1 Allgemeine Zufallsvariablen

Def. (Verteilungsfunktion)

$$t \mapsto F_X(t) := \mathrm{P}[X \le t] := \mathrm{P}[\{\omega \in \Omega \mid X(\omega) \le t\}].$$

Jede Verteilungsfunktion F_X hat die folgenden Eigenschaften:

1) F_X ist i) wachsend, d.h., es gilt $F_X(s) \le F_X(t)$ für $s \le t$, und ii) rechtsstetig, d.h., es gilt $F_X(u) \to F_X(t)$ für $u \to t$ mit $u > t$.
2) $\lim_{t\to-\infty} F_X(t) = 0$, und $\lim_{t\to+\infty} F_X(t) = 1$.

Umgekehrt kann man zeigen, dass jede Funktion F mit den Eigenschaften 1) und 2) die Verteilungsfunktion F_X einer Zufallsvariablen X ist.

Kennt man die Verteilung kennt man die Verteilungsfunkt. und umgekehrt $\mu_X(B) := \mathrm{P}[X \in B]$ und $F_X(t) = \mu_X((-\infty, t])$.

Def. (gemeinsame Verteilungsfunktion)
ZV $X_1, \dots, X_n$, Abb. $F\colon \mathbb{R}^n \to [0,1]$,

$$(x_1, \dots, x_n) \mapsto F(x_1, \dots, x_n) := \mathrm{P}[X_1 \le x_1, \dots, X_n \le x_n].$$

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

Falls $X_1, \dots, X_n$ eine gemeinsame Dichte f haben, so gilt analog zum diskreten Fall

i) $f(x_1, \dots, x_n) \ge 0$, und $= 0$ ausserh. v. $\mathcal{W}(\dots)$.
ii) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n)\, dx_n \dots dx_1 = 1$.
iii) $\mathrm{P}[(X_1, \dots, X_n) \in A] = \int \cdots \int_{(x_1,\dots,x_n)\in A} f(x_1, \dots, x_n)\, dx_n \dots dx_1$ für $A \subseteq \mathbb{R}^n$.

Def. (Verteilungsfkt. der Randverteilung von X)
ZV X, Y, $F_X\colon \mathbb{R} \to [0,1]$

$$x \mapsto F_X(x) := \mathrm{P}[X \le x] = \mathrm{P}[X \le x, Y < \infty] = \lim_{y\to\infty} F(x,y)$$

Com. $F_Y\colon \mathbb{R} \to [0,1]$ analog.

Def. (Dichtefunktion einer Randverteilung)
ZV X, Y mit gemeinsamer Dichte $f(x,y)$
Randverteilung von X ist $f_X\colon \mathbb{R} \to [0, \infty)$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$$

Com. $f_Y(y)$ analog.

Def. (Unabhängigkeit) Die ZV $X_1, \dots, X_n$ heissen unabhängig, falls gilt

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

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

Def. (Faltung von stetigen Zufallsvariablen)
Sind X und Y zusätzlich unabhängig, so ist $f(x,y) = f_X(x) f_Y(y)$ und damit

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy$$

d.h. $f_Z = (f_X * f_Y)(z)$ ist die Faltung f_X und f_Y.

Def. (Bedingte Gewichtsfkt. und Vert.)
ZV X, Y, gegeben

$$p_{X|Y}(x \mid y) := \mathrm{P}[X = x \mid Y = y] = \frac{\mathrm{P}[X = x, Y = y]}{\mathrm{P}[Y = y]} = \frac{p(x,y)}{p_Y(y)}$$

für $p_Y(y) > 0$ und 0 sonst. Das legt dann auch die bedingte Verteilung von X gegeben $Y = y$ fest.

Lem. (Affine Transformation von ZV)
Sei $g(x) = ax + b$ mit $a > 0$, $b \in \mathbb{R}$ und $Y = g(X) = aX + b$. Es folgt

$$\begin{aligned} F_Y(t) &= \mathrm{P}[Y \le t] = \mathrm{P}[aX + b \le t] \\ &= \mathrm{P}\left[X \le \frac{t-b}{a}\right] = F_X\left(\frac{t-b}{a}\right) \end{aligned}$$

und damit nach der Kettenregel

$$f_Y(t) = \frac{d}{dt} F_Y(t) = \frac{1}{a} f_X\left(\frac{t-b}{a}\right).$$

Analog bis auf ein Vorzeichen geht das auch mit $a < 0$.

Ex. (Affine Transform. von normalvert. ZV)
Ist $X \sim \mathcal{N}(\mu, \sigma^2)$ und $Y = aX + b$, $a > 0$, $b \in \mathbb{R}$, so ist $Y \sim \mathcal{N}(b + a\mu, a^2\sigma^2)$. (Bew: obige Formel.)

Ex. (Nichtlineare Transformation 1)
Für $g(x) = x^2$ ist $Y = X^2$ und

$$\begin{aligned} F_Y(t) &= \mathrm{P}[Y \le t] = \mathrm{P}[X^2 \le t] = \mathrm{P}\left[-\sqrt{t} \le X \le \sqrt{t}\right] \\ &= \mathrm{P}\left[X \le \sqrt{t}\right] - \mathrm{P}\left[X \le -\sqrt{t}\right] = F_X(\sqrt{t}) - F_X(-\sqrt{t}), \end{aligned}$$

falls X stetig ist. Mit d. Kettenregel erhält man $f_Y(t)$.

Ex. (Nichtlineare Transformation 2)
Für $g(x) = \frac{1}{x}$ ist $Y = \frac{1}{X}$ und

$$\begin{aligned} F_Y(t) &= \mathrm{P}[Y \le t] = \mathrm{P}\left[\frac{1}{X} \le t\right] \overset{(*)}{=} \mathrm{P}\left[X \ge \frac{1}{t}\right] \\ &= 1 - \mathrm{P}\left[X \le \frac{1}{t}\right] = 1 - F_X\left(\frac{1}{t}\right), \end{aligned}$$

falls X stetig ist. $f_Y(t) \to$ Ableiten und Kettenregel.
(*): Keine Mult. Einfach Bruch umkehren.

Thm. (Vert. mit Umkehrfkt. und $X \sim \mathcal{U}(0,1)$)
Sei F eine stetige und streng monoton wachsende Verteilungsfunktion, mit Umkehrfunktion F^{-1}. Ist $X \sim \mathcal{U}(0,1)$

und $Y = F^{-1}(X)$, so hat Y gerade die Verteilungsfunktion F.

Ex. (Simulation der Exponentialverteilung) Um eine Exponentialverteilung mit Parameter λ zu simulieren, nehmen wir die zugehörige Verteilungsfunktion $F(t) = 1 - e^{-\lambda t}$ für $t \geq 0$. Ihre Inverse F^{-1} erhalten wir via

$$y = F(t) = 1 - e^{-\lambda t} = 1 - e^{-\lambda F^{-1}(y)}$$

als

$$F^{-1}(y) = t = -\frac{\ln(1-y)}{\lambda}.$$

Mit $X \sim \mathcal{U}(0,1)$ ist also

$$Y := F^{-1}(X) = -\frac{\ln(1-X)}{\lambda} \sim Exp(\lambda).$$

2 Ungleichungen und Grenzwertsätze

2.1 Ungleichungen

Thm. (Markov Ungleichung) Sei X eine Zufallsvariable und ferner $g\colon \mathcal{W}(X) \to [0,\infty)$ eine wachsende Funktion. Für jedes $c \in \mathbb{R}$ mit $g(c) > 0$ gilt dann

$$\mathrm{P}[X \geq c] \leq \frac{\mathrm{E}[g(X)]}{g(c)}.$$

Cor. (Chebyshev-Ungleichung) Sei Y eine Zufallsvariable mit endlicher Varianz. Für jedes $b > 0$ gilt dann

$$\mathrm{P}[|Y - \mathrm{E}[Y]| \geq b] \leq \frac{\mathrm{Var}[Y]}{b^2},$$

bzw.

$$\mathrm{P}[|Y - \mathrm{E}[Y]| \leq b] \geq 1 - \frac{\mathrm{Var}[Y]}{b^2}.$$

Die obigen Ungleichungen illustrieren die Bedeutung der Varianz als Streunungsmass: Je kleiner die Varianz, desto eher (mit desto grösserer Wahrscheinlichkeit) liegen die Werte von X nahe beim Erwartungswert $\mathrm{E}[X]$.

