungen hervorgerufen wurden. Ist dies signifikant oder eher Zufall? Führen Sie dazu einen statistischen Test durch und gehen Sie dabei wie folgt vor:
 - i) Formulieren Sie ein geeignetes Modell.
 - ii) Formulieren Sie eine geeignete Null- und Alternativhypothese. Ist der Test ein- oder zweiseitig durchzuführen? Begründen Sie kurz!

- iii) Formulieren Sie die Teststatistik und berechnen Sie den Verwerfungsbereich des Tests für ein Signifikanzniveau von 5%.
- iv) Wie entscheidet der Test?

b) Angenommen, der wahre Anteil der unter Nebenwirkungen leidenden Patienten bei der Einnahme der Tabletten sei 20%. Wie gross ist in diesem Fall die Wahrscheinlichkeit eines Fehlers 2. Art?

Lösung:

a) Mittels eines Binomial-Tests möchten wir testen ob sich die Anteile der unter Nebenbeschwerden leidenden Patienten bei der Einnahme von Tabletten und Pulver signifikant unterscheiden. Wir gehen dabei nun wie folgt vor:

- Laut Aufgabenstellung ist die Anzahl Patienten, welche frei von Nebenwirkungen sind, gegeben durch $X \sim Bin(50, p)$. Dabei bezeichnet der Erfolgsparameter p die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person keine Nebenbeschwerden hat bei der Einnahme des neuen Medikaments.
- H_0 : Die Tabletten bringen keine signifikante Veränderung bzgl. Nebenwirkungen mit sich, d.h. es sind gleichviele Patienten welche frei sind von Nebenbeschwerden wie bei der Einnahme des Pulvers, also $X \sim Bin(50, p)$ mit $p = 0.6$.
 H_A : Das neue Medikament bringt eine signifikante Veränderung bzgl. Nebenwirkungen mit sich, d.h. es hat signifikant mehr oder signifikant weniger Patienten welche frei sind von Nebenbeschwerden, also $X \sim Bin(50, p)$ mit $p \neq 0.6$.
 Der Test ist zweiseitig durchzuführen. Wir möchten herausfinden ob eine signifikante *Veränderung* vorliegt. Es ist nicht wichtig ob es signifikant mehr oder signifikant weniger Patienten sind.
- Unter H_0 ist $X \sim Bin(n = 50, p = 0.6)$. Da wir zweiseitig testen, verwerfen wir H_0 , falls extrem viel mehr *oder* extrem viel weniger Patienten unter Beschwerden leiden, d.h. der Verwerfungsbereich hat die Form $V = V_1 \cup V_2$, wobei $V_1 = \{k \in \mathbb{N}_0 : k \leq k_1\}$ und $V_2 = \{k \in \mathbb{N}_0 : k \geq k_2\}$ für noch zu bestimmende $k_1, k_2 \in \mathbb{N}_0$.
 Für einen Test auf dem Niveau $\alpha = 5\%$ soll also gelten $P_{H_0}(X \in V) \leq 0.05$. Da zweiseitig getestet wird, haben wir $P_{H_0}(X \in V) = P_{H_0}(\{X \in V_1\} \cup \{X \in V_2\}) = P_{H_0}(X \in V_1) + P_{H_0}(X \in V_2)$ und somit muss gelten $P_{H_0}(X \in V_i) \leq 0.05/2 = 0.025$ für $i = 1, 2$.

V_1 : Gesucht ist k_1 so, dass $P_{H_0}(X \in V_1) = P_{H_0}(X \leq k_1) \leq 2.5\%$. Durch Ablesen aus der Binomialtabelle erhalten wir $P_{H_0}(X \leq 23) = 0.0314$ $P_{p=0.6}(X \leq 22) = 0.0160$ und somit also $k_1 = 22$.

V_2 Analog, suchen wir k_2 so, dass $P_{H_0}(X \in V_2) = P_{H_0}(X \geq k_2) \leq 2.5\%$. Mittels Tabelle findet man $P_{H_0}(X \geq k_2 = 37) = 1 - P_{H_0}(X \leq 36) = 1 - 0.9720 = 0.0280$ und $P_{H_0}(X \geq k_2 = 38) = 1 - P_{H_0}(X \leq 37) = 1 - 0.9867 = 0.0133$ und somit $k_2 = 38$.

Der Verwerfungsbereich ist somit gegeben durch $V = \{\dots, 20, 21, 22\} \cup \{38, 39, 40, \dots\}$.

