Mathematische Statistik
SS 2002

Inhalt

§1 Einführung

Kapitel 1: Schätztheorie
§2 Grundbegriffe der Schätztheorie
§3 Lineare Modelle

Kapitel 2: Testtheorie
§4 Grundbegriffe der Testtheorie
§5 Verteilungen einiger wichtiger Statistiken
§6 Konfidenzintervalle

Kapitel 3: Optimale Testverfahren
§7 Exponentialfamilien und Familien mit monotonen Dichtequotienten
§8 Das Testen zweier einfacher Hypothesen
§9 Einseitige Tests bei monotonen Dichteqotienten
§10 Invarianz
§11 Invariante Tests in linearen Modellen

§1 Einführung.

Was ist Statistik ? Die Statistik beschäftigt sich mit Prozessen der Informationsgewinnung durch geeignete Datenerhebungen und der Interpretation der gewonnen Ergebnisse. Eine solche Interpretation könnte etwa zu der Erkenntnis führen, daß jemand, der viel raucht, verstärkt mit Krebserkrankungen rechnen muß.

In der Umgangssprache versteht man unter Statistik meist eine **Zusammenfassung und Darstellung von Daten**, beispielsweise in der Meinungsforschung oder in der Medizin. Diese Form der Statistik wird **deskriptive Statistik** genannt.

Die **Mathematische Statistik** geht von einer abstrakten Formulierung eines konkreten Problems in einem mathematischen Modell aus. Ihr Ziel ist es, in dem Modell Verfahren zu entwickeln, die es in einer konkreten Situation gestatten, aufgrund gegebener Daten sogenannte **statistische Entscheidungen** zu treffen. Man spricht auch von **schließender** oder **analytischer Statistik**. Angesichts unvollständiger Information werden die Entscheidungen mit dem Risiko von Fehlentscheidungen verbunden sein. Die Diskussion der Konsequenz von möglichen Fehlentscheidungen und die Präzisierung des Risikos wird zum Verständnis fundamentaler statistischer Ideen führen. Bei dem Bemühen, die auftretende Fehler in einem geeigneten Sinn klein zu halten, werden Optimierungsmethoden verwendet.

Man hat die folgende Ausgangssituation. Ein Experiment liefert ein Ergebnis x, auch **Stichprobe** genannt, aus einem Raum S versehen mit einer σ–Algebra $\mathfrak{S}$. (In der W–theorie verwendet man ω und Ω anstelle von x und S.)

Der Ausgang des Experiments birgt eine **doppelte Ungewißheit** in sich:

1. Gewisse Parameter sind unbekannt (z.B. Erwartungswert und Varianz).
2. Selbst bei bekannten Parametern ist das Resultat des Experiments nicht vorhersagbar.

Problematik 2 bezieht sich auf die Zufälligkeit des Experiments und läßt sich mit Mitteln der W–Theorie erfassen. Sind die Parameter bekannt, so läßt sich das Experiment durch eine W–Verteilung beschreiben. Dies ist eine **Grundannahme der Statistik**.

Problematik 1 bedeutet eine unvollständige Information und ist charakteristisch für die Statistik . Sie wird dadurch berücksichtigt, daß man eine **Familie $\mathfrak{P}$ von W–Maßen** auf $(S,\mathfrak{S})$ vorausgesetzt wird. Das Gesamtexperiment wird sich aus n Einzelexperimenten zusammensetzen. Dabei wird es sich oft um Wiederholungen eines Experiments unter gleichen Bedingungen handeln. Man wird in der Regel umso mehr Information über das W–Maß, daß ein Experiment beschreibt, erhalten, je öfter man dieses wiederholt.

Definition 1. $(S,\mathcal{S},\mathfrak{P})$ heißt ein **statistischer Raum**, wenn $(S,\mathcal{S})$ ein Meßraum und $\mathfrak{P}$ eine Familie von W–Maßen auf $(S,\mathcal{S})$ ist. Dabei heißen $(S,\mathcal{S})$ der **Stichprobenraum** und $\mathfrak{P}$ die **Verteilungsannahme**. Es wird stets die folgende Situation vorliegen: $S := \overset{n}{\underset{1}{\times}} S'$, d.h. $S \ni x = (x_1,...,x_n)$, $\mathcal{S} := \overset{n}{\underset{1}{\otimes}} \mathcal{S}'$; dabei heißt n der **Stichprobenumfang**. Dann sei $X_i : S \mapsto S'$ die Projektion mit $X_i(x) = x_i$.

Versieht man $(S,\mathcal{S})$ mit einem $P \in \mathfrak{P}$, so ist X_i also eine Zva auf dem W–Raum $(S,\mathcal{S},P)$ mit Werten in S' und beschreibt das Ergebnis des i–ten Einzelexperiment. Die Einzelexperimente werden stets so ausgeführt, daß sie sich nicht gegenseitig beinflussen und somit durch unabhängige Zva beschrieben werden können. Deswegen sei hier an den kanonischen W–Raum erinnert, auf dem n unabhängige Zva mit vorgegebenen Verteilungen definiert sind.

Erinnerung. Zu geg. W–Räumen $(S',\mathcal{S}',P_i)$ sei definiert: $S := \overset{n}{\underset{1}{\times}} S'$, $\mathcal{S} := \overset{n}{\underset{1}{\otimes}} \mathcal{S}'$,
$X_i : S \mapsto S'$ die i–te Projektion. Dann sind äquivalent:

(i) $X_1,...,X_n$ sind unabhängige Zva auf $(S,\mathcal{S},P)$ mit $X_i \sim P_i$;

(ii) $P := \overset{n}{\underset{1}{\times}} P_i$.

Dabei bedeutet $X_i \sim P_i$: P_i ist die Verteilung PX_i^{-1} von X_i unter P**.** Ferner werden wir wie üblich iid für unabhängig und identisch verteilt (independent, identically distributed) schreiben.

Beispiel 1.

Zur Heilung einer bestimmten Krankheit wird ein Serum entwickelt. Das Experiment bestehe in einer Anwendung des Serums bei n Patienten, wobei sich als Stichprobe x ein Vektor $x = (x_1,...,x_n)$ ergibt mit

$$x_i = \begin{cases} 1 & \text{geheilt wird} \\ \quad \text{, falls i–ter Patient} & \\ 0 & \text{krank bleibt.} \end{cases}$$

Der statistische Raum $(S,\mathcal{S},\mathfrak{P})$ ließe sich etwa wählen als:

$S = \{0,1\}^n$, $\mathcal{S}$ Potenzmenge auf S, $\mathfrak{P} = \{ P_\vartheta = \overset{n}{\underset{i=1}{\times}} P'_\vartheta := \overset{n}{\underset{i=1}{\times}} b(1,\vartheta),\ \vartheta \in [0,1] \}$.

Unter jedem P_ϑ sind die Zva $X_1,...,X_n$ also iid mit $X_i \sim b(1,\vartheta)$; X_i beschreibt dabei das Ergebnis der Behandlung beim i–ten Patienten. []

Beispiel 2.

Eine Maschine füllt Zucker in Tüten ab. Das tatsächliche Tütengewicht schwankt um den Mittelwert a mit einer Streuung $\sigma^2>0$, wobei a der Justierung und σ^2 einer Maschinenkonstante entspricht. Zur Kontrolle der Justierung wird eine Stichprobe $x = (x_1,...,x_n)$ vom Umfang n entnommen. Hierbei sei x_i das Gewicht der i–ten Tüte.

Als Stichprobenraum wählt man etwa $(S,\mathfrak{S}) = (\mathbb{R}^n,\mathfrak{B}_n)$ und für die Verteilungsannahme

$\mathfrak{P} = \{P_\vartheta = \bigotimes_{i=1}^n P'_\vartheta := \bigotimes_{i=1}^n N(a,\sigma_o^2);\ \vartheta=a \in \Theta = (0,\infty)\}$, falls σ_o^2 bekannt ist, sonst:

$\mathfrak{P} = \{ P_\vartheta = \bigotimes_{i=1}^n P'_\vartheta := \bigotimes_{i=1}^n N(a,\sigma^2);\ \vartheta=(a,\sigma^2) \in \Theta = (0,\infty)\times(0,\infty)\}$.

Unter jedem P_ϑ sind die Zva $X_1,...,X_n$ also iid mit $X_i \sim N(a,\sigma_o^2)$ bzw. $X_i \sim N(a,\sigma^2)$; X_i beschreibt dabei das Ergebnis der Messung des Gewichts der i–ten Tüte. []

Die Normalverteilung wurde als erste zur Beschreibung von Abweichungen von einem Sollwert benutzt, obgleich die Symmetrie und der Träger $\mathbb{R}$ in praktischen Problemen oft nicht gerechtfertigt sind. Die Normalverteilung hieß ursprünglich **Fehlergesetz** und entstand aus der Arbeit der Astronomen und Mathematiker des 18. und 19. Jahrhundert bei dem Bemühen, Beobachtungsfehler zu berücksichtigen.

Die W–theorie befaßt sich in der Regel mit Aussagen darüber, was bei einem zufallsabhängigen Experiment zu erwarten ist, also mit Arten von Prognosen. Etwa: In soundsoviel Prozent der Fälle ist mit einem bestimmten Ereignis zu rechnen.

Die Statistik zieht hingegen mit Hilfe einer Stichprobe Rückschlüsse auf das W–Maß, das das Experiment am besten beschreibt. Dabei versetzt man sich in die Lage, daß irgendein $P\in\mathfrak{P}$ das "wahre" W–Maß ist, und überlegt, was unter P zu erwarten ist, welche Stichproben etwa mit welcher Häufigkeit zu beobachten sein werden, und vergleicht die Prognose mit der tatsächlich eingetretenen Stichprobe.

Einerseits baut die Statistik auf der W–theorie auf gemäß der Grundannahme der Statistik. Andererseits muß bei der Festlegung eines w–theoretischen Modells für einen konkreten Vorgang das geeignete W–Maß oft erst mit statistischen Verfahren bestimmt werden.

Definition 2. $\mathfrak{P}$ heißt **parametrisiert** gemäß $\mathfrak{P} = \{P_\vartheta , \vartheta\in\Theta\}$, wenn die Zuordnung $\Theta \ni \vartheta \mapsto P_\vartheta \in \mathfrak{P}$ bijektiv ist. Die Menge Θ heißt dann **Parameterraum**. Ein **Parameter** ρ ist eine Abbildung $\rho : \Theta \mapsto R$ von dem Parameterraum Θ in eine beliebige Menge R.

Die Parametrisierung $\mathfrak{P} = \{P_\vartheta , \vartheta\in\Theta\}$ ist in der Statistik sehr zweckmäßig und immer möglich.
Im Beispiel 2 wäre zu $\vartheta=(a,\sigma^2)$ dann $\rho(\vartheta)=a$ ein möglicher Parameter von $\mathfrak{P}$.

Wichtig für einen Informationsgewinn über ein Experiment sind wie erwähnt Wiederholungen, die unabhängig voneinander unter gleichen Bedingungen durchgeführt werden.

Obige Beispiele sind sogenannte Einstichprobenprobleme. Werden Experimente unter k verschiedenen Versuchsbedingungen durchgeführt, so spricht man von k–Stichprobenproblemen im Sinne der folgenden Definition.

Definition 3.

(i) Ein statistischer Raum $(S,\mathfrak{S},\mathfrak{P} = \{P_\vartheta\ ,\ \vartheta\in\Theta\})$ beschreibt ein **Einstichprobenproblem**, wenn n Einzelexperimente unter der gleichen Versuchsbedingung gemacht werden. Dann sind unter jedem W–Maß P_ϑ $X_1,...,X_n$ iid Zva. Dann existiert also eine Familie von W–Maßen $\mathfrak{P}' = \{P'_\vartheta\ ,\ \vartheta\in\Theta\}$ auf $(S',\mathfrak{S}')$ mit $X_i \sim P'_\vartheta$ unter $\vartheta \in \Theta$, also

$$E_\vartheta[g(X_i)] = \int g(x')\, P'_\vartheta[dx'] \quad \text{im Falle des Existenz des Erwartungswertes.}$$

Dann gilt $\mathfrak{P} = \{P_\vartheta = \bigtimes_{i=1}^{n} P'_\vartheta,\ \vartheta \in \Theta\}$ und wie immer $S = \bigtimes_{i=1}^{n} S'$, $\mathfrak{S} = \bigotimes_{i=1}^{n}\mathfrak{S}'$.

(ii) Ein statistischer Raum $(S,\mathfrak{S},\mathfrak{P} = \{P_\vartheta\ ,\ \vartheta\in\Theta\})$ beschreibt ein **k–Stichprobenproblem**, wenn k Versuchsbedingungen untersucht werden und jeweils n_i Einzelexperimente unter der i–ten Versuchsbedingung gemacht werden mit $n = \sum n_i$. Man benutzt dann die Indizierung $(X_1,...,X_n) = (X_{11},...,X_{1n_1},...,X_{k1},...,X_{kn_k})$. Dabei beschreibt X_{ij} das j–te Ergebnis unter der i–ten Versuchsbedingung. Für $x\in S$ bietet sich entsprechend die Bezeichnung $x = (x_{11},...,x_{1n_1},...,x_{k1},...,x_{kn_k})$ an. Die Verteilungsannahme besagt gerade:

unter P_ϑ sind $X_{11},...,X_{1n_1},...,X_{k1},...,X_{kn_k}$ unabhängige Zva und $X_{i1},...,X_{in_i}$ sind jeweils identisch verteilt. [Dann existieren also Familien $\mathfrak{P}_i$ auf $(S',\mathfrak{S}')$ mit

$\mathfrak{P} = \{P_\vartheta = \bigtimes_{i=1}^{k} (\bigtimes_{j=1}^{n_i} P_{i\vartheta})\ ;\ P_{i\vartheta}\in\mathfrak{P}_i,\ \vartheta \in \Theta\}$ und wie immer $S = \bigtimes_{i=1}^{n} S'$, $\mathfrak{S} = \bigotimes_{i=1}^{n}\mathfrak{S}'$.]

Jedes $P_\vartheta \in \mathfrak{P}$ läßt sich also darstellen in der Form $P_\vartheta = P_{1\vartheta}\times...\times P_{1\vartheta}\times\ .\ .\ .\times P_{k\vartheta}\times...\times P_{k\vartheta}$; unter P_ϑ gilt somit $X_{ij} \sim P_{i\vartheta}$ und

$$E_\vartheta[g(X_{ij})] = \int g(x')\, P'_{i\vartheta}[dx'] \quad \text{im Falle des Existenz des Erwartungswertes.}$$

Der häufigste Fall eines k–Stichprobenproblems bei k>1 ist der Fall k=2. Dabei handelt es sich dann in der Regel um den Vergleich zweier Behandlungsmethoden.

Beispiel 3. Es sollen zwei Seren ausprobiert und ihre Wirksamkeit verglichen werden. Dazu werden n_i Patienten mit dem Serum i behandelt. Für den Stichprobenumfang ergibt sich $n=n_1+n_2$ und für den statistischer Raum bietet sich folgendes an:

$$S = \{0,1\}^n \text{ und } \mathfrak{P} = \{\ P_\vartheta = \bigtimes_{i=1}^{n_1} b(1,\vartheta_1) \times \bigtimes_{i=1}^{n_2} b(1,\vartheta_2)\ ;\ \vartheta=(\vartheta_1,\vartheta_2)\in [0,1]^2\ \}.$$

Dabei wird S natürlich mit der Potenzmenge $\mathfrak{S}$ als σ–Algebra versehen. Die Verteilungsannahme besagt gerade: unter $\vartheta \in \Theta$ sind $X_{11},...,X_{1n_1},X_{21},...,X_{2n_2}$ $\{0,1\}$–wertige unabhängige Zva auf $(S,\mathfrak{S},P_\vartheta)$ und $X_{i1},...,X_{in_i}$ sind jeweils identisch verteilt mit $X_{ij} \sim b(1,\vartheta_i)$, also

$$E_\vartheta[g(X_{ij})] = g(0)\cdot(1-\vartheta_i) + g(1)\cdot\vartheta_i.$$

Eine weitere Interpretation wäre die folgende. Man untersucht Krebserkrankungen bei Rauchern und Nichtrauchern . Dann wäre ϑ_i die Wahrscheinlichkeit für eine Krebserkrankung bei einem Raucher (i=1) bzw. bei einem Nichtraucher (i=2). $[]$

Das Ziel der Mathematischen Statistik ist, wie bereits erwähnt, Verfahren zu entwickeln, die es ermöglichen, statistische Entscheidungen zu fällen. Es sollen drei mögliche Verfahren angegeben werden.

(i) (**Schätzverfahren**) Zu einem Parameter $\rho : \Theta \mapsto R$ ist ein **Schätzer** als Abbildung $\hat{\rho} : S \mapsto R$ definiert. Ein Schätzer legt also aufgrund einer Stichprobe x einen Schätzwert $\hat{\rho}(x)$ fest.

Im Beispiel 1 wäre $\hat{\rho}(x) := \frac{1}{n}\sum_1^n x_i$, die mittlere Anzahl der Erfolge, ein geeigneter Schätzer. Die Herleitung und das Studium von Schätzern ist Aufgabe der **Schätztheorie**.

