Angewandte Statistik

8. November 2001

Inhaltsverzeichnis

1 Information, Entscheidung, Statistik **1**

1.1 Daten und Information 1

1.2 Entscheidung bei Unsicherheit 2

1.3 Kontinuierliche Entscheidungsprobleme 3

2 Beschreibende Statistik **4**

2.1 Empirische Verteilungsfunktion 4

2.2 Diskrete Größen 4

2.3 Kontinuierliche Größen 4

2.4 Statistische Kenngrößen eindimensionaler Häufigkeitsverteilungen 4

2.4.1 Lageparameter 4

2.4.2 Streuungsparameter 7

2.5 Schiefe und Exzess eindimensionaler, empirischer Verteilungen 9

2.6 2-Dimensionale Häufigkeitsverteilungen 9

2.6.1 Kontingenzdaten 10

2.6.2 Empirischer Korrelationskoeffizient 10

2.7 Indexzahlen 11

2.7.1 Einfache Indexzahlen 11

2.7.2 Zeitreihen 11

2.7.3 Indexzahlen (zusammengesetzte Indizes) 12

3 Grafische Methoden zu Beurteilung empirischer Verteilungen **14**

3.1 Normalverteilungsnetz 14

3.2 Wahrscheinlichkeitsnetze für logarithmische normalverteilte Merkmale 15

3.3 Zuverlässigkeitsanalyse und Ausfallsraten 15

3.4 Lebensdauer-Netz (Weibull-Verteilung) 17
3.5 Anpassung von Verteilungsmodellen 18
3.6 Ergänzung zu Schätzungen 20

4 Bayes´sche Anteilsschätzung 23

4.1 Gamma-Funktion 23
4.2 Gamma-Verteilung Gam(α, β) $\alpha > 0$, $\beta > 0$ 23
4.3 Beta-Verteilung 1.Art Be(α, β) $\alpha > 0$, $\beta > 0$ 24
4.4 Bayes´sches Theorem 25
4.5 A-posteriori Dichte für einfache Anteile 27
4.6 HPD-Intervalle für θ 28
4.7 Bayes´sche Anteilsschätzung bei mehrfachen Alternativen 31
4.8 Konjugierte Verteilungsfamilien 33
4.9 Suffizienz 34

5 Statistische Qualitätskontrolle 35

5.1 Einfache Stichprobenpläne 35
5.2 Festlegung einfacher Stichprobenpläne 38
5.3 Zweifache Stichprobenpläne 40
5.4 Sequentielle Stichprobenpläne 40
5.5 Kontrollkarten 41

6 Regressionsanalyse 46

6.1 Regression 1.Art 46
6.2 Regression 2.Art 50
6.3 Stochastische Regressionsanalyse 54
6.4 Regressionsgeraden 2.Art bei Normalverteilung 55
6.5 Tests für den Parameter von Regressionsgeraden bei Normalverteilung 60
6.6 Tests für Regressionskurven bei Normalverteilung 62
6.7 Test auf Regressionsgerade 62
6.8 Multiple lineare Regression 65
6.9 Multiple lineare Regression bei Normalverteilung 69
6.10 Bayes´sche Regressionsanalyse 72

7 Varianzanalyse **75**

7.1 Grundlagen der Varianzanalyse (ANOVA) 75

7.2 Einfache Varianzanalyse (1 Faktor) 76

7.3 Zweifache Varianzanalyse 78

1 Information, Entscheidung, Statistik

Für gute Organisation bzw. vernünftige Entscheidungen

Beispiel: Wasserbedarf eines Stadt: Zahlenangaben, Meßvorgänge

Statistik ist die zahlenmäßige Erfassung und Beschreibung von Phänomenen.

Beispiel:

- Volkszählung (Zensus)
- Stichprobenerhebung (Mikrozensus, Sozialstatistik)
- Prognosen (z.B. Bevölkerung)

1.1 Daten und Information

Aus Erhebungen bzw. empirischen Untersuchungen (Versuche) erhält man Rohdaten, oft auch nur Daten genannt. Beispiel:

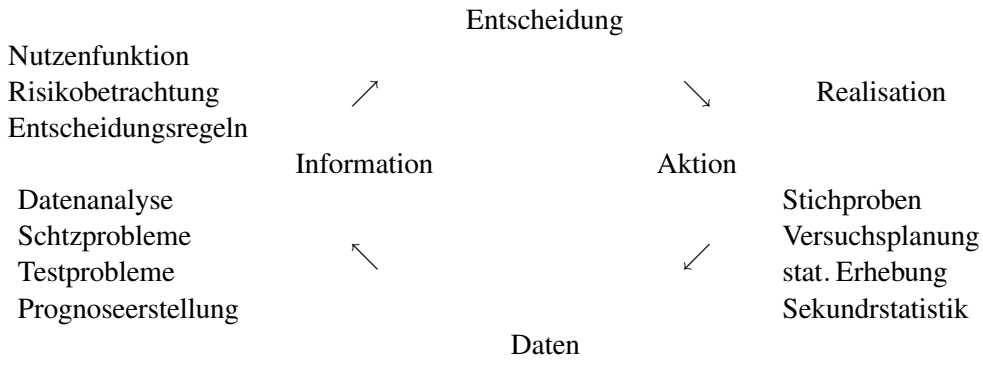
- Messung von Wartezeiten
- Messung von Lebensdauern

Informationsgewinnung: Daten oft in großer Zahl, daher unübersichtlich; schwierig zu vergleichen. Als Entscheidungsgrundlage nicht geeignet. Konzentration der Daten in Tabellen, Diagrammen und Ermittlung von Verteilungen.

Bemerkung: Kontrolle auf Vollständigkeit/Plausibilität

Raffung von Daten: z.B. Börsenindex

Information-Feedback-Cycle



1.2 Entscheidung bei Unsicherheit

Die stochastische Beschreibung von Entscheidungen, welche mit Nutzen bzw. Verlust verbunden sind.

Beispiel: Planung einer Straße, m mögliche Varianten, k verschiedene mögliche Klassen von Fahrzeughäufigkeiten.

$\theta_1, ..., \theta_k$ k verschiedene Zustnde (Fahrzeughufigkeiten)

$d_1, ..., d_m$ mgliche Entscheidungen (Fahrbahnbreiten)

$\widetilde{\theta}$ stochastische Gre, die den Zustand beschreibt

$p(\theta_1), ..., p(\theta_k)$ Wahrscheinlichkeiten der $\theta_1, ..., \theta_k \hat{=}$Wahrscheinlichkeitsverteilung von $\widetilde{\theta}$

$U(\theta_i,\, d_j) \geq 0$ Nutzenfunktion; Nutzen der Entscheidung d_j, falls System im Zustand θ_i ist.

$$\bar{U}(d_j) = \mathbb{E}U(\widetilde{\theta}, d_j) = \sum_{i=1}^{k} U(\theta_i, d_j) \cdot p(\theta_i)$$

zu erwartender Nutzen der Entscheidung d_j.

Die optimale Entscheidung d^* ist jenes d_j, für das $\bar{U}(d_j)$ maximal ist (Bayes´sche Entscheidung).

Bemerkung: Mann kann diese Probleme auch für kontinuierliche Zustände und Entscheidungsmöglichkeiten beschreiben (vgl. Abschnitt 1.3).

Beispiel: Ausnahme/Ablehnung einer Warensendung

N Losumfang

θ	Anteil der schlechten Stcke im Los
d_0	Entscheidung Annahme
d_1	Entscheidung Ablehnung

Entscheidung ist abhängig von einer Stichprobe: $x_1, ..., x_n$ mit $n \ll N$, $d = \delta(x_1, ..., x_n)$, $\delta \widehat{=}$Entscheidungsregel
Aus der Stichprobe erhält man Aussagen über θ (vgl. 4).

K	$=$	Verlust fr ein schlechtes Stck
G		Gewinn fr ein gutes verkauftes Stck

Nutzen:

$U(\theta, d)$		falls Anteil θ
$U(\theta, d_0)$	$=$	$(1-\theta)NG - \theta NK$
$U(\theta, d_1)$	$=$	0

Entscheidungskriterium: zu erwartender Nutzen $\mathbb{E}[U(\tilde{\theta}, d)]$
Erwartungsbildung mittels der Verteilung von $\tilde{\theta}$: in diesem Fall: a-posteriori Verteilung $\pi(\theta|x_1, ..., x_n)$ (vgl. Abschnitt 4)

$$\tilde{\theta} \sim U_{0,1} \quad \Rightarrow \quad \tilde{\theta}|_{x_1,...,x_n} \widehat{=} \text{Be}(\cdot, \cdot)$$

Entscheidung: $d_0 \widehat{=} \bar{U}(d_0) > 0$
Frage: $\bar{U}(d_j) =?$

1.3 Kontinuierliche Entscheidungsprobleme

Die Menge der möglichen Werte für θ ist kontinuierlich: $\theta \in \Theta$. Daher ist die Verteilung von $\tilde{\theta}$ ebenfalls kontinuierlich mit Dichtefunktion $g(\theta)$. Der erwartende Nutzen ist:

$$\bar{U}(d) = \mathbb{E}[U(\tilde{\theta}, d)] = \int_{\Theta} U(\theta\, d) \cdot g(\theta) d\theta$$

Bemerkung: Häufig ist $g(\theta)$ eine a-posteriori-Dichte $\pi(\theta|D)$.

Die optimale Entscheidung d^* ist dann jene, mit größtem, erwartenden Nutzen, d.h.:

$$\bar{U}(d^*) = \max_{d \in \mathcal{D}} \bar{U}(d) \qquad \mathcal{D} \widehat{=} \text{Menge aller mglichen Entscheidungen}$$

2 Beschreibende Statistik

Erhält man aus Erhebungen oder Versuchen eine große Anzahl von Daten, so sollen diese zu einer überschaubaren Information zusammengefasst werden: Konzentration in Kennzahlen, Tabellen, Diagrammen und empirischen Verteilungen.

2.1 Empirische Verteilungsfunktion

2.2 Diskrete Größen

2.3 Kontinuierliche Größen

2.4 Statistische Kenngrößen eindimensionaler Häufigkeitsverteilungen

Für Beobachtungen $x_1, ..., x_n$ eines 1-dimensionalen Merkmales sucht man irgendwie charakterisierende Größen der Verteilung, z.B. für die Mitte (Lageparameter) oder für das Streuverhalten der Verteilung (Streuparameter).

2.4.1 Lageparameter

Mittlerer Wert

$$x_1, ..., x_n \quad \rightarrow \quad \bar{x}_n := \frac{\sum_{i=1}^{n} x_i}{n} \qquad \text{Stichprobenmittel}$$

Sind die Daten gruppiert mit Klassenmitten z_j und Häufigkeiten H_j so gilt:

$$\bar{z} = \frac{1}{n} \sum_{j=1}^{k} z_j \cdot H_j \quad j = 1..k \qquad \text{empirischer Mittelwert}$$

Bei der Einteilung in Klassen kommt es zu einem Verlust von Information.

Geometrisches Mittel

$$\bar{x}_g := \sqrt[n]{x_1 \cdot x_2 \cdot .... \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Beispiel: Wachstumsfaktoren bei Gemeindegrößen

Harmonisches Mittel

z.B. für Durchschnittsgeschwindigkeiten

$x_1, ..., x_n \qquad x_i > 0 \quad \text{oder} \quad x_i < 0 \quad \forall i$

$$\bar{x}_h := \frac{h}{\sum\limits_{i=1}^{n} \frac{1}{x_i}}$$

Beispiel: Teilstrecken L_i mit x_i Geschwindigkeit durchfahren, Zeiten t_i
Gesamtstrecke $L = \sum_{i=1}^{n} L_i$ und Gesamtzeit $t = \sum_{i=1}^{n} t_i$
Durchschnittsgeschwindigkeit:

$$\begin{aligned} x &= \frac{l}{t} = \frac{\sum l_i}{\sum t_i} = \frac{\sum l_i}{\sum \frac{l_i}{x_i}} = \\ &= \frac{1}{\sum_{i=1}^{n} \frac{l_i}{x_i \cdot \sum l_i}} = \frac{1}{\sum \frac{w_i}{l_i}} \qquad \text{mit } w_i = \frac{l_i}{\sum l_i} \end{aligned}$$

Sind die Daten in k Klassen gruppiert mit Klassenmitten z_j und Häufigkeiten H_j so gilt:

$$\bar{z}_{gh} := \frac{n}{\sum_{j=1}^{k} \left(\frac{1}{z_j} \cdot H_j \right)}$$

Empirischer Median

Für Einzeldaten $x_1, ..., x_n$ wird bei ungerader Anzahl, d.h. $n = 2 \cdot k + 1$ der mittlere Wert $x_{(k+1)}$ als Median definiert. Bei gerader Anzahl $n = 2 \cdot k$ ist er definiert als $\frac{x_{(k)} + x_{(k+1)}}{2}$.

Für gruppiert Daten ermittelt man das Summenpolygon. Der empirische Median $x_{0.5}$ ist definiert als das 0.5-Fraktile des Summenpolygons (vgl. Abbildung 2.1)

Abbildung 2.1: Empirischer Median

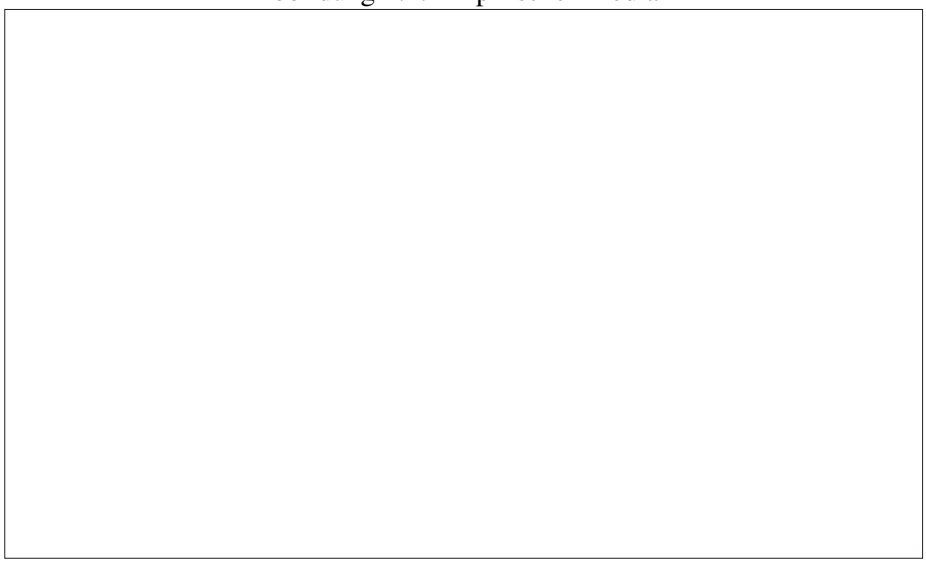


Abbildung 2.2: Empirischer Modus für diskrete, empirische Verteilungen

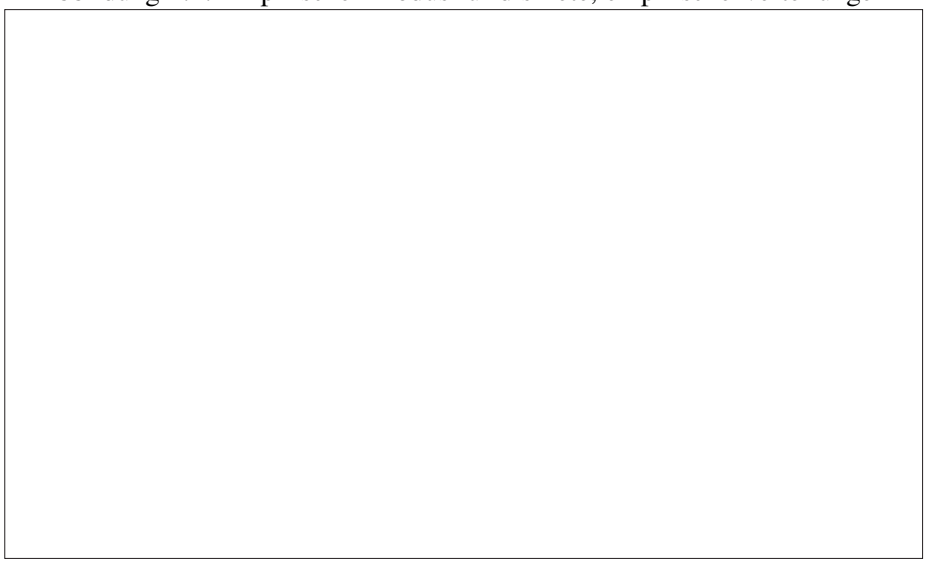
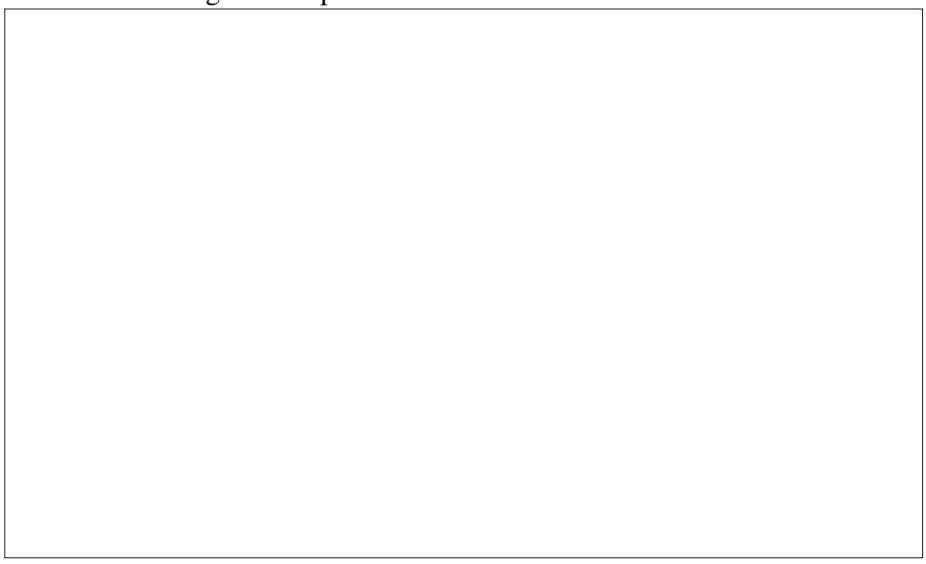


Abbildung 2.3: Empirischer Modus für kontinuierliche Größen



Empirischer Modus (Modalwert)

Für diskrete, empirische Verteilungen ist der empirische Modus als der häufigste Wert definiert, falls dieser existiert (vgl. Abbildung 2.2).

Für kontinuierliche Größen vgl. Abbildung 2.3.

2.4.2 Streuungsparameter

Spannweite

$$x_{(n)} - x_{(1)}$$

Quartilabstand

$$x_{0.75} - x_{0.25}$$

Bemerkung: Im Intervall $[x_{0.25}, x_{0.75}]$ liefen 50% aller Daten

Mittlere absolute Abweichung

$$MAD := \frac{1}{n} \sum_{i=1}^{n} |x_i - x_{0.5}|$$

Bemerkung: Berechnet man die Summe $\sum_{i=1}^{n} |x_i - a|$ für ein beliebiges $a \in \mathbb{R}$, so gilt:

$$\sum_{i=1}^{n} |x_i - x_{0.5}| \leq \sum_{i=1}^{n} |x_i - a| \qquad \forall a \in \mathbb{R}$$

Bei gruppierten Daten mit k Gruppen und den Gruppenmitten z_j sowie den Häufigkeiten H_j gilt:

$$MAD := \frac{1}{n} \sum_{j=1}^{n} |z_j - z_{0.5}| \cdot H_j$$

Empirische Varianz

Bei vollständigen Daten gilt:

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

Bei unvollständigen Daten gilt:

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

Bei Stichproben erfolgt die Division durch $n - 1$ zwecks Unverzerrtheit. Es gilt:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 \leq \sum_{i=1}^{n} (x_i - a)^2 \qquad \forall a \in \mathbb{R}$$

Für gruppierte Daten mit Gruppenmitten z_j und Häufigkeiten H_j, $j = 1(1)k$ gilt:

$$S_g^2 = \frac{1}{n} \sum_{j=1}^{k} (z_j - \bar{z})^2 \cdot H_j = \sum_{j=1}^{k} (z_j - \bar{z})^2 \cdot h_j$$

Die empirische Streuung bzw. empirische Standardabweichung ist definiert als:

$$S := +\sqrt{S^2}$$

Empirischer Variationskoeffizient (dimensionslos)

$$VK = \frac{S}{\bar{x}}$$

2.5 Schiefe und Exzess eindimensionaler, empirischer Verteilungen

Die Abweichung einer „Verteilung“ von der symmetrischen Form wird manchmal durch die sog. Schiefe beschrieben. Für gruppierte Daten gilt:

$$\gamma_1 = \frac{1}{n \cdot S^3} \sum_{j=1}^{k} (z_j - \bar{z})^3 \cdot H_j$$

Der Wert von γ_1 kann positiv oder negativ werden:

rechtschief	$\gamma_1 > 0$
linksschief	$\gamma_1 > 0$
symmetrisch	$\gamma_1 = 0$

Es Maß für die „Spitzigkeit“ (Stauchung bzw. Überhöhung) der Verteilungsform ist der sog. Exzess:

$$\gamma_2 = \left[\frac{1}{n \cdot S^4} \sum_{j=1}^{k} (z_j - \bar{z})^4 \cdot H_j \right] - 3$$

Auch γ_2 kann positive oder negative Werte annehmen.

überhöht	$\gamma_2 > 0$
gestaucht	$\gamma_2 < 0$
Ähnlich einer Normalverteilung	$\gamma_2 = 0$

Bemerkung: Pearson-Diagramm: $\beta_2 = \gamma_2 + 3$, $\beta_1 = \gamma_1$ (vgl. Abschnitt 3)

2.6 2-Dimensionale Häufigkeitsverteilungen

Zur Untersuchung von Zusammenhängen zwischen 1-dimensinalen Merkmalen. Je nach der Art der Größen X und Y werden verschiedene Zusammenhangsmaße verwendet.

2.6.1 Kontingenzdaten

X und Y sind qualitativer Art, d.h. diskret aber nicht der Größe nach zu ordnen.

X l Ausprägungen
Y m Ausprägungen

Daten: Paare $(x_k,\, y_k)$, $k = 1(1)n$

x \ y	b_1	b_2	b_j	b_m	$n_{i\bullet} = \sum_{j=1}^{m} n_{ij}$
a_1	n_{11}	n_{12}	n_{1j}	n_{1m}	$n_{1\bullet}$
a_2	n_{21}	n_{22}	n_{2j}	n_{2m}	$n_{2\bullet}$
a_i	n_{i1}	n_{i2}	n_{ij}	n_{im}	$n_{i\bullet}$
a_l	n_{l1}	n_{l2}	n_{lj}	n_{lm}	$n_{l\bullet}$
$n_{\bullet j} = \sum_{i=1}^{l} n_{ij}$	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet j}$	$n_{\bullet m}$	n

n_{ij}=Anzahl jener Paare mit Ausprägung $x_k = a_i$, $y_k = b_j$

Bemerkung: Für Abhängigkeitsuntersuchungen.

