Stochastik für Data Science

Wahrscheinlichkeitsverteilungen

Prof. Dr. Wolfgang Bischof

Augsburg

27. Februar 2024

Inhaltsverzeichnis

I Spezielle Verteilungen 2
1 Wiederholung 2
2 Bernoulli- und Binomialverteilung 2
3 Poisson-Verteilung 4
4 Hypergeometrische Verteilung 6
5 Diskrete Gleichverteilung 7
6 Kontinuierliche Gleichverteilung 8
7 Normalverteilung 9
8 Exponentialverteilung 11
9 Chiquadrat-Verteilung 13
10 t-Verteilung 14

I Spezielle Verteilungen

1 Wiederholung

Die Verteilung einer Zufallsvariable X ist im diskreten Fall durch ihre Wahrscheinlichkeitsfunktion $p(x_i)$ und im kontinuierlichen Fall durch ihre Dichtefunktion $f(x)$ gegeben. Es gilt:

$$P(X \in A) = \begin{cases} \sum\limits_{i:x_i \in A} p(x_i) \\ \int\limits_A f(x)\ dx \end{cases}$$

$$E[g(X)] = \begin{cases} \sum\limits_i g(x_i)p(x_i) \\ \int\limits_{\mathbb{R}} g(x)f(x)\ dx \end{cases}$$

2 Bernoulli- und Binomialverteilung

Situationen:

(i) Versuch oder Experiment mit den Ausgängen „Erfolg“ oder „Misserfolg“ (Medikament hilft/hilft nicht; Elfmeter mit/ohne Torerfolg; Münzwurf)

(ii) n unabhängige Versuche, jeder mit Ausgang „Erfolg“ oder „Misserfolg“

Wahrscheinlichkeitsfunktion:

(i) „Bernoulli-Verteilung“ [Jakob Bernoulli 1655 - 1705, schweizer Mathematiker]

$$\begin{aligned} X \sim B_{1,p} \ :\Leftrightarrow\ & P(X=0) = p_{B_{1,p}}(0) = 1-p \\ & P(X=1) = p_{B_{1,p}}(1) = p \end{aligned}$$

wobei $p \in (0,1)$ die Wahrscheinlichkeit für „Erfolg“ ist.

(ii) „Binomialverteilung“

$$X \sim B_{n,p} \ :\Leftrightarrow\ P(X=i) = p_{B_{n,p}}(i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, \ldots, n$$

wobei $\binom{n}{i} = \frac{n!}{i!(n-i)!}$ die Anzahl der verschiedenen Teilmengen mit i Elementen aus einer Menge mit n Elementen ist (vgl. Abschnitt II.4).

Erwartungswert und Varianz:

$$\begin{aligned} X \sim B_{n,p} \ \Rightarrow\ & \mathrm{E}[X] = n \cdot p \\ & \mathrm{Var}[X] = n \cdot p \cdot (1-p) \end{aligned}$$

Beweis: Es gilt $X = \sum\limits_{i=1}^{n} X_i$ mit $X_i \sim B_{1,p}$, wobei

$$X_i = \begin{cases} 1 & \text{wenn der } i\text{-te Versuch ein Erfolg war,} \\ 0 & \text{sonst.} \end{cases}$$

Da

$$\mathrm{E}[X_i] = p \cdot 1 + (1-p) \cdot 0 = p \quad \text{and} \quad \begin{aligned} \mathrm{Var}[X_i] &= \mathrm{E}[X_i^2] - \mathrm{E}[X_i]^2 \\ &= \mathrm{E}[X_i] - p^2 = p\,(1-p) \end{aligned}$$

und da die X_i's unabhängig sind, gilt:

$$\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i] = n \cdot p$$

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = n \cdot p \cdot (1-p)$$

#

Graphen: Siehe R-Datei `Verteilungsplots.R`

Wichtige R-Befehle:

$\underset{\substack{\uparrow \\ \text{Zeichen für} \\ \text{Dichte = Wahrsch.} \\ \text{funktion.}}}{\texttt{d}}\overbrace{\texttt{binom}}^{\text{Abk. der Verteilung}}\texttt{(i,n,p)} = p_{B_{n,p}}(i) = P(X = i)$ mit $X \sim B_{n,p}$