(ii) (**Testverfahren**) Der Parameterraum sei disjunkt zerlegt in zwei Teilmengen: $\Theta = H \cup K$.
Im Beispiel 1 könnte etwa $H = [0,\vartheta_o]$ und $K = (\vartheta_o,1]$ sein, wenn man sich dafür interessiert, ob die Heilungschance größer als ϑ_o ist.
Soll im Beispiel 2 der Sollwert für das Gewicht einer Tüte gerade 500 g sein, so wäre $H = \{500\}$ bei bekannter Varianz bzw. $H = \{(500,\sigma^2);\sigma^2>0\}$ bei unbekannter Varianz und $K = \Theta\backslash H$ eine geeignete Zerlegung. Man möchte dann wissen, ob die Abfüllmaschine richtig justiert ist oder nicht.
Im Beispiel 3 wäre durch $H = \{(\vartheta_1,\vartheta_2);\ \vartheta_1 \le \vartheta_2\}$ eine geeignete Zerlegung von Θ gegeben.
In der **Testtheorie** stellt die Zerlegung $\Theta = H \cup K$ ein **Testproblem** dar. Man definiert einen Test $\varphi : S \mapsto \{0,1\}$ mit der Maßgabe: $\varphi(x)=0$ oder 1 bedeutet eine Entscheidung für H bzw. K aufgrund der vorliegenden Stichprobe x.
Im Beispiel 1 würde von einem Test φ die naheliegende Eigenschaft erwartet, daß man sich für K entscheidet, also dafür daß die Heilungschance groß ist, wenn die Anzahl der Heilerfolge groß ist. Der Test hätte dann die Gestalt

$$\varphi(x) = \begin{cases} 1 & \\ \text{falls} & \sum x_i > c \\ 0 & \end{cases};$$

dabei ist es eine Aufgabe der Statistik, c geeignet festzulegen.
(iii) Sei $\rho:\Theta\mapsto R$ ein Parameter. Ein **Bereichsschätzer** (**Konfidenzbereich**) für ρ gibt aufgrund einer Stichprobe x einen Bereich $C(x) \subset R$ als Schätzung an. Für den Fall $R\subset\mathbb{R}$ heißt $C(x) = [\underline{\lambda}(x),\bar{\lambda}(x)]$ ein **Konfidenzintervall** für $\rho(\vartheta)$. Eine Angabe der Form $\rho(\vartheta) = \hat{\rho}\pm\varepsilon$ ist gerade als Angabe eines Konfidenzintervalls $[\hat{\rho}-\varepsilon,\hat{\rho}+\varepsilon]$ für $\rho(\vartheta)$ zu interpretieren.

Zwischen guten Tests und guten Bereichsschätzern besteht ein enger Zusammenhang.

Kapitel 1. Schätztheorie.

§2 Grundbegriffe der Schätztheorie.

Es liege wieder die Situation aus §1 Definition 1 vor.

Definition 1. Es sei ein Parameter $\rho : \Theta \mapsto R$ gegeben. Dann heißt

(i) ρ **reeller Parameter**, falls $R \subset \mathbb{R}$,

(ii) $\hat{\rho}$ **Schätzer** oder **Schätzfunktion** für ρ, falls $\hat{\rho} : S \mapsto R$.

Beispiel 1.

Es liege ein Einstichprobenproblem vor im Sinne von §1 Definition 3.

Es gelte: $S' \subset \mathbb{R}$, $\mathfrak{S}' = S' \cap \mathfrak{B}_1$, $E_\vartheta[|X_i|^m] < \infty \ \forall \vartheta \in \Theta$, i=1,...,n und m=1 bzw. 2. Dann sind

$a(\vartheta) := E_\vartheta[X_i]$ im Falle m=1 und

$\sigma^2(\vartheta) := \mathrm{Var}_\vartheta[X_i]$ im Falle m=2 reelle Parameter.

Als Schätzer für a und σ^2 bieten sich an:

das **Stichprobenmittel** $\bar{X} := \bar{X}_{(n)} := \frac{1}{n} \sum_{i=1}^n X_i =: \hat{a}$ und

die **Stichprobenstreuung** $S^2 := S^2_{(n)} := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 =: \hat{\sigma}^2$. []

[Der Wahl von 1/(n–1) statt 1/n wird bald erklärt werden.] []

Es sollen nun einige Prinzipien, nach denen man Schätzer auswählen kann, sowie einige Gütekriterien für Schätzer vorgestellt werden.

2 A: Maximum–Likelihood–Schätzer

Es liege wieder ein Einstichprobenproblem vor im Sinne von §1 Definition 3.

Liegen für die Verteilungsannahme Dichten vor, so kann ein wichtiges Prinzip für das Schätzen von ϑ eingeführt werden [also für den Fall $\rho(\vartheta)=\vartheta$, $\vartheta \in \Theta = R$].

Es sei dabei μ' ein σ–endliches Maß auf S' und $p'(\cdot,\vartheta)$ eine μ'–Dichte von P'_ϑ, d.h.

$P'_\vartheta(dx') = p'(x',\vartheta)\, \mu'(dx') \ \forall \vartheta \in \Theta$.

Dann gilt (vgl. W.th. I 11.17): $P_\vartheta \overset{n}{\underset{i=1}{\times}} P'_\vartheta$ hat eine μ–Dichte $p(\cdot,\vartheta)$ mit

$\mu := \overset{n}{\underset{i=1}{\times}} \mu'$ und $p(x,\vartheta) := \prod_{i=1}^n p'(x_i,\vartheta)$ also: $P_\vartheta[dx] = (\overset{n}{\underset{i=1}{\times}} P'_\vartheta)[dx] = p(x,\vartheta)\, \mu[dx]$.

Zwei wichtige und typische Fälle sind:

(i) Es sei $S' \in \mathfrak{B}_d$ und μ' die Restriktion des d–dimensionalen Lebesgue–Maßes auf S'. Dann sind $p'(\cdot,\vartheta)$ und $p(\cdot,\vartheta)$ gerade Lebesgue–Dichten.

(ii) Es sei S' eine abzählbare Menge und μ' das Zählmaß auf S'. Dann sind $p'(x',\vartheta) = P'_\vartheta[\{x'\}]$ und $p(x,\vartheta) = P_\vartheta[\{x\}]$ gerade Zähldichten.

Definition 2. In der obigen Situation heißt die Funktion $p(x,\cdot)$ auf Θ **Likelihood–Funktion.**

Jeder Wert $t =: \hat{\vartheta}(x) \in \Theta$ mit

$$p(x,t) := \max_{\vartheta\in\Theta} p(x,\vartheta)$$

heißt **Maximum–Likelihood–Schätzung** für ϑ zu gegebener Stichprobe x, und im Falle der Existenz heißt $\hat{\vartheta}$ **Maximum–Likelihood–Schätzer** (MLS). (Dieser ist nicht eindeutig definiert, wenn mehrere Maximumstellen der Likelihood–Funktion vorliegen.)

Im Falle von Zähldichten wird einer Stichprobe x nach dem Maximum–Likelihood–Prinzip also gerade ein solcher Parameter als Schätzwert zugeordnet, unter dem x mit der maximalen W. eintritt.

Anstelle der Likelihood–Funktion kann man auch die sogenannte **log–Likelihood–Funktion** $\log p(x,\cdot)$ über $\vartheta\in\Theta$ maximieren. Dann wird aus dem Produkt $\Pi_{i=1}^{n}\, p'(x_i,\vartheta)$ eine in der Regel einfacher zu behandelnde Summe. Im Falle $\vartheta=(\vartheta_1,...,\vartheta_d)\in\Theta \subset \mathbb{R}^d$ führen oft Lösungen der **Likelihood–Gleichungen** zum Ziel:

$$\frac{\partial}{\partial\vartheta_i} \log p(x,\vartheta) = 0 \quad 1\le i\le d\ .$$

2 B. Erwartungstreue.

Definition 3. Ein mb Schätzer $\hat{\rho} : S \mapsto R \subset \mathbb{R}$ für einen reellen Parameter $\rho : \Theta \mapsto R$ heißt **erwartungstreu (unverzerrt)**, wenn gilt: $E_\vartheta[\hat{\rho}] = \rho(\vartheta) \quad \forall\, \vartheta \in \Theta.$

Ein erwartungstreuer Schätzer liefert, unabhängig davon welcher Parameter $\vartheta\in\Theta$ zugrunde liegt, im Mittel den richtigen Wert $\rho(\vartheta)$, weist also keinen systematischen Fehler auf.

Beispiel 1 (Fortsetzung)**.**

Das Stichprobenmittel $\bar{X}$ und die Stichprobenstreuung S^2 sind erwartungstreue Schätzer für $a(\vartheta)$ bzw. $\sigma^2(\vartheta)$. Der Faktor $\frac{1}{n-1}$ anstelle von $\frac{1}{n}$ bei der Stichprobenstreuung erklärt sich also aus der Forderung der Erwartungstreue.

Die Aussage für $\bar{X}$ folgt aus: $E_\vartheta[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} E_\vartheta[X_i] = a(\vartheta)$.

Der Beweis für S^2 ist eine Übungsaufgabe.

Erwartungstreue Schätzer können noch schlecht sein, wie das Beispiel $\hat{a} := X_1$ zeigt. []

Beispiel 2 (Taxiproblem).
In einer Stadt gibt es N Taxis, die die Nummern 1 bis N tragen. Zur Schätzung von N beobachtet man die Nummern von n Taxis. Dann eignet sich folgendes Modell:
Die n Beobachtungen werden ausgedrückt durch n unabhängige Zva $X_1,...,X_n$, wobei die Verteilung der X_i die Gleichverteilung $U(\{1,...,N\})$, $N\in\mathbb{N}$, ist. Dann ist

$$\hat{N} := \max(X_1,...,X_n) + \min(X_1,...,X_n) - 1$$

erwartungstreuer Schätzer für $\vartheta = N$. Der Beweis ist eine Übungsaufgabe.
Offensichtlich haben die Zva $N - X_i + 1$, $1\le i\le n$, (unter jedem $\vartheta = N$) die gleiche gemeinsame Verteilung wie die X_i. Damit hat dann auch $\min(X_1,...,X_n) - 1$ die gleiche Verteilung wie $N - \max(X_1,...,X_n)$ und damit den gleichen Erwartungswert (unter jedem $\vartheta = N$). $[]$

2 C. Konsistenz.
Der Begriff der Konsistenz spielt bei großen Stichprobenumfängen eine Rolle. Dazu betrachtet man eine Folge von statistischen Räumen $\{(S_n,\mathfrak{S}_n,\mathfrak{P}_n),\ n\in\bar{\mathbb{N}} := \mathbb{N}\cup\{\infty\}\}$ mit

$$S_n = \bigtimes_{i=1}^{n} S',\ \mathfrak{S}_n = \bigotimes_{i=1}^{n}\mathfrak{S}',\ \mathfrak{P}_n = \{P_\vartheta^n = \bigtimes_{i=1}^{n} P'_\vartheta;\ \vartheta\in\Theta\}\ .$$

Definition 4. Gegeben sei ein Parameter $\rho : \Theta \mapsto R$. Ist für $n\in\mathbb{N}$ eine Abbildung $\hat{\rho}_n : S_n \mapsto R$ gegeben, so heißt die Folge $(\hat{\rho}_n, n\in\mathbb{N})$ ein **asymptotischer Schätzer.**

Definition 5a. Ein asymptotischer Schätzer $(\hat{\rho}_n, n\in\mathbb{N})$ heißt **konsistent** für den reellen Parameter ρ, wenn $\hat{\rho}_n$ mb ist und $(\hat{\rho}_n)$ unter ϑ nach W. gegen $\rho(\vartheta)$ konvergiert, d.h.

$$P_\vartheta^n[\,|\hat{\rho}_n - \rho(\vartheta)| > \varepsilon\,] \to 0 \quad (n\to\infty)\ \ \forall\,\varepsilon>0,\ \forall\,\vartheta\in\Theta$$

Für manche Betrachtungen ist es günstig, wenn man sich ein n–mal durchgeführtes Experiment als ein unendlich oft durchgeführtes Experiment vorstellt, wobei man nur die ersten n Versuche auswertet. Mit $(S_\infty,\mathfrak{S}_\infty,\mathfrak{P}_\infty)$ hat man so einen gemeinsamen W–Raum für alle n. Zu $\hat{\rho}_n$ definiert man dann $\hat{\rho}_{(n)} : S_\infty \mapsto R$ gemäß $\hat{\rho}_{(n)}(x_1,x_2,...) := \hat{\rho}_n(x_1,...,x_n)$. Dann gilt also

$$P_\vartheta^\infty[\hat{\rho}_{(n)}\in\Gamma] = P_\vartheta^n[\hat{\rho}_n\in\Gamma]\ \ \forall\,\vartheta\in\Theta\ .$$

Definition 5b. $(\hat{\rho}_n,\ n\in\mathbb{N})$ heißt **stark konsistent** für den reellen Parameter ρ, wenn $\hat{\rho}_n$ mb ist und $(\hat{\rho}_{(n)})$ unter ϑ f.s. gegen $\rho(\vartheta)$ konvergiert, d.h.

$$P_\vartheta^\infty[\hat{\rho}_{(n)} \to \rho(\vartheta) \text{ für } n\to\infty] = 1 \quad \forall\,\vartheta\in\Theta.$$

Der Vorteil des Begriffs der Konvergenz nach W. ist, daß die zugrundeliegenden W–Maße von n abhängen können und man so auf dem einfacheren Raum S_n bleiben kann. Der Vorteil des Begriffs der f.s. Konvergenz ist ,daß er man auf einer Menge vom Maß eins mit der üblichen Konvergenz in $R \subset \mathbb{R}$ rechnen kann.

Beispiel 1. (Fortsetzung)

(i) $(\bar{X}_{(n)}, n\in\mathbb{N})$ ist stark konsistent für $a(\vartheta)$;

(ii) $(S^2_{(n)}, n\in\mathbb{N})$ ist stark konsistent für $\sigma^2(\vartheta)$.

Der Beweis von (i) folgt direkt aus dem starken Gesetz der großen Zahlen; der Beweis von (ii) ist eine Übungsaufgabe.

2 D. Der mittlere quadratische Fehler.

Als Gütemaß für einen Schätzer $\hat{\rho}$ zu einem reellen Parameter ρ läßt sich der mittlere quadratische Fehler heranziehen, also:

(*) $E_\vartheta[(\hat{\rho} - \rho(\vartheta))^2] = \mathrm{Var}_\vartheta[\hat{\rho}] + (\rho(\vartheta) - E_\vartheta[\hat{\rho}])^2$, $\vartheta\in\Theta$.

Während die Erwartungstreue von $\hat{\rho}$ dafür sorgt, daß die Verteilung von $\hat{\rho}$ richtig zentriert ist, wird hier gefordert, daß die Verteilung von $\hat{\rho}$ möglichst eng um den richtigen Parameter konzentriert ist und zwar simultan für alle $\vartheta\in\Theta$.
Die Minimierung des mittleren quadratischen Fehlers in einer Klasse von erwartungstreuen Schätzern entspricht gerade der Minimierung der Varianz. Die Situation wird im nächsten §3 vorliegen.

Ist der asymptotische Schätzer $(\hat{\rho}_n)$ **asymptotisch erwartungstreu**, d.h. :

$$\lim E_\vartheta[\hat{\rho}_n] = \rho(\vartheta) \quad \forall\, \vartheta \in \Theta,$$

und gilt: $\lim \mathrm{Var}_\vartheta[\hat{\rho}_n] = 0$, so folgt aus (*): $(\hat{\rho}_n)$ ist **L^2–konsistent**, d.h.

$$E_\vartheta[\{\hat{\rho}_n - \rho(\vartheta)\}^2] \to 0,$$

und damit die Konsistenz von $(\hat{\rho}_n)$ wegen

$$P_\vartheta[\,|\hat{\rho}_n - \rho(\vartheta)| > \varepsilon] \le \frac{1}{\varepsilon^2} E_\vartheta[\{\hat{\rho}_n - \rho(\vartheta)\}^2].$$

§3 Lineare Modelle.

Die linearen Modelle sind einerseits so flexibel, daß sie viele empirische Sachverhalte hinreichend genau wiedergeben, und anderseits so einfach, daß sie eine bequeme mathematische Behandlung gestatten. Es handelt sich i.a. nicht mehr um ein Einstichprobenproblem, und es ist auch keine Zerlegung in Einstichprobenprobleme möglich, da die Varianz als gemeinsamer Parameter auftritt.