Beispiel: Vier-Felder-Tafel

x \ y	b_1	b_2	
a_1	n_{11}	n_{12}	$n_{1\bullet}$
a_2	n_{21}	n_{22}	$n_{2\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	n

Der sog. Yule´sche Assoziationskoeffizient ist definert als:

$$Q = \frac{n_{11} \cdot n_{22} - n_{12} \cdot n_{21}}{n_{11} \cdot n_{22} + n_{12} \cdot n_{21}} \qquad -1 \leq Q \leq 1$$

2.6.2 Empirischer Korrelationskoeffizient

X und Y sind metrische Merkmale.

$$r := \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right] \cdot \left[\sum_{i=1}^{n} (y_i - \bar{y})^2\right]}}$$

Bemerkung:

- $-1 \leq r \leq 1$
- r ändert sich nicht, wenn man anstelle der Wertepaare $(x_i,\, y_i)$ Lineartransformationen $(u_i,\, v_i)$ mit $u_i = \frac{x_i - x_0}{c}$ und $v_i = \frac{y_i - y_0}{c}$ verwendet.
- Falls $|r| = 1$, so liegen alle Paare $(x_i,\, y_i)$ auf einer Geraden in der $(x,\, y)$-Ebene.

$r = +1$ Positive lineare Abhngigkeit.
Falls x steigt, so steigt auch y.

$r = -1$ Negative lineare Abhngigkeit.
Falls x steigt, so sinkt y.

$r = 0$ Keine lineare Abhngigkeit.
Bedeutet NICHT keine Abhngigkeit.

2.7 Indexzahlen

2.7.1 Einfache Indexzahlen

Statistische Kennzahlen, die nebengeordnete Größen auf eine von ihnen oder auf einen Durchschnitt beziehen.

- Meßzahlen des sachlichen Vergleiches.
- Meßzahlen des örtlichen Vergleiches.
- Meßzahlen des zeitlichen Vergleiches.

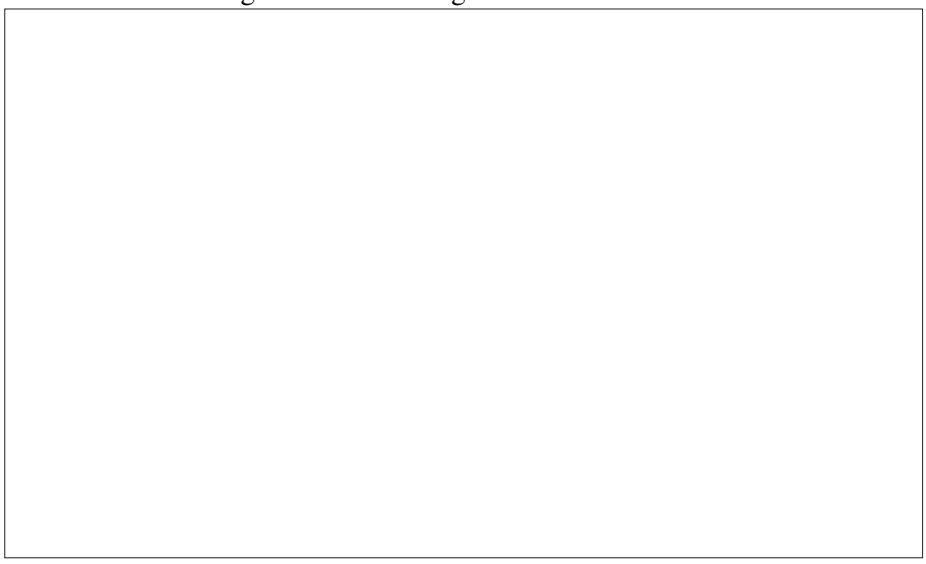
2.7.2 Zeitreihen

Zeitlich geordnete Folge statistische Beobachtungen $(A_t,\, t = 0, 1, 2, ...)$. Darauf ermittelt man eine Folge von Meßzahlen.

Zeitreihe der Meßzahlen mit Basisjahr 0:

$$I_{0t} := \frac{A_t}{A_0}; \quad t = 0, 1, 2, 3, ...$$

Abbildung 2.4: Meßzahlen gewonnen aus einer Zeitreihe



2.7.3 Indexzahlen (zusammengesetzte Indizes)

Hat man mehrere sachlich zusammenhängende Reihen von Daten, so möchte man deren Verlauf oft jeweils durch eine Zahl beschreiben. Dies erfolt durch sog. Indexzahlen.

Wertindex: Aussage über die relative Änderung des Wertes eines Bündels von Waren oder Dienstleistungen.

$1, ..., n$	Warennummern der n Waren des Warenkorbes
$p_{t1}, ..., p_{tn}$	Preise der Waren zu Berichtsperiode t
$q_{t1}, ..., q_{tn}$	Mengen der Waren zu Berichtsperioden t

Mit der Basisperiode $t = 0$ ergibt sich der Wert des Warenkorbes in der Periode t als:

$$\sum_{i=1}^{n} p_{ti} \cdot q_{ti}$$

Der Wertindex ist definiert als:

$$I_{0t}^{W} := \frac{\sum_{i=1}^{n} p_{ti} \cdot q_{ti}}{\sum_{i=1}^{n} p_{0i} \cdot q_{0i}}$$

Der Preisindex nach Paasche ist definiert als:

$$ {}_pI_{0t}^P := \frac{\sum_{i=1}^n p_{ti} \cdot q_{ti}}{\sum_{i=1}^n p_{0i} \cdot q_{ti}} $$

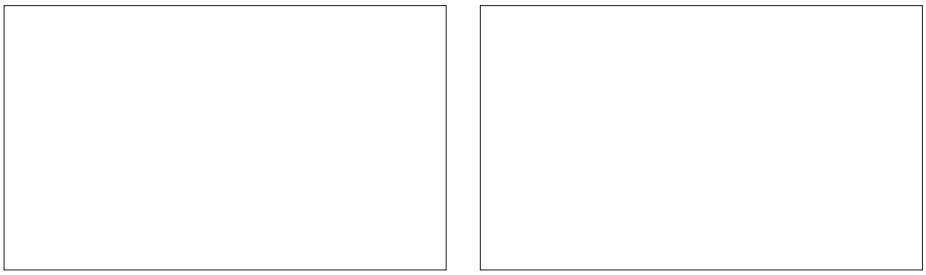
Der Preisindex nach Laspeyres (ÖSTAT) ist definiert als:

$$ {}_LI_{0t}^P := \frac{\sum_{i=1}^n p_{ti} \cdot q_{0i}}{\sum_{i=1}^n p_{0i} \cdot q_{0i}} $$

3 Grafische Methoden zu Beurteilung empirischer Verteilungen

Zur Beurteilung der Verteilungsart von empirisch gegebenen Merkmalen (stoch. Größen) konstruiert man sog. Wahrscheinlichkeitspapiere (oder Netzte).

Dabei transformiert man die Darstellungsebene der Verteilungsfunktion so, dass die Bilder der entsprechenden Verteilungsfunktion in Geraden übergeführt werden.



$$(x,\, y) \qquad \rightarrow \qquad (t,\, z)$$

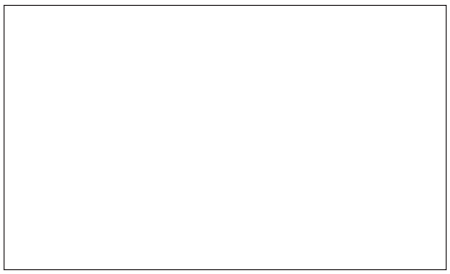
$t = \Phi(x) \qquad z = \psi(y)$

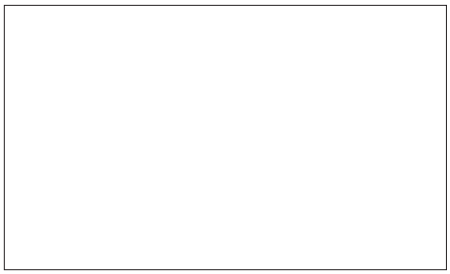
Das bekannteste solche Papier ist jenes für die Normalverteilung, genannt:

3.1 Normalverteilungsnetz

Ist $X \sim N(\mu,\, \sigma^2)$ mit Verteilungsfunktion $F(x) = W\{X \leq x\} = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad \forall x \in \mathbb{R}$, so wird folgende Transformation durchgeführt:

$$\begin{aligned} t &= x \\ z &= \Phi^{-1}(y) \\ u_p &= \text{p} - \text{Fraktile der } N(0,1) \end{aligned}$$





Es gilt: Ist x_p das p-Fraktile der $N(\mu, \sigma^2)$ so gilt folgender Zusammenhang:

$$u_p = \frac{x_p - \mu}{\sigma} = \frac{1}{\sigma} x_p - \frac{\mu}{\sigma}$$

Beweis:

$$p = W\{X \leq x_p\} = W\left\{\frac{x - \mu}{\sigma} \leq \frac{x_p - \mu}{\sigma}\right\}$$

3.2 Wahrscheinlichkeitsnetze für logarithmische normalverteilte Merkmale

Das Bild der Verteilungsfunktion von $\ln X$ ist im Normalverteilungsnetz eine Gerade. Teilt man die Abszisse logarithmisch, so kann man direkt die Werte von X eintragen. Die Ordinate ist wie beim Wahrscheinlichkeitspapier for Normalverteilungen geteilt (Fraktile der $N(0, 1)$).

Trägt man die Punkte der empirischen Verteilungsfunktion in das logarithmische Wahrscheinlichkeitsnetz ein, so deutet ein annähernd geradliniger Verlauf auf eine $LN(\mu, \sigma^2)$ hin.

3.3 Zuverlässigkeitsanalyse und Ausfallsraten

Lebensdauern X mit kontinuierlichen Verteilungen und Dichtefunktion $f(\cdot)$, $f(x) = 0 \quad \forall x < 0$. Gesucht ist die altersabhängige Ausfallswahrscheinlichkeit

$$W\{x \leq X \leq x + \Delta x \,|\, X > x\} = ?$$

Bemerkung: Bei Lebensdauern wird meist nicht die Verteilungsfunktion sondern die Zuverlässigkeitsfunktion

$$R(x) = W\{X > x\} = 1 - F(x) \quad \forall x >\geq 0$$

betrachtet. Damit erhält man für obige altersabhängige Ausfallswahrscheinlichkeit:

$$\begin{aligned} W\{x < X \leq x + \Delta x \,|\, X > x\} &= \frac{W\{x < X \leq x + \Delta x\}}{R(x)} \\ &= \frac{\int_x^{x+\Delta x} f(t)dt}{R(x)} \\ &= \frac{f(\xi) \cdot \Delta x}{R(x)} \end{aligned}$$

Für kleine Δx ist der letzte Ausdruck ungefähr $\frac{f(x)}{R(x)} \cdot \Delta x$, d.h. die altersabhängige Ausfallswahrscheinlichkeit ist nahezu proportional zu $\frac{f(x)}{R(x)}$. Diese altersabhängige Größe $\lambda(x) := \frac{f(x)}{R(x)}$ heißt Ausfallsrate.

Abbildung 3.1: Ausfallsrate

Beispiel: Für die Exponentialverteilung gilt:

$$\begin{aligned} f(x) &= \frac{1}{\tau} \cdot e^{-\frac{x}{\tau}} \mathrm{I}_{(0,\infty)}(x) \\ R(x) &= e^{-\frac{x}{\tau}} \quad \forall x > 0 \\ \lambda(x) &= \frac{1}{\tau} \qquad \text{konstant!} \end{aligned}$$

Für Ausfallsraten der Form $\lambda(x) = c \cdot x^\alpha$ verwendet man die Weibull-Verteilung.

3.4 Lebensdauer-Netz (Weibull-Verteilung)

$$X \sim \text{Wei}(\tau, \beta)$$

$$F(t) = W\{X \leq t\} = 1 - R(t) = 1 - e^{-\left(\frac{t}{\tau}\right)^{\beta}} \quad \forall x > 0$$

Abbildung 3.2: Weibull-Verteilung

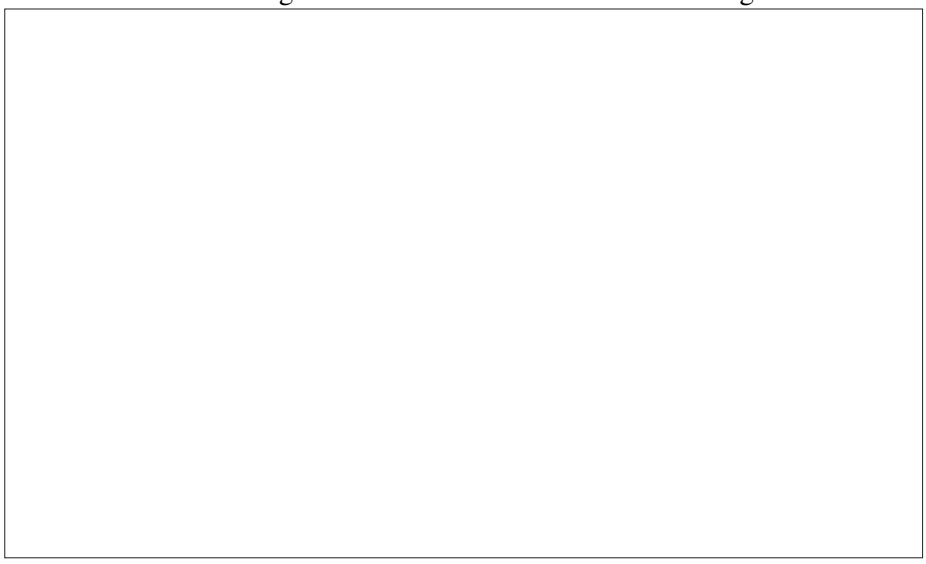
Die Transformation des $(t,\ z)$-Koordinatensystems in das (x, y)-Koordinatensystem geschieht folgendermaßen:

$$\begin{aligned} x &= \ln t \\ y &= \ln\left[\ln\left(\frac{1}{1-z}\right)\right] \end{aligned}$$

Es gilt: Die Bilder von Verteilungsfunktionen von Weibull-verteilten stochastischen Größen sind im $(x,\ y)$-Koordinatensystem Geraden, denn:

$$\begin{aligned} y &= \ln\left[\ln\left(\frac{1}{1-F(t)}\right)\right] = \\ &= \beta \cdot (\ln t - \ln \tau) = \\ &= \beta \cdot x - \beta \cdot \ln \tau \end{aligned}$$

Abbildung 3.3: Transformierte Weibull-Verteilung



3.5 Anpassung von Verteilungsmodellen

Für Modellrechnungen und Prognosen benötigt man „die“ Wahrscheinlichkeitsverteilung der entsprechenden stochastischen Größe.

Bemerkung: Dazu dienen Tests, Parameterschätzungen (Konfidenzbereiche) und Prognoseverteilungen. Auch das sogenannte Pearson-Diagramm ist eine Entscheidungshilfe.

Zur Kontrolle des sog. Pearson-Diagramms benötigt man analog zur Schiefe und Exzess empirischer Verteilungen für Wahrscheinlichkeitsverteilungen stochastische Größen.

Definition: Ist X eine stochastische Größe mit existierendem $\mathbb{E}(X^r)$, so schreibt man:

$$\begin{aligned} \mu_r^{'} &= \mathbb{E}(X^r) \\ \mu_r &= \mathbb{E}[(X - \mathbb{E}X)^r] \\ \mu_1^{'} &= \mu = \mathbb{E}(X) \\ \mu_2 &= \text{Var}X \end{aligned}$$

Vergleiche Abschnitt 2 (Schiefe, Spitzigkeit).

Schiefe: Es gilt:

$$\begin{aligned} \mu_3 &:= \mathbb{E}(X - \mu)^3 = \mu_3^{'} - 3 \cdot \mu_2^{'}\mu + 2 \cdot \mu^3 \\ \sqrt{\beta_1} &:= \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}} \text{ ... Ma fr die Schiefe relativ zu Spannweite} \end{aligned}$$

Beispiel: Exponentialverteilung $\sqrt{\beta_1} = 2$

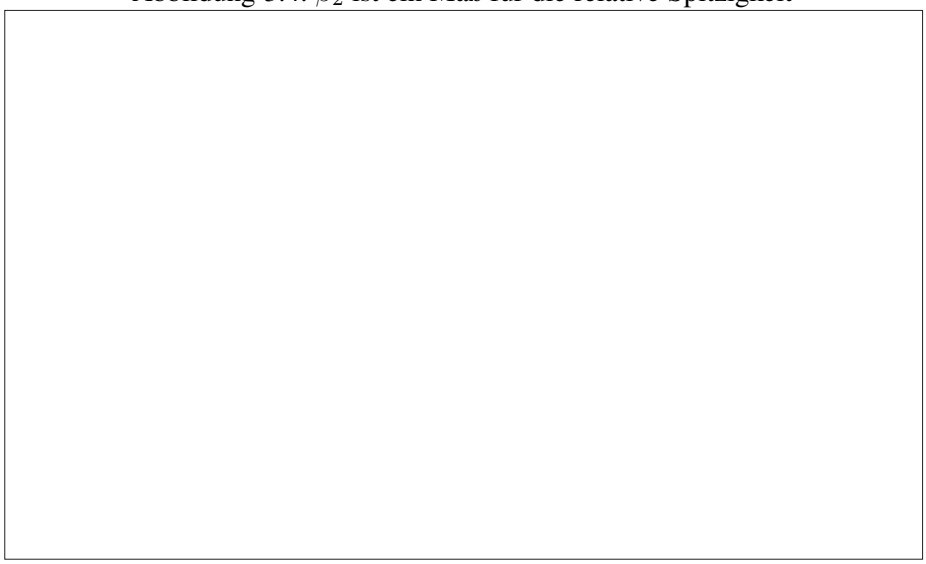
Für unimodale (1 Maximum) Verteilungen ist μ_4 ein Maß für die Spitzigkeit:

$$\beta_2 := \frac{\mu_4}{\mu_2} \text{ ... Ma fr die relative Spitzigkeit (Wlbung)}$$

Exzess: $\gamma_2 = \beta_2 - 3$

Beispiel: Normalverteilung, Gleichverteilung (vgl. Abbildung 3.4)

Abbildung 3.4: β_2 ist ein Maß für die relative Spitzigkeit



Trägt man die Werte $(\beta_1,\ \beta_2)$ in einem Diagramm auf, so erhält man Charakteristika für Verteilungstypen: sog. Pearson-Diagramm.

Definition: Ist $X_1, ..., X_n$ eine Stichprobe einer Verteilung W, so heißt $M_r^{'} := \frac{1}{n}\sum_{i=1}^{n} X_i^r$ das r-te Stichprobenmoment bezüglich 0 und $M_r := \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^r$ das r-te Stichprobenmoment bezüglich $\bar{X}_n$.

Satz 3.1 Ist $X_1, ..., X_n$ eine Stichprobe von X mit $\exists\mu_r^{'}$ und $\exists\mu_{2r}^{'}$, so folgt:

- $\mathbb{E}M_r^{'} = \mu_r^{'}$ (unverzerrte Schätzfunktion)
- $\text{Var}M_r^{'} = \frac{1}{n}\cdot\left[\mathbb{E}\left(X^{2r}\right)\cdot\mathbb{E}\left(X^r\right)^2\right] = \frac{1}{n}\cdot\left[\mu_{2r}^{'} - \left(\mu_r^{'}\right)^2\right]$

Korollar 3.1 Ist $X_1, ..., X_n$ eine Stichprobe von X und $\exists\text{Var}X = \sigma^2$ so folgt:

$$\mathbb{E}\bar{X}_n = \mathbb{E}X \qquad \text{und} \qquad \text{Var}\bar{\text{X}}_\text{n} = \frac{\sigma^2}{n}$$

Beweis: Sonderfall r=1

Anwendung des Pearsondiagrammes

Aus der Stichprobe $x_1, ..., x_n$ schätzt man die entsprechenden Momente. Daraus berechnet man Näherungswerte b_1 und b_2 für β_1 und β_2 und trägt das Paar $(b_1,\, b_2)$ in das Pearsondiagramm ein. Aus der Lage des Punktes erhält man einen Anhaltspunkt für den Verteilungstyp.

3.6 Ergänzung zu Schätzungen

Für die Stichprobenvarianz

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

gilt:

- $S_n^2 = \frac{n}{n-1} \cdot M_2$
- $S_n^2 = \frac{1}{2n\cdot(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(X_i - X_j\right)^2$
  $\sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2 = \cdots = \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^{n} X_i\right)^2$
  $\frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(X_i - X_j\right)^2 = -\|-$

Satz 3.2 Ist $X_1, ..., X_n$ eine Stichprobe von X mit $\exists \mathbb{E}X^4$ und $\exists \mathrm{Var}X = \sigma^2$ so gilt:

- $\mathbb{E}S_n^2 = \sigma^2$
- $\mathrm{Var}S_n^2 = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)$ fr $n > 1$

Satz 3.3 Ist X eine 1-dimensionale stochastische Größe mit endlicher Varianz σ^2 und $X_1, ..., X_n$ eine Stichprobe von X, so ist

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

asymptotisch $N(0, 1)$-verteilt.

Beweis: Zentraler Grenzverteilungssatz

Bemerkung: $\bar{X}_n$ ist näherungsweise normalverteilt mit Mittel μ und der Varianz $\left(\frac{\sigma^2}{n}\right)$.

