ulm university universität
uulm

Vorlesungsskript

Angewandte Stochastik II

Dr. Katharina Best

Wintersemester 2010/2011

Inhaltsverzeichnis

1 Grundideen der statistischen Datenanalyse

Im alltäglichen Sprachgebrauch versteht man unter *Statistik* eine Darstellung bzw. *Beschreibung* von Ergebnissen zusammengestellter Daten und Fakten unterschiedlichster Art, wie z. B. politischen Umfragen, ökonomischen Kenngrößen, Marktforschungsumfragen, klinischen Studien, Einwohnerdaten, usw.

Das hängt mit dem Ursprung des Wortes Statistik zusammen, das vom lateinischen *statisticum*, den Staat betreffend, abstammt. Insofern war die amtliche Statistik die erste beschreibende Statistik.

Gegenstand der mathematischen Statistik hingegen sind die Beschreibung und Untersuchung der Eigenschaften und Gesetzmäßigkeiten von Datensätzen. Dabei werden aus mehr oder weniger großen *Stichproben* Aussagen über die *Grundgesamtheit* gewonnen. Die untersuchten Daten sind dabei einerseits *reale Daten*, wie sie sich aus der Beobachtung von z. B. Vorgängen bzw. Strukturen in der Natur, Technik oder Wirtschaft ergeben, oder andererseits *synthetische Daten*, die bei der Simulation solcher Vorgänge bzw. Strukturen durch Algorithmen erzeugt werden.

Die grundlegende Idee der Statistik ist die *stochastische Modellierung* der Daten und Fragestellungen.

Dabei nutzt die Statistik Begriffe und Ergebnisse der Wahrscheinlichkeitsrechnung, wie sie in der Veranstaltung *Angewandte Statistik I* im Sommersemester 2010 behandelt wurden.

Dazu zählen

- Ereignisse und Wahrscheinlichkeit,
- Zufallsvariablen und Verteilung,
- stochastische Unabhängigkeit,
- Gesetz der großen Zahlen,
- zentraler Grenzwertsatz,

als wichtigste genannt.

1.1 Stichproben

Die genaueste und einfachste Antwort auf statistische Fragen erhält man, wenn man alle Elemente der Grundgesamtheit auf die interessierende Größe, das *Merkmal* untersucht, also eine *Vollerhebung* durchführt. Das ist jedoch meist nicht praktikabel und auch nicht nötig. Oft reicht es, aus einer Grundgesamtheit eine Stichprobe zu ziehen.

Im Hinblick auf die Aussagekraft der mathematischen Statistik, ist es erforderlich, dass diese Stichprobe hinsichtlich des Untersuchungsmerkmals möglichst repräsentativ ist und keinerlei Tendenzen aufzeigt.

So ist z. B. zur Untersuchung der Lebensverhältnisse von Personen mit einem Haushaltseinkommen von unter 1000 € die einseitige Befragung in einem Studentenwohnheim nicht geeignet.

Diese Eigenschaft der Repräsentativität ist keine, die die einzelne Stichprobe auszeichnet, denn da müsste man die Grundgesamtheit kennen, sondern eine Eigenschaft des Verfahrens zur Auswahl der Stichprobenelemente.

Zufallsstichproben

Angenommen, wir haben n Daten $x_1, x_2, \dots, x_n$. Dabei muss x_i nicht eine Zahl enthalten, sondern kann ein Vektor sein mit beispielsweise den Einträgen für Alter, Einkommen, Wohnort, Haushaltsgröße bei Einwohnerdaten. In dieser Vorlesung behandeln wir jedoch hauptsächlich eindimensionale Merkmale, also $x_i \in \mathbb{R}$ für alle $1 \leq i \leq n$.

Wir nehmen an, dass den Daten ein stochastisches Modell $X = (X_1, \dots, X_n)$ zu Grunde liegt. Dabei ist X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, also

$$X = (X_1, \dots, X_n) : \Omega \mapsto \mathcal{X}.$$

Definition 1.1 (Stichprobe): (a) $X = (X_1, \dots, X_n)$ heißt *Zufallsstichprobe*.
(b) Der Bildbereich $\mathcal{X}$ heißt *Stichprobenraum*.
(c) Jede Realisierung $X(\omega) = (x_1, \dots, x_n)$ der Zufallsvariable X heißt *(konkrete) Stichprobe*.
(d) Die Dimension n von X wird *Stichprobenumfang* genannt.

Anmerkung: Oft sind $X_1, \dots, X_n$ unabhängige, identisch verteilte Zufallsvariablen. Um von Varianz sprechen zu können, setzen wir $\mathbb{E}(X_i^2) < \infty$ für $1 \leq i \leq n$ voraus.

Gegenüber der Wahrscheinlichkeitsrechnung ist jetzt in der Statistik neu, dass das zu Grunde liegende Wahrscheinlichkeitsmaß $\mathbb{P}$ nicht gegeben ist. Vielmehr ist es Aufgabe der Statistik, das Wahrscheinlichkeitsmaß $\mathbb{P}$ bzw. die Verteilung von X über ihre Verteilungsfunktion F zu bestimmen, *schätzen* genannt.

Anmerkung: Statt einer Wahrscheinlichkeitsverteilung ist eine Familie $\mathcal{P}$ von Wahrscheinlichkeitsverteilungen gegeben. Diese lässt sich oft parametrisieren, d. h. $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, wobei Θ der Parameterraum ist.

Beispiel 1.2: Von einigen Schraubenschachteln haben sich die Aufkleber abgelöst. Man geht davon aus, dass die Schraubenbreite in jeder Schachtel normalverteilt ist und die Varianz gleich ist, sagen wir σ^2. In diesem Fall ist $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}\}$ mit dem freien Parameter μ.

Anmerkung: Zur Beschreibung von Verteilungen werden charakteristische Eigenschaften der Lage und Form der Häufigkeitsverteilungen verwendet. Dazu gehören verschiedene Mittel, wie das arithmetische oder das geometrische Mittel, empirische Streuung, oder auch die Schiefe der Häufigkeitsverteilung, um nur einige zu nennen.

Um diesen Aufgaben zu begegnen, benötigen wir Funktionen auf dem Stichprobenraum, die uns die Bewertungen der Stichprobe liefern.

Definition 1.3 (Stichprobenfunktion und Statistik): Eine messbare Abbildung $\varphi : \mathcal{X} \mapsto \mathbb{R}^m$ heißt *Stichprobenfunktion*. Auf die Zufallsstichprobe $X = (X_1, \ldots, X_n)$ angewendet, wird die Zufallsvariable $\varphi(X_1, \ldots, X_n)$ betrachtet und eine *Statistik* genannt.

Anmerkung: In der Schätztheorie wird eine Statistik *Schätzer* genannt, bei statistischen Tests spricht man von *Teststatistik*.

Im Folgenden wollen wir uns zuerst zwei Beispiele von Stichprobenfunktionen ansehen, das Stichprobenmittel und die Stichprobenvarianz. Seien von nun an (falls nicht anders angegeben) $X_1, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen auf $\mathbb{R}$ und $X = (X_1, \ldots, X_n)$ die Zufallsstichprobe.

1.2 Beispiele für Stichprobenfunktionen

1.2.1 Stichprobenmittel

Zunächst beschäftigt uns die Frage nach der Bestimmung des Erwartungswertes $\mu = \mathbb{E}X_1$ der Stichprobenvariablen aus der Stichprobe. Sei $\sigma^2 = \mathbb{V}\mathrm{ar}X_1$.

Definition 1.4 (Stichprobenmittel): Sei $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ mit $\varphi(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$ die Stichprobenfunktion, die einer Stichprobe ihr arithmetisches Mittel zuweist. Die Zufallsvariable

$$\overline{X}_n = \varphi(X) = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1.1}$$

heißt *Stichprobenmittel* der Zufallsstichprobe $(X_1, \ldots, X_n)$.

Aus dem Gesetz der großen Zahlen können wir den folgenden Satz über das Stichprobenmittel formulieren.

Satz 1.5: In Abhängigkeit des Erwartungswertes μ und der Varianz σ^2 der Stichprobenvariablen X_1 lassen sich

$$\mathbb{E}\overline{X}_n = \mu \tag{1.2}$$

und

$$\mathbb{V}\mathrm{ar}\overline{X}_n = \frac{\sigma^2}{n}, \tag{1.3}$$

der Erwartungswert und die Varianz von $\overline{X}_n$, angeben.

Wie wir sehen, ist $\overline{X}_n$ also ein geeigneter *Schätzer* für den Erwartungswert, der keinen *systematischen Fehler* macht. Er kann jedoch noch recht ungenau sein; über die *Schätzgenauigkeit* gibt uns die Varianz Auskunft. Wir können an dieser auch ablesen, dass die Schätzgenauigkeit mit wachsendem n verbessert wird.

Über die Verteilung von $\overline{X}_n$ lassen sich aus dem zentralen Grenzwertsatz Schlüsse ziehen.

Satz 1.6: Für alle $z \in \mathbb{R}$ gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z) \tag{1.4}$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist.

Damit können wir durch Betragsbildung die Wahrscheinlichkeit für die Abweichung des Stichprobenmittels $\overline{X}_n$ von μ um einen bestimmten Wert $\varepsilon > 0$ angeben.

1.2.2 Stichprobenvarianz

Als nächstes widmen wir uns der Frage nach der Bestimmung der Varianz $\mathbb{V}\mathrm{ar} X_1$ der Stichprobenvariablen aus der Stichprobe.

Definition 1.7 (Stichprobenvarianz): Wir betrachten die Stichprobenfunktion $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ gegeben durch $\varphi(x_1, \ldots, x_n) = \frac{1}{n-1} \sum\limits_{i=1}^{n} (x_i - \overline{x}_n)^2$. Die Zufallsvariable

$$S_n^2 = \varphi(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \tag{1.5}$$

heißt *Stichprobenvarianz* der Zufallsstichprobe $(X_1, \ldots, X_n)$.

Wir gehen analog zum Stichprobenmittel vor und geben den Erwartungswert und die Varianz von S_n^2 an.

Satz 1.8: In Abhängigkeit des Erwartungswertes μ und der Varianz σ^2 der Stichprobenvariablen X_1 errechnen sich der Erwartungswert und die Varianz von S_n^2 als

$$\mathbb{E} S_n^2 = \sigma^2 \tag{1.6}$$

und

$$\mathbb{V}\mathrm{ar} S_n^2 = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right) \tag{1.7}$$

wobei $\mu_4 = \mathbb{E} X_1^4$ das vierte Moment und $\sigma^4 = (\sigma^2)^2$ die quadrierte Varianz der Stichprobenvariablen X_1 bezeichnen.

Auch hier machen wir demnach keinen systematischen Fehler. Ebenfalls analog erhalten wir

Satz 1.9: Falls $\mathbb{E} X_1^4 < \infty$ ist, gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4}/\sqrt{n}} \leq z\right) = \Phi(z) \tag{1.8}$$

für alle $z \in \mathbb{R}$.

1.2.3 Empirische Verteilungsfunktion

Auch die Verteilungsfunktion F der Stichprobenvariablen X_1 lässt sich aus der Stichprobe schätzen. Der Fall ist nur insofern komplizierter, als wir eine Zufallsvariablen für alle Werte $y \in \mathbb{R}$ definieren, d. h. eine Familie von Zufallsvariablen haben.

Notation: Für eine Menge A gibt $|A|$ die *Mächtigkeit* von A an, d. h. die Anzahl der Elemente der Menge A.

Definition 1.10 (Empirische Verteilungsfunktion): Wir betrachten für alle $y \in \mathbb{R}$ die Stichprobenfunktion $\varphi : \mathbb{R}^n \times \mathbb{R} \mapsto [0,1]$ gegeben durch $\varphi(x_1, \ldots, x_n)(y) = \frac{1}{n} \left| \{ i : 1 \leq i \leq n \text{ mit } x_i \leq y \} \right|$, die die relative Häufigkeit der Stichprobenwerte angibt, welche den betrachteten Wert y nicht überschreiten. Wir erhalten die Familie $\left\{ \hat{F}_n(y) : y \in \mathbb{R} \right\}$ von Zufallsvariaben $\hat{F}_n(y) : \Omega \mapsto [0,1]$ mit

$$\hat{F}_n(y) = \varphi(X)(y) = \frac{1}{n} \left| \{ i : 1 \leq i \leq n \text{ mit } X_i \leq y \} \right| \tag{1.9}$$

und nennen sie die *empirische Verteilungsfunktion* der Zufallsstichprobe $(X_1, \ldots, X_n)$.

Anmerkung: Eine solche Familie von Zufallsvariablen ist ein *empirischer Prozess* und ist eine Klasse *stochastischer Prozesse*.

Auch bei der empirischen Verteilungsfunktion können wir Aussagen über Erwartungswert, Varianz und Verteilung treffen.

Satz 1.11: Für alle $y \in \mathbb{R}$ gelten folgende Eigenschaften:
(i) Die Zufallsvariable $n\hat{F}_n(y)$ ist binomialverteilt mit den Parametern n und $p = F(y)$, d. h. für $k = 0, 1, \ldots, n$ gilt

$$\mathbb{P}\left(n\hat{F}_n(y) = k \right) = \binom{n}{k} \left(F(y) \right)^k \left(1 - F(y) \right)^{n-k}, \tag{1.10}$$

(ii) insbesondere erhalten wir damit den Erwartungswert

$$\mathbb{E}\hat{F}_n(y) = F(y) \tag{1.11}$$

und die Varianz

$$\mathbb{V}\mathrm{ar}\hat{F}_n(y) = \frac{1}{n} F(y) \left(1 - F(y) \right). \tag{1.12}$$

(iii) Falls $0 < F(y) < 1$, so gilt für alle $z \in \mathbb{R}$ außerdem

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\hat{F}_n(y) - F(y)}{\sqrt{F(y)\left(1 - F(y)\right)} / \sqrt{n}} \leq z \right) = \Phi(z). \tag{1.13}$$

Beweis Die Zufallsvariable $n\hat{F}_n$ kann als die Anzahl der Erfolge beim n-maligen Münzwurf mit identischen Erfolgswahrscheinlichkeiten $p = F(x)$ aufgefasst werden. Daraus ergeben sich die Eigenschaften (i) und (ii), für (iii) wird dann der zentrale Grenzwertsatz für binomialverteilte Zufallsvariablen benötigt. □

Wir sehen, dass auch hier kein systematischer Fehler gemacht wird.

2 Parameterschätzung

Im folgenden wollen wir die Schätzung der Parameter systematischer betrachten. Dabei wollen wir sowohl untersuchen, wie gut der Schätzer an den wahren Wert heranreicht, also seine *Güte* beurteilen, als auch Methoden zur Konstruktion eines Schätzers kennenlernen.

2.1 Punktschätzung

Wir gehen von einer parametrischen Familie $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ aus, die *identifizierbar* ist, d. h. die Abbildung $\theta \mapsto P_\theta$ ist bijektiv, also aus $\theta_1 \neq \theta_2$ folgt, dass $P_{\theta_1} \neq P_{\theta_2}$, und umgekehrt. Sei dabei $\Theta \subset \mathbb{R}^m$ der Parameterraum und $\theta = (\theta_1, \ldots, \theta_m)$ der Parameter(vektor) des Wahrscheinlichkeitsmaßes P_θ mit der Verteilungsfunktion F_θ.

Sei $X = (X_1, \ldots, X_n)$ die Zufallsstichprobe und seien $X_1, \ldots, X_n$ unabhängig, identisch verteilt (i.i.d.) auf $\mathbb{R}$ mit der Verteilungsfunktion F_θ, die von dem unbekannten Parameter θ abhängt. Das Wahrscheinlichkeitsmaß $\mathbb{P}$ ist dann gegeben durch

$$\mathbb{P}\left(\left\{\omega : X_1(\omega) \leq y_1, \ldots, X_n(\omega) \leq y_n\right\}\right) = F_\theta(y_1) \cdot \ldots \cdot F_\theta(y_n) \tag{2.1}$$

für $y_1, \ldots, y_n \in \mathbb{R}$. Damit haben wir auch sofort über Grenzwertbildung in $n-1$ Dimensionen eine Aussage für die einzelnen Stichprobenvariablen X_i der Zufallsstichprobe

$$\lim_{\substack{y_j \to \infty \\ \text{für } j \neq i}} \mathbb{P}\left(\left\{\omega : X_1(\omega) \leq y_1, \ldots, X_n(\omega) \leq y_n\right\}\right) = \mathbb{P}\left(\left\{\omega : X_i(\omega) \leq y_i\right\}\right) = F_\theta(y_i) \tag{2.2}$$

mit $y_i \in \mathbb{R}$ für alle $1 \leq i \leq n$.

Notation: Damit können wir, um zu betonen, dass $\mathbb{P}$ kanonisch von dem Parameter θ abhängt, und sich des Unterschiedes zwischen P_θ und $\mathbb{P}$ bewusst seiend, die Notation $\mathbb{P} = \mathbb{P}_\theta$ einführen, analog für den Erwartungswert und die Varianz bezüglich $\mathbb{P}_\theta$ die Bezeichnungen $\mathbb{E}_\theta$ resp. $\mathbb{Var}_\theta$.

Wir wollen aus einer Stichprobe den Parameter(vektor) θ schätzen. Man nennt dann die Stichprobenfunktion $\hat{\theta} : \mathbb{R}^n \to \mathbb{R}^m$ einen *Punktschätzer*. Bevor wir uns damit beschäftigen, lernen wir noch einige Eigenschaften von Schätzern kennen.

2.1.1 Eigenschaften von Schätzern

Ein Schätzer soll keine systematischen Fehler machen, d. h. weder über- noch unterschätzen.

Definition 2.1 (Erwartungstreue und Bias): Ein Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für den Parameter θ, für den $\mathbb{E}_\theta \left|\hat{\theta}(X_1, \dots X_n)\right| < \infty$ für alle $\theta \in \Theta$ gilt, heißt *erwartungstreu*, falls

$$\mathbb{E}_\theta \hat{\theta} = \theta \tag{2.3}$$

für alle $\theta \in \Theta$. Er heißt *asymptotisch erwartungstreu*, falls

$$\lim_{n \to \infty} \mathbb{E}_\theta \hat{\theta} = \theta \tag{2.4}$$

für alle $\theta \in \Theta$.

Der *Bias* (oder die *Verzerrung*) des Schätzers ist gegeben durch

$$\mathrm{Bias}(\hat{\theta}) = \mathbb{E}_\theta \hat{\theta} - \theta, \tag{2.5}$$

strenggenommen $\mathrm{Bias}_\theta(\hat{\theta})$.

Ist $\hat{\theta}$ erwartungstreu, so gilt $\mathrm{Bias}(\hat{\theta}) = 0$.

Beispiel 2.2: Seien die Stichprobenvariablen X_i normalverteilt und sei

$$\mathcal{P} = \left\{ \mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0 \right\}$$

die parametrisierte Familie. Dann ist $\theta = (\mu, \sigma^2)$, also $m = 2$, und aus (1.2) und (1.6) folgt, dass der Schätzer $\hat{\theta} = (\overline{X}_n, S_n^2)$ für θ erwartungstreu ist.

Anmerkung: An dieser Stelle verstehen wir auch, wieso für die Varianz die Stichprobenvarianz S_n^2 dem kanonischer erscheinenden Schätzer $\frac{1}{n} \sum\limits_{i=1}^{n} (X_i - \overline{X}_n)^2$ gegenüber bevorzugt verwendet wird.

