

§ 8 Unabhängigkeit, Verteilungsfunktion, Erwartungswert und Varianz

Im Folgenden sind immer X_1, \dots, X_n Zufallsvariablen mit
 $X_i : \Omega \rightarrow W, \quad W \subseteq \mathbb{R}, \quad 1 \leq i \leq n.$

Definition (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:

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

\uparrow

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

Bemerkung 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 $a_1, \dots, a_n \in W$.

Notation Sei $X : \Omega \rightarrow W$ eine Zufallsvariable.

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

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

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

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

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

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

Satz

Seien X_1, \dots, X_n unabhängige Zufallsvariablen. Dann gilt:

- (a) Falls $X_i \sim \text{Bin}(1, p)$ $\Rightarrow X_1 + \dots + X_n \sim \text{Bin}(n, p)$
- (b) Falls $X_i \sim \text{Poi}(\lambda)$ $\Rightarrow X_1 + \dots + X_n \sim \text{Poi}(n \cdot \lambda)$

Definition

Sei X eine Zufallsvariable. Dann definieren wir

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

$F(a)$ 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:

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

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

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

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

wobei: $F(a^-) = \lim_{x \nearrow a} F(x).$

Ist X endlich oder diskrete mit Wertebereich $\{a_1, a_2, a_3, \dots\}$

Dann gilt:

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

Definition (Erwartungswert einer diskreten Zufallsvariable)

Sei X eine endliche oder diskr. Zufallsvariable mit Wertebereich $\{a_1, a_2, \dots\}$

Der Erwartungswert von X ist

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

Bemerkung Seien x_1, \dots, x_N Beobachtungen eines Merkmals 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:

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

In anderen Worten: 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:

$$\Omega = \{ (w_1, w_2, w_3, \dots) \mid 1 \leq w_i \leq n \}$$

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

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

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

Satz

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

Satz (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!

Zurück zum Beispiel:

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

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

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

Für $n=150$ haben wir $H_n = 5.6$.

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

Definition (Varianz)

Die **Varianz** einer Zufallsvariablen X ist definiert als

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

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

Satz

Seien X, Y Zufallsvariablen

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

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

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

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

(5) Tschbyeff - Ungleichung: $P(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}$

Gesetz der grossen Zahlen

Sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen, die alle die gleiche Verteilung haben (wir sagen, dass die X_i iid sind).

Sei $\mu := \mathbb{E}(X_1) = \mathbb{E}(X_2) = \dots$. Dann gilt:

$$P(\{\omega \in \Omega \mid \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i(\omega) = \mu\}) = 1$$

→ Mit Wahrscheinlichkeit = 1 strebt der empirische Mittelwert gegen μ .