Beispiel: $X \sim Ex_1,\ f(x) = e^{-x}\mathrm{I}_{(0,\,\infty)}(x)$

Anwendung Anteilsschätzung

Gesucht ist ein approximatives Konfidenzintervall für den Parameter θ einer Alternativverteilung. Für

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

gilt nach dem Satz von MOIVRE:

$$W\left\{\frac{\bar{X}_n - \theta}{\sqrt{\theta\,(1-\theta)\,/n}} \leq x\right\} \rightarrow_{n\rightarrow\infty} \Phi(x)$$

Dies ist ein Sonderfall des zentralen Grenzverteilungssatzes. Der Ausdruck

$$\frac{\bar{X}_n - \theta}{\sqrt{\theta(1-\theta)/n}}$$

ist approximativ $\mathrm{N}(0,\,1)$-verteilt. Das bedeutet:

$$\bar{X}_n \dot{\sim} \mathrm{N}\left(\mathbb{E}\bar{X}_n = \theta,\, \mathrm{Var}X = \frac{\theta(1-\theta)}{n}\right)$$

$$\mathrm{W}\left\{-u_{\frac{1+\gamma}{2}} \leq \frac{\bar{X}_n - \theta}{\sqrt{\theta(1-\theta)/n}} \leq +u_{\frac{1+\gamma}{2}}\right\} \doteq \gamma$$

mit u_α=α-Fraktile der $\mathrm{N}(0,\,1)$. Man kann diese Doppelungleichung umschreiben als

$$(X_n - \theta)^2 \leq \frac{\theta(1-\theta)}{n} u^2_{\frac{1+\gamma}{2}}$$

$$\left(\bar{X}_n - \theta\right)^2 - \frac{\theta(1-\theta)}{n} \cdot u^2_{\frac{1+\gamma}{2}} \leq 0$$

Nun sucht man die untere Nullstelle $\underline{\theta}\left(\bar{X}_n,\,\gamma\right)$ und die obere Nullstelle $\bar{\theta}\left(\bar{X}_n,\,\gamma\right)$:

$$\begin{aligned}
\underline{\theta}\left(\bar{X}_n,\,\gamma\right) &= \frac{2n\bar{X}_n + u^2 - u\cdot\sqrt{4n\bar{X}_n - 4n\bar{X}_n^2 + u^2}}{2\,(n+u^2)} \\
\bar{\theta}\left(\bar{X}_n,\,\gamma\right) &= \frac{2n\bar{X}_n + u^2 + u\cdot\sqrt{4n\bar{X}_n - 4n\bar{X}_n^2 + u^2}}{2\,(n+u^2)}
\end{aligned}$$

Division durch $2n$ und Vernachlässigung aller Glieder der Form $\frac{c}{n}$ ergibt als Näherung das approximative Konfidenzintervall mit Überdeckungswahrscheinlichkeit γ für θ

$$\left[\bar{X}_n - u_{\frac{1+\gamma}{2}}\sqrt{\frac{\bar{X}_n\left(1-\bar{X}_n\right)}{n}},\; \bar{X}_n - +_{\frac{1+\gamma}{2}}\sqrt{\frac{\bar{X}_n\left(1-\bar{X}_n\right)}{n}}\right].$$

Bemerkung: Dabei ist nur jener Teil des Intervalles zu berücksichtigen, der im Intervall $[0,\ 1]$ liegt. Die Form des Konfidenzintervalles ist naheliegent, da $\bar{X}_n$ eine gute Schätzfunktion für den $\mathbb{E}X_n = \theta$ ist und

$$\frac{\bar{X}_n(1-\bar{X}_n)}{n}$$

eine Schätzfunktion für

$$\text{Var}\bar{X}_n = \frac{\theta(1-\theta)}{n}$$

ist.

Beispiel: Stichprobe von $X \sim A_\theta$ vom Umfang n=130 und $\sum_{i=1}^{n} x_i$=28 und $\gamma = 0.95$. Gesucht ist ein approximatives Konfidenzintervall für θ.

Bemerkung: Vergleich mit HPD-Intervall für $\pi(\cdot)\hat{=}\text{U}_{0,1} = \text{Be}(1,\ 1)$ (vgl. Abschnitt 4).

4 Bayes´sche Anteilsschätzung

In der Bayes´schen Statistik werden alle unbekannten Größen durch stochastische Größen mit zugehöriger Wahrscheinlichkeitsverteilung beschrieben. Für ein stochastisches Modell $X \sim f(\cdot|\theta)$ mit Parameterraum Θ die Unsicherheit bezüglich des Parameters θ, der durch eine stochastiches Größe $\tilde{\theta}$ beschrieben wird, mittels einer sog. a-priori-Verteilung $\pi(\theta)$ von $\tilde{\theta}$ ausgedrückt.

Beispiel: $\pi(\cdot)\hat{=}\mathrm{U}_{0,1} \quad \Rightarrow \quad \pi(\theta) = \mathrm{I}_{(0,1)}(\theta)$
Dies ist ein Sonderfall der Funktion

$$f(\theta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}\mathrm{I}_{(0,1)}(\theta) \quad \text{mit } \alpha = \beta = 1,$$

welches die Dichtefunktion der Beta-Verteilung $\mathrm{Be}(\alpha, \beta)$ darstellt.

Frage: Kann man durch eine Stichprobe (Daten D) bessere Informationen über θ erhalten?

Bemerkung: Die Aktualisierung erfolgt über das Bayes´sche Theorem (vgl. Abschnitt 4.4).

4.1 Gamma-Funktion

$$\Gamma(x) \qquad \forall x > 0$$

$$\Gamma(n+1) = n! \qquad \Gamma(1) = 1 \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

4.2 Gamma-Verteilung $\mathbf{Gam}(\alpha, \beta) \quad \alpha > 0,\ \beta > 0$

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1} \cdot e^{\frac{-x}{\beta}}}{\Gamma(\alpha) \cdot \beta^{\alpha}} \mathrm{I}_{(0,\infty)}(x)$$

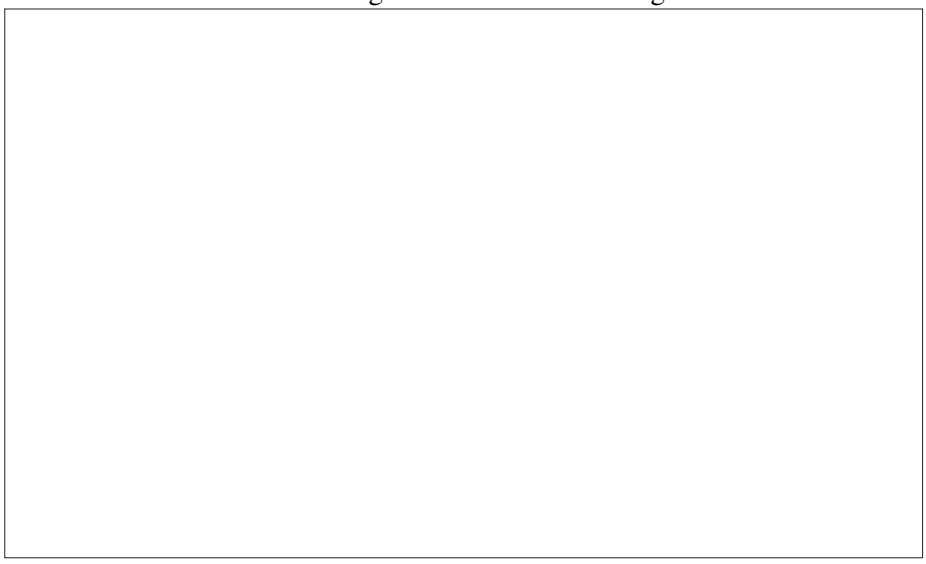
Falls $X \sim \mathrm{Gam}(\alpha, \beta)$ dann gilt:

$$\mathbb{E}X = \alpha \cdot \beta \qquad \mathrm{Var}X = \alpha \cdot \beta^2$$

Anwendung: Lebensdauern

Bemerkung: Die Gamma-Verteilung ist eine Verallgemeinerung der Exponential-Verteilung und der χ^2_n-Verteilung.

Abbildung 4.1: Gamma-Verteilung



Satz (Additionstheorem) für Gamma-Verteilungen

Falls $X_i \sim \text{Gam}(\alpha_i, \beta),\ i = 1, ..., n$ so gilt:

$$\sum_{i=1}^{n} X_i \sim \text{Gam}\left(\sum_{i=1}^{n} \alpha_i, \beta\right)$$

Anwendung: Lebensdauern von erneuerbaren Systemen (Kalte Reserven).

Bemerkung: Die Verteilungsfunktion zur Gamma-Verteilung ist durch die sog. unvollständige Gamma-Funktion gegeben:

$$F_X(x|\alpha, \beta) = W\{X \leq x\} = \int_o^x \frac{t^{\alpha-1} e^{-t/\beta}}{\Gamma(\alpha)\beta^\alpha} \mathrm{d}t$$

Diese Funktion ist tabelliert.

4.3 Beta-Verteilung 1.Art $\mathbf{Be}(\alpha, \beta) \quad \alpha > 0,\ \beta > 0$

$$\mathrm{f}(x|\alpha, \beta) \propto x^{\alpha-1} (1-x)^{\beta-1} \mathrm{I}_{(0,1)}(x)$$

Damit eine Dichtefunktion entsteht, bedient man sich einer normierenden Konstante:

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \mathrm{d}x =: \mathrm{B}(\alpha, \beta) \text{ ... Betafunktion}$$

Für B gilt:

$$\mathrm{B}(\alpha,\beta) = 2 \cdot \int_0^{\pi/2} \sin^{2\alpha-1}\varphi \cdot \cos^{2\beta-1}\varphi \mathrm{d}\varphi$$

Es gilt:

$$\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

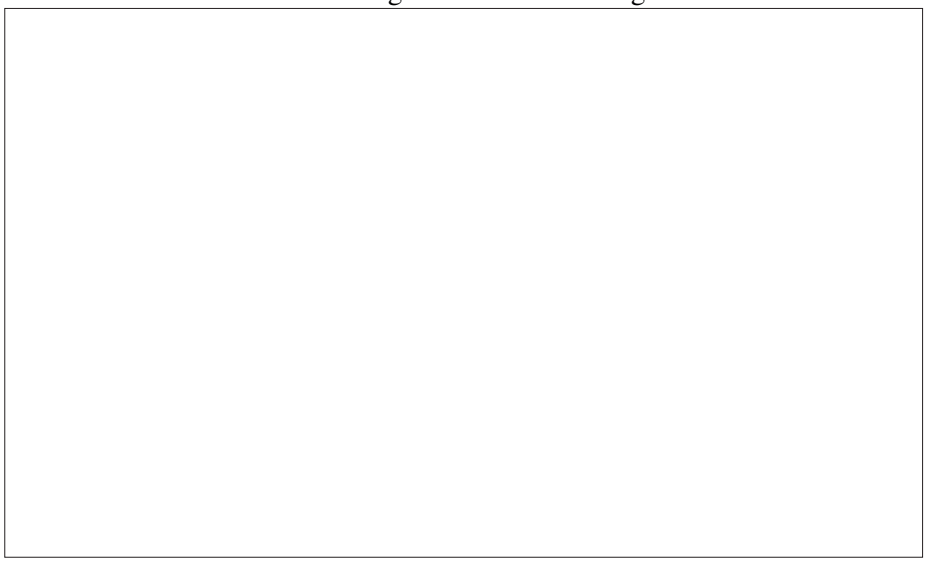
Bemerkung: Die Verteilungsfunktion der Be-Verteilung 1.Art ist durch die sog. unvollständige Beta-Funktion gegeben:

$$\mathrm{F}_X(x|\alpha,\beta) = \int_0^x \frac{t^{\alpha-1}(1-t)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\mathrm{d}t \quad 0 < x < 1$$

Diese Funktion ist tabelliert.

Anwendung: Anteilsschätzung im Fall einfacher Alternativen.

Abbildung 4.2: Beta-Verteilung



4.4 Bayes´sches Theorem

Dient dazu, um von der a-priori Einschätzung zur a-posteriori Einschätzung zu gelangen.

Vor Erhebung der Daten: a-priori Dichte $\pi(\theta)$, gemeinsame Dichte von $(X, \tilde{\theta})$:

$$g(x, \theta) = f(x|\theta) \cdot \pi(\theta)$$

Nach Erhebung *einer* Beobachtung x von X erhält man die durch $X = x$ bedingte Dichte von $\tilde{\theta}$:

$$\pi(\theta|x) := \frac{g(x, \theta)}{\int_\Theta g(x, \theta)\mathrm{d}\theta} = \frac{f(x|\theta) \cdot \pi(\theta)}{\int_\Theta f(x|\theta) \cdot \pi(\theta)\mathrm{d}\theta}$$

Für mehrere Beobachtungen $x_1, ..., x_n$ von X muß man die durch $X_1 = x_1$ bis $X_n = x_n$ bedingte Dicht von $\tilde{\theta}$ berechnen:

$$\pi(\theta|x_1, ..., x_n) = \frac{g(x_1, ..., x_n, \theta)}{\int_\Theta g(x_1, ..., x_n, \theta)\mathrm{d}\theta}$$

Für die gemeinsame Dichte $g(x_1, ..., x_n, \theta)$ von $(X_1, ..., X_n, \tilde{\theta})$gilt:

$$g(x_1, ..., x_n, \theta) = f_n(x_1, ..., x_n|\theta) \cdot \pi(\theta)$$

und im Fall einer vollständigen Stichprobe (unabhängig wie X verteilte X_i) gilt für die Dichte f_n von $(X_1, ..., X_n)$

$$f_n(x_1, ..., x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Damit folgt für die a-posteriori Dichte

$$\begin{aligned}
\pi(\theta|x_1, ..., x_n) &= \frac{[\prod_{i=1}^n f(x_i|\theta)] \cdot \pi(\theta)}{\int_\Theta \underbrace{\left[\prod_{i=1}^{n} f(x_i|\theta)\right]}_{l(\theta; x_1, ..., x_n)} \cdot \pi(\theta)\mathrm{d}\theta} \\
&= \frac{l(\theta; x_1, ..., x_n) \cdot \pi(\theta)}{\underbrace{\int_\Theta l(\theta; x_1, ..., x_n) \cdot \pi(\theta)\mathrm{d}\theta}_{c \in \mathbf{R}\,(\text{konstant})}} \\
&= c \cdot l(\theta; x_1, ..., x_n) \cdot \pi(\theta) \\
&\propto \pi(\theta) \cdot l(\theta; x_1, ..., x_n)
\end{aligned}$$

Bemerkung: Für nicht vollständige Daten D schreibt man

$$\pi(\theta|D) \propto \pi(\theta) \cdot l(\theta; D)$$

$$D = (x_1, ..., x_n; w_1, ..., w_n)$$

Dabei bezeichnet $w_1, ..., w_n$ sog. „Withdrawings“, z.B. zurückgezogene Flugzeugmotoren.

4.5 A-posteriori Dichte für einfache Anteile

Als a-priori Verteilung bietet sich die $\mathrm{Be}(\alpha, \beta)$ an, insbesondere $\mathrm{Be}(1,1) = \mathrm{U}_{0,1}$. Für die Punktwahrscheinlichkeiten der A_θ gilt

$$p(x|\theta) = \theta^x(1-\theta)^{1-x} \quad \text{fuer } x \in \{0,1\}.$$

Für n Beobachtungen sind die gemeinsamen Punktwahrscheinlichkeiten einer vollständigen Stichprobe vom Umfang n gegeben durch

$$\begin{aligned} p_n(x_1,...,x_n|\theta) &= \prod_{i=1}^{n} p(x_i|\theta) \\ &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \\ &= l(\theta; x_1,...,x_n). \end{aligned}$$

Für a-priori Verteilungen $\pi(\theta) \hat{=} \mathrm{Be}(\alpha, \beta)$ gilt

$$\pi(\theta) \propto \theta^{\alpha-1} \cdot (1-\theta)^{\beta-1} \mathrm{I}_{(0,1)}(\theta).$$

Aus dem Bayes´schem Theorem folgt:

$$\begin{aligned} \pi(\theta|x_1,...,x_n) &\propto \pi(\theta) \cdot l(\theta; x_1,...,x_n) \\ &\propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \theta^{\sum_{i=1}^n x_i} \cdot (1-\theta)^{n-\sum_{i=1}^n x_i} \\ &\propto \theta^{(\alpha+\sum_{i=1}^n x_i)-1} \cdot (1-\theta)^{(\beta+n-\sum_{i=1}^n x_i)-1} \\ &\hat{=} \mathrm{Be}(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i) \end{aligned}$$

Daraus berechnet man nun Punktschätzungen für den Anteil θ (falls ein solcher existiert).

Definition: Der a-posteriori Bayes-Schätzer $\hat{\theta}$ für θ ist der Erwartungswert von $\tilde{\theta}$ a-posteriori, d.h.

$$\hat{\theta} = \mathbb{E}_{\pi(\theta|\underline{x})}\tilde{\theta} = \int_o^1 \theta \cdot \pi(\theta|x_1,...,x_n)\mathrm{d}\theta.$$

Es gilt: Für $\theta \sim \mathrm{Be}\left(1 + \sum_{i=1}^n x_i,\ 1 + n - \sum_{i=1}^n x_i\right)$ ist

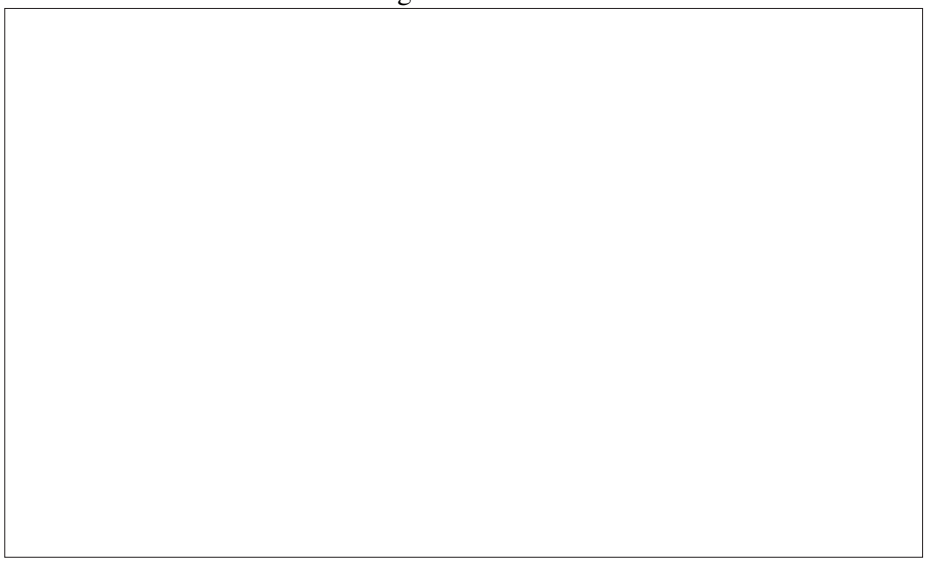
$$\mathbb{E}_{\pi(\theta|\underline{x})}\tilde{\theta} = \frac{\sum_{i=1}^n x_i + 1}{n+2}.$$

4.6 HPD-Intervalle für θ

Gegeben eine a-posteriori Dichte $\pi(\theta|D)$, $\theta \in \Theta$ ist ein HPD-Intervall mit Sicherheit $1-\alpha$ - bei Existenz - ein Intervall $\Theta^* \in \Theta$ mit:

- $\mathrm{W}_{\pi(\theta|D)}\left\{\tilde{\theta} \in \Theta^*\right\} = 1-\alpha$
- $\pi(\theta|D) \geq c \quad \forall \theta \in \Theta^*$, wobei c die größtmögliche derartige Konstante ist

Abbildung 4.3: HPD-Intervall



Bemerkung: Analog für mehrdimensionale Parameter.

Bemerkung: Ein (einfacher) Alternativversuch kann auf zwei Arten ausgehen, aber die Wahrscheinlichkeit der zweiten Art ist durch jene der ersten Art bestimmt.

$$\theta = \mathrm{W}\{X = 1.\text{Art}\} \qquad 1-\theta = \mathrm{W}\{X = 2.\text{Art}\}$$

Auch darstellbar als Versuch, der auf zwei Arten ausgeht ist:

$$\mathrm{W}\{X = 1.\text{Art}\} = \theta_1 \qquad \mathrm{W}\{\mathrm{X} = 2.\text{Art}\} = \theta_2 \qquad \text{mit } \theta_1 + \theta_2 = 1$$

Führt man jetzt diesen Versuch n-mal durch und zählt danach Y_1 der Ausgänge erster Art als auch Y_2 der Ausgänge zweiter Art, so erhält man eine 2-dimensionale stochastische Größe.

$$(Y_1, Y_2) \sim \mathrm{M}_{n;\theta_1,\theta_2} \qquad \text{Multinomialverteilung}$$

Abbildung 4.4: HPD-Intervall für einen zweidimensionalen Parameter

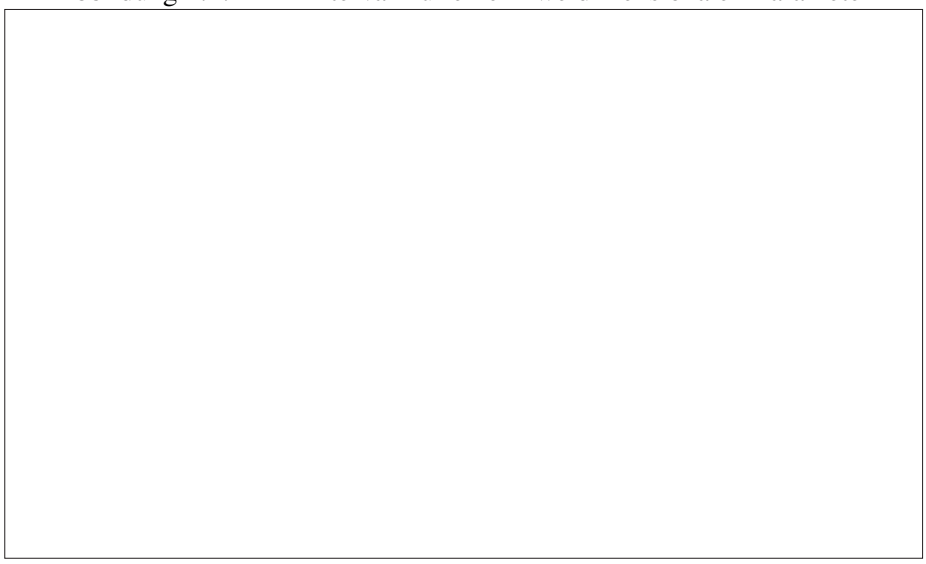
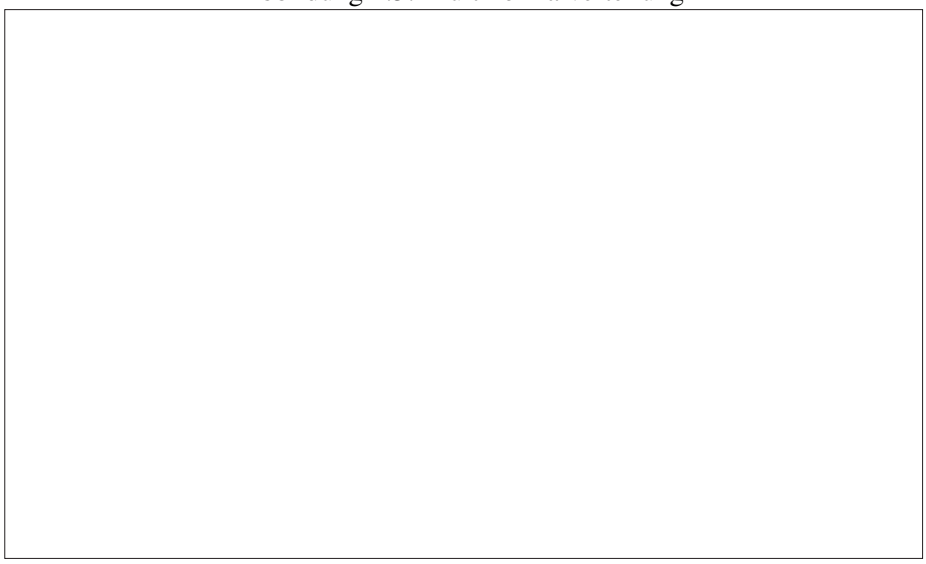


Abbildung 4.5: Multinomialverteilung



Bemerkung: Der Merkmalraum von (Y_1, Y_2) liegt auf einer Geraden in $\mathbf{R}$ (linearer Teilraum).