Definition 2.3 (Erwartete mittlere quadratische Abweichung): Der *mittlere quadratische Fehler*, englisch *mean squared error*, eines Schätzers $\hat{\theta}(X_1, \dots, X_n)$ für θ , für den $\mathbb{E}_\theta \left(\hat{\theta}(X_1, \dots X_n)^2 \right) < \infty$ für alle $\theta \in \Theta$ gilt, ist gegeben durch

$$\mathrm{MSE}(\hat{\theta}) := \mathbb{E}_\theta \left(|\hat{\theta} - \theta|^2 \right) \tag{2.6}$$

und lässt sich für $m = 1$ in der Form

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{Var}_\theta(\hat{\theta}) + \mathrm{Bias}_\theta(\hat{\theta})^2 \tag{2.7}$$

ausdrücken.

Ist $\hat{\theta}$ erwartungstreu, so gilt $\mathrm{MSE}(\hat{\theta}) = \mathbb{Var}_\theta(\hat{\theta})$.

Definition 2.4: Seien $\hat{\theta}_1(X_1, \dots, X_n)$ und $\hat{\theta}_2(X_1, \dots, X_n)$ zwei Schätzer für θ, für die für alle $\theta \in \Theta$ gilt $\mathbb{E}_\theta \left(\hat{\theta}(X_1, \dots X_n)^2 \right) < \infty$ und $\mathbb{E}_\theta \left(\hat{\theta}(X_1, \dots X_n)^2 \right) < \infty$. Dann sagt man, dass $\hat{\theta}_1$ *effizienter ist als* $\hat{\theta}_2$, falls

$$\mathrm{MSE}(\hat{\theta}_1) \leq \mathrm{MSE}(\hat{\theta}_2) \qquad \text{für alle } \theta \in \Theta \tag{2.8}$$

gilt. Sind die Schätzer erwartungstreu, ist im Fall $m = 1$ der Schätzer $\hat{\theta}_1$ *effizienter als* $\hat{\theta}_2$, falls

$$\mathbb{Var}_\theta(\hat{\theta}_1) \leq \mathbb{Var}_\theta(\hat{\theta}_2) \qquad \text{für alle } \theta \in \Theta \tag{2.9}$$

gilt.

Statt *effizienter* werden auch die Bezeichnungen *besser* oder *wirksamer* verwendet.

Definition 2.5 (Bester erwartungstreuer Schätzer): Ein erwartungstreuer Schätzer $\hat{\theta}(X_1, \ldots, X_n)$ für θ, für den $\mathbb{E}_\theta \left(\hat{\theta}(X_1, \ldots X_n)^2 \right) < \infty$ für alle $\theta \in \Theta$ gilt, ist der *beste erwartungstreue Schätzer* oder *effizienteste Schätzer*, falls er in der Klasse der erwartungstreuen Schätzer die kleinste erwartete mittlere quadratische Abweichung besitzt.

Für $m = 1$ ist der Schätzer $\hat{\theta}(X_1, \ldots, X_n)$ der beste erwartungstreue Schätzer, falls für jeden erwartungstreuen Schätzer $\tilde{\theta}(X_1, \ldots, X_n)$ gilt, dass

$$\mathbb{Var}_\theta \hat{\theta} \leq \mathbb{Var}_\theta \tilde{\theta} \qquad \text{für alle } \theta \in \Theta \tag{2.10}$$

d. h. $\hat{\theta}$ hat in der Klasse aller erwartungstreuen Schätzer für θ die minimale Varianz.

Definition 2.6 (Konsistenz): Ein Schätzer $\hat{\theta}(X_1, \ldots, X_n)$ für θ heißt

(i) *schwach konsistent*, falls $\hat{\theta} \xrightarrow{\mathrm{P}} \theta$ für $n \to \infty$, d. h. $\hat{\theta}$ stochastisch gegen θ konvergiert, d. h.

$$\lim_{n \to \infty} \mathbb{P}_\theta \left(|\hat{\theta}(X_1, \ldots, X_n) - \theta| > \epsilon \right) = 0 \qquad \text{für alle } \theta \in \Theta. \tag{2.11}$$

(ii) *stark konsistent*, falls $\hat{\theta} \xrightarrow{\text{f.s.}} \theta$ für $n \to \infty$, d. h. $\hat{\theta}$ fast sicher gegen θ konvergiert, d. h.

$$\mathbb{P}_\theta \left(\lim_{n \to \infty} \hat{\theta}(X_1, \ldots, X_n) = \theta \right) = 1 \qquad \text{für alle } \theta \in \Theta. \tag{2.12}$$

Definition 2.7 (Asymptotische Normalverteilung): Sei $\hat{\theta}(X_1, \ldots, X_n)$ ein Schätzer für θ.

(i) Der Fall $m = 1$. Falls $0 < \mathbb{Var}_\theta \hat{\theta} < \infty$ für alle $\theta \in \Theta$ und

$$\frac{\hat{\theta}(X_1, \ldots, X_n) - \mathbb{E}_\theta \hat{\theta}(X_1, \ldots, X_n)}{\sqrt{\mathbb{Var}_\theta \hat{\theta}(X_1, \ldots, X_n)}} \xrightarrow[n \to \infty]{\mathrm{d}} Y \sim \mathcal{N}(0, 1) \tag{2.13}$$

dann ist $\hat{\theta}$ *asymptotisch normalverteilt*.

(ii) Der Fall $m > 1$. Die Kovarianzmatrix $\mathbb{Cov}_\theta \hat{\theta}$ des Schätzers $\hat{\theta}$ sei für alle $n \in \mathbb{N}$ und für alle $\theta \in \Theta$ positiv definit. Bezeichne mit $\left(\mathbb{Cov}_\theta \hat{\theta} \right)^{1/2}$ die positiv definite Matrix L, für die $L \cdot L = \mathbb{Cov}_\theta \hat{\theta}$ gilt. Sei weiterhin

$$\left(\mathbb{Cov}_\theta \hat{\theta}(X_1, \ldots, X_n) \right)^{-1/2} \left(\hat{\theta}(X_1, \ldots, X_n) - \mathbb{E}_\theta \hat{\theta}(X_1, \ldots, X_n) \right) \xrightarrow[n \to \infty]{\mathrm{d}} Y \sim \mathcal{N}(0, \mathbb{I}_m) \tag{2.14}$$

wobei $\mathbb{I}_m$ die m-dimensionale Einheitsmatrix ist (und 0 den m-dimensionalen Nullvektor bezeichnet), dann ist $\hat{\theta}$ *asymptotisch normalverteilt*.

Anmerkung: Das Stichprobenmittel und die Stichprobenvarianz sind nach (1.4) und (1.8) asymptotisch normalverteilt. Außerdem sind sie sowohl schwach als auch stark konsistent, was wir nicht gezeigt haben.

2.1.2 Methoden der Punktschätzung

Momentenmethode

Bevor wir uns der Konstruktion des Schätzers widmen, müssen wir noch einige Bezeichnungen definieren.

Definition 2.8 (Moment und Stichprobenmoment): Seien $X, X_1, \ldots, X_n$ i. i. d. Zufallsvariablen und gelte $\mathbb{E}|X_i|^k < \infty$ für $k \in \mathbb{N}$. Dann ist

$$\mu_k := \mathbb{E}(X^k) \tag{2.15}$$

das *k-te Moment von X* und

$$\hat{\mu}_k := \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{2.16}$$

das *k-te Stichprobenmoment* oder das *k-te empirische Moment* der Stichprobe $X = (X_1, \ldots, X_n)$.

Wir gehen von einer Zufallsstichprobe $X = (X_1, \ldots, X_n)$ mit $X_1, \ldots, X_n$ i. i. d. aus.

Definition 2.9 (Momentenmethode-Schätzer): Es gelte $\mathbb{E}_\theta|X_i|^r$ für ein $r \geq m$ und seien die Momente $\mu_1, \ldots, \mu_r$ als Funktionen des Parametervektors $\theta = (\theta_1, \ldots, \theta_m) \in \Theta$ gegeben, d. h. es existieren für $1 \leq k \leq m$ Borel-messbare Funktionen g_k so dass $\mu_k = g_k(\theta_1, \ldots, \theta_m)$. Falls das *Momenten-Gleichungssystem*

$$\hat{\mu}_k = g_k(\theta), \qquad \text{für } 1 \leq k \leq r \tag{2.17}$$

eindeutig lösbar bezüglich θ ist, so heißt die Lösung

$$\hat{\theta}(X_1, \ldots, X_n) \tag{2.18}$$

Momentenschätzer oder *M-Schätzer* von θ.

Beispiel 2.10 (Momentenschätzer bei der Normalverteilung): Seien $X_1, \ldots, X_n$ normalverteilt mit Erwartungswert μ und Varianz σ^2, sei $\theta = (\mu, \sigma^2)$ der Parametervektor, für den wir einen M-Schätzer suchen. Für die Momente gelten folgende Gleichungen

$$\begin{aligned} g_1(\mu, \sigma^2) &= \mathbb{E}_\theta X_1 = \mu, \\ g_2(\mu, \sigma^2) &= \mathbb{E}_\theta X_1^2 = \mathbb{V}\mathrm{ar}_\theta X_1 + (\mathbb{E}_\theta X_1)^2 = \sigma^2 + \mu^2. \end{aligned} \tag{2.19}$$

Damit ergibt sich das Gleichungssystem

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} X_i &= \hat{\mu} \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 &= \hat{\sigma}^2 + \hat{\mu}^2 \end{aligned} \tag{2.20}$$

womit wir den Schätzer $(\hat{\mu}, \hat{\sigma}^2)$ mit

$$\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n, \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{2}(X_i - \overline{X}_n)^2 = \frac{n-1}{n} S_n^2 \end{aligned} \tag{2.21}$$

erhalten. Wir wissen schon, dass $\hat{\mu}$ erwartungstreu ist. Weiterhin gilt

$$\mathbb{E}_\theta \hat{\sigma}^2 = \frac{n-1}{n}\mathbb{E}_\theta S_n^2 = \frac{n-1}{n}\sigma^2,$$

weswegen $\hat{\sigma}^2$ nicht erwartungstreu ist.

Anmerkung: (a) Bei manchen Verteilungsfamilien besitzt das Gleichungssystem (2.18) für $r = m$ keine eindeutige Lösung uns es sind $r > m$ Gleichungen notwendig. Das ist z. B. der Fall, wenn einige Funktionen φ_i konstant sind.
(b) Die in Abschnitt 2.1.1 kennengelernten Eigenschaften gelten nicht allgemein, insbesondere sind die M-Schätzer nicht immer erwartungstreu.
(c) Die Momentenmethode liefert oft nicht den besten Schätzer, lässt sich aber im Allgemeinen gut berechnen und ist universell anwendbar.
(d) Unter gewissen Regularitätsbedingungen ist der M-Schätzer asymptotisch normalverteilt.

Maximum-Likelihood Schätzer

Eine weitere Methode der Parameterschätzung ist die Maximum-Likelihood-Methode.

Anmerkung: Ziel ist es, die Parameter so zu bestimmen, dass eine möglichst gute Anpassung des Modells an die beobachteten Daten erreicht wird. Anders gesagt gehen wir davon aus, dass die beobachteten Daten wahrscheinlich waren und maximieren bei der Maximum-Likelihood-Methode die Wahrscheinlichkeit für das Eintreten des Ereignisses. Wir suchen den Parameter des parametrischen Modells, der die höchste Wahrscheinlichkeit für die beobachteten Daten liefert.

Im Folgenden beschränken wir uns auf die Fälle, bei denen die Stichprobenvariablen entweder diskret sind, dann existiert eine Zähldichte p_θ, oder absolut stetig, dann existiert eine Dichtefunktion f_θ.

Anmerkung: Zur Erinnerung sei festgehalten, dass eine Zufallsvariable X diskret ist, wenn $\mathbb{P}(X \in C) = 1$ für eine abzählbare Menge $C \subset \mathbb{R}$ gilt. Die Zähldichte $\{p(y) : y \in C\}$ ist definiert durch $p(y) = \mathbb{P}(X = y)$ für $y \in C$.

Eine Zufallsvariable ist absolut stetig, wenn die Verteilungsfunktion F_X eine Integraldarstellung besitzt, i. e. $F_X(y) = \int_{-\infty}^{y} f_X(x)dx$ für alle $y \in \mathbb{R}$ gilt. Die Funktion f_X ist die Dichte der Verteilung.

Definition 2.11 (Likelihood-Funktion): (i) Seien die Stichprobenvariablen $X_1, \ldots, X_n$ diskret, sei $p_\theta(x)$ die Zähldichte und sei $x = (x_1, \ldots, x_n) \in \mathcal{X}$ gegeben. Dann ist die *Likelihoodfunktion* $L_x : \Theta \to \mathbb{R}$ gegeben durch
$$L_x(\theta) := \mathbb{P}_\theta(x) = p_\theta(x_1) \cdot \ldots \cdot p_\theta(x_n). \tag{2.22}$$
(ii) Seien die Stichprobenvariablen $X_1, \ldots, X_n$ absolut stetig, sei $f_\theta(x)$ die Dichtefunktion und sei die Stichprobe $x = (x_1, \ldots, x_n) \in \mathcal{X}$ gegeben. Dann ist die *Likelihoodfunktion* $L_x : \Theta \to \mathbb{R}$ gegeben durch
$$L_x(\theta) := \mathbb{P}_\theta(x) = f_\theta(x_1) \cdot \ldots \cdot f_\theta(x_n). \tag{2.23}$$

Wir wollen den Wert der Likelihood-Funktion maximieren, um einen Parameter zu erhalten, so dass die Stichprobe mit möglichst großer Wahrscheinlichkeit auftreten kann.

Definition 2.12 (Maximum-Likelihood-Schätzer): Sei $\hat{\theta} : \mathbb{R}^n \to \Theta \subset \mathbb{R}^m$ eine Stichprobenfunktion mit
$$L_x(\hat{\theta}) = \max_{\theta \in \Theta} L_x(\theta) \qquad \text{für alle } x \in X. \tag{2.24}$$
Dann wird $\hat{\theta}(X)$ der *Maximum-Likelihood-Schätzer* oder *ML-Schätzer* von θ genannt.

Wir verdeutlichen das Prinzip an einem Beispiel.

Beispiel 2.13: Wir betrachten eine Warenlieferung von 12 Exemplaren eines Artikels, bei der wir den Hersteller nicht kennen. Wir wissen, dass nur drei potentiell mögliche Hersteller in Frage kommen. Deren Lieferungen weisen folgende Ausschussanteile aus: $\theta_1 = 0,05$ (erster Hersteller), $\theta_2 = 0,10$ (zweiter Hersteller) und $\theta_3 = 0,15$ (dritter Hersteller).

Bei der Prüfung der Lieferung wird festgestellt, dass der Ausschuss genau ein Exemplar ist.

Wir wollen jetzt den vermutlichen Hersteller der vorliegenden Warenlieferung ermitteln.

Dazu betrachten wir die Stichprobe $X = (X_1, \dots, X_{12})$ mit

$$X_i(\omega) = \begin{cases} 1, & \text{falls das } i\text{-te Exemplar Ausschuss ist,} \\ 0 & \text{falls das } i\text{-te Exemplar in Ordnung ist.} \end{cases}$$

und die parametrische Familie $\mathcal{P} := \{\mathrm{Bin}(1, \eta) : \eta \in \{\theta_1, \theta_2, \theta_3\}\}$ mit den drei Bernoulli-Verteilungen zu den Parametern θ_1, θ_2 und θ_3.

Der Schätzer ist die Stichprobenfunktion $\hat{\theta} : \{0, 1\}^{12} \to \Theta$, die so gewählt ist, dass für alle Stichproben $x = (x_1, \dots, x_{12})$ mit $|\{i : x_i = 1\}| = 1$ die Likelihoodfunktion

$$L_x(\theta) = \mathbb{P}_\theta(\{X = x\}) = \theta(1-\theta)^{11}$$

maximal ist.

Dazu betrachten wir die Tabelle

θ	$L_x(\theta)$
$0,05$	$0,028$
$0,10$	$0,031$
$0,15$	$0,025$

aus der folgt, dass das Maximum bei $0,031$ liegt und somit $\hat{\theta}(x) = \theta_2$ ist, d. h. vermutlich war der zweite Hersteller (mit dem Ausschussanteil $0,10$) der Lieferant.

Anmerkung: (a) Bei manchen parametrischen Verteilungsfamilien ist der Maximum-Likelihood-Schätzer nicht eindeutig bestimmt.
(b) Erster Ansatz für das Optimierungsproblem ist Differentiation, die bereits oft zum Erfolg führt.
(c) Häufig gibt es keine analytische Lösung. Dann werden numerische Verfahren, z. B. die Newton-Raphson Methode verwendet. Als Startpunkt der Iteration eignet sich der Momentemethode-Schätzer.

Oft ist vorteilhafter, sich statt der Likelihoodfunktion die Loglikelihoodfunktion anzusehen.

Definition 2.14: Sei $x = (x_1, \dots, x_n) \in \mathcal{X}$ gegeben und $L_x : \Theta \to \mathbb{R}$ die *Likelihoodfunktion*. Dann ist $\log L_x(\theta)$ die *Loglikelihoodfunktion* der Stichprobe x.

Anmerkung: (a) Da der Logarithmus monoton ist, gilt

$$\operatorname{argmax} L_x(\theta) = \operatorname{argmax} \log L_x(\theta) \tag{2.25}$$

die Likelihoodfunktion und die Loglikelihoodfunktion nehmen ihr Maximum an selber Stelle an.
(b) Der Vorteil ist, dass das Produkt in der Likelihoodfunktion in eine Summe in der Loglikelihoodfunktion übergeführt wird, die oft einfacher handhabbar ist. Dies gilt insbesondere bei numerischen Betrachtungen.

Im Folgenden betrachten wir einige parametrische Familien und die entsprechenden ML-Schätzer.

Beispiel 2.15 (Bernoulli-Verteilung): Die parametrische Familie der Bernoulli-Verteilungen ist $\mathcal{P} = \{\mathrm{Bin}(1,\eta) : \eta \in [0,1]\}$. Für die Zähldichte gilt

$$p_\eta(y) = \eta^y (1-\eta)^{1-y} \qquad \text{für } y \in \{0,1\}$$

und die Likelihoodfunktion ist definiert als

$$L_x(\eta) = \prod_{i=1}^{n} \eta^{x_i}(1-\eta)^{1-x_i} \qquad \text{für } x = (x_1,\ldots,x_n) \in \mathcal{X} = \{0,1\}^n$$

Wir wollen jetzt das Maximum bestimmen: Falls $x = (\underbrace{0,\ldots,0}_{n \text{ mal}})$, so hat $L_x(\eta)$ bei $\eta = 0$ ein Maximum. Analog ist bei $x = (\underbrace{1,\ldots,1}_{n \text{ mal}})$ der Wert $\eta = 1$ das Maximum.

Andernfalls können wir die Loglikelihoodfunktion betrachten,

$$\log L_x(\eta) = \left(\sum_{i=1}^{n} x_i\right) \log \eta + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-\eta)$$

die im Intervall $(0,1)$ stetig ist mit den Grenzwerten

$$\lim_{\eta\to 0} \log L_x(\eta) = -\infty \qquad \text{und} \qquad \lim_{\eta \to 1} \log L_x(\eta) = -\infty,$$

woraus folgt, dass $L_x(\eta)$ in dem Intervall $(0,1)$ ein Maximum besitzt.