2.2 Das Gesetz der grossen Zahlen (Asymptotik von Zufallsvariablen)

Thm. (schwaches Gesetz der grossen Zahlen) Sei $X_1, X_2, \dots$ eine Folge von unabhängigen (es genügt schon: paarweise unkorreliert) Zufallsvariablen, die alle den gleichen Erwartungswert $\mathrm{E}[X_i] = \mu$ und die gleiche Varianz $\mathrm{Var}[X_i] = \sigma^2$ haben. Sei $\overline{X}_n = \frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Dann konvergiert $\overline{X}_n$ für $n \to \infty$ in Wahrscheinlichkeit/stochastisch gegen $\mu = \mathrm{E}[X_i]$, d.h.

$$\mathrm{P}\left[\left|\overline{X}_i - \mu\right| > \epsilon\right] \xrightarrow{n\to\infty} 0 \qquad \text{für jedes } \epsilon > 0.$$

Anschaulich bedeutet die obige Aussage, dass mit beliebig grosser Wahrscheinlichkeit $1-\delta$ der Wert von $\overline{X}_n$ für hinreichend grosse n beliebige nahe (bis auf ϵ) bei μ liegt.

Thm. (starkes Gesetz der grossen Zahlen) Sei $X_1, X_2, \dots$ eine Folge von unabhängigen Zufallsvariablen, die alle dieselbe Verteilung haben, und ihr Erwartungswert $\mu = \mathrm{E}[X_i]$ sei endlich. Für $\overline{X}_n = \frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^n X_i$ gilt dann

$$\overline{X}_n \xrightarrow{n\to\infty} \mu \quad P\text{-fastsicher } (P\text{-f.s.}),$$

d.h.

$$\mathrm{P}\left[\left\{\omega \in \Omega \,\middle|\, \overline{X}_n(\omega) \xrightarrow{n\to\infty} \mu\right\}\right] = 1.$$

2.3 Der Zentrale Grenzwertsatz (Asymptotik von Verteilungen)

Thm. (Zentraler Grenzwertsatz, ZGS) Sei $X_1, X_2, \dots$ eine Folge von i.i.d. Zufallsvariablen mit $\mathrm{E}[X_i] = \mu$ und $\mathrm{Var}[X_i] = \sigma^2$. Für die Summe $S_n = \sum_{i=1}^n X_i$ gilt dann

$$\lim_{n\to\infty} \mathrm{P}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right] = \Phi(x) \quad \text{für alle } x \in \mathbb{R},$$

wobei Φ die Verteilungsfunktion der $\mathcal{N}(0,1)$-Verteilung ist.

Com. Wie im Beweis des schwachen GGZ kann man nachrechnen, dass S_n Erwartungswert $\mathrm{E}[S_n] = n\mu$ und Varianz $\mathrm{Var}[S_n] = n\sigma^2$ hat. Also hat die Grösse

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

Erwartungswert 0 und Varianz 1; sie heisst deshalb auch die Standardisierung von S_n.

Ex. (Normalapprox. von Summen von ZV:) Für praktische Anwendungen schreibt man die Aussage des ZGS meistens in der Form

$$\mathrm{P}[S_n^* \leq x] \approx \Phi(x) \quad \text{oder} \quad S_n^* \overset{\text{approx.}}{\sim} \mathcal{N}(0,1) \quad \text{für } n \text{ gross}$$

wobei das Symbol $\overset{\text{approx.}}{\sim}$ für "ist approximativ verteilt gemäss" steht. Für S_n oder $\overline{X}_n = \frac{1}{n}S_n$ zurückübersetzt heisst das dann

$$S_n \overset{\text{approx.}}{\sim} \mathcal{N}(n\mu, n\sigma^2) \quad \text{bzw.} \quad \overline{X}_n \overset{\text{approx.}}{\sim} \mathcal{N}(\mu, \frac{1}{n}\sigma^2),$$

und das wird benutzt, um Wahrscheinlichkeiten für diese Grössen (d.h. für S_n bzw. $\overline{X}_n$) approximativ als Wahrscheinlichkeiten für eine entsprechende Normalverteilung zu berechnen.

Ex. (Normalapprox. für die Binomialvert.) Ist n gross, so wird es schwierig die Binomialkoeffizienten bei der Dichte/Verteilungsfunktion der Binomialverteilung zu berechnen. Deshalb fasst man die Binomialverteilung (sofern es möglich ist) als Summe von i.i.d. $X_1, \dots, X_n$ sin $Be(p)$ Zufallsvariablen auf, d.h.

$$S_n = \sum_{i=1}^n X_i \sim Bin(n,p), \quad \mathrm{E}[X_i] = p, \quad \mathrm{Var}[X_i] = p(1-p).$$

Die Approximation ergibt sich dann mithilfe des ZGS für grosse n folgendermassen:

$$S_n \overset{\text{approx.}}{\sim} \mathcal{N}(n\,\mathrm{E}[X_i], n\,\mathrm{Var}[X_i]) = \mathcal{N}(np, np(1-p))$$

und natürlich ist die standardisierte Zufallsvariable folgendermassen verteilt:

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} = \frac{S_n - \mathrm{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \overset{\text{approx.}}{\sim} \mathcal{N}(0,1).$$

Eine etwas genauere Approximation für die Binomialverteilung erhält man, wenn man zusätzlich noch die sogenannte Kontinuitätskorrektur benutzt:

$$\begin{aligned} \mathrm{P}[a < S_n \le b] &= \mathrm{P}\left[\frac{a-np}{\sqrt{np(1-p)}} < S_n^* \le \frac{b-np}{\sqrt{np(1-p)}}\right] \\ &\approx \Phi\left(\frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Korrektur $+\frac{1}{2}$: Zentrieren der Stäbe im Histogramm macht Approximation genauer.

2.4 Grosse Abweichungen und Chernoff-Schranken

Def. (Momenterzeugende Funktion) einer Zufallsvariable X ist

$$M_X(t) = \mathrm{E}\left[e^{tX}\right] \quad \text{für } t \in \mathbb{R};$$

das ist immer wohldefiniert in $[0,\infty]$, kann aber $+\infty$ werden.

Thm. Seien $X_1, \dots, X_n$ i.i.d. Zufallsvariablen, für welche die momenterzeugende Funktion $M_X(t)$ für alle $t \in \mathbb{R}$ endlich ist. Für jedes $b \in \mathbb{R}$ gilt dann

$$\mathrm{P}[S_n \ge b] \le \exp\left(\inf_{t\in\mathbb{R}} (n \log M_X(t) - tb)\right).$$

Thm. Seien $X_1, \dots, X_n$ unabhängig mit $X_i \sim Be(p_i)$ und $S_n = \sum_{i=1}^n X_i$. Sei ferner $\mu_n := \mathrm{E}[S_n] = \sum_{i=1}^n p_i$ und $\delta > 0$. Dann gilt

$$\mathrm{P}[S_n \ge (1+\delta)\mu_n] \le \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^{\mu_n}.$$

Com. Sind $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} Be(p)$, so ist $S_n \sim Bin(n,p)$ und man kann es für die Abschätzung von Restwahrscheinlichkeiten von der Binomialverteilung anwenden.