Es gilt:

$$\mathrm{W}\left\{\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right\} = \frac{n!}{y_1! \cdot y_2!} \cdot \theta_1^{y_1} \cdot \theta_2^{y_2}$$

mit $y_1 + y_2 = n$ und $\theta_1 + \theta_2 = 1$.

Bemerkung: Für den allgemeinen Fall „mehrfacher“ Alternativen (K mögliche Ausgänge eines Einzelversuches) ist die Beschreibung der a-priori Verteilung in etwas anderer Form nützlich.
Dazu betrachtet man (als 2-dimensionale Form der $\mathrm{Be}(\alpha_1, \alpha_2)$) die sog. Dirichlet-Verteilung (2-dimensional) mit Dichte

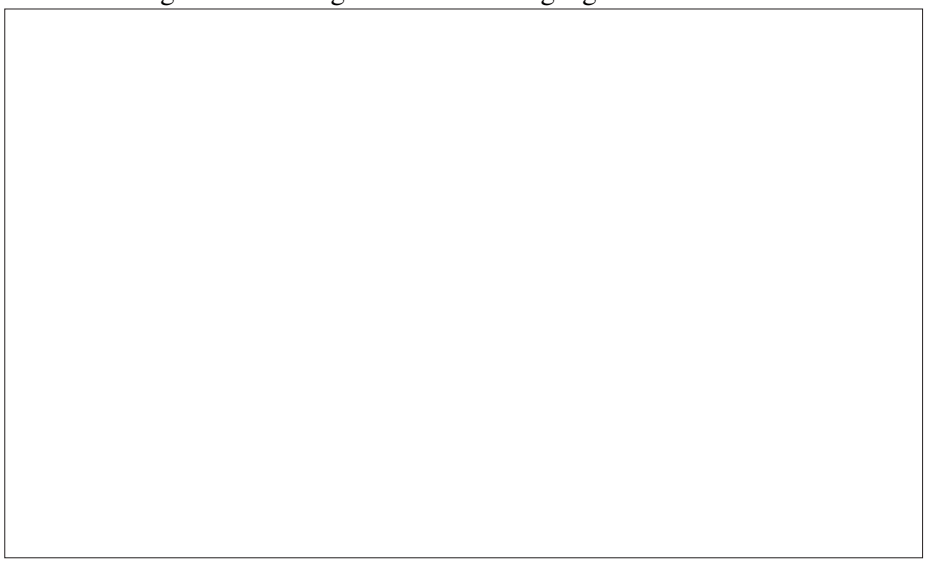
$$\pi(\theta_1, \theta_2 | \alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \cdot \Gamma(\alpha_2)} \cdot \theta_1^{\alpha_1 - 1} \cdot \theta_2^{\alpha_2 - 1}$$

für $\theta_i \geq 0$ und $\theta_1 + \theta_2 = 1$ und $\alpha_i = 0$.

Abbildung 4.6: Dirichlet-Verteilung

Bemerkung: Im Fall einfacher Alternativen betrachtet man meist das 1-dimensionale Problem Y_1 bzw. θ_1.
Im Fall von drei möglichen Versuchsausgängen eines Einzelversuches mit den entsprechenden Wahrscheinlichkeiten $\theta_1, \theta_2, \theta_3$ mit $\theta_1 + \theta_2 + \theta_3 = 1$ ist der Merkmalraum von $\left(\tilde{\theta}_1, \tilde{\theta}_2 \tilde{\theta}_3\right)$ eine Ebene des $\mathbf{R}^3$.

Abbildung 4.7: Drei mögliche Versuchausgänge eines Einzelversuches



4.7 Bayes´sche Anteilsschätzung bei mehrfachen Alternativen

Ein Einzelversuch kann auf K Arten ausgehen mit W $= \theta_j$, $k = 2$, $\sum_{j=1}^{k} \theta_j = 1$. Bei n-facher Durchführung eines solchen Versuches erhält man $X_1, ..., X_n$. Danach stellt man fest, wie oft die k verschiedenen Arten beobachtet wurden:

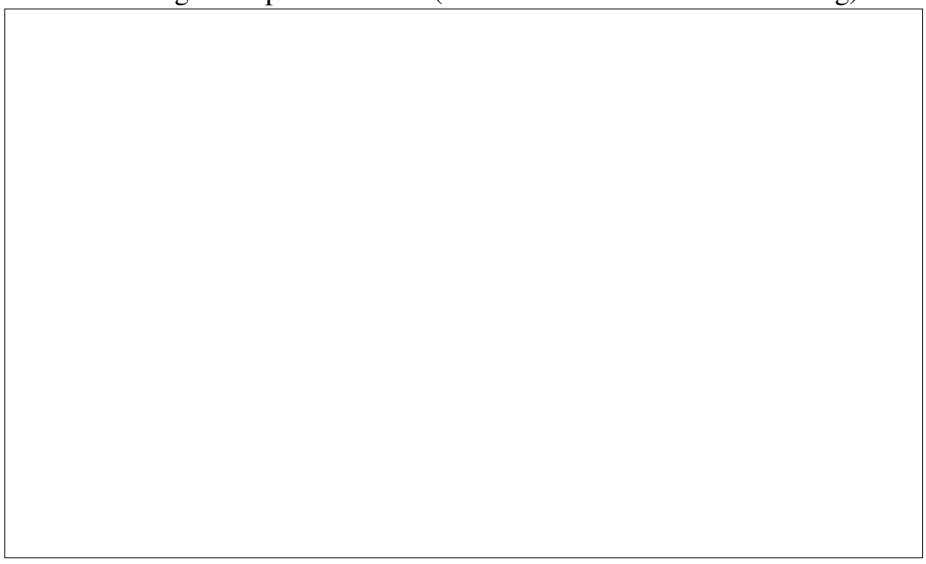
$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix} \qquad Y_j = \text{Anzahl der Ausgaenge auf die } j - \text{te Art}$$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix} \sim \mathrm{M}_{n;\theta_1,...,\theta_k} \qquad \text{Multinomialverteilung}$$

$$\mathrm{W}\left\{ \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} \right\} = \frac{n!}{y_1! \cdot ... \cdot y_k!} \cdot \theta_1^{y_1} \cdot ... \cdot \theta_k^{y_k} \qquad \text{fuer } \begin{cases} \sum_{j=1}^{k} y_j = n \\ y_j \in \mathbf{N}_0 \end{cases}$$

Bemerkung: Der Merkmalraum von diesem stochastischen Vektor $(Y_1, ..., Y_k)^T$ ist ein Teilraum einer $(k-1)$-dimensionalen Hyperebene des $\mathbf{R}^k$.

Abbildung 4.8: Spezialfall k=3 (2-dimensionale diskrete Verteilung)



Spezialfall k=3: Wegen $y_3 = n - y_1 - y_2$ kann dies als zweidimensionales Problem behandelt werden.

Als a-priori Verteilung für $\tilde{\theta} = \left(\tilde{\theta}_1, ..., \tilde{\theta}_k\right)$ verwendet man die k-dimensionale Dirichlet-Verteilung Dir$(\alpha_1, ..., \alpha_k)$.

$$\pi(\theta_1, ..., \theta_k | \alpha_1, ..., \alpha_k) = \frac{\Gamma\left(\sum_{j=1}^k \alpha_j\right)}{\prod_{j=1}^k \Gamma(\alpha_j)} \cdot \prod_{j=1}^k \theta_j^{\alpha_j - 1}$$

$$\begin{cases} \alpha_j \geq 0 \\ \theta_j \geq 0 \\ \sum_{j=1}^k \theta_j = 1 \end{cases}$$

Der Merkmalraum von $\left(\tilde{\theta}_1, ..., \tilde{\theta}_k\right)$ ist eine Teilmenge einer $(k-1)$-dimensionalen Hyperebene des $\mathbf{R}^k$. Die Parameter $\alpha_1, ..., \alpha_k$ der a-priori Verteilung heißen Hyperparameter.
Die k-dimensionale uniforme Verteilung auf der $(k-1)$-dimensionalen Hyperebene $\sum_{j=1}^k \theta_j = 1$ von $\Theta \subseteq \mathbf{R}^k$ enthält man mit $\alpha_1 = \alpha_2 = ... = \alpha_k = 1$.

Spezialfall k=3:

$$\pi(\theta_1, \theta_2, \theta_3 | \alpha_1, \alpha_2, \alpha_3) > 0 \quad \text{nur fuer} \quad \theta_1 + \theta_2 + \theta_3 = 1$$

Für die Aktualisierung der Verteilung $\pi(\theta_1, ..., \theta_k | \alpha_1, ..., \alpha_k)$ nach Beobachtung einer Stichprobe $x_1, ..., x_n$ und den entsprechenden Werten $y_1, ..., y_k$ verwendet man das Bayes´sche Theorem:

$$\pi(\theta_1, ..., \theta_k | \alpha_1, ..., \alpha_k) \propto \prod_{j=1}^k \theta_j^{\alpha_j - 1}$$

$$
\begin{aligned}
l(\theta_1, ..., \theta_k; y_1, ..., y_k) &\propto \prod_{j=1}^{k} \theta_j^{y_j} \\
\pi(\theta_1, ..., \theta_k | D) &\propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1} \cdot \prod_{j=1}^{k} \theta_j^{y_j} \\
&\propto \prod_{j=1}^{k} \theta_j^{\alpha_j + y_j - 1} \\
\pi(\theta_1, ..., \theta_k | D) &\hat{=} \text{Dir}(\alpha_1 + y_1, ..., \alpha_k + y_k)
\end{aligned}
$$

Spezialfall k=3:

$$\pi(\theta_1, \theta_2, \theta_3 | D) \hat{=} \text{Dir}(\alpha_1 + y_1, \alpha_2 + y_2, \alpha_3 + y_3)$$

Daraus ermittelt man eine Verteilung für $\left(\tilde{\theta}_1, \tilde{\theta}_2\right)$ durch $\theta_3 = 1 - \theta_1 - \theta_2$.

Beispiel: Meinungsumfrage Hainburg

- a-priori von $\theta_1 =$ pro, $\theta_2 =$ contra, $\theta_3 =$ unentschlossen
- a-posteriori Dichte
- HPD-Bereiche
- W $\{\text{pro} > \text{contra}\}$

4.8 Konjugierte Verteilungsfamilien

Anteilsschätzung:

- a-priori Be(α, β)
- a-posteriori Be $(\alpha + \sum_{i=1}^{n} x_i,\ \beta + n - \sum_{i=1}^{n} x_i)$

Analog bei mehrfachen Alternativen:

- a-priori Dir$(\alpha_1, ..., \alpha_k)$
- a-posteriori Dir$(\alpha_1 + y_1, ..., \alpha_k + y_k)$

Allgemeiner für die stochastischen Modelle un zugehörigen a-priori Familien.

Definition: Ist $X \sim \text{W}_\theta$, $\theta \in \Theta$ ein stochastisches Modell und $\mathcal{P}$ eine Familie von a-priori Verteilungen für $\tilde{\theta}$, sodaß für beliebige Stichproben von X die a-posteriori Verteilung von $\tilde{\theta}$ wieder ein Element von $\mathcal{P}$ ist, so spricht man von konjugierten Verteilungsfamilien.

Beispiel: $X \sim \text{Ex}_\tau$, $\tau > 0$; $\mathcal{P} = \{\text{Gam}(\alpha, \beta) : \alpha > 0, \beta > 0\}$

Beispiel: $X \sim \text{P}_\mu$, $\mu > 0$; $\mathcal{P} = \{\text{Gam}(\alpha, \beta) : \alpha > 0, \beta > 0\}$

Bemerkung: Man spricht auch von „einer zu einem stochastischen Modell konjugierten a-priori Familie“.

4.9 Suffizienz

Eine aus einer Stichprobe $X_1, ..., X_n$ berechnete Statistik $s(X_1, ..., X_n)$ heißt suffizient für einen Parameter θ der Verteilung von X, falls zur Berechnung der a-posteriori Verteilung $\pi(\theta|x_1, ..., x_n)$ die Kenntnis von $s(x_1, ..., x_n)$ hinreicht.

Beispiel: $\sum_{i=1}^n X_i$ für die Alternativ-Verteilung

Beispiel: $(Y_1, ..., Y_k)$, $\text{M}_{n;\theta_1,...,\theta_k}$

$$\text{W}\left\{ \left(\begin{array}{c} Y_1 \\ \vdots \\ Y_k \end{array} \right) = \left(\begin{array}{c} y_1 \\ \vdots \\ y_k \end{array} \right) \middle| n, \theta_1, ..., \theta_k \right\} = \text{p}(y_1, ..., y_k|\theta_1, ..., \theta_k) =$$

$$= \frac{n!}{y_1! \cdot ... \cdot y_k!} \cdot \theta_1^{y_1} \cdot ... \cdot \theta_k^{y_k} \cdot \text{I}_{\{\sum_{j=1}^k y_j = n\} \bigcap \{y_j \in \mathbf{N}_0\}} (y_1, ..., y_k)$$

Dazu konjugiert die Dirichlet-Verteilung

$$\begin{aligned} \text{f}(x_1, ..., x_k|\alpha_1, ..., \alpha_k) &= \frac{\Gamma(\alpha_1 + ... + \alpha_k)}{\Gamma(\alpha_1) \cdot ... \cdot \Gamma(\alpha_k)} \cdot x_1^{\alpha_1 - 1} \cdot ... \cdot x_k^{\alpha_k - 1} \cdot \text{I}_{D^*}(x_1, ..., x_n) \\ D^* &= \left\{ (x_1, ..., x_k) : x_i \geq 0 \bigwedge \sum_{i=1}^k x_i = 1 \right\} \end{aligned}$$

$$\pi(\theta_1, ..., \theta_k|y_1, ..., y_k) = \text{Dir}(\alpha_1 + y_1, ..., \alpha_k + y_k)$$

In diesem Beispiel ist $(Y_1, ..., Y_k)$ eine suffiziente Statistik für $\underline{\theta} = (\theta_1, ..., \theta_k)$.

Bemerkung: Auch $(Y_1, ..., Y_{k-1})$ ist suffizient.

5 Statistische Qualitätskontrolle

Entscheidung, ob eine Warensendung (Los) genügend gute Stücke enthält.

N	Losumfang
n	Stichprobenumfang

Entscheidungsfindung:

- Risiko für Lieferanten und Käufer zumutbar
- Laufende Überwachung eines Produktionsvorganges

5.1 Einfache Stichprobenpläne

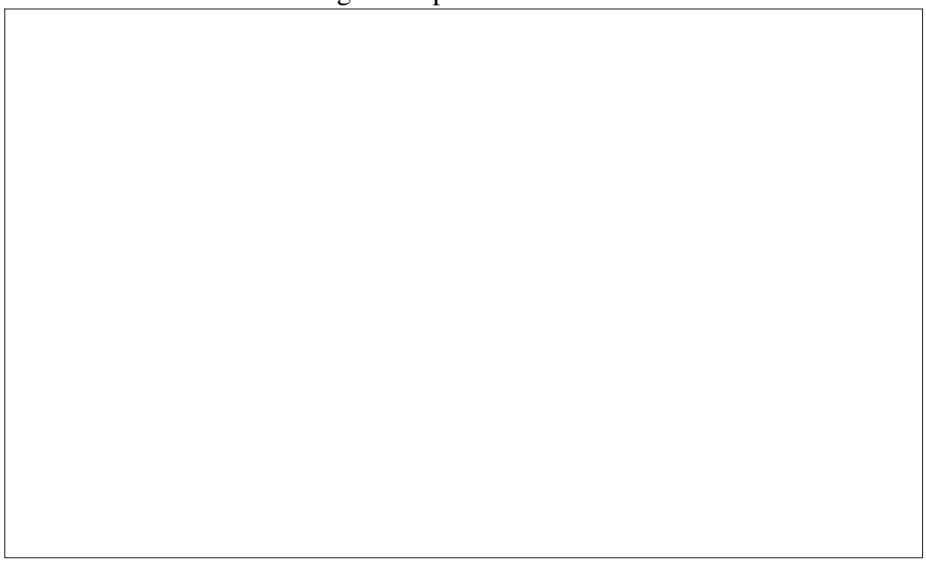
θ	Anteil der schlechten Stücke (bzw. A=Anzahl der schlechten Stücke)
x	Anzahl der schlechten Stücke in der Stichprobe
X	stochastische Größe, die x vor der Erhebung der Stichprobe beschreibt
(n, c)	c=Annahmekennzahl (wieviel schlechte Stücke maximal)

Entscheidung: Nehme Los an, falls $x \leq c$.

Bemerkung: c so zu wählen, dass gute Lose mit großer Wahrscheinlichkeit angenommen werden und schlechte Lose mit sehr kleiner Wahrscheinlichkeit angenommen werden.
Dazu dient die sog. „Operationscharakteristik“ OC; eine Funktion von θ (auch Kennkurve).

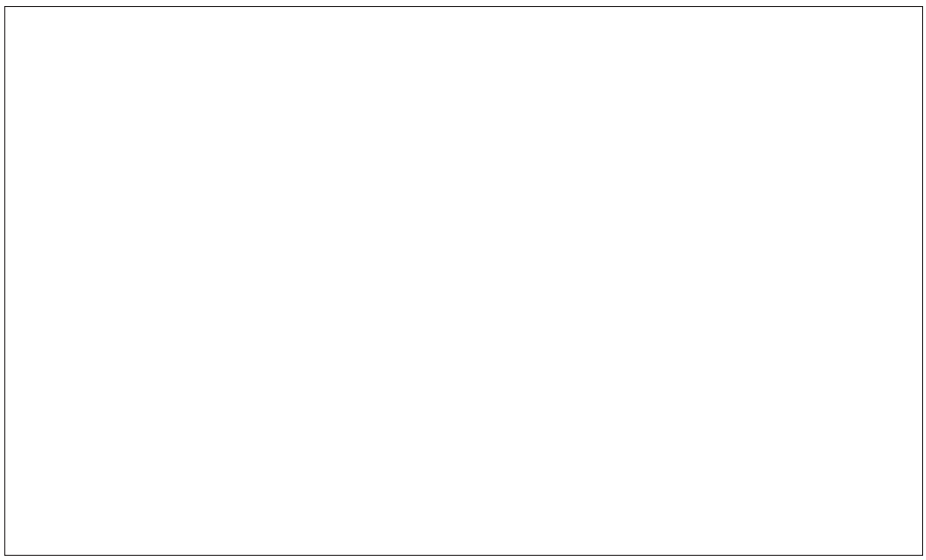
$$\begin{aligned} \mathrm{W}(\theta) &= \mathrm{W}\left\{\text{Los angenommen} \,|\, \text{Schlechtanteil} = \theta\right\} \\ &= \mathrm{W}\left\{X \leq c \,|\, \text{Schlechtanteil} = \theta\right\} \end{aligned}$$

Abbildung 5.1: Operationscharakteristik



Eigenschaften der Operationscharakteristik

In konkreten Produktionen ist oft ein bestimmter (kleiner) Schlechtanteil p zulässig. Lose mit Schlechtanteil $\theta \leq p$ sollten angenommen werden, solche mit $\theta > p$ sollten abgelehnt werden. Eine ideale OC wäre daher:



Dies ist mit konkreten Stichproben im allgemeinen nicht erreichbar. Die OC kann folgenderma-

ßen berechnet werden:

$$\mathrm{W}(\theta) = \mathrm{W}\{X \leq c \,|\, \theta\} = \sum_{a=0}^{c} \mathrm{W}\{X = a \,|\, \theta\}$$

Bei klassischen Stichproben gilt:

$$X \sim H_{N,A,n} \qquad \text{mit } \theta = \frac{A}{N}$$

und

$$\begin{aligned} \mathrm{W}\{X = a\} &= \frac{\binom{A}{a}\binom{N-A}{n-a}}{\binom{N}{n}} \qquad \text{fuer } a_1 \leq a \leq a_2 \\ \text{mit } a_1 &= \max(0, n-(N-A)) \\ a_2 &= \min(n, A) \end{aligned}$$

Wenn der Losumfang N groß ist ($n < \frac{N}{10}$) kann für praktische Zwecke die $H_{N,A,n}$ durch die $B_{n,\frac{A}{N}}$ approximiert werden, d.h.

$$\mathrm{W}\{X = a\} \approx \binom{n}{a} \cdot \theta^a \cdot (1-\theta)^{n-a} \qquad \text{fuer } a = 0(1)n.$$

Wenn n relativ groß ist und obiges gilt, sowie $n \cdot \theta \leq 5$ so kann man die Binomialvertelung $B_{n,\theta}$ durch die Poisson-Verteilung $\mathrm{P}_{n\cdot\theta}$ approximieren:

$$\mathrm{W}\{X = a\} \approx \frac{(n \cdot \theta)^a \cdot e^{-n\theta}}{a!}$$

Bemerkung: Die OC ist abhängig von n, c und θ.

Beispiel: Große Warensendung von Schrauben, einfacher Stichprobenplan $n = 80$, $c = 2$.

1. Falls der Schlechtanteil p=0.01 ist, untersuche man ob dieser Stichprobenplan Lose mit guter Qualität, d.h. $\theta \leq p$, mit hoher Wahrscheinlichkeit annimmt.
   Hier ist die Poisson-Approximation zulässig, da $n < \frac{N}{10}$ und $n \cdot \theta = 0.8 < 5$.

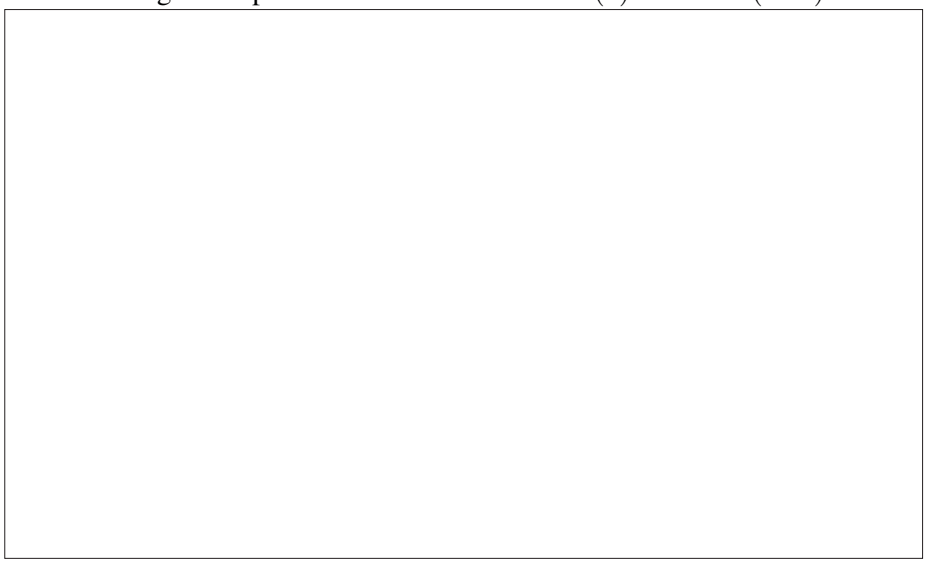
   $$\begin{aligned} \mathrm{W}(\theta) &= \mathrm{W}\{X \leq 2\} \approx \sum_{a=0}^{2} \frac{(80 \cdot \theta)^a}{a!} \cdot e^{-80\cdot\theta} \\ &= e^{-80\cdot\theta}\left[1 + 80 \cdot \theta + \frac{(80 \cdot \theta)^2}{2!}\right] \end{aligned}$$

   Für $\theta = p = 0.01$ ergibt sich $\mathrm{W}(0.01) = 0.953$. Solche Lose werden also mit hoher Wahrscheinlichkeit angenommen.

