# Wahrscheinlichkeitsrechnung und Statistik für Studierende der Informatik

Vorlesung 15

# Klausurinfo

**WANN?** **Freitag 7.3. 10.00-11.30**

**Bitte seien Sie spätestens um 9.45 am Hörsaalgebäude!**

**WO?** **LX 1205**

**LuDi:** **Sondertermine 4.-6.3.**

# Mitbringen

- Stifte (kein rot, kein Bleistift)
- Studierendenausweis

# Erlaubte Hilfsmittel

- Sie dürfen ein handschriftlich beidseitig beschriebenes DIN A4 Blatt mitnehmen.
- Ein unbeschriftetes Wörterbuch (D↔E)
- Weitere Hilfsmittel sind nicht zugelassen.
- Insbesondere ist kein Taschenrechner erlaubt!

## Weitere Hinweise

- Begründen Sie Ihr Vorgehen. Ihre Antworten müssen nur mit dem Wissen aus der Vorlesung nachvollziehbar sein.
- Es genügt meistens, wenn Sie Ihre Ergebnisse in Form von Rechenausdrücken angeben.
- Sollte am Ende eine explizite natürliche/ganze/reelle Zahl stehen, so wird dies in der Aufgabenstellung deutlich gemacht.

# Hypothesentest

**Definition 5.26.** Wir zerlegen den Parameterbereich $\Theta$ in zwei nichtleere disjunkte Teile

$$\Theta = \Theta_0 \cup \Theta_1 \ .$$

Ein *statistischer Test* oder *Hypothesentest* ist eine Entscheidungsregel, die innerhalb des vorgegebenen Modell-Rahmens für jede mögliche Stichprobe $x$ festlegt, ob man sich für die

$$\textit{Nullhypothese} \;\; H_0 : \quad \text{es gilt } \vartheta \in \Theta_0$$

oder für die

$$\textit{Gegenhypothese (Alternative)} \;\; H_1 : \quad \text{es gilt } \vartheta \in \Theta_1$$

entscheidet. Die Zerlegung von $\Theta$ impliziert eine Zerlegung

$$\mathcal{K}_0 \cup \mathcal{K}_1$$

des Stichprobenraums und damit folgende Entscheidungsregel:

$$\text{"Ist } x \in \mathcal{K}_0, \text{ so entscheide für } H_0.\text{"}$$

$$\text{"Ist } x \in \mathcal{K}_1, \text{ so entscheide für } H_1.\text{"}$$

Man sagt auch: "Zu testen ist die Hypothese $H_0$ gegen die Alternative $H_1$". $\mathcal{K}_0$ heißt *Annahmebereich* des Tests und $\mathcal{K}_1$ heißt *Verwerfungsbereich* oder auch *kritischer Bereich*. Die Hypothese $H_0$ wird häufig auch *Nullhypothese* genannt.

# Hypothes

**Definition 5.27.** Wir übernehmen die Bezeichnungen aus Definition 5.26. Gilt (in Wirklichkeit) $\vartheta \in \Theta_0$ und man entscheidet sich für die Gegenhypothese, so spricht man von einem *Fehler erster Art*. Gilt dagegen $\vartheta \in \Theta_1$ und man entscheidet sich für die Nullhypothese, so spricht man von einem *Fehler zweiter Art*. Man kann dies folgendermaßen zusammenfassen:

| | | Wirklichkeit | |
|---|---|---|---|
| | | $H_0 : \vartheta \in \Theta_0$ | $H_1 : \vartheta \in \Theta_1$ |
| Entschei-dung | für $H_0$ | richtige Entscheidung | Fehler 2. Art |
| | für $H_1$ | Fehler 1. Art | richtige Entscheidung |

Um die Wahrscheinlichkeit für eine falsche Entscheidung möglichst klein zu halten, ist eine sog. *Gütefunktion* $g$ mit kleinen Werten auf $\Theta_0$ und großen Werten auf $\Theta_1$ wünschenswert. Wir definieren $g : \Theta \to [0, 1]$ durch

$$g(\vartheta) := p_\vartheta(X \in \mathcal{K}_1) \ , \quad \vartheta \in \Theta.$$

