

1 Beschreibende/Deskriptive Statistik

1.1 Begriffe

]

1.1.1 Beschreibende/Deskriptive Statistik

Beobachtete Daten werden durch geeignete statistische Kennzahlen charakterisiert und durch geeignete Grafiken anschaulich gemacht.

1.1.2 Schließende/Induktive Statistik

Aus beobachtete Daten werden Schlüsse gezogen und diese im Rahmen vorgegebener Modelle der Wahrscheinlichkeitstheorie bewertet.

1.1.3 Grundgesamtheit

Ω : Grundgesamtheit ω : Element oder Objekt der Grundgesamtheit diskret (< 30 Ausprägungen), stetig (≥ 30 Ausprägungen), univariat ($p=1$), multivariat ($p>1$)

1.2 Lagemaße

1.2.1 Modalwerte x_{mod}

Am häufigsten auftretende Ausprägungen (insbesondere bei qualitativen Merkmalen)

1.2.2 Mittelwert

R: $mean(x)$
Schwerpunkt der Daten.
Empfindlich gegenüber Ausreißern.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

1.3 Median

R: $median(x)$
Liegt in der Mitte der sortierten Daten x_i .
Unempfindlich gegenüber Ausreißern.

$$x_{0.5} = \begin{cases} x_{\frac{n+1}{2}}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), & \text{falls } n \text{ gerade} \end{cases} \quad (1)$$

1.4 Streuungsmaße

1.4.1 Spannweite

$$\max x_i - \min x_i$$

1.4.2 Stichprobenvarianz s^2

R: $var(x)$

Verschiebungssatz:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} (\sum_{i=1}^n x_i^2 - n \bar{x}^2)$$

$n\bar{x}^2$) Gemittelte Summe der quadratischen Abweichung vom Mittelwert

1.4.3 Stichprobenstandardabweichung

R: $sd(x)$

$s = \sqrt{s^2}$ Streuungsmaß mit gleicher Einheit wie beobachteten Daten x_i . \bar{x} minimiert die "quadratische Verlustfunktion" oder die Varianz gibt das Minimum der Fehlerquadrate an.

1.5 p-Quantile

R: $quantile(x, p)$. Teilt die sortierten Daten x_i ca. im Verhältnis p : $(1-p)$ d.h. $\hat{F}(x_p) \approx p$; 1. Quartil = 0.25-Quantil; Median = 0.5-Quantil; 3. Quartil = 0.75-Quantil;

1.6 Interquartilsabstand I

$I = x_{0.75} - x_{0.25}$. Ist ein weiterer Streuungsparameter.

1.7 Chebyshev

$\frac{N(S_k)}{n} > 1 - \frac{1}{k^2}$, für alle $k \geq 1$ \bar{x} der Durchschnitt, $s > 0$ die Stichproben-Standardabweichung von Beobachtungswerten x_1, \dots, x_n . Sei $S_k = \{i, 1 \leq i \leq n : |x_i - \bar{x}| < k \cdot s\}$; Für eine beliebige Zahl $k \geq 1$ liegen mehr als $100 \cdot (1 - \frac{1}{k^2})$ Prozent der Daten im Intervall von $\bar{x} - ks$ bis $\bar{x} + ks$. **Speziell:** Für $k=2$ liegen mehr als 75% der Daten im 2s-Bereich um \bar{x} . Für $k=3$ liegen mehr als 89% der Daten im 3s-Bereich um \bar{x} . **Komplement Formulierung:** $\bar{S}_k = \{i || x_i - \bar{x}| \geq k \cdot s\}$; $\frac{N(\bar{S}_k)}{n} \leq \frac{1}{k^2}$; Die Ungleichheit liefert nur eine sehr grobe Abschätzung, ist aber unabhängig von der Verteilung der Daten. **Empirische Regeln** 68% der Daten im Bereich um $\bar{x} \pm s$. 95% um $\bar{x} \pm 2s$. 99.7% um $\bar{x} \pm 3s$.

1.8 Korrelation

Grafische Zusammenhang zwischen multivariaten Daten x und y durch ein Streudiagramm. Kennzahlen zur Untersuchung des Zusammenhangs:

1.8.1 Empirische Kovarians

$$R: cov(x, y); s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} (\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y})$$

1.8.2 Empirische Korrelationskoeffizient r

R: $cor(x, x)$; $r = \frac{s_{xy}}{s_x s_y}$; Näherungsweise lin. Zusammenhang zw. x und y , falls $|r| \approx 1$.

