Wahrscheinlichkeit und Statistik Zusammenfassung

Fabian Hahn, Dino Wernli

29. September 2008

Wahrscheinlichkeitstheorie

Grundlegende Kombinatorik

Die folgende Tabelle bezieht sich auf das Urnenmodell:

ziehe n aus k	geordnet	ungeordnet
mit Zurücklegen	n^k	$\binom{n+k-1}{k}$
ohne Zurücklegen	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Ereignisraum Ω

Menge aller möglichen **Elementarereignisse** ω des betrachteten Zufallsexperiments.

Beispiel: Würfel: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Ereignis A

Teilmenge des *Ereignisraums* Ω ($A \subseteq \Omega$), also eine Kollektion aus *Elementarereignissen*. Kann also auch mehrere "Ergebnisse" des Zufallsexperiments umfassen.

Beispiel: Gerader Würfelwurf: $A = \{2, 4, 6\} \subseteq \Omega$

Beobachtbare Ereignisse $\mathcal{F}$

Ereignisse, die man wirklich beobachten, also messen kann. Es gilt $\mathcal{F} \subseteq 2^\Omega$. Normalerweise, also wenn Ω endlich oder überabzählbar ist, wählt man $\mathcal{F} = 2^\Omega$, also alle möglichen Ereignisse des *Ereignisraums*.

Mengenoperationen auf Ereignissen

Mengenoperationen auf Ereignissen funktionieren genau so, wie man es sich auch intuitiv vorstellen würde:

- Schnittmenge $A \cap B$: *Ereignis*, dass A und B gleichzeitig eintreten.
- Vereinigung $A \cup B$: *Ereignis*, dass A oder B (oder auch beide) eintreten.
- Komplement $\overline{A}$: *Ereignis*, dass A nicht eintritt.

Wahrscheinlichkeitmass P

Funktion, welche einem *Ereignis* einen Wert zwischen 0 und 1 zuordnet, also eine Abbildung $P : \mathcal{F} \to [0, 1]$.

Rechenregeln

- $P[\Omega] = 1, \quad P[\emptyset] = P[\overline{\Omega}] = 0$
- $P[A \cup B] = P[A] + P[B]$ falls $A \cap B = \emptyset$ (A und B heissen dann **disjunkt**), ansonsten verwende Inklusion-Exklusions-Prinzip:
  $P[A \cup B] = P[A] + P[B] - P[A \cap B]$
- Falls alle A_i *paarweise disjunkt* sind, gilt:

$$P\left[\bigcup_i A_i\right] = \sum_i P[A_i] \tag{1}$$

- $P[\overline{A}] = 1 - P[A]$
- $A \subseteq B \to P[A] \leq P[B]$

Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$

Ein *Wahrscheinlichkeitsraum* charakterisiert ein Zufallsexperiment und umfasst einen *Grundraum* Ω, die *beobachtbaren Ereignisse* $\mathcal{F}$ und eine *Wahrscheinlichkeitsabbildung* P.

Laplace-Raum

Ein *Laplace-Raum* ist ein endlicher *Wahrscheinlichkeitsraum*, in welchem alle *Elementarereignisse* gleich wahrscheinlich sind. Das *Wahrscheinlichkeitsmass* P in diesem Raum heisst **diskrete Gleichverteilung**. Für alle *Ereignisse* $A \subseteq \Omega$ gilt:

$$P[A] = \frac{|A|}{|\Omega|} \tag{2}$$

Bedingte Wahrscheinlichkeit

Die *bedingte Wahrscheinlichkeit* von B gegeben A entspricht der Wahrscheinlichkeit, dass B eintritt, wenn man schon weiss, dass A eingetreten ist. Es gilt:

$$P[B|A] = \frac{P[A \cap B]}{P[A]} \tag{3}$$

$$P[A \cap B] = P[A] \cdot P[B|A] = P[B] \cdot P[A|B] \tag{4}$$

Die **Pfadregel**, nach der Wahrscheinlichkeiten in einem **Wahrscheinlichkeitsbaum** multipliziert werden, um die Wahrscheinlichkeit eines "Blattes" zu erhalten, entspricht einer Verkettung *bedingter Wahrscheinlichkeiten.*

Satz der totalen Wahrscheinlichkeit

Sei A_i eine *disjunkte* Zerlegung von Ω, also $\bigcup_{i=1}^n A_i = \Omega$ *disjunkt*. Für beliebige *Ereignisse* B gilt dann:

$$P[B] = \sum_{i=1}^n P[A_i \cap B] = \sum_{i=1}^n P[A_i] \cdot P[B|A_i] \tag{5}$$

Der Satz bedeutet gerade, dass man die *Wahrscheinlichkeit* eines Ereignisses "unten" in einem *Wahrscheinlichkeitsbaum* bekommen kann, wenn man die einzelnen Pfade addiert, die zu ihm führen.

Satz von Bayes

Mit dem *Satz von Bayes* lassen sich *bedingte Wahrscheinlichkeiten* "umkehren". Sei A_i eine *disjunkte* Zerlegung von Ω, also $\bigcup_{i=1}^{n} A_i = \Omega$ *disjunkt* und B ein weiteres Ereignis. Dann gilt für jedes k:

$$P[A_k|B] = \frac{P[A_k \cap B]}{P[B]} = \frac{P[A_k] \cdot P[B|A_k]}{\sum_{i=1}^{n} P[A_i] \cdot P[B|A_i]} \tag{6}$$

Unabhängigkeit

Zwei *Ereignisse* A und B heissen *unabhängig*, falls gilt:

$$P[A \cap B] = P[A] \cdot P[B] \tag{7}$$

Vergleich mit (4) liefert: $(P[A] \neq 0,\ P[B] \neq 0)$

$$A, B \text{ unabhängig} \Leftrightarrow P[B|A] = P[B] \Leftrightarrow P[A|B] = P[A] \tag{8}$$

Anschaulich bedeutet *Unabhängigkeit*, dass das Eintreten des einen Ereignisses keinen Einfluss auf die Wahrscheinlichkeit des anderen hat. Deshalb ist auch die *Kovarianz* $\text{Cov}(X,Y)$ zweier Zufallsvariablen X und Y gleich Null, falls sie *unabhängig* sind (**Achtung:** Umkehrung gilt nicht!).