Definition 1. $(S,\mathfrak{S},\mathfrak{P})$ heißt ein **lineares Modell**, wenn folgende Situation vorliegt:

$(S,\mathfrak{S}) = (\mathbb{R}^n,\mathfrak{B}_n)$, $X_i : S \mapsto \mathbb{R}$ ist die i–te Projektion, L_k sei ein k–dim. linearer Teilraum des $\mathbb{R}^n$. Die Verteilungsannahme $\mathfrak{P} = \{P_\vartheta,\vartheta\in\Theta\}$ sei derart, daß unter ϑ $X_1,...,X_n$ unabhängige [es reichte sogar unkorrelierte] Zva sind mit:

$E_\vartheta[X_i] =: a_i(\vartheta) =: a_i$, $a := a(\vartheta) := (a_1,...,a_n)^\top \in L_k$,
$Var_\vartheta[X_i] = \sigma^2(\vartheta) =: \sigma^2 > 0$, $1\le i\le n$.
Dabei sei L_k minimal gewählt, d.h. $\exists$ k lin. unabhängige Vektoren $a(\vartheta_i)$, $1\le i\le k$.

Die Eigenschaft, daß $Var_\vartheta[X_i]$ unabhängig von i ist, wird als **Homogenität** bezeichnet.

Wir schreiben $X := (X_1,...,X_n)^\top$ für die Gesamtstichprobe; dann ordnet $X : \mathbb{R}^n \mapsto \mathbb{R}^n$ gerade jedem $x \in \mathbb{R}^n$ den zugehörigen Spaltenvektor zu, und es gilt:

(3.1) $E_\vartheta[X] = a(\vartheta)$.

3.2 Bemerkung. In einem linearen Modell hat man also gerade die folgende Verteilungsannahme:

$P_\vartheta = \bigotimes_{i=1}^n P'_{i,\vartheta}$ für gewisse W–Maße $P'_{i,\vartheta}$ auf $S'= \mathbb{R}$,
$a(\vartheta) := (a_1,...,a_n)^\top \in L_k$ mit $a_i(\vartheta) := \int x' P'_{i,\vartheta}[dx']$,
$\sigma^2(\vartheta) := \int (x'-a_i)^2 P'_{i,\vartheta}[dx']$ unabhängig von i.

3.3 Bemerkung. Ein wichtiges Beispiel ist ein **lineares Modell mit Normalverteilungsannahme**. Dabei ist

$\mathfrak{P} = \{ P_\vartheta = \bigotimes_{i=1}^n N(a_i,\sigma^2) ; \vartheta=(a,\sigma^2) \in \Theta = L_k\times J\}$.

Man wählt $J=(0,\infty)$ im Falle unbekannter Varianz und $J=\{\sigma_0^2\}$ im Falle bekannter Varianz.
Man hat

$X_i = a_i + \sigma W_i$ mit $W_i = (X_i - a_i)/\sigma \sim N(0,1)$; $W_1,...,W_n$ sind unabhängig.

Die letztere Darstellung wird auch oft zur Definition eines linearen Modells gewählt, wobei die N(0,1)–Verteilung durch eine beliebige standardisierte Verteilung Q ersetzt werden, also:

(*) $\int x\, Q[dx] = 0$, $\int x^2\, Q[dx] = 1$ $\forall\, Q \in \mathfrak{Q}$,
oder auch durch eine Familie $\mathfrak{Q}$ von standardisierten Verteilungen. []

Definition 2. Gegeben sei ein lineares Modell $(S,\mathfrak{S},\mathfrak{P})$.

Dann heißt $\hat{a} : S \mapsto L_k$ **Kleinste–Quadrate–Schätzer (KQS)** für a, wenn gilt:

$$\|X-\hat{a}\|^2 = \min_{a\in L_k} \|X-a\|^2.$$

Der KQS paßt also den Schätzwert der Beobachtung x möglichst gut an im Sinne der Norm. Bekanntlich gilt:

3.4 Lemma. (a) $\hat{a}$ ist die Projektion von $\mathbb{R}^n$ auf L_k;

(b) $\hat{a}$ ist erwartungstreu, d.h. $E_\vartheta[\hat{a}] = a(\vartheta) \in L_k \ \forall \vartheta \in \Theta$.

Beweis: a) ist bekannt. b) Da $\hat{a}$ eine lineare Abbildung ist, folgt nämlich mit (a):

$E_\vartheta[\hat{a}] = E_\vartheta[\hat{a}\circ X] = \hat{a}(E_\vartheta[X]) = \hat{a}\circ a = a \in L_k$. []

Beispiel 1. Einstichprobenfall. Es sei $(S,\mathfrak{S},\mathfrak{P})$ ein lineares Modell mit $a_i = \mu \ \forall$ i.

Dann gilt: $a = \mu e$ und $L_k := \mathfrak{L}(e)$ mit $e=(1,...,1)^\top$, $k = \dim L_k = 1$.

Jeder Vektor $a\in L_k$ läßt sich darstellen in der Form $a = \mu \cdot e$ (für ein $\mu \in \mathbb{R}$) und somit auch $\hat{a}$ als

$$\hat{a} = \hat{\mu} \cdot e \text{ für ein } \hat{\mu} : S \mapsto \mathbb{R};$$

$\hat{\mu}$ ist ein Schätzer für den Parameter μ. Dabei ist $\hat{\mu}(x)$ die Minimumsstelle von $\mu \mapsto \sum_{i=1}^n (x_i-\mu)^2$. Die Lösung ergibt sich leicht als (vgl. Beispiel 2 unten):

$$\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^n x_i, \text{ d.h. } \hat{\mu} = \bar{X}. \ []$$

Wir interessieren uns in diesem §3 für lineare erwartungstreue Schätzer.

Definition 3. $\hat{\rho}$ heißt ein **linearer Schätzer** für einen reellen Parameter ρ, wenn gilt: $\hat{\rho} = q^\top \cdot X$ für einen Vektor $q\in\mathbb{R}^n$. Vektoren werden dabei stets als Spaltenvektoren aufgefaßt.

3.5 Bemerkung. Erwartungstreue lineare Schätzer existieren nur für gewisse lineare reelle Parameter ρ. Denn ist $\hat{\rho} = q^\top \cdot X$ ein linearer erwartungstreuer Schätzer für ρ, so folgt:

$\rho(\vartheta) = E_\vartheta[q^\top \cdot X] = q^\top \cdot E[X] = q^\top \cdot a$. ρ ist also notwendig eine lineare Funktion von a.

Nur solche linearen Parameter können wir also im Folgenden betrachten.

3.6 Satz von Gauß–Markoff. Gegeben sei ein lineares Modell $(S,\mathfrak{S},\mathfrak{P})$. Sei $\hat{a}$ der KQS für a und ρ ein Parameter der Gestalt $\rho(\vartheta) = c^\top \cdot a$ für ein $c\in\mathbb{R}^n$. Dann ist $\hat{\rho} := c^\top \cdot \hat{a}$ linearer erwartungstreuer Schätzer für ρ mit kleinster Varianz unter allen $\vartheta \in \Theta$ und als solcher eindeutig bestimmt. $\hat{\rho}$ heißt **Gauß–Markoff–Schätzer** (GMS).

Beweis.(i) Für $q\in\mathbb{R}^n$ sei $q_{(k)}$ $[=\hat{a}(q)]$ die Projektion von q auf L_k. Dann kann man o.E. $c = c_{(k)}$ annehmen wegen $\rho(\vartheta) = c_{(k)}^{\top}\cdot a$ und $\hat{\rho} := c_{(k)}^{\top}\cdot\hat{a}$.

(ii) $\hat{\rho}$ ist linear. Dies folgt aus: $\hat{\rho} = c_{(k)}^{\top}\cdot\hat{a} = c_{(k)}^{\top}\cdot X$ gemäß 3.4.

(iii) $\hat{\rho}$ ist erwartungstreu; denn es gilt nach 3.4: $E_\vartheta[\hat{\rho}] = c^{\top}\cdot E[\hat{a}] = c^{\top}\cdot a = \rho(\vartheta)$.

(iv) Ist $q^{\top}\cdot X$ irgendein linearer erwartungstreuer Schätzer für $\rho(\vartheta)$, so gilt: $q_{(k)} = c_{(k)}$.

Wie in 3.5 gilt nämlich: $\rho(\vartheta) = E_\vartheta[q^{\top}\cdot X] = q^{\top}\cdot a = q_{(k)}^{\top}\cdot a$ und $\rho(\vartheta) = c_{(k)}^{\top}\cdot a$ für k unabhängige $a = a(\vartheta_i) \in L_k$, $1\le i\le k$. Daraus folgt $(q_{(k)}-c_{(k)})^{\top}\cdot a(\vartheta_i) = 0$, $1\le i\le k$, und somit $(q_{(k)}-c_{(k)})^{\top}\cdot a = 0 \quad \forall\, a\in L_k$.Damit hat man: $q_{(k)} = c_{(k)}$.

(v) Für die Kovarianzmatrix von X gilt bekanntlich:

$\mathfrak{Cov}_\vartheta[X] = \left(E_\vartheta[(X_i-a_i)\cdot(X_j-a_j)],\ 1\le i,j\le n\right) = E_\vartheta[(X-a)\cdot(X-a)^{\top}]$, also hier $\mathfrak{Cov}_\vartheta[X] = \sigma^2\cdot I_n$, wobei die I_n die n–dim. Einheitsmatrix ist [vgl. W–th. I 14.21b, 16.3].

(vi) $\mathrm{Var}_\vartheta[q^{\top}\cdot X] = E_\vartheta[(q^{\top}\cdot X-q^{\top}\cdot a)^2] = E_\vartheta[(q^{\top}\cdot(X-a))^2] = E_\vartheta[q^{\top}\cdot(X-a)\cdot(X-a)^{\top}\cdot q]$

$= q^{\top}\cdot E_\vartheta[(X-a)\cdot(X-a)^{\top}]\cdot q = q^{\top}\cdot(\sigma^2\cdot I_n)\cdot q = \sigma^2\cdot\|q\|^2 = \sigma^2(\|q_{(k)}\|^2+\|q-q_{(k)}\|^2)$

$= \sigma^2(\|c_{(k)}\|^2+\|q-q_{(k)}\|^2)$ nach (iv).

$\mathrm{Var}_\vartheta[q^{\top}\cdot X]$ ist also genau dann minimal, wenn $q=q_{(k)}=c_{(k)}$ gilt.

Dann gilt nach (i) auch: $q^{\top}\cdot X = c_{(k)}^{\top}\cdot X = c^{\top}\cdot\hat{a}$. []

3.7 Satz. $(S,\mathfrak{S},\mathfrak{P})$ sei ein lineares Modell. Für k<n ist dann

$$\hat{\sigma}^2 := \frac{1}{n-k}\,\|X-\hat{a}\|^2$$ ein (nicht linearer) erwartungstreuer Schätzer für σ^2
und heißt **Residualschätzer**.

Der **Beweis** ist eine Übungsaufgabe.

Der Name Residualschätzer rührt daher, daß hier der zu â orthogonale Anteil der Gesamtstichprobe X zum Schätzen benutzt wird. Unter Annahmen an die dritten und vierten Momente kann man auch eine Optimalität von $\hat{\sigma}^2$ zeigen [vgl. Witting (1985), S. 455–456].

Zur Berechnung der Projektion â auf L_k betrachten wir folgende Situation für $(S,\mathfrak{S},\mathfrak{P})$:

(**3.8**) Es seien $b_1,...,b_k$ linear unabhängige (Spalten–) Vektoren in $\mathbb{R}^n$ mit $L_k := \mathfrak{L}(b_1,...,b_k)$. Dann gilt für $a(\vartheta) \in L_k$ eine Darstellung $a(\vartheta) = \sum_{i=1}^{k}\alpha_i(\vartheta)\cdot b_i$ mit einen Parameter $\alpha = (\alpha_1,...,\alpha_k)^{\top} \in \mathbb{R}^k$. Man hat also $L_k = \{B\alpha;\ \alpha\in\mathbb{R}^k\}$ mit $B = (b_1,...,b_k) \in \mathbb{R}^{n\times k}$. Für den KQS â gilt folglich eine Darstellung $\hat{a} = \sum_{i=1}^{n}\hat{\alpha}_i\cdot b_i$ für gewisse $\hat{\alpha}_i : S \mapsto \mathbb{R}$, die als Schätzer für die Parameter α_i aufgefaßt werden können, $1\le i\le k$.

Die Parameter α_i sind Beispiel für lineare Parameter der Form $c^\top \cdot a$.

3.9 Satz. In der Situation von (3.8) existiert $(B^\top B)^{-1}$ und es gilt:

(0) Die Parameter α_i sind lineare Parameter (der Form $c_i^\top \cdot a$).

(a) Die $\hat{\alpha}_i$, $1 \le i \le k$, sind durch die folgenden **Normalengleichungen** eindeutig bestimmt:

$$b_i^\top \cdot X = \sum_{j=1}^{k} \hat{\alpha}_j \, b_i^\top \cdot b_j \,, \quad i=1,...,k;$$

(b) $\hat{a} = B \cdot (B^\top \cdot B)^{-1} \cdot B^\top \cdot X = B\hat{\alpha}$ [$(B^\top \cdot B)^{-1} \cdot B^\top$ heißt **Pseudoinverse** von B];

(c) $\hat{\alpha}_i : S \mapsto \mathbb{R}$ ist GMS für den Parameter α_i, $1 \le i \le k$.

Beweis: Sei $\ell_i : L_k \mapsto \mathbb{R}$ durch $\ell_i(a) = \ell_i(\sum \alpha_j \, b_j) = \alpha_i$ definiert. Dann ist ℓ_i linear und gemäß 3.6 ist $\hat{\alpha}_i = \ell_i(\hat{a}) : S \mapsto \mathbb{R}$ GMS für α_i. Die Existenz von $(B^\top \cdot B)^{-1}$ folgt aus "$\alpha = 0 \Leftrightarrow B\alpha = 0 \Leftrightarrow B^\top \cdot B\alpha = 0$"; diese Beziehung ergibt sich aus "$B^\top \cdot B\alpha = 0 \Rightarrow \alpha^\top B^\top \cdot B\alpha = \|B\alpha\|^2 = 0$". Dann erhält man (0) auch aus $(B^\top B)^{-1} B^\top a = (B^\top B)^{-1} B^\top B\alpha = \alpha$. Der Rest ist eine Übungsaufgabe. []

Beispiel 1. (Fortsetzung)

Hier ist also $\bar{X}$ gemäß 3.9 GMS für $\mu = E[X_i]$ und

$\hat{\sigma}^2 = \frac{1}{n-1} \|X - \hat{a}\|^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = S^2$ nach 3.7 erwartungstreuer Schätzer für σ^2 in Übereinstimmung mit Beispiel 1 aus §2. []

Beispiel 2. Ein **k–Stichprobenproblem.** Für das vorliegende k–Stichprobenproblem schreiben wir wieder $X = (X_{11},...,X_{1n_1},...,X_{k1},...,X_{kn_k})^\top$ mit $n := \sum_{i=1}^{k} n_i$. Es ist $E_\vartheta[X_{ij}] = \mu_i$ unabhängig von j, also $a = (\mu_1,...,\mu_1, \ldots ,\mu_k,...,\mu_k)^\top \in \mathbb{R}^n$ mit $\mu_i \in \mathbb{R}$. Dabei heißt μ_i auch **Effekt** der i–ten Behandlung (Versuchsbedingung).

Die Vektoren $b_i := (0,...,0,1,...,1,0,...,0)^\top$ (mit jeweils n_i Einsen), $1 \le i \le k$, erfüllen die Bedingungen von (3.8) und es gilt: $a = \sum_{i=1}^{n} \mu_i \, b_i$ (also $\alpha = (\mu_1,...,\mu_k)^\top$), $\hat{a} = \sum_{i=1}^{n} \hat{\mu}_i \, b_i$.

Die Berechnung der $\hat{\mu}_i$ kann nach 3.9 über die folgenden Normalengleichungen erfolgen.

$$b_i^\top \cdot X = \sum_j \hat{\mu}_j \, b_i^\top \cdot b_j = \hat{\mu}_i \, b_i^\top \cdot b_i \quad 1 \le i \le k$$

$$\Leftrightarrow \quad \sum_{j=1}^{n_i} X_{ij} = \hat{\mu}_i \cdot n_i \quad 1 \le i \le k \quad \Leftrightarrow \quad \hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} =: \bar{X}_i \,, \; 1 \le i \le k.$$

Somit ist $\hat{\mu}_i = \bar{X}_i$ GMS für μ_i und $\hat{a} = (\bar{X}_1,...,\bar{X}_1, \ldots ,\bar{X}_k,...,\bar{X}_k)$ KQS für a.
Für den Residualschätzer ergibt sich:

$$\hat{\sigma}^2 := \frac{1}{n-k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 = \sum_{i=1}^{k} \frac{n_i - 1}{n-k} S_i^2 \quad \text{mit } S_i^2 := \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \,.$$

Man sagt hier, daß die Versuchsbedingungen durch einen **qualitativen Faktor** bestimmt sind. []

Beispiel 3. Lineare Regression in einer Regressionsvariablen s.

Die Versuchsbedingungen seien durch einen **quantitativen Faktor** bestimmt, für den die Werte $s_1,...,s_n \in \mathbb{R}$ gegeben sind. Dabei beschreibe s_i etwa eine Dosierung beim i–ten Versuch.