3 Statistische Grundideen

Def. (Stichprobe, Stichprobenumfang) Die Gesamtheit der Beobachtungen $x_1, \dots, x_n$ oder Zufallsvariablen $X_1, \dots, X_n$ nennt man oft eine Stichprobe; die Anzahl n heisst dann der Stichprobenumfang.

4 Schätzer

Def. (Erwartungstreu) Ein Schätzer T heisst erwartungstreu für ϑ, falls gilt $\mathrm{E}_\vartheta[T] = \vartheta$; im Mittel (über alle denkbaren Realisationen ω) schätzt T also richtig. Allgemein heisst $\mathrm{E}_\vartheta[T] - \vartheta$ der Bias (oder erwartete Schätzfehler) von T; erwartungstreu (auf Englisch "unbiased") bedeutet also, dass der Bias Null ist.

Def. (Mean Squared Error, MSE) Der mittlere quadratische Schätzfehler ("mean squared error", MSE) ist definiert als

$$\mathrm{MSE}_\vartheta[T] := \mathrm{E}_\vartheta\left[(T-\vartheta)^2\right].$$

Def. (Konsistent) Eine Folge von Schätzern $T^{(n)}$, $n \in \mathbb{N}$ heisst konsistent für ϑ, falls $T^{(n)}$ für $n \to \infty$ in P_ϑ-Wahrscheinlichkeit gegen ϑ konvergiert, d.h. für jedes $\vartheta \in \Theta$ gilt

$$\lim_{n\to\infty} \mathrm{P}_\vartheta\left[\left|T^{(n)} - \vartheta\right| > \epsilon\right] = 0 \qquad \text{für jedes } \epsilon > 0.$$

(Das setzt anschaulich voraus, dass man beliebig viele Daten haben könnte.)

Com. Mithilfe der Rechenregeln für Erwartungswerte kann man den MSE zerlegen in

$$\begin{aligned} \mathrm{MSE}_\vartheta[T] &= \mathrm{E}_\vartheta\left[(T-\vartheta)^2\right] \\ &= \mathrm{E}_\vartheta\left[T^2\right] - 2\vartheta\,\mathrm{E}_\vartheta[T] + \vartheta^2 \\ &= \mathrm{E}_\vartheta\left[T^2\right] - (\mathrm{E}_\vartheta[T])^2 + (\mathrm{E}_\vartheta[T])^2 - 2\vartheta\,\mathrm{E}_\vartheta[T] + \vartheta^2 \\ &= \mathrm{Var}_\vartheta[T] + (\mathrm{E}_\vartheta[T] - \vartheta)^2, \end{aligned}$$

also in die Summe aus der Varianz des Schätzers T und dem Quadrat des Bias. Für erwartungstreue Schätzer sind Varianz und MSE also dasselbe (weil der zweite Summand wegfällt).

4.1 Maximum-Likelyhood-Methode

Def. (Likelyhood-Funktion)

$$L(x_1, \dots, x_n; \vartheta) := \begin{cases} p(x_1, \dots, x_n; \vartheta) & \text{im diskreten Fall,} \\ f(x_1, \dots, x_n; \vartheta) & \text{im stetigen Fall.} \end{cases}$$

$\log L(x_1, \dots, x_n; \vartheta)$ heisst log-Likelihood-Funktion.

Def. (Maximum-Likelyhood-Schätzer) Der ML-Schätzer T für ϑ wird dadurch definiert, dass er

$$\vartheta \mapsto L(x_1, \dots, x_n; \vartheta) \qquad \text{als Funktion von } \vartheta$$

maximiert.

Meistens sind $X_1, \dots, X_n$ i.i.d. unter P_ϑ; die Likelihood-Funktion L ist dann ein Produkt, und es ist bequemer, statt L die log-Likelihood-Funktion $\log L$ zu maximieren, weil diese eine Summe ist. Statt zu maximieren sucht man ferner meistens nur Nullstellen der (partiellen) Ableitung(en) nach den Komponenten von ϑ.

Def. (Momentenschätzer) Der Schätzer $T = (T_1, T_2)$, wobei $T_1 = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}_n$, $T_2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - \left(\overline{X}_n\right)^2$ ist der sogenannte Momentenschätzer für $(\mathrm{E}_\vartheta[X], \mathrm{Var}_\vartheta[X])$ in jedem Modell P_ϑ, wo $X_1, \dots, X_n$ i.i.d. sind.

Cor. (Momentenschätzer nicht erwartungstreu) Der Momentenschätzer hat aber den allgemeinen Nachteil, dass er nicht erwartungstreu für $(\mathrm{E}_\vartheta[X], \mathrm{Var}_\vartheta[X])$ ist. Zwar ist für jedes ϑ $\mathrm{E}_\vartheta[T_1] = \mathrm{E}_\vartheta[X]$; aber $\mathrm{E}_\vartheta[T_2] = \frac{n-1}{n}\mathrm{E}_\vartheta[X]$. Um einen erwartungstreuen Schätzer T' für $(\mathrm{E}_\vartheta[X], \mathrm{Var}_\vartheta[X])$ zu haben benutzt man deshalb meistens $T_1' = T_1 = \overline{X}_n$,
$T_2' = \frac{n}{n-1}T_2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2\right) - \frac{n}{n-1}\left(\overline{X}_n\right)^2$.

Def. (Empirische Stichprobenvarianz) Für T_2' benutzt man oft auch die Notation

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

und nennt S^2 die empirische Stichprobenvarianz.

4.2 Verteilungsaussagen

In vielen Situationen ist es nützlich oder nötig die Verteilung (unter P_ϑ, für jedes $\vartheta \in \Omega$) eines Schätzers zu kennen. Exakte allgemeine Aussagen gibt es dazu nur wenige. Einen allgemeine approximativen Zugang liefert der Zentrale Grenzwertsatz (siehe vorher). Für normalverteilte Stichproben hat man aber folgenden Satz:

Thm. Seien $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Dann gilt:

1) $\overline{X}_n$ ist normalverteilt gemäss $\sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$, und $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$. (Richtige Var. im Nenner nehmen!)
2) $\frac{n-1}{\sigma^2}S^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ is χ^2-verteilt mit $n-1$ Freiheitsgraden.
3) $\overline{X}_n$ und S^2 sind unabhängig.
4) Der Quotient

$$\frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{S/\sigma} = \frac{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1}\frac{n-1}{\sigma^2}S^2}}$$

ist t-verteilt mit $n-1$ Freiheitsgraden.

5 Tests

Grundidee: Wir haben schon eine Vermutung, wo in Θ der richtige (aber unbekannte) Parameter ϑ liegen könnte, und wollen diese mit Hilfe der Daten überprüfen ("testen").

Def. (Test, Teststatistik) Die Zufallsvariable $T = t(X_1, \dots, X_n)$ heisst Teststatistik. Die Entscheidungsregel bzw. der Test ist $I_{\{t(x_1,\dots,x_n) \in K\}}$. D.h. verwerfe H_0 falls $I(\omega) = 1$.