Die Ereignisse A_i mit $i \in \{1, 2, \ldots, n\}$ heissen *unabhängig*, wenn für jede Teilfamilie T von A_i gilt:

$$1 \leq j_1 < \ldots < j_k \leq n, \quad T = A_{j_1} \cap \ldots \cap A_{j_k} \tag{9}$$

$$P[T] = \prod_{m=1}^{k} P[A_{j_m}]$$

Achtung: Paarweise Unabhängigkeit bedeutet nicht, dass alle Elemente unabhängig sind. Gegenbeispiel Würfelwurf: A = erste Zahl gerade, B = zweite Zahl gerade, C = beide gerade oder beide ungerade.

Zufallsvariablen

Zufallsvariable X

Eine *Zufallsvariable* ist eine Funktion $X : \Omega \to \mathbb{R}$, welche jedem *Elementarereignis* eine Zahl zuordnet, welche repräsentiert, wofür wir uns bei unserem Zufallsexperiment überhaupt interessieren.

Beispiele: Augenzahl bei einem Würfelwurf, Anzahl Köpfe bei k-maligem Münzwurf, Gewicht einer aus einem Sack gezogenen Kartoffel.

Eine *Zufallsvariable* mit abzählbarem Wertebereich heisst **diskret**, ansonsten ist sie (bei uns meist) **stetig**.

Indikatorfunktion $\mathbf{1}$

Die *Indikatorfunktion* ist eine spezielle *Zufallsvariable*, welche angibt, ob ein Ereignis A eingetreten ist.

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases} \tag{10}$$

Gewichtsfunktion p

Die *Gewichtsfunktion* $p_X(t) : \mathcal{W}(X) \to [0,1]$ ordnet jedem Wert t, den eine *Zufallsvariable* X annehmen kann, seine *Wahrscheinlichkeit* zu. Viel intuitiver aber auch erlaubt ist die Schreibweise $P[X = t]$, wobei $t \in \mathcal{W}(X)$.
Anschaulich gibt $P[X = t]$ auf die Frage Auskunft: "Mit welcher *Wahrscheinlichkeit* nimmt X den Wert t an?"

Achtung: Die *Gewichtsfunktion* ist nur für *diskrete Zufallsvariablen* sinnvoll (denn bei einer *stetigen* Verteilung hat ein Einzelereignis immer die *Wahrscheinlichkeit* 0)!

Dichtefunktion

Die *Dichtefunktion* f_X ist das Analagon zur *Gewichtsfunktion* für *stetige Zufallsvariablen*. Sie gibt die Änderung ihrer *Verteilungsfunktionen* an und charakterisiert sie so vollständig.

Eigenschaften

$$\forall t \in \mathcal{W}(X) : f_X(t) \geq 0 \tag{11}$$

$$\forall t \notin \mathcal{W}(X) : f_X(t) = 0 \tag{12}$$

$$\int_{-\infty}^{\infty} f_X(t)\, dt = 1 \tag{13}$$

Transformation

Sei X eine *Zufallsvariable* mit *Dichtefunktion* f_X und $Y = g(X)$ bijektiv. Dann hat Y die *Dichtefunktion*:

$$f_Y(t) = \left| \frac{d}{dt} g^{-1}(t) \right| \cdot f_X\left(g^{-1}(t)\right) \tag{14}$$

Sei zum Beispiel $Y = aX + b$. Dann gilt: $f_Y(t) = \frac{1}{a} \cdot f_X\left(\frac{t-b}{a}\right)$

Verteilung μ

Die *Verteilung* $\mu_X : B \to [0,1]$ mit $B \subseteq \mathcal{W}(X)$ ist das Analogon zu *Gewichts-* bzw. *Dichtefunktion* für Bereiche (Teilmengen) der Wertemenge einer *Zufallsvariablen* X.
Anschaulich beantwortet $\mu_X(B)$, wie gross die *Wahrscheinlichkeit* ist, dass X einen Wert aus B annimmt.
Bei *stetigen Zufallsvariablen* ist B oft Teilintervall von $\mathcal{W}(X)$.

Verteilungsfunktion F

Die *Verteilungsfunktion* $F_X : \mathcal{W}(X) \to [0,1]$ gibt über die *Wahrscheinlichkeit* Auskunft, mit welcher eine *Zufallsvariable* X höchstens einen bestimmten Wert erreicht. Intuitiver und erlaubt ist die Schreibweise $P[X \leq t]$, wobei $t \in \mathcal{W}(X)$.

Eigenschaften

- F_X ist monoton wachsend und rechtsstetig.
- Es gilt:

$$\lim_{t \to -\infty} F_X(t) = 0 \tag{15}$$

$$\lim_{t \to \infty} F_X(t) = 1 \tag{16}$$

- Wenn X *diskret*:

$$F_X(t) = \sum_{\substack{i \in \mathcal{W}(X) \\ i \leq t}} p_X(i) = \sum_{\substack{i \in \mathcal{W}(X) \\ i \leq t}} P[X = i] \tag{17}$$

- Wenn X *stetig*:

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\ dx \tag{18}$$

- Für *diskrete* und *stetige* X gilt gleichermassen:

$$P[a < X \leq b] = F_X(b) - F_X(a) \tag{19}$$

- Bei *stetigen* X darf man innerhalb von $P[\ldots]$-Blöcken die Zeichen $<, \leq$ sowie $>, \geq$ frei vertauschen

Quantil und Median

Für $\alpha \in (0,1)$ heisst $q_\alpha = F_X^{-1}(\alpha)$ das **α-Quantil** der *Verteilungsfunktion* F. Es bedeutet, dass die *Zufallsvariable* X mit *Wahrscheinlichkeit* α Werte zwischen $-\infty$ und dem α-*Quantil* q_α annimmt.
Das $\frac{1}{2}$-Quantil einer *Verteilung* heisst **Median**.

Erwartungswert E

Der *Erwartungswert* zeigt, welchen Wert eine *Zufallsvariable* bei vielen Wiederholungen des Zufallsexperiments durchschnittlich annehmen wird. Oft schreibt man anstatt $E[X]$ auch μ (**Achtung:** nicht mit *Verteilung* verwechseln!).

$$\text{diskret:} \quad E[X] = \sum_{t \in \mathcal{W}(X)} t \cdot P[X = t] \tag{20}$$

$$\text{stetig:} \quad E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\ dx \tag{21}$$

Erlaubte Transformationen

- Für $Y = g(X)$, diskret gilt:

$$E[Y] = \sum_{t \in \mathcal{W}(X)} g(t) \cdot P[X = t] \tag{22}$$

- Für $Y = g(X)$, stetig gilt:

$$E[Y] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\ dx \tag{23}$$

- Für $Z = g(X) \cdot h(Y)$ mit X, Y *unabhängig* gilt:

$$E[Z] = E[g(X)] \cdot E[h(Y)] \tag{24}$$

Achtung: Umkehrung gilt nicht!