Um das Maximum zu berechnen, differenzieren wir nach η und erhalten die Gleichung

$$\frac{\partial}{\partial \eta} \log L_x(\eta) = \left(\sum_{i=1}^{n} x_i\right) \frac{1}{\eta} - \left(n - \sum_{i=1}^{n} x_i\right) \frac{1}{1-\eta} = 0,$$

welche die eindeutig bestimmte Lösung

$$\hat{\eta}(x) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

besitzt. Der ML-Schätzer für den Parameter η ist gegeben durch

$$\hat{\eta}(X) = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n$$

Analog gehen wir bei der Binomialverteilung vor.

Beispiel 2.16 (Binomialverteilung): Sei $m \in \mathbb{N}$ gegeben. Die parametrische Familie der Binomialverteilungen mit einem Umfang von m ist $\mathcal{P} = \{\mathrm{Bin}(m,\eta) : \eta \in [0,1]\}$. Für die Zähldichte gilt

$$p_\eta(y) = \binom{m}{y} \eta^y (1-\eta)^{m-y} \qquad \text{für } y \in \{0,1,\ldots,m\}$$

und die Likelihoodfunktion ist definiert als

$$L_x(\eta) = \prod_{i=1}^{n} \binom{m}{x_i} \eta^{x_i} (1-\eta)^{m-x_i} \qquad \text{für } x = (x_1, \ldots, x_n) \in \mathcal{X} = \{0, 1, \ldots, m\}^n.$$

Nach ähnlichen Vorüberlegungen wie im Beispiel 2.15, differenzieren wir nach η und erhalten die Gleichung

$$\frac{\partial}{\partial \eta} \log L_x(\eta) = \left(\sum_{i=1}^{n} x_i \right) \frac{1}{\eta} - \left(n \cdot m - \sum_{i=1}^{n} x_i \right) \frac{1}{1-\eta} = 0,$$

und somit den ML-Schätzer

$$\hat{\eta}(X) = \frac{\frac{1}{n} \sum_{i=1}^{n} X_i}{m} = \frac{\overline{X}_n}{m}.$$

Beispiel 2.17 (Normalverteilung): Wir betrachten die parametrische Familie der Normalverteilungen $\mathcal{P} = \left\{ \mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0 \right\}$. Für die Dichtefunktion gilt

$$f_{\mu,\sigma^2}(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2 \right) \qquad \text{für } y \in \mathbb{R}$$

und die Likelihoodfunktion ist definiert als

$$L_x(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right) \qquad \text{für } x \in \mathbb{R}^n$$

und die Loglikelihoodfunktion ist

$$\log L_x(\mu, \sigma^2) = -n \log\left(\sqrt{2\pi}\sigma \right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \qquad \text{für } x \in \mathbb{R}^n$$

Wir widmen uns zuerst dem Parameter μ. Die Abbildung $\mu \mapsto \log L_x(\mu, \sigma^2)$ ist stetig. Außerdem gilt

$$\lim_{\mu \to -\infty} \log L_x(\mu, \sigma^2) = -\infty \qquad \text{und} \qquad \lim_{\mu \to \infty} \log L_x(\mu, \sigma^2) = -\infty,$$

woraus folgt, dass $\mu \mapsto L_x(\mu, \sigma^2)$ in $\mathbb{R}$ ein Maximum besitzt. Differentiation nach μ liefert die Gleichung

$$\frac{\partial}{\partial \mu} \log L_x(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0,$$

welche die Lösung $\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}_n$ besitzt. Da $\frac{\partial^2}{\partial \mu^2} \log L_x(\mu, \sigma^2) = -\frac{n}{\sigma^2} < 0$ ist, ist die gefundene Lösung auch wirklich das Maximum, und zwar für jedes fest vorgegebene $\sigma^2 > 0$.

Jetzt untersuchen wir den Parameter σ^2. Die Abbildung $\sigma^2 \mapsto \log L_x(\overline{x}_n, \sigma^2)$ ist stetig für $\sigma^2 > 0$. Da für absolut stetige Zufallsvariablen $\mathbb{P}(X_1 = \ldots = X_n) = 0$ ist, gilt weiter

$$\lim_{\sigma^2 \to 0} \log L_x(\overline{x}_n, \sigma^2) = -\infty \qquad \text{und} \qquad \lim_{\sigma^2 \to \infty} \log L_x(\overline{x}_n, \sigma^2) = -\infty,$$

woraus folgt, dass $\sigma^2 \mapsto L_x(\overline{x}_n, \sigma^2)$ in $(0, \infty)$ ein Maximum besitzt. Differentiation nach σ^2 liefert die Gleichung

$$\frac{\partial}{\partial (\sigma^2)} \log L_x(\overline{x}_n, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \overline{x}_n)^2 = 0,$$

welche wegen der absoluten Stetigkeit die eindeutig bestimmte Lösung $\hat{\sigma}^2(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2$ besitzt. Die berechneten ML-Schätzer für die Parameter μ und σ^2 sind

$$\hat{\mu}(X) = \overline{X}_n \qquad \text{resp.} \qquad \hat{\sigma}^2(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

In den bisher betrachteten Beispielen erhielten wir mittels der Maximum-Likelihood-Methode Schätzer, die wir bereits von früher kannten. Dass die Methode auch neue Schätzer zur Verfügung stellt, sehen wir an der Verteilungsfamilie der Gamma-Verteilungen (siehe Übungsaufgabe).

Methode der kleinsten Quadrate

Die Methode der kleinsten Quadrate ist ein wichtiges Verfahren der Parameterschätzung, dem das folgende Szenario zu Grunde liegt:

Eine Zufallsvariable Y setzt sich aus einer Funktion h, welche von einer nicht zufälligen Größe x (z. B. auf $\mathbb{R}$) abhängig ist, und zufälligen Schwankungen Z zusammen

$$Y(x,\omega) := h(x) + Z(\omega) \qquad \text{für alle } x \in \mathbb{R} \text{ und alle } \omega \in \Omega, \tag{2.26}$$

wobei Ω der Grundraum des Wahrscheinlichkeitsraums ist, auf dem Z definiert ist.

Annahmen Wir gehen von folgenden Bedingungen aus:

$$\mathbb{E}Z = 0, \qquad \text{d. h. } \mathbb{E}Y = h, \tag{2.27}$$

$$\mathbb{V}\mathrm{ar}Z = \sigma^2 \qquad \text{(unabhängig von } x\text{)} \tag{2.28}$$

Zu den Werten $x_1, \ldots, x_n$ werden die Werte $y_1 := Y_1(x_1), \ldots, y_n := Y_n(x_n)$ beobachtet; diese immer auch mit den zufälligen Schwankungen. Definiere die Vektoren $x = (x_1, \ldots, x_n)$ und $y = (y_1, \ldots, y_n)$.

Dabei ist zu beachten, dass im Gegensatz zu den bisherigen Modellannahmen, hier Y_i sowohl unterschiedliche Verteilungen haben (können) und im Allgemeinen auch nicht unabhängig sind.

Zu schätzen ist die Funktion h. Diese wird häufig als eine aus der Funktionenschar $\mathcal{F}_\Theta = \{f_\theta : \theta \in \Theta\}$ modelliert, weswegen wir es erneut mit einer Parameterschätzung zu tun haben. Deswegen betrachte h_θ.

> **Beispiel 2.18:** Betrachte ein Bremsexperiment, bei dem x die Geschwindigkeit und Y den Bremsweg darstellt. Dan ist h_θ die Funktion für den mittleren Bremsweg in Abhängigkeit von der Geschwindigkeit und F_θ ist die Menge aller Polynome.

Vorgehen: Definiere die Summe der quadratischen Abweichungen

$$Q(\tilde{h}) := \sum_{i=1}^{n}\Big(y_i - \tilde{h}(x_i)\Big)^2 \tag{2.29}$$

für $\tilde{h} \in \mathcal{F}_\Theta$.

Kleinste-Quadrate Schätzer ist die Funktion $\hat{h} := h_{x,\hat{\theta}} \in \mathcal{F}_\Theta$, für die $Q(\hat{h})$ minimal ist, also

$$\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}}\, Q(h_\theta) = \underset{\theta\in\Theta}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - h_\theta(x_i)\right)^2 \tag{2.30}$$

Anmerkung: (a) Oft ist h linear. Dann ist der Kleinste-Quadrate-Schätzer der beste unter allen linearen erwartungstreuen Schätzern.
(b) Häufig werden die Z_i als normalverteilt modelliert, somit sind auch Y_i normalverteilt. Dann stimmen die Kleinste-Quadrate-Schätzungen mit der ML-Schätzung überein.
(c) Wir werden im Rahmen der Regression auf die Kleinste-Quadrate-Schätzer noch einmal eingehen.

Bayes Schätzer

Motivation ist in der subjektivistischen Wahrscheinlichkeitsauffassung begründet, d. h.
(i) die Wahrscheinlichkeit ist ein Maß für die Unsicherheit und hängt vom Kenntnisstand des Betrachters ab;
(ii) es gibt keine Trennung von zufälligen Größen (Zufallsvariable) und nicht zufälligen Größen (Parametern).

Vorgehen Der unbekannte Parameterwert θ wird als Realisation einer Zufallsvariablen Z aufgefasst. Die Verteilung von Z stellt den jeweiligen Kenntnisstand des Betrachters über den Parameter dar. Die Wertemenge von Z ist Θ.

Wir haben es mit zwei Verteilungen für Z zu tun:

a priori Verteilung $\mathbb{P}_{pri}$ entspricht den Vorkenntnissen des Beobachters über die Verteilung von Z vor der Beobachtung der Stichprobe.

a posteriori Verteilung $\mathbb{P}_{post}$ entspricht dem neuen Kkenntnisstand des Beobachters über die Verteilung von Z nach der Beobachtung der Stichprobe.

Das Vorgehen entspricht dem Lernen aus Beobachtungen. Die Verbindung zwischen der a priori Verteilung und der Verteilung der Stichprobe einerseits und der a postieriori Verteilung andererseits liefert die Formel von Bayes.

Satz 2.19 (Satz von Bayes): Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und es seien $B_1, \dots, B_k \in \mathcal{F}$ eine Partition von Ω. Weiterhin gelte $\mathbb{P}(B_i) > 0$ für alle $1 \leq i \leq n$. Dann gilt

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

für alle $A \in \mathcal{F}$ mit $\mathbb{P}(A) > 0$ und alle $1 \leq k \leq n$.

Zur Schätzung des Parameters wird dann eine Verlustfunktion für den erwarteten Verlust, dass der Parameter einen bestimmten Wert annimmt, definiert.

Definition 2.20: Eine messbare Funktion $l : \Theta^2 \to \mathbb{R}_+$ ist eine Verlustfunktion.

Definition 2.21: Ein Schätzer $\hat{\theta}$ heißt *Bayes-Schätzer* des Parameters θ, falls

$$\hat{\theta}(x,y) = \operatorname*{argmin}_{\theta\in\Theta} \mathbb{E}_{post} l(Z,\theta) \tag{2.31}$$

existiert und eindeutig ist.

Beispiel 2.22 (Quadratische Verlustfunktion): In diesem Fall ist $l(Z,\theta) := (Z-\theta)^2$, dann gilt

$$\begin{aligned}\min_{\theta\in\Theta} \mathbb{E}_{post} l(Z,\theta) &= \min_{\theta\in\Theta} \mathbb{E}_{post}(Z-\theta)^2 \\ &= \min_{\theta\in\Theta}(\mathbb{E}_{post} Z^2 + \underbrace{\theta^2 - 2\theta\,\mathbb{E}_{post} Z}_{\text{Parabel mit Minimum bei } \mathbb{E}_{post} Z})\end{aligned}$$

Dann ist also $\hat{\theta} = \operatorname*{argmin}_{\theta\in\Theta} \mathbb{E}_{post}(Z-\theta)^2 = \mathbb{E}_{post} Z$.

Vorgehen bei diskreten Zufallsvariablen Seien die Stichprobenvariablen $X_1,\dots,X_n$ diskret, Z ebenfalls. Die Version des Satzes von Bayes für diskrete Zufallsvariablen liefert die a posteriori Verteilung $\mathbb{P}_{post}$ durch die Zähldichte

$$\mathbb{P}_{post}(Z=\theta) = \mathbb{P}(Z=\theta|X_1=x_1,\dots,X_n=x_n) = \frac{\mathbb{P}(X_1=x_1,\dots,X_n=x_n|Z=\theta)\cdot\mathbb{P}_{pri}(Z=\theta)}{\sum\limits_{z\in\Theta}\mathbb{P}(X_1=x_1,\dots,X_n=x_n|Z=z)\cdot\mathbb{P}_{pri}(Z=z)}.$$

Beispiel 2.23: Es seien N Kugeln in einer Urne, davon M einer ausgewiesenen Eigenschaft, sagen wir rot, wobei M unbekannt ist. Wir betrachten Ziehen mit Zurücklegen. Dann kann der Parameter η dafür eine rote Kugel zu ziehen die Werte $0, 1/N, 2/N, \dots, 1$ annehmen. Da sonst keine Information über η vorhanden ist, wird Gleichverteilung als a priori Verteilung angenommen, also $\mathbb{P}_{pri}(Z=\eta) = \frac{1}{N+1}$ für alle $\eta = 0, 1/N, \dots, 1$.

Nun ziehen wir n-mal. Dann sind $X_i \sim \text{Bin}(1,\eta)$ Bernoulli-verteilt, falls $Z=\eta$ ist. Der Wertebereich von X_i ist $\{0,1\}$ und es gilt

$$\mathbb{P}(X_i=x_i|Z=\eta) = \eta^{x_i}(1-\eta)^{1-x_i}$$

für alle $i=1,\dots,n$. Da X_i unabhängig sind, haben wir

$$\mathbb{P}(X_1=x_1,\dots,X_n=x_n|Z=\eta) = \prod_{i=1}^n \mathbb{P}(X_i=x_i|Z=\eta) = \eta^{\sum x_i}(1-\eta)^{n-\sum x_i}$$

als gemeinsame Verteilung von $X_1,\dots,X_n$ falls Z den Wert η hat.

Die a posteriori Verteilung ist dann

$$\begin{aligned}\mathbb{P}_{post}(Z=\eta) &= \mathbb{P}(Z=\eta|X_1=x_1,\dots,X_n=x_n) \\ &= \frac{\mathbb{P}(X_1=x_1,\dots,X_n=x_n|Z=\eta)\cdot\mathbb{P}_{pri}(Z=\eta)}{\sum\limits_{z\in\{0,1/N,\dots,1\}}\mathbb{P}(X_1=x_1,\dots,X_n=x_n|Z=z)\cdot\mathbb{P}_{pri}(Z=z)} \\ &= \frac{\eta^{\sum x_i}(1-\eta)^{n-\sum x_i}\cdot\frac{1}{N+1}}{\sum\limits_{z\in\{0,1/N,\dots,1\}} z^{\sum x_i}(1-z)^{n-\sum x_i}\frac{1}{N+1}} \\ &= \frac{\eta^{\sum x_i}(1-\eta)^{n-\sum x_i}}{\sum\limits_{z\in\{0,1/N,\dots,1\}} z^{\sum x_i}(1-z)^{n-\sum x_i}},\end{aligned}$$

wobei sich auf Grund der Gleichverteilung als a posteriori Verteilung der Term $\frac{1}{N+1}$ kürzen lässt.

Beobachtung: Die a posteriori Verteilung hängt nicht von den einzelnen X_i sondern nur von der Funktion $g(X_1, \ldots, X_n) = \sum_i X_i$ ab. Zur Komplexitätsreduktion wird dann die a posteriori Verteilung bezüglich einer Funktion der Stichprobe berechnet.

Beispiel 2.24: Wir konkretisieren das vorherige Beispiel. Sei $N = 8, n = 4$ und $\sum x_i = 3$. Dann ist

$$\mathbb{P}_{post}(Z = \eta) = \frac{\eta^3(1-\eta)}{\sum\limits_{i=0}^{8} (i/8)^3(1 - i/8)} = \frac{1024}{399}\eta^3(1-\eta)$$

Das liefert uns die folgenden Werte:

η	0	$1/8$	$2/8$	$3/8$	$4/8$	$5/8$	$6/8$	$7/8$	$8/8$
$\mathbb{P}_{pri}$	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$
$\mathbb{P}_{post}$	0	0,004	0,030	0,085	0,160	0,235	0,271	0,215	0

und wählen deswegen $\hat{\eta} = 6/8$ als das η mit der höchsten a posteriori Wahrscheinlichkeit für den Bayes-Schätzer.

Vorgehen bei stetigen Zufallsvariablen Hier seien die Stichprobenvariablen $X_1, \ldots, X_n$ stetig, Z ebenfalls. Die Version des Satzes von Bayes für stetige Zufallsvariablen liefert die a posteriori Verteilung $\mathbb{P}_{post}$ mit der Dichte

$$f_{post}(\theta) = f(\theta|x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n|\theta) \cdot f_{pri}(\theta)}{\sum\limits_{z\in\Theta} f(x_1, \ldots, x_n|z) \cdot f_{pri}(z)}\,.$$

2.2 Prüfverteilungen

Aus der Normalverteilung lassen sich weitere wichtige Verteilungen gewinnen. Da diese sowohl für die Intervallschätzung als auch später für die Testverfahren benötigt werden, seien sie hier eingefügt.

Bevor wir beginnen, definieren wir den Begriff der Quantilfunktion.

Definition 2.25: Sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion. Dann ist die Umkehrabbildung $F^{-1} : [0,1] \to \mathbb{R}$ mit

$$F^{-1}(y) = \inf\{x : F(x) \geq y\} \tag{2.32}$$

die *Quantilfunktion* von F. Sei weiter $\alpha \in (0,1)$. Dann bezeichnet man die Zahl $F^{-1}(\alpha)$ als das α-*Quantil* von F.

Anmerkung: Die Quantilfunktion existiert, da die Verteilungsfunktion F monoton wachsend und rechststetig ist.

Beispiel 2.26 (Quantilfunktion der Normalverteilung): Bezeichne Φ die Verteilungsfunktion der Standardnormalverteilung. Dann wird das α-Quantil mit z_α für alle $\alpha \in (0,1)$ notiert. Es gilt $\Phi(z_\alpha) = \alpha$.

Wegen der Symmetrieeigenschaft der Standardnormalverteilung ist

$$\Phi(-x) = 1 - \Phi(x)$$

für alle $x \in \mathbb{R}$ und somit

$$z_\alpha = -z_{1-\alpha} \tag{2.33}$$

für alle $\alpha \in (0,1)$.

2.2.1 χ^2-Verteilung

Definition 2.27: Seien $X_1, \dots, X_n$ unabhängige und identisch $\mathcal{N}(0,1)$ verteilte Zufallsvariablen. Dann wird die Verteilung der Summe ihrer Quadrate

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

als *Chi-Quadrat-verteilt mit n Freiheitsgraden* bezeichnet, kurz χ_n^2-verteilt.

Anmerkung: Eine χ^2-verteilte Zufallsvariable ist nicht symmetrisch. Für kleine n ist die Dichte deutlich linksteil. Für wachsendes n nähert sich die χ^2-Verteilung als Folge des zentralen Grenzwertsatzes der Normalverteilung an.

Satz 2.28: Sei Z_n eine χ^2-verteilte Zufallsvariable mit n Freiheitsgraden, sei F_{Z_n} ihre Verteilungsfunktion. Dann ist

$$\begin{aligned} \mathbb{E}Z_n &= n, \\ \mathbb{V}\mathrm{ar}Z_n &= 2n, \end{aligned}$$

außerdem gilt für große $n > 30$ und alle $\alpha \in (0,1)$ die folgende Approximation der Quantilfunktion

$$F_{Z_n}^{-1}(\alpha) = \frac{1}{2}\left(z_\alpha + \sqrt{2n-1}\right)^2,$$

wobei z_α die Quantilfunktion $\mathcal{N}(0,1)$-Verteilung ist.

