§ 8 Unabhängigkeit, Verteilungsfunktion, Erwartungswert und Varianz

Im Folgenden sind immer $X_1, \dots, X_n$ Zufallsvariablen mit

$$X_i : \Omega \longrightarrow W, \qquad W \subseteq \mathbb{R}, \qquad 1 \leq i \leq n.$$

<u>Definition</u> (Unabhängigkeit von Zufallsvariablen)

Die Zufallsvariablen $X_1, \dots, X_n$ heißen stochastisch unabhängig, falls für alle Intervalle $A_1, \dots, A_n \subseteq W$ gilt:

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

$$= P(\{\omega \in \Omega \mid X_1(\omega) \in A_1, \dots, X_n(\omega) \in A_n\})$$

<u>Bemerkung</u> Wenn $X_1, \dots, X_n$ endliche oder diskrete Zufallsvariablen sind, dann sind sie stoch. unabhängig, genau dann wenn

$$P(X_1 = a_1, \dots, X_n = a_n) = \prod_{i=1}^{n} P(X_i = a_i)$$

für alle $a_1, \dots, a_n \in W$.

<u>Notation</u> Sei $X : \Omega \to W$ eine Zufallsvariable.

Falls X gleichverteilt auf $W = \{1, \dots, n\}$ ist, schreiben wir

$$X \sim \mathrm{Unif}(\{1, \dots, n\})$$

Falls X die Binomialverteilung mit Parametern n, p hat, schreiben wir

$$X \sim \mathrm{Bin}(n, p) \qquad \left(\text{d.h. } P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}\right)$$

Falls X die Poissonverteilung mit Parameter $\lambda > 0$ hat, schreiben wir

$$X \sim Poi(\lambda) \qquad \left(\text{d.h. } P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!} \right).$$

Satz

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen. Dann gilt:

(a) Falls $X_i \sim Bin(1,p)$ $\Rightarrow$ $X_1 + \ldots + X_n \sim Bin(n,p)$

(b) Falls $X_i \sim Poi(\lambda)$ $\Rightarrow$ $X_1 + \ldots + X_n \sim Poi(n \cdot \lambda)$

Definition

Sei X eine Zufallsvariable. Dann definieren wir

$$F(a) := P(X \le a) = P(\{\omega \in \Omega \mid X(\omega) \le a\})$$

$F(a)$ heißt (kumulative) Verteilungsfunktion von X.

Bemerkung

Die Verteilungsfunktion F einer Zufallsvariable X charakterisiert ihre Verteilung vollständig. D.h.:

Seien $a < b$ reelle Zahlen. Dann gilt:

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a)$$

$$P(a \le X \le b) = P(X \le b) - P(X < a) = F(b) - F(a-)$$

$$P(a \le X < b) = P(X < b) - P(X < a) = F(b-) - F(a-)$$

$$P(a < X < b) = P(X < b) - P(X \le a) = F(b-) - F(a)$$

wobei: $F(a-) = \lim\limits_{x \nearrow a} F(x)$.

Ist X endlich oder diskrete mit Wertebereich $\Omega = \{a_1, a_2, a_3, \dots\}$
Dann gilt:

$$F(a) = \sum_{i:\, a_i \leq a} P(X = a_i)$$

<u>Definition</u> (Erwartungswert einer diskreten Zufallsvariable)
Sei X eine endliche oder diskr. Zufallsvariable mit Wertebereich $\Omega = \{a_1, a_2, \dots\}$
Der Erwartungswert von X ist

$$\mathbb{E}(X) := \sum_i a_i \cdot P(X = a_i).$$

<u>Bemerkung</u> Seien $x_1, \dots, x_N$ Beobachtungen eines <u>Merkmals</u> mit Wertebereich $\{a_1, \dots, a_M\}$ (wie in Vorlesungen 1+2), und sei $f_j := \frac{1}{N} \# \{ i \mid x_i = a_j \}$.
Dann definiert $f = (f_1, \dots, f_M)$ eine Zufallsverteilung auf dem Wertebereich $\{a_1, \dots, a_M\}$ durch $P(X = a_i) := f_i$.
Dann gilt:

$$\text{Mittelwert } \frac{1}{N}(x_1 + \dots + x_N) = \mathbb{E}(X).$$

<u>In anderen Worten</u>: falls die f_i die Verteilung einer gesuchten Zufallsvariable X approximieren, so dass $P(X = a_i) \approx f_i$. $\Rightarrow$ Mittelwert $\approx \mathbb{E}(X)$.