- Für $Y = a + \sum_{i=1}^{n} b_i X_i$ gilt:

$$E[Y] = a + \sum_{i=1}^{n} b_i \cdot E[X_i] \tag{25}$$

Erwartungswert einer Funktion mehrerer ZV

Seien $X_1, \ldots, X_n$ *Zufallsvariablen* mit *gemeinsamer Gewichtsfunktion* p bzw. *gemeinsamer Dichtefunktion* f und $Y = g(X_1, \ldots, X_n)$. Dann gilt je nach Fall:

$$E[Y] = \sum_{x_1, \ldots, x_n} g(x_1, \ldots, x_n) \cdot p(x_1, \ldots, x_n) \tag{26}$$

$$= \int \cdots \int g(x_1, \ldots, x_n) \cdot f(x_1, \ldots, x_n)\ dx_n \cdots dx_1 \tag{27}$$

Varianz Var

Die *Varianz* ist ein Mass für die Streuung einer *Zufallsvariable*. Sie stellt den Übergang zwischen der *Zufallsvariable* selbst und ihrer *Standardabweichung* dar. Oft schreibt man deshalb anstatt $Var[X]$ auch σ^2.

$$Var[X] = E[Y^2], \text{ wobei } Y = X - E[X] \tag{28}$$

Daraus folgt:

- Falls $E[X^2] < \infty$:

$$Var[X] = E[X^2] - E[X]^2 \tag{29}$$

- X diskret:

$$Var[X] = \sum_{t \in \mathcal{W}(X)} (t - \mu)^2 \cdot P[X = t] \tag{30}$$

- X stetig:

$$Var[X] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f_X(x)\ dx \tag{31}$$

- Für $Z = a + bX$ gilt:

$$Var[Z] = b^2 \cdot Var[X] \tag{32}$$

- Für $X_1, X_2, \ldots X_n$ *unabhängig* gilt:

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] \tag{33}$$

Standardabweichung σ

Die *Standardabweichung* ist das Mass für die Abweichung einer *Zufallsvariable* von ihrem *Erwartungswert*.

$$\sigma = \sqrt{\sigma^2} = \sqrt{Var[X]} \tag{34}$$

Kovarianz Cov

Die *Kovarianz* ist ein Mass für den Zusammenhang zwischen zwei *Zufallsvariablen* X und Y bzw. ihrer Streuung:

$$\begin{aligned} Cov(X,Y) &= E\left[(X - E[X])\,(Y - E[Y])\right] \\ &= E[X \cdot Y] - E[X] \cdot E[Y] \\ &= \frac{Var[X+Y] - Var[X] - Var[Y]}{2} \end{aligned} \tag{35}$$

Eigenschaften

$$Cov(X,X) = Var[X] \tag{36}$$

$$Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z) \tag{37}$$

$$\forall a \in \mathbb{R} : Cov(X, a) = 0 \tag{38}$$

$$\forall b \in \mathbb{R} : Cov(X, bY) = b \cdot Cov(X,Y) \tag{39}$$

Korrelation ϱ

Wie die *Kovarianz* ist auch die *Korrelation* ein Mass für den Zusammenhang zwischen zwei Zufallsvariablen X und Y. Im Gegensatz zu dieser ist die *Korrelation* jedoch dimensions- und einheitenlos und eignet sich deshalb für Vergleiche verschiedener Zusammenhänge.

$$\varrho = \frac{Cov(X,Y)}{\sqrt{Var[X] \cdot Var[Y]}} \tag{40}$$

Eigenschaften

- $-1 \leq \varrho(X,Y) \leq 1$
- $|\varrho(X,Y)| = 1$ gilt genau dann, wenn es Konstanten $a, b \in \mathbb{R}$, $b \neq 0$ gibt sodass $P[Y = a + bX] = 1$. X und Y heissen dann **perfekt korreliert**.

Diskrete Verteilungen

Diskrete Gleichverteilung

Einfachste Verteilung, bei welcher alle n Werte $\in \mathbb{Z}$ zwischen zwei Schranken $a \in \mathbb{Z}$ und $b \in \mathbb{Z}$ gleich wahrscheinlich sind.

Eigenschaften

$$P[X = t] = \begin{cases} \frac{1}{n} & a \leq t \leq b \\ 0 & \text{sonst} \end{cases} \tag{41}$$

$$F_X(t) = P[X \leq t] = \begin{cases} 0 & t < a \\ \frac{t-a+1}{n} & a \leq t \leq b \\ 1 & t > b \end{cases} \tag{42}$$

$$\mu_X(B) = \frac{|B|}{n} \tag{43}$$

$$\mu = E[X] = \frac{a+b}{2} \tag{44}$$

$$\sigma^2 = Var[X] = \frac{n^2-1}{12} \tag{45}$$

Beispiel: Ergebnis eines Münz- oder Würfelwurfs.

Bernoulli-Verteilung

Eine *Bernoulli-verteilte Zufallsvariable* mit Parameter p nimmt genau die beiden Werte 0 und 1 mit den *Wahrscheinlichkeiten* $1-p$ und p an. Also ist sie auch die *Indikatorfunktion* $\mathbf{1}_A$ für ein *Ereignis* A mit $P[A] = p$.