2. Wie groß ist die Wahrscheinlichkeit der Annahme des Loses, falls $\theta = 0.05$ ist ?
   Für $\theta = 0.05$ ergibt sich $\mathrm{W}(0.05) = 0.238$, d.h. ein solcher Los wird nur mit einer Wahrscheinlichkeit von 23.8% angenommen.

Übung: Im obigen Beispiel berechne man $\mathrm{W}(\theta)$ für $\theta = 0(0.01)0.10$ und zeichne die OC. Für größerer Werte von θ sollten die Binomial-Wahrscheinlichkeiten berechnet werden $(80 \cdot 0.1 = 8 > 5)$.

Abbildung 5.2: Operationscharakteristik für $\mathrm{W}(\theta)$ mit $\theta = 0(0.01)0.10$



5.2 Festlegung einfacher Stichprobenpläne

Man gibt zwei Punkte der OC vor und zwar für p_1 (gute Lose, $\theta \leq p_1$)

$$\mathrm{WS}(p_1) = 1 - \alpha \tag{5.1}$$

und p_2 (schlechte Lose, $\theta \geq p_2$)

$$\mathrm{WS}(p_2) = \beta \tag{5.2}$$

wobei α und β kleine Zahlen sind.

- Produzentenrisiko α: der Produzent möchte das Lose von hoher Qualität ($\theta \leq p_1$) mit hoher Wahrscheinlichkeit $\mathrm{WS}(1 - \alpha)$ angenommen werden.

  Bemerkung: Beziehung zur Irrtumswahrscheinlichkeit α bei statistischen Tests (Fehler 1.Art)

- Konsumentenrisiko β: der Käufer möchte das Lose von schlechter Qualität ($\theta \geq p_2$) mit kleiner Wahrscheinlichkeit WS(β) angenommen werden.

  Bemerkung: Fehler 2.Art, Schärfe $1-\beta$ bei statistischen Tests

Die Werte für α und β sind vom Losumfang N abhängig; je größer das N ist, desto wichtige ist es keine falsche Entscheidung zu treffen. Da n und $c \in \mathbf{N}_0$, können i.a. die Gleichungen 5.1 und 5.2 nicht exakt erfüllt sein. Man wählt für n und c solche werte, dass die OC möglichst nahe an den Punkten $(p_1, 1-\alpha)$ und $(p_2, 1-\beta)$ liegt und ferner gilt:

$$\mathrm{W}(\mathrm{p}_1) \geq 1-\alpha \qquad \text{und} \qquad \mathrm{W}(p_2) \leq \beta.$$

Zur Bestimmung von n und c ist folgendes hilfreich:

$$\mathrm{W}(\theta) = \mathrm{W}\left\{Y > 2 \cdot n \cdot \theta\right\} \qquad \text{wobei} \quad Y \sim \chi^2_{2(c+1)}$$

Beweis: Durch sukzessive partielle Integration sieht man die Gültigkeit von

$$\int_{2n\theta}^{\infty} \frac{y^c e^{-y/2}}{c! \cdot 2^{c+1}} \mathrm{d}y = \sum_{k=0}^{c} \frac{(n\theta)^k}{k!} \cdot e^{-n\theta} = \mathrm{W}(\theta)$$

$$\frac{y^c e^{-y/2}}{c! \cdot 2^{c+1}} = \underbrace{\frac{y^{\frac{2(c+1)}{2}-1} \cdot e^{-y/2}}{\underbrace{\Gamma(c+1)}_{c!} \cdot 2^{\frac{2(c+1)}{2}}}}_{\text{Dichte von } \chi^2_{2(c+1)}}$$

Zur Bestimmung von n und c hat man zwei Gleichungen:

$$1-\alpha \quad \leq \quad \mathrm{W}(p_1) = \mathrm{W}\left\{Y \geq 2np_1 \,|\, Y \sim \chi^2_{2(c+1)}\right\} \tag{5.3}$$

$$\beta \quad \geq \quad \mathrm{W}(p_2) = \mathrm{W}\left\{\mathrm{Y} \geq 2np_2 \,|\, Y \sim \chi^2_{2(c+1)}\right\} \tag{5.4}$$

In einer Tabelle der χ^2-Verteilung sucht man Fraktile:

$$\chi^2_{\bullet;\,\alpha} \qquad \text{und} \qquad \chi^2_{\bullet;\,1-\beta}$$

sodaß gilt:

$$\begin{aligned} 2np_1 &= \chi^2_{2(c+1)\,;\,\alpha} \\ 2np_2 &= \chi^2_{2(c+1)\,;\,1-\beta} \end{aligned}$$

Dies ist dann der Fall, wenn

$$\frac{\chi^2_{\bullet;\,1-\beta}}{\chi^2_{\bullet;\,\alpha}} = \frac{p_2}{p_1}$$

Man sucht jene Anzahl von Freiheitsgraden, sodaß der Quotient der Fraktilen möglichst nahe bei $\frac{p_2}{p_1}$ liegt (es kommen nur geraden Anzahlen in Betracht).
Zur Bestimmung von n müssen die Gleichungen 5.3 und 5.4 erfüllt sein. Daraus erhält man für n zwei Ungleichungen. Die kleinste ganze Zahl, die beide Ungleichungen erfüllt, ist der gesuchte Stichprobenumfang.

Bemerkung: Für die praktive Anwendung gibt es Tabellen und sog. „Normogramme“.

5.3 Zweifache Stichprobenpläne

Hier wird eine Stichprobe vom Umfang n_1 gezogen (x_1=Anzahl der schlechten Stücke in dieser Stichprobe). Falls:

$x_1 \leq c_1$	Annahme des Loses
$x_1 \geq d_1$	Ablehnung
$c_1 < x_1 < d_1$	Zweite Stichprobe mit Umfang n_2

Dabei bezeichnet n_2 die Anzahl der schlechten Stücke in der zweiten Stichprobe. Falls:

$x_1 + x_2 \leq c_2$	Annahme des Loses
anderenfalls	Ablehnung

Man erhält einen zweifachen Stichprobenplan $(n_1, c_1, d_1, n_2, c_2)$.

Bemerkung: Die Berechnung mehrfacher Stichprobenpläne ist kompliziert. Es gibt aber Tabellen. Der Vorteil ist, dass der zu erwartende Stichprobenumfang bei pratkisch gleicher OC kleiner ist.

5.4 Sequentielle Stichprobenpläne

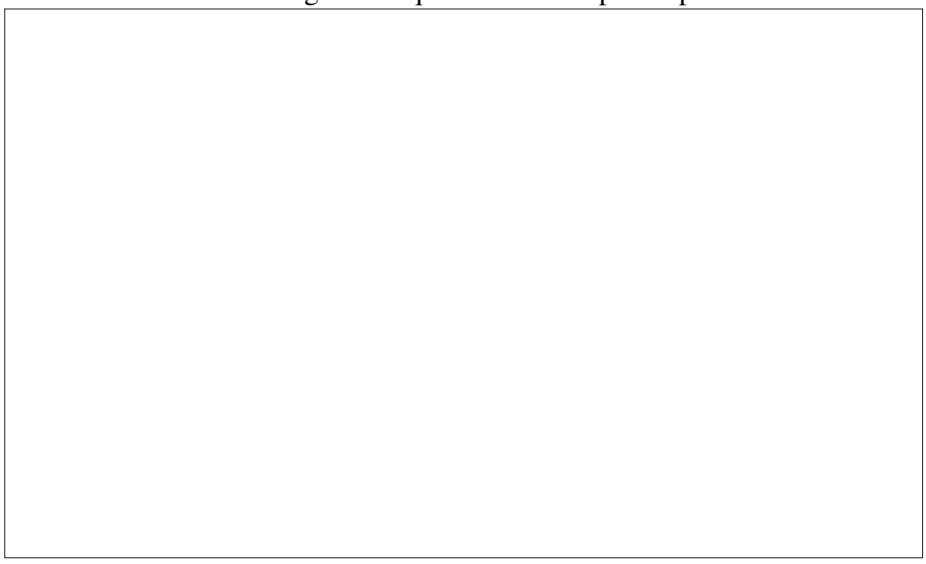
Man entnimmt hintereinander einzelne Elemente des Loses. Dabei bezeichnet x_n die Anzahl der schlechten Stücke, nachdem man insgesamt n Stücke geprüft hat.

Man benötigt einen sog. Prüfplan (a, b, c)

$x_n \leq c \cdot n - a$	Annahme
$x_n \geq c \cdot n + b$	Ablehnung
$c \cdot n - a < x_n < c \cdot n + b$	Weiterprüfen

Die Berechnung von a, b, c erfolgt aus den vorgegebenen Werte p_1, p_2, α, β der OC mit Hilfe des sog. approximativen Sequentialtest von A. Wald:

Abbildung 5.3: Sequentieller Stichprobenplan



$$A = \ln \frac{1-\alpha}{\beta} \quad B = \ln \frac{1-\beta}{\alpha} \quad P = \ln \frac{p_2}{p_1} \quad Q = \ln \frac{1-p_1}{1-p_2}$$
$$a = \frac{A}{P+Q} \quad b = \frac{B}{P+Q} \quad c = \frac{Q}{P+Q}$$

5.5 Kontrollkarten

Zur laufenden Überachung eines Produktionsprozesses. Dazu wird eine Größe x gemessen, die in Abhängigkeit der Zeit variiert. Dabei nimmt man meist an, dass diese eine Realisierung einer stochastischen Größe X ist, die (approximativ) normalverteilt ist.

Dabei sind gewisse Schwankungen von X durchaus zulässig, solange sie in einem gewissen Bereich (Intervall) bleiben.

Beispiel:

1. Druckfestigkeit von Beton
2. Feuchtigkeit von Bauholz
3. Durchmesser von Schrauben

Qualitätsänderungen entsprechen Änderungen von zumindest einem Parameter. Hier kann sich eine kleine Streuung (Varianz) insofern günstig auswirken, da man in diesem Fall mit dem Erwartungswert μ_0 der Produktion näher an die Qualitätsgrenze herangehen kann.

Es gibt verschiedene Arten von Kontrollkarten, je nachdem was aufgetragen wird:

- Einzelmeßwerte: x-Karten
- Mittelwerte: $\bar{x}$-Karten
- Streuungen: s-Karten

Damit können Schwankungen im Produktionsprozeß erkannt werden.

Bezeichnungen:

ML	Mittellinie	
T_o	obere Toleranzgrenze	
T_u	untere Toleranzgrenze	
OKG	obere Kontrollgrenze	Eingriff notwendig
UKG	untere Kontrollgrenze	Eingriff notwendig
OWG	obere Warngrenze	
UWG	untere Warngrenze	

Ur- oder Einzelwertkarte

In regelmäßigen Abständen werden kleine Stichproben vom Umfang n (etwa 5) gezogen und die Messwerte in der Kontrollkarte eingetragen. ML, KG_n und WG_n werden aus einem Verlauf von k (20-30) Stichproben aus je m Einzelwerten bestimmt.

$x_{j1}, ..., x_{jm}$	Einzelwerte der j − Stichprobe	
$\bar{x}_j$	Stichprobenmittel	der $j-$ ten Stichprobe
s_j	Stichprobenstreuung	der $j-$ ten Stichprobe
R_j	Spannweite $x_{max} - x_{min}$	der $j-$ ten Stichprobe

$$\bar{\bar{x}} = \frac{\bar{x}_1 + ... + \bar{x}_k}{k} \quad \text{Gesamtmittel}$$

$$\bar{R} = \frac{R_1 + ... + R_k}{k} \quad \text{Mittelwert der Spannweiten}$$

$$\bar{s} = \frac{s_1 + ... + s_k}{k} \quad \text{Mittelwert der Stichprobenstreuungen}$$

Die verschiedenen Grenzen ergeben sich aus den Bedingungen für die Messgrößen zum Zeitpunkt t:

$$\mathrm{W}\{G_u \leq X_1(t), ..., X_n(t) \leq G_o\} \geq 1 - \alpha$$

Bei einseitigen Fragestellungen (z.B. Mindestgewicht) gilt:

$$\mathrm{W}\{G_u \leq X_1(t), ..., X_n(t)\} \geq 1 - \alpha$$

Dabei bezeichnet α die Verwerfungswahrscheinlichkeit, falls der Prozess unter Kontrolle ist.

Bemerkung: Für $\alpha = 0.01$ Kontrollgrenzen, für $\alpha = 0.05$ Warngrenzen.

Bestimmung der Grenzen:

$$\begin{aligned}
\mathrm{W}\{G_u \le X_1(t), ..., X_n(t) \le G_o\} &= \mathrm{W}\left(\bigcap_{\mathrm{i}=1}^{n} \{G_u \le X_i(t) \le G_o\}\right) \\
&= \prod_{i=1}^{n} \mathrm{W}\{G_u \le X_i(t) \le G_o\} \\
&= (\mathrm{W}\{G_u \le X(t) \le G_o\})^n = 1 - \alpha \\
\mathrm{W}\{G_u \le X \le G_o\} &= \sqrt[n]{1-\alpha} \\
G_u &= \frac{1 - \sqrt[n]{1-\alpha}}{2} - \text{Fraktiles der Vtlg. von } X \\
G_o &= 1 - \frac{1 - \sqrt[n]{1-\alpha}}{2} - \text{Fraktiles der Vtlg. von } X
\end{aligned}$$

Diese Fraktilwerte können wegen der Normalverteilungsannahme leicht berechnet werden. Praktisch werden die Grenzen mittels Tabellen ermittelt.

$\bar{x}$-Karte

Wenn der Prozess unter Kontrolle ist, gilt

$$\bar{X}_n \sim N\left(\mu_0, \frac{\sigma_0^2}{n}\right) \quad , \sigma^2 = \frac{\sigma_0^2}{n}.$$

Damit kann man Warn- und Kontrollgrenzen ermitteln.

Die KG_n werden entweder mittels der Sicherheitswahrscheinlichkeit α (meist 0.01) oder mittels der 3σ-Regel festgelegt

$$\bar{\bar{x}} \pm \frac{3\sigma}{\sqrt{n}}.$$

Für zweiseitige Kontrollkarten gilt:

$$\begin{aligned}
OKG &:= \inf_y \left\{\mathrm{W}\left\{\bar{X}_n > y\right\}\right\} \le \frac{\alpha}{2} \\
UKG &:= \sup_y \left\{\mathrm{W}\left\{\bar{X}_n < y\right\}\right\} \le \frac{\alpha}{2}
\end{aligned}$$

Bei Normalverteilungen verwendet man oft

$$OKG - ML = ML - UKG = 3\sqrt{\mathrm{Var}\bar{X}_n}.$$

Dabei wird $\mathrm{Var}\bar{X}_n$ geschätzt bzw. wird eine Schätzung $\bar{s}$ für $\sqrt{\mathrm{Var}\ \bar{X}_n}$ benützt.

ML	WG	KG	KG
	0.05	$\alpha = 0.01$	3σ
$\bar{\bar{x}}$	$\bar{\bar{x}} \pm \frac{1.96}{\sqrt{n}}\sigma$	$\bar{\bar{x}} \pm \frac{2.58}{\sqrt{n}}\sigma$	$\bar{\bar{x}} \pm \frac{3}{\sqrt{n}}\sigma$

Begründung:

$$
\begin{aligned}
\bar{X}_n \sim \mathrm{N}\left(\mu, \sigma^2\right) &\rightarrow \mathrm{W}\left\{\left|\bar{X}_n - \mu\right| \leq G\right\} = 1 - \alpha \\
\mathrm{W}\left\{\left|\bar{X}_n - \mu\right| \leq G\right\} &= \mathrm{W}\left\{\mu - G \leq \bar{X}_n \leq \mu + G\right\} \\
&= \mathrm{W}\left\{\frac{\mu - G - \mu}{\sigma} \leq \underbrace{\frac{\bar{X}_n - \mu}{\sigma}}_{\sim \mathrm{N}(0,1)} \leq \frac{\mu + G - \mu}{\sigma}\right\} \\
&= \Phi\left(\frac{G}{\sigma}\right) - \Phi\left(-\frac{G}{\sigma}\right) \\
&= 2\Phi\left(\frac{G}{\sigma}\right) - 1 \\
&= 1 - \alpha \\
\Phi\left(\frac{G}{\sigma}\right) &= 1 - \frac{\alpha}{2} \\
\frac{G}{\sigma} &= z_{1-\frac{\alpha}{2}} \qquad \text{Fraktile der } \mathrm{N}(0,1)
\end{aligned}
$$

Für $\alpha = 0.05$ gilt $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$.
Für $\alpha = 0.01$ gilt $z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.58$.

s-Karte

Zur Überwachung der Streuung (Varianz) von X gilt bei Normalverteilung

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

$$
\begin{aligned}
\mathrm{W}\left\{\chi^2_{n-1;\frac{\alpha}{2}} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \chi^2_{n-1;1-\frac{\alpha}{2}}\right\} &= 1 - \alpha \\
\mathrm{W}\left\{\frac{\sigma^2}{n-1}\chi^2_{n-1;\frac{\alpha}{2}} \leq S_n^2 \leq \frac{\sigma^2}{n-1}\chi^2_{n-1;1-\frac{\alpha}{2}}\right\} &= 1 - \alpha \\
\mathrm{W}\left\{\sigma\sqrt{\frac{\chi^2_{n-1;\frac{\alpha}{2}}}{n-1}} \leq S_n \leq \sigma\sqrt{\frac{\chi^2_{n-1;1-\frac{\alpha}{2}}}{n-1}}\right\} &= 1 - \alpha
\end{aligned}
$$

Für die Grenzen ergibt sich somit

$$\text{obere} \qquad \sigma\sqrt{\frac{\chi^2_{n-1;1-\frac{\alpha}{2}}}{n-1}}$$

$$\text{untere} \qquad \sigma\sqrt{\frac{\chi^2_{n-1;\frac{\alpha}{2}}}{n-1}}$$

Da σ mittels $\bar{s}$ geschätzt wird, erhält man folgende Grenzen:

- Warngrenzen $\alpha = 0.05$

$$\frac{\bar{s}}{c_n}\sqrt{\frac{\chi^2_{n-1;0.025}}{n-1}} \qquad \frac{\bar{s}}{c_n}\sqrt{\frac{\chi^2_{n-1;0.975}}{n-1}}$$

- Kontrollgrenzen

$$\frac{\bar{s}}{c_n}\sqrt{\frac{\chi^2_{n-1;0.005}}{n-1}} \qquad \frac{\bar{s}}{c_n}\sqrt{\frac{\chi^2_{n-1;0.995}}{n-1}}$$

Bemerkung: Es werden verschiedene Typen von QR-Karten (QS-Karten) verwendet ($\tilde{x}$-Karten, R-Karten, Summenkarten CUSUM).

6 Regressionsanalyse

Anwendung zur Beschreibung (kausaler) aber nicht deterministischer Zusammenhänge.

- Zwischen stochastischen Größen
- Zwischen Einstellgrößen (unabhängige Größen) und stochastischen Größen

Beispiel: Ist (X, Y) ein 2-dimensionaler Vektor der Körpergröße und -gewicht beschreibt.

Beispiel: Versuch, dessen Ausgang von einer deterministischen einstellbaren Größe abhängt.

Einstellgröße x: Abhängige stochastische Größe Y_x.

6.1 Regression 1.Art

$$\begin{aligned} \mathbb{E}[Y|X=x] &= y(x) \\ y(x) &= \sum_{j=0}^{k} \theta_j \cdot x^j \qquad \theta_j \text{ Parameter} \end{aligned}$$

Dies ist eine sog. „polynomische Regression der Ordnung k“.

Sonderfall k=1:

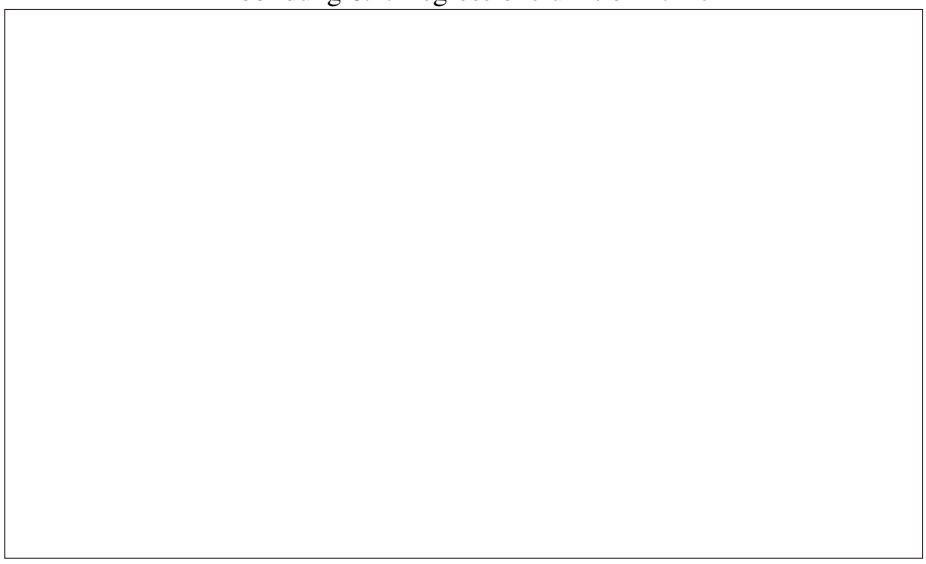
$$\mathbb{E}[Y|X=x] = \theta_0 + \theta_1 \cdot x \,.....\, \text{Regressionsgeraden}$$

Satz 6.1

$$\begin{aligned} (X,Y) &\sim \mathrm{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \\ \mathbb{E}[Y|X=x] &= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) \\ \mathbb{E}[X|Y=y] &= \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2) \end{aligned}$$

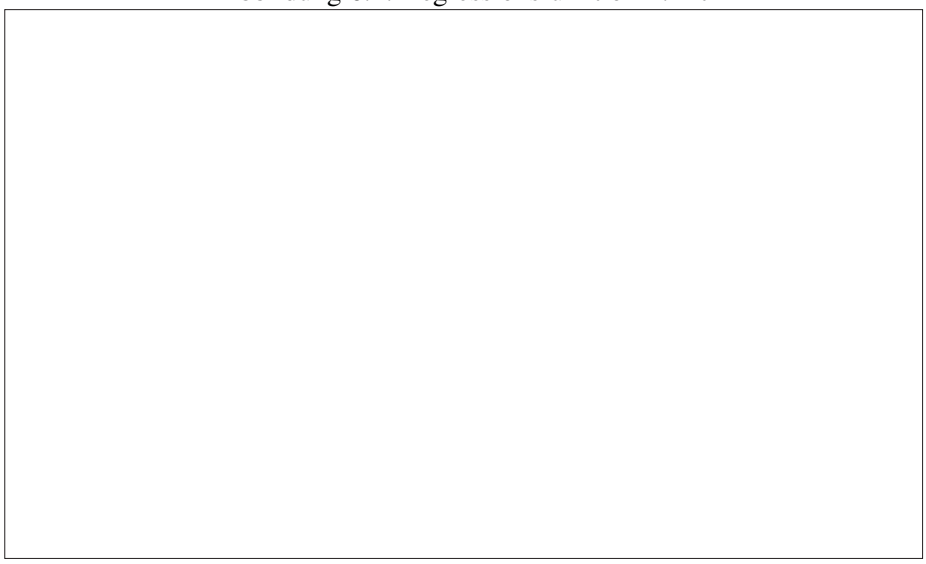
d.h. bei 2-dimensionalen Normalverteilungen hat man immer Regressionsgeraden.