Grundsätzlich versucht man Tests immer so zu konstruieren, dass die folgenden Wahrscheinlichkeiten möglichst klein werden. Eine Möglichkeit dafür ist das folgende **asymmetrische Vorgehen**: Wähle ein Signifikanzniveau α und versuche in zwei Schritten die folgenden Fehlerwahrscheinlichkeiten möglichst klein zu machen

a) **Fehler 1. Art:** H_0 wird zu unrecht abgelehnt (nicht akzeptiert). Um diese Fehlerwhs. zu minimieren muss man schauen, dass die folgende Whs. möglichst klein (kleiner als das vorgegebene α) wird:

$$\sup \mathrm{P}_{\vartheta \in \Theta_0}[T \in K] < \alpha \quad \text{(Signifikanzniveau)}$$

b) **Fehler 2. Art:** H_0 wird zu unrecht akzeptiert (nicht verworfen). Um diese Fehlerwhs. zu minimieren muss man schen, dass die folgende Whs möglichst klein, bzw. die Macht des Tests $\beta(\vartheta)$ möglichst gross wird:

$$\inf \mathrm{P}_{\vartheta \in \Theta_A}[T \notin K] = 1 - \sup \underbrace{\mathrm{P}_{\vartheta \in \Theta_A}[T \in K]}_{=\beta(\vartheta) \quad (\vartheta \in \Theta_A)}$$

Dieses asymmetrische Vorgehen impliziert folgendes:

- Es ist schwieriger H_0 zu verwerfen. Deshalb wählt man in der Praxis für H_0 die Negation der Aussage (gelingt es trotzdem H_0 zu verwerfen kann man zuversichtlicher sein).
- Es kann durchaus sein, dass man durchs Vertauschen von H_0 und H_A unterschiedliche Testergebnisse erhält.

Com. Es ist nicht immer möglich die Verteilung von T unter jedem $\vartheta \in \Theta_A$ für die Macht des Tests zu bestimmen. Dann gibt man sich damit zufrieden, dass man nur die Verteilung von T unter $\vartheta \in \Theta_A$ zum Einhalten des Signifikanzniveaus bestimmt, um das Signifikanzniveau einzuhalten.

Def. (P-Wert) Der Wert wird bestimmt durch die gezogene Stichprobe. Er deutet an, wie wahrscheinlich es ist, ein solches Stichprobenergebnis oder ein extremeres zu erhalten, wenn die Nullhypothese wahr ist. Ein häufiges Missverständnis ist die Gleichsetzung dieser Aussage mit der falschen Behauptung, der p-Wert würde angeben, wie wahrscheinlich die Nullhypothese bei Erhalt dieses Stichprobenergebnisses ist. Mit dem p-Wert wird also angedeutet, wie extrem das Ergebnis ist: je kleiner der p-Wert, desto mehr spricht das Ergebnis gegen die Nullhypothese.

	Links	Beidseitig	Rechts
p-Wert=	$\mathrm{P}_{\vartheta_0}[X \le x]$	$\begin{cases} \mathrm{P}_{\vartheta_0}[x \ge X], & T(\omega) > \mu_0 \\ \mathrm{P}_{\vartheta_0}[X \ge x], & T(\omega) < \mu_0 \end{cases}$	$\mathrm{P}_{\vartheta_0}[x \le X]$

Falls wir ein Signifikanzlevel α a priori gegeben haben, dann gilt also

- verwerfe H_0, falls p-Wert $\le \alpha$, und
- akzeptiere H_0, falls p-Wert $> \alpha$.

5.1 Konstruktion von Tests

Def. (Likelihood Quotient) Sei $L(x_1, \ldots, x_n; \vartheta)$ die Likelihood-Funktion; für diskrete X_i ist das also im Modell P_ϑ die Wahrscheinlichkeit $\mathrm{P}_\vartheta[X_1 = x_1, \ldots, X_n = x_n]$, die beobachteten Werte $x_1, \ldots, x_n$ zu erhalten. Für $\vartheta_0 \in \Theta_0$ und $\vartheta_A \in \Theta_A$ betrachten wir den Likelihood-Quotienten

$$R(x_1, \ldots, x_n; \vartheta_0, \vartheta_A) := \frac{L(x_1, \ldots, x_n; \vartheta_0)}{L(x_1, \ldots, x_n; \vartheta_A)}.$$

Ist dieser Quotient klein, so ist der Nenner wesentlich grösser als der Zähler. Das bedeutet also, dass die Beobachtungen $x_1, \ldots, x_n$ als Resultate im Modell P_{ϑ_A} deutlich wahrscheinlicher sind als im Modell P_{ϑ_0}; die Daten sprechen also gegen ϑ_0 im Vergleich zu ϑ_A.
Es liegt deshalb nahe, als Teststatistik $T := R(X_1, \ldots, X_n; \vartheta_0, \vartheta_A)$ und als kritischen Bereich $K := [0, c)$ zu wählen, wenn man ϑ_0 gegen ϑ_A testen will; man verwirft dann also die Hypothese H_0, wenn der Quotient R zu klein wird. Wegen $R \geq 0$ genügt $[0, c)$ statt $(-\infty, c)$.

Der letzte Schritt den wir allgemein noch machen müssen ist, um den Kritischen Bereich K passend zum gewünschten Niveau α festlegen zu können, brauchen wir die Verteilung der Teststatistik T unter der Hypothese H_0, d.h. in jedem Modell P_ϑ mit $\vartheta \in \Theta_0$. Evtl.

Sind Hypothese und Alternative beide einfach (was leider eher selten der Fall ist), so ist dieser Test optimal, wie das folgende Resultat zeigt.

Thm. (Neyman Pearson-Lemma) Sei $\Theta_0 = \{\vartheta_0\}$ und $\Theta_A = \{\vartheta_A\}$. Wie oben sei $T := R(x_1, \ldots, x_n; \vartheta_0, \vartheta_A)$ und $K := [0, c)$, sowie $\alpha^* = \mathrm{P}_{\vartheta_0}[T \in K] = \mathrm{P}_{\vartheta_0}[T \leq c]$. Der Likelihood-Quotienten-Test mit Teststatistik T und kritischem Bereich K ist dann im folgenden Sinn optimal: Jeder andere Test mit Signifikanzniveau $\alpha \leq \alpha^*$ hat kleinere Macht bzw. eine grössere Wahrscheinlichkeit für einen Fehler 2. Art.
(Etwas formaler: Ist (T', K') ein anderer Test mit $\mathrm{P}_{\vartheta_0}[T' \in K'] \leq \alpha^*$, so gilt auch $\mathrm{P}_{\vartheta_A}[T' \in K'] \leq \mathrm{P}_{\vartheta_A}[T \in K]$.)

Def. (Verallgemeinerter Likelihood-Quotient)

$$R(x_1, \ldots, x_n) := \frac{\sup_{\vartheta \in \Theta_0} L(x_1, \ldots, x_n; \vartheta)}{\sup_{\vartheta \in \Theta_A} L(x_1, \ldots, x_n; \vartheta)}$$

oder auch

$$\tilde{R}(x_1, \ldots, x_n) := \frac{\sup_{\vartheta \in \Theta_0} L(x_1, \ldots, x_n; \vartheta)}{\sup_{\vartheta \in \Theta_A \cup \Theta_0} L(x_1, \ldots, x_n; \vartheta)}$$

Die Teststatistik ist dann $T_0 := R(X_1, \ldots, X_n)$ bzw. $\tilde{T} := \tilde{R}(X_1, \ldots, X_n)$ mit krit. Ber. $K_0 := [0, c_0)$.