Hier wird als Ansatz eine lineare Abhängigkeit vom quantitativen Faktor vorausgesetzt:

$a_i = \mu + \nu \cdot s_i$, $\mu,\nu \in \mathbb{R}$, wobei o.E. die Werte s_i nicht alle gleich sein sollen.

Der Parameter ν heißt **Regressionskoeffizient**.

Der k–Stichprobenfall $(s_1,...,s_n) = (s^1,...,s^1,...,s^k,...,s^k)$ ist ein Spezialfall.

Es ist $L_k = \mathfrak{L}(e,s)$ und $a = \mu\, e + \nu\, s$ mit $e:=(1,...,1)^\top$ ($\hat{=}\ b_1$) und $s:=(s_1,...,s_n)^\top$ ($\hat{=}\ b_2$ und $\alpha \hat{=} (\mu,\nu)^\top$). Nach Voraussetzung sind e und s linear unabhängig.

Zur Bestimmung des KQS $\hat{a} = \hat{\mu} \cdot e + \hat{\nu} \cdot s$ sind die Normalengleichungen zu lösen:

$$e^\top \cdot X = \hat{\mu} \cdot e^\top \cdot e + \hat{\nu} \cdot e^\top \cdot s \quad \text{und} \quad s^\top \cdot X = \hat{\mu} \cdot s^\top \cdot e + \hat{\nu} \cdot s^\top \cdot s$$

$$\Leftrightarrow \quad \Sigma X_i = \hat{\mu} \cdot n + \hat{\nu} \cdot \Sigma s_i \quad \text{und} \quad \Sigma s_i X_i = \hat{\mu} \cdot \Sigma s_i + \hat{\nu} \cdot \Sigma s_i^2$$

$$\Leftrightarrow \quad \bar{X} = \hat{\mu} + \hat{\nu} \cdot \bar{s} \quad \text{und} \quad \overline{sX} = \hat{\mu} \cdot \bar{s} + \hat{\nu} \cdot \overline{s^2}$$

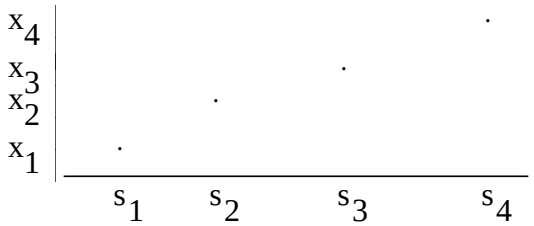
$$\text{mit } \bar{s} := \frac{1}{n}\Sigma s_i\,,\ \overline{s^2} := \frac{1}{n}\Sigma s_i^2,\ \overline{sX} := \frac{1}{n}\Sigma s_i X_i.$$

Es folgt: $\hat{\mu} = \bar{X} - \bar{s} \cdot \hat{\nu}$ und $\hat{\nu} = (\overline{sX} - \bar{s}\bar{X})/(\overline{s^2} - \bar{s}^2)$, wobei $\overline{s^2} - \bar{s}^2 > 0$ gilt

aufgrund der Schwarzschen Ungleichung. $\hat{\nu}$ heißt **empirischer Regressionskoeffizient**.

Gemäß 3.9 sind $\hat{\mu}$ und $\hat{\nu}$ GMS für μ und ν und $\hat{a} = (\hat{\mu}+\hat{\nu} \cdot s_1,...,\hat{\mu}+\hat{\nu} \cdot s_n)^\top$ KQS für a.

Graphische Interpretation: Für $\hat{\mu}$ und $\hat{\nu}$ gilt: $\Sigma(X_i - \hat{\mu} - \hat{\nu} \cdot s_i)^2 = \min_{\mu,\nu} \Sigma(X_i - \mu - \nu \cdot s_i)^2$. Man paßt also den Meßwerten x_1 (zu s_1),,..,x_n (zu s_n) eine Gerade $x = \hat{\mu} + \hat{\nu} \cdot y$ "möglichst gut " an. Diese Gerade heißt **Regressionsgerade**:



In einem historischen Versuch wurden 1903 die Abhängigkeit der Körpergröße s_i vom Vater und der Größe x_i des Sohnes für $n = 1078$ Vater–Sohn–Paare gemessen. Als Schätzwerte ergaben sich $\hat{\mu} = 85{,}67$, $\hat{\nu} = 0{,}516$.

Dabei ist die Geradengleichung $x = \hat{\mu} + \hat{\nu} \cdot y = y - (1-\hat{\nu})(y - \hat{\mu}/(1-\hat{\nu})) = y - 0{,}484(y-177)$.

Die Tatsache "$\hat{\nu}<1$" führt zu $-(1-\hat{\nu})<0$. Dies wurde als "Rückschritt" oder "Rückkehr zum Durchschnitt" interpretiert und führte zum Namen "Regression". Einem Wert $y>177$ etwa wird nämlich ein Wert $x<y$ zugeordnet und entsprechend für $y<177$.

Für den Residualschätzer ergibt sich hier:

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum(X_i - \hat{\mu} - \hat{\nu}\cdot s_i)^2 = \frac{1}{n-2}\sum[X_i - \bar{X} - \hat{\nu}(s_i - \bar{s})]^2.$$

Beispiel 4. Zwei qualitative Faktoren.

Die Versuchsbedingungen seien bestimmt durch zwei Faktoren A und B, die jeweils über m bzw. r Stufen variiert werden können. Es werden $n_{ij} \in \mathbb{N}_0$ Beobachtungen gemacht zur Stufenkombination (i,j). Die Stichprobe ist also beschreibbar durch Zva

$$X_{ij\ell} \quad \ell = 1,...,n_{ij},\ i = 1,...,m,\ j = 1,...,r.$$

Die Matrix $\left(n_{ij},\ 1\le i\le m,\ 1\le j\le r\right)$ heißt **Versuchsplan**. Bei $n_{ij}>0 \ \forall$ (i,j) heißt der Versuchsplan **vollständig**. Ist $n_{ij} = K \in \mathbb{N}$ unabhängig von (i,j), so heißt der Versuchsplan **balanciert.** Sei

$$E_\vartheta[X_{ij\ell}] = \mu_{ij} = \mu_{ij}(\vartheta),\ \mathrm{Var}_\vartheta[X_{ij\ell}] = \sigma^2 = \sigma^2(\vartheta) \text{ unabhängig von (i,j).}$$

Dann heißt μ_{ij} auch der (unbekannte) **Effekt** der Stufenkombination (i,j).

Ohne weitere Annahmen liegt ein k–Stichprobenproblem vor mit $k = m\cdot r$. Man hat also hier den Parameter $\alpha = \left(\mu_{ij}\right)$ und entsprechend $\hat{\alpha} = \left(\hat{\mu}_{ij}\right)$. Es ergeben sich aber neue Fragestellungen wie etwa die Frage nach den Einflüssen der einzelnen Faktoren und ihrer Wechselwirkung. Als GMS hat man nach Beispiel 2:

$$\hat{\mu}_{ij} = \frac{1}{n_{ij}}\sum_\ell X_{ij\ell}\,.$$

[Dabei existiert eine lineare Abbildung $\ell_{ij} : L_k \mapsto \mathbb{R}$ mit $\ell_{ij}(a) = \mu_{ij}$, $\ell_{ij}(\hat{a}) = \hat{\mu}_{ij}$.]

Man interessiert sich nun für die folgenden Parameter:

$\mu_{ij} =: \mu + \alpha_i + \beta_j + \gamma_{ij}$ mit

$\mu := \frac{1}{m\cdot r}\sum_{ij}\mu_{ij}$ allgemeiner (mittlerer) Effekt,

$\alpha_i := \frac{1}{r}\sum_j \mu_{ij} - \mu$ Haupteffekt des Faktors A auf Stufe i,

$\beta_j := \frac{1}{m}\sum_i \mu_{ij} - \mu$ Haupteffekt des Faktors B auf Stufe j,

$\gamma_{ij} := (\mu_{ij} - \mu) - \alpha_i - \beta_j$ Wechselwirkungseffekt zwischen A und B bei der Stufenkombination (i,j).

Dabei sind die Größen jedoch als grobe Näherung von dem zu sehen, was ihre Namen ausdrücken. Bei einer entsprechenden Reparametrisierung hat man $1 + m + r + m\cdot r$ Parameter; diese müssen jedoch die folgenden **Reparametrisierungsnebenbedingungen** erfüllen:

$\sum\alpha_i = 0$, $\quad\quad \sum\beta_j = 0$,

(Zentrierung, in der Summe heben sich die Haupteffekte auf)

$\sum_i \gamma_{ij} = 0 \ \forall j$, $\quad\quad \sum_j \gamma_{ij} = 0 \ \forall i$.

Man spricht von einem **Modell ohne Wechselwirkung**, falls $\gamma_{ij} = 0 \ \forall$ i,j.

Nach dem Satz 3.6 von Gauß–Markoff ergeben sich als GMS:

$$\hat{\mu} := \frac{1}{m \cdot r} \sum_{ij} \hat{\mu}_{ij} \ ,$$
$$\hat{\alpha}_i := \frac{1}{r} \sum_j \hat{\mu}_{ij} - \hat{\mu} \ ,$$
$$\hat{\beta}_j := \frac{1}{m} \sum_i \hat{\mu}_{ij} - \hat{\mu} \ ,$$
$$\hat{\gamma}_{ij} := (\hat{\mu}_{ij} - \hat{\mu}) - \hat{\alpha}_i - \hat{\beta}_j \ .$$

[Betrachten wir etwa den Parameter μ, so existiert offenbar eine lineare Abbildung $f : \mathbb{R}^{m \times r} \mapsto \mathbb{R}$ mit $f((\mu_{ij})) = \mu = f((\ell_{ij}(a))$. Gemäß 3.6 ist dann $f((\ell_{ij}(\hat{a})) = f((\hat{\mu}_{ij})) = \hat{\mu}$ GMS für μ.]

Bei der Bildung von $\hat{\mu}$ haben die $\hat{\mu}_{ij}$ alle gleiches Gewicht; es wird nicht berücksichtigt, auf welchem Stichprobenumfang n_{ij} die Schätzung $\hat{\mu}_{ij}$ beruht. In der Literatur wird deshalb auch der Schätzer $\frac{1}{n} \cdot \sum_{i,j,\ell} X_{ij\ell}$ gewählt. Dieser ist aber ein GMS für den Parameter $\frac{1}{m \cdot r} \sum_{ij} n_{ij} \cdot \mu_{ij}$, welcher nun besser schätzbar ist, jedoch von dem Versuchsplan abhängt. Die letztere Eigenschaft ist nicht sehr plausibel. [ ]

3.10 Bemerkung. Bei der Parameterdarstellung $a = \sum_{i=1}^{k} \alpha_i \cdot b_k = B\alpha$ spricht man von **linearen Regression** mit **Designmatrix** B.

Von einem **allgemeinen Regressionsmodell** spricht man, wenn der Parameter a eine k–dim. Fläche F_k im $\mathbb{R}^n$ durchläuft, also etwa $a=a(\alpha_1,...,\alpha_k)$ gilt mit $\alpha=(\alpha_1,...,\alpha_k) \in \mathbb{R}^k$.

Dann führt die Bestimmung des KQS â auf das Optimierungsproblem

$$\|X - \hat{a}\|^2 = \min_{\alpha \in \mathbb{R}^k} \|X - a(\alpha)\|^2.$$

Ist $a(\cdot)$ glatt, so kann man wieder die Normalengleichungen aufstellen, die gerade das Verschwinden der partiellen Ableitungen $\frac{\partial}{\partial \alpha_i}$ fordern. [ ]

Ein Beispiel für die Situation in 3.10 ist das folgende:

Beispiel 5. Schätzung von Zinskurven.

Der **Nullkuponpreis** $\beta(0,t)$ ist der Wert z.Zt. 0 von einer Geldeinheit, die man z.Zt. t erhält. [D.h. $\beta(0,t)$ ist der Diskontierungsfaktor.] Die **Zinsstruktur** ist gegeben durch die **Zinskurve** $t \mapsto \beta(0,t)$. Ein **Zahlungsstrom** $(z_1,z_2,...)$ beschreibt eine Folge von Auszahlungen z_t in t. Der Wert eines Vertrages, der einen Zahlungsstrom $(z_1,z_2,...)$ garantiert, ist dann

$$\sum_t z_t \cdot \beta(0,t) \ .$$

Der Wert eines Vertrages mit Zahlungsstrom (0,...,0,1,0,...) (1 an Stelle t) ist also gerade $\beta(0,t)$. Solche Verträge existieren aber selten, sondern man beobachtet Preise/Werte X_i, zu denen Verträge mit Zahlungsströmen $(z_1^i,z_2^i,...)$ gehandelt werden. Dabei reichen die Preise X_i i.a. nicht aus, um die Zinskurve zu berechnen. Die Gründe können in der numerischen Instabilität und in dem Mangel an Daten liegen. Ein weiterer wichtiger Grund ist, daß man Abweichungen W_i des

tatsächlichen Preises X_i von dem theoretischen Preis beobachtet gemäß:

$$X_i = \sum_t z_t^i \cdot \beta(0,t) + W_i.$$

Man macht nun den **statistischen Ansatz**, daß diese Abweichungen W_i durch unabhängige Zva mit Erwartungswert 0 und Varianz σ^2 modelliert werden. Dann sind die X_i unabhängige Zva mit Erwartungswert $a_i = \sum_t z_t^i \cdot \beta(0,t)$ und Varianz $\sigma^2>0$. Die Zahlungsströme $(z_1^i, z_2^i, ...)$ entsprechen dabei gerade den Versuchsbedingungen.

Dazu macht man einen allgemeinen Regressionsansatz:

$$a = a(r,\alpha) = \begin{pmatrix} a_1(r,\alpha) \\ a_2(r,\alpha) \\ . \end{pmatrix} \text{ mit einem k+1–dim.Parameter } (r,\alpha)\text{ , } r > 0,\ \alpha = (\alpha_1,...,\alpha_k)^\top \in \mathbb{R}^k.$$

Dann berechnet sich der KQS aus dem Optimierungsproblem:

$$\|X - \hat{a}\|^2 = \min_{r>0} \min_{\alpha\in\mathbb{R}^k} \|X - a(r,\alpha)\|^2.$$

Ein spezieller Ansatz von Vasicek & Fong, J. Finance (1982) ist nun der folgende, wobei die Minimierung über $\alpha\in\mathbb{R}^k$ mit den oben entwickelten Methoden durchführen kann:

$$\beta(0,t) := \sum_{j=1}^k \alpha_j e^{-j\cdot r\cdot t} = \sum_{j=1}^k \alpha_j (e^{-r\cdot t})^j.$$

Wegen $\beta(0,t) = e^{-rt}[\alpha_1 + o(1)]$ (für $t\to\infty$) kann der Parameter r als **langfristige Zinsrate** interpretiert werden. Es folgt:

$$a_i(r,\alpha) = \sum_t z_t^i \sum_{j=1}^k \alpha_j e^{-j\cdot r\cdot t} = \sum_{j=1}^k \alpha_j \sum_t z_t^i e^{-j\cdot r\cdot t} =: \sum_{j=1}^k \alpha_j b_{ij}(r)$$

mit $b_j(r) = \begin{pmatrix} b_{1j}(r) \\ b_{2j}(r) \\ .. \end{pmatrix}$. Mit $B(r) := \begin{pmatrix} b_1(r) & b_k(r) \end{pmatrix}$ und $a = a(r,\alpha) = B(r)\alpha$ führt die Bestimmung des KQS â auf das Optimierungsproblem:

$$\min_{r>0} \min_{\alpha\in\mathbb{R}^k} \|X - a(r,\alpha)\|^2 = \min_r \min_\alpha \|X - B(r)\alpha\|^2.$$

Nach den Ergebnissen 3.9 zur linearen Regression hat man:

$$\min_r \min_\alpha \|X - B(r)\alpha\|^2 = \min_r \|X - B(r)(B(r)^\top B(r))^{-1} B(r) X\|^2,$$

wenn $\{b_1(r),...,b_k(r)\}$ linear unabhängig sind. Dabei gilt für den Spaltenvektor $b_j(r)$:

$$b_j(r) = \left(\sum_t z_t^i e^{-j\cdot r\cdot t}\right) = \sum_t e^{-j\cdot r\cdot t} \begin{pmatrix} z_t^i \end{pmatrix} . [\,]$$

Kapitel 2. Testtheorie.

§4 Grundbegriffe der Testtheorie.

Gegeben sei ein statistischer Raum $(S,\mathfrak{S},\mathfrak{P})$ mit $\mathfrak{P} = \{P_\vartheta; \vartheta\in\Theta\}$.

Definition 1. Ein **Testproblem** ist gegeben durch die disjunkte Aufspaltung des Parameterraums $\Theta = H \cup K$ in die **Hypothese** H und die **Alternative** K.

Definition 2. Es sei $\rho : \Theta \mapsto R$ ein reeller Parameter und $r_o \in R$ fest.
(a) Gilt $H = \{\vartheta\in\Theta;\ \rho(\vartheta) \le r_o\}$, so spricht man von einem **einseitigen Testproblem** und schreibt dafür: $H: \rho(\vartheta) \le r_o$, $K: \rho(\vartheta) > r_o$.
(b) Gilt $H = \{\vartheta\in\Theta;\ \rho(\vartheta) = r_o\}$, so spricht man von einem **zweiseitigen Testproblem** und schreibt dafür: $H: \rho(\vartheta) = r_o$, $K: \rho(\vartheta) \ne r_o$.