Satz 2.29 (Additionssatz): Die Summe zweier unabhängiger Zufallsvariablen, die χ_n^2-verteilt resp. χ_m^2-verteilt sind, ist χ_{n+m}^2-verteilt.

2.2.2 t-Verteilung

Definition 2.30: Seien X eine $\mathcal{N}(0,1)$-verteilte und Z_n eine von X unabhängige χ_n^2-verteilte Zufallsvariable. Dann wird die Verteilung der Zufallsvariablen

$$T := \frac{X}{\sqrt{\frac{1}{n}Z_n}}$$

als *t-Verteilung mit n Freiheitsgraden* bezeichnet.

Satz 2.31: Sei T eine t-verteilte Zufallsvariable mit n Freiheitsgraden, sei F_T ihre Verteilungsfunktion. Dann ist

$$\begin{aligned} \mathbb{E}T &= 0 \\ \mathbb{V}\mathrm{ar}T &= \frac{n}{n-2} \end{aligned}$$

für $n \geq 3$ und es gilt für alle $\alpha \in (0,1)$

$$F_T^{-1}(\alpha) = 1 - F_T^{-1}(1-\alpha),$$

was der Symmetrie der Verteilung geschuldet ist.

2.2.3 *F*-Verteilung

Definition 2.32: Seien X und Y zwei unabhängige Zufallsvariablen, die χ^2_m-verteilt resp. χ^2_n-verteilt sind. Dann wird die Verteilung der Zufallsvariablen

$$Z := \frac{\frac{1}{m}X}{\frac{1}{n}Y}$$

als *F-Verteilung mit den Freiheitsgraden m und n* bezeichnet.

Satz 2.33: Sei Z eine F-verteilte Zufallsvariable mit den Freiheitsgraden m und n. Dann gilt

$$\begin{aligned} \mathbb{E}Z &= \frac{n}{n-2}, \\ \mathbb{V}\mathrm{ar}Z &= \frac{2n^2(n+m-2)}{m(n-4)(n-2)^2}. \end{aligned}$$

Falls $F_{m,n}$ die Verteilungsfunktion von Z ist und $F_{n,m}$ die Verteilungsfunktion einer F-verteilten Zufallsvariablen mit den Freiheitsgraden n und m, so lassen sich die Quantilfunktionen über

$$F_{m,n}^{-1}(\alpha) = \frac{1}{F_{n,m}^{-1}(1-\alpha)}$$

für alle $0 < \alpha < 1$ miteinander in Bezug setzen.

Das wichtigste Beispiel zur Verwendung der Prüfverteilungen sei an dieser Stelle als Satz formuliert.

Satz 2.34: Seien $X_1, \ldots, X_n$ unabhängige und identisch $\mathcal{N}(\mu, \sigma^2)$-verteilte Zufallsvariablen, z. B. eine Zufallsstichprobe. Dann sind

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{und} \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

voneinander stochastisch unabhängig und es gilt

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \sigma^2/n\right) \qquad \text{und} \qquad \frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1}.$$

Beweis Wir zeigen nur die Verteilungsaussage. Der erste Teil gilt wegen der Linearität des Erwartungswertes und der Unabhängigkeit der Zufallsvariablen, da

$$\mathbb{V}\mathrm{ar}(\overline{X}_n) = \mathbb{V}\mathrm{ar}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

Der zweite Teil folgt aus folgender Überlegung: Es gilt

$$\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\overline{X}_n)^2 + \left(\frac{1}{\sigma/\sqrt{n}}(\overline{X}_n-\mu)\right)^2 \tag{2.34}$$

nach Teilung durch σ^2 der Identität

$$\begin{aligned}
\sum_{i=1}^n (X_i-\mu)^2 &= \sum_{i=1}^n \Big((X_i-\overline{X}_n)+(\overline{X}_n-\mu)\Big)^2 \\
&= \sum_{i=1}^n \Big((X_i-\overline{X}_n)^2 + 2(X_i-\overline{X}_n)(\overline{X}_n-\mu) + (\overline{X}_n-\mu)^2\Big) \\
&= \sum_{i=1}^n (X_i-\overline{X}_n)^2 + 2(\overline{X}_n-\mu)\sum_{i=1}^n (X_i-\overline{X}_n) + n(\overline{X}_n-\mu)^2 \\
&= \sum_{i=1}^n (X_i-\overline{X}_n)^2 + n(\overline{X}_n-\mu)^2
\end{aligned}$$

Der Term auf der linken Seite von (2.34) ist als Summe von n Quadraten $\mathcal{N}(0,1)$ verteilter Zufallsvariablen χ^2_n-verteilt. Der zweite Summand des rechten Terms ist aus dem gleichen Grund χ^2_1-verteilt. Nun gilt, dass $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\overline{X}_n)^2$ wegen des Additionssatzes 2.29 χ^2_{n-1}-verteilt ist. □

2.3 Intervallschätzung

Der Punktschätzer $\hat{\theta}$ eines Parameters, ist normalerweise nicht mit diesem identisch. Zu dem Schätzer selbst ist es notwendig zu wissen, welche Präzision das Schätzverfahren mitbringt. Bei erwartungstreuen Schätzern ist die Varianz des Schätzers ein Maß für die Präzision.

Einen anderen Ansatz verfolgt die Intervallschätzung. Statt eines einzelnen Wertes ist das Ergebnis der Schätzung ein Intervall. Dabei wird die Wahrscheinlichkeit, dass das Intervall den Parameter nicht enthält kontrolliert.

Anmerkung (Annahmen in diesem Abschnitt): Wir gehen von einer parametrischen Verteilungsfamilie $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^m\}$ mit den Verteilungsfunktionen F_θ aus. Wir wollen vorerst nur die j-te Komponente des Parameters $\theta = (\theta_1, \ldots, \theta_m)$ bestimmen.

Für die Intervallschätzung benötigen wir zwei Stichprobenfunktionen, $\theta_u(X_1,\ldots,X_n)$ für die untere und $\theta_o(X_1,\ldots,X_n)$ für die obere Intervallgrenze.

Definition 2.35 (Konfidenzintervall): Sei $\alpha \in (0,1)$ beliebig, aber fest gewählt. Dann heißt das zufällige Intervall $\big(\theta_u(X_1,\ldots,X_n), \theta_o(X_1,\ldots,X_n)\big)$ ein *Konfidenzintervall zum Niveau* $(1-\alpha)$ für die Parameterkomponente θ_j , falls

$$\mathbb{P}_\theta\left(\theta_u(X_1,\ldots,X_n) \le \theta_o(X_1,\ldots,X_n)\right) = 1, \tag{2.35}$$

$$\mathbb{P}_\theta\Big(\theta_u(X_1,\ldots,X_n) \le \theta_j \le \theta_o(X_1,\ldots,X_n)\Big) \ge 1-\alpha \tag{2.36}$$

gelten. Alternativ zu (2.50) kann die Bedingung

$$\inf_{\theta \in \Theta} \mathbb{P}_\theta \left(\theta_u(X_1, \ldots, X_n) \leq \theta_j \leq \theta_o(X_1, \ldots, X_n) \right) = 1 - \alpha \tag{2.37}$$

oder

$$\lim_{n \to \infty} \mathbb{P}_\theta \left(\theta_u(X_1, \ldots, X_n) \leq \theta_j \leq \theta_o(X_1, \ldots, X_n) \right) \geq 1 - \alpha \tag{2.38}$$

betrachtet werden, wobei man im letzteren Fall vom *asymptotischen Konfidenzintervall* spricht.

Damit haben wir ein zweiseitiges Konfidenzintervall.

Anmerkung (Irrtumswahrscheinlichkeit): Das Konfidenzniveau $1 - \alpha$ bestimmt, mit welcher Wahrscheinlichkeit ein Intervall entsteht, das den unbekannten Parameter θ_j enthält. Umgekehrt ist α die Wahrscheinlichkeit ein Intervall zu enthalten, das θ_j nicht enthält. Deswegen wird α auch *Irrtumswahrscheinlichkeit* genannt.

Anmerkung (Vorgehen): Zur Bestimmung des Konfidenzintervalls
(i) gibt man sich ein α vor,
(ii) wählt man zwei Stichprobenfunktionen θ_u und θ_o, die die Bedingung (2.49) und eine der Bedingungen (2.50)–(2.38) erfüllen,
(iii) berechnet die Werte $\theta_u(X_1, \ldots, X_n)$ und $\theta_o(X_1, \ldots, X_n)$

Anmerkung: Wir werden uns im folgenden auf symmetrische Konfidenzintervalle beschränken. Damit ist die Wahrscheinlichkeit, dass θ unterhalb der unteren Grenze θ_u liegt, dieselbe wie die Wahrscheinlichkeit, dass θ oberhalb der oberen Grenze θ_o liegt.

Anmerkung: Will man die Abweichung der Schätzung vom wahren Wert nur in eine Richtung kontrollieren, so kann man eine der Grenzen durch $-\infty$ bzw. ∞ ersetzen. Dann erhält man ein einseitiges Konfidenzintervall. Das weitere Vorgehen ist analog.

Ob ein zweiseitiges oder ein einseitiges Konfidenzintervall angebracht ist, hängt von der Fragestellung ab.

Beispiel 2.36: (a) These: Frauen sind intelligenter als Männer. Hier ist ein einseitiges Konfidenzintervall zu verwenden.
(b) These: Frauen sind genauso intelligent wie Männer. Hier wird ein zweiseitiges Konfidenzintervall berechnet.

2.3.1 Eine-Stichprobe-Probleme

Konfidenzintervall für den Erwartungswert μ bei bekannter Varianz σ^2

Sei $X = (X_1, \ldots, X_n)$ die Zufallsstichprobe mit iid Stichprobenvariablen und sei $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ unbekannt und σ^2 bekannt ist. Dann besitzt $\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n}$ eine $\mathcal{N}(0,1)$-Verteilung.

Sei α die vorgegebene Irrtumswahrscheinlichkeit. Dann ergibt sich, dass

$$\mathbb{P}\left(z_{\alpha/2} \leq \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \leq z_{1-\alpha/2} \right) = 1 - \alpha. \tag{2.39}$$

und mit (2.33) folgt

$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \leq z_{1-\alpha/2}\right) = 1-\alpha,$$

was umgeformt wird

$$\mathbb{P}\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \overline{X}_n - \mu \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha,$$

und weiter zu

$$\mathbb{P}\left(\overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

Damit sind mit den Stichprobenfunktionen

$$\theta_u(X_1,\ldots,X_n) = \overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \text{und} \qquad \theta_o(X_1,\ldots,X_n) = \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{2.40}$$

die Gleichungen (2.37) und (2.39) äquivalent. Da die Standardabweichung bekannt ist, sind θ_u und θ_o konkret berechenbar.

Die Länge des Zufallsintervalls $[\theta_u, \theta_o]$ ist

$$L(X_1,\ldots,X_n) = \theta_o(X_1,\ldots,X_n) - \theta_u(X_1,\ldots,X_n) = 2\,z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{2.41}$$

und damit nicht zufällig. Zu einer vorgegebenen Länge l ist dann

$$n = \left\lceil \left(\frac{2\sigma z_{1-\alpha/2}}{l}\right)^2 \right\rceil \tag{2.42}$$

der benötigte Stichprobenumfang bei festem σ^2 und fester Irrtumswahrscheinlichkeit α.

Beispiel 2.37: Mittels eines Schneidegerätes wird ein Papierband in Stücke einer bestimmten Länge μ geschnitten, die stufenlos eingestellt werden kann. Dabei ist die Längenungenauigkeit unabhängig von der eingestellten Länge. Die geschnittene Papierlänge wird als normalverteilt mit dem Erwartungswert μ und der Standardabweichung $\sigma = 2,4$ (in mm) angesehen.

Es werden $n = 9$ Stücke geschnitten und ihre Länge nachgemessen. Die Werte sind

$$184,2 \quad 182,6 \quad 185,3 \quad 184,5 \quad 186,2 \quad 183,9 \quad 185,0 \quad 187,1 \quad 184,4$$

Zur Irrtumswahrscheinlichkeit $\alpha = 0,01$ soll eine Intervall-Schätzung durchgeführt werden. Aus den Gleichungen (2.40) ergibt sich die Realisierung $[182,74; 186,86]$ des Zufallsintervalls. Dieses hat die Länge $4,12$.

Um eine Länge von höchstens 3 zu erhalten, benötigen wir nach (2.42) eine Stichprobe der Länge mindestens 17.

Konfidenzintervall für den Erwartungswert μ bei unbekannter Varianz σ^2

Sei $X = (X_1, \dots, X_n)$ die Zufallsstichprobe mit iid Stichprobenvariablen und sei $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ und σ^2 beide unbekannt sind. Da wir jetzt σ nicht kennen, können wir das Konfidenzintervall aus (2.40) nicht verwenden.

Wir nutzen deswegen aus, dass

$$\frac{\overline{X}_n - \mu}{S_n}\sqrt{n}$$

t-verteilt mit $n-1$ Freiheitsgraden ist. Bezeichne mit $t_{n-1,\alpha/2}$ das $\alpha/2$-Quantil der t-Verteilung, $t_{n-1,1-\alpha/2}$ entsprechend. Dann gilt

$$\mathbb{P}\left(t_{n-1,\alpha/2} \le \frac{\overline{X}_n - \mu}{S_n}\sqrt{n} \le t_{n-1,1-\alpha/2}\right) = 1 - \alpha$$

und wegen der Symmetrie

$$\mathbb{P}\left(-t_{n-1,1-\alpha/2} \le \frac{\overline{X}_n - \mu}{S_n}\sqrt{n} \le t_{n-1,1-\alpha/2}\right) = 1 - \alpha,$$

was umgeformt

$$\mathbb{P}\left(\overline{X}_n - t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}} \le \mu \le \overline{X}_n + t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}}\right) = 1 - \alpha$$

ergibt. Die Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) = \overline{X}_n - t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}} \qquad \text{und} \qquad \theta_o(X_1, \dots, X_n) = \overline{X}_n + t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}} \tag{2.43}$$

liefern ein symmetrisches Konfidenzintervall zum Niveau $1-\alpha$.

Im Gegensatz zu (2.41) gilt hier für die Länge

$$L(X_1, \dots, X_n) = \theta_o(X_1, \dots, X_n) - \theta_u(X_1, \dots, X_n) = 2\, t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}}$$

nicht, dass sie vom Zufall unabhängig wäre. Deswegen haben die Realisierungen des Konfidenzintervalls unterschiedliche Längen.

Ein asymmetrisches Konfidenzintervall für μ zum Niveau $1-\alpha$ erhalten wir, wenn wir

$$\mathbb{P}\left(t_{n-1,\alpha_2} \le \frac{\overline{X}_n - \mu}{S_n}\sqrt{n} \le t_{n-1,1-\alpha_1}\right) = 1 - \alpha$$

für $\alpha_1 + \alpha_2 = \alpha$ betrachten. Die Stichprobenfunktionen sind dann

$$\theta_u(X_1, \dots, X_n) = \overline{X}_n - t_{n-1,1-\alpha_1}\frac{S_n}{\sqrt{n}} \qquad \text{und} \qquad \theta_o(X_1, \dots, X_n) = \overline{X}_n + t_{n-1,1-\alpha_2}\frac{S_n}{\sqrt{n}}. \tag{2.44}$$

Insbesondere ergibt sich daraus mit $\alpha_1 = 0$ das einseitige Kofidenzintervall $\big(-\infty, \theta_o(X_1, \dots, X_n)\big)$ zum Niveau $1-\alpha$ mit

$$\theta_o(X_1, \dots, X_n) = \overline{X}_n + t_{n-1,1-\alpha}\frac{S_n}{\sqrt{n}}.$$

Konfidenzintervall für die Varianz σ^2 bei bekanntem Erwartungswert μ

Sei $X = (X_1, \dots, X_n)$ die Zufallsstichprobe mit iid Stichprobenvariablen und sei $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ bekannt und σ^2 diesmal unbekannt sind.

Wir wollen σ^2 schätzen. Dazu verwenden wir die Zufallsvariable

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

mit der Eigenschaft

$$\frac{n}{\sigma^2} \overline{S}_n^2 \sim \chi_n^2$$

und können das asymmetrische Konfidenzintervall mit $\alpha_1 + \alpha_2 = \alpha \in (0,1)$ ansetzen mit

$$\mathbb{P}\left(\chi^2_{n,\alpha_2} \leq \frac{n}{\sigma^2} \overline{S}_n^2 \leq \chi^2_{n,1-\alpha_1} \right) = 1 - \alpha \tag{2.45}$$

und erhalten die Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) = \frac{n\overline{S}_n^2}{\chi^2_{n,1-\alpha_1}} \qquad \text{und} \qquad \theta_o(X_1, \dots, X_n) = \frac{n\overline{S}_n^2}{\chi^2_{n,\alpha_2}}. \tag{2.46}$$

Konfidenzintervall für die Varianz σ^2 bei unbekanntem Erwartungswert μ

Sei $X = (X_1, \dots, X_n)$ die Zufallsstichprobe mit iid Stichprobenvariablen und sei $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ und $\sigma^2 > 0$ unbekannt sind.

Wir wollen σ^2 schätzen. Wir wissen, dass

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1}$$

ist und deswegen können wir für das asymmetrische Konfidenzintervall mit $\alpha_1 + \alpha_2 = \alpha \in (0,1)$ die Gleichung

$$\mathbb{P}\left(\chi^2_{n-1,\alpha_2} \leq \frac{n-1}{\sigma^2} S_n^2 \leq \chi^2_{n-1,1-\alpha_1} \right) = 1 - \alpha \tag{2.47}$$

ansetzen, wobei χ^2_{n-1,α_2} das α_2-Quantil der χ^2-Verteilung mit $n-1$ Freiheitsgraden bezeichnet. Damit ergeben sich die Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) = \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha_1}} \qquad \text{und} \qquad \theta_o(X_1, \dots, X_n) = \frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha_2}}. \tag{2.48}$$

Wählen wir $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, so erhalten wir wieder ein symmetrisches Konfidenzintervall.

Zwei-Stichproben-Probleme

Wir können das Szenario verallgemeiern, wenn wir gleichzeitig zwei Stichproben $(X_{11}, \dots, X_{1n_1})$ und $(X_{21}, \dots, X_{2n_2})$ betrachten.

Annahmen Generell gibt es da zwei mögliche Fälle
(a) Die Stichproben sind untereinander unabhängig. Dann müssen die Längen n_1, n_2 nicht gleich sein.
(b) Die Stichproben hängen voneinander ab. Dann gilt $n_1 = n_2$ und wir definieren $X_i := (X_{1i}, X_{2i})$.
Es stellt sich die Frage, wie wir die Definition 2.35 für das neue Szenario erweitern können: Wir gehen davon aus, dass $X_{11}, \dots, X_{1n_1}$ unabhängig und identisch verteilt sind, das Gleiche gilt für $X_{21}, \dots, X_{2n_2}$.

Zu schätzen ist ein Funktionswert $g(\theta)$ des Parameters $\theta = (\theta_1, \dots, \theta_m)$, wobei $g : \Theta :\to \mathbb{R}$ eine eindimensionale messbare Abbildung des Parameterwertes ist.