Schreibweise: $X \sim \text{Be}(p)$

Eigenschaften

$$P[X = t] = \begin{cases} 1-p & t = 0 \\ p & t = 1 \end{cases} \tag{46}$$

$$\mu = E[X] = p \tag{47}$$

$$\sigma^2 = Var[X] = p \cdot (1-p) \tag{48}$$

Binomialverteilung

Die *Binomialverteilung* beschreibt die Anzahl der Erfolge bei n *unabhängigen* 0-1-Experimenten mit *Erfolgswahrscheinlichkeit* p. Daraus folgt:

$$X = \sum_{i=1}^{n} Y_i \quad \text{mit} \quad Y_i \sim \text{Be}(p) \tag{49}$$

Schreibweise: $X \sim \text{Bin}(n,p)$

Eigenschaften

$$P[X = t] = \binom{n}{t} \cdot p^t \cdot (1-p)^{n-t} \tag{50}$$

$$\mu = E[X] = n \cdot p \tag{51}$$

$$\sigma^2 = Var[X] = n \cdot p \cdot (1-p) \tag{52}$$

Beispiel: Anzahl Köpfe beim 10-maligen Münzwurf

Geometrische Verteilung

Die *geometrische Verteilung* mit Parameter p beschreibt die Wartezeit auf den ersten Erfolg bei einer Folge unabhängiger 0-1-Experimente mit *Erfolgswahrscheinlichkeit* p.

$$X = \inf\{i \in \mathbb{N} | Y_i = 1\} \text{ mit } Y_i \sim \text{Be}(p) \tag{53}$$

Schreibweise: $X \sim \text{Geom}(p)$

Eigenschaften

$$P[X = t] = p \cdot (1-p)^{t-1} \tag{54}$$

$$F_X(t) = P[X \leq t] = 1 - (1-p)^t \tag{55}$$

$$\mu = E[X] = \frac{1}{p} \tag{56}$$

$$\sigma^2 = Var[X] = \frac{1}{p^2} - \frac{1}{p} \tag{57}$$

Gedächtnislosigkeit / forgetfulness property:

$$P[X \geq t + s \mid X \geq s] = P[X \geq t] \tag{58}$$

Die *Wahrscheinlichkeit*, dass man noch einen weiteren Zeitabschnitt t warten muss, hängt nicht davon ab, wie lange man schon gewartet hat.

Beispiel: Wartezeit auf Kopf bei wiederholtem Münzwurf

Negativbinomialverteilung

Die *Negativbinomialverteilung* mit Parametern r und p beschreibt die Wartezeit auf den r-ten Erfolg bei einer Folge unabhängiger 0-1-Experimente mit *Erfolgswahrscheinlichkeit* p. Daraus folgt $\text{NB}(1,p) \equiv \text{Geom}(p)$.

Schreibweise: $X \sim \text{NB}(r,p)$

Eigenschaften

$$P[X = t] = \binom{t-1}{r-1} \cdot p^r \cdot (1-p)^{t-r} \tag{59}$$

$$\mu = E[X] = \frac{r}{p} \tag{60}$$

$$\sigma^2 = Var[X] = \frac{r \cdot (1-p)}{p^2} \tag{61}$$

Für $X_1, \ldots, X_r \sim \text{Geom}(p)$ *i.i.d.* gilt:

$$X_1 + X_2 + \ldots + X_r \sim \text{NB}(r,p) \tag{62}$$

Beispiel: Wartezeit auf dritten Kopf beim wiederholtem Münzwurf

Hypergeometrische Verteilung

Eine *hypergeometrisch* verteilte *Zufallsvariable* beschreibt die Anzahl "guter" Kugeln, wenn man aus einer Urne mit n Kugeln, davon r "gute", m Mal ohne Zurücklegen zieht.

Eigenschaften

$$P[X = t] = \frac{\binom{r}{t} \cdot \binom{n-r}{m-t}}{\binom{n}{m}} \tag{63}$$

$$\mu = E[X] = \frac{r \cdot m}{n} \tag{64}$$

$$\sigma^2 = Var[X] = \frac{r \cdot m}{n} \cdot \left(1 - \frac{m}{n}\right) \cdot \frac{n-r}{n-1} \tag{65}$$

Poisson-Verteilung

Die *Poisson-Verteilung* mit Parameter $\lambda \in \mathbb{R}^+$ ist eine Approximation der *Binomialverteilung* für kleine p und grosse n, wobei $\lambda = np$. Die Approximation ist gut für $np^2 \leq 0.05$ und damit ein Modell für seltene Ereignisse.

Schreibweise: $X \sim \mathrm{P}(\lambda)$

Eigenschaften

$$P[X = t] = e^{-\lambda} \cdot \frac{\lambda^t}{t!} \tag{66}$$

$$\mu = E[X] = \lambda \tag{67}$$

$$\sigma^2 = Var[X] = \lambda \tag{68}$$

Beispiel: Anzahl eingehender Druckaufträge in $n = 3600$ Sekunden mit einer Jobwahrscheinlichkeit von $p = 1\%$ pro Sekunde, wenn pro Sekunde maximal ein Job eintreffen kann.

Stetige Verteilungen

Stetige Gleichverteilung

Bei der *stetigen Gleichverteilung* oder auch **uniformen Verteilung** sind alle Werte zwischen zwei Schranken a und b gleich Wahrscheinlich.

Schreibweise: $X \sim \mathcal{U}(a, b)$

Eigenschaften

$$f_X(t) = \begin{cases} \frac{1}{b-a} & a \leq t \leq b \\ 0 & \text{sonst} \end{cases} \tag{69}$$

$$F_X(t) = \begin{cases} 0 & t < a \\ \frac{t-a}{b-a} & a \leq t \leq b \\ 1 & t > b \end{cases} \tag{70}$$

$$\mu = E[X] = \frac{a+b}{2} \tag{71}$$

$$\sigma^2 = Var[X] = \frac{(b-a)^2}{12} \tag{72}$$

Exponentialverteilung

Die *Exponentialverteilung* mit Parameter $\lambda \in \mathbb{R}^+$ ("Ausfallrate") ist das stetige Analogon zur *geometrischen Verteilung* und beschreibt eine Lebensdauer oder eine Wartezeit.

Schreibweise: $X \sim \mathrm{Exp}(\lambda)$

Eigenschaften

$$f_X(t) = \begin{cases} 0 & t < 0 \\ \lambda \cdot e^{-\lambda t} & t \geq 0 \end{cases} \tag{73}$$

$$F_X(t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-\lambda t} & t \geq 0 \end{cases} \tag{74}$$

$$\mu = E[X] = \frac{1}{\lambda} \tag{75}$$

$$\sigma^2 = Var[X] = \frac{1}{\lambda^2} \tag{76}$$

Gedächtnislosigkeit / forgetfulness property:

$$P[X > t + s \mid X > s] = P[X \geq t] \tag{77}$$

Normalverteilung

Die *Normalverteilung* mit *Erwartungswert* $\mu \in \mathbb{R}$ und *Varianz* $\sigma^2 \in \mathbb{R}^+$ modelliert Summen oder Mittelwerte vieler *i.i.d.* Zufallsvariablen (dies folgt aus dem *zentralen Grenzwertsatz*). Meistens beschreibt sie die Ergebnisse von Messungen "aus der Natur".

