Mathematik II für Biologen

Stetige Verteilungen, Unabhängigkeit & ZGS

Stefan Keppeler

14. Juni 2013

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

stetig verteilte Zufallsvariable X

- Verteilungsfunktion $F_X : \mathbb{R} \to \mathbb{R}_0^+$, stetig und diffbar
- Dichte $f_X := F_X'$

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

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

Jede Dichte erfüllt

- $f_X(t) \geq 0 \quad \forall\, t \in \mathbb{R}$ (damit F_X monoton wachsend)
- $\int_{-\infty}^{+\infty} f_X(s)\,\mathrm{d}s = 1$ (Normierung)

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

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

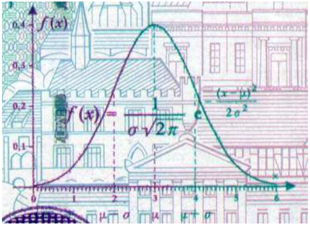
die Varianz als

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

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

Definition: X heißt normalverteilt mit Parametern $\mu \in \mathbb{R}$ und σ^2 ($X \sim \mathcal{N}(\mu, \sigma^2)$) falls X die Dichte

$$f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}} \qquad \text{hat.}$$



- **Skizze:**
- Beispiele für normalverteilte Größen: später!

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Eigenschaften der Normalverteilung

- Falls $X \sim \mathcal{N}(\mu, \sigma^2)$, so gilt $E[X] = \mu$ und $\mathrm{Var}(X) = \sigma^2$.
- Keine elementare Formel für F_X
- Einige wichtige Werte der Verteilungsfunktion Φ der Standardnormalverteilung $\mathcal{N}(0,1)$:

$$\Phi(0) = \tfrac{1}{2} \qquad \Phi(1) \approx 0{,}84$$
$$\Phi(1{,}64) \approx 95\% \qquad \Phi(1{,}96) \approx 97{,}5\%$$

  (siehe Tabellen oder MATLAB-Befehl `normcdf`)
- Standardisisierung:

  Falls $X \sim \mathcal{N}(\mu, \sigma^2)$ so gilt $Z := \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

  Beispiel: $X \sim \mathcal{N}(2,5)$, $P[X \geq 7] =?$

Definition: X mit Wertebereich $[0,\infty)$ heißt exponentialverteilt mit Parameter $\alpha > 0$ ($X \sim \mathrm{Expo}(\alpha)$), falls

$$f_X(t) = \begin{cases} \alpha\, \mathrm{e}^{-\alpha t} & \text{für} \quad t \geq 0 \\ 0 & \text{für} \quad t < 0 \end{cases}$$

und damit

$$F_X(t) = \begin{cases} 1 - \mathrm{e}^{-\alpha t} & \text{für} \quad t \geq 0 \\ 0 & \text{für} \quad t < 0 \end{cases} .$$

Skizze:

- Falls $X \sim \mathrm{Expo}(\alpha)$, so ist $E[X] = \frac{1}{\alpha}$ und $\mathrm{Var}(X) = \frac{1}{\alpha^2}$.
- **Beispiele:** $X =$ "Wartezeit" bis zum/zur nächsten
  - Zerfall (vgl. ÜA 30, WS 12/13),
  - Erdbeben (vgl. Schüttelhausen, letze Woche),
  - Mutation,

  also $X =$ "Lebensdauer".

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

...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 Z höchstens t
$= P[Z \leq t] = F_Z(t)$ also $Z \sim \mathrm{Expo}(\lambda)$

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Definition: X mit Wertebereich $[a, b]$ ($a < b$) heißt uniform verteilt (gleichverteilt) auf $[a, b]$, falls

$$f_X = \begin{cases} \frac{1}{b-a} & \text{für } a \leq t \leq b \\ 0 & \text{sonst} \end{cases}$$

und damit

$$F_X(t) = \begin{cases} 0 & \text{für } t < a \\ \frac{t-a}{b-a} & \text{für } a \leq t \leq b \\ 1 & \text{für } t > b \end{cases} .$$

Skizze:

Beispiel:
$X =$ Rundungsfehler üblicherweise gleichverteilt auf $[0, 1]$
Anwendung:
MATLAB-Befehl `rand` erzeugt in $[0, 1]$ gleichverteilte Zufallszahlen.