$\underset{\substack{\uparrow \\ \text{Zeichen für} \\ \text{Verteilungsfunktion} \\ \text{(=prob. distr. fct.)}}}{\texttt{p}}\texttt{binom(i,n,p)} = F_{B_{n,p}}(i) = P(X \leq i)$ mit $X \sim B_{n,p}$

$\underset{\substack{\uparrow \\ \text{Quantil}}}{\texttt{q}}\texttt{binom(q,n,p)}$ = kleinstes „i“ mit $\mathrm{P}(X \leq i) \geq q$ mit $X \sim B_{n,p}$

$\underset{\substack{\uparrow \\ \text{Zufallszahl} \\ \text{(=random number)}}}{\texttt{r}}\texttt{binom(m,n,p)}$ erzeugt m $B_{n,p}$-verteilte Zufallszahlen

Beispiel:
Ein Kommunikationssystem besteht aus n Komponenten, von denen jede unabhängig von den anderen mit der Wahrscheinlichkeit p funktioniert. Das Gesamtsystem ist funktionsfähig, wenn mindestens die Hälfte der Einzelkomponenten funktioniert.

(a) Berechnen Sie jeweils die Wahrscheinlichkeit dafür, dass ein System aus 5 Komponenten und dass ein System aus 3 Komponenten funktioniert, wenn $p = 10\%$!

(b) Für welche Werte von p ist die Wahrscheinlichkeit, dass ein System aus 5 Komponenten funktioniert, höher als bei einem System aus 3 Komponenten?

Lösung: Die Zufallsvariable X gebe die Anzahl der funktionierenden Komponenten an.

(a)
$$\begin{aligned} P(\text{„5-Komp.-Sys. funkt.„}) &\underset{\substack{\uparrow \\ X \sim B_{5;0,1}}}{=} P(X \geq 3) = 1 - P(X \leq 2) \\ &= 1 - \texttt{pbinom(2,5,0.1)} = 0,856\% \end{aligned}$$

$$P(\text{„3-Komp.-Sys. funkt.„}) \underset{\substack{\uparrow \\ X \sim B_{3,0.1}}}{=} P(X \geq 2) = 1 - P(X \leq 1) \underset{\substack{\uparrow \\ \texttt{R}}}{=} 2,8\%$$

(b) $$P(\text{„5-Komp.-Sys. funkt.„}) \underset{\substack{\uparrow \\ X \sim B_{5,p}}}{=} P(X \geq 3) =$$
$$= \binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p) + \binom{5}{5} p^5$$

$$P(\text{„3-Komp.-Sys. funkt.„}) = P(X \geq 2) =$$
$$= \binom{3}{2} p^2 (1-p) + \binom{3}{3} p^3$$

Bed.:

$$\binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p) + \binom{5}{5} p^5 > \binom{3}{2} p^2 (1-p) + \binom{3}{3} p^3$$

$$\Leftrightarrow 10p(1-p)^2 + (1-p)(5p^2-3) + p \underbrace{(p^2-1)}_{=(p+1)(p-1)} > 0$$
$$\Leftrightarrow (1-p)(10p - 10p^2 + 5p^2 - 3 - p - p^2) > 0$$
$$\Leftrightarrow (1-p)(-6p^2 + 9p - 3) > 0$$
$$\Leftrightarrow 3(p-1)(2p^2 - 3p + 1) > 0$$
$$\underset{\substack{\uparrow \\ \text{Hilfsrechnung:}}}{\Leftrightarrow} 3(p-1)^2(2p-1) > 0$$

$$\begin{array}{l} (\ 2p^2 - 3p + 1\ ) : (p-1) = 2p - 1 \\ \underline{-\ 2p^2 + 2p} \\ \qquad\quad -\ p + 1 \\ \qquad\quad \underline{\ \ p - 1} \\ \qquad\qquad\quad 0 \end{array}$$

$$\Leftrightarrow p > \frac{1}{2} \text{ und } p \neq 1$$

3 Poisson-Verteilung

Situationen:
Die Poisson-Verteilung stellt eine Näherung für die Binomialverteilung mit großem n und kleinem p dar (häufig kennt man nur $n \cdot p$). Details siehe Ross, S. 137f.