Schreibweise: $X \sim \mathcal{N}(\mu, \sigma^2)$

Eigenschaften

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}} \tag{78}$$

$$F_X(t) = \text{tabelliert durch } \Phi(t) \tag{79}$$

$$\mu = E[X] = \mu \tag{80}$$

$$\sigma^2 = Var[X] = \sigma^2 \tag{81}$$

Beispiel: Gewicht von Kürbissen

Standardnormalverteilung

$\mathcal{N}(0, 1)$ heisst *Standardnormalverteilung*. Ihre Dichte $F_X(t) \equiv \Phi(t)$ ist tabelliert.

Eigenschaft:

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y = \frac{X-\mu}{\sigma}$. Dann gilt $Y \sim \mathcal{N}(0, 1)$ und:

$$P[X \leq x] = P\left[Y \leq \frac{x-\mu}{\sigma}\right] = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{82}$$

Chiquadratverteilung

Die *Chiquadratverteilung* mit m Freiheitsgraden taucht bei bestimmten Tests auf und entsteht durch Quadrieren und Summieren von m *standardnormalverteilten Zufallsvariablen.*

Schreibweise: $X \sim \chi^2_m$

Eigenschaft:

$$\chi^2_m \equiv \sum_{i=1}^{m} X_i^2 \quad \text{mit} \quad X_i \sim \mathcal{N}(0, 1) \tag{83}$$

t-Verteilung

Die *t-Verteilung* mit m Freiheitsgraden ist symmetrisch und tritt bei Tests auf.

Schreibweise: $X \sim t_m$

Eigenschaften

$$f_X(t) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\cdot\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{x^2}{m}\right)^{-\frac{m+1}{2}} \quad (84)$$

$$\mu = E[X] = 0 \text{ für } n > 1 \quad (85)$$

$$\sigma^2 = Var[X] = \frac{n}{n-2} \text{ für } n > 2 \quad (86)$$

Simulation

Sei F eine stetige, streng monoton wachsende und bijektive *Verteilungsfunktion* mit Umkehrfunktion F^{-1}. Sei $X \sim \mathcal{U}(0,1)$ gleichverteilt und $Y = F^{-1}(X)$.
Dann hat Y gerade die *Verteilungsfunktion* F.

Anwendung: PC liefert immer uniforme Zufallszahlen zwischen 0 und 1. Wir wollen aber eine andere Verteilung haben.

Beispiel Exponentialverteilung

$$s = F(t) = 1 - e^{-\lambda t} \quad (87)$$

$$\text{Umkehrfunktion:} \quad t = F^{-1}(s) = -\frac{\log(1-s)}{\lambda}$$

Sei nun x ein zufällig generierter Wert der *Zufallsvariable* X mit $X \sim \mathcal{U}(0,1)$. Aufgrund obigen Formeln gilt:

$$y = F^{-1}(x) = -\frac{\log(1-x)}{\lambda} \sim \text{Exp}(\lambda) \quad (88)$$

Gemeinsame Verteilungen

Funktionen

Analog zum Fall für eine *Zufallsvariable* existieren für gemeinsame Verteilungen:

- **Gemeinsame Gewichtsfunktion**, falls alle X_i diskret:

$$p(x_1,\dots,x_n) = P[X_1 = x_1,\dots,X_n = x_n] \quad (89)$$

- **Gemeinsame Dichtefunktion**, falls alle X_i stetig:

$$f(x_1,x_2,\dots,x_n) \quad (90)$$

- **Gemeinsame Verteilungsfunktion**

$$F(x_1,\dots,x_n) = P[X_1 \le x_1,\dots,X_n \le x_n] \quad (91)$$

Im diskreten Fall gilt:

$$F(x_1,x_2,\dots,x_n) \quad (92)$$

$$= \sum_{\substack{t_1\in\mathcal{W}(X_1)\\ t_1\le x_1}} \cdots \sum_{\substack{t_n\in\mathcal{W}(X_n)\\ t_n\le x_n}} P[X_1 = t_1,\dots,X_n = x_n] \quad (93)$$

Im stetigen Fall gilt:

$$F(x_1,x_2,\dots,x_n) \quad (94)$$

$$= \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f(t_1,\dots,t_n)\; dt_n\cdots dt_1$$

Darstellung für $n = 2$

- Im diskreten Fall: Darstellung als Tabelle
- Im stetigen Fall: Darstellung als Flächengraph

Randverteilungen

Möchte man ausgehend von einer gemeinsamen Verteilung bestimmte *Zufallsvariablen* unbetrachtet lassen und quasi “Einzelverteilungen” herausfinden, dann:

Gewichtsfunktion: unbetrachtete ZV wegsummieren:

$$P[X=x] = \sum_{t_1\in\mathcal{W}(Y_1)}\cdots\sum_{t_{n-1}\in\mathcal{W}(Y_{n-1})} P[X=x, Y_1=y_1,\dots,Y_{n-1}=y_{n-1}] \quad (95)$$

Dichtefunktion: unbetrachtete ZV wegintegrieren:

$$f_X(x) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x,y_1,\dots,y_{n-1})\; dy_1\cdots dy_{n-1} \quad (96)$$

Verteilungsfunktion: Limes bilden:

$$F_X(x) = \lim_{y_1\to\infty}\cdots\lim_{y_{n-1}\to\infty} F(x,y_1,\dots,y_{n-1}) \quad (97)$$

Unabhängigkeit von Zufallsvariablen

Die *Zufallsvariablen* $X_1, X_2, \dots, X_n$ heissen *unabhängig*, falls:

$$F(x_1,\dots,x_n) = F_{X_1}(x_1)\cdots F_{X_n}(x_n) \quad (98)$$

Dies ist je nach Fall (diskret oder stetig) äquivalent zu:

$$P[X_1=x_1,\dots,X_n=x_n] \quad (99)$$

$$= P[X_1=x_1]\cdots P[X_n=x_n]$$

$$f(x_1,\dots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n) \quad (100)$$

Sind Zufallsvariablen gleich verteilt und unabhängig, so schreibt man **i.i.d.** für “independent identically distributed”.