Beispiel (Pokemon - Sammelkarten)

Frage: Wieviele Karten muss ein Sammler im Mittel kaufen, um eine Serie von $n = 150$ Karten zu erhalten.

Annahmen: (1) Die Karten werden einzeln gekauft.

(2) Jede Karte hat die gleiche Wahrscheinlichkeit gezogen.

Unter den Annahmen:

$$\Omega = \{ (\omega_1, \omega_2, \omega_3, \dots) \mid 1 \leq \omega_i \leq n \}$$

⇝ Urnenmodell mit Zurücklegen und mit Beachtung der Reihenfolge.

Wir definieren:

$X_i :=$ Anzahl Karten die gekauft werden müssen, um eine neue Karte zu erhalten, nachdem bereits $(i-1)$ verschiedene Karten gezogen wurden.

⇝ $P(X_i = k) = \left(\frac{i-1}{n}\right)^{k-1} \cdot \frac{n-(i-1)}{n}$, d.h. X_i ist geometrisch verteilt mit Parameter $p = \frac{n-i+1}{n}$

Die Zufallsvariable $X := X_1 + X_2 + \dots + X_n$ misst die Anzahl der Karten, die gezogen werden, um alle Karten zu bekommen.

Die Anzahl der Karten, die im Mittel gezogen werden muss ist

$$\mathbb{E}(X).$$

Satz

Sei $X \sim \mathrm{Geo}(p)$. Dann ist $\mathbb{E}(X) = \frac{1}{p}$.

<u>Satz</u> (Eigenschaften von Erwartungswerten)

Seien X, Y Zufallsvariablen.

(1) $\mathbb{E}(aX+b) = a \cdot \mathbb{E}(X)+b$, für $a, b \in \mathbb{R}$.

(2) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

(3) Falls X und Y unabhängig sind, gilt: $\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\mathbb{E}(Y)$

Im Allgemeinen gilt (3) nicht!

<u>Zurück zum Beispiel:</u>

Erinnerung:
- $\mathbb{E}(X)$ ist die erwartete Anzahl Karten, die gekauft werden müssen, um alle n Karten zu erhalten.
- $X = X_1 + \dots + X_n$
- $X_i \sim \mathrm{Geo}\left(\frac{n-i+1}{n}\right)$

Es gilt:
$$\mathbb{E}(X) = \mathbb{E}(X_1 + \dots + X_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)$$
$$= \frac{n}{n} + \frac{n}{n-1} + \dots + \frac{n}{1}$$
$$= \sum_{i=1}^{n} \frac{n}{n-i+1}$$
$$= n \cdot \underbrace{\sum_{i=1}^{n} \frac{1}{n-i+1}}_{=: H_n}$$

$\Rightarrow \quad \mathbb{E}(X) = n \cdot H_n.$

Für $n = 150$ haben wir $H_n = 5.6$.

$\rightsquigarrow$ Man muss 5.6 mal mehr Karten kaufen, um alle $n=150$ zu sammeln.

Definition (Varianz)

Die Varianz einer Zufallsvariablen X ist definiert als

$$\mathrm{Var}(X) := \mathbb{E}\left((X-\mathbb{E}(X))^2\right).$$

Die Standardabweichung von X ist $\sqrt{\mathrm{Var}(X)}$.

Satz

Seien X, Y Zufallsvariablen

(1) $\mathrm{Var}(X) \geq 0$

(2) $\mathrm{Var}(aX+b) = a^2 \cdot \mathrm{Var}(X)$ für $a,b \in \mathbb{R}$.

(3) Falls X und Y unabhängig sind, gilt: $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

(4) $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

(5) Tschebycheff - Ungleichung: $P(|X-\mathbb{E}(X)| \geq \varepsilon) \leq \dfrac{\mathrm{Var}(X)}{\varepsilon^2}$

Gesetz der großen Zahlen

Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen, die alle die gleiche Verteilung haben (wir sagen, dass die X_i iid sind). Sei $\mu := \mathbb{E}(X_1) = \mathbb{E}(X_2) = \ldots$. Dann gilt:

$$P\left(\left\{\omega \in \Omega \;\middle|\; \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} X_i(\omega) = \mu\right\}\right) = 1$$

$\rightsquigarrow$ Mit Wahrscheinlichkeit $=1$ strebt der empirische Mittelwert gegen μ.