> **Anmerkung:** Im letzten Abschnitt hatten wir den Fall betrachtet, bei dem g die Projektion auf eine Koordinate ist, also $g(\theta) = \theta_j$ für ein $1 \le j \le n$.

Dazu betrachten wir die Stichprobenfunktionen $\theta_u : \mathbb{R}^{n_1+n_2} \to \mathbb{R}$ und $\theta_o : \mathbb{R}^{n_1+n_2} \to \mathbb{R}$.

Definition 2.38: Sei $\alpha \in (0,1)$ beliebig, aber fest gewählt. Dann definieren die Stichprobenfunktionen θ_u und θ_o das zufällige Intervall $\big(\theta_u(X_1, \dots, X_n), \theta_o(X_1, \dots, X_n)\big)$. Dieses heißt ein *Konfidenzintervall zum Niveau* $(1-\alpha)$ für $g(\theta)$, falls

$$\mathbb{P}_\theta \left(\theta_u(X_1, \dots, X_n) \le \theta_o(X_1, \dots, X_n)\right) = 1, \tag{2.49}$$

$$\mathbb{P}_\theta \left(\theta_u(X_1, \dots, X_n) \le g(\theta) \le \theta_o(X_1, \dots, X_n)\right) \ge 1 - \alpha \tag{2.50}$$

für alle $\theta \in \Theta$ gilt.

Konfidenzintervall für die Differenz zweier Erwartungswerte bei bekannten Varianzen

Annahmen
(a) Seien n_1, n_2 beliebig, jedoch fest vorgegeben, d. h. wir betrachten die Zufallsstichproben $(X_{11}, \dots, X_{1n_1})$ und $(X_{21}, \dots, X_{2n_2})$.
(b) Seien $X_i = (X_{i1}, X_{i2})$ normalverteilte Zufallsvariablen und die Komponenten X_{i1} und X_{i2} unabhängig, jedoch nicht notwendigerweise identisch verteilt, d. h. $X_{11} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $X_{21} \sim \mathcal{N}(\mu_2, \sigma_2^2)$, wobei $\mu_1, \mu_2 \in \mathbb{R}$ unbekannt sind, $\sigma_1^2, \sigma_2^2 > 0$ bekannt.
(c) Daraus folgt insbesondere, dass $\overline{X}_{1n_1} := \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}$ und $\overline{X}_{2n_1} := \frac{1}{n_2}\sum_{i=1}^{n_1} X_{2i}$ unabhängig sind.
(d) Damit ist der unbekannte Parametervektor $\theta = (\mu_1, \mu_2)$ und $g(\theta) = \mu_1 - \mu_2$.
Wir wissen, dass

$$\overline{X}_{1n_1} \sim \mathcal{N}(\mu_1, \sigma_1^2/n_1) \qquad \text{und} \qquad \overline{X}_{2n_1} \sim \mathcal{N}(\mu_2, \sigma_2^2/n_2),$$

woraus wir mit der Unabhängigkeit auf

$$\overline{X}_{1n_1} - \overline{X}_{2n_1} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

schließen. Damit können wir für das Konfidenzintervall für jedes $\alpha \in (0,1)$ nach Normierung

$$\mathbb{P}\left(z_{\alpha/2} \le \frac{(\overline{X}_{1n_1} - \overline{X}_{2n_1}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \le z_{1-\alpha/2}\right) = 1 - \alpha \tag{2.51}$$

ansetzen. Nach Umformung haben wir

$$\mathbb{P}\left(\overline{X}_{1n_1}-\overline{X}_{2n_1}-z_{1-\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\le\mu_1-\mu_2\le\overline{X}_{1n_1}-\overline{X}_{2n_1}+z_{1-\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)=1-\alpha$$

bzw. die Stichprobenfunktionen

$$\theta_u(X_1,\dots,X_n):=\overline{X}_{1n_1}-\overline{X}_{2n_1}-z_{1-\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\qquad\text{und}\qquad\theta_o(X_1,\dots,X_n):=\overline{X}_{1n_1}-\overline{X}_{2n_1}+z_{1-\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\tag{2.52}$$

und damit ein Konfidenzintervall für $\mu_1-\mu_2$ zum Niveau $1-\alpha$.

Beispiel 2.39 (Lasermessung): Zwei Laser sollen auf gleich Wellenlänge überprüft werden. Siehe hierzu die Simulation und den Test der Laser im R-Dokument, welches sich auf der Website befindet.

Konfidenzintervall für den Quotienten zweier Varianzen bei unbekannten Erwartungswerten

Annahmen Wir gehen analog zu oben vor.
(a) Seien n_1, n_2 wieder beliebig, jedoch fest vorgegeben, d. h. wir betrachten die Zufallsstichproben $(X_{11},\dots,X_{1n_1})$ und $(X_{21},\dots,X_{2n_2})$.
(b) Seien $X_i=(X_{i1},X_{i2})$ normalverteilte Zufallsvariablen und die Komponenten X_{i1} und X_{i2} unabhängig, jedoch nicht notwendigerweise identisch verteilt, d. h. $X_{11}\sim\mathcal{N}(\mu_1,\sigma_1^2)$ und $X_{21}\sim\mathcal{N}(\mu_2,\sigma_2^2)$, wobei $\mu_1,\mu_2\in\mathbb{R}$ wie auch $\sigma_1^2,\sigma_2^2>0$ unbekannt sind.
(c) Daraus folgt in diesem Fall, dass $S_{1n_1}^2:=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1i}-\overline{X}_{1n_1})^2$ und $S_{2n_2}^2:=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(X_{2i}-\overline{X}_{2n_2})^2$ unabhängig sind.
(d) Damit ist der unbekannte Parametervektor $\theta=(\sigma_1^2,\sigma_2^2)$ und $g(\theta)=\sigma_1^2/\sigma_2^2$.
Auch hier können wir beim Bekannten ansetzen, und zwar ist

$$\frac{(n_1-1)S_{1n_1}^2}{\sigma_1^2}\sim\chi_{n_1-1}^2\qquad\text{und}\qquad\frac{(n_2-1)S_{2n_2}^2}{\sigma_2^2}\sim\chi_{n_2-1}^2$$

weswegen mit der Unabhängigkeit

$$\frac{S_{2n_2}^2/\sigma_2^2}{S_{1n_1}^2/\sigma_1^2}\sim F_{n_2-1,n_1-1}$$

gilt. Für jedes $\alpha\in(0,1)$ gilt dann

$$\mathbb{P}\left(F_{n_2-1,n_1-1,\alpha/2}\le\frac{S_{2n_2}^2/\sigma_2^2}{S_{1n_1}^2/\sigma_1^2}\le F_{n_2-1,n_1-1,1-\alpha/2}\right)=1-\alpha,$$

und nach Umformung

$$\mathbb{P}\left(\frac{S_{1n_1}^2}{S_{2n_2}^2}F_{n_2-1,n_1-1,\alpha/2}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{S_{1n_1}^2}{S_{2n_2}^2}F_{n_2-1,n_1-1,1-\alpha/2}\right)=1-\alpha,$$

woraus wir die Stichprobenfunktionen

$$\theta_u(X_1,\dots,X_n):=\frac{S_{1n_1}^2}{S_{2n_2}^2}F_{n_2-1,n_1-1,\alpha/2}\qquad\text{und}\qquad\theta_o(X_1,\dots,X_n):=\frac{S_{1n_1}^2}{S_{2n_2}^2}F_{n_2-1,n_1-1,1-\alpha/2}\tag{2.53}$$

ablesen und damit ein Konfidenzintervall für σ_1^2/σ_2^2 zum Niveau $1-\alpha$ erhalten.

Konfidenzintervall für die Differenz der Erwartungswerte bei verbundenen Stichproben

Problemstellung Hier betrachten wir den Fall, bei dem die Zufallsstichproben $X_1 = (X_{11}, \ldots, X_{1n_1})$ und $X_2 = (X_{21}, \ldots, X_{2n_2})$ nicht voneinander unabhängig sind. Das ist immer dann der Fall, wenn wir eine Vorher-Nachher-Observation haben. Allgemeiner immer dann, wenn X_{1i} und X_{2i} auch stochastisch abhängig sind.

Annahmen
(a) Im Gegensatz zu den vorherigen Beispielen ist hier $n_1 = n_2$.
(b) Sei n beliebig, jedoch fest vorgegeben, d. h. wir betrachten die Zufallsstichproben $(X_{11}, \ldots, X_{1n})$ und $(X_{21}, \ldots, X_{2n})$.
(c) Die Komponenten X_{i1} und X_{i2} müssen nicht unabhängig sein.
(d) Sei $\mathbb{E}X_{11} = \mu_1$ und $\mathbb{E}X_{21} = \mu_2$, wobei $\mu_1, \mu_2 \in \mathbb{R}$ unbekannt sind.
(e) Des weiteren seien die Differenzen $Y_i := X_{1i} - X_{2i}$ unabhängig und identisch normalverteilt mit $Y_1 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma^2)$ mit $\sigma^2 > 0$ (unbekannt).
Mit den Eigenschaften des Satzes 2.34 für die Zufallsstichprobe $(Y_1, \ldots, Y_n)$ wissen wir, dass

$$\overline{Y}_n \sim \mathcal{N}\left(\mu_1 - \mu_2, \sigma^2/n\right) \qquad \text{und} \qquad \frac{n-1}{\sigma^2} S^2_{Y,n} \sim \chi^2_{n-1} \qquad \text{und} \qquad \frac{\sqrt{n}\left(\overline{Y}_n - (\mu_1 - \mu_2)\right)}{S_{Y,n}} \sim t_{n-1},$$

und können für das Konfidenzintervall zum Niveau $\alpha \in (0,1)$ die Gleichung

$$\mathbb{P}\left(t_{n-1,\alpha/2} \leq \frac{\sqrt{n}\left(\overline{Y}_n - (\mu_1 - \mu_2)\right)}{S_{Y,n}} \leq t_{n-1,1-\alpha/2} \right) = 1 - \alpha$$

ansetzen, die umgeformt

$$\mathbb{P}\left(\overline{Y}_n - t_{n-1,1-\alpha/2} \frac{S_{Y,n}}{\sqrt{n}} \leq \mu_1 - \mu_2 \leq \overline{Y}_n + t_{n-1,1-\alpha/2} \frac{S_{Y,n}}{\sqrt{n}} \right) = 1 - \alpha$$

ist und die Stichprobenfunktionen

$$\theta_u(Y_1, \ldots, Y_n) = \overline{Y}_n - t_{n-1,1-\alpha/2} \frac{S_{Y,n}}{\sqrt{n}} \qquad \text{und} \qquad \theta_o(Y_1, \ldots, Y_n) = \overline{Y}_n + t_{n-1,1-\alpha/2} \frac{S_{Y,n}}{\sqrt{n}} \tag{2.54}$$

liefert.

Beispiel 2.40: Wir wollen feststellen, ob sich der Luftdruck mit der Höhe ändert. Dazu müssen wir mehrmals Messungen durchführen. Wir messen einen Monat lang jeden Tag am Münsterplatz in Ulm wie auch auf der Besucherplattform des Münsterturms. Da sich das Wetter in dieser Zeit ändert, was wir nicht beeinflussen können, und dies den Luftdruck verändert, haben wir es mit verbundenen Stichproben zu tun.

Anmerkung: Das Szenario der verbundenen Stichprobe wird immer dann eingesetzt, wenn die Grundbedingungen nicht beeinflussbar sind und wir somit für die Zufallsstichprobe keine identische Verteilung annehmen können. Die Idee ist, eine kontrollierte Veränderung eines anderen Aspektes zu erzeugen, um einen Effekt festzustellen.

2.3.2 Asymptotische Konfidenzintervalle

Wir gehen von einer Zufallsstichprobe aus und zeigen, wie mittels des zentralen Grenzwertsatzes und des starken Gesetzes der großen Zahlen asymptotische Konfidenzintervalle aufgestellt werden können.

Konfidenzintervall für den Parameter η bei der Bernoulli-Verteilung

Seien $X_1, \ldots, X_n \sim \mathrm{Bin}(1, \eta)$, wobei $\eta \in (0,1)$ unbekannt und zu schätzen ist. Wir wissen, dass $\mathbb{E}X_1 = \eta$ und $\mathbb{V}\mathrm{ar}X_1 = \eta(1-\eta)$. Mit dem zentralen Grenzwertsatz folgt daraus für alle $\alpha \in (0,1)$, dass

$$\lim_{n\to\infty} \mathbb{P}\left(z_{\alpha/2} \le \sqrt{n}\frac{\overline{X}_n - \eta}{\sqrt{\eta(1-\eta)}} \le z_{1-\alpha/2}\right) = 1-\alpha.$$

Wir betrachten die doppelte Ungleichung genauer und erhalten mit $z := z_{1-\alpha/2}$ die äquivalente Bedingung

$$\begin{aligned}
\left|\frac{\overline{X}_n - \eta}{\sqrt{\eta(1-\eta)}}\right| &\le \frac{z}{\sqrt{n}} \\
\left(\overline{X}_n - \eta\right)^2 &\le \frac{z^2}{n}\left(\eta(1-\eta)\right) \\
\eta^2 - 2\eta\,\overline{X}_n + \overline{X}_n^2 &\le \frac{z^2}{n}\eta - \frac{z^2}{n}\eta^2 \\
\eta^2\left(1+\frac{z^2}{n}\right) - \eta\left(2\overline{X}_n + \frac{z^2}{n}\right) + \overline{X}_n^2 &\le 0 \\
\left(\frac{n+z^2}{n}\right)\left[\eta^2 - 2\eta\frac{\overline{X}_n + \frac{z^2}{2n}}{\frac{n+z^2}{n}} + \frac{n\overline{X}_n^2}{n+z^2}\right] &\le 0 \\
\eta^2 - 2\eta\frac{n\overline{X}_n + z^2/2}{n+z^2} + \frac{n\overline{X}_n^2}{n+z^2} &\le 0 \\
\left(\eta - \frac{n\overline{X}_n + z^2/2}{n+z^2}\right)^2 &\le \left(\frac{n\overline{X}_n + z^2/2}{n+z^2}\right)^2 - \frac{n\overline{X}_n^2}{n+z^2} \\
\left(\eta - \frac{n\overline{X}_n + z^2/2}{n+z^2}\right)^2 &\le \frac{n^2\overline{X}_n^2 + z^4/4 + n\overline{X}_n z^2}{(n+z^2)^2} - \frac{n^2\overline{X}_n^2 + n\overline{X}_n^2 z^2}{(n+z^2)^2} \\
\left|\eta - \frac{n\overline{X}_n + z^2/2}{n+z^2}\right| &\le \frac{z\sqrt{z^2/4 + n\overline{X}_n(1-\overline{X}_n)}}{n+z^2},
\end{aligned}$$

aus welcher die Stichprobenfunktionen

$$\theta_u(X_1,\ldots,X_n) := \frac{1}{n+z_{1-\alpha/2}^2}\left(n\overline{X}_n + \frac{z_{1-\alpha/2}^2}{2} - z_{1-\alpha/2}\sqrt{\frac{z_{1-\alpha/2}^2}{4} + n\overline{X}_n(1-\overline{X}_n)}\right) \tag{2.55}$$

$$\theta_u(X_1,\ldots,X_n) := \frac{1}{n+z_{1-\alpha/2}^2}\left(n\overline{X}_n + \frac{z_{1-\alpha/2}^2}{2} + z_{1-\alpha/2}\sqrt{\frac{z_{1-\alpha/2}^2}{4} + n\overline{X}_n(1-\overline{X}_n)}\right) \tag{2.56}$$

folgen, die ein asymptotisches Konfidenzintervall $(\theta_u(X_1,\dots,X_n),\theta_o(X_1,\dots,X_n))$ für η zum Niveau $1-\alpha$ ergeben.

Konfidenzintervall für den Parameter λ bei der Poisson-Verteilung

Seien $X_1,\dots,X_n$ die Zufallsstichprobe mit $X_1 \sim \mathrm{Poi}(\lambda)$ mit $\lambda > 0$ unbekannt. Wir wissen, dass $\mathbb{E}X_1 = \lambda$ und $\mathbb{V}\mathrm{ar}X_1 = \lambda$ gilt. Dann folgt mit dem zentralen Grenzwertsatz analog zum vorherigen Abschnitt für alle $\alpha \in (0,1)$, dass

$$\lim_{n\to\infty} \mathbb{P}\left(z_{\alpha/2} \le \sqrt{n}\frac{\overline{X}_n - \lambda}{\sqrt{\lambda}} \le z_{1-\alpha/2} \right) = 1-\alpha$$

und nach ähnlichen Umformungen die Stichprobenfunktionen

$$\begin{aligned} \theta_u(X_1,\dots,X_n) &:= \overline{X}_n + \frac{z_{1-\alpha/2}}{2n} - z_{1-\alpha/2}\sqrt{\frac{z^2_{1-\alpha/2}}{4n^2} + \frac{\overline{X}_n}{n}} \\ \theta_o(X_1,\dots,X_n) &:= \overline{X}_n + \frac{z_{1-\alpha/2}}{2n} + z_{1-\alpha/2}\sqrt{\frac{z^2_{1-\alpha/2}}{4n^2} + \frac{\overline{X}_n}{n}} \end{aligned}$$

für ein asymptotisches Konfidenzintervall $(\theta_u(X_1,\dots,X_n),\theta_o(X_1,\dots,X_n))$ für λ zum Niveau $1-\alpha$.

Anmerkung: Aus dem starken Gesetz der großen Zahlen wissen wir, dass $\overline{X}_n \overset{\text{f.s.}}{\to} \lambda$. Damit können wir ein weiteres asymptotisches Konfidenzintervall ansetzen durch

$$\lim_{n\to\infty} \mathbb{P}\left(z_{\alpha/2} \le \sqrt{n}\frac{\overline{X}_n - \lambda}{\sqrt{\overline{X}_n}} \le z_{1-\alpha/2} \right) = 1-\alpha \tag{2.57}$$

und erhalten die Stichprobenfunktionen

$$\begin{aligned} \theta_u(X_1,\dots,X_n) &= X_n - \frac{z^2_{1-\alpha/2}}{\sqrt{n}}\sqrt{X_n}, \\ \theta_o(X_1,\dots,X_n) &= X_n + \frac{z^2_{1-\alpha/2}}{\sqrt{n}}\sqrt{X_n}. \end{aligned}$$

Die Einfachheit der Terme ist dadurch erkauft, dass $\theta_u(X_1,\dots,X_n)$ auch negativ sein kann, was eine sehr großzügige untere Schranke für $\lambda > 0$ darstellt.

Konfidenzintervall für die Differenz zweier Erwartungswerte bei der Poisson-Verteilung

Hier betrachten wir ein asymptotisches Zwei-Stichproben-Szenario. Details siehe Übungsblatt xxx Aufgabe yyy.

3 Statistische Tests

3.1 Prinzipien des Testens

Ein statistischer Test ist ein Verfahren, mit dem wir an Hand von Stichproben Hypothesen über Eigenschaften der unbekannten Verteilungsfunktion F der Stichprobenvariablen $X_1, \dots, X_n$ prüfen.

Vorgehen
(a) Wir konstruieren eine Entscheidungsregel, nach der eine Hypothese verworfen oder im ursprünglichen Wortsinn (weiterhin) angenommen, also beibehalten, wird.
(b) Die Entscheidungsregel wird so gewählt, dass die Wahrscheinlichkeiten möglicher Fehlentscheidungen minimiert werden.
(c) Die Entscheidung treffen wir dann abhängig von den beobachteten Daten, d. h. der Realisierung der Zufallsstichprobe.
(d) Analog zu den Konfidenzintervallen geben wir uns ein Signifikanzniveau α vor.