(i) Die Anzahl der Druckfehler auf einer Seite in einem Buch mit:

n	Anzahl der Zeichen auf einer Seite
p	Wsk., dass ein Zeichen falsch gedruckt wurde
$n \cdot p$	Rate der Zeichen, die auf einer Seite falsch gedruckt wurden (zum Beispiel: 2 pro Seite)

(ii) Anzahl der Kunden, die in einer bestimmten Zeiteinheit (Tag, Stunde, ...) eine Postfiliale betreten mit:

n	Anzahl der Menschen im Einzugsbereich der Postfiliale
p	Wsk., dass ein Kunde die Postfiliale betritt
$n \cdot p$	Rate der Menschen, die die Postfiliale üblicherweise pro Tag betreten

(iii) Die Anzahl der Transistoren, die am ersten Tag der Benutzung ausfallen.

Wahrscheinlichkeitsfunktion

„Poisson-Verteilung“ [Siméon Denis Poisson 1781 - 1840, franz. Mathematiker]

$$X \sim P_\lambda \;:\Leftrightarrow\; \mathrm{P}(X = i) = p_{P_\lambda}(i) = e^{-\lambda}\frac{\lambda^i}{i!} \quad (\lambda > 0; i \in \mathbb{N}_0)$$

$$\sum_{i=0}^{\infty} p_{P_\lambda}(i) = 1 \qquad \text{(Übung 12.1)}$$

Erwartungswert und Varianz:

$$X \sim P_\lambda \;\Rightarrow\; \begin{aligned} \mathrm{E}[X] &= \lambda \\ \mathrm{Var}[X] &= \lambda \end{aligned}$$

Beweis: Übung 12.1.

Graphen: siehe R-Datei.

R-Befehle: `dpois(i,`λ`)` $= p_{P_\lambda}(i)$ usw.

Beispiel:
Angenommen, durchschnittlich drei Kunden rufen pro Tag bei einer Hotline an. Berechnen Sie die Wahrscheinlichkeit, dass heute genau ein Kunde anruft. Wie groß ist die Wahrscheinlichkeit, dass heute mehr als 10 Kunden anrufen?

Lösung:

4 Hypergeometrische Verteilung

Situationen:

(i) Ziehen ohne Zurücklegen aus einer endlichen Menge, deren Elemente in zwei Kategorien eingeteilt werden können, wie z.B. akzeptabel/Fehler, männlich/weiblich oder erwerbstätig/-arbeitslos.

(ii) Ein konkretes Beispiel: Ein Behälter enthält $N + M$ Batterien, von denen N akzeptabel und die anderen M defekt sind.
Eine Stichprobe mit n Batterien wird zufällig gezogen (ohne Zurücklegen). [Welche Verteilung passt zu der gleichen Situation mit Zurücklegen?]

Wahrscheinlichkeitsfunktion:

$$X \sim H_{M,N,n} \quad :\Leftrightarrow \quad P(X = i) = \frac{\binom{M}{i}\binom{N}{n-i}}{\binom{M+N}{n}} \quad i = 0, \ldots, min(M, n)$$

M: Objekte vom Typ 1
N: Objekte vom Typ 2
n: Größe der Stichprobe
X: Anzahl der Elemente vom Typ 1 in der Stichprobe

Erwartungswert und Varianz:

$$X \sim H_{M,N,n} \quad \Rightarrow \quad \mathrm{E}[X] = \frac{M}{M+N} \cdot n$$
$$\mathrm{Var}[X] = n \cdot \frac{M}{M+N} \cdot \left(1 - \frac{M}{M+N}\right) \cdot \left(1 - \frac{n-1}{M+N-1}\right)$$

Beweis: Ross, S. 144 f.

Vergleich mit dem Erwartungswert und der Varianz der Binomialverteilung mit $p = \frac{M}{M+N}$:

$$X \sim B_{n,M/(M+N)} \quad \Rightarrow \quad \mathrm{E}[X] = \frac{M}{M+N} \cdot n$$
$$\mathrm{Var}[X] = n \cdot \frac{M}{M+N} \cdot \left(1 - \frac{M}{M+N}\right)$$

Bemerkung:
Für $n > 1$ ist der Faktor $\left(1 - \frac{n-1}{M+N-1}\right)$ kleiner als 1 und konvergiert gegen 1, wenn $M + N$ für festes n gegen ∞ geht.

