

Mathematik für Informatiker III

(Frank Hoffmann)

1. σ -Algebra

- (a) Beweisen Sie, dass jede endliche σ -Algebra eine gerade Anzahl von Ereignissen hat.

Lösung: Da σ -Algebren abgeschlossen gegen Komplementbildung sind, gehört mit jedem Ereignis A auch A^c zur σ -Algebra. Da für verschiedene Ereignisse A, B auch die Komplemente A^c und B^c verschieden sind, bildet die Menge der Paare $\{A, A^c\}$ eine Partition der σ -Algebra.

- (b) Kann es eine solche Algebra mit genau 6 Ereignissen geben? Begründung!

Lösung: Nein. Angenommen die σ -Algebra \mathcal{F} hätte genau 6 paarweise verschiedene Elemente: $\mathcal{F} = \{\emptyset, \Omega, A, A^c, B, B^c\}$.

Fall 1: Die Ereignisse $B \cap A$ und $B \cap A^c$ sind beide nichtleer. Dann kann es nicht sein, dass sowohl $B \cap A$ echte Teilmenge von A als auch $B \cap A^c$ echte Teilmenge von A^c ist, weil dann $B \cap A$ ein neues nichtleeres Ereignis ist, was zu \mathcal{F} gehören müsste, es aber nicht tut. Also muss A echte Teilmenge von B sein. Dies kann aber auch nicht sein, denn dann ist $B \setminus A$ nicht in \mathcal{F} .

Fall 2: Sei o.B.d.A. $B \cap A$ leer. Damit ist B echte Teilmenge von A^c und der Widerspruch besteht darin, dass $A^c \setminus B$ nicht in \mathcal{F} liegt.

2. Ziehen ohne Zurücklegen, Hypergeometrische Verteilung I

- (a) Aus einer Kiste mit N Kugeln, davon w weiße und s schwarze mit $w + s = N$, werden gleichzeitig (!) n Kugeln gezogen. Was ist die Wahrscheinlichkeit Pr_k , dass davon genau k weiß sind? Die Elementarereignisse sind also n -elementige (ungeordnete) Mengen, die wir als gleichverteilt annehmen.

Hinweis: Die Zahlen Pr_k heißen hypergeometrische Verteilung auf $\{0, \dots, n\}$.

- (b) Analysieren Sie die Wahrscheinlichkeit, dass beim Skat jeder der 3 Spieler genau ein Ass bekommt, also im Skat das 4. Ass liegt.

3. Hypergeometrisch II

In einem Teich leben N Fische. Es werden $r = 1000$ Fische gefangen und rot markiert. Nach einer Weile (N hat sich nicht verändert) werden wieder $n = 2000$ Fische gefangen und nachgezählt, dass davon $k = 100$ rot sind. Wir nehmen wieder die Voraussetzungen der hypergeometrischen Verteilung an. Daraus kann man die Größe von N schätzen. Argumentieren Sie, dass $N = 2900$ und $N = 10^6$ unrealistisch sind. Der sogenannte Maximum-Likelihood-Schätzer beruht auf der Analyse von $\frac{Pr_k(N)}{Pr_k(N-1)}$ und benutzt dasjenige N^* , für welches $Pr_k(N^*)$ maximal ist. Tun Sie dies. Zeigen Sie dazu zunächst, dass der Quotient für $Nk < nr$ größer 1 ist und ansonsten kleiner als 1.

Lösung:

Gemäß hypergeometrischer Verteilung ist

$$p_k(N) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}$$

die Wahrscheinlichkeit, aus einem Teich mit N Fischen, von denen genau r rot sind, in einem Fang von n Fischen genau k rote Fische zu finden. Nach dem Fang von 2000 Fischen, von denen 100 rot sind, wissen wir, dass der Teich mindestens 2900 Fische enthält. Angenommen, der Teich enthält tatsächlich nur 2900 Fische, d.h. wir haben zufällig alle nichtmarkierten Fische gefangen. Dann ist ein Ereignis eingetreten, das die Wahrscheinlichkeit

$$p_k(2900) = \frac{\binom{1000}{100}}{\binom{2900}{2000}} \approx 3 \cdot 10^{-639}$$

(mit MuPAD berechnet) hat, also praktisch ausgeschlossen ist. (Es ist z.B. noch etwas wahrscheinlicher, im Lotto 89 mal hintereinander 6 Richtige zu haben: $\binom{49}{6}^{-89} \approx 10^{-636}$). Eine feste Obergrenze für N gibt es nicht, aber ein sehr großes N ist ebenfalls unrealistisch: z.B. ist

$$p_k(10^6) \approx 10^{-132}.$$

Wir suchen nun ein N^* , so dass $p_k(N^*)$ möglichst groß ist. Dazu rechnen wir aus:

$$\frac{p_k(N)}{p_k(N-1)} = \frac{(N-r)(N-n)}{N(N-r-n+k)},$$

wobei $N > n + r - k = 2900$, also $N(N-n-r+k) > 0$, so dass

$$\begin{aligned} p_k(N)/p_k(N-1) < 1 &\iff (N-r)(N-n) < N(N-n-r+k) \\ &\iff N^2 - Nn - Nr + rn < N^2 - Nn - Nr + Nk \\ &\iff rn < Nk, \end{aligned}$$

und analog $p_k(N)/p_k(N-1) > 1 \iff rn > Nk$. Die Folge p_k ist also links von $\frac{rn}{k}$ monoton steigend und rechts davon monoton fallend, d.h. der maximale Wert wird bei

$$N^* = \frac{rn}{k} = 20.000$$

angenommen. Das Ergebnis stimmt also mit der Lösung der Verhältnisgleichung

$$(\text{rote Fische}) : (\text{alle Fische}) \approx k : n$$

überein.