Beispiel 3.1: Wir haben zwei äußerlich nicht unterscheidbare Münzen, deren Wahrscheinlichkeit η Kopf zu werfen bei $0,4$ bzw. $0,5$ liegt. Wir werfen eine der beiden Münzen 1000 mal zur Bestimmung von η. Dazu stellen wir die Hypothese $\eta = 0.5$ auf. Wir wollen diese ablehnen, falls

$$Z = \{\text{Anzahl von Würfen mit dem Ausgang Kopf}\} < 450$$

und behalten sie weiter bei, falls $Z \geq 450$.

Falls $Z \sim \text{Bin}(1000, 0,5)$, dann ist $\mathbb{P}_{0,5}(Z < 450) = \mathbb{P}_{0,5}(Z \leq 449) = 7,0 \cdot 10^{-4}$; und falls $Z \sim \text{Bin}(1000, 0,4)$, dann ist $\mathbb{P}_{0,4}(Z \geq 450) = 1 - \mathbb{P}_{0,4}(Z < 450) = 1 - \mathbb{P}_{0,5}(Z \leq 449) = 7,4 \cdot 10^{-4}$.

Wir haben damit die Fehler der Fehlentscheidung. Diese sind nicht gleich, da unter anderem die Varianzen unterschiedlich sind, 250 resp. 240.

Es ist möglich, einen Fehler zu reduzieren. Sagen wir, die Hypothese wird beibehalten, falls $Z < 430$ und sonst verworfen. Dann ist $\mathbb{P}_{0,5}(Z < 430) = 4,0 \cdot 10^{-6}$ und $\mathbb{P}_{0,5}(Z \geq 430) = 0,029$.

3.2 Aufbau eines statistischen Tests

Sei $\mathcal{P}$ die Familie der möglichen Verteilungsfunktionen F der Stichprobenvariablen $X_1, \dots, X_n$, die wir als unabhängig und identisch verteilt annehmen.

Hypothese Wir bilden eine Zweier-Partition von $\mathcal{P}$, also $\mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1, \mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$ und betrachten die Hypothesen

$$H_0 : F \in \mathcal{P}_0$$
$$H_1 : F \in \mathcal{P}_1$$

dafür dass die (unbekannte) Verteilungsfunktion F in $\mathcal{P}_0$ bzw. $\mathcal{P}_1$ liegt. Wir nennen H_0 auch die *Nullhypothese* und H_1 die *Alternativhypothese*.

Die Hypothesen heißen *einfach*, falls $|\mathcal{P}_0| = 1$ bzw. $|\mathcal{P}_1| = 1$, ansonsten *zusammengesetzt* (Im Beispiel 3.1 waren sowohl die Null- als auch die Alternativhypothese einfach.)

Entscheidungsregel Der Stichprobenraum wird deshalb ebenfalls in zwei (messbare) Mengen zerlegt, K und $K^c \subset \mathbb{R}^n$. Dabei wird K der *kritische Bereich* genannt und die Nullhypothese wird verworfen, falls $(x_1, \ldots, x_n) \in K$.

Dazu definieren wir uns eine Stichprobenfunktion $\varphi : \mathcal{X} \to \{0, 1\}$ durch

$$\varphi(x_1, \ldots, x_n) = \mathbb{1}_K(x_1, \ldots, x_n) \tag{3.1}$$

und sagen, dass H_0 verworfen wird, falls $\varphi(x_1, \ldots, x_n) = 1$ ist; falls die Stichprobe also in der Menge K landet.

Definition 3.2: Sei $\alpha \in (0, 1)$ fest und sei $\mathcal{P}$ eine Verteilungsfamilie. Dann ist die Funktion φ aus 3.1 ein *Test zum Niveau* α, falls

$$\mathbb{P}_F \left(\varphi(X_1, \ldots, X_n) = 1\right) \leq \alpha$$

für alle $F \in \mathcal{P}_0$. Gilt zusätzlich

$$\mathbb{P}_F \left(\varphi(X_1, \ldots, X_n) = 1\right) \geq \alpha$$

für alle $F \in \mathcal{P}_1$, so heißt φ *unverfälschter Test zum Niveau* α.

Anmerkung: (a) Der *Fehler erster Art* ist es, die Nullhypothese abzulehnen, obwohl sie zutrifft, und seine Wahrscheinlichkeit ist

$$\alpha_n(F) = \mathbb{P}_F \left((X_1, \ldots, X_n) \in K\right) = \mathbb{P}_F \left(\varphi(X_1, \ldots, X_n) = 1\right) \tag{3.2}$$

für alle $F \in \mathcal{P}_0$.
(b) Der *Fehler zweiter Art* ist es, die Nullhypothese nicht abzulehnen, obwohl sie nicht zutrifft, und seine Wahrscheinlichkeit ist

$$\beta_n(F) = \mathbb{P}_F \left((X_1, \ldots, X_n) \notin K\right) = \mathbb{P}_F \left(\varphi(X_1, \ldots, X_n) = 0\right) \tag{3.3}$$

für alle $F \in \mathcal{P}_1$.
(c) Wir interessieren uns für Tests, für die zu einem vorgegebenen α die Wahrscheinlichkeit $\beta_n(F)$ für den Fehler zweiter Art möglichst klein ist.

Parametrische Verteilungsfamilie Ist die Verteilungsfamilie $\mathcal{P}$ parametrisch, so können wir die Struktur und Entscheidung des Tests über die Parameter der zugehörigen Verteilungen angeben. Sei $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ gegeben, so definieren wir zu $\mathcal{P}_0$ und $\mathcal{P}_1$ eine Partition $\{\Theta_0, \Theta_1\}$ von Θ so dass

$$\mathcal{P}_0 = \{P_\theta \in \mathcal{P} : \theta \in \Theta_0\} \qquad \text{und} \qquad \mathcal{P}_1 = \{P_\theta \in \mathcal{P} : \theta \in \Theta_1\} \tag{3.4}$$

gilt und formulieren die Hypothesen

$$H_0 : \theta \in \Theta_0,$$
$$H_1 : \theta \in \Theta_1.$$

Beispiel 3.3: Sei $\mathcal{P} = \left\{ \mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R} \right\}$ für ein vorgegebenes σ^2. Betrachte die Hypothesen

$$H_0 : \mu = 0$$
$$H_1 : \mu \neq 0$$

wobei H_0 einfach und H_1 zusammengesetzt sind.

Definition 3.4: Sei $\alpha \in (0,1)$ beliebig aber fest und sei $\mathcal{P}$ eine parametrische Verteilungsfamilie. Dann ist die Funktion φ aus 3.1 ein *Parametertest zum Niveau* α, falls

$$\mathbb{P}_\theta \left(\varphi(X_1, \ldots, X_n) = 1 \right) \leq \alpha$$

für jedes $\theta \in \Theta_0$. Gilt zusätzlich

$$\mathbb{P}_\theta \left(\varphi(X_1, \ldots, X_n) = 1 \right) \geq \alpha$$

für alle $\theta \in \Theta_1$, so heißt φ *unverfälschter Parametertest zum Niveau* α.

Bei parameterischen Verteilungsfamilien können wir zusätzlich definieren:

Definition 3.5: (i) Die Funktion

$$\alpha_n : \Theta \to [0,1]$$
$$\alpha_n(\theta) = \mathbb{P}_\theta \left(\varphi(X_1, \ldots, X_n) = 1 \right)$$

heißt *Gütefunktion* des Parametertests φ.

(ii) Die Funktion $\alpha_n : \Theta_1 \to [0,1]$ (eingeschränkt auf Θ_1) heißt *Macht* des Parametertests φ.

(iii) Der Test zum Niveau α heißt *konsistent*, falls $\lim_{n\to\infty} \alpha_n(\theta) = 1$ für jedes $\theta \in \Theta_1$.

(iv) Seien φ_1 und φ_2 zwei Parametertests zum Niveau α. Falls die Macht von φ_1 größer ist als die Macht von φ_2 für alle $\theta \in \Theta_1$, d. h.

$$\mathbb{P}_\theta \left(\varphi_1(X_1, \ldots, X_n) = 1 \right) \geq \mathbb{P}_\theta \left(\varphi_2(X_1, \ldots, X_n) = 1 \right) \qquad \text{für alle } \theta \in \Theta_1,$$

so sagen wir, dass φ_1 *besser* ist als φ_2.

Anmerkung: Ein Test ist unverfälscht, wenn für die Gütefunktion $\alpha_n(\theta)$ für alle $\theta \in \Theta_1$ die Ungleichung $\alpha_n(\theta) \geq \alpha$ gilt.

Vorgehen zur Bestimmung des kritischen Bereichs K eines Parametertests zum Niveau α.

(a) Definiere die Teststatistik $T : \mathcal{X} \to \mathbb{R}$ so dass

$$T(X_1, \ldots, X_n | \theta) \sim F$$

für alle $\theta \in \Theta_0$ und eine Verteilung F, die nicht von θ abhängt.

(b) Bestimme $c > 0$ so, dass

$$\mathbb{P}_\theta \left(\left| T(X_1, \ldots, X_n) \right| > c \right) < \alpha$$

für alle $\theta \in \Theta_0$ gilt.

(c) Damit ist die Menge

$$K = \Big\{ (x_1, \ldots, x_n) : \big| T(x_1, \ldots, x_n) \big| > c \Big\}$$

der kritische Bereich eines Parametertests zum Niveau α.
Manchmal ist ein asymmetrisch definierter Bereich sinnvoller, da er die Macht des Tests erhöhen kann. Ersetze dann (b) und (c) durch
(b′) Bestimme $c_1, c_2 \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ mit $c_1 < c_2$ so, dass

$$\mathbb{P}_\theta \left(c_1 \leq T(X_1, \ldots, X_n) \leq c_2 \right) \geq 1 - \alpha$$

für alle $\theta \in \Theta_0$ gilt.
(c′) Dann ist die Menge

$$K = \mathbb{R}^n \setminus \left\{ (x_1, \ldots, x_n) : c_1 \leq T(x_1, \ldots, x_n) \leq c_2 \right\}$$

der kritische Bereich des asymmetrischen Parametertests zum Niveau α.

Anmerkung (Statistische Tests und Konfidenzintervalle): Die Komplementärmenge K^c des kritischen Bereichs

$$K = \Big\{ (x_1, \ldots, x_n) : \big| T(x_1, \ldots, x_n) \big| \leq c \Big\}$$

bzw.

$$K = \left\{ (x_1, \ldots, x_n) : c_1 \leq T(x_1, \ldots, x_n) \leq c_2 \right\}$$

führte zur Konstruktion des Konfidenzintervalls. Während wir dabei ein zufälliges Intervall für eine feste Größe, den Parameterwert $g(\theta)$ erstellen, betrachten wir bei den Tests ein festes Intervall (denn wir haben eine Hypothese) für eine zufällige Größe.

3.3 Parametertests bei Normalverteilung

Wir nehmen im ganzen Abschnitt an, dass die Stichprobenvariablen $X_1, \ldots, X_n$ i.i.d. und normalverteilt sind, d. h. $\mathcal{P} = \Big\{ \mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0 \Big\}$.

Test des Erwartungswertes μ (bei bekannter Varianz σ^2)

Annahme Gelte $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$ unbekannt, $\sigma^2 > 0$ bekannt und sei $X = (X_1, \ldots, X_n)$ die Stichprobe.

Wir geben uns ein $\mu_0 \in \mathbb{R}$ vor, auf dem die Hypothese beruht.

Zweiseitiger Test Die Hypothesen sind

$$\begin{aligned} &H_0 : \mu = \mu_0 \qquad && \Theta_0 = \{\mu_0\} \\ &H_1 : \mu \neq \mu_0 \qquad && \Theta_1 = \mathbb{R} \setminus \{\mu_0\} \end{aligned}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \to (0, \infty)$ mit

$$T(x_1, \ldots, x_n) = \sqrt{n}\frac{\overline{x}_n - \mu_0}{\sigma} \qquad \text{und es gilt} \qquad T(X) = \sqrt{n}\frac{\overline{X}_n - \mu_0}{\sigma} \sim \mathcal{N}(0,1) \tag{3.5}$$

unter der Nullhypothese, wobei $c = z_{1-\alpha/2}$ ist und die Schwellenwerte $-c_1 = c_2 = c$ gewählt werden. Dann gilt für alle $\mu \in \Theta_0$, also für μ_0, dass

$$\mathbb{P}_{\mu_0}\left(c_1 \le T(X) \le c_2\right) = 1 - \mathbb{P}_{\mu_0}\left(\left|T(X)\right| > z_{1-\alpha/2}\right) = 1 - \alpha. \tag{3.6}$$

Der kritische Bereich ist dann die Menge

$$\begin{aligned} K &= \left\{(x_1, \ldots, x_n) : \left|T(x_1, \ldots, x_n)\right| > c\right\} \\ &= \mathbb{R}^n \setminus \left\{(x_1, \ldots, x_n) : \left|T(x_1, \ldots, x_n)\right| \le z_{1-\alpha/2}\right\} \\ &= \mathbb{R}^n \setminus \left\{(x_1, \ldots, x_n) : -z_{1-\alpha/2} \le \frac{\sqrt{n}}{\sigma}(\overline{x}_n - \mu_0) \le z_{1-\alpha/2}\right\} \\ &= \mathbb{R}^n \setminus \left\{(x_1, \ldots, x_n) : \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \le \overline{x}_n \le \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right\}, \end{aligned}$$

wobei die Menge K^c ebenfalls die Bedingung des symmetrischen Konfidenzintervalls für μ bei bekanntem σ^2 zum Niveau $1 - \alpha$ bei Normalverteilung darstellt.

Beispiel 3.6: Wir erinnern uns an die Papiermaschine aus Beispiel 2.37. Zu $\alpha = 0{,}01$ wollen wir die Hypothesen

$$\begin{aligned} &H_0 : \mu = \mu_0 = 182 && \Theta_0 = \{182\} \\ &H_1 : \mu \neq 182 && \Theta_1 = \mathbb{R} \setminus \{182\} \end{aligned}$$

testen. Dann ist $\left|T(x_1, \ldots, x_9)\right| = 3{,}5 > 2{,}576 = z_{0{,}995}$. Diese Nullhypothese wird abgelehnt.

Anmerkung (Gütefunktion): Wir untersuchen die Gütefunktion $\alpha_n : \to [0,1]$ dieses Tests

$$\begin{aligned} \alpha_n(\mu) &= \mathbb{P}_\mu\left(\left|T(X_1, \ldots, X_n)\right| > z_{1-\alpha/2}\right) \\ &= 1 - \mathbb{P}_\mu\left(-z_{1-\alpha/2} \le T(X_1, \ldots, X_n) \le z_{1-\alpha/2}\right) \\ &= 1 - \left(\Phi\left(z_{1-\alpha/2} - \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right) - \Phi\left(-z_{1-\alpha/2} - \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right)\right) \end{aligned}$$

Hieraus folgt

$$\begin{aligned} \alpha_n(\mu_0) &= \alpha \\ \alpha_n(\mu) &> \alpha \qquad \text{für } \mu \neq \mu_0 \end{aligned}$$

und

$$\begin{aligned} &\mu < \mu_0 : && \lim_{n\to\infty} \alpha_n(\mu) = 1 \\ &\mu > \mu_0 : && \lim_{n\to\infty} \alpha_n(\mu) = 1 \end{aligned}$$

Der Test ist deshalb unverfälscht und konsistent.

Einseitiger Test Die Hypothesen sind

$$\begin{aligned} &H_0 : \mu = \mu_0 && \Theta_0 = \{\mu_0\} \\ &H_1 : \mu > \mu_0 && \Theta_1 = \{\mu \in \mathbb{R} : \mu > \mu_0\} \end{aligned}$$

Mit der gleichen Teststatistik haben wir zwei Möglichkeiten zur Bestimmung des kritischen Bereichs, entweder

$$K = \Big\{(x_1, \dots, x_n) : \big|T(x_1, \dots, x_n)\big| > z_{1-\alpha/2}\Big\}$$

wie bisher oder

$$\widetilde{K} = \big\{(x_1, \dots, x_n) : T(x_1, \dots, x_n) > z_{1-\alpha}\big\}$$

als einseitigen kritischen Bereich.

Die zugehörigen Gütefunktionen α_n und $\tilde{\alpha}_n$ sind

$$\begin{aligned} \alpha_n &= 1 - \Phi\left(\sqrt{n}\frac{\mu_o - \mu}{\sigma} + z_{1-\alpha/2}\right) + \Phi\left(\sqrt{n}\frac{\mu_o - \mu}{\sigma} - z_{1-\alpha/2}\right) \\ \tilde{\alpha}_n &= 1 - \Phi\left(\sqrt{n}\frac{\mu_o - \mu}{\sigma} + z_{1-\alpha}\right) \end{aligned}$$

und es gilt $\alpha_n(\mu) < \tilde{\alpha}_n(\mu)$ für alle $\mu \in \Theta_1$. Deswegen bevorzugen wir den einseitigen kritischen Bereich $\widetilde{K}$.

Test des Erwartungswertes μ (bei unbekannter Varianz σ^2)

Annahme Gelte $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$, $\sigma^2 > 0$ unbekannt und sei $X = (X_1, \dots, X_n)$ die Stichprobe.

Wir geben uns ein $\mu_0 \in \mathbb{R}$ vor, auf dem die Hypothese beruht.

Zweiseitiger Test Die Hypothesen sind

$$\begin{aligned} &H_0 : \mu = \mu_0 && \Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\} \\ &H_1 : \mu \neq \mu_0 && \Theta_1 = \{(\mu, \sigma^2) : \mu \neq \mu_0, \sigma^2 > 0\} \end{aligned}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \to (0, \infty)$ mit

$$T(x_1, \dots, x_n) = \sqrt{n}\frac{\overline{x}_n - \mu_0}{s_n} \qquad \text{und es gilt} \qquad T(X) = \sqrt{n}\frac{\overline{X}_n - \mu_0}{S_n} \sim t_{n-1} \tag{3.7}$$

unter der Nullhypothese. Seien $c = t_{n-1,1-\alpha/2}$ und die Schwellenwerte $-c_1 = c_2 = c$. Dann gilt für alle $(\mu, \sigma^2) \in \Theta_0$, dass

$$\mathbb{P}_{(\mu_0, \sigma^2)}\big(c_1 \leq T(X) \leq c_2\big) = 1 - \mathbb{P}_{(\mu_0, \sigma^2)}\Big(\big|T(X)\big| > t_{n-1,1-\alpha/2}\Big) = 1 - \alpha. \tag{3.8}$$

Der kritische Bereich ist dann die Menge

$$\begin{aligned} K &= \Big\{(x_1,\dots,x_n) : \big|T(x_1,\dots,x_n)\big| > c\Big\} \\ &= \mathbb{R}^n \setminus \Big\{(x_1,\dots,x_n) : \big|T(x_1,\dots,x_n)\big| \le t_{n-1,1-\alpha/2}\Big\} \\ &= \mathbb{R}^n \setminus \left\{(x_1,\dots,x_n) : -t_{n-1,1-\alpha/2} \le \frac{\sqrt{n}}{s_n}(\overline{x}_n - \mu_0) \le t_{n-1,1-\alpha/2}\right\} \\ &= \mathbb{R}^n \setminus \left\{(x_1,\dots,x_n) : \mu_0 - \frac{s_n}{\sqrt{n}} t_{n-1,1-\alpha/2} \le \overline{x}_n \le \mu_0 + \frac{s_n}{\sqrt{n}} t_{n-1,1-\alpha/2}\right\}. \end{aligned}$$

Beispiel 3.7: Wir beschäftigen uns immer noch mit der Papiermaschine. Sei $\alpha = 0{,}05$ gewählt und die Hypothese

$$\begin{aligned} H_0 &: \mu = \mu_0 = 182 \\ H_1 &: \mu \neq \mu_0 \end{aligned}$$

Dann ist $\big|T(x_1,\dots,x_9)\big| = 6{,}396 > 2{,}306 = t_{8,0{,}975}$. Die Null-Hypothese wird abgelehnt.