Definition 3. Ein **deterministischer Test** φ ist eine mb Abbildung $\varphi : S \mapsto \{0,1\}$; dabei ist
$\{\varphi=0\}$ der **Annahmebereich** (von H) und
$\{\varphi=1\}$ der **Ablehnungsbereich** (von H $\hat{=}$ Annahmebereich von K).

Zunächst hatte man nur die Hypothese bei einem Testproblem formuliert; deswegen beziehen sich viele Begriffe wie in Definition 3 auf die Hypothese. Die Notwendigkeit, auch eine Alternative festzulegen, sah man erst später.

Definition 4. Eine (Test–) **Statistik** oder **Prüfgröße** T ist eine mb Abbildung $T : (S,\mathfrak{S}) \mapsto (D,\mathfrak{D})$.

Eine **typische Situation** ist die folgende:

$$\varphi = \begin{cases} 1 & T > c \\ \text{für} & \\ 0 & T \le c \end{cases}$$, wobei c **kritischer Wert** heißt.

Bei der Anwendung eines Tests unterscheidet man zwei Arten von Fehlern, die eine unterschiedliche Bedeutung haben.

Definition 5. (a) Ein Fehler 1. Art ist eine Entscheidung gegen H (also für K), obgleich H richtig ist; ein Fehler 2. Art ist eine Entscheidung für H, obgleich H nicht richtig ist.
(b) Die Abbildung $\beta_\varphi = \beta : \Theta \mapsto [0,1]$ mit $\beta(\vartheta) = P_\vartheta[\varphi=1] = E_\vartheta[\varphi]$ heißt **Gütefunktion** von φ. Man nennt $\beta(\vartheta)$, $\vartheta\in H$, **Fehlerwahrscheinlichkeiten** (Fw) **1. Art** und $1-\beta(\vartheta)$, $\vartheta\in K$, **Fw. 2. Art**.

Der Minimierung der Fw. 2. Art entspricht die Maximierung von $\varphi \mapsto \beta_\varphi(\vartheta)$, $\vartheta\in K$. Es gibt einen Test, der die Fw 1. Art minimiert, nämlich $\varphi \equiv 0$; und es gibt einen Test, der die Fw 2. Art minimiert, nämlich $\varphi \equiv 1$. Die Minimierung der Fw 1. und 2. Art widersprechen sich also. Einen Ausweg aus diesem Dilemma ergibt sich auf Grund der Tatsache, daß die beiden Fehler ein unterschiedliches Gewicht haben.

I.a. wird das Testproblem so formuliert, daß H einer etablierten Theorie und K neuen Effekten entsprechen. Beispielsweise wählt man H beim Testen eines alten gegen ein neues Medikament als den Parameterbereich, bei dem das alte nicht schlechter als das neue ist. Der Fehler 1. Art ist gemäß der Wahl von H und K der schwerwiegendere; man nimmt also eine konservative Haltung ein. Man beschränkt sich dann auf Tests, deren Fw 1. Art ein **Niveau** $\alpha \in (0,1)$ (die sogenannte **Irrtumswahrscheinlichkeit**) nicht überschreiten. Solche Tests heißen **Tests zum Niveau** α. Man gibt sich also ein $\alpha \in (0,1)$ vor und fordert von φ: $\beta_\varphi(\vartheta) \le \alpha \ \forall\, \vartheta \in H$. Üblich sind $\alpha=0{,}05$ oder $\alpha=0{,}01$. Unter dieser Nebenbedingung versucht man, die Fw 2. Art zu minimieren.

Beispiel 1. Einseitiges Testproblem bei einer Binomialverteilung.
Man betrachtet ein Einstichprobenproblem: Unter $\vartheta \in \Theta$ seien $X_1,...,X_n$ iid Zva mit $X_i \sim b(1,\vartheta)$. [D.h. $S = \bigotimes_{i=1}^n \{0,1\}$, $\mathfrak{P} = \{ P_\vartheta = \bigotimes_{i=1}^n b(1,\vartheta) \,.\, \vartheta \in \Theta = [0,1] \}$]. Dann interessiert das Testproblem

$$H : \vartheta \le \vartheta_o \,,\ K: \vartheta > \vartheta_o \quad \text{mit der Statistik} \quad T := \sum_{i=1}^n X_i.$$

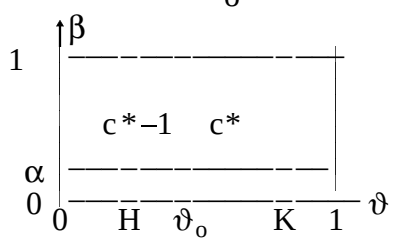
Dabei beschreibt T die Anzahl der Erfolge. Betrachten wir nun Tests der Form:

$$\varphi_c := \begin{cases} 1 & T>c \\ \text{für} & \\ 0 & T\le c \end{cases} \qquad \text{mit } c = -1,0,1,...,n.$$

Dann gilt:

- $T \sim b(n,\vartheta)$ unter ϑ,
- $\beta_c(\vartheta) := P_\vartheta[T>c] := \sum_{i=c+1}^n \binom{n}{i} \vartheta^i (1-\vartheta)^{n-i}$,
- $\beta_c \ge \beta_{c+1}$ auf Θ,
- $\frac{d}{d\vartheta} \beta_c(\vartheta) > 0$ auf $(0,1)$ für $0 \le c \le n-1$,
- φ_c ist Test z.N. $\alpha \ \Leftrightarrow\ \beta_c(\vartheta) \le \alpha \ \forall\, \vartheta \le \vartheta_o \ \Leftrightarrow\ \beta_c(\vartheta_o) \le \alpha$.
  $\Leftrightarrow c \ge c^*$ mit $c^* := \min \{c;\ \beta_c(\vartheta_o) \le \alpha\}$.

$c^* = \min \{ c;\ P_{\vartheta_o}[T>c] \le \alpha\} =: b(n,\vartheta_o)_\alpha$ heißt **α–Fraktil** von $b(n,\vartheta_o)$.



Anhand der Zeichnung sieht man, daß unter allen Tests φ_c , $c \ge c^*$, φ_{c^*} die kleinste Fw 2. Art hat, d.h. $1-\beta_{c^*}(\vartheta) \le 1-\beta_c(\vartheta) \ \forall\, \vartheta \in K, \forall\, c \ge c^*$. [ ]

Definition 6. Es sei Q ein W–Maß auf $(\mathbb{R},\mathfrak{B}_1)$ mit Vf F und $0<\alpha<1$. Dann ist

$F^-(s) := \inf \{t \in \mathbb{R};\ F(t) \ge s\}$ die (verallgemeinerte) Inverse von F, und man nennt

$Q_\alpha := \inf \{t \in \mathbb{R};\ Q[(t,\infty)] \le \alpha\} = F^-(1-\alpha)$ das **α–Fraktil** von Q.
Es sei $N_\alpha := N(0,1)_\alpha$ das α–Fraktil von N(0,1).

4.1 Lemma. Sei Q_α das α–Fraktil und F Vf von Q; dann gilt:

(i) $Q[(Q_\alpha,\infty)] \le \alpha \le Q[[Q_\alpha,\infty)]$, d.h. $F(Q_\alpha-0) \le 1-\alpha \le F(Q_\alpha)$.

(ii) $Q[(Q_\alpha,\infty)] = Q[[Q_\alpha,\infty)] = \alpha$, d.h. $F(Q_\alpha-0) = 1-\alpha = 1-F(Q_\alpha)$,
falls $P[\{Q_\alpha\}] = 0$, falls also F in Q_α stetig ist.

In der Regel gilt in dem obigen Beispiel: $\beta_{c^*}(\vartheta_o) < \alpha$, die vorgebene Irrtums–W. α wird also nicht voll ausgeschöpft. Als Ausweg haben sich die Mathematiker folgende Erweiterung des Testbegriffs einfallen lassen, durch den die Menge aller Tests eine konvexe Menge wird.

Definition 7.
Ein **(randomisierter) Test** ist eine mb Abbildung $\varphi : S \mapsto [0,1]$ verbunden mit der Vorschrift: entscheide dich für K bzw. H mit der W. $\varphi(x)$ bzw. $1-\varphi(x)$.

Die Entscheidung mit Hilfe eines Tests φ umfaßt i.a. zwei Experimente:

(i) das ursprüngliche Experiment der Stichprobenentnahme, die durch den statistischen Raum $(S,\mathfrak{S},\mathfrak{P})$ beschrieben wird,

(ii) eine Randomisierung der Entscheidung bei Vorliegen der Stichprobe x, d.h. das Ziehen einer Zufallszahl aus dem Einheitsintervall. Dieses Experiment soll durch eine U(0,1)–verteilte Zva Z beschrieben werden können.
Bei $Z \le \varphi(x)$ entscheidet man sich für K, bei $Z > \varphi(x)$ für H.

(Ist $\varphi(x) \in \{1/6,...,5/6\}$, so kann das Ziehen einer Zufallszahl offenbar auch durch Würfeln ersetzt werden.)

Zur Beschreibung dieses zweistufigen Experiments in einem mathematischen Modell sei $(S,\mathfrak{S},P)$ und φ gegeben. Dann definiert man

$\hat{S} := S \times (0,1]$, $\hat{\mathfrak{S}} := \mathfrak{S} \otimes (0,1] \cap \mathfrak{B}_1$, $\hat{P}_\vartheta := P_\vartheta \times U(0,1)$ als W–Maß auf $\hat{\mathfrak{S}}$.

Ferner seien $\hat{X} : \hat{S} \mapsto S$ und $\hat{Z} : \hat{S} \mapsto (0,1]$ die Projektionen auf S bzw. $(0,1]$; diese Zva beschreiben dann die Stichprobenentnahme und das Ziehen einer Zufallszahl.
Es gilt für die Entscheidung für K (nach Def. des Produktmaßes):

$$\hat{P}_\vartheta[\hat{Z} \le \varphi\circ\hat{X}] = P_\vartheta \times U(0,1)[\{(x,z);\, z \le \varphi(x)\}] = \int U(0,1)[(0,\varphi(x)]\ ]\ P_\vartheta[dx]$$
$$= \int \varphi\, dP_\vartheta = E_\vartheta[\varphi] =: \beta(\vartheta).$$

Somit beschreibt $\beta\,|_H$ die Fw 1. Art und $(1-\beta)\,|_K$ die Fw 2. Art.
Man kann auch ein W–Modell konstruieren, in dem $\varphi(x)$ die bedingte W. für eine Entscheidung für K bei gegebener Stichprobe x ist. Dann folgt wieder sofort, daß $\beta(\vartheta) = \int \varphi(x)\, dP_\vartheta$ die absolute W. für eine Entscheidung für K ist, die vor Entnahme der Stichprobe x berechnet wird.

Definition 8. Sei $0<\alpha<1$.

(a) $\varphi \in \Phi :\Leftrightarrow \varphi$ ist Test (im Sinne von Def. 7).

(b) $\beta : \Theta \mapsto [0,1]$ mit $\beta(\vartheta) = E_\vartheta[\varphi]$ heißt **Gütefunktion** $\beta = \beta_\varphi$ von $\varphi \in \Phi$.

(c) $\varphi \in \Phi_\alpha \quad :\Leftrightarrow \varphi \in \Phi, \beta_\varphi(\vartheta) \le \alpha \ \forall \vartheta \in H$

$\Leftrightarrow$: φ ist **Test z.N. α**.

α heißt **(Signifikanz–) Niveau**.

Wie bereits erwähnt, wird bei den Fehlern 1. und 2. Art ein asymmetrische Betrachtungsweise angewandt. Es werden nur Tests zugelassen, deren Fw 1. Art ein Niveau α nicht überschreiten. Unter dieser Nebenbedingung wird ein Test gesucht, der die Fw 2. Art minimiert, dessen Gütefunktion auf K also maximal ist oder zumindest überhalb einer gewissen Schranke liegt.

Definition 9. Seien $\varphi,\varphi^* \in \Phi_\alpha$, $\mathfrak{M} \subset \Phi_\alpha$.

(a) φ^* heißt **schärfster Test in** $\mathfrak{M}$, wenn gilt:

$$\varphi^* \in \mathfrak{M} \quad \text{und} \quad \beta_{\varphi^*}(\vartheta) = \max_{\varphi \in \mathfrak{M}} \beta_\varphi(\vartheta), \ \forall \vartheta \in K.$$

(b) φ heißt **unverfälschter** Test z.N. α, wenn gilt:

$$\varphi \in \Phi_\alpha \ \text{und} \ \beta_\varphi(\vartheta) \ge \alpha \ \forall \vartheta \in K \ \Leftrightarrow: \ \varphi \in \tilde{\Phi}_\alpha.$$

Bei einem unverfälschten Test wird also die Forderung erhoben, daß die Fw 2. Art den i. a. recht großen Wert $1-\alpha$ nicht überschreiten. Dann wird die Entscheidung für K unter K mindestens so oft getroffen wie unter H. Im Beispiel 1 ist i.a. keiner der Tests ein unverfälschter Test z.N. α. Aus $\beta_\varphi(\vartheta_o) < \alpha$ folgt nämlich aus Stetigkeitsgründen $\beta_\varphi(\vartheta) < \alpha$ für gewisse $\vartheta > \vartheta_o$. Man sieht auch, daß aus Stetigkeitsgründen die untere Schranke α für $\beta\big|_K$ nicht größer gewählt werden kann.

(4.2) Für den Test $\varphi_\alpha :\equiv \alpha$ gilt: $\varphi_\alpha \in \tilde{\Phi}_\alpha$.

(4.3) Ist φ^* schärfster Test in $\mathfrak{M}$ und gilt: $\varphi_\alpha \equiv \alpha \in \mathfrak{M}$, so folgt: $\varphi^* \in \tilde{\Phi}_\alpha$.

(4.4) Φ ist eine konvexe Menge.

Die Forderung der Unverfälschtheit ist in Hinblick auf (4.2) noch eine schwache Forderung.

Beispiel 1. (Fortsetzung) **Übergang zu einem randomisierten Tests.**

Sei $c^* =: b(n,\vartheta_o)_\alpha$ wieder das α–Fraktil zu $b(n,\vartheta_o)$.

Wegen $\beta_{c^*-1}(\vartheta_o) > \alpha \ge \beta_{c^*}(\vartheta_o)$, existiert ein $\gamma \in [0,1)$ mit

$$\gamma \cdot \beta_{c^*-1}(\vartheta_o) + (1-\gamma) \cdot \beta_{c^*}(\vartheta_o) = \alpha = E_{\vartheta_o}[\gamma \cdot \varphi_{c^*-1} + (1-\gamma) \cdot \varphi_{c^*}] =: E_{\vartheta_o}[\varphi^*].$$

Dabei ist $\varphi^* = \begin{cases} 1 & T > c^* \\ \gamma & \text{für } T = c^* \\ 0 & T < c \end{cases}$.

Aus der Isotonie von β_{c^*-1} und β_{c^*} ergibt sich die Isotonie von $\beta^*(\vartheta) := E_\vartheta[\varphi^*]$.

Wegen $\beta^*(\vartheta_o) = \alpha$ ist φ^* sowohl ein Test z.N. α als auch unverfälscht, d.h.: $\varphi^* \in \tilde{\Phi}_\alpha$.
Die Entscheidung erfolgt also durch Vergleich des Wertes der Prüfgröße T(x) mit dem kritischen Wert c*.
Für T(x)>c* bzw. <c* erfolgt die Entscheidung für K bzw. H wie bei einem deterministischen Test. Ist jedoch T(x)=c*, so liefert das Experiment keine klare Entscheidung zugunsten von K oder H. Man führt dann, wie bereits erwähnt, ein Zusatzexperiment durch.
Der Wert γ bestimmt sich aus:

(4.5) $$P_{\vartheta_o}[T>c^*] + \gamma \cdot P_{\vartheta_o}[T=c^*] = \alpha \quad \text{mit} \quad \gamma=0 \Leftrightarrow P_{\vartheta_o}[T>c^*] = \alpha.$$

In §9 kann der folgende Satz gezeigt werden, nach dem φ^* optimal in der größtmöglichen Klasse Φ_α von zugelassenen Test ist.

