

Mathematik II für Biologen  
Zufallsvariable: Verteilungen & Kennzahlen

Stefan Keppeler

4. Juni 2008

## Zufallsvariable

Definition

Kennzahlen: Erwartungswert

Kennzahlen: Varianz

Kennzahlen: Erwartungstreue

Verteilungsfunktion

## Poisson-Verteilung

## Stetige Zufallsvariable

Definition

Kennzahlen

**Definition:** Eine Funktion  $X : \Omega \rightarrow \mathbb{R}$  heißt **Zufallsvariable** (Zufallsgröße). Sie heißt **diskret**, falls sie nur endlich viele oder abzählbar unendlich viele Werte annehmen kann.

**Bemerkung:** Jede Teststatistik ist eine Zufallsvariable, z.B. Spermasexing:  $\Omega = \{\sigma, \varphi\}^{12}$ ,  $X = \#\varphi$ , d.h. z.B.  $X(\varphi, \varphi, \varphi, \varphi, \sigma, \varphi, \varphi, \varphi, \varphi, \sigma, \varphi, \varphi) = 10$

**Kennzahlen:** Erwartungswert  $E[X]$ , Varianz  $\text{Var}[V]$ , Standardabweichung, Median...

**Vorsicht:** Nicht verwechseln mit den Begriffen für Stichproben! (siehe auch später...)

**Definition:** Der **Erwartungswert** einer diskreten Zufallsvariablen  $X$  ist

$$E[X] := \sum_k k \cdot P[X = k]$$

**Beispiele:**

- ▶  $X$ : Ergebnis eines fairen Würfelwurfs

$$E[X] = \sum_{k=1}^6 k \cdot \frac{1}{6} = \frac{1 + 2 + \dots + 6}{6} = 3,5$$

- ▶  $X \sim \text{Bin}(n, p)$ :  $E[X] = np$

**Beweis:** 

**Linearität** ( $X, Y$ : Zufallsvariablen,  $a, b \in \mathbb{R}$ )

$$E[aX + bY] = aE[X] + bE[Y],$$

insbesondere  $E[a + bX] = a + bE[X]$ .

**Beweis:** 



**Definition:** Die (theoretische) **Varianz** einer diskreten Zufallsvariablen  $X$  ist

$$\text{Var}(X) := E[(X - E[X])^2]$$

und ihre **Standardabweichung** ist  $\sigma(X) := \sqrt{\text{Var}(X)}$ .

Es gilt:  $\text{Var}(X) = E[X^2] - (E[X])^2$

**Beweis:** 

**Beispiele:**

- Wurf eines fairen Würfels

$$\begin{aligned}\text{Var}(X) &= \sum_{k=1}^6 (k - \frac{7}{2})^2 \cdot \frac{1}{6} = \frac{(-\frac{5}{2})^2 + (-\frac{3}{2})^2 + \dots + (\frac{5}{2})^2}{6} \\ &= \frac{35}{12} \approx 2,9 \quad \Rightarrow \quad \sigma(X) \approx 1,7\end{aligned}$$

- $X \sim \text{Bin}(n, p)$ :  $\text{Var}(X) = np(1 - p)$



- ▶ Stichprobe:  $\{x_1, \dots, x_n\}$
- ▶ Betrachte die  $x_i$  als Zufallsvariablen  $X_i$ , die alle gleich verteilt sind, insbesondere

$$E[X_i] = E[X] \quad \text{und} \quad \text{Var}(X_i) = \text{Var}(X) \quad \forall i = 1, \dots, n.$$

- ▶ Dann sind der Mittelwert  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  und

$$\text{die empirische Varianz } s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

erwartungstreu, d.h.

$$E[\bar{x}] = E[X] \quad \text{und} \quad E[s_x^2] = \text{Var}(X).$$

Beweis: 

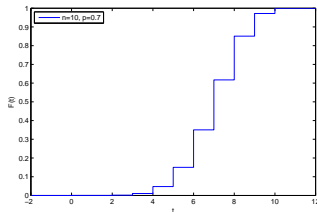
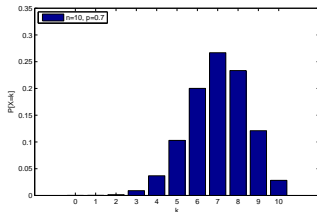
**Definition:** Die (theoretische kumulative) **Verteilungsfunktion**  $F : \mathbb{R} \rightarrow [0, 1]$  einer Zufallsvariable  $X$  ist gegeben durch

$$F(t) := P[X \leq t]$$

## Beispiele:

▶ Wurf eines fairen Würfels 

▶  $X \sim B(10; 0,7)$



**Definition:** Eine Zufallsvariable  $X : \Omega \rightarrow \mathbb{N}_0$  heißt **Poisson-verteilt** mit Parameter  $\lambda > 0$  falls

$$P[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Man schreibt  $X \sim \text{Pois}(\lambda)$