6 Konfidenzbereiche

Grundidee: Wie in vorherigen Abschnitt suchen wir aus einer Familie $(P_\vartheta)_{\vartheta \in \Theta}$ von Modellen eines, das zu unseren Daten $x_1, \ldots, x_n$ passt. Ein Schätzer für ϑ gibt uns dabe einen einzelnen zufälligen möglichen Parameterwert. Weil es schwierig ist, mit diesem einen Wert den richtigen (aber unbekannten) Paremeter zu treffen, suchen wir nun stattdessen eine (zufällige) Teilmenge des Parameterbereichs, die hoffentlich den wahren Parameter enthält.

Def. (Konfidenzbereich) Etwas formaler ist ein Konfidenzbereich für ϑ zu Daten $x_1, \ldots, x_n$ eine Menge $C(x_1, \ldots, x_n) \subseteq \Theta$; in den meisten Fällen ist das ein Intervall, dessen Endpunkte von $x_1, \ldots, x_n$ abhängen. Ersetzen wir die Daten $x_1, \ldots, x_n$ durch die sie generierenden Zufallsvariablen, so ist $C(X_1, \ldots, X_n)$ also eine zufällige Teilmenge von Θ, mit Realisierung $C(\omega) = C(X_1(\omega), \ldots, X_n(\omega))$ bei einem festen ω. Ein solches C heisst ein Konfidenzbereich zum Niveau $1 - \alpha$, falls gilt

$$\mathrm{P}_\vartheta[\vartheta \in C(X_1, \ldots, X_n)] \geq 1 - \alpha \qquad \text{für alle } \vartheta \in \Theta$$

d.h. in jedem Modell erwischt man den Parameter mit grosser Wahrscheinlichkeit.

Seien die Daten die Realisationen von Zufallsvariablen $X_1, \ldots, X_n$, die alle unter P_ϑ i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$ sind. Die offensichtlichen Schätzer für (μ, σ^2) sind $T = (\overline{X}_n, S^2)$. Das Konfidenzintervall wird um diese Schätzer herum angesetzt.

a) μ gesucht, σ^2 bekannt: Wir wissen, dass in diesem Fall $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ unter P_ϑ. Also ist das Konfidenzintervall für μ zum Niveau $1 - \alpha$

$$C(X_1, \ldots, X_n) = \left[\overline{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right].$$

b) μ gesucht, σ^2 gesucht: Wir wissen, dass in diesem Fall $\frac{\overline{X}_n - \mu}{S/\sqrt{n}} \sim t_{n-1}$ unter P_ϑ. Also ist das Konfidenzintervall für μ zum Niveau $1 - \alpha$

$$C(X_1, \ldots, X_n) = \left[\overline{X}_n - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X}_n + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right].$$

Weil $\frac{1}{\sigma^2}(n-1)S^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \sim \chi^2_{n-1}$ unter P_ϑ ist, ist das Konfidenzinterval für σ^2 zum Nieveau $1 - \alpha$

$$C(X_1, \ldots, X_n) = \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right].$$

c) μ bekannt, σ^2 gesucht: Wir bestimmen das Konfidenzinterval für σ^2 zum Niveau $1 - \alpha$ genau wie bei b).

Wichtig: Man bemerke, dass das Konfidenzintervall in allen Fällen schmäler wird, sobald der Stichprobenumfang n grösser wird!

Com. Können wir keine genauen Verteilungsaussagen über die Zufallsvariablen $X_1, \ldots, X_n$ machen, so ist es schwieriger exakte Konfidenzintervalle zu erhalten. In allgemeinen Situationen kann man oft nur approximative Konfidenzintervalle mit Hilfe des Zentralen Grenzwertsatzes bekommen (siehe vorher).

Vom Konfidenzbereich zur Hypothese:

$$\begin{array}{c} C(X_1, \ldots, X_n) \text{ ist} \\ \text{Konfidenzbereich für} \\ \vartheta \text{ zum Niveau } 1-\alpha \end{array} \Longrightarrow \begin{array}{c} I_{\{\vartheta_o \notin C(X_1,\ldots,X_n)\}} \text{ ist} \\ \text{Test für } \vartheta = \vartheta_0 \\ \text{zum Niveau } \alpha \end{array}$$

Von der Hypothese zum Konfidenzbereich:

$$\begin{array}{c} I_{\{t(x_1,\ldots,x_n) \in K_{\vartheta_0}\}} \\ \text{Test für } \vartheta = \vartheta_0 \\ \text{zum Niveau } \alpha \end{array} \Longrightarrow \begin{array}{c} C(X_1, \ldots, X_n) := K_\vartheta^c \\ \text{ist Konfidenzbereich für} \\ \vartheta \text{ zum Niveau } 1-\alpha \end{array}$$

7 Sonstiges

Drawing Elements from a Set: The table shows the umber of possibilities for selecting k elements from a set of size n (the example set is $\{a, b, c\}$ thus $n = 3$; and $k = 2$):

	Ordered		Unordered	
	Number	Examples	Number	Examples
w. rep.	n^k (Variat.)	aa ab ac ba bb bc ca cb cc	$\binom{n+k-1}{k}$	aa ab ac bb bc cc
w.o. rep.	$n^{\underline{k}}$	ab ac ba bc ca cb	$\binom{n}{k}$ (Kombin.)	ab ac bc

Identitäten $\binom{n}{k} = \binom{n}{n-k}$, $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$, $(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$, $\sum_{k=0}^n \binom{n}{k} = 2^n$, $\sum_{k=0}^n (-1)^k \binom{n}{k} = 0$,

Geometrische Reihe

$$S_n = a_0 \sum_{k=0}^n q^k = a_0 \frac{1-q^{n+1}}{1-q} = a_0 \frac{q^{n+1}-1}{q-1}$$

$$S = a_0 \sum_{k=0}^\infty = \frac{1}{1-q} \qquad \text{falls } |q| < 1$$

Wichtig: Auf die Indizes achten! Sie beginnt bei $k = 0$ somit ist das erste Glied 1. Falls bei $k = 1$ begonnen wird muss man -1 addieren!

Arithmetische Reihe

$$S_n = \sum_{k=0}^n (k \cdot d + a_0) = (a_0 + a_n) \cdot \frac{(n+1)}{2}$$

wobei

$$a_i = i \cdot d + a_0 \quad \forall i \geq 1$$

und wobei d die Differenz zwischen zwei Folgegliedern ist.

Ex. (Summe der Zahlen von 0 bis n)

$$\sum_{k=0}^n k = n \cdot \frac{(n+1)}{2}$$

Ex. (Summe der ersten n ungeraden Zahlen)

$$\sum_{k=1}^n (2k-1) = \sum_{k=0}^{n-1} (2k+1) = 1 + 3 + 5 + 7 + \cdots + (2n-1) = n^2$$

Summe von Quadratzahlen
Die Summe der ersten n Quadratzahlen ist für alle $n \geq 1$:

$$\sum_{k=1}^n k^2 = \frac{n \cdot (n+1) \cdot (2n+1)}{6}.$$

Summe von spezieller Potenzreihe
Die Summe für die folgende Potenzreihe für ein q mit $|q| < 1$ ist:

$$\sum_{k=1}^\infty k \cdot q^{k-1} = \sum_{k=1}^\infty \frac{d}{dq}(q^k) = \frac{d}{dq}\left(\sum_{k=1}^\infty q^k\right) = \frac{d}{dq}\left(\frac{1}{1-q} - 1\right) = \frac{1}{(1-q)^2}.$$

Com. Oft substituiert man $q := (1-p)$.