1.8.3 Regressionsgerade y

$$y = mx + t \text{ mit } m = r \cdot \frac{s_y}{s_x} \text{ und } t = \bar{y} - m \cdot \bar{x}$$

2 Wahrscheinlichkeitsrechnung

2.1 Begriffe

Ergebnisraum Ω : Menge aller möglichen Ergebnisse eines Experiments

Elementarereignis $\omega \in \Omega$: einzelnes Element von Ω

Ereignis $E \subseteq \Omega$: beliebige Teilmenge des Ergebnisraums Ω heißt sicheres Ereignis, \emptyset heißt unmögliches Ereignis

Vereinigung $E \cup F$: Ereignis E oder Ereignis F treten ein. $\bigcup_{i=1}^n E_i$: mindestens ein Ereignis E_i tritt ein.

Schnitt $E \cap F$: Ereignis E und Ereignis F treten ein.

$\bigcap_{i=1}^n E_i$ alle Ereignisse E_i treten ein. **Generereignis** $\bar{E} = \Omega / E$: Ereignis E tritt nicht ein (Komplement von E)

Disjunkte Ereignisse E und F: $E \cap F = \emptyset$

2.2 De Morgan'schen Regeln

$$\overline{E_1 \cup E_2} = \bar{E}_1 \cap \bar{E}_2$$

$$\overline{E_1 \cap E_2} = \bar{E}_1 \cup \bar{E}_2$$

2.3 Wahrscheinlichkeit

$$0 \leq P(E) \leq 1; P(\Omega) = 1;$$

$$P(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i), \text{ falls } E_i \cap E_j = \emptyset \text{ für } i \neq j$$

2.3.1 Satz 2.1

$$P(\bar{E}) = 1 - P(E)$$

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

(Übungsaufgabe!!! Ergänzen)

2.4 Laplace-Experiment

Zufallsexperimente mit n gleich wahrscheinlichen Elementarereignissen. Dann berechnet sich die Wahrscheinlichkeit $P(E)$ für $E \subseteq \Omega$ aus:

$$P(E) = \frac{\text{Anzahl der für } E \text{ günstigen Ereignisse}}{\text{Anzahl der möglichen Ereignisse}} = \frac{\text{Mächtigkeit von } E}{\text{Mächtigkeit von } \Omega} = \frac{|E|}{|\Omega|} \text{ text}$$

2.5 Kombinatorik

2.5.1 Allgemeines Zählprinzip

Anzahl der Möglichkeiten für ein k -stufes Zufallsexperiment mit n_i Varianten im i -ten Schritt: $n_1 \cdot n_2 \cdot \dots \cdot n_k$

2.5.2 Permutationen

Anzahl einer n -elementigen Menge n -maliges Ziehen ohne Zurücklegen mit Beachtung der Reihenfolge: **n unterscheidbare Elemente:** $n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1$
k Klassen mit je n_i nicht unterscheidbaren Elementen $n = \sum_{i=1}^k n_i$: $\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$

2.5.3 Abzähl k-elementigen Teilmengen einer n-elementigen Menge k-maliges Ziehen aus einer n-elementigen Menge

ohne Zurücklegen = $k \leq n$.

mit Zurücklegen = $k > n$ möglich.

mit Beachtung der Reihenfolge, ohne

Zurücklegen: $\frac{n!}{(n-k)!}$

ohne Beachtung der Reihenfolge, ohne

Zurücklegen: $\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$

mit Beachtung der Reihenfolge, mit Zurücklegen: n^k

ohne Beachtung der Reihenfolge, mit Zurücklegen: $\binom{n+k-1}{k}$

2.6 Bedingte Wahrscheinlichkeit

$$P(E|F) = P_F(E) = \frac{|E \cap F|}{|F|} = \frac{P(E \cap F)}{P(F)}$$

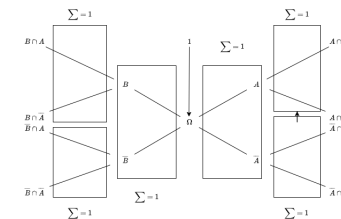
$$P(E \cap F) = P(F) \cdot P(E|F)$$

$$P(E \cap F) = P(F|E) \cdot P(E)$$

2.6.1 Satz 2.2

$$P(E \cap F) = P(E|F) \cdot P(F)$$

$$P(E \cap F) = P(F|E) \cdot P(E)$$



2.6.2 Satz der totalen Wahrscheinlichkeit

Sei $\Omega = \bigcup_{i=1}^n E_i$ mit $E_i \cap E_j = \emptyset$ für $i \neq j$ d.h. die Ereignisse bilde eine disjunkte Zerlegung bzw. eine Partition von Ω . Somit gilt:

$$P(F) = \sum_{i=1}^n P(F \cap E_i) = \sum_{i=1}^n P(F|E_i) \cdot P(E_i)$$

Summe der Äste des Wahrscheinlichkeitsbaums zu allen Schnitten $F \cap E_i$



2.6.3 Vierfeldertafel

$$P(F) = P(F \cap E) + P(F \cap \bar{E})$$

$$P(\bar{F}) = P(\bar{F} \cap E) + P(\bar{F} \cap \bar{E})$$

	E	\bar{E}	
F	$P(F \cap E)$	$P(F \cap \bar{E})$	$P(F)$
\bar{F}	$P(\bar{F} \cap E)$	$P(\bar{F} \cap \bar{E})$	$P(\bar{F})$
	$P(E)$	$P(\bar{E})$	1

Satz 2.2 $P(E \cap F) = P(E) \cdot P(F|E)$
 $P(F|E) = \frac{P(E \cap F)}{P(E)}$ **Tafel** $P(F) = P(F \cap E) + P(F \cap \bar{E})$
 $P(\bar{E}) = P(E) - P(E \cap F)$; $P(\bar{F}|E) = 1 - P(F|E)$

2.6.4 Formel von Bayes

Hilfreich, wenn man man $P(F|E_i)$ kennt, aber nicht $P(E_k|F)$ **Satz 2.4** $P(E_k|F) = \frac{P(F|E_k) \cdot P(E_k)}{\sum_{i=1}^n P(F|E_i) \cdot P(E_i)}$

$$\frac{P(F|E_k) \cdot P(E_k)}{\sum_{i=1}^n P(F|E_i) \cdot P(E_i)}$$

Nur Nenner! $P(F)$ aus dem Satz der totalen Wahrscheinlichkeit.

2.6.5 Stochastische Unabhängigkeit

Uebung Die Ereignisse E und F heißen (stochastisch) unabhängig, wenn die Information über das Eintreten des einen Ereignisses die Wahrscheinlichkeit für das Eintreten des anderen Ereignisses nicht ändert, d.h. falls $P(E|F) = P(E)$ oder $P(E \cap F) = P(E) \cdot P(F)$

$$P(E \cap F) = P(E) \cdot P(F)$$

$$P(E \cap F) = P(E) \cdot P(F)$$

Es gilt Falls die Ereignisse E, F unabhängig sind, dann sind auch:

$$E, \bar{F}$$

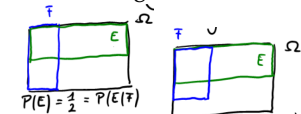
$$\bar{E}, F$$

$$\bar{E}, \bar{F}$$

unabhängig **Bemerkung**

• Stochastische Unabhängigkeit bedeutet nicht notwendigerweise eine kausale Abhängigkeit

• Veranschaulichung mit Venn Diagramm



stoch. unabhängig $P(E) = \frac{1}{2} = P(F)$

$P(E \cap F) = \frac{1}{4} < P(E) \cdot P(F)$

• $A, B \neq \emptyset$ und $A \cap B = \emptyset$

$$P(A \cap B) \stackrel{?}{=} P(A) \cdot P(B)$$

$\emptyset \neq P(A) \cdot P(B)$ da $P(A) > 0$ und $P(B) > 0$

$\Rightarrow A, B$ stochastisch abhängig

3 Zufallsvariable

Abbildung des abstrakten Ergebnisraums Ω auf \mathbb{R} . Eine Abbildung $X: \Omega \rightarrow \mathbb{R}$, $\omega \mapsto X(\omega)$ heißt Zufallsvariable (ZV). $x \in \mathbb{R}$ heißt Realisation der ZV X .