Bedingte Verteilungen

Bedingte Gewichtsfunktion

Die *bedingte Gewichtsfunktion* einer *Zufallsvariablen* X, gegeben dass $Y = y$, ist definiert als:

$$P[X=x \mid Y=y] = \frac{P[X=x, Y=y]}{P[Y=y]} \quad (101)$$

X und Y sind *unabhängig* genau dann, wenn $\forall x \in \mathcal{W}(X)$ gilt:

$$P[X=x \mid Y=y] = P[X=x] \quad (102)$$

Bedingte Dichtefunktion

Die *bedingte Dichtefunktion* einer *Zufallsvariablen* X, gegeben $Y = y$, ist definiert als:

$$f(x \mid y) = \frac{f(x,y)}{f(y)} \quad (103)$$

X und Y sind *unabhängig* genau dann, wenn $\forall x$ gilt:

$$f(x \mid y) = f(x) \quad (104)$$

Grenzwertsätze

Schwaches Gesetz der grossen Zahlen

Seien $X_1, X_2, \ldots$ *unabhängige* oder zumindest paarweise *unkorrelierte Zufallsvariablen* mit gleichem *Erwartungswert* μ und gleicher Varianz σ^2. Sei $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Dann gilt:

$$\forall \varepsilon > 0 : \lim_{n\to\infty} P\left[|\overline{X}_n - \mu| > \varepsilon\right] = 0 \tag{105}$$

Starkes Gesetz der grossen Zahlen

Seien $X_1, X_2, \ldots$ *i.i.d. Zufallsvariablen* mit endlichem *Erwartungswert* μ. Sei $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Dann gilt:

$$P\left[\lim_{n\to\infty} \overline{X}_n = \mu\right] = 1 \tag{106}$$

Markow-Ungleichung

Sei X eine Zufallsvariable, $a \in \mathbb{R}^+$ und $h : \mathbb{R} \to \mathbb{R}^+$ eine Funktion. Dann gilt:

$$P[h(X) \geq a] \leq \frac{E[h(X)]}{a} \tag{107}$$

Chebyshev-Ungleichung

Sei Y *Zufallsvariable* mit *Erwartungswert* μ und *Varianz* σ^2:

$$\forall \varepsilon > 0 : P\left[|Y - \mu| > \varepsilon\right] \leq \frac{\sigma^2}{\varepsilon^2} \tag{108}$$

Anwendung: Sehr leicht berechenbare Schranke für jede beliebige Verteilung.

Chernoff-Schranken

Seien $X_1, X_2, \ldots, X_n$ *i.i.d.* mit $X_i \sim \mathrm{Be}(p)$. Sei $S_n = \sum_{i=1}^n X_i \sim \mathrm{Bin}(n,p)$ und damit $E[S_n] = np$.

$$\forall \varepsilon > 0 : P\left[S_n \geq (1+\varepsilon)E[S_n]\right] \leq \left(\frac{e^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}}\right)^{E[S_n]} \tag{109}$$

Anwendung: Viel einfacher zu rechnen als die *Verteilungsfunktion* der *Binomialverteilung*, liefert im Gegensatz zur *Poisson-Approximation* eine exakte Schranke.

Zentraler Grenzwertsatz

Seien $X_1, X_2, \ldots$ *i.i.d. Zufallsvariablen* mit $E[X_i] = \mu$ und $Var[X_i] = \sigma^2$. Sei $S_n = \sum_{i=1}^n X_i$. Dann gilt:

$$\lim_{n\to\infty} \left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right] = \Phi(t) \tag{110}$$

Oft wird folgende Abkürzung benutzt:

$$S_n^* = \frac{S_n - E[S_n]}{\sqrt{Var[S_n]}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \tag{111}$$

Die entstandene *Zufallsvariable* S_n^* heisst **Standardisierung** von S_n. Der *zentrale Grenzwertsatz* lässt sich so schreiben als:

$$\begin{aligned} n \text{ gross} \;\Leftrightarrow\;& S_n^* \overset{\text{approx}}{\sim} \mathcal{N}(0,1) \\ \Leftrightarrow\;& P[S_n^* \leq t] \approx \Phi(t) \\ \Leftrightarrow\;& S_n \overset{\text{approx}}{\sim} \mathcal{N}(n\mu, n\sigma^2) \end{aligned} \tag{112}$$

Für $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ gilt: $\overline{X}_n \overset{\text{approx}}{\sim} \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$

Statistik

Stichproben

In der Statistik möchte man aus beobachteten Daten Rückschlüsse auf die *Verteilungen* ziehen, welche diese Daten generiert haben.
Eine von einem *Elementarereignis* ω erzeugte Datenmenge heisst *Stichprobe*. Ihre Daten $x_1, x_2, \ldots, x_n$ heissen **Realisierungen** und werden als Werte $X_1(\omega), \ldots, X_n(\omega)$ von *Zufallsvariablen* $X_1, \ldots, X_n$ betrachtet.

Schätzer

Betrachtet man jede der zu charakterisierenden *Zufallsvariablen* $X_1, \ldots, X_n$ als abhängig von Parametern ϑ_i, so reduziert sich das Ziel darauf, diese Parameter $\vartheta_1, \ldots, \vartheta_m$ zu schätzen.
Ein *Schätzer* T_i ist eine *Zufallsvariable*, die in Abhängigkeit der Daten genau einen der gesuchten Parameter ϑ_i schätzt.

Erwartungstreue

Ein *Schätzer* T_i heisst *erwartungstreu* für ϑ_i, falls $E[T_i] = \vartheta_i$. Im Mittel über alle *Realisationen* schätzt er also richtig.