Reihenentwicklung von e^x
$e^x = \sum_{n=0}^\infty \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$,

Partielle Integration
$\int_a^b f'(x) g(x)\ dx = [f(x)g(x)]_a^b - \int_a^b f(x) g'(x)\ dx$

Substitution
Ist f stetig und g ein Diffeomorphismus (in beiden Richtungen stetig differenzierbare Abbildung), dann gilt
$\int_a^b f(g(x)) g'(x)\ dx = \int_{g(a)}^{g(b)} f(y)\ dy$
$y = g(x) \iff x = g^{-1}(y)$
$\frac{dy}{dx} = g'(x) \iff dy = g'(x) dx$

$a + b = \max(a, b) + \min(a, b)$

Beim Berechnen der Varianz verwendet man oft den Trick dass $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \mathrm{E}[X(X-1)] + \mathrm{E}[X] - (\mathrm{E}[X])^2$, weil der Erwartungswert $\mathrm{E}[X(X-1)]$ in manchen Fällen einfacher als $\mathrm{E}[X^2]$ zu berechnen ist.

Diskrete Verteilung μ_X	Beschrieb von X	$\mathcal{W}(X) =$	Gewichtsfunktion $p_X(k) = \mathrm{P}[X = k] =$	Verteilungsfunktion $F_X(t) = \mathrm{P}[X \leq t] =$	$\mathrm{E}[X] =$	$\mathrm{Var}[X] =$
Gleichverteilung $X \sim \mathcal{U}_T$	Alle n Elementarereignisse der Trägermenge T sind gleich wahrscheinlich.	$\{x_1, x_2, \ldots, x_n\}$	$\begin{cases} \frac{1}{n} & x \in \mathcal{W}(X), \\ 0, & x \notin \mathcal{W}(X). \end{cases}$	$\dfrac{\lvert\{x_k \in \mathcal{W}(X) \mid x_k \leq t\}\rvert}{n}$	$\dfrac{1}{n}\sum_{i=1}^{n} x_i$	$\frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - (x_n)^2\right)$
Bernoulli $X \sim Be(p)$	Anzahl der Erfolge X bei einem einzelnen 0-1-Experiment A_1 mit Erfolgsparameter p $X = Y_1 = I_{A_1}$ (Bem.: $X^2 = X$)	$\{0,1\}$ 1=Erfolg, 0=Misserfolg.	$p^k(1-p)^{1-k}$	$\begin{cases} 0 & t < 0 \\ 1-p & 0 \leq t < 1 \\ 1 & t \geq 1 \end{cases}$	p	$p(1-p)$
Binomial $X \sim Bin(n,p)$ Rekursionsformel: $\mathrm{P}[X = k+1] =$ $\frac{p}{1-p}\frac{n-k}{k+1}\mathrm{P}[X = k]$	Anzahl der Erfolge X bei n unabhängigen 0-1-Experimenten mit Erfolgsparameter p. Sind $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} Be(p)$, so ist $X = \sum_{i=1}^{n} I_{A_i} = \sum_{i=1}^{n} Y_i \sim Bin(n,p).$ Approx.: $X \overset{\text{approx.}}{\sim} \mathcal{P}(np)$ gut falls $np^2 \leq 0.05$.	$\{0, 1, \ldots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$\sum_{k=0}^{\lfloor t \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Geometrisch $X \sim Geom(p)$ "gedächtnislos"	Wartezeit X auf den ersten Erfolg bei einer unendlichen Folge von 0-1-Experimenten mit Erfolgsparameter p. Spezialfall von $NB(1,p)$ $X = \inf\{i \in \mathbb{N} \mid A_i \text{ tritt ein}\}$ $= \inf\{n \in \mathbb{N} \mid Y_i = 1\}$	$\{1, 2, 3, \ldots\}$	$(1-p)^{k-1}p$	$1 - (1-p)^{\lfloor t \rfloor}$ $\mathrm{P}[X > t] = (1-p)^{\lfloor t \rfloor}$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
Negativbinomial $X \sim NB(r,p)$	Wartezeit X auf den r-ten Erfolg bei einer unendlichen Folge von 0-1-Experimenten mit Erfolgsparameter p. $X = \inf\left\{k \in \mathbb{N} \,\middle\vert\, \sum_{i=1}^{k} I_{A_i} = r\right\}$ $= \inf\left\{k \in \mathbb{N} \,\middle\vert\, \sum_{i=1}^{k} Y_i = r\right\}$ Sind die ZV $X_1, \ldots, X_r \overset{\text{i.i.d.}}{\sim} Geom(p)$, so ist $X := X_1 + \cdots + X_r \sim NB(r,p)$.	$\{r, r+1, r+2, \ldots\}$	$\binom{k-1}{r-1} p^r (1-p)^{k-r}$	$\sum_{k=r}^{\lfloor t \rfloor} \binom{k-1}{r-1} p^r (1-p)^{k-r}$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$
Hypergeometrisch $X \sim HypG(n,r,m)$ (r wie "richtig")	In einer Urne seien n Gegenstände, davon r vom Typ 1 und $n-r$ vom Typ 2. Man zieht ohne Zurücklegen m der Gegenstände; die Zufallsvariable X beschreibt die Anzahl der Gegenstände vom Typ 1 in dieser Stichprobe vom Umfang m. $T = T_1 \cup T_2$, $\lvert T\rvert = n$, $\lvert T_1\rvert = r$, $\lvert T_2\rvert = n-r$ $S \subseteq T$, $\lvert S\rvert = m$ $X = \lvert\{t \in S \mid t \in T_1\}\rvert$	$\{0, 1, \ldots, \min(m,r)\}$	$\dfrac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$	$\sum_{k=0}^{\lfloor t \rfloor} \dfrac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$	$\dfrac{mr}{n}$	$m\frac{r}{n}\left(1-\frac{r}{n}\right)\frac{n-m}{n-1}$
Poisson $X \sim \mathcal{P}(\lambda)$	Anzahl Ereignisse X in einem festen Zeitintervall. Der Param. $\lambda > 0$ ist dabei die durchschnittl., bzw. die erwartete Anz. Ereignisse in dem Intervall.	$\{0, 1, 2, \ldots\}$	$e^{-\lambda}\dfrac{\lambda^k}{k!}$	$e^{-\lambda}\sum_{k=0}^{\lfloor t \rfloor} \dfrac{\lambda^k}{k!}$	λ	λ