• Diskrete ZV: $X(\Omega) = x_1, \dots, x_2 (n \in \mathbb{N})$; z.B. X = "Augensumme beim Würfeln"

• Stetige ZV: $X(\Omega) \subseteq \mathbb{R}$; "z.B. Körpergröße eines Menschen"

3.1 Verteilungsfunktion-allg.

Die Wahrscheinlichkeit $P(B)$ für ein Ereignis B in \mathbb{R} wird zurückgeführt auf die Wahrscheinlichkeit der entsprechenden Ereignisse in Ω . Für jedes $X \in \mathbb{R}$ ist die

Verteilungsfunktion $F: \mathbb{R} \rightarrow [0,1]$ einer ZV X definiert durch:
 $F(x) = P(X \leq x)$

- $0 \leq F(x) \leq 1$
- $\lim_{x \rightarrow -\infty} F(x) = 0$ $\lim_{x \rightarrow \infty} F(x) = 1$
- monoton wachsend
- $P(X > x) = 1 - F(x)$
- $P(a < X \leq b) = F(b) - F(a)$

3.2 Diskrete ZVs
Für eine diskrete ZV X mit $X(\Omega) = x_1, \dots, x_n$ (n endlich oder abzählbar unendlich) ist die Wahrscheinlichkeitsfunktion definiert durch:

$$p(x) = \begin{cases} P(X = x_i), & \text{falls } x_i \in X(\Omega) \\ 0, & \text{sonst} \end{cases} \quad (2)$$

Es gilt:

- $F(x) = (P(X \leq x) = \sum_{x_i \leq x} p(x_i)$
- $F(x)$ ist eine rechtseitig stetige **Treppenfunktion** mit **Sprüngen** bei der Realisation von x_i .

3.3 Stetige ZVs
Stetige ZV X ist die Wahrscheinlichkeitsdichte $f: \mathbb{R} \rightarrow [0, \infty[$ definiert durch

$$P(a < X < b) = \int_a^b f(x) dx$$

Es gilt:

- $F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$ und $F'(x) = f(x)$
- $F(x)$ ist stetig & $P(a < X \leq b) = P(a \leq X \leq b)$ wegen $P(X = a) = 0$

3.4 Verteilungsfunktion

$\int_{\text{Untergrenze}}^x$ Es wird normal mit - Integriert.

3.5 Zusammenfassung

3.5.1 Diskrete ZV

- Wahrscheinlichkeitsverteilung $p(x) \sum_{i=1}^n p(x_i) = 1$ x_i ist Realisation der ZV.
- Verteilungsfunktion $F(x)$ ist rechtsseitig stetige **Treppenfunktion**. **Sprunghöhen:** $P(X = x_i) = F(x_i) - \lim_{x \rightarrow x_i^-} F(x) \neq 0$
- $P(a < X \leq b) = F(b) - F(a) \neq P(a \leq X \leq b)$

3.5.2 Stetige ZV

- Dichtefunktion $f_X: \mathbb{R} \rightarrow [0, \infty[$ $\int_{-\infty}^{\infty} f(x) dx = 1$
- Verteilungsfunktion $F(x)$ ist stetig mit $F'(x) = f(x); P(X = x_i) = 0$
- $P(a < X \leq b) = F(b) - F(a) = P(a \leq X \leq b) = F(b) - F(a) = P(a < X < b)$

3.6 Erwartungswert

Der Erwartungswert $E[X]$ einer ZV X ist der **Schwerpunkt** ihrer Verteilung oder der durchschnittliche zu erwartende Wert der ZV.

- diskrete ZV: $E[X] = \sum_{i=1}^n x_i \cdot p(x_i)$
- stetige ZV: $E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$

ZV ist konstant. $E[X]$ verhält sich linear. Eigenschaften von $E[X]$:

- $E[b] = b$
- $E[aX + b] = aE[X] + b$
- $E[\underbrace{X_1 + \dots + X_n}_{\text{unabhängig}}] = \sum_{i=1}^n E[X_i]$
- $\sum_{i=1}^n x_i$

3.6.1 Satz 3.1

Sei $Y = g(X)$ eine Funktion der ZV X . Dann gilt:

- für diskrete ZV: $E[g(X)] = \sum_{i=1}^n g(x_i) \cdot p(x_i)$
- für stetige ZV: $E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx$. Das vertauschen von E und g nur bei **linearen** Funktionen möglich. $\Rightarrow g(E[X])$

3.7 Varianz

Die Varianz einer ZV X mit μ ist ein quadratisches Streuungsmaß. $\sigma^2 = Var[X] = E[(X - \mu)^2]$ falls X stetig $\int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)$

$g(X)$

Die Standardabweichung $\sigma = \sqrt{Var[X]}$ hat im Gegensatz zur Varianz die gleiche Dimension von der ZV X .

