

Mathematik II für Biologen

Stetige Verteilungen, Unabhängigkeit & ZGS

Stefan Keppeler

11. Juni 2010

Stetige Zufallsvariable

Wiederholung

Eigenschaften & Kennzahlen

Normalverteilung

Definition

Eigenschaften, Standardisierung

Exponentialverteilung

Zusammenhang von Poisson- & Exponentialverteilung

Gleichverteilung

Unabhängigkeit

Gesetz der großen Zahlen

Rechenregeln

Zentraler Grenzwertsatz

stetig verteilte Zufallsvariable X

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

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

- ▶ Wahrscheinlichkeiten

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

Jede Dichte erfüllt

▶ $f_X(t) \geq 0 \quad \forall t \in \mathbb{R}$ (damit F_X monoton wachsend)

▶ $\int_{-\infty}^{+\infty} f_X(s) ds = 1$ (Normierung)

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

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

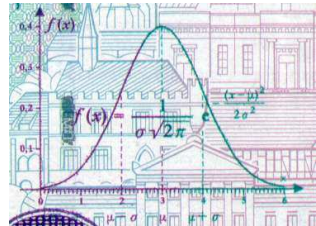
die **Varianz** als

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

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

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

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



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

Eigenschaften der Normalverteilung

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

$$\Phi(0) = \frac{1}{2}$$

$$\Phi(1) \approx 0,84$$

$$\Phi(1,64) \approx 95\%$$

$$\Phi(1,96) \approx 97,5\%$$

(siehe Tabellen oder MATLAB-Befehl `normcdf`)

- ▶ **Standardisierung:**

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

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

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

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

und damit

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

Skizze: 

- ▶ Falls $X \sim \text{Expo}(\alpha)$, so ist $E[X] = \frac{1}{\alpha}$ und $\text{Var}(X) = \frac{1}{\alpha^2}$.
- ▶ **Beispiele:** $X =$ “Wartezeit” bis zum/zur nächsten
 - ▶ Zerfall (vgl. ÜA 28, WS 09/10),
 - ▶ Erdbeben (vgl. Schüttelhausen, letzte Woche),
 - ▶ Mutation,

also $X =$ “Lebensdauer”.



...zurück nach Schüttelhausen:

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

gilt auch für andere Zeiträume:

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

$$P[Y \geq 1] = 1 - P[Y = 0] = 1 - e^{-\lambda t}$$

= Wahrscheinlichkeit für mindestens ein Beben im Zeitraum der Länge t

= Wahrscheinlichkeit dafür, dass Wartezeit Z höchstens t

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



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

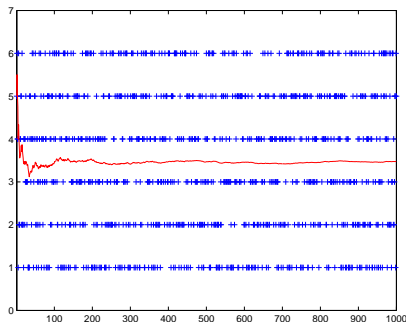
X und Z sind unabhängig

Y und Z sind nicht unabhängig

Aussage: Der Durchschnitt $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ vieler unabhängiger Zufallsvariablen X_1, \dots, X_n , die alle dieselbe Verteilung haben (iid¹ – viele gleichartige aber unabhängige Messungen), d.h. insbesondere $E[X_j] = E[X_k] \forall j, k$, strebt für $n \rightarrow \infty$ 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
```



¹independent and identically distributed



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

Es gelten **immer**:

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

Falls X, Y **unabhängig** sind, gilt außerdem

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

Falls X, Y **unabhängig** und

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

Merke:

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

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

Aussage des Zentralen Grenzwertsatzes:

Die Summe vieler unabhängiger Zufallsvariablen, die alle die gleiche Verteilung haben (also iid), ist ungefähr normalverteilt.

Beispiele:

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

Begründung:



Anwendung:

Faustregel für Annahmehbereich
beim (zweiseitigen) Binomialtest



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

