Mathematik II für Biologen

Zufallsvariable: Verteilungen & Kennzahlen

Stefan Keppeler

12. Juni 2015

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Definition: Eine Funktion $X : \Omega \to \mathbb{R}$ 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: $\Omega = \{♂, ♀\}^{12}$, $X = \#♀$,
d.h. z.B. $X(♀, ♀, ♀, ♀, ♂, ♀, ♀, ♀, ♀, ♂, ♀, ♀) = 10$

Kennzahlen: Erwartungswert $E[X]$, Varianz $\mathrm{Var}(X)$, Standardabweichung, Median...

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

Definition: Der Erwartungswert einer diskreten Zufallsvariable X ist

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

Beispiele:

- X: Ergebnis eines fairen Würfelwurfs

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

- $X \sim \mathrm{Bin}(n, p)$: $\quad E[X] = np$ **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

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

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

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

Beispiele:

- Wurf eines fairen Würfels

$$\mathrm{Var}(X) = \sum_{k=1}^{6} (k - \tfrac{7}{2})^2 \cdot \frac{1}{6} = \frac{(-\frac{5}{2})^2 + (-\frac{3}{2})^2 + \ldots + (\frac{5}{2})^2}{6}$$

$$= \frac{35}{12} \approx 2{,}9 \qquad \Rightarrow \quad \sigma(X) \approx 1{,}7$$

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

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

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

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

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

erwartungstreu, d.h.

$$E[\overline{x}] = E[X] \qquad \text{und} \qquad E[s_x^2] = \mathrm{Var}(X)\,.$$

Beweis:

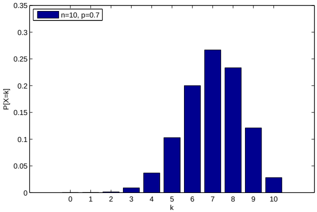
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

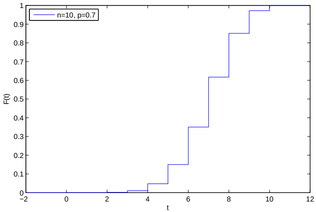
Definition: Die (theoretische kumulative) Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ einer Zufallsvariable X ist gegeben durch

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

Beispiele:

- Wurf eines fairen Würfels
- $X \sim B(10; 0{,}7)$





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

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

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

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

Beispiele:

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

Grund: $\mathrm{Bin}(n, p = \frac{\lambda}{n}) \xrightarrow[n \to \infty]{} \mathrm{Pois}(\lambda)$

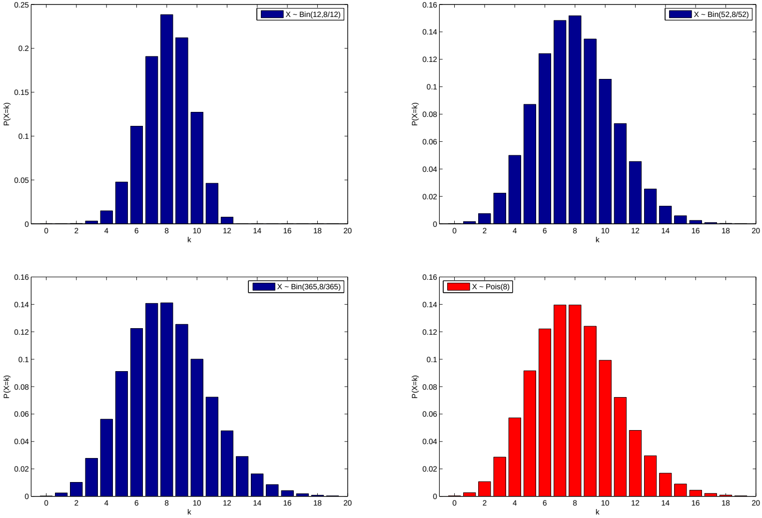
Beispiel:

- 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 benötigen: $E[X] = 8$

Teile Jahr in gleich große Zeitabschnitte:
Wahrscheinlichkeit für Beben

- im Januar: $\frac{8}{12}$, im Februar: $\frac{8}{12}$... $\rightsquigarrow X \sim \mathrm{Bin}(12, \frac{8}{12})$
- in Woche 1: $\frac{8}{52}$, in Woche 2: $\frac{8}{52}$... $\rightsquigarrow X \sim \mathrm{Bin}(52, \frac{8}{52})$
- am 1. Januar: $\frac{8}{365}$, am 2. Januar: $\frac{8}{365}$... $\rightsquigarrow X \sim \mathrm{Bin}(365, \frac{8}{365})$

...stündlich, minütlich... immer feiner $\rightsquigarrow X \sim \mathrm{Pois}(8)$


X ~ Bin(12,8/12)
X ~ Bin(52,8/52)
X ~ Bin(365,8/365)
X ~ Pois(8)
P(X=k)
k

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

Beispiel: "Wartezeit" in Schüttelhausen

Definition: Eine Zufallsvariable X heißt stetig verteilt, falls eine stetige und diffbare Funktion $F_X : \mathbb{R} \to \mathbb{R}_0^+$, 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:

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

$$P[a \leq X \leq b] = \int_a^b f_X(s)\,\mathrm{d}s = F_X(b) - F_X(a)\,.$$

Skizze

Beispiel: ...zurück nach Schüttelhausen:

- $X = \#$ Erdbeben in einem Jahr, $E[X] = 8 =: \lambda$,
  also $X \sim \mathrm{Pois}(\lambda)$, d.h. $P[X = k] = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$

gilt auch für andere Zeiträume:

- $X_1 = \#$ Beben in einem halben Jahr, $E[X] = \frac{\lambda}{2} = 4$
  also $X_1 \sim \mathrm{Pois}(\frac{\lambda}{2})$, d.h. $P[X_1 = k] = \frac{(\frac{\lambda}{2})^k}{k!}\mathrm{e}^{-\frac{\lambda}{2}}$
- $X_2 = \#$ Beben in fünf Jahren, $E[X] = 5\lambda$
  also $P[X_2 = k] = \frac{(5\lambda)^k}{k!}\mathrm{e}^{-5\lambda}$
- $Y = \#$ Beben in Zeitintervall der Länge t (in Jahren)
  $E[Y] = \lambda t$, also $P[Y = k] = \frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t}$

$P[Y \geq 1] = 1 - P[Y = 0] = 1 - \mathrm{e}^{-\lambda t}$
$=$ Wahrscheinlichkeit für mindestens ein Beben im Zeitraum der Länge t
$=$ Wahrscheinlichkeit dafür, dass Wartezeit T höchtens t
$= P[T \leq t] = F_T(t)$ und damit auch $f_T(t) = \lambda\mathrm{e}^{-\lambda t}$

EBERHARD KARLS UNIVERSITÄT TÜBINGEN