Definiton: Zwei Zufallsvariablen heißen unabhängig, falls $\forall\ A, B \subseteq \mathbb{R}$ gilt

$$P[\{X \in A\} \cap \{Y \in B\}] = P[\{X \in A\}] \cdot P[\{Y \in B\}]$$

Beispiel:

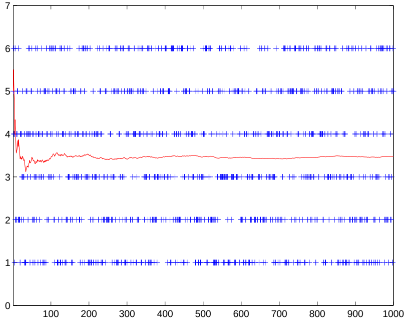
- $X =$ Ergebnis eines Würfelwurfs
- $Y =$ Temperatur
- $Z =$ Niederschlagsmenge

X und Y sind unabhängig
X und Z sind unabhängig
Y und Z sind nicht unabhängig

Aussage: Der Durchschnitt $\overline{X} = \frac{1}{n}(X_1 + \ldots + X_n)$ vieler unabhängiger Zufallsvariablen $X_1, \ldots, X_n$, die alle dieselbe Verteilung haben (iid[1] – viele gleichartige aber unabhängige Messungen), d.h. insbesondere $E[X_j] = E[X_k]\ \forall j, k$, strebt für $n \to \infty$ gegen $E[X_j]$.

Beispiel:
n-facher Würfelwurf
MATLAB-Code

```
N=1000;
n=1:N;
X=unidrnd(6,1,N);
Xquer=cumsum(X)./n;
fig=plot(n,X,'+')
hold on
plot(n,Xquer,'r')
hold off
```



[1] independent and identically distributed

Rechenregeln: X, Y Zufallsvariablen, $c \in \mathbb{R}$
Es gelten immer:

- $E[cX] = cE[X]$
- $E[X + Y] = E[X] + E[Y]$
- $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$

Falls X, Y unabhängig sind, gilt außerdem

- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Falls X, Y unabhängig und

- $X \sim \mathrm{Bin}(n, p)$, $Y \sim \mathrm{Bin}(m, p) \Rightarrow X + Y \sim \mathrm{Bin}(n + m, p)$ (laut Definition!)
- $X \sim \mathrm{Pois}(\lambda)$, $Y \sim \mathrm{Pois}(\mu) \Rightarrow X + Y \sim \mathrm{Pois}(\lambda + \mu)$
- $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$

  $\Rightarrow X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Merke:

Die Summe unabhängig $\left\{\begin{matrix} \mathrm{Bin}(\cdot, p) \\ \mathrm{Pois}(\cdot) \\ \mathcal{N}(\cdot,\cdot) \end{matrix}\right\}$-verteilter Zufallsvariabler

ist wieder $\left\{\begin{matrix} \mathrm{Bin}(\cdot, p) \\ \mathrm{Pois}(\cdot) \\ \mathcal{N}(\cdot,\cdot) \end{matrix}\right\}$-verteilt.

Aussage des Zentralen Grenzwertsatzes:
Die Summe vieler unabhängiger Zufallsvariabler, die alle die gleiche Verteilung haben (also iid), ist ungefähr normalverteilt.

Beispiele:

- $\mathrm{Bin}(n,p) \approx \mathcal{N}(np, np(1-p))$ für große n

  Begründung:

  Anwendung:
  Faustregel für Annahmebereich
  beim (zweiseitigen) Binomialtest

- $\mathrm{Pois}(\lambda) \approx \mathcal{N}(\lambda, \lambda)$ falls λ groß