Vorlesung 8a

Zweistufige Zufallsexperimente

Teil 6:

Bedingte Wahrscheinlichkeiten

(Buch S. 115-117)

Definition.

Seien E_1, E_2 Ereignisse. Dann ist die

bedingte Wahrscheinlichkeit von E_2, gegeben E_1,

definiert als

$$\mathbf{P}(E_2 \,|\, E_1) := \frac{\mathbf{P}(E_2 \cap E_1)}{\mathbf{P}(E_1)}$$

... *die Wahrscheinlichkeit von E_2, wenn man schon weiß, dass E_1 eingetreten ist.*

Dies passt in den Rahmen von Teil 4,
als Verteilungsgewicht $\mathbf{P}(I_{E_2} = 1 \,|\, I_{E_1} = 1)$
der bedingten Verteilung von I_{E_2}, gegeben $\{I_{E_1} = 1\}$.

Beispiel:

T sei Geom(p)-verteilt. Für $k, \ell \in \mathbb{N}_0$ berechnen wir die bedingte Wahrscheinlichkeit von $\{T - k > \ell\}$, gegeben $\{T > k\}$.

Wegen $\{T - k > \ell\} \cap \{T > \ell\} = \{T > k + \ell\}$ ist

$$\mathbf{P}(T - k > \ell \,|\, T > k) = q^{k+\ell}/q^k = q^\ell \; .$$

Die bedingte Verteilung von $T - k$, gegeben $\{T > k\}$,

ist somit (wieder) gleich Geom(p).

Die Kenntnis, dass T größer als k ausfällt,

ändert also die Verteilung der verbleibenden Wartezeit nicht.

Man nennt diese Eigenschaft die

Gedächtnislosigkeit der geometrischen Verteilung.

Bedingte Verteilung der verbleibenden Wartezeit: Gedächtnislosigkeit der Exponentialverteilung

Für exponentialverteiltes T zum Parameter λ gilt für r, $s > 0$

$$\mathbf{P}(T > r + s \,|\, T > r) = e^{-\lambda s} \ .$$

Die bedingte Verteilung von $T - r$, gegeben $\{T > r\}$, ist somit gleich Exp(λ).

Die Kenntnis, dass T einen Wert größer als r annimmt, ändert also die Verteilung der verbleibenden Wartezeit nicht.

Für zwei Ereignisse E_1, E_2 lautet die

Formel von Bayes:

$$\mathbf{P}(E_1|E_2) = \frac{\mathbf{P}(E_1)P(E_2|E_1)}{\mathbf{P}(E_1)P(E_2|E_1) + \mathbf{P}(E_1^c)P(E_2|E_1^c)}$$

Beispiel:

In einer Population haben 0.1% eine bestimmte Krankheit.

Bei einer bestimmten Reihenuntersuchung wird

eine kranke Person in 100% der Fälle positiv getestet,

eine gesunde Person in 1%.

Wie groß ist die Wahrscheinlichkeit, dass

eine positiv getestete Person wirklich krank ist?

Hier ist erst einmal ein Rezept

für eine intuitive Überschlagsrechnung:

In einer Population von 1000

sind 999 gesund und einer krank. Von den 999 Gesunden

werden ca 10 (falsch) positiv diagnostiziert.

Von 11 positiv Diagnostizierten

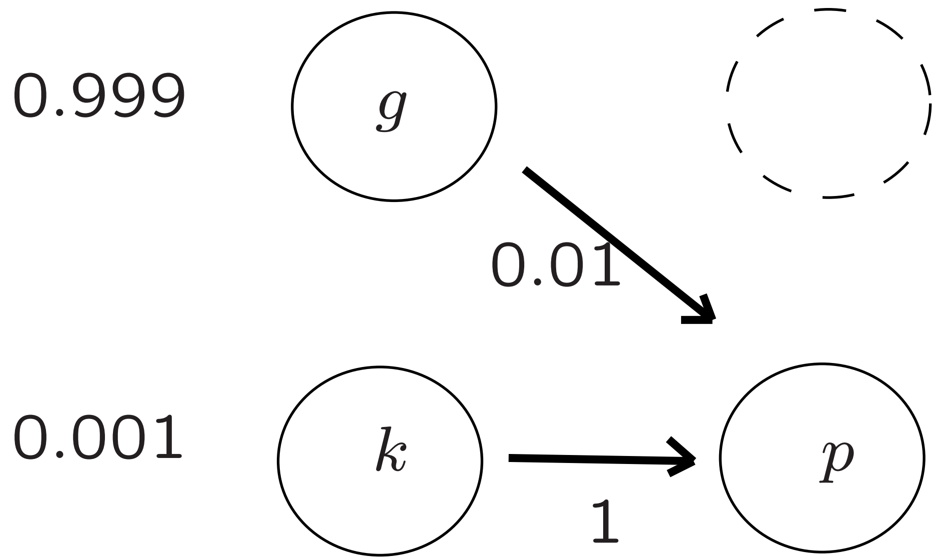
ist also im Schnitt nur einer krank.

Hier ist eine **Formalisierung:**

X_1 sei der Gesundheitszustand (mit Werten in $S_1 = \{g, k\}$),

X_2 der Testbefund (mit Werten in $S_2 = \{p, n\}$).

(X_1, X_2) entsteht über ein zweistufiges Experiment:



$$\mathbf{P}(X_1 = k \,|\, X_2 = p) = \frac{\mathbf{P}(X_1 = k, X_2 = p)}{\mathbf{P}(X_2 = p)}$$

$$= \frac{0.001 \cdot 1}{0.999 \cdot 0.01 + 0.001 \cdot 1} \approx \frac{1}{11}$$