Anmerkung: Die Entscheidungen in den Tests des Erwartungswertes bei bekanntem σ^2 bzw. bei unbekanntem σ^2 können sich unterscheiden.

Einseitiger Test Die Hypothesen sind

$$\begin{aligned} H_0 &: \mu = \mu_0 & \Theta_0 &= \{(\mu,\sigma^2) : \mu = \mu_0, \sigma^2 > 0\} \\ H_1 &: \mu > \mu_0 & \Theta_1 &= \{(\mu,\sigma^2) : \mu > \mu_0, \sigma^2 > 0\} \end{aligned}$$

mit gleicher Teststatistik, dem Schwellenwert $c = t_{n-1,1-\alpha}$ und dem resultierenden kritischen Bereich $K = \big\{(x_1,\dots,x_n) : T(x_1,\dots,x_n) > t_{n-1,1-\alpha}\big\}$.

Test der Varianz σ^2 (bei bekanntem Erwartungswert μ)

Annahme Gelte $X_1,\dots,X_n \sim \mathcal{N}(\mu,\sigma^2)$ mit $\mu \in \mathbb{R}$ bekannt, $\sigma^2 > 0$ unbekannt und sei $X = (X_1,\dots,X_n)$ die Stichprobe.

Wir geben uns ein $\sigma_0 > 0$ vor, auf dem die Hypothese beruht.

Zweiseitiger Test Die Hypothesen sind

$$\begin{aligned} H_0 &: \sigma^2 = \sigma_0^2 & \Theta_0 &= \{\sigma_0^2\} \\ H_1 &: \sigma^2 \neq \sigma_0^2 & \Theta_1 &= (0,\infty) \setminus \{\sigma_0^2\} \end{aligned}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \to (0,\infty)$ mit

$$T(x_1,\dots,x_n) = \frac{n}{\sigma_0^2}\tilde{s}_n^2 \qquad \text{und es gilt} \qquad T(X) = \frac{n}{\sigma_0^2}\tilde{S}_n^2 \sim \chi_n^2 \tag{3.9}$$

unter der Nullhypothese, wobei

$$\tilde{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \qquad \text{und} \qquad \tilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \tag{3.10}$$

ist, mit den Schwellenwerten $c_1 = \chi^2_{n,\alpha/2}$ und $c_2 = \chi^2_{n,1-\alpha/2}$.

Es gilt für alle $\sigma^2 \in \Theta_0$, also für σ_0^2, dass

$$\mathbb{P}_{\sigma_0^2}\left(c_1 \le T(X) \le c_2\right) = 1-\alpha, \tag{3.11}$$

und der kritische Bereich ist

$$\begin{aligned} K &= \mathbb{R}^n \setminus \left\{(x_1,\dots,x_n) : c_1 \le T(x_1,\dots,x_n) \le c_2\right\} \\ &= \mathbb{R}^n \setminus \left\{(x_1,\dots,x_n) : \chi^2_{n,\alpha/2} \le \frac{n}{\sigma_0^2}\tilde{s}_n^2 \le \chi^2_{n,1-\alpha/2}\right\} \end{aligned}$$

Einseitiger Test Die Hypothesen sind

$$\begin{aligned} &H_0 : \sigma^2 = \sigma_0^2 && \Theta_0 = \{\sigma_0^2\} \\ &H_1 : \sigma^2 > \sigma_0^2 && \Theta_1 = \{\sigma^2 : \sigma^2 > \sigma_0^2\} \end{aligned}$$

Mit der gleichen Teststatikstik und den Schwellenwerten $c_1 = -\infty$ und $c := c_2 = \chi^2_{n,1-\alpha}$ gilt (3.11) analog und wir haben den kritischen Bereich

$$\begin{aligned} \widetilde{K} &= \mathbb{R}^n \setminus \left\{(x_1,\dots,x_n) : T(x_1,\dots,x_n) \le c\right\} \\ &= \left\{(x_1,\dots,x_n) : \frac{n}{\sigma_0^2}\tilde{s}_n^2 > \chi^2_{n,1-\alpha}\right\}. \end{aligned}$$

Die zugehörige Gütefunktion α_n ist

$$\begin{aligned} \alpha_n(\sigma^2) &= 1 - \mathbb{P}_{\sigma^2}\left(c_1 \le T(X) \le c_2\right) \\ &= 1 - \mathbb{P}_{\sigma^2}\left(\chi^2_{n,\alpha/2} \le \frac{n}{\sigma_0^2}\tilde{S}_n^2 \le \chi^2_{n,1-\alpha/2}\right) \\ &= 1 - \mathbb{P}_{\sigma^2}\left(\frac{\sigma_0^2}{\sigma^2}\chi^2_{n,\alpha/2} \le \frac{n}{\sigma^2}\tilde{S}_n^2 \le \frac{\sigma_0^2}{\sigma^2}\chi^2_{n,1-\alpha/2}\right), \end{aligned}$$

und da $\frac{n}{\sigma^2}\tilde{S}_n^2 \sim \chi_n^2$ gilt,

$$= 1 - \left(\mathbb{P}_{\chi_n^2}\left(\frac{\sigma_0^2}{\sigma^2}\chi^2_{n,1-\alpha/2}\right) - \mathbb{P}_{\chi_n^2}\left(\frac{\sigma_0^2}{\sigma^2}\chi^2_{n,\alpha/2}\right)\right).$$

Sie hat in $\sigma^2 = \sigma_0^2$ kein Minimum, da $\alpha_n(\sigma_0^2) = \alpha$ aber $\alpha_n(\sigma^2) < \alpha$ für $\sigma_0^2 > \sigma^2 \in \Theta_1$. Weiterhin ist

$$\begin{aligned} \tilde{\alpha}_n &= \mathbb{P}_{\sigma^2}\left(T(X_1,\dots,X_n) > \chi^2_{n,1-\alpha}\right) \\ &= \mathbb{P}_{\sigma^2}\left(\frac{n\tilde{S}_n^2}{\sigma^2} > \frac{\sigma_0^2}{\sigma^2}\chi^2_{n,1-\alpha}\right), \end{aligned}$$

da $\frac{\sigma_0^2}{\sigma^2}\chi^2_{n,1-\alpha} \sim \chi^2_n$, haben wir

$$= 1 - \mathbb{P}_{\chi^2_n}\left(\frac{\sigma_0^2}{\sigma^2}\chi^2_{n,1-\alpha}\right) \geq \alpha$$

für alle $\sigma^2 \geq \sigma_0^2$ wobei die Gleichheit nur für $\sigma^2 = \sigma_0^2$ gilt.

Aus diesen Rechnungen folgt, dass der zweiseitige Test nicht unverfälsch ist, der einseitige schon.

Test der Varianz σ^2 (bei unbekanntem Erwartungswert μ)

Annahme Gelte $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ unbekannt und sei $X = (X_1, \ldots, X_n)$ die Stichprobe.

Wir geben uns ein $\sigma_0 > 0$ vor, auf dem die Hypothese beruht.

Zweiseitiger Test Die Hypothesen sind

$$\begin{aligned} &H_0 : \sigma^2 = \sigma_0^2 && \Theta_0 = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\} \\ &H_1 : \sigma^2 \neq \sigma_0^2 && \Theta_1 = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \neq \sigma_0^2\} \end{aligned}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \to (0, \infty)$ mit

$$T(x_1, \ldots, x_n) = \frac{n-1}{\sigma_0^2} s_n^2 \qquad \text{und es gilt} \qquad T(X) = \frac{n-1}{\sigma_0^2} S_n^2 \sim \chi^2_{n-1} \tag{3.12}$$

unter der Nullhypothese mit den Schwellenwerten $c_1 = \chi^2_{n-1,\alpha/2}$ und $c_2 = \chi^2_{n-1,1-\alpha/2}$. Es gilt für alle $(\mu, \sigma^2) \in \Theta_0$, dass

$$\mathbb{P}_{(\mu,\sigma_0^2)}\left(c_1 \leq T(X) \leq c_2\right) = 1 - \alpha, \tag{3.13}$$

unabhängig von μ und der kritische Bereich ist

$$\begin{aligned} K &= \mathbb{R}^n \setminus \left\{(x_1, \ldots, x_n) : c_1 \leq T(x_1, \ldots, x_n) \leq c_2\right\} \\ &= \mathbb{R}^n \setminus \left\{(x_1, \ldots, x_n) : \chi^2_{n-1,\alpha/2} \leq \frac{n-1}{\sigma_0^2} s_n^2 \leq \chi^2_{n-1,1-\alpha/2}\right\} \\ &= \mathbb{R}^n \setminus \left\{(x_1, \ldots, x_n) : \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha/2} \leq \frac{n-1}{\sigma^2} s_n^2 \leq \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,1-\alpha/2}\right\} \end{aligned}$$

Einseitiger Test Die Hypothesen sind

$$\begin{aligned} &H_0 : \sigma^2 = \sigma_0^2 && \Theta_0 = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\} \\ &H_1 : \sigma^2 > \sigma_0^2 && \Theta_1 = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > \sigma_0^2\} \end{aligned}$$

Mit der gleichen Teststatikstik und den Schwellenwerten $c_1 = -\infty$ und $c := c_2 = \chi^2_{n-1,1-\alpha}$ gilt (3.13) analog und wir haben den kritischen Bereich

$$\begin{aligned}\widetilde{K} &= \mathbb{R}^n \setminus \{(x_1,\dots,x_n) : T(x_1,\dots,x_n) \le c\} \\ &= \left\{(x_1,\dots,x_n) : \frac{n-1}{\sigma_0^2} s_n^2 > \chi^2_{n-1,1-\alpha}\right\}.\end{aligned}$$

Die zugehörigen Gütefunktionen α_n resp. $\tilde{\alpha}_n$ sind

$$\begin{aligned}\alpha_n(\sigma^2) &= 1 - \mathbb{P}_{(\mu,\sigma^2)}\left(c_1 \le T(X) \le c_2\right) \\ &= 1 - \mathbb{P}_{(\mu,\sigma^2)}\left(\chi^2_{n-1,\alpha/2} \le \frac{n-1}{\sigma_0^2} S_n^2 \le \chi^2_{n-1,1-\alpha/2}\right)\end{aligned}$$

und

$$\begin{aligned}\tilde{\alpha}_n(\sigma^2) &= \mathbb{P}_{(\mu,\sigma^2)}\left(T(X) > c_2\right) \\ &= \mathbb{P}_{(\mu,\sigma^2)}\left(\frac{n-1}{\sigma_0^2} S_n^2 > \chi^2_{n-1,1-\alpha}\right).\end{aligned}$$

Der Argumentation beim Test der Varianz bei bekanntem Erwartungswert folgend, ist der einseitige Test unverfälscht, der zweiseitige aber nicht.

Anmerkung: Es gibt unverfälschte und gleichmässig effizienteste Tests, mit denen die zweiseitigen Tests der letzten zwei Szenarien approximativ übereistimmen.

3.4 Äquivalenztest

Unter dem Begriff Äquivalenz versteht man die Gleichheit eines Parameters θ mit einem Wert θ_0 bis auf *praktisch irrelevante* Abweichungen. Dazu legen wir einen *Äquivalenzbereich* $A = [\theta_0 - \varepsilon_1, \theta_0 + \varepsilon_2]$ fest, häufig symmetrisch, bzw. $A = \left[(1-\beta_1)\theta_0, \frac{1}{1-\beta_2}\theta_0\right]$ im multiplikativen Fall.

Dieses Szenario führt zu den Hypothesen

$$\begin{aligned}H_0 &: \theta \notin A, \\ H_1 &: \theta \in A,\end{aligned}$$

was bei Ablehnung der Nullhypothese die Äquivalenz bedeutet. Das Vorgehen beschreibt demnach die gegensätzliche Fragestellung zu den bisher kennengelernten Tests.

Problem Der Nachteil dieser Fragestellung ist in der Schwierigkeit gespiegelt, eine geeignete Teststatistik mit bekannter Verteilung unter der Nullhypothese zu finden.

Häufig wird der direkte Äquivalenztest durch eine Hintereinaderschaltung einseitiger Tests ersetzt. Dabei muss jedoch das Konfidenzniveau der einzelnen Tests angepasst werden.

Test auf Äquivalenz von Erwartungswerten (bei bekanntem σ^2) unter der Normalverteilungsannahme

Annahmen $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ und $\mu \in \mathbb{R}$ unbekannt, $\sigma^2 > 0$ bekannt.

Wir gehen uns ein $\mu_0 \in \mathbb{R}$ vor, auf dem die Hypothese beruht, außerdem $\varepsilon_1, \varepsilon_2 > 0$ mit $A = [\mu_0 - \varepsilon_1, \mu_0 + \varepsilon_2]$ und stellen die Hypothese

$$H_0 : \mu \notin A,$$
$$H_1 : \mu \in A,$$

auf. Wir können alternativ standardisieren und haben mit $\varepsilon'_{1,2} := \frac{\varepsilon_{1,2}}{\sigma/\sqrt{n}}$ den Äquivalenzbereich

$$A' = \left[\frac{\mu_0}{\sigma/\sqrt{n}} - \varepsilon'_1, \frac{\mu_0}{\sigma/\sqrt{n}} + \varepsilon'_2\right]$$

und die Hypothesen

$$\begin{aligned} H_0 &: \frac{\mu}{\sigma/\sqrt{n}} \notin A', &\text{also}&\quad \frac{\mu-\mu_0}{\sigma/\sqrt{n}} \leq -\varepsilon'_1 \text{ oder } -\varepsilon'_2 \leq \frac{\mu-\mu_0}{\sigma/\sqrt{n}}, \\ H_1 &: \frac{\mu}{\sigma/\sqrt{n}} \in A', &\text{also}&\quad -\varepsilon'_1 \leq \frac{\mu-\mu_0}{\sigma/\sqrt{n}} \leq \varepsilon'_2. \end{aligned}$$

Nun wissen wir, dass

1. $\overline{X}_n \sim \mathcal{N}\left(\mu, \sigma^2/n\right)$, somit $\frac{\overline{X}_n-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}\left(\frac{\mu-\mu_0}{\sigma^2/n}, 1\right)$ und es gilt

$$\mathbb{P}\left(\overline{X}_n - \mu_0 > 0\right) = \mathbb{P}\left(\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} > 0\right) = \Phi\left(\right).$$

Beweis Betrachte $Y \sim \mathcal{N}(0,1)$ und $Z \sim \mathcal{N}(\mu, 1)$. Dann gilt

$$\mathbb{P}(Z \geq \mu) = \mathbb{P}(Y \geq 0)$$

bzw. um μ verschoben

$$\begin{aligned} \mathbb{P}(Z \geq 0) &= \mathbb{P}(Y \geq -\mu) \\ &= \mathbb{P}(Y \leq \mu) = \Phi(\mu). \end{aligned}$$

□

2. Weiter gilt unter der Nullhypothese, dass

$$\mathbb{P}\left(X_1 - \mu_0 > 0\right) \not\approx \mathbb{P}\left(X_1 - \mu_0 < 0\right)$$

ist. Würden wir $\mathbb{P}\left(X_1 - \mu_0 > 0\right) \approx \mathbb{P}\left(X_1 - \mu_0 < 0\right)$ annehmen, so gäbe es $\tilde{\varepsilon}_1, \tilde{\varepsilon}_2 > 0$, so dass

$$1/2 - \tilde{\varepsilon}_1 \leq \mathbb{P}(\overline{X}_n - \mu_0 > 0) \leq 1/2 + \tilde{\varepsilon}_2$$

und mit oben

$$1/2 - \tilde{\varepsilon}_1 \leq \Phi\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right) \leq 1/2 + \tilde{\varepsilon}_2,$$

was

$$-\varepsilon_1' = \Phi^{-1}\left(1/2 - \tilde{\varepsilon}_1\right) \leq \frac{\mu - \mu_0}{\sigma/\sqrt{n}} \leq \Phi^{-1}\left(1/2 + \tilde{\varepsilon}_2\right) = \varepsilon_2'$$

entspricht. Dabei sind $\tilde{\varepsilon}_1 := 1/2 - \Phi\left(-\varepsilon_1'\right)$ und $\tilde{\varepsilon}_2 := \Phi\left(\varepsilon_2'\right) - 1/2$.

Damit können wir die maximalen Abweichungen zwischen der Wahrscheinlichkeit für eine positive respektive negative Differenz und $1/2$ ist begrenzt.

$$\Delta_1 := \mathbb{P}\left(\overline{X}_n - \mu_0\right) - 1/2 \leq \tilde{\varepsilon}_2 = \Phi\left(\varepsilon_2'\right) - 1/2$$
$$\Delta_2 := 1/2 - \mathbb{P}\left(\overline{X}_n - \mu_0\right) \leq \tilde{\varepsilon}_1 = 1/2 - \Phi\left(-\varepsilon_1'\right).$$

Sei z. B. $\varepsilon_1' = \varepsilon_2' = 0{,}1$. Dann gilt $\Delta_1 = \Delta_2 = \max\{\Delta_1, \Delta_2\} = 0{,}0398$.

3.5 p-Wert versus Konfidenzniveau

Der p-Wert einer Teststatistik einer Stichprobe liefert die Wahrscheinlichkeit, dass unter der Nullhypothese eine Stichprobe mindestens so extrem ist wie bei der untersuchten. Er gewann erst mit der Verwendung von Computern und Statistiksoftware an Bedeutung, als es möglich wurde, diese Wahrscheinlichkeit zu berechnen.

Sei $z = T(x_1, \ldots, x_n)$ die ausgewertete konkrete Stichprobe. Dann ist der p-Wert bei rechtsseitigem Test

$$p_r = p_r(x_1, \ldots, x_n) = \mathbb{P}_{H_0}\left(T(X_1, \ldots, X_n) \geq z\right),$$

bei linksseitigem Test

$$p_l = p_l(x_1, \ldots, x_n) = \mathbb{P}_{H_0}\left(T(X_1, \ldots, X_n) \leq z\right),$$

und bei beidseitigem Test

$$p = 2 \cdot \min\{p_l, p_r\}.$$

Vorgehen Zuerst wird ein Signifikanzniveau α festgelegt und wir sagen

$$(x_1, \ldots, x_n) \in K, \qquad \text{falls} \qquad p(x_1, \ldots, x_n) \leq \alpha,$$

was zur Entscheidung führt.

3.6 Asymptotische Tests

3.6.1 χ^2-Anpassungstest

Annahme Seien $X, X_1, \ldots, X_n$ i. i. d. mit $X \sim F$ für eine Verteilung F.

Wir geben uns Intervalle $A_1 = (a_1, b_1], \dots, A_r = (a_r, b_r]$ vor, die paarweise disjunkt sind.

Desweiteren betrachten wir Vektoren $(p_1, \dots, p_r)$ mit $0 < p_i < 1$ und $\sum_{i=1}^r p_i = 1$.