Konsistenz

Erweitert man einen *Schätzer* so, dass er weitere *Realisierungen* derselben *Zufallsvariablen* berücksichtigt, heisst er *konsistent*, falls er für eine grössere Anzahl n immer genauer wird:

$$P\left[\lim_{n\to\infty} T_i^{(n)} = \vartheta_i\right] = 1 \tag{113}$$

Maximum-Likelihood-Methode

Die **Likelihood-Funktion** L gibt die *Wahrscheinlichkeit* einer Stichprobe unter einem Parameter ϑ wieder:

$$L(x_1, \ldots, x_n; \vartheta) = \begin{cases} p(x_1, \ldots, x_n; \vartheta) & \text{diskreter Fall} \\ f(x_1, \ldots, x_n; \vartheta) & \text{stetiger Fall} \end{cases} \tag{114}$$

Sind die *Zufallsvariablen* X_i *i.i.d.*, so gilt:

$$L(x_1, \ldots, x_n; \vartheta) = \begin{cases} \prod_{i=1}^n p_X(x_i; \vartheta) & \text{diskreter Fall} \\ \prod_{i=1}^n f_X(x_i; \vartheta) & \text{stetiger Fall} \end{cases} \tag{115}$$

$\log L(x_1, \ldots, x_n; \vartheta)$ heisst **log-Likelihood-Funktion**.
Der **Maximum-Likelihood-Schätzer** T^{ML} für ϑ ist dadurch definiert, dass er die *Likelihood-Funktion* L über ϑ maximiert. Dadurch schätzt er ϑ genau so, dass die gegebene *Stichprobe* unter ϑ am wahrscheinlichsten ist.
Da es sich bei L meistens um ein Produkt handelt und die log-Funktion streng monoton steigend ist, kann man auch $\log L$ über ϑ minimieren, was einfacher zu rechnen ist.

Häufig gebrauchte ML-Schätzer

- $X_i \sim \mathrm{Be}(p)$:

$$T_p = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}_n \tag{116}$$

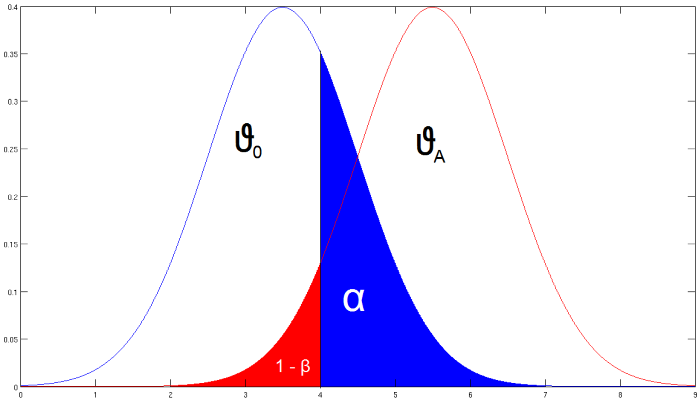

Abbildung 1: Ein Test mit Fehlern erster und zweiter Art

- $X_i \sim \mathcal{N}(\mu, \sigma^2)$:

$$T_\mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n \tag{117}$$

$$T_{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \tag{118}$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \left(\overline{X}_n\right)^2$$

Achtung: T_{σ^2} ist nicht erwartungstreu:

$$E[T_{\sigma^2}] = \frac{n-1}{n} Var[X] \tag{119}$$

Deshalb benutzt man meistens die **empirische Stichprobenvarianz** S^2:

$$S^2 = T'_{\sigma^2} = \frac{n}{n-1} T_{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \tag{120}$$

Tests

Ausgehend von einer *Stichprobe* möchte man zwei Verteilungsparameter (bzw. Parameterbereiche) gegeneinander abwägen: Eine **Nullhypothese** Θ_0 und eine **Alternativhypothese** Θ_A. Man wählt nun eine *Zufallsvariable* T namens **Teststatistik** abhängig von ϑ, welche die Daten der *Stichprobe* auswertet. Dabei wollen wir die *Nullhypothese* ablehnen, wenn der realisierte Wert $T(x_1, \ldots, x_n)$ in einen **kritischen Bereich** oder **Verwerfungsbereich** K fällt.
Die Entscheidung kann auf zwei Arten falsch herauskommen:

- **Fehler 1.Art**: Die *Nullhypothese* wird abgelehnt, obwohl sie eigentlich richtig wäre. Die *Wahrscheinlichkeit* für einen *Fehler 1. Art* heisst **Signifikanzniveau** α:

$$\alpha = P_{\vartheta_0}[T \in K], \quad \vartheta_0 \in \Theta_0 \tag{121}$$

- **Fehler 2.Art**: Die *Nullhypothese* wird angenommen, obwohl sie falsch wäre. Die *Wahrscheinlichkeit* für einen *Fehler 2. Art* beträgt $1-\beta$, wobei β als die **Macht** des Tests bezeichnet wird:

$$\beta = P_{\vartheta_A}[T \in K], \quad \vartheta_A \in \Theta_A \tag{122}$$

Da es schwieriger ist, eine Hypothese zu verwerfen, wählen wir als *Nullhypothese* das Gegenteil der eigentlich gewünschten Aussage. Ausserdem wird immer ein *Signifikanzniveau* α gegeben. Falls sich der *kritische Bereich* z.B. am "rechten Rand" von T befindet, berechnet sich der **kritische Wert** c, also dessen untere Grenze (exklusiv), zu:

$$P_{\vartheta_0}[T \geq c] = \alpha \quad \text{nach } c \text{ auflösen} \tag{123}$$

Bei einer zweiseitigen *Alternativhypothese* sind gegebenenfalls Symmetrien auszunutzen.
Die Wahrscheinlichkeit unter der *Nullhypothese*, den erhaltenen Wert für die *Teststatistik* zu bekommen, heisst **P-Wert**. Ist er kleiner als α, so verwirft man die *Nullhypothese*.

z-Test

Der *z-Test* ist ein Test für *Erwartungswert* bei bekannter *Varianz* σ^2. Es seien also $X_1, \ldots, X_n \sim \mathcal{N}(\vartheta, \sigma^2)$ *i.i.d.*. Wir wollen die *Nullhypothese* $\vartheta = \vartheta_0$ testen. *Teststatistik*:

$$T = \frac{\overline{X}_n - \vartheta_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1) \tag{124}$$

Den *kritischen Bereich* K bestimmt man abhängig von der *Alternativhypothese* Θ_A:

- $\vartheta_A > \vartheta_0$ einseitig: $K = (c_>, \infty)$ mit

$$c_> = (1-\alpha)\text{-Quantil} = \Phi^{-1}(1-\alpha) \tag{125}$$

- $\vartheta_A < \vartheta_0$ einseitig: $K = (-\infty, c_>)$ mit

$$c_< = \alpha\text{-Quantil} = \Phi^{-1}(\alpha) \tag{126}$$

- $\vartheta_A \neq \vartheta_0$ zweiseitig: $K = (-\infty, c_{\neq}) \cup (c_{\neq}, \infty)$

$$c_{\neq} = \left(1 - \frac{\alpha}{2}\right)\text{-Quantil} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{127}$$

t-Test

Der *t-Test* ist ein Test für *Erwartungswert* bei unbekannter *Varianz*. Es seien also $X_1, \ldots, X_n \sim \mathcal{N}(\vartheta, \sigma^2)$ und wir möchten die *Nullhypothese* $\vartheta = \vartheta_0$ testen. *Teststatistik*:

$$T = \frac{\overline{X}_n - \vartheta_0}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1} \tag{128}$$

Dabei bezeichnet S^2 den *Schätzer* für die *empirische Stichprobenvarianz*. Der *kritische Bereich* K lässt sich analog zum *z-Test* mit den *Quantilen* der t_{n-1}-Verteilung bestimmen.