Graphen: Siehe R-Datei.

R-Befehle: `dhyper(i,M,N,n)` $= p_{H_{M,N,n}}(i)$ usw..

Beispiel:
Aus einem Behälter mit 20 gebrauchten Bauteilen werden zufällig sechs Bauteile für ein System ausgewählt. Das aus ihnen zusammengesetzte Sechs-Komponenten-System ist funktionsfähig, wenn mindestens vier der sechs Bauteile fehlerfrei sind. Wie groß ist die Wahrscheinlichkeit, dass das System funktionsfähig ist, wenn 15 der 20 Bauteile fehlerfrei sind?

Die Zufallsvariable X bezeichne die *Anzahl der fehlerfreien Bauteile*. Dann gilt:

$X \sim$

5 Diskrete Gleichverteilung

Situationen:

(i) Werfen eines fairen Würfels.

(ii) Ziehen eines Loses aus einer Urne mit n Losen.

Wahrscheinlichkeitsfunktion:

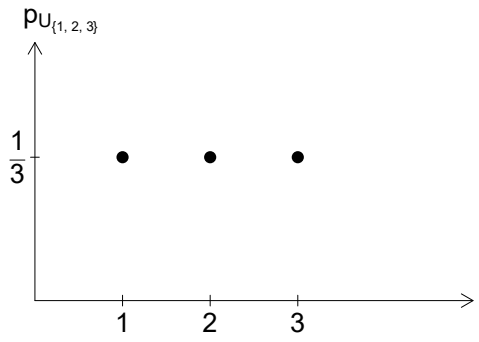
$$X \sim \underset{\substack{\uparrow \\ \text{[Uniform]}}}{U_{\{x_1,\ldots,x_n\}}} \quad :\Leftrightarrow \quad P(X = x_i) = p_{U_{\{x_1,\ldots,x_n\}}}(x_i) = \frac{1}{n} \quad \text{für alle } i = 1, \ldots, n$$

Erwartungswert und Varianz:

$$\mathrm{E}[X] = \sum_{i=1}^{n} x_i \cdot \frac{1}{n} = \bar{x} \qquad \text{„Durchschnitt der Werte } x_1, \ldots, x_n\text{„}$$

$$\mathrm{Var}[X] = \sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \frac{1}{n}$$

Graphen:



R-Befehle:

- keine d,p,q,r-Familie, weil die Wahrscheinlichkeiten leicht selbst berechnet werden können

- Generierung von n auf $\{a, \ldots, b\}$ gleichverteilten Zufallszahlen ($a, b \in \mathbb{Z}$):
  `sample(c(a:b),n,replace=TRUE)`

Beispiel: Hunderttausend mal Würfeln und zählen, wie viele Sechsen man gewürfelt hat.

```
sum( sample(c(1:6),100000,replace=TRUE) == 6 )
```

JITT kont. Vtlg.n

6 Kontinuierliche Gleichverteilung

Situationen:

(i) Ankunftszeit eines Gastes, der sagt, dass er zwischen 20:00 und 21:00 Uhr kommt.

(ii) Zufallszahl im Intervall $[0, 1]$.

Wahrscheinlichkeitsdichte:

$$X \sim U_{[a,b]} = U_{(a,b)} \quad :\Leftrightarrow f_{U_{[a,b]}}(x) = \frac{1}{b-a} \cdot \mathbb{1}_{[a,b]}(x)$$

$$:= \frac{1}{b-a} \cdot \begin{cases} 1 & \text{falls } x \in [a,b] \\ 0 & \text{sonst} \end{cases}$$

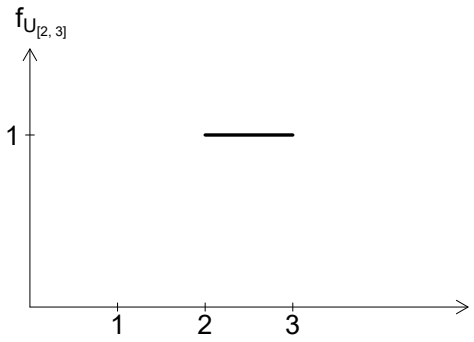
$$\Leftrightarrow \mathrm{P}(X \in A) = \int_A \frac{1}{b-a} \cdot \mathbb{1}_{[a,b]}(x)\ dx$$