Stetige Verteilung μ_X	**Beschrieb von** X	$\mathcal{W}(X) =$	**Dichtefunktion** $f_X(t) = (\neq \mathrm{P}[X = t])$	**Verteilungsfunktion** $F_X(t) = \mathrm{P}[X \leq t] =$	$\mathrm{E}[X]=$	$\mathrm{Var}[X] =$
Gleichverteilung $X \sim \mathcal{U}(a,b)$	Zufällige Wahl eines Punktes in $[a,b]$.	$[a,b]$	$\begin{cases} \frac{1}{b-a}, & \text{für } a \leq t \leq b, \\ 0, & \text{sonst.} \end{cases}$	$\begin{cases} 0, & \text{für } t < a, \\ \frac{t-a}{b-a}, & \text{für } a \leq t \leq b, \\ 1, & \text{für } t > b. \end{cases}$	$\frac{a+b}{2}$	$\frac{1}{12}(b-a)^2$
Exponential $X \sim Exp(\lambda)$ "gedächtnislos" $\mathrm{P}[X > t+s \mid X > s] = \mathrm{P}[X > t]$	Analogon der geometrischen Verteilung. Modell für Wartezeiten oder Lebensdauern. Wartezeit, bis zum eintreffen eines Ereignisses. $\lambda > 0$ steht für die erwartete Anzahl Ereignisse pro Einheitsintervall.	$[0,\infty)$	$\begin{cases} \lambda e^{-\lambda t}, & \text{für } t \geq 0, \\ 0, & \text{für } t < 0. \end{cases}$	$\begin{cases} 1 - e^{-\lambda t}, & \text{für } t \geq 0, \\ 0, & \text{für } t < 0. \end{cases}$	$\frac{1}{\lambda}$ (MTBF)	$\frac{1}{\lambda^2}$
Standard-Normal $X \sim \mathcal{N}(0,1)$	Glockenkurve mit Erwartungswert $\mu = 0$ und Varianz $\sigma^2 = 1$.	$\mathbb{R}$	$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$	$\Phi(t) = \int_{-\infty}^{t} \varphi(s)\, ds$	0	1
Normalverteilung, NV, oder Gauss-Verteilung mit Parameter $\mu \in \mathbb{R}$, $\sigma^2 > 0$ $X \sim \mathcal{N}(\mu, \sigma^2)$	Ist $Y \sim \mathcal{N}(\mu,\sigma^2)$, so ist $X := \frac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1)$	$\mathbb{R}$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma}\varphi(\frac{t-\mu}{\sigma})$	$\mathrm{P}\left[\frac{X-\mu}{\sigma} \leq \frac{t-\mu}{\sigma}\right] = \Phi(\frac{t-\mu}{\sigma})$	μ	σ^2
Cauchy $X \sim Cauchy(s,t)$	Breitenparameter $s > 0$ und Lageparameter t, $\lvert t\rvert < \infty$	$\mathbb{R}$	$\frac{1}{\pi} \cdot \frac{s}{s^2 + (x-t)^2}$	$\frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-t}{s}\right)$	existiert nicht (nicht abs. integ.)	existiert nicht
Pareto $X \sim Par(\alpha, x_0)$	Whs., dass Schadengrösse X grösser als $x_0 > 0$ ist für Param. $\alpha > 0$ ("Excess-of-Loss"). $X \sim \mathcal{U}(0,1) \implies \frac{1}{X} \sim Par(1,1)$.	$[x_0,\infty)$	$\begin{cases} \frac{\alpha x_0^\alpha}{t^{\alpha+1}} & t \geq x_0 \\ 0 & t < x_0 \end{cases}$	$1 - (\frac{x_0}{t})^\alpha$	$\begin{cases} x_0 \frac{\alpha}{\alpha-1} & \alpha > 1, \\ \infty & \alpha \leq 1. \end{cases}$	$\begin{cases} x_0^2 \frac{\alpha}{(\alpha-2)(\alpha-1)^2} & \alpha > 2, \\ \infty & 1 < \alpha \leq 2. \end{cases}$
Chi-Quadrat-Verteilung mit n Freiheitsgraden $X \sim \chi_n^2$	Sind $Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, so ist $X := \sum_{i=1}^{n} Y_i^2 \sim \chi_n^2$.	$[0,\infty)$	Für $n = 2$ ergibt das $Exp(\frac{1}{2})$. Spezialfall einer $Ga(\alpha,\lambda)$-Verteilung mit $\alpha = \frac{1}{2}$ und $\lambda = \frac{1}{2}$.		n	$2n$
Student-t-Verteilung mit n Freiheitsgraden $X \sim t_n$	Sind Y und Z unabhängig mit $Y \sim \mathcal{N}(0,1)$ und $Z \sim \chi_n^2$, so ist $X := \frac{Y}{\sqrt{\frac{1}{n}Z}} \sim t_n$.	$\mathbb{R}$	Für $n = 1$ ist das eine Cauchy-Verteilung, und für $n \to \infty$ erhält man eine $\mathcal{N}(0,1)$-Verteilung. Ist wie die Normalverteilung glockenförmig und symmetrisch um 0, aber langschwänziger, und zwar umso mehr, je kleiner n ist.		0 für $n > 1$	$\frac{n}{n-2}$ für $n > 2$

8 Standard-Tests (für (approximativ) normalverteilte ZV)

8.1 Einstichproben-Tests

	z-Test	t-Test
Beschrieb:	Normalverteilung, Test für Erwartungswert bei bekannter Varianz	Normalverteilung, Test für Erwartungswert bei unbekannter Varianz
Stichprobe:	$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\vartheta, \sigma^2)$ unter P_ϑ	$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ unter $P_{\vec{\vartheta}}$
Bekannt:	σ^2 (Varianz)	(nichts)
Unbekannt ϑ:	$\vartheta = \mu$	$\vec{\vartheta} = (\mu, \sigma^2)$
Hypothese H_0:	$\vartheta = \vartheta_0$ (einfach) $\Theta_0 = \{\vartheta_0\}$	$\mu = \mu_0$ (zusammengesetzt) $\Theta_0 = \{\mu_0\} \times (0, \infty) = \left\{ \vec{\vartheta} = (\mu, \sigma^2) \,\middle\vert\, \mu = \mu_0 \right\}$
Alternative H_A:	1) $\vartheta > \vartheta_0$ 2) $\vartheta < \vartheta_0$ 3) $\vartheta \neq \vartheta_0$	1) $\mu > \mu_0$ 2) $\mu < \mu$ 3) $\mu \neq \mu_0$
Teststatistik T:	$T = \dfrac{\overline{X}_n - \vartheta_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ unter P_{ϑ_0}	$T = \dfrac{\overline{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$ unter $P_{\vec{\vartheta}_0}$ $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$
Krit. Bereich K:	1) $[c_>, \infty)$ 2) $(-\infty, c_<]$	3) $(-\infty, -c_{\neq}] \cup [c_{\neq}, \infty)$
Krit. Wert c:	1) $c_> = \Phi^{-1}(1-\alpha) = z_{1-\alpha}$ 2) $c_< = -\Phi^{-1}(1-\alpha) = -z_{1-\alpha}$ 3) $c_{\neq} = \Phi^{-1}(1-\frac{\alpha}{2}) = z_{1-\frac{\alpha}{2}}$ ($z_{1-\alpha} = (1-\alpha)$-Quantil der $\mathcal{N}(0,1)$-Vert.)	1) $c_> = t_{n-1,1-\alpha}$ 2) $c_< = t_{n-1,\alpha} = -t_{n-1,1-\alpha}$ 3) $c_{\neq} = t_{n-1,1-\frac{\alpha}{2}}$ ($t_{m,\gamma} = \gamma$-Quantil der t_m-Vert.)
Verwerfe H_0 falls:	1) $\overline{X}_n > c_> \dfrac{\sigma}{\sqrt{n}} + \vartheta_0, \quad T > z_{1-\alpha}$ 2) $\overline{X}_n < c_< \dfrac{\sigma}{\sqrt{n}} + \vartheta_0, \quad T < -z_{1-\alpha}$ 3) $\left\vert \dfrac{\overline{X}_n - \vartheta_0}{\sigma/\sqrt{n}} \right\vert > c_{\neq}, \quad \vert T \vert > z_{1-\frac{\alpha}{2}}$	1) $\overline{X}_n > c_> \dfrac{S}{\sqrt{n}} + \vartheta_0, \quad T > t_{n-1} 1 - \alpha$ 2) $\overline{X}_n < c_< \dfrac{S}{\sqrt{n}} + \vartheta_0, \quad T < -t_{n-1} 1 - \alpha$ 3) $\left\vert \dfrac{\overline{X}_n - \vartheta_0}{S/\sqrt{n}} \right\vert > c_{\neq}, \quad \vert T \vert > t_{n-1,1-\frac{\alpha}{2}}$

unabh. $\overset{\not\Leftarrow}{\Longrightarrow}$ paarw. unabh. $\Longrightarrow$ paarw. unkorr.
Unabhängigkeit von ZV bleibt bei Transf. erhalten!