Abbildung 6.1: Regressionsfunktion 2.Art



$$\begin{aligned} \psi(x) &= \mathbb{E}Y_x \\ Y_x &= \psi(x) + \underbrace{U_x}_{\mathbb{E}U_x = 0} \end{aligned}$$

Abbildung 6.2: Regressionsfunktion 1.Art



Bemerkung:

- $\tan(\phi) = \frac{\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2} \cdot \frac{1-\rho^2}{\rho}$

- Falls X unabhängig von Y ist, sind die Regressionslinien parallel zu den Regressionsachsen.

Aus einer Stichprobe (X_i, Y_i), $i = 1(1)n$ kann die Regressionsfunktion $X \to \mathbb{E}[Y|X = x]$ geschätzt werden (vgl. Abbildung 6.3).

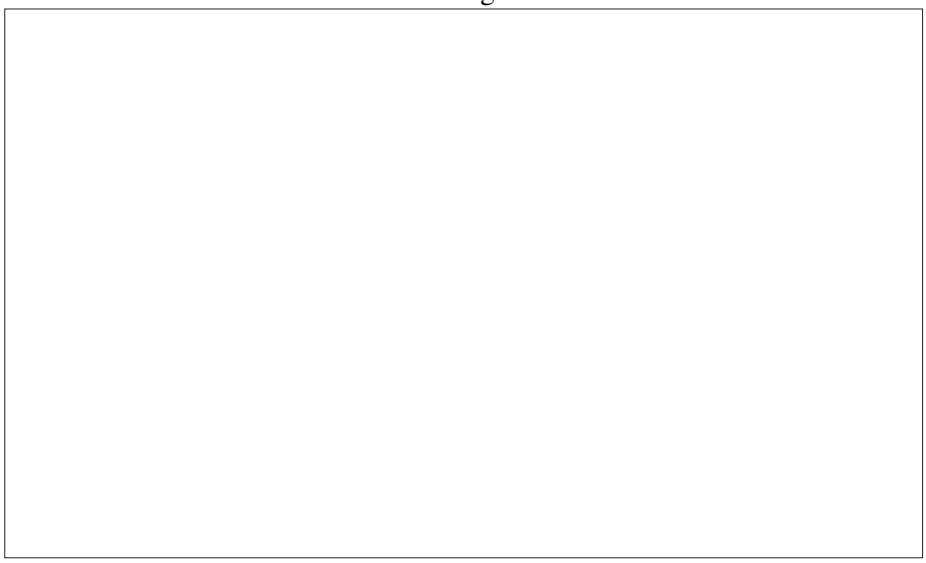
Definition: Bedingte Mittelwerte $g(x_i)$

$$g(x_i) = \bar{y}_j = \frac{1}{n_j} \cdot \sum_{k=1}^{n_j} y_{jk} = \frac{y_{j\bullet}}{n_j} \qquad j = 1(1)n$$

Abbildung 6.3 zeigt eine empirische Regressionsfunktion 1.Art von Y bezüglich X (auf den Werten $x_1, ..., x_m$) definiert). Durch die Punkte $(x_j, g(x_j))$, $j = 1(1)n$ gehende Kurven heißen *empirische Regressionskurven 1.Art*.

Bemerkung: Für größeren Datemunfang n bildet man Klasseneinteilungen für die x-Werte und trägt die bedingten y-Mittelwerte über den Klassenmitten auf. So erhält man oft u.U. glattere Regressionskurven.

Abbildung 6.3:



Konkrete Stichprobe (x_i, y_i), $i = 1(1)n$, m Werte von x_i verschieden.

$$\begin{array}{ll} (x_1, y_{11}), ..., (x_1, y_{1n_1}) & x_1, ..., x_m \text{ verschieden} \\ (x_2, y_{21}), ..., (x_2, y_{2n_2}) & \\ (x_m, y_{m1}), ..., (x_m, y_{mn_m}) & \sum_{j=1}^{n} n_j = n \end{array}$$

Satz 6.2 Die empirische Regressionsfunktion von y bezüglich x ist unter allen $(x_1, ..., x_m)$ definierten Funktionen $f(\cdot)$ jene, für welche die Summe der Quadrate aller y-Differenzen minimal ist, d.h.

$$\sum_{j=1}^{m}\sum_{k=1}^{n_j}[y_{jk} - g(x_j)]^2 \equiv \sum_{j=1}^{m}\sum_{k=1}^{n_j}[y_{jk} - f(x_j)]^2$$

Beweis: $g(x_j)$ ist das arithmetische Mittel der $y_{j1}, ..., y_{jn_j}$; für festes j gilt wegen der bekannten Eigenschaften der Abstandsquadratsumme:

$$\sum_{k=1}^{n_j}[y_{jk} - g(x_j)]^2 \leq \sum_{k=1}^{n_j}[y_{jk} - c_j]^2 \qquad \forall c_j \neq g(x_j)$$

Da alle m Summanden nicht negativ sind, folgt die Behauptung.

Bemerkung: Empirische Regressionsfunktion 1.Art von X bezüglich Y werden analog berechnet.

6.2 Regression 2.Art

Abhängigkeit einer Größe Y von einer nicht stochastischen Größe (Einstellgröße) X.

Ausgleichsrechnung

Beobachtete Paare (x_i, y_i), $i = 1(1)n$ liegen annähernd auf einer Geraden

$$y = \alpha + \beta x.$$

Die „Parameter“ α und β können mit der Methode der kleinstenm Abstandsquadratsumme bestimmt werden:

$$\text{QS} = \sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 \rightarrow \text{Min} \tag{6.1}$$

Jene Werte $\hat{\alpha}$ und $\hat{\beta}$ für die QS in Gleichung 6.1 minimal ist, sind die Parameter der Ausgleichsgeraden $y = \hat{\alpha} + \hat{\beta}x$.

Die Werte $\hat{\alpha}$ und $\hat{\beta}$ erhält man durch die notwendigen Bedinungen

$$\frac{\partial \text{QS}}{\partial \alpha} = 0 \qquad \frac{\partial \text{QS}}{\partial \beta} = 0$$

Satz 6.3 Als Lösung ergibt sich

$$\begin{aligned}\hat{\alpha} &= \frac{\left(\sum_{i=1}^n x_i^2\right)\cdot\left(\sum_{i=1}^n y_i\right)^2 - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n x_i \cdot y_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ \hat{\beta} &= \frac{n\left(\sum_{i=1}^n x_i y_i\right) - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n\right)^2}\end{aligned}$$

Mit der Bezeichnung

$$s_{x,y} = \frac{1}{n-1}\left[\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}\right] = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

gilt:

$$\hat{\beta} = \frac{s_{x,y}}{s_x^2} \qquad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Die Ausgleichsgerade ergibt sich zu:

$$y - \bar{y} = \hat{\beta}(x - \bar{x}) = \frac{s_{xy}}{s_x^2}(x - \bar{x})$$

Bemerkung: Die Ausgleichsgerade (Regressionsgerade) geht durch den Punkt $(\bar{x}, \bar{y})$ und hat die Steigung $\frac{s_{xy}}{s_x^2}$.

Bemerkung: P.S. de Laplace hat für das Regressionsmodell $y = \theta x_i + \epsilon_i$ die Absolutbeträge herangezogen:

$$\hat{\theta} = \arg_\theta \min \sum_{i=1}^n |y_i - \theta x_i|$$

Allgemeine Ausgleichskurven

$$y = \psi(x; \alpha_1, ..., \alpha_l)$$

Mittels der Methode der kleinsten Abstandsquadratsumme aus Daten $x_i, y_i,\ i = 1(1)n$

$$\text{QS} = \sum_{i=1}^n [y_i - \psi(x_i; \alpha_1, ..., \alpha_l)]^2 \to \text{Min}$$

ermittelt man die „Parameter“ $\hat{\alpha}_1, ..., \hat{\alpha}_l$ der Ausgleichskurve

$$y = \psi(x; \hat{\alpha}_1, ..., \hat{\alpha}_l).$$

Notwendige Bedingung:

$$\frac{\partial \text{QS}}{\partial \alpha_k} = -2\sum_{i=1}^n [y_i - \psi(x_i; \alpha_1, ..., \alpha_l)]\frac{\partial \psi}{\partial \alpha_k} = 0 \qquad k = 1(1)l$$

Sonderfall: $y = a + bx + cx^2$

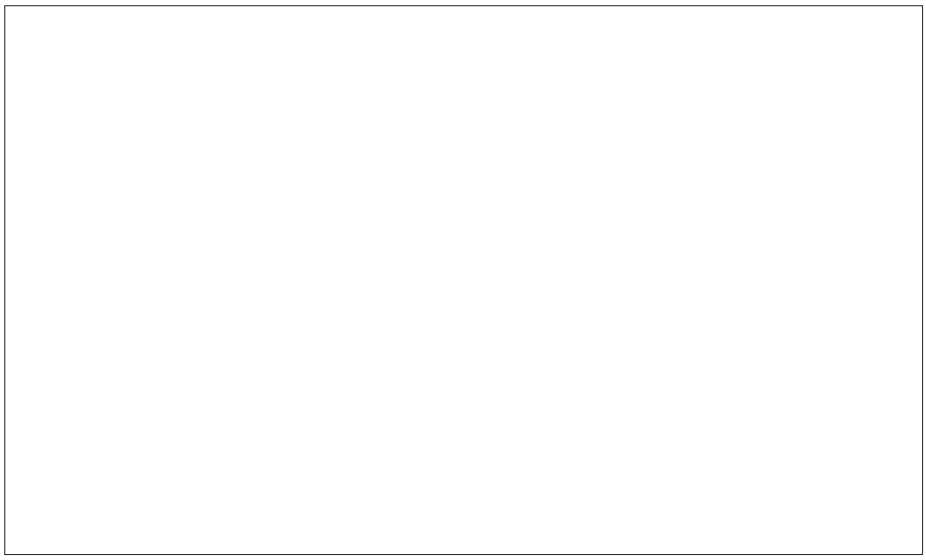
Multiple Ausgleichsrechnung

Für funktionale Abhängigkeiten von k reelen Variablen

$$y = \psi(x_1, ..., x_k; \theta) \qquad \theta = \begin{pmatrix} \theta_1 \\ \theta_m \end{pmatrix} \in \mathbf{R}^m$$

und Beobachtungen $(x_{i1}, ..., x_{ik}; y_i)$; $i = 1(1)n$ sind Auslgeichsflächen gesucht.

Sonderfall k=1:



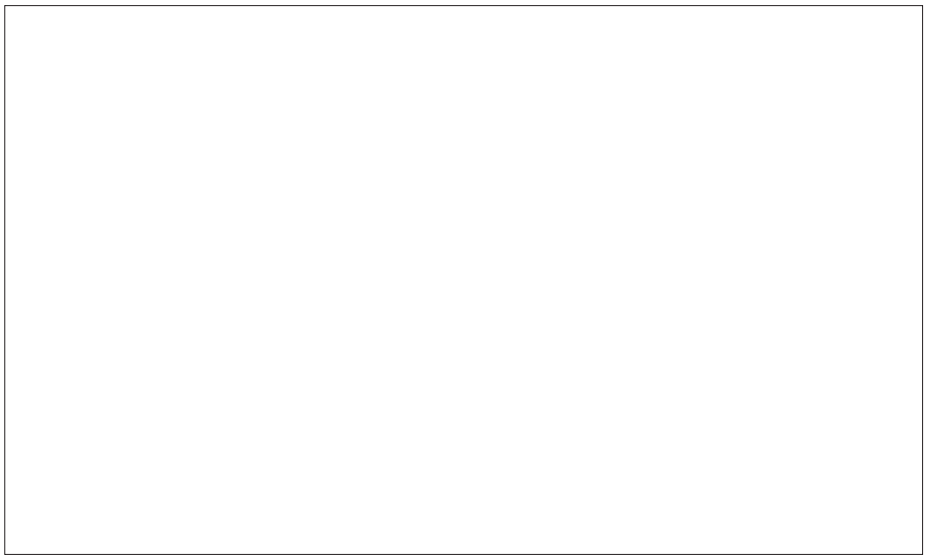
Spezialfall: Lineare Funktion

$$y = \theta_0 + \sum_{j=1}^{k} \theta_j x_j \qquad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \\ \theta_k \end{pmatrix} \qquad m = k + 1$$

Notation:

$$\begin{aligned} \underline{x} &= (x_1, ..., x_k) \\ \dot{\underline{x}} &= (1, x_1, ..., x_k) \\ y &= \dot{\underline{x}} \cdot \theta \end{aligned}$$

Spezialfall k=2:



Mit Hilfe der Methode der kleinsten Abstandsquadratsumme für n Beobachtungen

$$\text{QS} = \sum_{i=1}^{n} \left[y_i - \theta_0 - \sum_{j=1}^{k} \theta_j x_{ij} \right]^2 \rightarrow \text{Min} \tag{6.2}$$

können über die sog. Gauß´schen Normalgleichungen die Lösungen $\hat{\theta}_0, \hat{\theta}_1, ..., \hat{\theta}_k$ gefunden werden. Schreibt man für die Matrix der Einstellgrößen

$$\mathcal{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & & & & \vdots \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

$$\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$$

$$\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

so erhält 6.2 die Form

$$\text{QS} = \left(\underline{y} - \mathcal{X}\theta\right)^T \left(\underline{y} - \mathcal{X}\theta\right).$$

Satz 6.4 Als Lösungsvektor θ ergibt sich für Rang$\mathcal{X} = k + 1$ und $n > k + 1$

$$\hat{\theta} = \left(\mathcal{X}^T \mathcal{X}\right)^{-1} \mathcal{X}^T \underline{Y}.$$

6.3 Stochastische Regressionsanalyse

Für Streeungsuntersuchungen betrachtet man y als stochastische Größe Y_x, die von der Einstellgröße x abhängt:

$$Y_x = \psi(x) + U_x$$

Die Werte y_i in den Datenpaaren (x_i, y_i) werden als Realisierungen von der stochastischen Größen Y_i, die den Y_{x_i} entsprechen, betrachtet. Meist wird identisches Streuverhalten vorrausgesetzt, d.h.

$$\text{Var}Y_x \equiv \sigma^2.$$

1-dimensionaler Fall $x_i \in \mathbf{R}$

Satz 6.5 (Gauß-Markoff) Sind $Y_1, ..., Y_n$ unkorreliert und gilt $\mathbb{E}Y_i = \alpha + \beta \cdot x_i$, sowie $\text{Var}Y_i \equiv \sigma^2$, $\forall i = 1(1)n$ und sind die $x_1, ..., x_n$ bekannt und nicht alle gleich und sind α, β und σ^2 unbekannt, so gilt:
Die mittels der Ausgleichsparameter $\hat{\alpha}$ und $\hat{\beta}$ konstruierten Schätzfunktionen A und B für α und β

$$\begin{aligned} A &= \frac{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n Y_i\right) - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n x_i Y_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ &\uparrow \\ &\textbf{SG!} \\ &\downarrow \\ B &= \frac{n\left(\sum_{i=1}^n x_i Y_i\right) - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n Y_i\right)}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \end{aligned}$$

sind die effizienten linearen Schätzfunktionen (linear in Y_i) für α bzw. β und es gilt:

$$\begin{aligned} \text{Var}A &= \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\sigma^2 \\ \text{Var}B &= \frac{n}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\sigma^2 \\ \text{Cor}(A,B) &= -\frac{\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i^2\right)}\sigma^2 \end{aligned}$$

Eine unverzerrte Schätzfunktion für σ^2 ist

$$S^2 = \frac{\sum_{i=1}^n (Y_i - A - Bx_i)^2}{n-2}$$

Bemerkung:

1. $A = \bar{Y} - B\bar{x}$

2. $n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 = n\sum_{i=1}^{n} (x_i - \bar{x}_n)^2$
3. Matrixschreibweise

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_n \end{pmatrix} \qquad \mathcal{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad \theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$\begin{pmatrix} A \\ B \end{pmatrix} \hat{=} \hat{\theta} - \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \left(\mathcal{X}^T \mathcal{X}\right)^{-1} \mathcal{X}^T \underline{Y}$$

4. Der Wert $\mathbb{E}Y_x$ der Regressionsfunktion an der Stelle x wird durch $\widehat{\mathbb{E}_Q Y_x} = A + B \cdot x$ unverzerrt geschätzt.
   Für diese Schätzfunktion gilt: $A + B \cdot x = \bar{Y} + B\,(x - \bar{x})$

6.4 Regressionsgeraden 2.Art bei Normalverteilung

Sind in Satz 6.5 die stochastischen Größen Y_i normalverteilt, d.h.

$$Y_i \sim \mathrm{N}\left(\alpha + \beta x_i, \sigma^2\right),$$

so kann man einfache Konfidenzbereiche für α, β, σ^2 angebenund stochastische Tests konstruieren.

Satz 6.6 Gilt zu den Vorraussetzungen von Satz 6.5 noch $Y_i \sim \mathrm{N}\left(\alpha + \beta x_i, \sigma^2\right)$ so folgt:

1. $\hat{\alpha}$ und $\hat{\beta}$ sind auch die plausiblen Schätzwerte für α und β.
2. $A \sim \mathrm{N}\left(\alpha, \mathrm{Var}A\right)$ und $B \sim \mathrm{N}\left(\beta, \mathrm{Var}B\right)$
3. $\frac{\sum_{i=1}^{n}(Y_i - A - Bx_i)^2}{\sigma^2} = \underbrace{\frac{(n-2)\,S^2}{\sigma^2}}_{\substack{\text{unabh. von} \\ A \text{ und } B}} \sim \chi^2_{n-2}$

Beweis:

1. Lösungen der Plausibilitätsgleichungen

$$\frac{\partial}{\partial \alpha} \ln l\left(\alpha, \beta, \sigma^2; y_1, ..., y_n, x_1, ..., x_n\right) = 0$$

2. $Y_1, ..., Y_n$ sind unabhängig normalverteilt, daher gilt:
   a) $\underline{Y} = \begin{pmatrix} Y_1 \\ Y_n \end{pmatrix}$ ist normalverteilt

b) $\begin{pmatrix} A \\ B \end{pmatrix} = \left(\mathcal{X}^T\mathcal{X}\right)^{-1}\mathcal{X}^T \cdot \underline{Y}$ ist normalverteilt

c) A und B sind normalverteilt

In Verbindung mit Satz 6.5 ergibt sich die Behauptung 2.

Konstruktion von Konfidenzintervallen für $\alpha, \beta, \sigma^2, \sigma$:

Satz 6.7 Unter den Vorraussetzungen von Satz 6.6 erhält man folgende Konfidenzintervalle mit Überdeckungswahrscheinlichkeit $1-\gamma$:

1. Für α
$$\left[A - t_{n-2;1-\frac{\gamma}{2}}\sqrt{s^2\frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}};\ A + t\ldots\right]$$
2. Für β
$$\left[B - t_{n-2;1-\frac{\gamma}{2}}\sqrt{s^2\frac{n}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}};\ B + t\ldots\right]$$
3. Für σ^2
$$\left[\frac{(n-2)\,s^2}{\chi^2_{n-2;1-\frac{\gamma}{2}}};\frac{(n-2)\,s^2}{\chi^2_{n-2;\frac{\gamma}{2}}}\right]$$
4. Für σ
$$\left[\sqrt{\frac{(n-2)\,s^2}{\chi^2_{n-2;1-\frac{\gamma}{2}}}};\sqrt{\frac{(n-2)\,s^2}{\chi^2_{n-2;\frac{\gamma}{2}}}}\right]$$

Beweis:

1. $\frac{A-\alpha}{\sqrt{\mathrm{Var}A}} \sim \mathrm{N}(0,1)$ und Satz 6.6 Punkt 3:
$$\frac{(n-2)\,S^2}{\sigma^2} \sim \chi^2_{n-2}$$
Wenn $X \sim \mathrm{N}(0,1)$ und $Y \sim \chi^2_n$ dann sind die beiden Größen unabhängig und somit gilt:
$$\begin{aligned}\frac{X}{\sqrt{\frac{Y}{n}}} &\sim t_n\\ \frac{(A-\alpha)\,\sigma}{\sqrt{\mathrm{Var}A}\sqrt{S^2}} &\sim t_{n-2}\end{aligned}$$
2. Analog zu 1

3. $\frac{(n-2)S^2}{\sigma^2}$ ist eine Pivot-Größe

$$\mathrm{W}\left\{\chi^2_{n-2;\frac{\gamma}{2}} \leq \frac{(n-2)\,S^2}{\sigma^2} \leq \chi^2_{n-2;1-\frac{\gamma}{2}}\right\} = 1-\gamma$$

4.

$$\mathrm{W}\left\{\sqrt{\chi^2_{n-2;\frac{\gamma}{2}}} \leq \sqrt{\frac{(n-2)\,S^2}{\sigma^2}} \leq \sqrt{\chi^2_{n-2;1-\frac{\gamma}{2}}}\right\} = 1-\gamma$$

Bemerkung:

1. $\frac{n}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{1}{\sum_{i=1}^n (x_i-\bar{x})^2}$
2. Mann kann auch 2-dimensionale Konfidenzbereiche für den Parametervektor (α, β) konstruieren.