4. False Positives

Ein Krebstest liefert bei 96% der Untersuchten ein positives Ergebnis, falls diese tatsächlich Krebs haben, bei 94% ein negatives Ergebnis, falls diese keinen Krebs haben. Bei einem Patienten XYZ, in dessen Altersgruppe 0.5% aller Patienten Krebs haben, verläuft der Test positiv. Frage: Was ist die Wahrscheinlichkeit, dass er tatsächlich Krebs hat?

Lösung: Wir definieren das Ereignis K = Patient hat Krebs und T sei Ereignis, dass Test positiv verläuft. Gesucht ist die bedingte Wahrscheinlichkeit $Pr(K|T)$. Dann ist

$$\begin{aligned} Pr(K|T) &= \frac{Pr(K \cap T)}{Pr(T)} = \frac{Pr(K \cap T)}{Pr(T|K) \cdot Pr(K) + Pr(T|K^c) \cdot Pr(K^c)} = \\ &= \frac{0.005 \cdot 0.96}{0.005 \cdot 0.96 + 0.995 \cdot 0.06} = 0.074 \end{aligned}$$

Beim Umformen des Nenners haben wir den Satz über die totale Wahrscheinlichkeit benutzt.

Fazit: Um eine recht seltene Krankheit zuverlässig mit einem Test zu erkennen, darf der Test nur sehr wenige "false positives" liefern.

5. 2 Modellierungen

Wir betrachten einen Kreis mit Radius 1, in den ein gleichseitiges Dreieck einbeschrieben ist. Was ist die Wahrscheinlichkeit, dass eine zufällige Sehne des Kreises länger ist als die Seitenlänge des Dreiecks?

1. Modellierung: Der Grundraum besteht aus allen geordneten Paaren (ϕ, ψ) mit $\phi, \psi \in [0, 2\pi)$ und $\psi \leq \phi$. Bestimmen Sie die Menge G der günstigen Elementarereignisse. Gleichverteilung angenommen berechnet sich die gesuchte Wahrscheinlichkeit als Fläche von G geteilt durch die Fläche des Grundraums.

Lösung: So ein einbeschriebenes gleichseitiges Dreieck zerlegt den Vollkreis in 3 Sektoren zu je $2\pi/3$. Zwei Winkel $\psi \leq \phi$ gehören zu G , falls man das Dreieck nicht so drehen kann, dass beide Winkel in ein und denselben Sektor fallen. Also $G = \{\phi, \psi \in [0, 2\pi) | 2\pi/3 < \psi - \phi < 4\pi/3\}$. Die Fläche von G ist $2\pi^2/3$. Der Grundraum hat die Größe $2\pi^2$, also ist $Pr(G) = 1/3$.

2. Modellierung: Die Sehne ist durch ihren Mittelpunkt eindeutig festgelegt.

Tipp: Die günstigen Fälle haben etwas zu tun mit dem in das gleichseitige Dreieck einbeschriebenen Kreis.

Lösung: Der Radius des einbeschriebenen Kreises ist $\sin \pi/6 = 1/2$. Eine Sehne gehört zum günstigen Ereignis G , falls ihr Mittelpunkt in diesem einbeschriebenen Kreis liegt. Um sich davon zu überzeugen, drehe man das gleichseitige Dreieck so, dass eine Seite parallel zur Sehne liegt. Wenn der Sehnenmittelpunkt im Kreis mit Radius $1/2$ liegt, so liegt die Sehne zwischen der Dreiecksseite und dem Kreismittelpunkt und ist somit länger als die Seite. Man beachte, dass die Strecke vom Kreismittelpunkt zum Mittelpunkt der Sehne immer senkrecht auf der Sehne steht. Die Fläche von G ist $\pi/4$, die des Grundraums (der Kreis mit Radius 1) ist π und damit ist $Pr(G) = 1/4$.

Fazit: Dieses Paradoxon lässt sich nur dadurch erklären, dass nicht beide Modellierungen ein und dasselbe Zufallsexperiment beschreiben.

6. Unabhängigkeit

Zwei faire Würfel werden gleichzeitig gespielt. Wir betrachten die 3 Ereignisse

A_1 = Der erste Würfel zeigt eine ungerade Augenzahl.

A_2 = Der zweite Würfel zeigt eine ungerade Augenzahl.

A_3 = Die Augensumme der zwei Würfel ist ungerade.

Untersuchen Sie diese Ereignisse auf paarweise bzw. auf totale Unabhängigkeit!

7. **Teilmuster** Sei P ein 0–1–String der Länge k . Zeigen Sie, dass die Wahrscheinlichkeit, dass P in einem zufälligen 0–1–String der Länge n nicht als Teilstring vorkommt, mit wachsendem n gegen 0 geht.