Erwartungswert und Varianz:

$$\begin{aligned} \mathrm{E}[X] &= \int_{\mathbb{R}} x \cdot \frac{1}{b-a}\ \mathbb{1}_{[a,b]}(x)\ dx = \\ &= \int_a^b \frac{1}{b-a} \cdot x\ dx = \\ &= \frac{1}{b-a} \left[\frac{1}{2}x^2\right]_a^b = \\ &= \frac{1}{2(b-a)}(b^2 - a^2) = \frac{b+a}{2} \end{aligned}$$

$$\mathrm{Var}[X] = \frac{(b-a)^2}{12} \qquad \text{(Übung 12.4)}$$

Graphen:



R-Befehle: `dunif(x,a,b)` $= f_{U_{[a,b]}}(x)$ usw.

Beispiel:
An einer bestimmten Haltestelle fahren ab 7:00 Uhr jede Viertelstunde Busse ab, also um 7:00, 7:15, 7:30, 7:45, usw.. Ein Fahrgast kommt zu einem zufälligen Zeitpunkt zwischen 7:00 und 7:30 an. Berechnen Sie die Wahrscheinlichkeit, dass er

(a) weniger als 5 Minuten oder,

(b) mindestens 12 Minuten auf den Bus warten muss.

Lösung:
$X :=$ *Minuten, die der Fahrgast nach 7:00 Uhr ankommt*
$\Rightarrow \; X \sim U_{[0,30]}$

(a)
$$\begin{aligned} P(\textit{Wartezeit} < 5\ \textit{min.}) &= P(X \in (10,15] \cup X \in (25,30]) = \\ &= \int_{10}^{15} \frac{1}{30}\, dx + \int_{25}^{30} \frac{1}{30}\, dx = \\ &= \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \end{aligned}$$

(b)
$$\begin{aligned} P(Wartezeit \geq 12) &= P(X \in (0,3] \cup X \in (15,18]) = \\ &= \int_{0}^{3} \frac{1}{30}\, dx + \int_{15}^{18} \frac{1}{30}\, dx = \frac{1}{5} \end{aligned}$$

7 Normalverteilung

Situationen:

- Zufällige Größen, die sich als Summe oder arithmetisches Mittel von vielen unabhängigen zufälligen Einflussgrößen beschreiben lassen
- IQ
- Messfehler

Warnung: Nicht jede zufällige Größe, die sich als Summe oder arithmetisches Mittel von vielen zufälligen Einflussgrößen beschreiben lässt, ist normalverteilt:

- Einkommen in Deutschland
- die Verkaufszahlen von Buchtiteln
- die Versicherungsschäden nach einer Katastrophe
- Körpergröße

sind nicht normalverteilt, da die Einflussgrößen nicht unabhängig sind.

Wahrscheinlichkeitsdichte:

$$X \sim N_{\mu,\sigma^2} \quad :\Leftrightarrow \quad f_{N_{\mu,\sigma^2}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{„Normalverteilung“}$$

Erwartungswert und Varianz:

$$\begin{aligned} \mathrm{E}[X] &= \mu \\ \mathrm{Var}[X] &= \sigma^2 \end{aligned}$$

Beweis:

Betrachte

$$\begin{aligned} \mathrm{E}[X-\mu] &= \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \\ &\underset{\substack{\uparrow \\ u := (x-\mu)/\sigma \\ du = 1/\sigma \; dx}}{=} \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \cdot e^{-\frac{u^2}{2}} \, du = \qquad dx = \sigma \; du \\ &= \frac{\sigma}{\sqrt{2\pi}} \left[e^{-\frac{u^2}{2}} \right]_{-\infty}^{\infty} = 0 \end{aligned}$$

$\mathrm{Var}[X]$ siehe Ross S. 154f.

R-Befehle: `dnorm(x,`μ`,`$\underset{\substack{\uparrow \\ [\texttt{nicht } \sigma^2!]}}{\sigma}$`)` $= f_{N_{\mu,\sigma^2}}(x)$ usw.