Beispiele

t-Test

Aufgabe

Ein Waschmittelhersteller bringt 5kg-Packungen in den Umlauf. Die Konsumentenschutzorganisation kauft 25 Packungen. Es ergibt sich ein Mittel von $\overline{X} = 4.9$ kg und eine empirische Stichprobenvarianz $S^2 = 0.1\text{kg}^2$. Die einzelnen Gewichte seien durch unabhängige $\mathcal{N}(\mu, \sigma^2)$ Zufallsvariablen beschrieben.

1. Wie lauten die Hypothesen H_0 und H_A ?
2. Wie ist $(\overline{X} - 5)/(S/5))$ verteilt unter H_0 ?
3. Führen Sie den t-Test auf dem 5%-Niveau durch. Wie gross ist der P-Wert? Wird die Nullhypothese verworfen?
4. Berechnen Sie das 95%-Vertrauensintervall für den in 3. beschriebenen Test.

Lösung

1. $H_0 : \mu = 5$, $H_1 : \mu < 5$
2. Verteilung: $\sim t_{5^2-1} = t_{24}$
3. T-Statistik: $T_1 = \frac{\overline{X}-\mu_0}{\sqrt{\frac{S^2}{n}}} = -1.581$. P-Wert $= P_{\mu_0}[T < c] = 0.0653$. Gesucht ist ein c, sodass $P_{\mu_0}[T < c] = 0.05$ bzw $P_{\mu_0}[T > -c] = 0.95$. Dies ergibt $c = -1.711$. Weil $T_1 = -1.581 > c$ wird die Nullhypothese nicht verworfen.
4. Die Formel für das Vertrauensintervall ergibt $I_{0.95} = [4.792, 5.008]$. Da unser Test aber einseitig ist, wird das Intervall $I_{0.95} = [-\infty, 5.008]$.

Folge von Zufallsvariablen

Aufgabe

Seien $X_1, \ldots, X_n$ iid. Poisson-verteilte Zufallsvariablen mit Parameter $\lambda = 2$. Gegeben sei nun die Folge:

$$\frac{2n+1}{n^2} \sum_{i=1}^{n} X_i$$

Wie gross ist der Grenzwert?

Lösung

$$\frac{2n+1}{n^2} \sum_{i=1}^{n} X_i = \frac{2n+1}{n} \cdot \frac{1}{n} \sum_{i=1}^{n} X_i$$

Der zentrale Grenzwertsatz liefert für $Y \sim \mathcal{N}(2, 4/n)$

$$\lim_{n \to \infty} \frac{2n+1}{n^2} \sum_{i=1}^{n} X_i = 2 \cdot E[Y] = 4$$

Simulation(1)

Aufgabe

Die Zufallsvariable X sei Pareto-verteilt und habe die Verteilungsfunktion:

$$F_X(x|x_0, a) = \begin{cases} 1 - \left(\frac{x}{x_0}\right)^{-\alpha} & \alpha > 1, x \geq x_0 \\ 0 & \text{sonst} \end{cases}$$

$$f_X(x|x_0, a) = \begin{cases} \alpha x_0^{\alpha} x^{-\alpha-1} & \alpha > 1, x \geq x_0 \\ 0 & \text{sonst} \end{cases}$$

Ein auf [0,1] uniform verteilter Zufallsgenerator habe nun die Werte 0.237 und 0.733 generiert. Berechne daraus 2 Realisierungen der einer Pareto-verteilten Zufallsvariable mit $\alpha = 2$ und $x_0 = 2 \cdot 10^6$.

Lösung

Die Inverse F_X^{-1} der Verteilungsfunktion, ausgewertet an den generierten Stellen $\in [0, 1]$ ergibt die Realisierungen der Verteilung F_X.

$$\text{für} x \geq x_0 : F_X^{-1}(y) = \frac{x_0}{(1-y)^{1/\alpha}}$$

Jetzt müssen nur noch die beiden Werte eingesetzt werden.

Simulation(2)

Aufgabe

Die Cauchy-Verteilung hat die Dichte

$$f(x) = \frac{1}{\pi \cdot (1 + x^2)} \text{ für } x \in \mathcal{R}$$

Konstruieren Sie eine Cauchy-verteile Stichprobe vom Umfang 4 aus der folgenden Exp(1)

$$x_1 = 0.417, \quad x_2 = 0.448, \quad x_3 = 0.114, \quad x_4 = 0.665$$

Lösung

Die Verteilungsfunktion von $X \sim \text{Exp}(1)$ ist $F(x) = 1 - e^{-x}$. Wenn man auf exponentialverteilte Werte die Verteilungsfunktion der Exponentialverteilung anwendet, erhält man $[0, 1]$-uniform-verteilte Werte. Dies liefert in diesem Beispiel durch einsetzen der x_i in F die uniform verteilten y_i:

$$y_1 = 0.341, \quad y_2 = 0.361, \quad y_3 = 0.107, \quad y_4 = 0.486$$

Nun muss man nur noch die y_i in F^{-1} der Cauchy-Verteilung einsetzen für Cauchy-verteilte z_i.

$$F(x) = \int_{-\infty}^{x} f(x)dx = \frac{1}{2} + \frac{1}{\pi} \arctan(x)$$

$$\to F^{-1}(y) = \tan(\pi \cdot (y - \frac{1}{2}))$$

$$z_1 = -0.544, \quad z_2 = -0.465, \quad z_3 = -2.836, \quad z_4 = -0.043$$