$Y = g(X) \Longrightarrow \mathrm{E}[Y] = \mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$
(Absolute Konvergenz nötig)
$Y = \sum_{\ell=1}^{m} X_\ell \Longrightarrow \mathrm{E}[Y] = a + \sum_{\ell=1}^{m} b_\ell \, \mathrm{E}[X_\ell]$.
$\forall \omega \in \Omega : X(\omega) \leq Y(\omega) \Longrightarrow \mathrm{E}[X] \leq \mathrm{E}[Y]$
$\mathcal{W}(X) = \mathbb{N}_0 \Rightarrow \mathrm{E}[X] = \sum_{j=1}^{\infty} \mathrm{P}[X \geq j] = \sum_{\ell=0}^{\infty} P[X > \ell]$.
$X_i, \ldots, X_n$ unabh. $\Rightarrow \mathrm{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathrm{E}[X_i]$

$\mathrm{E}[\mathrm{E}[X]] = \mathrm{E}[X]$
$\mathrm{E}[\mathrm{E}[X]\,\mathrm{E}[X]] = (\mathrm{E}[X])^2$
$\mathrm{E}[\mathrm{E}[X]\, X] = \mathrm{E}[X]\,\mathrm{E}[X] = (\mathrm{E}[X])^2$
Com. EW ist innerhalb von EW eine Konstante.

$\mathrm{E}[X^2] < \infty \Rightarrow \mathrm{Var}[X] := \mathrm{E}[(X - \mathrm{E}[X])^2]$ (definiert)
(mit $g(X) = (X - \mathrm{E}[X])^2$, mittlere q. Abw. von EW)
$\mathrm{E}[X^2] < \infty \Rightarrow\Rightarrow \mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$
$\mathrm{E}[X^2] < \infty \Rightarrow\Rightarrow \mathrm{Var}[aX + b] = a^2 \, \mathrm{Var}[X]$
$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j)$
$\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + 2\,\mathrm{Cov}(X_1, X_2)$.
X_i, X_j p.w. unkorr. $\Rightarrow \mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$

$\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$ oder
$\mathrm{Cov}(X, Y) = \mathrm{E}[(X_1 - \mathrm{E}[X_1])(X_2 - \mathrm{E}[X_2])]$
X, Y unabh. $\Longrightarrow \mathrm{Cov}(X, Y) = 0$
X, Y unabh. $\Longrightarrow \mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$
$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
$\mathrm{Cov}(X, X) = \mathrm{Var}[X]$,
$\mathrm{Cov}(X, Y + Z) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z)$
$\mathrm{Cov}(X, a) = 0$ für $a \in \mathbb{R}$
$\mathrm{Cov}(c + aX, d + bY) = ab\,\mathrm{Cov}(X, Y)$

8.2 Zweistichproben-Tests

Gepaart $(m = n)$	**Ungepaart** $(m \neq n \vee m = n)$
Normalverteilung, Test für Erwartungswert bei unbekannter Varianz	Normalverteilung, Test für Erwartungswert bei unbekannter Varianz

$$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X, \sigma^2) \text{ unter } P_\vartheta \text{ (oder } P_{\vec{\vartheta}})$$

$$Y_1, \dots, Y_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_Y, \sigma^2) \text{ unter } P_\vartheta \text{ (oder } P_{\vec{\vartheta}})$$

$$X_1, \dots, X_n,\ Y_1, \dots, Y_m \text{ sind unabh. unter } P_\vartheta \text{ (oder } P_{\vec{\vartheta}})$$

Varianz σ^2 ist in beiden Stichproben dieselbe

Gepaart:

Man bildet paarweise Differenzen:

$$Z_i := X_i - Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X - \mu_Y, 2\sigma^2)$$

a) σ^2 bekannt: z**-Test**

$$T = \frac{\overbrace{(\overline{X}_n - \overline{Y}_n)}^{\overline{Z}_n} - \overbrace{(\mu_X - \mu_Y)}^{\mu_{\vartheta_0}}}{\sigma\sqrt{\frac{2}{n}}} \sim \mathcal{N}(0,1)$$

unter jedem P_ϑ.

b) σ^2 unbekannt: t**-Test**

$$T = \frac{(\overline{X}_n - \overline{Y}_n) - (\mu_X - \mu_Y)}{S\sqrt{\frac{2}{n}}} \sim t_n$$

unter jedem $P_{\vec{\vartheta}}$, wobei

$$S^2 = \frac{1}{2}(S_X^2 + S_Y^2) = \frac{1}{2}S_Z^2.$$

$S_X^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 \quad (S_Y,\ S_Z \text{ analog})$

Ungepaart:

Man kann nicht mehr paarweise Differenzen bilden, auch wenn zufällig $m = n$ ist.

a) σ^2 bekannt: z**-Test**

$$T = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{N}(0,1)$$

unter jedem P_ϑ.

b) σ^2 unbekannt: t**-Test**

$$T = \frac{(\overline{X}_n - \overline{Y}_m) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

unter jedem $P_{\vec{\vartheta}}$, wobei (S=Schätzer für σ^2)

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

$S_Y^2 = \frac{1}{m-1}\sum_{i=1}^m (Y_i - \overline{Y}_m)^2$

$$S^2 = \frac{1}{m+n-2}\left((n-1)S_X^2 + (m-1)S_Y^2\right).$$

Hier sieht man auch die vielen Möglichkeiten $f_X(x)$ zu berechnen:

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{d}{dx}\lim_{y\to\infty} F(x,y)$$
$$= \frac{d}{dx}\lim_{y\to\infty} \overbrace{\left(\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\ dv\ du\right)}^{=F(x,y)}$$
$$= \frac{d}{dx}\left(\int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u,v)\ dv\ du\right) = \int_{-\infty}^{\infty} f(x,v)\ dv.$$

Thm. (Additionsregel)

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

Lem. Für beliebige $A \subseteq \Omega$ ist dann in einem Laplace-Raum:

$$P[A] = \frac{\text{Anz. Elementarereignisse in } A}{\text{Anz. Elementarereignisse in } \Omega} = \frac{|A|}{|\Omega|} = \frac{\text{“günstige”}}{\text{“mögliche”}}$$

Com. Abzählen: Kombinatorik

Thm. (Formel von Bayes) Ist $A_1, \dots, A_n$ eine Zerlegung von Ω mit $P[A_i] > 0$ für $i = 1, \dots, n$ und B ein Ereignis mit $P[B] > 0$, so gilt für jedes k

$$\mathrm{P}[A_k \mid B] \overset{\text{Def}}{=} \frac{\mathrm{P}[A_k \cap B]}{\mathrm{P}[B]} = \frac{\overbrace{\mathrm{P}[B \mid A_k]\,\mathrm{P}[A_k]}^{=\mathrm{P}[A_k \cap B]\text{ (Mult. Regel)}}}{\underbrace{\sum_{i=1}^{n} \mathrm{P}[B \mid A_i]\,\mathrm{P}[A_i]}_{=\mathrm{P}[B]\text{ (Satz. d. tot. Whs.)}}}.$$

Def. ((Stoch.) Unabh. Ereignisse)

$$P[A \cap B] = P[A]P[B]. \quad \text{bzw.} \quad P\left[\bigcap_{i=1}^{m} A_{k_i}\right] = \prod_{i=1}^{m} P[A_{k_i}].$$

Ist $P[A] = 0$ oder $P[B] = 0$, so sind A und B immer unabhängig. Für $P[A] \neq 0$ gilt

$$A, B \text{ unabhängig} \iff P[B|A] = P[B].$$