Wichtige Eigenschaften: (für die Beweise siehe Ross S. 155 ff)

Beginn JITT-Lücke

1. $$\begin{aligned} X \sim N_{\mu,\sigma^2} \quad &\Rightarrow \quad aX + b \sim N_{a\mu+b, a^2\sigma^2} \\ &\Rightarrow \quad \frac{X-\mu}{\sigma} \sim N_{0,1} \qquad \text{„Standardnormalverteilung“} \end{aligned}$$

   Die Verteilungsfunktion von $N_{0,1}$ wird mit $\Phi(x)$ bezeichnet.

2. $X_1 \sim N_{\mu_1,\sigma_1^2}$, $X_2 \sim N_{\mu_2,\sigma_2^2}$ st. unabh. $\Rightarrow$ $X_1+X_2 \sim N_{\mu_1+\mu_2,\sigma_1^2+\sigma_2^2}$

Graph: siehe Normal-density.png [Prägen Sie sich die σ-Bereiche ein!]

Beispiel:
Laut Klimadatenbank kann die Niederschlagsmenge N_j, die im Jahr j in Augsburg fallen wird, durch eine normalverteilte Zufallsvariable mit Erwartungswert 64 $[cm]$ und Varianz 36 $[cm^2]$ modelliert werden.

(a) Berechnen Sie die Wahrscheinlichkeit, dass es nächstes Jahr mehr als 70 cm regnet.

(b) Berechnen Sie die Wahrscheinlichkeit, dass die Gesamtniederschlagsmenge in den nächsten beiden Jahren mehr als 140 cm beträgt. Nehmen Sie für diese und die nächste Teilaufgabe an, dass die Niederschlagsmengen für die nächsten beiden Jahre unabhängig sind.

(c) Berechnen Sie die Wahrscheinlichkeit, dass die Niederschlagsmenge im nächsten Jahr mehr als 16 cm über der des übernächsten Jahres liegt.

Lösung: [selber]

Ende JITT-Lücke

8 Exponentialverteilung

Situationen:

- Wartezeit auf ein bestimmtes Ereignis (Kunde kommt an Warteschlange an, Erdbeben, neuer Krieg, ...)
  (vgl. Ross, Abschnitt 5.6, Seite 160ff)

Wahrscheinlichkeitsdichte:

$$X \sim Exp_\lambda :\Leftrightarrow$$
$$f_{Exp_\lambda}(x) = \lambda \cdot e^{-\lambda x} \cdot \mathbb{1}_{[0,\infty)}(x)$$

Erwartungswert und Varianz:

$$\begin{aligned}
\mathrm{E}[X] &= \int_0^\infty \underbrace{x}_{=G} \cdot \lambda \cdot \underbrace{e^{-\lambda x}}_{=f} \, dx = \\
&= \lambda \left[x \cdot \left(-\frac{1}{\lambda} e^{-\lambda x} \right) \right]_0^\infty - \lambda \int_0^\infty \underbrace{1}_{=G'} \cdot \left(-\frac{1}{\lambda} e^{-\lambda x} \right) \, dx \\
&= 0 + \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty \\
&= \frac{1}{\lambda}
\end{aligned}$$

$$\mathrm{Var}[X] = \frac{1}{\lambda^2} \qquad \text{(analog)}$$

Charakteristische Eigenschaft:

$$X \sim Exp_\lambda \quad \Rightarrow \quad \mathrm{P}(X > s + t | X > t) = \mathrm{P}(X > s)$$

„X ist gedächtnislos“
[Die Restlebensdauer eines t Jahre alten Transistors ist genauso verteilt wie die Lebensdauer eines neuen Transistors.]

Beweis: Übung 12.5

Graphen: siehe R-Datei

R-Befehl: `dexp(x,`λ`)` $= f_{Exp_\lambda}(x)$ usw.

Beispiel:
Die Laufleistung eines Autos (in Kilometern), bevor die Batterie defekt ist, sei exponentialverteilt mit Erwartungswert 10000 km. Eine Person will eine 5000 km lange Reise antreten. Wie groß ist die Wahrscheinlichkeit, dass sie die Reise beenden kann, ohne die Batterie zu ersetzen?