- Testentscheidung: Von den ausgewählten 50 Patienten hatten bei der Einnahme des neuen Medikaments 35 über Nebenwirkungen geklagt. Somit verwerfen wir H_0 nicht, denn $35 \notin V$.

b) Gesucht ist der Fehler 2. Art des Binomial-Tests aus a) für die Alternativhypothese $H_A : X \sim Bin(50, 1 - 0.2 = 0.8)$. Der Fehler 2. Art ist die Wahrscheinlichkeit H_0 nicht zu verwerfen unter der Annahme dass H_A korrekt ist, d.h. $P(\text{Fehler 2. Art}) = P_{H_A}(X \notin V)$.

Laut Binomial-Tabelle (für $p = 0.8$) gilt mit dem korrekten Verwerfungsbereich aus a) $V = \{\dots, 21, 22\} \cup \{38, 39, 40, \dots\}$ also

$$P_{H_A}(X \notin V) = P_{H_A}(23 \leq X \leq 37) = \underbrace{P_{H_A}(X \leq 37)}_{= 0.186} - \underbrace{P_{H_A}(X \leq 22)}_{= 0} = 0.186.$$

Alternativlösung: Es kann auch eine *Normalapproximation* verwendet werden.

- a)
- $X \sim N(\mu, \sigma)$ mit $\mu = np = 50p$.
 - $H_0 : \mu = 50 \cdot 0.6 = 30$ und $H_A : \mu \neq 30$, zweiseitig mit gleicher Begründung wie oben.
 - Für einen Test auf dem Niveau $\alpha = 5\%$ soll also gelten $P_{H_0}(X \in V) \leq 0.05$. Da zweiseitig getestet wird, muss gelten $P_{H_0}(X \leq k_1) \leq 0.05/2 = 0.025$ und $P_{H_0}(X \geq k_2) \leq 0.05/2 = 0.025$ für noch zu findende Werte k_1 und k_2 . Mittels Normalapproximation erhalten wir:

$$P_{H_0}(X \leq k_1) = P_{H_0}\left(\frac{X - \mu}{\sigma} \leq \frac{k_1 - \mu}{\sigma}\right) \approx \Phi\left(\frac{k_1 - \mu}{\sigma}\right)$$

und somit gilt mit Hilfe der NV-Tabelle

$$\frac{k_1 - \mu}{\sigma} \leq \Phi^{-1}(0.025) \approx -1.96 \Rightarrow k_1 \leq 30 - \sqrt{121} \cdot 1.96 \approx 23.21$$

Analog gilt für k_2 :

$$\frac{k_2 - \mu}{\sigma} \geq \Phi^{-1}(0.975) \approx 1.96 \Rightarrow k_2 \leq 30 + \sqrt{121} \cdot 1.96 \approx 36.79$$

Damit ist der Verwerfungsbereich V gegeben durch $V = \{x \in \mathbb{R} : x \leq k_1\} \cup \{x \in \mathbb{R} : x \geq k_2\}$.

Analog können wir auch den Verwerfungsbereich für p anstatt X angeben: Verwerfungsbereich = ausserhalb des Vertrauensintervalls:

$$VI = \frac{x}{n} \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{x}{n} \left(1 - \frac{x}{n}\right) \frac{1}{n}} = \begin{cases} [0.57, 0.82], \text{ falls } H_0 : p = 0.6, \\ [0.26, 0.53], \text{ falls } H_0 : p = 0.4, \end{cases}$$

- Testentscheidung: $35 \notin V$ (resp. $0.7 \notin [-\infty, 0.57] \cup [0.82, \infty]$) oder $0.4 \notin [-\infty, 0.26] \cup [0.53, \infty]$. Die Nullhypothese wird also nicht verworfen.

- b) Fehler 2.Art: $H_A : X \sim N(\mu, \sigma)$ mit $\mu_A = np = 50 \cdot 0.8 = 40$ Damit gilt

$$\begin{aligned} P(\text{Fehler 2. Art}) &= P_{H_A}(X \notin V) = P_{H_A}(k_1 \leq X \leq k_2) \\ &= P_{H_A}(X \leq k_2) - P_{H_A}(X \leq k_1) \\ &= P_{H_A}\left(\frac{X - \mu}{\sigma} \leq \frac{k_2 - \mu}{\sigma}\right) - P_{H_A}\left(\frac{X - \mu}{\sigma} \leq \frac{k_1 - \mu}{\sigma}\right) \\ &\approx \Phi\left(\frac{k_2 - \mu}{\sigma}\right) - \Phi\left(\frac{k_1 - \mu}{\sigma}\right) \\ &\approx \Phi\left(\frac{36.79 - 40}{\sqrt{8}}\right) - \Phi\left(\frac{23.21 - 40}{\sqrt{8}}\right) \\ &\approx 1 - \Phi(1.135) - (1 - \Phi(5.936)) \\ &\approx 1 - 0.8719 \approx 0.1281. \end{aligned}$$