- $Var[b] = 0$
- $Var[aX + b] = a^2 Var[X]$

3.7.1 Satz 3.2

$Var[X] = E[X^2] - (E[X])^2$ Beim Minuend wird beim Erwartungswert nur das einfach stehende x quadriert **nicht** $f(x)$!!!

3.8 Kovarianz, Standardisierung

Sei X eine ZV mit μ und σ . Dann ist $Z = \frac{X - \mu}{\sigma} = \frac{x}{\sigma} - \frac{\mu(\text{konstant})}{\sigma}$

3.9 Kovarianz
Eigenschaften:

- $Cov[X, Y] = Cov[Y, X]$
- $Cov[X, X] = Var[X]$
- $Cov[aX, Y] = aCov[X, Y]$

Die Kovarianz zweier ZV (X, Y) ist definiert durch $Cov[X, Y] = E[(X - E[X])(Y - E[Y])]$ Die Kovarianz beschreibt die Abhängigkeit zweier ZV X und Y . Je stärker diese Korrelieren, desto (betragsmäßig) größer ist die Kovarianz. Falls X, Y stochastisch unabhängig $\Rightarrow Cov[X, Y] = 0$

3.10 Satz 3.3
 $Cov[X, Y] = E[XY] - E[X] \cdot E[Y]$

3.10.1 Varianz einer Summe von ZV

- $Var[X_1 + \dots + X_n] = \sum_{i=1}^n Var[X_i] + 2Cov[X_1, X_2]$
- Falls X_i, X_j paarweise unabhängig !!!: $Var[X_1 + \dots + X_n] = \sum_{i=1}^n Var[X_i]$

3.11 Overview μ, σ

3.11.1 $E[X]$

$E[aX + b] = AE[X] + b; EX_1 + \dots + E_n = \sum_{i=1}^n E[X_i]$
Falls X_1, X_2 unabhängig:
 $E[X_i] = \mu \Rightarrow E[\bar{X}] = E[\frac{1}{n}(X_1 + \dots + X_n)] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \cdot n \cdot \mu = \mu$

3.11.2 Varianz

$Var[aX + b] = a^2 Var[X]$
Falls X_i, X_j paarweise unabhängig:
 $Var[X_1 + \dots + X_n] = \sum_{i=1}^n Var[X_i]$
 $Var[X_i] = \sigma^2 \Rightarrow Var[\bar{X}] = Var[\frac{1}{n}(X_1 + \dots + X_n)] = \frac{1}{n^2} \sum_{i=1}^n Var[X_i] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$

3.12 Quantile

Sei X eine ZV mit Verteilungsfunktion $F(x)$ und $0 < p < 1$. Dann ist das p -Quantil definiert als der Wert $x_p \in \mathbb{R}$ für den gilt:
 $F(x_p) \geq p$. p -Quantil einer stetigen ZV mit **streng monoton wachsenden** $F(x): x_p = F^{-1}(p)$ d. h. umkehrbar.

4 Spezielle Verteilung
4.1 Diskrete Verteilung

4.1.1 Bernoulli-Verteilung

Indikatorvariable mit den Werten 1 bei Erfolg und 0 bei Misserfolg; **Wahrscheinlichkeit:** $P(X = 1) = p, P(X = 0) = 1 - p$; **Verteilung:** $X \sim B_{1,p}$ p ist Erfolgswahrscheinlichkeit; $E[X] = p = \sum x_i \cdot p(x_i) = 1 \cdot p(1); Var[X] = p(1-p) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)$;

4.1.2 Binomialverteilung

Anzahl der Erfolge beim n -maligen Ziehen mit Zurücklegen; **Wahrscheinlichkeit** $P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, k \in \{0, 1, \dots, n\}$; **Verteilung** $X \sim B_{n,p}$; $E[X] = np$; $Var[X] = np(1 - p)$; **R:** $dbinom(k, n, p) = P(X = k) \triangleq$ Wahrscheinlichkeits-/Dichtefunktion; $pbinom(k, n, p) = F(k) \triangleq$ Verteilungsfunktion; $qbinom(q, n, p) \triangleq q$ -Quantil; $rbinom(k, n, p) \triangleq$ binomialverteilte Zufallszahlen;