4.6 Satz. Einseitiger Binomialtest. Unter $\vartheta \in \Theta = (0,1)$ seien $X_1,...,X_n$ iid mit $X_i \sim b(1,\vartheta)$. Ferner sei $0<\vartheta_o<1$, $T := \sum_{i=1}^n X_i$, $c^* := b(n,\vartheta_o)_\alpha$ und γ wie in (4.5). Dann ist für das einseitige Testproblem H: $\vartheta \le \vartheta_o$, K: $\vartheta > \vartheta_o$ folgender Test schärfster Test in Φ_α:

$$\varphi^* := \begin{cases} 1 & T > c^* \\ \gamma \text{ bei } & T = c^* \\ 0 & T < c^* \end{cases}.$$

Es soll nun ein **Zahlenbeispiel** betrachtet werden mit $n = 14$, $\vartheta_o = 0{,}25$, $\alpha=0{,}01$ und $T(x)=8$.
Dabei ist: $\beta_c(\vartheta_o) = P_{\vartheta_o}[T>c] = 1$ für c=–1, bzw. = 0,9822 (c=0), = 0,8922 (c=1), = 0,7189 (c=2), = 0,4787 (c=3), = 0,2585 (c=4), = 0,1117 (c=5), = 0,0383 (c=6), = 0,0103 (c=7), = 0,0027 (c=8), = 0,0003 (c=9), = 0,0000 (c=10) [vgl. Tafel für b(n,1/4)].
Demnach ist c* = 8; bei Benutzung des Test $\varphi_{c^*} = 1_{\{T>8\}}$ erfolgt die Entscheidung für H.
Für φ^* ergibt sich $\gamma = 0{,}963$ aus (4.5) und $P_\vartheta[T=c] = P_\vartheta[T>c-1] - P_\vartheta[T>c]$.
Benutzt man φ^*, so erfolgt also die Entscheidung für K bei T(x)=8 in 96,3 % aller Fälle.
Bei einer Entscheidung für K heißt die Antwort:

Es ist z.N. α statistisch gesichert, daß die Erfolgs–W. ϑ größer als 0,25 ist.

Für $\alpha = 0{,}0103$ wäre $\varphi^* = \varphi_7$ und somit deterministisch.
Für $\alpha = 0{,}001$ ist c*=9; bei T(x)=8 würde man sich für H entscheiden. (Hier ist sowohl die Schranke für die zugelassenen Fw als auch der Stichprobenumfang n klein.)
Bei einer Entscheidung für H heißt die Antwort:

Es ist z.N. $\alpha = 0{,}001$ nicht statistisch gesichert, daß die Erfolgs–W. $\vartheta > 0{,}25$ ist.

Die Antwort darf **nicht** heißen: "Es ist z.N. $\alpha = 0{,}001$ statistisch gesichert, daß $\vartheta \le 0{,}25$ ist". Die Fw 2. Art sind für $\vartheta \in K$ mit $\vartheta \approx \vartheta_o$, nämlich $\approx 1-\alpha$. Unter φ^* sind die Fw 2. Art $\ge 0{,}8$ für $\vartheta \in (0{,}25\ ,\ 0{,}5]$. Eine Entscheidung für H bedeutet nur, daß die Beobachtung im Rahmen des Niveaus α nicht im Widerspruch zu H steht.

Um die Abhängigkeit der Entscheidung von n zu studieren, machen wir eine asymptotische Betrachtung. Es sei $W_n := \sqrt{\frac{n}{\vartheta(1-\vartheta)}} \cdot (\frac{1}{n} T - \vartheta)$. Nach dem Zentralen Grenzwertsatz konvergiert unter ϑ die Verteilung von W_n gegen N(0,1) (im Sinne der schwachen Konvergenz). Da die Vf zu N(0,1) streng monoton und stetig ist, gilt für die zugehörigen inversen Vfn: $F_n^-(1-\alpha) \to F^-(1-\alpha) = N_\alpha$, $0<\alpha<1$ [vgl. W–th.I, 17.10]. Dies ist die Konvergenz der Fraktile. Man sieht leicht, daß für das Fraktil $F_n^-(1-\alpha)$ der Verteilung von W_n gilt:

$F_n^-(1-\alpha) = \sqrt{\frac{n}{\vartheta(1-\vartheta)}} \cdot (\frac{1}{n} b(n,\vartheta)_\alpha - \vartheta) \left[\to N_\alpha\right]$; eine eindeutige Entscheidung für K erfolgt bei:

$$T > c^* \Leftrightarrow \bar{X}_{(n)} > \frac{1}{n} b(n,\vartheta_o)_\alpha = \vartheta_o + N_\alpha \cdot \sqrt{\frac{1}{n} \cdot \vartheta_o(1-\vartheta_o)} + o(n^{-\frac{1}{2}}).$$

Mit wachsendem n braucht die mittlere Anzahl $\bar{X}$ der Erfolge also den Wert ϑ_o um einen immer kleineren Betrag zu überschreiten. []

Nach diesem Beispiel können nun zwei Sätze vorgestellt werden, die auf die Situation von Beispiel 2 in §1 anwendbar sind und die erst in §§ 9,10 bewiesen werden können. In beiden Sätzen ist die Verteilungsannahme dadurch charakterisiert, daß gilt:

Unter $\vartheta = a \in \Theta = \mathbb{R}$ sind $X_1,...,X_n$ iid mit $X_i \sim N(a,\sigma_o^2)$ für ein $\sigma_o^2>0$.

4.7 Satz. Einseitiger Gauß–Test. Sei $a_o \in \mathbb{R}$, , $0<\alpha<1$. Unter der obigen Verteilungsannahme verwendet man für das einseitige Testproblem $H_1: a \le a_0$, $K_1: a>a_o$ die Prüfgröße

$T := \sqrt{n} \frac{\bar{X} - a_o}{\sigma_o}$ und den Test $\varphi_1^* := \begin{cases} 1 & \text{für } T > N_\alpha \Leftrightarrow \bar{X} > a_0 + N_\alpha\sqrt{\sigma_o^2/n} \\ 0 & T \le N_\alpha \end{cases}$.

Es gilt: φ_1^* ist schärfster Test in Φ_α und damit unverfälscht.

Die Unverfälschtheit von φ_1^* kann auch direkt gezeigt werden mit Hilfe von:

4.8 Lemma. Es sei φ_1^* wie in 4.7. Dann gilt für die zu φ_1^* gehörige Güterfunktion:

$\beta_{\varphi_1^*}(a) = 1 - F(N_\alpha - \sqrt{n} \cdot \frac{a-a_o}{\sigma_o})$, wobei F die Vf zu N(0,1) ist.

Der **Beweis** ergibt sich aus: $T = \sqrt{n} \frac{\bar{X} - a}{\sigma_o} + \sqrt{n} \cdot \frac{a-a_o}{\sigma_o}$.

Mit $1-F(N_\alpha) = \alpha$ und der Isotonie von $\beta_{\varphi_1^*}(a)$ folgt dann $\varphi_1^* \in \tilde{\Phi}_\alpha$. []

Für den zweiseitigen Fall betrachten wir die Spiegelung $\pi : S \mapsto S$ bei a_o, also

(4.9) $\pi(x_1,...,x_n) = (2a_o - x_1,...,2a_o - x_n)$, $(x_1,...,x_n) \in \mathbb{R}^n$.

Dann gilt

(4.10) $T \circ \pi = -T$

(4.11a) $\beta_{\varphi\circ\pi}(a_o+\delta) = \beta_\varphi(a_o-\delta)$ für die Gütefunktion eines jeden Tests φ.

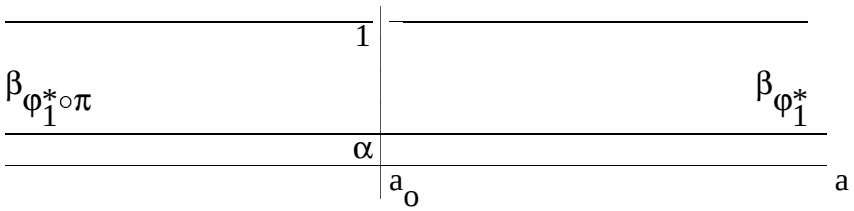
Dies ergibt sich aus:

(4.11b) $X_1,...,X_n$ iid mit $X_i \sim N(a_o+\delta) \Leftrightarrow 2a_o-X_1,...,2a_o-X_n$ iid mit $2a_o-X_i \sim N(a_o-\delta)$.

Es soll nun das zweiseitige Testproblem betrachtet werden:

$H_2: a = a_o$, $K_2: a \neq a_o$.

Dabei interessiert wieder die Existenz eines schärfsten Test φ_2^* in Φ_α. Wegen $\beta_{\varphi_1^*\circ\pi}(a_o) = \beta_{\varphi_1^*}(a_o) = \alpha$, sind φ_1^* und $\varphi_1^*\circ\pi$ auch Tests z.N. α für das zweiseitige Problem und somit Konkurrenten für φ_2^*. Somit muß die Gütefunktion von φ_2^* oberhalb von $\max(\beta_{\varphi_1^*},\beta_{\varphi_1^*\circ\pi})$ mit Gleichheit in a_o. Anschaulich betrachtet müßte sie dann einen Knick an der Stelle a_o haben. Andererseits ist sie aber leicht als differenzierbar nachzuweisen. In der Tat gilt:



4.12 Lemma. Es liege die obige Verteilungsannahme vor, und es sei $a_0 \in \mathbb{R}$, $0<\alpha<1$.

Dann existiert für das zweiseitige Testproblem H_2 gegen K_2 kein schärfster Test in Φ_α.

Der genaue **Beweis** ist eine Übungsaufgabe.

$\varphi_1^* = \mathbf{1}_{\{T>N_\alpha\}}$ berücksichtigt für die Ablehnung von H_2 nur große Werte von T, $\varphi_1^*\circ\pi = \mathbf{1}_{\{T<-N_\alpha\}}$ dagegen nur kleine. Zu große Werte von T sollten aber in gleicher Weise zur Ablehnung von H führen wie zu kleine. Um dieser Forderung Rechnung zu tragen, wollen wir uns hier bei der Suche nach einem schärfsten Test auf die Tests φ z.N. α mit $\varphi\circ\pi = \pi$ einschränken. Dann gilt der folgende

4.13 Satz. Zweiseitiger Gaußtest. Es liege die obige Verteilungsannahme vor, und es sei $a_0 \in \mathbb{R}$, $0<\alpha<1$. Für das zweiseitige Testproblem $H_2: a = a_o$, $K_2: a \neq a_o$ ist

$$\varphi_2^* := \begin{cases} 1 & \text{für } |T| > N_{\alpha/2} \\ 0 & |T| \leq N_{\alpha/2} \end{cases}$$ schärfster Test in $\mathfrak{M} := \{\varphi \in \Phi_\alpha;\ \varphi\circ\pi = \varphi\}$ und unverfälscht.

Der Beweis wird später geführt. Die Unverfälschtheit folgt hier aus (4.3). Daß $\varphi_2^* \in \tilde{\Phi}_\alpha$ gilt, kann auch elementar gezeigt werden. Wegen (4.10) gilt offenbar $\varphi_2^*\circ\pi = \varphi_2^*$.

Zum zweiseitigen Testproblem soll ein **Zahlenbeispiel** betrachtet werden:

Sei $a_o = 500$, $\sigma_o = 5{,}2$, $n = 7$, und $\alpha = 0{,}01$. Die Stichprobe sei:

x = (514 , 497 , 507 , 499 , 510 , 501 , 508).

Es ergeben sich folgende Werte: $\bar{X}(x) = 505{,}14$, $T(x) = \sqrt{7}\cdot\frac{5{,}14}{5{,}2} = 2{,}615$, $N_{\alpha/2} = N_{0{,}005} = 2{,}576$. Somit ist $\varphi_2^*(x) = 1$. Die Abweichung des Mittelwerts a von 500 ist **signifikant z.N. α** = 0,01.

§5 Verteilungen einiger wichtiger Statistiken.

Es soll an die folgenden Fakten über die Berechnnung von Lebesgue–Dichten erinnert werden:

5.1 Satz. (a) Sei $A \in \mathbb{R}^{n\times n}$ invertierbar, $b \in \mathbb{R}^n$, $T(x) := Ax+b$, $x\in\mathbb{R}^n$. Dann gilt:

$$\lambda^n T^{-1} = |\det A|^{-1} \cdot \lambda^n,$$

(b) Sind X und Y unabhängige Zva auf $(\Omega,\mathfrak{F},P)$ mit Werten in $(\mathbb{R}^n,\mathfrak{B}_n)$ bzw. $(E,\mathfrak{E})$, hat X λ^n–Dichte f, sind $A : E \mapsto \mathbb{R}^{n\times n}$, $b : E \mapsto \mathbb{R}^n$ mb, so daß $A(Y)^{-1}$ f.s. existiert, so hat $A(Y)X+b(Y)$ die λ^n–Dichte: $z \mapsto \int |\det A(y)|^{-1} \cdot f(A(y)^{-1}(z-b(y)))\, P^{Y^{-1}}[dy]$.

(c) Sind $X_1,...,X_n$ unabhängige Zva mit $X_i \sim N(a_i,\sigma^2)$, $1\le i\le n$ und ist A orthogonale Matrix, so gilt für $\begin{pmatrix}Y_1\\ \vdots\\ Y_n\end{pmatrix} := A \begin{pmatrix}X_1\\ \vdots\\ X_n\end{pmatrix}$, $\begin{pmatrix}b_1\\ \vdots\\ b_n\end{pmatrix} := A \begin{pmatrix}a_1\\ \vdots\\ a_n\end{pmatrix}$:
$Y_1,...,Y_n$ sind unabhängige Zva mit $Y_i \sim N(b_i,\sigma^2)$, $1\le i\le n$.

Beweis: a) Barner/Flohr (1983, S. 312); b) W–th. I 13.8; c) Hier kann man Teil (b) anwenden mit $f(A^{-1}x) = (2\pi\sigma^2)^{-n/2} \cdot \exp\{-\frac{1}{2\sigma^2}\|A^{-1}(x - Aa)\|^2\}$ (vgl. auch W–th. I 16.6k).

5.2. Spezialfälle von 5.1c sind Y=const., d.h. $|E|=1$, sowie
$A(Y)X+b(Y) = X\pm Y$ bzw. für n=1 $X \cdot Y$ bzw. $\mathbf{1}_{\{Y\neq 0\}} \cdot X/Y$.

5 A. Die χ^2–Verteilung.

Die χ^2–Verteilung ist gemäß $\chi^2_n := \Gamma_{1/2,n/2}$ definiert; $\chi^2_{n,\alpha}$ sei das α–Fraktil zu χ^2_n.

5.3 Satz. (a) Sind $X_1,...,X_n$ unabhängige Zva mit $X_i \sim N(0,1)$, $1\le i\le n$, so gilt: $\sum_{i=1}^n X_i^2 \sim \chi^2_n$;

(b) $\chi^2_m \star \chi^2_n = \chi^2_{m+n}$ für $m,n\in\mathbb{N}$;

(c) χ^2_n hat die Lebesgue–Dichte $h_n(t) := \text{const} \cdot t^{\frac{1}{2}n-1} \cdot \exp\{-\frac{1}{2}t\}$, $t>0$.

Beweis: W–th. I 12.17, 13.13.

Definition 1. Die **nichtzentrale χ^2_n–Verteilung** $\chi^2_n(a^2)$ mit **Nichtzentralitätsparameter** a^2 und n Freiheitsgraden wird folgendermaßen definiert:

$\chi^2_1(a^2)$ ist die Verteilung von X^2, falls $X \sim N(a,1)$, und
$\chi^2_n(a^2) := \chi^2_1(a^2) \star \chi^2_{n-1}$, $n\ge 2$.

Daß die Verteilung von X^2 nur von a^2 abhängt, folgt aus der Symmetrie von N(0,1): Es ist $W := X-a \sim N(0,1)$ und somit: $X^2 = (W+a)^2 \sim (-W+a)^2 = (W-a)^2$. Ferner gilt: $\chi^2_n(0) = \chi^2_n$.

5.4 Satz. (a) $\chi_1^2(a^2)$ hat die Lebesgue–Dichte

$h_{1,a^2}(t) := \frac{1}{2}\left[f(a-\sqrt{t}) + f(a+\sqrt{t})\right] \cdot \mathbf{1}_{(0,\infty)}(t) / \sqrt{t}$, wobei f die Dichte zu N(0,1) ist.

(b) Sind $X_1,...,X_n$ unabhängige Zva mit $X_i \sim N(a_i,1)$, so gilt:

$\sum_{i=1}^n X_i^2 \sim \chi_n^2(a^2)$ mit $a^2 = \sum_1^n a_i^2$.

Beweis. a) Sei W~N(0,1) und F die Vf zu N(0,1). Dann gilt:

$P\left[(W+a)^2 \le x\right] = P\left[\,|W+a| \le \sqrt{x}\right] = P\left[W+a \le \sqrt{x}\right] - P\left[W+a < -\sqrt{x}\right] = F(\sqrt{x}-a) - F(-\sqrt{x}-a) =$

$\int_0^x \frac{d}{dt}\left\{F(\sqrt{t}-a) - F(-\sqrt{t}-a)\right\} dt$. Nun ist $\frac{d}{dt}\{...\} = \left[f(\sqrt{t}-a) + f(-\sqrt{t}-a)\right]/2\sqrt{t}$.

b) Wähle eine orthogonale Matrix A, deren erste Zeile gleich $\frac{1}{a}(a_1,...,a_n)$ ist, und setze:

$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} := A \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$. Dann ergibt sich $E\left[\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}\right] = A \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}$ und $\sum X_i^2 = \sum Y_i^2$.