Hypothesen

$$H_0 : \mathbb{P}(X \in A_i) = \tilde{p}_i \text{ für alle } 1 \le i \le r \qquad \Theta_0 = \{(\tilde{p}_1, \dots, \tilde{p}_r)\}$$

$$H_1 : \mathbb{P}(X \in A_i) \ne \tilde{p}_i \text{ für ein } 1 \le i \le r \qquad \Theta_1 = \{(p_1, \dots, p_r) \in (0,1)^r : \sum_{i=1}^r p_i = 1, \exists i \text{ mit} p_i \ne \tilde{p}_i\}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \to [0, \infty)$ mit

$$T(x_1, \dots, x_n) = \sum_{j=1}^r \frac{\left(y_j(x_1, \dots, x_n) - n \cdot \tilde{p}_j\right)^2}{n \cdot \tilde{p}_j},$$

wobei $y_j(x_1, \dots, x_n) = \left|\left\{i : 1 \le i \le n : a_j < x_i \le b_j\right\}\right|$ die Anzahl der Stichprobenwerte $x_1, \dots, x_n$ angibt, die im j-ten Intervall liegen. Die Funktionen T und y lassen sich auf die Zufallsstichprobe anwenden und wir haben eine asymptotische Verteilungsaussage für die Zufallsvariable $T(X_1, \dots, X_n)$ mit

$$\lim_{n\to\infty} T(X_1, \dots, X_n) \sim \chi^2_{r-1} \tag{3.14}$$

und mit dem Schwellenwert $c = \chi^2_{r-1,1-\alpha}$, dass

$$\lim_{n\to\infty} \mathbb{P}_{(\tilde{p}_1,\dots,\tilde{p}_r)}\left(T(X_1, \dots, X_n) > \chi^2_{r-1,1-\alpha}\right) = \alpha. \tag{3.15}$$

Anmerkung: Die asymptotische Verteilung ergibt sich aus der näheren Untersuchung des Zufallsvektors $(Y_1, \dots, Y_n)$. Dieser ist multinomialverteilt mit den Parametern $n \in \mathbb{N}$ und $(\tilde{p}_1, \dots, \tilde{p}_r)$, d. h. für $k_1, \dots, k_r \in \mathbb{N}$ mit $\sum_{i=1}^r k_i = n$ gilt

$$\mathbb{P}(Y_1 = k_1, \dots, Y_r = k_r) = \frac{n!}{k_1! \cdot \ldots \cdot k_r!} p_1^{k_1} \cdot \ldots , \cdot p_r^{k_r}$$

wobei $p_j = \mathbb{P}(X \in A_j) = F(b_j) - F(a_j)$ ist, da

$$\mathbb{P}(X_1 \in A_{i_1}, \dots, X_n \in A_{i_n}) = \prod_{j=1}^n \left(F(b_{i_j}) - F(a_{i_j})\right)$$

wegen der Unabhängigkeit der X_i gilt. Dann ist jede der r Zufallsvariablen $Y_1, \dots, Y_r$ binomialverteilt, und damit asymptotisch normalverteilt, woraus (3.14) folgt.

Beispiel 3.8 (Wahlprognose): Anlässlich der bald stattfindenden Landtagswahl soll überprüft werden, ob sich das Wahlverhalten seit 2006 geändert hat. Die Verteilung, gegen die wir testen, basiert auf den Ergebnissen von 2006:

CDU	44,15	%
SPD	25,15	%
Grüne	11,69	%
FDP	10,65	%
andere	8,36	%

Emnid befragte in der zweiten Dezemberwoche (2010) 1001 Personen[1] und wir haben

CDU	407
SPD	191
Grüne	295
FDP	37
andere	71

Wir wenden hier den χ^2-Anpassungstest an, siehe R-Datei.

3.6.2 χ^2-Unabhängigkeitstest

Annahme Wir haben eine verbundene i. i. d. Stichprobe $X_1, \dots, X_n$, wobei $X_i = (X_{i_1}, X_{i_2})$, deren Verteilung der von $X = (\widetilde{X}_1, \widetilde{X}_2)$ folgt, also $X_1 \sim X$. Die zwei Einträge $\widetilde{X}_1$ und $\widetilde{X}_2$ untereinander müssen nicht unabhängig sein, sondern werden gerade auf Unabhängigkeit getestet.

Wir geben uns Intervalle $A_1, \dots, A_I$ und $B_1, \dots, B_J$ vor, die jeweils paarweise disjunkt sind.

Hypothesen

$$\begin{aligned} H_0 : p_{ij} &= \mathbb{P}\left((\widetilde{X}_1, \widetilde{X}_2) \in A_i \times B_j\right) = \mathbb{P}(\widetilde{X}_1 \in A_i)\mathbb{P}(\widetilde{X}_2 \in A_j) && \text{für alle } i, j \\ H_1 : p_{ij} &= \mathbb{P}\left((\widetilde{X}_1, \widetilde{X}_2) \in A_i \times B_j\right) \neq \mathbb{P}(\widetilde{X}_1 \in A_i)\mathbb{P}(\widetilde{X}_2 \in A_j) && \text{für mindestens ein Paar } (i, j) \end{aligned}$$

Wir definieren

$$N_{ij}(X_1, \dots, X_n) = \left|\left\{k : 1 \leq k \leq n : X_{k_1} \in A_i, X_{k_2} \in B_j\right\}\right| \qquad \text{für } 1 \leq i \leq I, 1 \leq j \leq J$$

die Anzahl der Stichprobenzufallsvariablen $X_1, \dots, X_n$, die im (i, j) Rechteck landen, und n_{ij} die Anzahl der (konkreten) Stichprobenwerte $x_1, \dots, x_n$ entsprechend. Diese Werte kann man tabellarisch anordnen; die resultierende Tabelle wird Kontingenztafel genannt. In dieser definieren wir Spalten-

$$n_{\bullet j} := \sum_{i=1}^{I} n_{ij} \qquad \text{und Zeilensummen} \qquad n_{i\bullet} := \sum_{j=1}^{J} n_{ij}$$

und die Wahrscheinlichkeiten, in dem (i, j) Rechteck (bzw. Zelle der Tabelle) zu landen führen zu den Marginalwahrscheinlichkeiten

$$p_{\bullet j} := \sum_{i=1}^{I} p_{ij} \qquad \text{und} \qquad p_{i\bullet} := \sum_{j=1}^{J} p_{ij},$$

weiter bezeichnen wir mit π diese zwei-dimensionale Verteilung. Für diese Wahrscheinlichkeiten haben wir die Schätzer[2]

$$\hat{p}_{ij} \overset{\text{unter } H_0}{=} \hat{p}_{i\bullet} \cdot \hat{p}_{\bullet j} \overset{ML}{=} \frac{n_{i\bullet}}{n} \cdot \frac{n_{\bullet j}}{n} \tag{3.16}$$

[1]Es wurden die relativen Ergebnisse veröffentlicht. Die Zahlen hier sind ein möglicher Ausgang der Umfrage.
[2]Diese sind die ML-Schätzer

Dann gilt unter der Nullhypothese für die geschätzte Anzahl in einem Rechteck

$$\hat{m}_{ij} = n\hat{p}_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}. \tag{3.17}$$

Die Teststatistik $T : \mathcal{X} \to [0, \infty)$ mit

$$T(x_1, \ldots, x_n) = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - \hat{m}_{ij}\right)^2}{\hat{m}_{ij}},$$

Dann ist $T(X_1, \ldots, X_n) \sim \chi^2_{(I-1)(J-1)}$, wobei sich die Anzahl der Freiheitsgrade ergibt aus $(I \cdot J - 1) - (I - 1) - (J - 1) = (I - 1)(J - 1)$.

Damit haben wir den Schwellenwert $c = \chi^2_{(I-1)(J-1),1-\alpha}$ und

$$\lim_{n\to\infty} \mathbb{P}_\pi \left(T(X_1, \ldots, X_n) > \chi^2_{(I-1)(J-1),1-\alpha}\right) = \alpha. \tag{3.18}$$

Beispiel 3.9: Wir nehmen folgende Umfrage an: n Wahlberechtigte in Baden-Württemberg werden befragt, welche Partei sie in der kommenden Landtagswahl wählen wollen und welche sie 2006 gewählt haben. Dann haben wir $I = 5$ (CDU, SPD, Grüne, FDP, andere) und auch $J = 5$ und die einzelnen Befragten werden im jeweiligen Rechteck zusammengezählt.

4 Lineare Regression

Wir wollen einen kurzen Einblick in die lineare Regression geben. Zur Erinnerung betrachten wir den Kleinste-Quadrate-Schätzer.

Definition 4.1: Seien $\binom{x_1}{y_1}, \ldots, \binom{x_n}{y_n}$ Wertepaare. Definiere die normierte Summe der quadratischen Abweichungen

$$Q(\tilde{h}) := \frac{1}{n} \sum_{i=1}^{n} \Big(y_i - \tilde{h}(x_i) \Big)^2 \tag{4.1}$$

für $\tilde{h} \in \mathcal{F}_\Theta$, einer parametrischen Funktionenfamilie.

Bei der linearen Regression wird im eindimensionalen Fall die Familie der Geraden

$$\mathcal{F}_\Theta = \Big\{ x \mapsto \alpha + \beta x : (\alpha, \beta) \in \Theta \subset \mathbb{R}^2 \text{ für } x \in \mathbb{R} \Big\}$$

betrachtet.

Notation: In diesem Fall ist das Argument von Q der Parametervektor, also $Q(\alpha, \beta)$.

Standardmodell Die einfachste Art des Zusammenhangs zwischen zwei Zufallsvariablen ist durch eine Gerade $h(x) = \alpha + \beta x$ gegeben. Dann ist

$$Y = h(X) + Z = \alpha + \beta X + Z \tag{4.2}$$

und wir nennen Y die erklärte Variable, X die erklärende Variable.

Normalverteilungsannahme Wir gehen von folgenden Bedingungen aus:

$$\mathbb{E}Z = 0, \qquad \text{d. h. } \mathbb{E}Y = h, \tag{4.3}$$

$$\mathbb{V}\mathrm{ar} Z = \sigma^2 \qquad \text{(unabhängig von } x\text{)} \tag{4.4}$$

Anmerkung: Wir nennen β den *Regressionskoeffizienten* von Y bezüglich X und verwenden diesen zur Vorhersage für Y aus X. Dazu greifen wir zurück auf die Korrelation und sagen, dass $\beta = \rho(X, Y) \frac{\sigma_X}{\sigma_Y}$ zur Vorhersage geeignet ist.

Kleinste-Quadrate Schätzer ist im Falle der linearen Regression die Funktion $\hat{h} \in \mathcal{F}_\Theta$, die zu einer gegebenen Stichprobe $(x, y) = \left(\binom{x_1}{y_1}, \ldots, \binom{x_n}{y_n} \right)$ den minimalem Wert für $Q(\hat{h})$ liefert, also der Parametervektor

$$\hat{\theta} = (\hat{\alpha}, \hat{\beta}) = \operatorname*{argmin}_{(\alpha,\beta) \in \Theta} Q(\alpha, \beta) = \operatorname*{argmin}_{\theta \in \Theta} \sum_{i=1}^{n} \big(y_i - \alpha - \beta x_i \big)^2 \tag{4.5}$$

Vorgehen: Um den Schätzer zu bestimmen, müssen wir $Q(\alpha, \beta)$ minimieren. Dazu wird differenziert und wir erhalten die Gleichungen

$$\begin{aligned}\frac{\partial Q(\alpha,\beta)}{\partial \alpha} &= -\frac{2}{n}\sum_{i=1}^{n}\big(y_i - (\alpha + \beta x_i)\big) = 0 \\ \frac{\partial Q(\alpha,\beta)}{\partial \beta} &= -\frac{2}{n}\sum_{i=1}^{n}\big(y_i - (\alpha + \beta x_i)\big)\, x_i = 0\end{aligned}$$

und damit

$$\begin{aligned}\hat\alpha &= \overline{y}_n - \hat\beta \overline{x}_n \\ \hat\beta &= \frac{\sum_{i=1}^n y_i x_i - \overline{y}_n \overline{x}_n}{\sum_{i=1}^n x_i^2 - \overline{x}_n^2} = \frac{\sum_{i=1}^n (x_i - \overline{x}_n)(y_i - \overline{y}_n)}{\sum_{i=1}^n (x_i - \overline{x}_n)^2}\end{aligned}$$

Die Schätzer sind

$$\begin{aligned}\hat\alpha(X,Y) &= \overline{Y}_n - \hat\beta \overline{X}_n \\ \hat\beta(X,Y) &= \frac{\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sum_{i=1}^n (X_i - \overline{X}_n)^2}\end{aligned}$$

Hier sehen wir schon einen Zusammenhang mit dem Korrelationskoeffizienten. Aus diesem Grund definieren wir

Definition 4.2: Sei $\binom{X_1}{Y_1}, \ldots, \binom{X_1}{Y_1}$ eine verbundene Stichproben. Dann ist

$$r_{X,Y,n} = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \overline{X}_n)^2 \cdot \sum_{i=1}^n (X_i - \overline{X}_n)^2}}$$

der *empirische* oder *Bravais-Pearson-Korrelationskoeffizient*. Die *empirische Kovarianz* ist durch

$$\widetilde{S}_{X,Y,n} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$$

und die *empirischen Standardabweichungen* sind durch

$$\widetilde{S}_{X,n} = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)} \qquad \text{resp.} \qquad \widetilde{S}_{Y,n} = \sqrt{\frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y}_n)}$$

gegeben.

Satz 4.3: Sei (X, Y) eine verbundene Stichprobe mit $0 < \mathbb{V}\mathrm{ar}(X), \mathbb{V}\mathrm{ar}(Y) < \infty$. Es gilt

(i) Die umnormierte Stichprobenkovarianz $S_{X,Y,n} := \frac{n}{n-1}\widetilde{S}_{X,Y,n}$ ist ähnlich der Stichprobenvarianz S_n^2 ein erwartungstreuer Schätzer für $\mathbb{C}\mathrm{ov}(X, Y)$.[1]

(ii) Sowohl $S_{X,Y,n}$ als auch $r_{X,Y,n}$ sind schwach und stark konsistent.

[1] Die Literatur ist da mit den Bezeichnungen unschlüssig. Wir nehmen die Tilde-Variante immer bei der Normierung mit n, bei $n-1$ aber ohne Tilde.

Dann gilt auch $r_{X,Y,n} = \frac{\widetilde{S}_{X,Y,n}}{\widetilde{S}_{X,n}\widetilde{S}_{Y,n}}$ und für unsere Schätzer haben wir

$$\hat{\alpha}(X,Y) = \overline{Y}_n - \hat{\beta}\overline{X}_n, \tag{4.6}$$

$$\hat{\beta}(X,Y) = r_{X,Y,n} \cdot \frac{\widetilde{S}_{Y,n}}{\widetilde{S}_{X,n}}, \tag{4.7}$$

und definieren die Zufallsvariable $\hat{Y} := \hat{\alpha} + \hat{\beta}X$, welche die beste lineare Vorhersage von Y durch X bezüglich der mittleren quadratischen Abweichung ist.

Anmerkung: Sind X und Y Zufallsvariable mit $\mathbb{E}(X) = \mathbb{E}(Y) = 0$ und weiter $\mathbb{Var}(X) = \mathbb{Var}(Y) = 1$, so können wir als beste lineare Vorhersage $\widetilde{Y}$ für Y durch X

$$\widetilde{Y} = \rho(X,Y)X$$

erhalten. Dann gilt

$$|\widetilde{Y}| = |\rho(X,Y)| \cdot |X| \leq |X|,$$

was die Namensgebung *Regression* begründet.

Beispiel 4.4: Galileo war der bekannteste Experimentator seiner Zeit. Eines der Themen, die ihn beschäftigten, waren die Fallgesetze. Um den Zusammenhang zwischen Fallhöhe h und Fallzeit t zu ermitteln, warf er verschiedene Steine von diversen Höhen. Mit der Gravitationskonstante g kennen wir den Zusammenhang

$$t = \sqrt{2\,\frac{h}{g}},$$

das heißt, wir haben einen linearen Zusammenhang zwischen $\sqrt{h}$ und t.

Die Ergebnisse von Galileo waren übrigens sehr ungenau. Zur Zeitmessung hat er beispielsweise seinen eigenen Puls gezählt.

Wenn wir eine sehr genaue Messung durchführen würden, so würde diese Linearisierung nicht gelten, da die Gravitationskonstante $g(h)$ höhenabhängig ist.

Hinweis: Bitte nicht nachmachen, da ein Stein, der jemandem aus mehreren Hundert Metern auf den Kopf fällt, wirklich Schaden anrichten kann.

Anmerkung: Betrachten wir den Regressionskoeffizienten genauer.

$$\beta = \rho(X,Y)\frac{\sigma_Y}{\sigma_X}$$

beschreibt nicht eindeutig den Zusammenhang von X und Y. So können zwei Paare von Zufallsvariablen $(X^{(1)}, Y^{(1)})$ und $(X^{(2)}, Y^{(2)})$ zwar den gleichen Regressionskoeffizienten besitzen, wobei das zweite Paar eine geringere Korrelation mit einer erhöhten Streuung von $Y^{(2)}$ kompensiert.

Um die Güte eines Regressionsmodells zu beurteilen, bezeichnen wir den Fehler

$$\hat{Z}_i = Y_i - \hat{Y}_i$$

als *Residuum*, welches wir weiter untersuchen. Damit können wir die Streuung zerlegen. Wir sagen

$$\underbrace{SQT}_{\text{total sum of squares}} = \underbrace{SQE}_{\text{explained sum of squares}} + \underbrace{SQR}_{\text{residual sum of squares}} \tag{4.8}$$

wobei

$$SQT = \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2 = n \cdot \widetilde{S}_{Y,n}$$
$$SQE = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y}_n)^2$$
$$SQR = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Das *Bestimmtheitsmaß*

$$R^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT}$$

beschreibt den Anteil der durch das Modell erklärten Varianz.

Lemma 4.5: Im Standardmodell ist

$$\hat{\sigma}^2(X,Y) := \frac{SQR}{n-2}$$

ein erwartungstreuer Schätzer für σ^2.

Die KQ-Schätzer sind linear in $Z_1, \ldots, Z_n$ und damit können wir auch eine Aussage über ihre Verteilung machen.

Verteilung der standardisierten Schätzer Unter der Normalverteilungsannahme gilt weiter

$$\frac{\hat{\alpha}(X,Y)}{\hat{\sigma}_{\hat{\alpha}}(X,Y)} \sim t_{n-2} \qquad \frac{\hat{\beta}(X,Y)}{\hat{\sigma}_{\hat{\beta}}(X,Y)} \sim t_{n-2}$$

mit

$$\hat{\sigma}_{\hat{\alpha}}(X,Y) = \hat{\sigma}(X,Y) \cdot \frac{\sqrt{\sum_{i=1}^{n} X_i^2}}{\sqrt{n \sum_{i=1}^{n} (X_i - \overline{X}_n)^2}} \qquad \hat{\sigma}_{\hat{\beta}}(X,Y) = \hat{\sigma}(X,Y) \cdot \frac{1}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}}$$

und es ist

$$\widehat{F} = \frac{SQE}{SQR/n-2} \sim F_{k,n-(k+1)}. \tag{4.9}$$

Damit lassen sich sowohl Konfidenzintervalle als auch Tests für die Parameter der Regressionsgerade erstellen.