Wahrscheinlichkeit und Statistik

Lösungen Serie 12

Lösung Version 1 (30. Mai 2024: basierend auf Serie Version 2 (29. Mai: Kleine Verbesserungen der Formulierungen), Version 1 (23. Mai))
Bitte stellt Fragen in den Übungen und/oder im Forum.
Bitte stell sicher, dass du die Webseite `https://kahoot.it/` in der Übung am **28. Mai** öffnen kannst.
Freiwillige Abgabe bis **30. Mai 8:00**. Nachher kann selbstständig mit dieser Lösung verglichen werden.

Aufgabe 12.1 [Test I: Gezinkter Würfel]
Bei einem Würfel mit sechs Seiten soll getestet werden, ob der Würfel gezinkt ist und eher auf der Sechs landet. Hierzu wird ein Experiment durchgeführt, bei dem zehn Mal gewürfelt und jeweils die beobachtete Augenzahl notiert wird. Wir gehen davon aus, dass alle Würfe unabhängig voneinander sind und die Wahrscheinlichkeit, eine 1, 2, 3, 4 oder 5 zu würfeln, gleich ist. Wir modellieren die Ausgänge der Würfe als eine Stichprobe $X_1, \dots, X_{10}$, wobei $X_i = 1$ bedeutet, dass der i-te Wurf eine Sechs ist, und $X_i = 0$ sonst. Wir erhalten folgende Resultate:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0	0	1	0	1	1	0	1	0	0

(a) Bestimme ein geeignetes Modell $(\mathbb{P}_\theta)_{\theta\in\Theta}$, d.h. einen Parameterraum und die Verteilungen von $X_1, \dots, X_{10}$ unter jedem $\mathbb{P}_\theta$.

(b) Bestimme eine geeignete Hypothese H_0 und Alternative H_1.

(c) Sei $T = \sum_{i=1}^{10} X_i$. Welcher Verteilung folgt T?

(d) Sei $\varphi(X) = \mathbb{1}_{\sum_{i=1}^{10} X_i \in (4,10]} = \mathbb{1}_{T\in(4,10]}$ ein Test. Berechne die Wahrscheinlichkeit für einen Fehler 1. Art des Tests $\varphi(X)$.[1]

(e) Beschreibe den Testentscheid basierend auf den obigen Resultaten.

Lösung 12.1

(a) Aus den Annahmen folgt, dass $X_1, \dots, X_{10}$ unabhängig und jeweils $\mathrm{Ber}(\theta)$-verteilt sind unter P_θ mit unbekanntem Erfolgsparameter $\theta = P[X_1 = 1] \in \Theta = [0, 1]$.

(b) Unsere Nullhypothese ist, dass der Würfel nicht gezinkt ist; also ist

$$H_0 : \theta = \frac{1}{6}, \quad \text{d.h. } \Theta_0 = \left\{\frac{1}{6}\right\}.$$

Die Alternativhypothese, dass der Würfel gezinkt ist, ist dann

$$H_1 : \theta > 1/6, \quad \text{d.h. } \Theta_1 = \left(\frac{1}{6}, 1\right].$$

(c) Da $X_1, \dots, X_{10}$ unabhängig und $\mathrm{Ber}(\theta)$-verteilt sind unter P_θ, ist $T \sim \mathrm{Bin}(10, \theta)$ unter P_θ.

[1] In der Literatur wird die Zufallsvaraible $T = \sum_{i=1}^{10} X_i$ auch als Teststatistik bezeichnet und die deterministische Menge $K = (4, 10]$ als Verwerfungsbereich, wenn der Test die Form $\varphi(X) = \mathbb{1}_{T\in K}$ hat.

(d) Die Wahrscheinlichkeit für einen Fehler 1. Art ist

$$\begin{aligned}
\mathrm{P}_{\frac{1}{6}}[\varphi(X)=1] = \mathrm{P}_{\frac{1}{6}}[T \in (4,10]] &= \mathrm{P}_{\frac{1}{6}}[T \geq 5] \\
&= \sum_{k=5}^{10} \binom{10}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k} \\
&= \frac{1}{6^{10}} \left(\binom{10}{5} 5^5 + \binom{10}{6} 5^4 + \binom{10}{7} 5^3 \right. \\
&\qquad \left. + \binom{10}{8} 5^2 + \binom{10}{9} 5^1 + \binom{10}{10} 5^0 \right) \\
&= 0.0155.
\end{aligned}$$

(e) Da $t(x_1,\dots,x_{10}) = t(X_1(\omega),\dots,X_{10}(\omega)) = T(\omega) = 4$ gilt, erhalten wir $\varphi(X(\omega)) = \mathbb{1}_{4\in(4,10]} = 0$. Somit verwerfen wir die Nullhypothese nicht.