Unter den Vorraussetzungen von Satz 6.6 gilt

$$A + Bx - \alpha - \beta \sim \mathrm{N}\left(0, \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right),$$

da

$$\begin{aligned}
\mathrm{Var}\,(A + Bx - \alpha - \beta) &= \mathrm{Var}\,(A + Bx) \\
\mathrm{Var}A + \mathrm{Var}\,(Bx) + 2x\mathrm{Cov}\,(A,B) &= \frac{nx^2 - 2x\sum_{i=1}^n x_i + \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\sigma^2 \\
&= \sigma^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right)
\end{aligned}$$

Gesucht ist nun ein Konfidenzintervall für $\mathbb{E}Y_x$ zu gegebenen x (vgl. Abbildung 6.4)

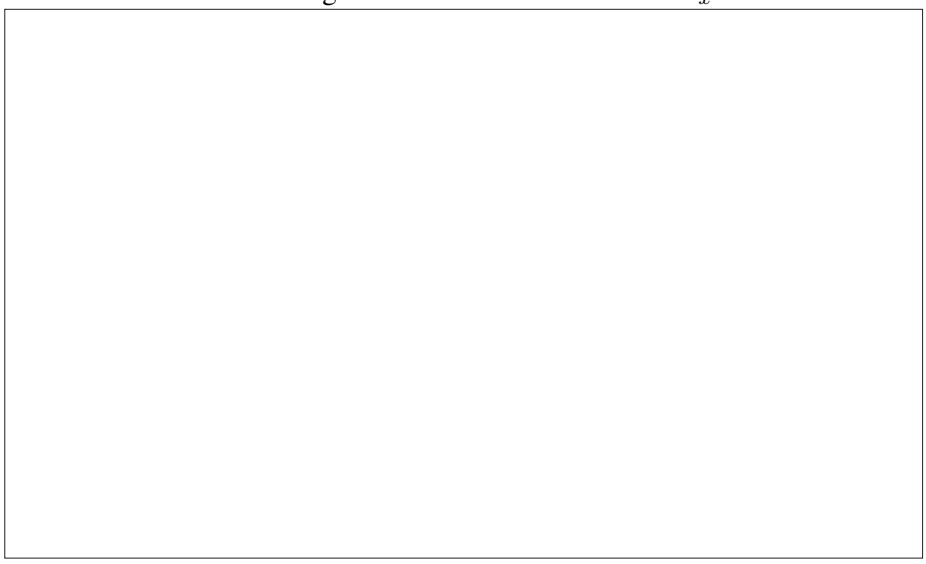
Satz 6.8 Unter den Vorraussetzungen von Satz 6.6 gilt:

$$\left.\begin{pmatrix} Y_1 - A - Bx_1 \\ \vdots \\ Y_n - A - Bx_n \end{pmatrix} \atop \begin{pmatrix} A \\ B \end{pmatrix}\right\} \text{unabhaengig normalverteilt}$$

Folgerung:

$$\begin{aligned}
Y_x - A - Bx &\sim \mathrm{N}\left(0, 1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right) \text{ da } \mathbb{E}\,(Y_x - A - B_x) = 0 \\
\mathrm{Var}\,(Y_x - A - B_x) &= \mathrm{Var}Y_x + \mathrm{Var}\,(A + Bx) \\
&= \sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right)
\end{aligned}$$

Abbildung 6.4: Konfidenzintervall für $\mathbb{E}Y_x$



Satz 6.9 Mit der Bezeichnung

$$\begin{aligned}\hat{Y}_x &:= \bar{Y} + B\,(x-\bar{x}) = A + Bx \\ D^2 &:= \sum_{i=1}^{n} \left[\left(Y_i - \bar{Y}\right) - B\,(x_i - \bar{x})\right]^2 = (n-2)\,S^2\end{aligned}$$

gilt:

$$\begin{aligned}T &= \sqrt{n-2}\frac{\bar{Y} + B\,(x-\bar{x}) - \alpha - \beta x}{\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)S_x^2}} \cdot D} \\ &= \sqrt{n-2}\frac{\hat{Y}_x - \alpha - \beta x}{\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)S_x^2}} \cdot D} \\ &\sim t_{n-2}\end{aligned}$$

Daraus konstruiert man ein Konfidenzintervall für $\alpha + \beta x$ an der Stelle x mit Überdeckungswahrscheinlichkeit $1-\gamma$:

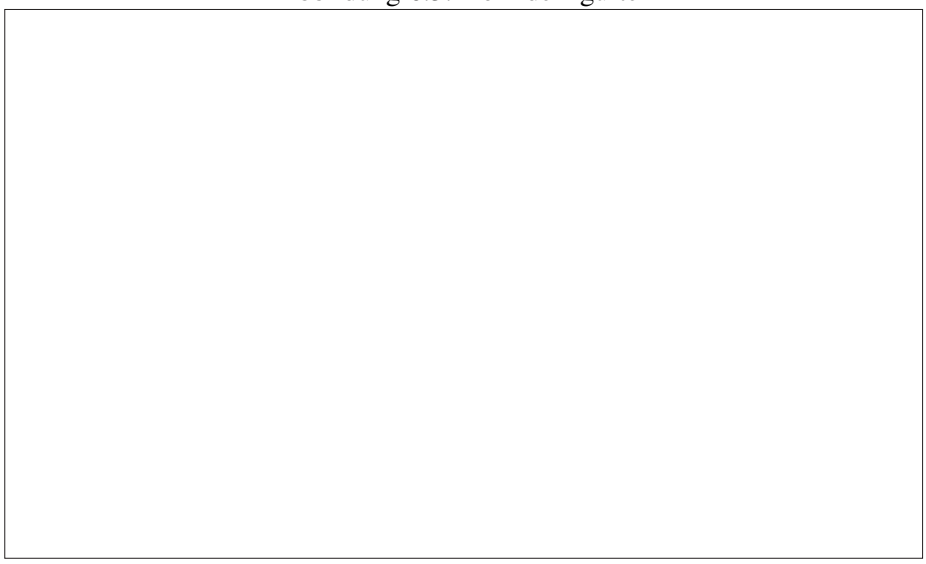
$$\left[\hat{Y}_x - t_{n-2;1-\frac{\gamma}{2}} \cdot \frac{D}{\sqrt{n-2}}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)\,S_x^2}}; \hat{Y}_x + \ldots\right]$$

Ist $y = \hat{\alpha} + \hat{\beta}x$ die empirische Regressionsgerade, so erhält man für jedes x das empirische Konfidenzintervall für $\mathbb{E}Y_x$:

$$\left[\hat{\alpha} + \hat{\beta}x - t_{n-2;1-\frac{\gamma}{2}} \cdot \frac{d}{\sqrt{n-2}}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)\,s_x^2}}; \hat{\alpha} + \hat{\beta}x + \ldots\right]$$

Lässt man x variieren, dann erhält man einen sog. Konfidenzgürtel (vgl. Abbildung 6.5)

Abbildung 6.5: Konfidenzgürtel



Prognoseintervalle für einen zu x gehörigen Wert der stochastischen Größe Y_x.

Satz 6.10 Unter den Vorraussetzungen dieses Abschnittes gilt

$$T = \frac{Y_x - A - Bx}{S\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \sim t_{n-2}$$

Beweis: Vergleiche Folgerung von Satz 6.8. Daraus erhält man durch Umformung von

$$\mathrm{W}\left\{-t_{n-2;1-\frac{\gamma}{2}} \leq T \leq t_{n-2;1-\frac{\gamma}{2}}\right\} = 1 - \gamma$$

ein Konfidenzintervall für Y_x:

$$\left[\hat{\alpha} + \hat{\beta}x - t_{n-2;1-\frac{\gamma}{2}} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}; \hat{\alpha} + \hat{\beta}x + \ldots\right]$$

Bemerkung: Es gibt auch (breitere) simultane Konfidenzbänder.

6.5 Tests für den Parameter von Regressionsgeraden bei Normalverteilung

Vorraussetzungen: $\mathbb{E}Y_x = \alpha+\beta x$, $\forall x \in$ Versuchsbereich und n Beobachtungen (x_i, y_i), $i = 1(1)n$ gilt:

$$\begin{aligned} S_A^2 &= \frac{\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n\right)^2} \cdot S^2 \\ S_B^2 &= \frac{n}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \cdot S^2 \end{aligned}$$

Tests für verschieden Hypothesen sind in folgender Tabelle angegeben.

T1

6.6 Tests für Regressionskurven bei Normalverteilung

Vorraussetzungen: $Y_x \sim \mathrm{N}(\mu(x), \sigma^2)$, σ^2 ist unabhängig von x, (x_ν, y_ν) , $\nu = 1(1)n$, $x_i^* \rightarrow$ m verschieden Werte für x mit Vielfachheit n_i

$$\bar{y}_i = \frac{1}{n_i} \sum_{i=1}^{n} y_{ik}$$

6.7 Test auf Regressionsgerade

Null-Hypothese: $\mathcal{H}_0 : \ \mu(x) = \alpha + \beta x$, wobei α und β nicht bekannt sein müssen.

$$\begin{aligned} \tilde{y} &= A + Bx = \bar{y} + B\,(x - \bar{x}) \\ \tilde{y}_i &= \bar{y} + B\,(x_i^* - \bar{x}) \end{aligned}$$

T2

Zerlegung der Summe:

$$\begin{aligned}
(y_{ik}-\tilde{y}_i)^2 &= [(y_{ik}-\bar{y}_i)+(\bar{y}_i-\tilde{y}_i)]^2 \\
&= (y_{ik}-\bar{y}_i)^2+(\bar{y}_i+\tilde{y}_i)^2+2\,(y_{ik}-\bar{y}_i)\,(\bar{y}_i-\tilde{y}_i) \\
q &= \underbrace{\sum_{i=1}^{m}\sum_{k=1}^{n_i}(y_{ik}-\tilde{y}_i)^2}_{q_2}+\underbrace{\sum_{i=1}^{m}\sum_{k=1}^{n_i}(\bar{y}_i-\tilde{y}_i)^2}_{q_1}+2\sum_{i=1}^{m}(\bar{y}_i-\tilde{y}_i)\underbrace{\sum_{k=1}^{n_i}(y_{ik}-\bar{y}_i)}_{=0}
\end{aligned}$$

q_2=Summe der quadrierten Abweichungen vom jeweiligen Mittel $\bar{y}_i$

q_1=Summe der quadrierten Abweichungen der Schätzwerte von den Mittelwerten

Satz 6.11 von Cockran: Sind $X_1,...,X_n$ unabhängig standard-normalverteilt und gilt

$$\sum_{k=1}^{n}X_k^2=\sum_{i=1}^{m}Q_i\,(X_i,...,X_n)$$

mit Q_i sind nicht negativ symmetrische quadratische Formen mit Rang $(Q_i)\leq r_i$ und $\sum_{i=1}^{m}r_i=n$ dann gilt:

$$\left.\begin{array}{l} Q_i\,(X_1,...,X_n)\sim\chi^2_{r_i} \\ Q_1\,(X_1,...,X_n),...,Q_m\,(X_1,...,X_n) \end{array}\right\}\text{ unabhaengig}$$

Satz 6.12 Für die den Beobachtungen (x_ν,y_ν) entsprechenden stochastischen Größen Y_{ik}, $\bar{Y}_i$ und $\tilde{Y}_i$ gilt:

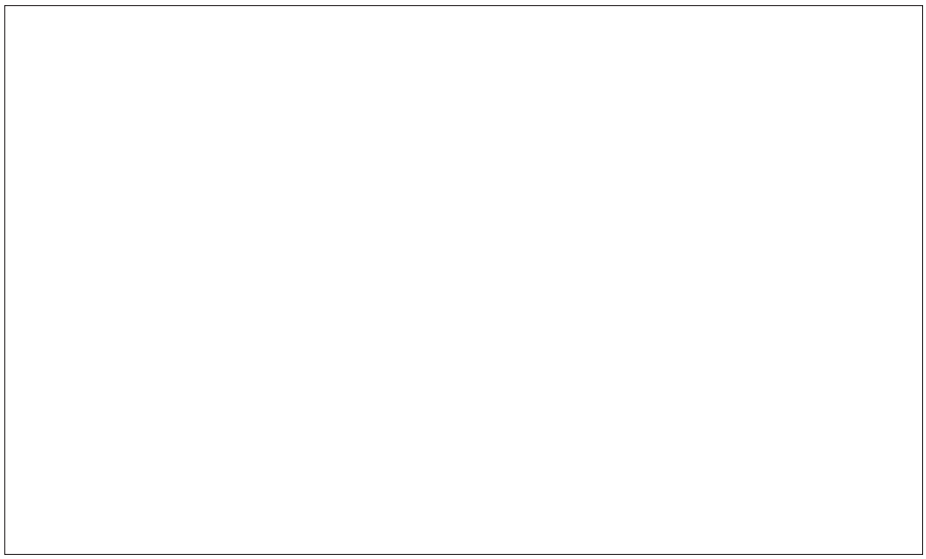
$$\begin{aligned}
\frac{1}{\sigma^2}\sum_{i=1}^{m}n_i\left(\bar{Y}_i-\tilde{Y}_i\right)^2 &= \frac{Q_1}{\sigma^2}\sim\chi^2_{m-2} \\
\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{k=1}^{n_i}\left(Y_{ik}-\bar{Y}_i\right)^2 &= \frac{Q_2}{\sigma^2}\sim\chi^2_{n-m}
\end{aligned}$$

Diese beiden Größen sind unabhängig. Daraus folgt, falls $\mathcal{H}_0$ richtig ist:

$$Z=\frac{Q_1/(m-2)}{Q_2/(n-m)}\sim F_{m-2;n-m}$$

Test: $\mathcal{H}_0$ wird abgelehnt, falls

$$\frac{q_1/(m-2)}{q_2/(m-n)}>F_{m-2;n-m;1-\gamma}$$



6.8 Multiple lineare Regression

k Einstellgrößen $x_1, ..., x_k$, $k+1$ unbekannte Parameter $\theta_0, \theta_1, ..., \theta_k$, Regressionsfunktion $\psi(x_1, ..., x_n) = \theta_0 + \sum_{j=1}^{k} \theta_j \cdot x_j$

$$\begin{aligned} \underline{x} &= (x_1, ..., x_k) \\ \dot{\underline{x}} &= (1, x_1, ..., x_k) \\ \theta &= \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix} \end{aligned}$$

Y: abhängige Größe

$$\begin{aligned} Y_{\underline{x}} &= \dot{\underline{x}} \cdot \theta + U_{\underline{x}} \text{ mit } \mathbb{E}U_{\underline{x}} = 0 \\ \text{Var}Y_{\underline{x}} &\equiv \sigma^2 \end{aligned}$$

Mit den Bezeichnungen der multiplen Ausgleichsrechnung (vgl. Abschnitt 6.2) gilt:

n Beobachtungen $(x_{i1}, x_{i2}, ..., x_{ik}; y_i)$ mit $i = 1(1)n$

$$\underline{y} = \mathcal{X}\theta + \underline{u} \quad \text{mit} \quad \mathcal{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & & x_{2k} \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

$$\underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

Schätzungen aus dem Ausgleichsproblem:

Frage: Statistische Qualität der Schätzer? Schätzung der Varianz σ^2?

Das System der Gleichungen hat die Gestalt

$$\underline{Y} = \mathcal{X}\theta + \underline{U}$$

Vorraussetzungen:

1. Der Rang$\mathcal{X} = k + 1$, daher $n > k$
2. VCov$(U_1, ..., U_n) = \sigma^2 \cdot I_n$ (I_n=1-Matrix)
3. Der Parameterraum darf keine Hyperebene der $\mathbf{R}^{k+1}$ sein.

Definition: Schätzfunktionen T_j für θ_j in einem multiplen linearen Regressionsmodell heißen effiziente lineare Schätzfunktionen, wenn gilt:

1. $T_j = \sum_{i=1}^{n} c_{ji} \cdot Y_i$, wobei die c_{ji} nur von den (bekannten) x_{ij} abhängen.
2. $\mathbb{E}_{\theta,\sigma^2} T_j = \theta_j$ (unverzerrt, erwartungstreu)
3. Var$_{\theta,\sigma^2} T_j$: minimal unter allen Schätzfunktionen, welche die ersten beiden Punkte erfüllen

Analog zu Satz 6.6:

Satz 6.13 (Gauß-Markoff) Unter den Vorraussetzungen:

$$\begin{aligned}
Y_1, ..., Y_n & \quad \text{paarweise unkorreliert} \\
\text{Var} Y_i & \equiv \sigma^2 \quad \forall i = 1(1)n \quad (\text{Homoskedastizitaet}) \\
\mathcal{X} & = \begin{pmatrix} 1 & \underline{x}_1 \\ 1 & \underline{x}_2 \\ \vdots & \vdots \\ 1 & \underline{x}_n \end{pmatrix} \quad \text{bekannte Einstellgroessen } \underline{x}_i,\ i = 1(1)n \\
\theta_0, \theta_1, ..., \theta_k & \quad \text{linear unabhaengig } (\text{nicht in einer Hyperebene der } \mathbf{R}^{k+1}) \\
\mathbb{E} Y_i & = \dot{\underline{x}}_i \cdot \theta = \theta_0 + \sum_{j=1}^{k} x_{ij} \cdot \theta_j \\
\text{Rang}\mathcal{X} & = k + 1
\end{aligned}$$

gilt:

1. $\underline{T} = \begin{pmatrix} T_0 \\ T_1 \\ \vdots \\ T_k \end{pmatrix} = \left(\mathcal{X}^T\mathcal{X}\right)^{-1}\mathcal{X}^T \cdot \underline{Y}$
   sind die eindeutig bestimmen effizienten linearen Schätzfunktionen T_j für θ_j $j = 0(1)k$.
2. VCov $(T_0, T_1, ..., T_k) = \left(\mathcal{X}^T\mathcal{X}\right)^{-1} \cdot \sigma^2$
3. $S^2 := \frac{(\underline{Y}-\mathcal{X}\underline{T})^T(\underline{Y}-\mathcal{X}\underline{T})}{h-k-1}$ ist eine unverzerrte Schätzfunktion für σ^2.

Beweis:

1. Extrema unter Nebenbedingungen. Unter den gegebenen Annahmen hat das Ausgleichsproblem $\exists 1$ Lösung:
$$\hat{\theta}_0, \hat{\theta}_1, ..., \hat{\theta}_k \leftrightarrow T_j,\ j = 0(1)k$$
2. Kovarianzen berechnet man mit Hilfe von
$$\text{Cov}\,(Y_r, Y_s) = \sigma^2\delta_{r,s}...$$
3. $$\begin{aligned}
\mathbb{E}\left[(n-k-1)\,S^2\right] &= \mathbb{E}\left[(\underline{Y}-\mathcal{X}\underline{T})^T(\underline{Y}-\mathcal{X}\underline{T})\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\left(Y_i - T_0 - \sum_{j=1}^{k} x_{ij}T_j\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\left\langle Y_i^2 - 2Y_i\left\{T_0 + \sum_{j=1}^{k} x_{ij}T_j\right\} + \left(T_0 + \sum_{j=1}^{k} x_{ij}T_j\right)^2\right\rangle\right]
\end{aligned}$$
   Einsetzen von $\underline{T} = \left(\mathcal{X}^T\mathcal{X}\right)^{-1}\mathcal{X}^T\underline{Y}$,....

Bemerkung: Die Varianzschätzung ist intuitiv; gemittelte quadrierte Abweichungen von der geschätzten Regressions-Hyperebene:

$$\begin{aligned}
y &= \hat{\theta}_0 + \sum_{j=1}^{k} x_j \cdot \hat{\theta}_j,\ x_j \in \mathbf{R} \\
s^2 &= \frac{1}{n-k-1} \cdot \sum_{i=1}^{n}\left[y_i - \hat{\theta}_0 - \sum_{j=1}^{k} x_{ij} \cdot \hat{\theta}_j\right]^2
\end{aligned}$$

Prognose im linearen Regressionsmodell

$$Y_{\underline{x}} = \underline{\dot{x}}\theta + U_{\underline{x}}$$

Aus Daten geschätzte Parameter $\hat{\theta} = \begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{pmatrix}$

Prognose zum Einstell-Vektor $\underline{x} = (x_1, ..., x_k)$:

$$\hat{Y}_{\underline{x}} := \underline{\dot{x}} \cdot \hat{\theta} = \hat{\theta}_0 + x_1 \cdot \hat{\theta}_1 + ... + x_k \cdot \hat{\theta}_k$$

Für diese Prognose gilt:

1. $\mathbb{E}\hat{Y}_{\underline{x}} = \mathbb{E}Y_{\underline{x}}$ (unverzerrt)
2. $\text{Var}\hat{Y}_{\underline{x}} = \underline{\dot{x}} \left(\mathcal{X}^T \mathcal{X}\right)^{-1} \cdot \underline{\dot{x}}^T \cdot \sigma^2$

Beweis:

1.

$$\begin{aligned} \mathbb{E}\hat{Y}_{\underline{x}} &= \mathbb{E}\left(\underline{\dot{x}}\hat{\theta}\right) \\ &= \mathbb{E}\left[\hat{\theta}_0 + \sum_{j=1}^{k} x_{\hat{j}} \cdot \hat{\theta}_j\right] \\ &= \mathbb{E}\hat{\theta}_0 + \sum_{j=1}^{k} x_j \cdot \mathbb{E}\hat{\theta}_j \\ &= \theta_0 + \sum_{j=1}^{k} x_j \cdot \theta_j \\ &= \mathbb{E}Y_{\underline{x}} \end{aligned}$$

2.

$$\begin{aligned} \text{Var}\hat{Y}_{\underline{x}} &= \text{Var}\left(\underline{\dot{x}}\hat{\theta}\right) \\ &= \text{Var}\left(\underbrace{\underline{\dot{x}}\left(\mathcal{X}^T\mathcal{X}\right)^{-1}\mathcal{X}^T}_{\text{Vektor } \underline{z}} \cdot \underline{Y}\right) \end{aligned}$$

$$
\begin{aligned}
&= \operatorname{Var}\left(\underline{z}\cdot\underline{Y}\right) = \operatorname{Var}\left(\sum_{i=1}^{n} z_i \cdot Y_i\right) \\
&= \sum_{i=1}^{n} \operatorname{Var}\left(z_i Y_i\right) = \sum_{i=1}^{n} z_i^2 \cdot \underbrace{\operatorname{Var} Y_i}_{\sigma^2} \\
&= \sigma^2 \cdot \underline{z} \cdot \underline{z}^T \\
&= \sigma^2 \left(\underline{\dot{x}}\left(\mathcal{X}^T\mathcal{X}\right)^{-1}\mathcal{X}^T\right)\cdot\left(\underline{\dot{x}}\left(\mathcal{X}^T\mathcal{X}\right)^{-1}\mathcal{X}^T\right)^T \\
&= \sigma^2 \cdot \underline{\dot{x}}\left(\mathcal{X}^T\mathcal{X}\right)^{-1}\mathcal{X}^T\mathcal{X}\left(\left(\mathcal{X}^T\mathcal{X}\right)^{-1}\right)^T\underline{\dot{x}}^T \\
&= \sigma^2 \cdot \underbrace{\left(\underline{\dot{x}}\cdot\left(\mathcal{X}^T\mathcal{X}\right)^{-1}\cdot\underline{\dot{x}}^T\right)^T}_{\in\mathbf{R}}
\end{aligned}
$$

6.9 Multiple lineare Regression bei Normalverteilung

Falls die abhängigen Größen $Y_1, ..., Y_n$ im Satz von Gauß-Markoff normalverteilt sind, gilt:

Satz 6.14 Wird zu den Vorraussetzungenvon Gauß-Markoff vorrausgesetzt, dass

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim \mathrm{N}\left(\mathcal{X}\theta, \sigma^2 I_n\right)$$

so folgt:

1. Die T_j sind auch die plausiblen Schätzfunktionen für die Regressionsparameter θ_j, $j = 0(1)k$
2. $$\begin{pmatrix} T_0 \\ T_1 \\ \vdots \\ T_k \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}, \left(\mathcal{X}^T\mathcal{X}\right)^{-1}\cdot\sigma^2\right)$$
3. $$\left.\begin{array}{l} \frac{(n-k-1)S^2}{\sigma^2} \sim \chi^2_{n-k-1} \\ \frac{(\underline{T}-\theta)^T\mathcal{X}^T\mathcal{X}(\underline{T}-\theta)}{\sigma^2} \sim \chi^2_{k+1} \end{array}\right\} \text{unabhaengig}$$

Beweis:

1. Plausibilitätsfunktion

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{\left(y_i-\theta_0-\sum_{j=1}^{k} x_{ij}\theta_j\right)^2}{2\sigma^2}}$$

2. $\underline{T} = \underbrace{\left(\mathcal{X}^T\mathcal{X}\right)^{-1}\mathcal{X}^T}_{\text{konstant}}\underline{Y}$ ist als Lineartransformation einer n-dimensionalen Normalverteilung auch normalverteilt.
3. Quadratsummen von normalverteilten stochastischen Größen (Division durch σ) ergeben χ^2-Verteilungen.