Lösung:

Sei $X :=$ *Restlebensdauer der Batterie* (in Tsd. km)

$$\underset{\textit{ist gedaechtnislos}}{\overset{\textit{Gesamtlebensdauer}}{\Rightarrow}} X \sim Exp_{\frac{1}{10}} \quad \Rightarrow P(X > 5) = 1 - P(X \leq 5) =$$
$$\underset{\substack{\uparrow \\ \texttt{R}}}{=} 1 - \texttt{pexp(5,0.1)} \approx 60,65\%$$

9 Chiquadrat-Verteilung

Definition:
Falls $Z_1, \ldots, Z_n$ unabhängig und standard-normalverteilte Zufallsvariable sind, so gilt:

$$X := Z_1^2 + \cdots + Z_n^2$$

ist „Chiquadrat-verteilt mit n Freiheitsgraden„.

Notation: $X \sim \chi_n^2$

Situationen: Quadrate von normalverteilten Abweichungen.

Eigenschaft:

$$X_1 \sim \chi_{n_1}^2, X_2 \sim \chi_{n_2}^2 \quad \Rightarrow \quad X_1 + X_2 \sim \chi_{n_1+n_2}^2$$

Beweis: Summe von $n_1 + n_2$-std.- normalverteilten Zufallsvariablen.

Wahrscheinlichkeitsdichte siehe Ross, S.171

Erwartungswert und Varianz:

$$\begin{aligned} \mathrm{E}[X] &\underset{\substack{\uparrow \\ \text{[klar nach Definition]}}}{=} n \\ \mathrm{Var}[X] &\underset{\substack{\uparrow \\ \text{siehe Ross, S. 172}}}{=} 2n \end{aligned}$$

Graphen: siehe R-Datei

R-Befehl: `dchisq(x,n)` $= f_{\chi_n^2}(x)$ usw.

Beispiel:
Es soll ein Zielpunkt im dreidimensionalen Raum getroffen werden. Nehmen Sie an, dass in jeder Koordinate ein normalverteilter Messfehler mit Erwartungswert 0 und Standardabweichung 2 auftritt.

Berechnen Sie die Wahrscheinlichkeit, dass der Abstand zwischen Messung und Zielpunkt mehr als 3 Meter beträgt.

Lösung:
Für den Abstand D gilt:

$$D^2 = X_1^2 + X_2^2 + X_3^2,$$

wobei X_i den Fehler in der i-ten Koordinate bezeichne. Damit ergibt sich:

$$\underset{\substack{\uparrow \\ X_i/2 \sim N_{0,1}}}{\Rightarrow} P(D^2 > 9) = P\left(\frac{X_1^2}{4} + \frac{X_2^2}{4} + \frac{X_3^2}{4} > \frac{9}{4}\right) =$$

$$= 1 - F_{\chi_3^2}\left(\frac{9}{4}\right) \underset{\substack{\uparrow \\ \texttt{R}}}{=} 1 - \texttt{pchisq(9/4, 3)} \approx 52.2\%$$

10 t-Verteilung

Motivation:
Die t-Verteilung wird in der Schätz- und Testtheorie benötigt.

Definition:
Seien $Z \sim N_{0,1},\ X \sim \chi^2_n$

Die Verteilung von

$$T_n = \frac{Z}{\sqrt{\frac{X^2}{n}}}$$

wird „t-Verteilung mit n Freiheitsgraden“ genannt.

Eigenschaften:

- $t_n \overset{n\to\infty}{\to} N_{0,1}$ (ohne Beweis)
- t_n ist achsensymmetrisch zur y-Achse

Erwartungswert und Varianz:

$$\begin{aligned} \mathrm{E}[T_n] &= 0 \ \text{ für } n > 1 \ \text{ (für } n = 1 \text{ existiert der Erwartungswert nicht)} \\ \mathrm{Var}[T_n] &= \frac{n}{n-2} \ \text{ für } n > 2 \end{aligned}$$

Graphen: siehe R-Datei

R-Befehl: `dt(x,n)` $= f_{t_n}(x)$ usw.

Beispiel:

Ende JITT kont. Vtlg.n

Vergleiche:

$$P(T_{12} \le 1.4) \underset{\substack{\uparrow \\ \texttt{R}}}{=} 0.9066$$

$$P(Z \le 1.4) \underset{\substack{\uparrow \\ \texttt{R}}}{=} 0.9192 \ \ (Z \sim N_{0,1})$$

[Achtung: Die Lücke in JITT zu den kont. Vtlg.n muss nun nachgeholt werden!]

ABC 28
Ü 12
Moodle-Test