Aufgabe 12.2 [Test II: Normalverteilung]
Seien $X_1,\dots,X_{12}$ unabhängig und je $\mathcal{N}(\mu,\sigma^2)$-verteilt unter $\mathbb{P}_\theta$, wobei $\theta=\mu$ ein unbekannter Parameter ist. Die Standardabweichung $\sigma = 0.0499$ ist bekannt. Wir haben folgende Daten für die Stichprobe gegeben:

x_1	x_2	x_3	x_4	x_5	x_6
1.00781	1.00646	1.00801	1.00833	1.00738	1.00687
x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1.00783	1.00936	1.00564	1.00543	1.00794	1.01060

Wir testen die Hypothese $H_0\colon \mu=\mu_0=1.0085$ gegen die Alternative $H_1 : \mu \neq \mu_0$.

(a) Bestimme a und b so, dass $T := \frac{\sum_{i=1}^{12} X_i + b}{a} \sim \mathcal{N}(0,1)$-verteilt ist unter $\mathbb{P}_{\mu_0}$.

(b) Wähle $K := (-\infty,-c_{\neq}) \cup (c_{\neq},\infty)$ für ein zu bestimmendes $c_{\neq} \geq 0$, sodass $\varphi(X) = \mathbb{1}_{\frac{\sum_{i=1}^{12} X_i + b}{a} \in K} = \mathbb{1}_{T\in K}$ ein Test mit Signifikantsniveau 5% ist (bei bestmöglicher Macht des Tests)[2]. Teste H_0 gegen H_1 für das Signifikantsniveau 5%.

(c) Berechne die Macht des oben durchgeführten Tests an der Stelle $\mu = 1.008$.

Lösung 12.2

(a) Wegen $\sum_{i=1}^{12} X_i \sim \mathcal{N}(12\mu_0, 12\sigma^2)$ unter P_{μ_0} ist

$$T = \frac{\sum_{i=1}^{12} X_i - 12\mu_0}{\sigma\sqrt{12}} \sim \mathcal{N}(0,1)$$

unter P_{μ_0}. Wir wählen also $a = \sigma\sqrt{12}$ und $b = -12\mu_0$.

(b) Sei $\alpha = 0.05$. Wir führen einen Test mit der Teststatistik T und dem Verwerfungsbereich K durch, verwerfen also die Hypothese, falls $|T| > c_{\neq}$ für ein noch zu bestimmendes $c_{\neq}$. Die Definition des Niveaus ergibt

$$\begin{aligned}
\alpha = \mathrm{P}_{\mu_0}[\varphi(X)=1] = \mathrm{P}_{\mu_0}[T\in K] = \mathrm{P}_{\mu}\big[T \notin [-c_{\neq}, c_{\neq}]\big] &= \mathrm{P}_{\mu_0}[T < -c_{\neq}] + \mathrm{P}_{\mu_0}[T > c_{\neq}] \\
&= \Phi(-c_{\neq}) + 1 - \Phi(c_{\neq}) = 2 - 2\Phi(c_{\neq}),
\end{aligned}$$

[2]Je grösser $c_{\neq} \geq 0$ gewählt wird desto besser (kleiner) wird das Signifikanzniveau α, aber desto schlechter (kleiner) wird die Macht des Tests $\varphi(X)$. Darum ist es das Ziel dieser Aufgabe das kleinstmögliche $c_{\neq} \geq 0$ zu finden sodass $\varphi(X) = \mathbb{1}_{\frac{\sum_{i=1}^{12} X_i + b}{a} \in K} = \mathbb{1}_{T\in K}$ ein Test mit Signifikantsniveau 5% ist.

da $T \sim \mathcal{N}(0,1)$-verteilt ist unter P_{μ_0}. Aus

$$\Phi(c_{\neq}) = 1 - \frac{\alpha}{2} = 1 - \frac{0.05}{2} = 0.975$$

folgt, dass $c_{\neq} = \Phi^{-1}(0.975) = z_{0.975} = 1.96$. Der realisierte Schätzwert ist

$$T(\omega) = t(x_1, \dots, x_{10}) = -0.00598;$$

also verwerfen wir die Hypothese nicht.

(c) Die Macht des Tests an der Stelle μ ist

$$\begin{aligned}
\mathrm{P}_\mu[\varphi(X) = 1] = \mathrm{P}_\mu[T \in K] = \mathrm{P}_\mu&\left[T \notin [-c_{\neq}, c_{\neq}]\right] \\
&= \mathrm{P}_\mu\left[\frac{\sum_{i=1}^{12} X_i - 12\mu}{\sigma\sqrt{12}} < -c_{\neq} + \frac{\sqrt{12}}{\sigma}(\mu_0 - \mu)\right] \\
&\quad + \mathrm{P}_\mu\left[\frac{\sum_{i=1}^{12} X_i - 12\mu}{\sigma\sqrt{12}} > c_{\neq} + \frac{\sqrt{12}}{\sigma}(\mu_0 - \mu)\right] \\
&= \Phi\left(-c_{\neq} + \frac{\sqrt{12}}{\sigma}(\mu_0 - \mu)\right) + 1 - \Phi\left(c_{\neq} + \frac{\sqrt{12}}{\sigma}(\mu_0 - \mu)\right) \\
&= \Phi(-1.93) + 1 - \Phi(1.99) = 0.0501.
\end{aligned}$$

Das ist ein niedriger Wert, und der Grund dafür ist, dass μ sehr nahe bei μ_0 liegt.