Gemäß 5.1c sind $Y_1,...,Y_n$ unabhängige Zva mit $Y_1 \sim N(a,1)$, $Y_2,...,Y_n \sim N(0,1)$.

Dann sind nach 5.3a Y_1^2 und $\sum_2^n Y_i^2$ unabhängige Zva mit $Y_1^2 \sim \chi_1^2(a^2)$ und $\sum_2^n Y_i^2 \sim \chi_{n-1}^2$ und die Aussage folgt direkt aus der Definition von $\chi_n^2(a^2)$. []

Es folgt jetzt eine Anwendung auf das Stichprobenmittel $\bar{X}$ und die Stichprobenstreuung S^2.

5.5 Satz. Seien $X_1,...,X_n$ i.i.d. Zva mit $X_i \sim N(a,\sigma^2)$ und n>1. Dann gilt:

(a) $\bar{X}$ und S^2 sind unabhängig.

(b) $\bar{X} \sim N(a,\frac{\sigma^2}{n}) \;\Leftrightarrow\; \sqrt{n}\,\frac{\bar{X}-a}{\sigma} \sim N(0,1)$; $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$;

Insbesondere ist die Verteilung von S^2 unabhängig von a.

Beweis. Setze $W_i := (X_i-a)/\sigma$; dann sind $W_1,...,W_n$ unabhängige Zva mit $W_i \sim N(0,1)$.

Sei A eine sogenannte Helmert–Transformation, d.h. eine orthogonale Matrix mit der 1. Zeile

$(1/\sqrt{n},...,1/\sqrt{n})$ und $\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} := A \begin{bmatrix} W_1 \\ \vdots \\ W_n \end{bmatrix}$.

Gemäß 5.1c sind $Y_1,...,Y_n$ i.i.d. Zva mit $Y_i \sim N(0,1)$ und es gilt:

(i) $Y_1 = \sqrt{n}\cdot\frac{1}{n}\sum_1^n W_i = \sqrt{n}\,\frac{\bar{X}-a}{\sigma}$;

(ii) $\sum_1^n Y_i^2 = \sum_1^n W_i^2 = \frac{1}{\sigma^2}\sum_1^n (X_i-a)^2 = \frac{1}{\sigma^2}\left[\sum_1^n (X_i-\bar{X})^2 + n(\bar{X}-a)^2\right] = (n-1)\,S^2/\sigma^2 + Y_1^2$;

(die hier benutzte Verschiebungsformel für S^2 ist eine Übungsaufgabe) somit

$\sum_2^n Y_i^2 = (n-1)\,S^2/\sigma^2$.

Die Aussagen von (b) folgen nun unmittelbar. Teil (a) ergibt aus $\bar{X} = f(Y_1)$ und $S^2 = g(Y_2,...,Y_n)$, wobei sich f und g aus den obigen Beziehungen berechnen. []

5.6 Satz (χ^2–Test für die Varianz). Unter $\vartheta=(a,\sigma^2) \in \mathbb{R}\times(0,\infty)$ seien $X_1,...,X_n$ iid mit $X_i \sim N(a,\sigma^2)$; $\sigma_0 > 0$, $0 < \alpha < 1$. Für das Testproblem $H_1 : \sigma^2 \le \sigma_0^2$, $K_1 : \sigma^2 > \sigma_0^2$ verwendet man die Prüfgröße $T := (n-1)S^2/\sigma_0^2$. Für das Testen von H_1 gegen K_1 ist dann der Test

$$\varphi_1 := \begin{cases} 1 & \text{für } T \overset{>}{\underset{\le}{}} \chi^2_{n-1,\alpha} \\ 0 & \end{cases}$$

unverfälschter Test z.N. α.

Die Optimalität von φ_1 und der zweiseitige Fall sollen hier nicht behandelt werden.

Beweis von 5.6. Sei $H_n(t) := \int_0^t h_n(s)\,ds$ die Verteilungsfunktion zu χ^2_n; H_n ist stetig und streng isoton auf $[0,\infty)$, somit gilt: $1 - H_{n-1}(c) = \alpha$ für $c := \chi^2_{n-1,\alpha}$. Die Gütefunktion von φ_1 ist

$$E_{a,\sigma^2}[\varphi_1] = P_{a,\sigma^2}[T > c] = P_{a,\sigma^2}[(n-1)S^2/\sigma^2 > (\sigma_0^2/\sigma^2)\cdot c] = 1 - H_{n-1}((\sigma_0^2/\sigma^2)\cdot c) =: \beta(\sigma^2)$$

gemäß Satz 5.5. Somit ist β stetig und streng isoton auf $(0,\infty)$ mit $\beta(\sigma_0^2) = \alpha$. Damit gilt auch: $\beta(\sigma^2) \le \alpha$ für $\sigma^2 \le \sigma_0^2$ und $\beta(\sigma^2) > \alpha$ für $\sigma^2 > \sigma_0^2$. []

5.6* Bemerkung. In der Situation von Satz 5.6 ist für das Testproblem $H_1^* : \sigma^2 \ge \sigma_0^2$, $K_1^* : \sigma^2 < \sigma_0^2$ der Test $\varphi_1^* := \begin{cases} 1 & \text{für } T \overset{<}{\underset{\ge}{}} \chi^2_{n-1,1-\alpha} \\ 0 & \end{cases}$ unverfälschter Test z.N. α.

Der Beweis verläuft wie der von Satz 5.6.

In 6.8 wird dieses Ergebnis benutzt, um eine sogenannte obere Konfidenzschranke für den Preis einer Option in einem Finanzmarktmodell herzuleiten. []

5 B. Die Student–Verteilung.

Definition 2. Sind X und Y unabhängige Zva mit $X \sim N(a,1)$ und $Y \sim \chi^2_n$, so heißt die Verteilung von $X / \sqrt{\frac{1}{n} Y}$ **nichtzentrale Student–Verteilung** $t_n(a)$ mit **Nichtzentralitätsparameter** $a \in \mathbb{R}$ (mit n Freiheitgraden). $t_n(0) =: t_n$ heißt (zentrale) Stundent–Verteilung .

$t_{n,\alpha}$ sei das α–Freaktil zu t_n.

5.7 Bemerkung. Ist $T : \mathbb{R}^2 \mapsto \bar{\mathbb{R}}$ die Abbildung $T(x,y) := x/\sqrt{\frac{1}{n} y}$, so kann $t_n(a)$ auch definiert werden als $t_n(a) := (N(a,1) \times \chi^2_n)T^{-1}$. []

5.8 Satz. (a) Die Verteilung t_n ist symmetrisch und hat die Lebesgue–Dichte:

$$x \mapsto \text{const} \cdot (1 + \tfrac{1}{n} x^2)^{-(n+1)/2};$$

(b) t_1 ist die Cauchy–Verteilung.

(c) $t_n(a)$ hat die Lebesgue–Dichte: $x \mapsto \text{const} \int_0^\infty y^n \exp\{-\tfrac{n}{2} y^2 (1 + \tfrac{1}{n} x^2) + xya\}\, dy$.

Beweis: Seien X und Y unabhängige Zva mit $X \sim N(a,1)$ und $Y \sim \chi^2_n$ und Dichten $f_{a,1}$ bzw. h_n..

Die Dichte g_n von $\sqrt{\frac{1}{n}\ Y}$ ergibt sich aus :

$P[\sqrt{\frac{1}{n}\ Y} \le x] = P[\ Y \le n\,x^2] = {}_0\int^{nx^2} h_n(y)\,dy = {}_0\int^{x} 2nz\ h_n(nz^2)\,dz = {}_0\int^{x} g_n(z)\,dz.$

Gemäß 5.2 ist also jetzt $x \mapsto {}_0\int^{\infty} f_{a,1}(xz)\ 2nz\ h_n(nz^2)\ z\ dz$ zu berechnen; dies ist eine Übungsaufgabe. []

Aus Satz 5.5 erhalten wir sofort:

5.9 Korollar. Seien $X_1,...,X_n$ i.i.d. Zva mit $X_i \sim N(a,\sigma^2)$ und $n>1$. Dann gilt:

$$\sqrt{n}\,\frac{\bar{X} - a}{S} \sim t_{n-1}\,.$$

Beweis: Es ist $T := \sqrt{n}\,\frac{\bar{X} - a}{S} = \sqrt{n}\,\frac{\bar{X} - a}{\sigma} \,/\, \sqrt{\frac{1}{n-1}\ (n-1)\cdot S^2/\sigma^2}$. Mit den Bezeichnungen von Beweis von Satz 5.5 gilt auch: $T = Y_1 \,/\, \sqrt{\frac{1}{n-1}\ \sum_2^n\ Y_i^2}$.[]

Nun kann ein weiterer Test vorgestellt werden, dessen zweiseitige Form in § 11 bewiesen wird. Dieser Test entspricht dem Gaußtest in dem Fall, daß die Varianz unbekannt ist. In der Teststatistik zum Gaußtest wird hier die bekannte Varianz σ_0^2 durch den Schätzer S^2 ersetzt.

5.10 Satz. (t–Test). Es sei $S = \mathbb{R}^n$, $\mathfrak{P} = \{ \bigotimes_{i=1}^{n} N(a,\sigma^2);\ \vartheta=(a,\sigma^2) \in \mathbb{R}\times(0,\infty)\}$, $a_0 \in \mathbb{R}$, $0<\alpha<1$.

Für die Testprobleme $H_1: a \le a_o$, $K_1: a > a_o$ und $H_2: a = a_o$, $K_2: a \ne a_o$

verwendet man die Prüfgröße $T := \sqrt{n}\,\frac{\bar{X} - a_o}{S}$. Definiert man Transformationen $\pi_c : S \mapsto S$ gemäß $\pi_c(a_o+u_1,...,a_o+u_n) = (a_o+c\cdot u_1,...,a_o+c\cdot u_n)$, so gilt:

(i) Für das Testen von H_1 gegen K_1 ist der Test

$$\varphi_1 := \begin{cases} 1 & \text{für } T \overset{>}{\le} t_{n-1,\alpha} \\ 0 & \end{cases}$$

schärfster Test in $\{\varphi \in \Phi_\alpha;\ \varphi\circ\pi_c = \varphi\ \ \forall\ c>0\}$ und damit unverfälscht.

(ii) Für das Testen von H_2 gegen K_2 ist der Test

$$\varphi_2 := \begin{cases} 1 & \text{für } |T| \overset{>}{\le} t_{n-1,\alpha/2} \\ 0 & \end{cases}$$

schärfster Test in $\{\varphi \in \Phi_\alpha;\ \varphi\circ\pi_c = \varphi\ \ \forall\ c\ne 0\}$ und damit unverfälscht.

Nach 5.8 und 5.9 gilt: $\beta_{\varphi_1}(a_o,\sigma^2) = \alpha = \beta_{\varphi_2}(a_o,\sigma^2)$. Daraus folgt: φ_2 ist Test z.N.α und mit Monotoniebetrachtungen auch: φ_1 ist Test z.N. α.

Zur Motivation der Forderung $\varphi \circ \pi_c = \varphi$ soll nur der Fall "c>0" betrachtet werden. Der Fall "c=–1" wurde bereits beim zweiseitigen Gaußtest betrachtet. Die Abweichungen vom Sollwert a_o sind relativ zur Varianz zu sehen. Die Unkenntnis über die Varianz führt dazu, daß man nicht weiß, ob Abweichungen von a_o als groß oder klein zu bewerten sollen. Deshalb sollen x und $x'=\pi_c(x)$, wo also bei x' die Abweichungen um den Faktor c aufgebläht sind, zur gleichen Entscheidung führen.

5 C. Die Fisher–Verteilung.

Die Fisher–Verteilung wird hier nicht direkt benötigt. Da aber die hier entwickelten t–Tests in der Literatur oft auch als spezielle F–Tests angegeben werden mit dem α–Fraktil der Fisher–Verteilung als kritischen Wert, sollen hier zumindest die Beziehungen zwischen der Fischer–Verteilung und der Student–Verteilung aufgezeigt werden.

Definition 3. Sind X und Y unabhängige Zva mit $X \sim \chi^2_m$, $Y \sim \chi^2_n$, so ist die Verteilung von

$\frac{1}{m}X / \frac{1}{n}Y$ die **Fisher–Verteilung** $F_{m,n}$ mit Freiheitsgraden m und n.
$F_{m,n,\alpha}$ ist das α–Fraktil zu $F_{m,n}$.

5.11 Satz. (a) Bei $V \sim t_n$ gilt $V^2 \sim F_{1,n}$.

(b) $(t_{n,\alpha/2})^2 = F_{1,n,\alpha}$.

Beweis: a) folgt sofort aus der Def. 2 von t_n und 5.3a.
b) Die Vf von V ist streng isoton und stetig. Also ist $\tau = t_{n,\alpha/2}$ bestimmmt durch $P[V>\tau]=\alpha/2$. Wegen der Symmetrie von t_n folgt: $P[V<-\tau] = P[-V>\tau]=\alpha/2$.
Somit hat man: $P[V^2>\tau^2] = P[V>\tau] + P[V<-\tau] = \alpha$.
Da auch die Vf von V^2 streng isoton und stetig auf $[0,\infty)$ ist, ist alles gezeigt. $[]$

5.12 Folgerung: Für den Test φ_2 aus 5.10 gilt:

$$\varphi_2 := \begin{cases} 1 & \text{für } T^2 > \\ 0 & \leq \end{cases} F_{1,n-1,\alpha} .$$

§6 Konfidenzintervalle.

Dieser §6 beschäftigt sich mit Schätzproblemen für einen reellen Parameter $\rho : \Theta \mapsto R \subset \mathbb{R}$ bei einem statistischen Raum $(S,\mathfrak{S},\mathfrak{P}=\{P_\vartheta,\vartheta\in\Theta\})$. Ein Ziel ist dabei die Bestimmung einer Fehlerschranke ε, sodaß der Schätzer $\hat\rho$ durch die Angabe $\rho = \hat\rho \pm \varepsilon$ ergänzt wird Diese Angabe soll bedeuten, daß der Fehler des Schätzers $|\rho - \hat\rho|$ mit großer W. kleiner als ε ist.

Definition 1. Sind $\underline{\lambda}$, $\overline{\lambda} : S \mapsto \bar{\mathbb{R}}$ mb, so heißt $[\underline{\lambda}\,,\,\overline{\lambda}]\cap R$ ein **Konfidenzintervall** für den reellen Parameter $\rho : \Theta \mapsto R \subset \mathbb{R}$. Ist $\overline{\lambda}\equiv\infty$, so heißt $\underline{\lambda}$ **untere Konfidenzschranke**. Allgemeiner heißt eine Abbildung C von S in die Potenzmenge von R **Konfidenzbereich**, falls
$\{x\in S;\ C(x) \ni r\} \in \mathfrak{S}\ \ \forall\, r\in R$.

Bei Anwendungen sind die interessierenden Konfidenzbereiche in der Regel Konfidenz–intervalle. Der allgemeinere Begriff des Konfidenzbereichs wird für die Formulierung des weiter unten folgenden Korrespondenzsatzes gebraucht.

Beispiel 1. Konfidenzintervall für den Mittelwert einer Normalverteilung.

Seien $X_1,...,X_n$ iid Zva mit $X_i \sim N(a,\sigma_o^2)$, $a = \vartheta \in R = \Theta$, R ein Intervall in $\mathbb{R}$ und $\rho(a) = a$. Dazu betrachten wir

$$\varphi_{2a} := \begin{cases} 1 & \sqrt{n}\,\dfrac{|\bar X - a|}{\sigma_o} > N_{\alpha/2} \\ 0 & \sqrt{n}\,\dfrac{|\bar X - a|}{\sigma_o} \le N_{\alpha/2} \end{cases},$$

also den zweiseitigen Gaußtest. Dieser ist nach 4.13 Test z.N. α für das Testproblem $H = \{a\}$, $K = R\setminus\{a\}$. Es folgt

$$P_a\Big[\ \sqrt{n}\,\frac{|\bar X - a|}{\sigma_o} > N_{\alpha/2}\ \Big] \le \alpha \text{ (sogar = } \alpha) \ \forall\, a$$

$$\Leftrightarrow \quad P_a\Big[\ \sqrt{n}\,\frac{|\bar X - a|}{\sigma_o} \le N_{\alpha/2}\ \Big] \ge 1-\alpha \text{ (sogar = } 1-\alpha) \ \forall\, a$$

$$\Leftrightarrow \quad P_a\big[\ \bar X - \varepsilon \le a \le \bar X + \varepsilon\ \big] \ge 1-\alpha \text{ (sogar = } 1-\alpha) \ \forall\, a \quad \text{mit} \quad \varepsilon := \sigma_o \cdot N_{\alpha/2}/\sqrt{n}.$$

Setzt man $[\underline{\lambda}\,,\,\overline{\lambda}] := [\ \bar X - \varepsilon\,,\,\bar X + \varepsilon\ ]$, so ist $[\underline{\lambda}\,,\,\overline{\lambda}]$ ein Konfidenzintervall mit

$$P_a\big[\ \underline{\lambda} \le a \le \overline{\lambda}\ \big] \ge 1-\alpha \ \text{ (sogar = } 1-\alpha) \ \forall\, a.$$

Eine andere Schreibweise wäre $a = \bar X \pm \varepsilon$. Man erkennt hier die Abhängigkeit der Abweichung ε von α, σ_o und n. []

Definition 2. Ein Konfidenzbereich C für ρ hat das **Konfidenzniveau** $1-\alpha$, falls gilt:

$$P_\vartheta\big[\ C \ni \rho(\vartheta)\ \big] \ge 1-\alpha \ \ \forall\, \vartheta\in\Theta \ \Leftrightarrow: C \in \mathfrak{C}_{1-\alpha}$$

Der in Beispiel 1 aufgezeigte Zusammenhang mit der Testtheorie kann auch im allgemeinen Fall hergestellt werden.