Satz 6.15 Unter den Vorraussetzungen des vorausstehenen Satzes gilt:

$$\left.\begin{array}{r} \underline{Y} - \mathcal{X}\underline{T} = \begin{pmatrix} Y_1 - T_0 - \sum_{j=1}^{k} x_{1j}T_j \\ \vdots \\ Y_n - T_0 - \sum_{j=1}^{k} x_{nj}T_j \end{pmatrix} \\ \hat{\theta} = \underline{T} = \begin{pmatrix} T_0 \\ T_1 \\ \vdots \\ T_k \end{pmatrix} \end{array}\right\} \text{unabhaengig normalverteilt}$$

Satz 6.16 Unter den Vorraussetzungen dieses Abschnittes und der Bezeichnung s^{jj}=j-tes Hauptdiagonalelement der Matrix $\left(\mathcal{X}^T\mathcal{X}\right)^{-1}$ gilt:

1. Konfidenzintervalle für θ_j mit Überdeckungswahrscheinlichkeit $1-\delta$:

$$\left[T_j - t_{n-k-1;1-\frac{\delta}{2}} \cdot \sqrt{s^{jj}} \cdot S\,,\, T_j + \ldots\right]$$

2. Konfidenzintervall für σ^2 mit Überdeckungswahrscheinlichkeit $1-\delta$:

$$\left[\frac{(n-k-1)\,S^2}{\chi^2_{n-k-1;1-\frac{\delta}{2}}}\,,\,\frac{(n-k-1)\,S^2}{\chi^2_{n-k-1;\frac{\delta}{2}}}\right]$$

Beweis:

1. Laut Satz 6.14 und Satz 6.15 gilt:

$$\left.\begin{array}{l} \frac{T_j-\theta_j}{\sqrt{s^{jj}}\cdot\sigma} \sim \mathrm{N}(0,1) \\ \frac{(n-k-1)S^2}{\sigma^2} \sim \chi^2_{n-k-1} \end{array}\right\} \text{unabhaengig}$$

$$\frac{T_j - \theta_j}{\sqrt{s^{jj}} \cdot S} \sim t_{n-k-1}$$

2. Laut Satz 6.14 gilt:

$$\frac{(n-k-1)\,S^2}{\sigma^2} \sim \chi^2_{n-k-1}$$

Rest analog zu ersterem.

Test auf Linearität einer Regressionsfunktion

Für normalverteilte abhängige Größen $Y_i \sim \mathrm{N}\left(\mu\left(\underline{x}_i\right), \sigma^2\right)$ kann man auf Linearität testen, wenn mindestens ein Einstellvektor $\underline{x}_i = (x_{i1}, ..., x_{ik})$, $i = 1(1)n$ vorliegt, zu dem mindestens zwei Beobachtungen $Y_{i,1}$ und $Y_{i,2}$ vorliegen.

Satz 6.17 Gilt für die abhängigen Größen $Y_i \sim \mathrm{N}\left(\dot{\underline{x}}_i \cdot \theta, \sigma^2\right)$ mit $\dot{\underline{x}} = (1, x_{i1}, ..., x_{ik})$

$$\theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix} \qquad \mathcal{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \quad \text{Rang}\mathcal{X} = k+1$$

und bildet man analog zu Abschnitt 6.6 die zu gleichen Einstellvektoren $\underline{x}_i$ gehörenden Mittelwerte der zugehörigen Y-Werte

$$\begin{array}{ll} \underline{x}_a,\ a = 1(1)l & \text{verschiedene } \underline{x}_i - \text{Werte} \\ n_a,\ \sum_{a=1}^{l} n_a = n & \text{mindestens } 1\, n_a \geq 2 \end{array}$$

$$\begin{aligned} \bar{Y}_a &= \frac{1}{n_a} \sum_{\underline{x}_i = \underline{x}_a} Y_i \\ Z &= \frac{(n-l) \sum_{a=1}^{l} n_a \cdot \left(\bar{Y}_a - T_0 - \sum_{j=1}^{k} x_{ij} T_j\right)^2}{(l-k-1) \sum_{a=1}^{l} \sum_{\{\underline{x}_i = \underline{x}_a\}} \left(Y_i - \bar{Y}_a\right)^2} \sim F_{l-k-1, n-l} \end{aligned}$$

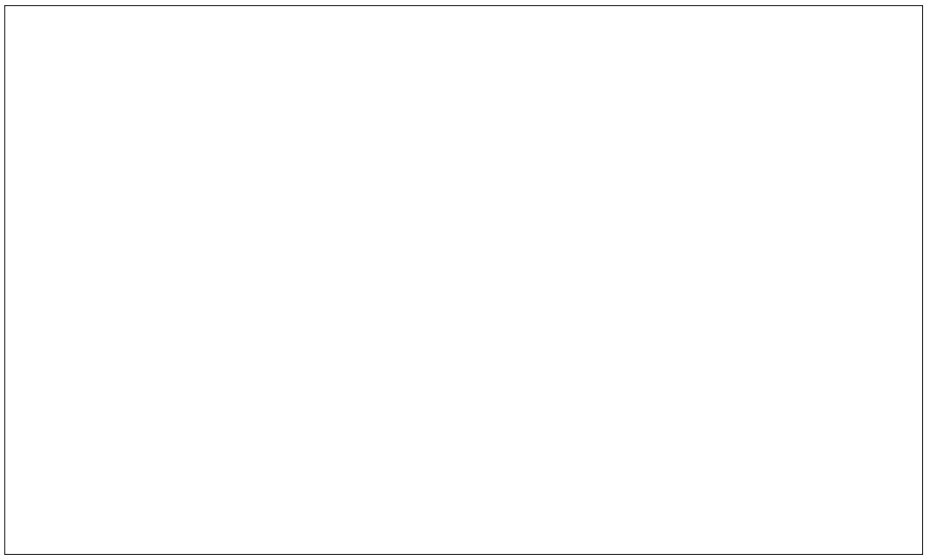
Test: Unter den Vorraussetzungen von Satz 6.17 ist ein Test für die Hypothese $\mathcal{H}_0$

$$\mathcal{H}_0 : \mathbb{E}Y_{\underline{x}} = \theta_0 + \sum_{j=1}^{k} x_{ij} \cdot \theta_j = \dot{\underline{x}}\theta$$

mit Überdeckungswahrscheinlichkeit α eines Fehlers 1.Art durch folgenden Verwerfungsraum V für die Teststatisik Z gegeben:

$$V = [F_{k-k-1, n-l; 1-\alpha}, \infty]$$

Begründung:



Bemerkung: Es ist sinnvoll für große Werte von Z zu verwerfen, da der Erwartungswert des Nenners unabhängig von der besonderen Art der Regressionsfunktion gleich σ^2 ist.
Der Erwartungswert des Zählers ist umso größer, je stärker die tatsächlichen $\mathbb{E}\mu_i$ von den hypothetischen Erwartungswerten abweichen.

Bemerkung: Der Satz 6.12 ergibt sich als Sonderfall.

6.10 Bayes´sche Regressionsanalyse

Auch die Parameter werden durch stochastische Größen beschrieben.

		a − priori
α	$\tilde{\alpha}$	$\pi(\alpha)$
β	$\tilde{\beta}$	$\pi(\beta)$
σ^2	$\tilde{\sigma}^2$	$\pi\left(\sigma^2\right)$
θ	$\tilde{\theta}$	$\pi(\theta)$

Regressionsgeraden

$\mathbb{E}Y_{\underline{x}} = \alpha + \beta x$, $\mathbf{Var}Y_x = \sigma^2$

A-priori Verteilung $\pi\left(\alpha, \beta, \sigma^2\right)$

habe die Dichte $f\left(y|x, \alpha, \beta, \sigma^2\right)$

Daten $D = (x_i, y_i)$, $i = 1(1)n$

Plausibilitätsfunktion $l\left(\alpha, \beta, \sigma^2; D\right)$

$$l\left(\alpha, \beta, \sigma^2; x_1, y_1, ..., x_n, y_n\right) = \prod_{i=1}^{m} f\left(y_i | x_i, \alpha, \beta, \sigma^2\right)$$

Mittels des Bayes´schen Theorems erhält man die a-posteriori Verteilung von $\left(\tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}^2\right)$

$$\pi\left(\alpha, \beta, \sigma^2 | D\right) \propto \pi\left(\alpha, \beta, \sigma^2\right) \cdot l\left(\alpha, \beta, \sigma^2; D\right)$$

Damit kann man HPD-Bereiche

- für einzelne Parameter
- gemeinsame HPD-Bereiche

konstruieren.

Außerdem: Prädiktivverteilungen (Prognoseverteilungen) $Y_x | D$ mit Prognosedichte $f_x\left(\bullet | D\right)$

$$f_x\left(y | D\right) = \int\int\int f\left(y | x, \alpha, \beta, \sigma^2\right) \cdot \pi\left(\alpha, \beta, \sigma^2 | D\right) \mathrm{d}\alpha \mathrm{d}\beta \mathrm{d}\sigma^2$$

Bemerkung: Mittels $f_x\left(\bullet | D\right)$ kann man Prognoseintervalle für Y_x ermitteln.

Spezialfall: $Y_x \sim \mathrm{N}\left(\alpha + \beta x, \sigma^2\right)$

$$\begin{aligned} f\left(y | x, \alpha, \beta, \sigma^2\right) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)^2} \\ \text{Daten } D &= (x_1, y_1, ..., x_n, y_n) \\ l\left(\alpha, \beta, \sigma^2; D\right) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)^2} \\ &= \frac{1}{\left(\sqrt{2\pi}\sigma\right)} e^{\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)^2\right]} \end{aligned}$$

Die Anwendung des Bayes´schen Theorems liefert die a-posteriori Verteilung

$$\pi\left(\alpha, \beta, \sigma^2 | D\right) \propto \pi\left(\alpha, \beta, \sigma^2\right) \cdot \frac{1}{\left(\sigma^2\right)^{n/2}} e^{\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)^2\right]}$$

Daraus berechnet man HPD-Bereiche und Prognosedichten sowie Prognoseintervalle.

Bemerkung: Häufig $\pi\left(\alpha, \beta, \sigma^2\right) = p\left(\alpha, \beta\right) \cdot q\left(\sigma^2\right)$

Multiple lineare Regression

$$\mathbb{E}Y_{\underline{x}} = \underline{\dot{x}}\theta \quad \text{mit} \quad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$$

$Y_{\underline{x}} \sim f\left(\bullet|\underline{x}, \theta, \sigma^2\right)$ $\quad$ $\mathrm{Var}Y_{\underline{x}} = \sigma^2$

$\underline{x} = (x_1, ..., x_k)$ $\quad$ $\underline{\dot{x}} = (1, x_1, ..., x_k)$

Daten $D = (x_i, y_i)\,,\ i = 1(1)n$ $\quad$ mit $\underline{x}_i = (x_{i1}, ..., x_{ik})$

Plausibilitätsfunktion $l\left(\theta, \sigma^2; D\right)$

Für vollständige Daten gilt:

$$l\left(\theta, \sigma^2; \underline{x}_1, \underline{y}_1, ..., \underline{x}_n, \underline{y}_n\right) = \prod_{i=1}^{n} f\left(y_i|\underline{x}_i, \theta, \sigma^2\right)$$

Aufgrund des Bayes´schen Theorems ergibt sich

$$\pi\left(\theta, \sigma^2|D\right) \propto \pi\left(\theta, \sigma^2\right) \cdot l\left(\theta, \sigma^2; D\right)$$

Daraus:

- HPD-Bereiche für Parameter
- Prognosedichten
- Prognoseintervalle

Sonderfall: $Y_{\underline{x}} \sim \mathrm{N}\left(\underline{\dot{x}}\theta, \sigma^2\right)$

$$f\left(y|\underline{x}, \theta, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y - \underline{\dot{x}} \cdot \theta)^2}{2\sigma^2}}$$

7 Varianzanalyse

Es soll untersucht werden, ob ein oder mehrere Größen (genannt Faktoren) Einfluß auf ein beobachtetes Merkmal haben. Die Varianzanalyse umfaßt Testverfahren, die auf einer Analyse des Streuverhaltens empirisch gegebener stochastischer Größen beruhen.

Annahme: Alle beobachtbaren stochastischen Größen werden als normalverteilt mit identischen Varianzen vorrausgesetzt.

7.1 Grundlagen der Varianzanalyse (ANOVA)

Mit jeder symmetrischen, quadratischen n-reihigen MAtrix $A = (a_{ij})$ ist durch

$$\underline{x} = (x_1, ..., x_n)^T \rightarrow Q(\underline{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \cdot x_i \cdot x_j = \underline{x}^T \cdot A \cdot \underline{x}$$

eine symmetrische quadratische Form Q in den n Variablen $x_1, ..., x_n$ festgelegt.

Der Rang von Q ist definiert als der Rang der Matrix A.

Eine symmetrische Matrix B - bzw. die zugeordnete lineare Transformation - heißt orthogonal, wenn

$$B^{-1} = B^T$$

Satz 7.1 (Hauptachsentransformation) Jede symmetrische quadratische Form $Q(\underline{x})$ von Rang r kann durch eine orthogonale Transformation in die Form

$$Q^*(\underline{y}) = \sum_{i=1}^{r} c_i y_i^2 \quad \text{mit } c_1, ..., c_r \neq 0$$

übergeführt werden. Ist Q nichtnegativ, d.h. $Q : \mathbf{R}^n \rightarrow [0, \infty)$, so gilt $c_1, ..., c_r > 0$.

Satz 7.2 (Satz von Cockran) Sind $X_1, ..., X_n$ unabhängig verteilt nach N $(0, 1)$ und

$$\sum_{k=1}^{n} X_k^2 = \sum_{i=1}^{m} Q_i(X_1, ..., X_n)$$

wobei die $Q_i(\cdots)$ nichtnegative symmetrische quadratische Formen mit Rang $(Q_i) \leq r_i$ und $\sum_{i=1}^{m} r_i = n$, so folgt:

1. $Q_i\left(X_1,...,X_n\right) \sim \chi^2_{r_i}$
2. $Q_1\left(X_1,...,X_n\right),...,Q_m\left(X_1,...,X_n\right)$ sind unabhängig.

7.2 Einfache Varianzanalyse (1 Faktor)

Es soll getestet werden, ob m stochastische Größen $X_1,...,X_m$ identische Erwartungswerte haben.

Nullhypothese: $\mathcal{H}_0: \ \mu_1 = \mu_2 = ... = \mu_m$

Zur stochastischen Größe X_i konkrete Stichprobe $x_{i1},...,x_{in_i}$ mit $n_1,...,n_m,\ \sum_{i=1}^m n_i = n$

Gruppe	Stichprobenwerte	Summe
1. Gruppe	$x_{11}, x_{12}, ..., x_{1n_1}$	$x_{1\bullet}$
2. Gruppe	$x_{21}, x_{22}, ..., x_{2n_2}$	$x_{2\bullet}$
$\vdots$		
i. Gruppe	$x_{i1}, x_{i1}, ..., x_{in_i}$	$x_{i\bullet}$
$\vdots$		
m. Gruppe	$x_{m1}, x_{m2}, ..., x_{mn_m}$	$x_{m\bullet}$
		$x_{\bullet\bullet}$

Gruppenwerte $x_{i\bullet} = \sum_{k=1}^{n_i} x_{ik} \quad i = 1(1)m$

Gruppenmittel $\bar{x}_i = \frac{x_{i\bullet}}{n_i}$

Gesamtmittel $\bar{x} = \frac{1}{n}\sum_{i=1}^m n_i \cdot \bar{x}_i = \frac{x_{\bullet\bullet}}{n}$

$s^2 = \frac{1}{n-1}\sum_{i=1}^m \sum_{k=1}^{n_i} \left(x_{ik} - \bar{x}\right)^2$

Quadratsumme $q = (n-1)\cdot s^2$

$$
\begin{aligned}
q &= \sum_{i=1}^m \sum_{k=1}^{n_i} \left(x_{ik} - \bar{x}\right)^2 \\
&= \sum_{i=1}^m \sum_{k=1}^{n_i} \left[\left(x_{ik} - \bar{x}_i\right) + \left(\bar{x}_i - \bar{x}\right)\right]^2 \\
&= \sum_{i=1}^m \sum_{k=1}^{n_i} \left(x_{ik} - \bar{x}_i\right)^2 + \sum_{i=1}^m \sum_{k=1}^{n_i} \left(\bar{x}_i - \bar{x}\right)^2 + \underbrace{2\sum_{i=1}^m \sum_{k=1}^{n_i} \left(x_{ik} - \bar{x}_i\right)\left(\bar{x}_i - \bar{x}\right)}_{0}
\end{aligned}
$$

$$
\sum_{i=1}^m \sum_{k=1}^{n_i} \left(x_{ik} - \bar{x}\right)\left(\bar{x}_i - \bar{x}\right) = \sum_{i=1}^m \left(\bar{x}_i - \bar{x}\right) \cdot \sum_{k=1}^{n_i} \left(x_{ik} - \bar{x}_i\right)
$$

$$= \underbrace{\sum_{i=1}^{m} (\bar{x}_i - \bar{x}) \underbrace{\left[\underbrace{\left(\sum_{k=1}^{n_i} x_{ik} \right)}_{n_i \cdot \bar{x}_i} - n_i \cdot \bar{x}_i \right]}_{0}}_{0}$$

Wegen

$$\sum_{k=1}^{n_i} (\bar{x}_i - \bar{x})^2 = n_i \cdot (\bar{x}_i - \bar{x})^2$$

gilt:

$$\begin{aligned} q &= \sum_{i=1}^{m} \sum_{k=1}^{n_i} (x_{ik} - \bar{x})^2 \\ &= \underbrace{\sum_{i=1}^{m} n_i (\bar{x}_i - \bar{x})^2}_{q_1} + \underbrace{\sum_{i=1}^{m} \sum_{k=1}^{n_i} (x_{ik} - \bar{x}_i)^2}_{q_2} \end{aligned}$$

q_1=Summe der Abweichungsquadrate zwischen den Gruppen

q_2 =Summe der Abweichungsquadrate innerhalb der Gruppen

Falls die Hypothese $\mathcal{H}_0 : \mu_1 = \mu_2 = ... = \mu_m = \mu$ richtig ist, gilt $X_{ik} \sim \mathrm{N}\left(\mu, \sigma^2\right)$ und $\underline{X} = (X_{11}, ..., X_{1n_1}; ...; X_{m1}, ..., X_{mn_m})$ ist eine Stichprobe von X mit Umfang $n = \sum_{i=1}^{m} n_i$.

Bemerkung: Eine unerzerrte Schätzfunktion für σ^2:

$$\frac{1}{n-1} Q = \frac{1}{n-1} \sum_{i=1}^{n} \sum_{k=1}^{n_i} \left(X_{ik} - \bar{X}_n \right)^2$$

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

$\frac{Q}{\sigma^2}$ hat den Rang $n-1$ und wegen $Q = Q_1 + Q_2$ folgt: $\frac{Q}{\sigma^2} = \frac{Q_1}{\sigma^2} + \frac{Q_2}{\sigma^2}$ und Rang$\frac{Q_1}{\sigma^2} \leq m-1$ und Rang$\frac{Q_2}{\sigma^2} \leq n-m$.

Nach dem Satz von Cockran folgt, dass $\frac{Q_1}{\sigma^2} \sim \chi^2_{n-1}$ und $\frac{Q_2}{\sigma^2} \sim \chi^2_{n-m}$ unabhängig sind.

Daraus folgt:

$$T = \frac{Q_1 / (m-1)\, \sigma^2}{Q_2 / (n-m)\, \sigma^2} = \frac{Q_1 / (m-1)}{Q_2 / (n-m)} \sim F_{m-1;n-m}$$

Verwerfungsraum für $\mathcal{H}_0$

$$V = \left\{ (x_{11}, ..., x_{mn_m}) : \frac{q_1/(m-1)}{q_2/(n-m)} \geq F_{m-1,n-m;1-\alpha} \right\}$$

mit einer Wahrscheinlichkeit α eines Fehlers 1.Art.

7.3 Zweifache Varianzanalyse

Es soll gleichzeitig der Einfluß zweier Faktoren („Einflußgrößen“) auf ein beobachtetes Merkmal untersucht werden.

Beispiel: Betonqualität abhängig von Kornmischung des Schotters und der Wasserbeigabe.

- Kornverteilung: Faktor A (Zeilenfaktor)
  mögliche Werte $\alpha_1, ..., \alpha_a$
- Wasserbeigabe: Faktor B (Spaltenfaktor)
  mögliche Werte $\beta_1, ..., \beta_b$

Zu jeder Variante des Zeilenfaktors A und des Spaltenfaktors B werden gleich viele c Beobachtungen gemacht.
x_{ijk}: k-ter beobachteter Wert zur i-ten Variante von A und j-ten Variante von B.

Beispiel: Jeder von a Geodäten führt mit jedem vom b Meßgeräten genau c Messungen am selben Objekt durch. Frage:

1. Gibt es bezüglich der Geodäten (Zeilenfaktor) bzw. zwischen den Geräten (Spaltenfaktor) Unterschiede?
2. Gibt es Geodäten, denen ein bestimmtes Gerät mehr liegt als ein anderes, d.h. ob es eine Wechselwirkung zwischen Geodäten und Messgeräten gibt.

Modell:

$$X_{ijk} = \mu + \alpha_i + \beta_i + \gamma_{ij} + \epsilon_{ijk} \qquad i = 1(1)a,\ j = 1(1)b,\ k = 1(1)c$$

Dabei bezeichnet γ_{ij} die Wechselwirkung.

Stochastisches Modell:

$$X_{ijk} = \mu + \alpha_i + \beta_i + \gamma_{ij} + U_{ijk} \qquad U_{ijk} \sim \mathrm{N}\left(0, \sigma^2\right)$$

Die Parameter genügen folgenden Bedingungen:

$$\mu \in \mathbf{R} \quad \begin{array}{ll} \sum_{i=1}^{a} \alpha_i = 0 & \sum_{j=1}^{b} \beta_i = 0 \\ \sum_{i=1}^{a} \gamma_{ij} = 0 & \forall j = 1(1)b \\ \sum_{j=1}^{b} \gamma_{ij} = 0 & \forall i = 1(1)a \end{array}$$

Kopien 6-7 bis 6-11