4.1.3 Hypergeometrische Verteilung

Anzahl der Erfolge beim **n -maligen Ziehen ohne Zurücklegen** aus einer Menge mit M Elementen, die Erfolg bedeuten, und N Elementen, die Misserfolg bedeuten. **Gesamtumfang** $= M + N$; **Wahrscheinlichkeit** $P(X = k) = \frac{\binom{M}{k} \cdot \binom{N}{n-k}}{\binom{M+N}{n}}, k \in \{0, 1, \dots, \min\{n, M\}\}$; **Verteilung** $X \sim H_{M,N,n}$; $E[X] = n \frac{M}{M+N}$; $\frac{M}{M+N} \triangleq$ Trefferwahrscheinlichkeit; $Var[X] = n \frac{M}{M+N} (1 - \frac{M}{M+N}) \frac{M+N-n}{M+N-1}$; $\rightarrow 1$ falls n klein im Verhältnis zu $M+N$; **R:** $dhyper(k, M, N, n) = P(X = k)$; $phyper(k, M, N, n) = F(k)$;

4.1.4 Poisson-Verteilung

Verteilung der seltenen Ereignisse Häufigkeit punktförmiger Ereignisse in einem Kontinuum. Die durchschnittlich zu erwartende Anzahl der Erfolge λ pro Maßeinheit (i. a. Zeiteinheit) sei bekannt. $k \in \mathbb{N}_0 \rightarrow$ **diskret Wahrscheinlichkeit** $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ mit $\sum_{k=0}^{\infty} P(X = k) = 1$, da $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$; **Verteilung** $X \sim P_{\lambda}$; $E[X] = \lambda$, da $\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} \lambda \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda$;

$Pois(k, \lambda) = \lambda R: dpois(k, \lambda) = P(X = k)$; $ppois(k, \lambda) = F(k)$;

4.1.5 Gleichverteilung

Alle Werte $\{x_1, \dots, x_n\}$ einer ZV X sind gleich wahrscheinlich; **Wahrscheinlichkeit** $P(X = x_k) = \frac{1}{n}$; **Verteilung** $X \sim U_{[x_1, \dots, x_n]}$; $E[X] = \frac{1}{n} \sum_{k=1}^n x_k = \bar{x}$; $Var[X] = \frac{1}{n} \sum_{k=1}^n x_k^2 - \bar{x}^2$; **R:** $sample(1, N, n) \triangleq n$ Zufallszahlen zwischen 1 und N

4.2 Gleichverteilung

4.2.1 Stetige Gleichverteilung

Zufallszahlen aus einem Intervall $[a, b]$; **Dichte:** $f(x) = \frac{1}{b-a}$ für $x \in [a, b]$; **Verteilung:** $X \sim U_{[a,b]}$; $E[X] = \frac{a+b}{2}$; $Var[X] = \frac{(b-a)^2}{12}$ **R:** $dunif(x, a, b) = f(x)$; $runif(n) \triangleq n$ Zufallszahlen zwischen 0 und 1; $runif(n, a, b) \triangleq n$ Zufallszahlen zwischen a und b ;

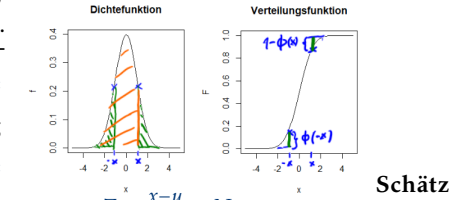
4.2.2 Normalverteilung

Beschreibt viele reale Situationen, ist insbesondere Grenzverteilung unabhängiger Summen; **Dichte:** $f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$; **Verteilung:** $X \sim N_{\mu, \sigma^2}$; $E[X] = \mu$; $Var[X] = \sigma^2$; **R:** $dnorm(x, \mu, \sigma) = f(x)$; $pnorm(x, \mu, \sigma) = F(x)$; $qnorm(q, \mu, \sigma) : q$ -Quantil; **Maximalstelle** von $f(x)$ bei $x = \mu$; **Wendestelle** von $f(x)$ bei $x = \mu \pm \sigma$; $E[aX + b] = aE[X] + b$; $Var[aX + b] = a^2 Var[X]$; $X \sim N_{\mu, \sigma^2} \Rightarrow aX + b \sim N_{a\mu + b, a^2 \sigma^2}$ und $\frac{X - \mu}{\sigma} \sim N_{0,1}$; $X_1 \sim N_{\mu_1, \sigma_1^2}$ und $X_2 \sim N_{\mu_2, \sigma_2^2} \Rightarrow X_1 + X_2 \sim N_{\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2}$; X_1, X_2 stochastisch unabhängig