$g$ ordnet jedem $\vartheta$ die sog. *Verwerfungswahrscheinlichkeit der Hypothese $H_0$ unter $p_\vartheta$* zu.

erster Art vor und legt $\mathcal{K}_1$ so fest, dass

$$g(\vartheta) \leq \alpha \quad \text{für alle } \vartheta \in \Theta_0$$

gilt. Ein solcher Test heißt *(Signifikanz-)Test zum (Signifikanz-)Niveau $\alpha$*. Dabei sind für $\alpha$ Werte aus dem Intervall $[0.01, 0.1]$ üblich.

# Hypothesentest

*„Wir sind jetzt mit dabei,*
*In jedem siebten Ei!"*

**Beispiel Ü-Ei: Ist wirklich in jedem siebten Ei eine lustige Figur?**

**Sei $p$ die Wahrscheinlichkeit für eine Figur im Ei.**

**Dann ist** $H_0: \; p \geq \frac{1}{7}$

$\boxed{\Theta_0 = \left[\frac{1}{7}, \infty\right)}$

**Gegenhypothese:** $H_1: \; p < \frac{1}{7}$

**Wir testen 1000 Eier. Wie viele Figuren dürfen höchstens in diesen Eiern sein, damit wir $H_0$ mit einem Signifikanzniveau von $\alpha = 0.025$ verwerfen können?**

**Sei $X$ die Zufallsvariable, die die Anzahl von Figuren misst.**
**Gesucht ist ein $c \in \{0,\ldots,1000\}$, so dass $P_p(X \leq c) \leq \alpha$ für alle $p \in \Theta_0$ gilt.**

**$X$ ist binomialverteilt mit $n = 1000$. Wenn $p \in \Theta_0$ ($H_0$ stimmt), dann ist $p \geq \frac{1}{7}$.**

**Es muss also** $P_p(X \leq c) \leq \sum_{k=0}^{c} \binom{1000}{k} \left(\frac{1}{7}\right)^k \left(\frac{6}{7}\right)^{1000-k} \leq \alpha$ **für alle** $p \geq \frac{1}{7}$ **gelten.**

# Hypothesentest

| $1-\beta$ | 0.900 | 0.950 | 0.975 | 0.990 | 0.995 |
|---|---|---|---|---|---|
| $c=\phi^{-1}(1-\beta)$ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 |

**Beispiel Ü-Ei: Ist wirklich in jedem siebten Ei eine lustige Figur?**

$$H_0: \ p \geq \frac{1}{7} \qquad H_1: \ p < \frac{1}{7}$$

**Wir testen 1000 Eier. Wie viele Figuren dürfen höchstens in diesen Eiern sein, damit wir $H_0$ mit einem Signifikanzniveau von $\alpha = 0.025$ verwerfen können?**

**Es muss also** $P_p(X \leq c) \leq \sum_{k=0}^{c} \binom{1000}{k} \left(\frac{1}{7}\right)^k \left(\frac{6}{7}\right)^{1000-k} \leq \alpha$ **gelten.**

$P_p(X \leq c) \leq \alpha$ **gilt für alle** $p \geq \frac{1}{7}$**, wenn es für** $p = \frac{1}{7}$ **gilt!**

**Approximation durch Normalverteilung:** $\Phi\left(\frac{c-np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{c-143}{\sqrt{122{,}5}}\right) \leq \alpha$

$$\Longleftrightarrow \ \Phi\left(\frac{143-c}{\sqrt{122{,}5}}\right) \geq 1-\alpha = 0{,}975 \qquad \Longleftrightarrow \ \frac{143-c}{\sqrt{122{,}5}} \geq 1{,}96$$

$$\Longleftrightarrow \ c \leq 121$$

**Bei weniger als $122$ Figuren können wir $H_0$ mit dem vorgegebenen Signifikanzniveau verwerfen.**