**Eigenschaften:**  $E[X] = \lambda$ ,  $\text{Var}(X) = \lambda$

**Beispiele:**

- ▶  $X = \#$  radioaktive Zerfälle pro Sekunde (für 1kg Plutonium)
- ▶  $X = \#$  Mutationen in festem Zeitraum
- ▶  $X = \#$  Erdbeben in festem Zeitraum

**Grund:**  $\text{Bin}(n, p = \frac{\lambda}{n}) \xrightarrow{n \rightarrow \infty} \text{Pois}(\lambda)$





## Beispiel:

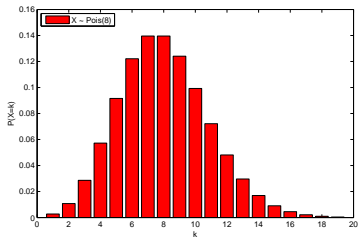
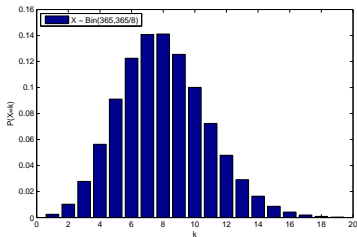
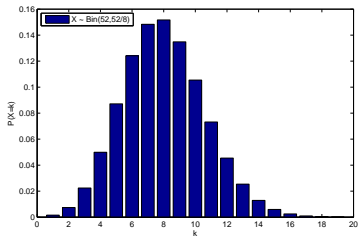
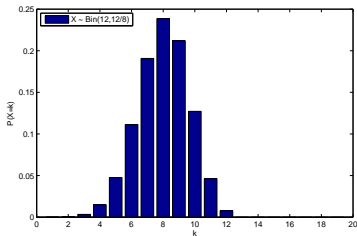
- ▶ In Schüttelhausen gibt es im Schnitt 8 Erdbeben pro Jahr.
- ▶ Annahme: Beben zu jeder Zeit gleich wahrscheinlich
- ▶ Wie ist  $X$ , die Anzahl der Erdbeben pro Jahr, verteilt?
- ▶ Wir möchten:  $E[X] = 8$

Teile Jahr in gleich große Zeitabschnitte:

Wahrscheinlichkeit für Beben

- ▶ im Januar:  $\frac{8}{12}$ , im Februar:  $\frac{8}{12} \dots \rightsquigarrow X \sim \text{Bin}(12, \frac{8}{12})$
  - ▶ in Woche 1:  $\frac{8}{52}$ , in Woche 2:  $\frac{8}{52} \dots \rightsquigarrow X \sim \text{Bin}(52, \frac{8}{52})$
  - ▶ am 1. Januar:  $\frac{8}{365}$ , am 2. Januar:  $\frac{8}{365} \dots \rightsquigarrow X \sim \text{Bin}(365, \frac{8}{365})$
- ...stündlich, minütlich... immer feiner  $\rightsquigarrow X \sim \text{Pois}(8)$





Nicht alle Zufallsvariablen sind diskret (“zählen” etwas), manche können beliebige reelle Werte annehmen, daher...

**Definition:** Eine Zufallsvariable  $X$  heißt **stetig verteilt**, falls eine stetige und diffbare Funktion  $F_X : \mathbb{R} \rightarrow \mathbb{R}_0^+$ , die **Verteilungsfunktion**, existiert mit

$$P[a \leq X \leq b] = F_X(b) - F_X(a).$$

Die Ableitung  $f_X := F'_X$  heißt **Dichte** von  $X$

**Zusammenhang:**

$$F_X(t) = \int_{-\infty}^t f_X(s) ds,$$

$$P[a \leq X \leq b] = \int_a^b f_X(s) ds = F_X(b) - F_X(a).$$

Skizze:



Jede Dichte erfüllt

- ▶  $f_X(t) \geq 0 \quad \forall t \in \mathbb{R}$
- ▶  $\int_{-\infty}^{+\infty} f_X(s) ds = 1$

Der **Erwartungswert** einer stetigen Zufallsvariable  $X$  mit Dichte  $f_X$  ist definiert als

$$E[X] := \int_{-\infty}^{+\infty} t f_X(t) dt,$$

die **Varianz** als

$$\text{Var}(X) := \int_{-\infty}^{+\infty} (t - E[X])^2 f_X(t) dt = E[X^2] - (E[X])^2,$$

und der **Median** erfüllt  $F_X(\text{med}) = \frac{1}{2}$ .