4.2.3 Standardnormalverteilung

Dichte: $\varphi(x) = \frac{1}{\sqrt{2}} e^{-\frac{1}{2}x^2}$; **Verteilung**

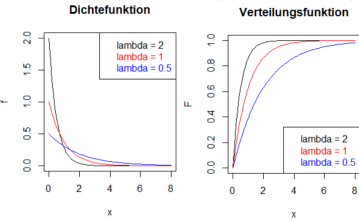
$\Phi(x) = \int_{-\infty}^x \varphi(t) dt$; **Quantile:** $\Phi(-x) = 1 - \Phi(x) \Rightarrow -x_p = x_{1-p}$ z.B. $-x_{0.25} = x_{0.75}$;



Schätz-
werte: $Z = \frac{x - \mu}{\sigma} \sim N_{0,1}$

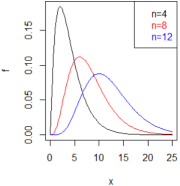
4.2.4 Exponentialverteilung

Modellierung von Lebensdauern, Wartezeiten Sei $Y_t \sim P_{\lambda t}$ im Intervall $[0, t]$ von t Zeiteinheiten, dann beschreibt die Exponentialverteilung die Wartezeit X bis zum Eintreten eines Ereignisses; **Dichte- und Verteilungsfunktion:** $f(x) = \lambda e^{-\lambda x} (x \geq 0)$ und $F(x) = 1 - e^{-\lambda x}$; **Verteilung:** $X \sim Exp_{\lambda}$; $E[X] = \frac{1}{\lambda} \Rightarrow$ Berechnung mit partieller Integration; $Var[X] = \frac{1}{\lambda^2}$; **R:** $dexp(x, \lambda) = f(x)$; $pexp(x, \lambda) = F(x)$; **Eigenschaft:** Eine exponentialverteilte ZV X ist gedächtnislos, d.h. $P(X > s + t) | X > t = P(X > s)$;



4.2.5 Chiquadrat-Verteilung

Z_1, \dots, Z_n seien unabhängige, standard-normalverteilte ZV $\Rightarrow X = Z_1^2 + \dots + Z_n^2$ hat Chiquadratverteilung mit n Freiheitsgraden; **Anwendungsmodell:** Summen unabhängiger, standardnormalverteilter ZV; **Verteilung:** $X \sim \chi_n^2$; $E[X] = n$; $Var[X] = 2n$; **R:** $dchisq(x, n) = f(x)$; $pchisq(x, n) = F(x)$; **Eigenschaft:** $X_1 \sim \chi_{n_1}^2$ und $X_2 \sim \chi_{n_2}^2 \Rightarrow X_1 + X_2 \sim \chi_{n_1+n_2}^2$



4.2.6 t-Verteilung

$Z \sim N_{0,1}$ und $X \sim \chi_n^2 \Rightarrow Y = \frac{Z}{\sqrt{\frac{X}{n}}}$ ist t-verteilt mit n Freiheitsgraden; **Anwendungsmodell:** Schätz- und Testverfahren bei unbekannter Varianz; **Verteilung:** $Y \sim t_n$; $E[Y] = 0$ für $n > 1$; $Var[Y] = \frac{n}{n-2}$ für $n > 2$; **R:** $dt(y, n)$; $pt(y, n)$; **Eigenschaft:** **ten:** Für $n \rightarrow \infty : t_n \rightarrow N_{0,1}$; Achsensymmetrie der Dichtefunktion $\Rightarrow -y_p = x_{1-p}$

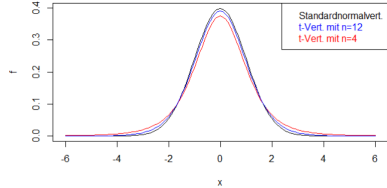


Abbildung Dichtefunktion