Für $r \in R$ definieren wir die Hypothese H(r) gemäß:

$$\vartheta \in H(r) \ :\Leftrightarrow \ \rho(\vartheta) = r \ .$$

Dann gilt: $C \in \mathfrak{C}_{1-\alpha} \ \Leftrightarrow \ P_\vartheta[\, C \ni r \,] \geq 1-\alpha \ \ \forall\, \vartheta \in H(r), \forall\, r \in R.$

Setze: $\varphi \in \Phi_{\alpha,r} :\Leftrightarrow \varphi$ ist deterministischer Test mit $\beta_\varphi(\vartheta) \leq \alpha \ \ \forall\, \vartheta \in H(r)$.

Ein Test $\varphi \in \Phi_{\alpha,r}$ ist also ein deterministischer Test z.N. α für das Testen von H(r) [gegen $K(r) := \Theta \setminus H(r)$].

Die Idee für die Konstruktion eines Konfidenzbereichs C ist nun die folgende:

Wähle für $r \in R$ einen Test $\varphi_r \in \Phi_{\alpha,r}$. Ausgehend von dieser Familie $(\varphi_r, r \in R)$ von Tests wird dann ein Konfidenzbereich C so definiert, daß C(x) gerade die Werte r enthält, für die die Hypothese $\rho(\vartheta) = r$ auf Grund der Stichprobe x unter dem Test φ_r nicht verworfen wird.

[Ist $A_r := \{\varphi_r = 0\}$ der Annahmebereich des Tests φ_r und setzt man $A := \{(x,r) \in S \times R;\ x \in A_r\}$, so ist A_r gerade der r–Schnitt von A, und man wählt nun C(x) als der x–Schnitt von A.]

6.1 Korrespondenzsatz. Ist $(\varphi_r, r \in R)$ eine Familie von deterministischen Tests mit $\varphi_r \in \Phi_{\alpha,r}$ und setzt man $C := \{r;\ \varphi_r = 0\}$, so ist $C \in \mathfrak{C}_{1-\alpha}$.

Beweis: Zunächst gilt: $\{C \ni r\} = \{\varphi_r = 0\} \in \mathfrak{S} \ \ \forall\, r \in R$, also ist C ein Konfidenzbereich. Ferner gilt: $P_\vartheta[C \ni r] = P_\vartheta[\varphi_r = 0] = 1 - \beta_{\varphi_r}(\vartheta) \geq 1-\alpha$ für $r = \rho(\vartheta) \Leftrightarrow \vartheta \in H(r)$, also hat C das Konfidenzniveau $1-\alpha$. []

6.1* Bemerkung. Es gilt sogar, daß die folgenden Abbildungen γ und $\bar{\gamma}$ zueinander inverse Bijektionen sind: $\gamma : \underset{r \in R}{\times} \Phi_{\alpha,r} \mapsto \mathfrak{C}_{1-\alpha}$ mit $\gamma((\varphi_r, r \in R)) = C := \{r;\ \varphi_r = 0\}$,

$\bar{\gamma} : \mathfrak{C}_{1-\alpha} \mapsto \underset{r \in R}{\times} \Phi_{\alpha,r}$ mit $\bar{\gamma}(C) = (\varphi_r, r \in R)$, so daß $\varphi_r := \begin{cases} 1 & r \notin C \\ 0 & \text{für } r \in C \end{cases}$.

Beispiel 2. Konfidenzschranke für die Erfolgs–W. einer Binomial–Verteilung.

Seien $X_1,\ldots,X_n$ iid Zva mit $X_i \sim b(1,\vartheta)$, $0 \leq \vartheta \leq 1$, $\rho(\vartheta) = \vartheta$. Dann ist

$$\varphi_\vartheta := \begin{cases} 1 \\ 0 \end{cases} \text{für } \sum X_i \begin{matrix} > \\ \leq \end{matrix} b(n,\vartheta)_\alpha$$

nach Beispiel 1 in §4 Test z.N. α für das Testen von $H = [0,\vartheta]$ gegen $K = (\vartheta,1]$ und somit auch für $H(\vartheta) = \{\vartheta\}$ [gegen $K(\vartheta)$]; also gilt insbesondere auch: $\beta_{\varphi_\vartheta}(\vartheta) \leq \alpha$, d.h. $\varphi_\vartheta \in \Phi_{\alpha,\vartheta}$. Nach dem Korrespondenzsatz 6.1 ist

$$C := \{\vartheta;\ \varphi_\vartheta = 0\} = \{\vartheta;\ \textstyle\sum X_i \leq b(n,\vartheta)_\alpha\} \in \mathfrak{C}_{1-\alpha}.$$

Die Gestalt von C zu erkennen ist hier schwierig, weil die Abhängigkeit von ϑ im kritischen Wert $b(n,\vartheta)_\alpha$ liegt. Es soll nun gezeigt werden, daß gilt:

$$(\,\underline{\lambda}\,, 1\,] \subset C \subset [\,\underline{\lambda}\,, 1\,] \text{ für eine mb Funktion } \underline{\lambda} : S \mapsto [0,1] \ .$$

Dann kann o.E. $C = [\,\underline{\lambda}\,,\,1\,]$ gewählt werden, und $\underline{\lambda}$ ist eine untere Konfidenzschranke.
Die Funktionen $\beta_c(\vartheta) = P_\vartheta[\,\sum X_i > c\,]$ wurden bereits als isoton in ϑ erkannt. Es folgt, daß
$\vartheta \mapsto b(n,\vartheta)_\alpha = \min\{c\in\mathbb{N};\, P_\vartheta[\,\sum X_i > c\,] \le \alpha\}$ isoton ist. Setze nun

$$\underline{\lambda} := \inf\{\,\vartheta;\, \vartheta \in C(x)\,\} = \inf\{\,\vartheta;\, b(n,\vartheta)_\alpha \ge \sum X_i\,\}.$$

Für große n ergibt sich eine Vereinfachung, wenn $b(n,\vartheta)_\alpha$ wie in §4 durch N_α ausgedrückt wird. []

Bisher wurde die Forderung gestellt, daß der "wahre" Wert $\rho(\vartheta)$ unter ϑ mit großer W. überdeckt wird. Dies leistet auch der Konfidenzbereich $C \equiv R$. Jetzt soll die W. betrachtet werden, daß ein falscher Wert $r \ne \rho(\vartheta)$ überdeckt wird.

Definition 3. Ein Konfidenzbereich $C \in \mathfrak{C}_{1-\alpha}$ heißt **unverfälscht**, wenn gilt:

$$P_\vartheta[\,C \ni r\,] \le 1-\alpha \quad \forall\, r\ne\rho(\vartheta)\ \forall\, \vartheta\in\Theta \iff:\ C \in \tilde{\mathfrak{C}}_{1-\alpha}.$$

Um wieder einen Zusammenhang zur Testtheorie herzustellen, betrachten wir neben der Hypothese $H(r) = \{\vartheta\in\Theta;\, \rho(\vartheta)=r\}$ eine Alternative $K(r) := \{\vartheta\in\Theta;\, \rho(\vartheta)\ne r\}$ und setzen

$$\tilde{\Phi}_{\alpha,r} := \{\,\varphi\in\Phi_{\alpha,r};\, \beta_\varphi(\vartheta) \ge \alpha\ \forall\ \vartheta\in K(r)\}$$

für die Klasse der deterministischen unverfälschten Tests z.N.α für das Testproblem H(r) gegen K(r).

6.2 Korrespondenzsatz. (Fortsetzung) Ist $(\varphi_r, r\in R)$ eine Familie von deterministischen Tests mit $\varphi_r \in \tilde{\Phi}_{\alpha,r}$ und setzt man $C := \{r;\, \varphi_r = 0\}$, so ist $C \in \tilde{\mathfrak{C}}_{1-\alpha}$.

Beweis. Es gilt:
$P_\vartheta[\,C \ni r\,] = P_\vartheta[\,\varphi_r = 0\,] = 1 - \beta_{\varphi_r}(\vartheta) \le 1-\alpha\ \ \forall\,\vartheta\in K(r)$, also $\forall\, r\ne\rho(\vartheta)$. []

6.2* Bemerkung. Ist γ wie in 6.1*, so ist (die Restriktion von) γ eine bijektive Abildung von $\times_{r\in R} \tilde{\Phi}_{\alpha,r}$ nach $\tilde{\mathfrak{C}}_{1-\alpha}$.

6.3 Korollar. Das in Beispiel 1 angegebene Konfidenzintervall ist ein unverfälschtes Konfidenzintervall zum Konfidenzniveau $1-\alpha$ für den Parameter a.

Beweis: Die Aussage folgt sofort aus 6.2 und 4.13. Man kann auch die folgende Beobachtung machen. Wie wir im Beispiel 1 gesehen haben, gilt:

$$P_a\left[\,-N_{\alpha/2} \le \sqrt{n}\,\frac{\bar{X}-a}{\sigma_o} \le N_{\alpha/2}\,\right] = P_a[\,\bar{X}-\varepsilon \le a \le \bar{X}+\varepsilon\,] = 1-\alpha\ \ \forall\, a \quad \text{mit}\ \ \varepsilon := \sigma_o\cdot N_{\alpha/2}/\sqrt{n}.$$

Weil $\sqrt{n}\,\frac{\bar{X}-a}{\sigma_o}$ eine positive Dichte hat, gilt dann auch für $\eta,\delta > 0$:

$P_a\left[-N_{\alpha/2} + \sqrt{n}\,\frac{\eta}{\sigma_o} \le \sqrt{n}\,\frac{\bar X - a}{\sigma_o} \le N_{\alpha/2} - \sqrt{n}\,\frac{\delta}{\sigma_o} \right] < 1-\alpha$, also:

(6.4) $\qquad P_a\left[\bar X - \varepsilon + \delta \le a \le \bar X + \varepsilon - \eta \right] < 1-\alpha \quad \forall\, a.$

Man kann (6.4) so interpretieren, daß man das Konfidenzintervall $[\bar X - \varepsilon\ ,\ \bar X + \varepsilon]$ nicht weiter verkleinern kann. Die Eigenschaft (6.4) ist eine weitere Möglichkeit, für Konfidenzintervalle die Unverfälschtheit zu erklären. $[\,]$

6.5 Korollar. Seien $X_1,...,X_n$ iid Zva mit $X_i \sim N(a,\sigma^2)$, $\vartheta=(a,\sigma^2) \in \mathbb{R}\times(0,\infty)\}$, $0<\alpha<1$,

$\bar X$ das Stichprobenmittel und S^2 die Stichprobenstreuung. Dann ist

$$\left[\bar X - \sqrt{\frac{1}{n}}\cdot S \cdot t_{n-1,\alpha/2}\ ,\ \bar X + \sqrt{\frac{1}{n}}\cdot S \cdot t_{n-1,\alpha/2} \right]$$

ein unverfälschtes Konfidenzintervall zum Konfidenzniveau $1-\alpha$ für den Parameter a.

Hier ist die Intervallänge über S abhängig vom Zufall, was störend sein kann..
Der **Beweis** verläuft analog zu dem von 6.3 mit 5.7(ii) anstelle von 4.13.

Eine weitere Eigenschaft von Tests war die Minimierung der Fw. 2. Art. Auch diese läßt sich zwanglos auf die Konfidenzbereiche übertragen. Dies soll hier aber nicht geschehen.

6.6 Beispiel. Obere Konfidenzschranke für die Varianz einer Normalverteilung.

Unter $\vartheta = (a,\sigma^2) \in \Theta := \mathbb{R}\times(0,\infty)$ [oder $\Theta := I\times(0,\infty)$ mit einem beliebigen Intervall in $\mathbb{R}$] seien $X_1,...,X_n$ iid Zva mit $X_i \sim N(a,\sigma^2)$ und $\rho : \Theta \mapsto (0,\infty)$, $\rho(\vartheta) = \sigma^2$. Dazu betrachten wir den Test

$$\varphi_{\sigma^2} := \begin{cases} 1 \\ 0 \end{cases} \begin{matrix} < \\ \ge \end{matrix} \chi^2_{n-1,1-\alpha} =: c\ ,$$

Dieser ist nach 5.6* unverfälschter Test z.N. α für das einseitige Testproblem

$H_1^*(\sigma^2) = \{\ \vartheta' \in \Theta;\ \rho(\vartheta') \ge \sigma^2\}$ gegen $K_1^*(\sigma^2) = \{\ \vartheta' \in \Theta;\ \rho(\vartheta') < \sigma^2\}$.

Dann ist φ_{σ^2} ist auch Test z.N. α für das Testproblem:

$H(\sigma^2) = \{\ \vartheta' \in \Theta;\ \rho(\vartheta') = \sigma^2\}$ gegen $K(\sigma^2)$.

Nach dem Korrespondenzsatz 6.1 ist

$$\mathfrak{C}_{1-\alpha} \ni C := \{\sigma^2;\ \varphi_{\sigma^2}=0\} = \{\sigma^2;\ (n-1)S^2/\sigma^2 \ge c\} = \{\sigma^2;\ (n-1)S^2/c \ge \sigma^2\}$$
$$= (0,(n-1)S^2/c] \quad \left[= (-\infty,(n-1)S^2/c] \cap (0,\infty)\right].$$

Somit ist $\bar\sigma^2 := (n-1)S^2/c$ eine obere Konfidenzschranke für σ^2, für die allerdings keine Unverfälschtheit gezeigt wurde. Der Beweis von 5.6* oder 5.6 zeigt aber, daß genauer

$P_{a,\sigma^2}\left[\ \bar\sigma^2 \ge \sigma^2\ \right] = 1-\alpha$ gilt. Gemäß 5.5 hat $\bar\sigma^2$ (unter $\vartheta = (a,\sigma^2)$) eine positive Dichte auf $(0,\infty)$.

Somit gilt auch für $\eta > 0$: $P_{a,\sigma^2}\left[\ \bar\sigma^2 \ge \sigma^2 + \eta\ \right] < 1-\alpha$, also:

(6.7) $\qquad P_{a,\sigma^2}\left[\ \bar\sigma^2 - \eta \ge \sigma^2\ \right] < 1-\alpha\ ,$ für $\eta > 0$.

Diese Eigenschaft kann man wieder so interpretieren, daß man die obere Konfidenzschranke $\bar{\sigma}^2$ nicht weiter verkleinern kann. Diese Eigenschaft entspricht der Eigenschaft 6.4 bei Konfidenzintervallen. $[]$

6.8 Anwendung auf die Black–Scholes–Formel.

Es soll die folgende **(Europäische Call–) Option** betrachtet werden. Eine Bank bietet z.Zt. t=0 die Option an, 1 Wertpapier (etwa 1 \$) zum Preis von K Euro in t=T zu kaufen. Der Marktpreis oder Kurs des Wertpapiers in t sei gegeben durch die Zva P(t). Ist $P(T) \le K$, so braucht man die Option nicht auszuüben. Der Gewinn des Käufers (Verlust der Bank) ist dann $(P(T) - K)^+ \ge 0$. Dafür muß der Käufer in t=0 eine **Prämie** π zahlen. Es sei nun die Annahme gemacht, daß $\{P(t),\ 0 \le t \le T\}$ durch eine sogenannte **geometrische Brownsche Bewegung** beschrieben wird, d.h.

(*) Für $0 \le t_1 < t_2 < \ldots < t_n \le T$, $n \in \mathbb{N}$, sind $\ln P(t_2) - \ln P(t_1), \ldots, \ln P(t_n) - \ln P(t_{n-1})$ unabhängige Zva und es gilt: $\ln P(s+t) - \ln P(s) \sim N(a \cdot t, \sigma^2 \cdot t)$.

Dann ist eine **faire Prämie** π nach Black–Scholes (1973) zu wählen als :

$$\pi(\sigma^2) := P(0) \cdot \Phi\left(\left[\ln(P(0)/K) + (r + \tfrac{1}{2}\sigma^2)T \right] / \sigma\sqrt{T} \right) - K \cdot e^{-rT} \cdot \Phi\left(\left[\ln(P(0)/K) + (r - \tfrac{1}{2}\sigma^2)T \right] / \sigma\sqrt{T} \right),$$

$$\pi(\sigma^2) := P(0) \cdot \Phi\left(\frac{\ln(P(0)/K) + (r + \tfrac{1}{2}\sigma^2) \cdot T}{\sigma \cdot \sqrt{T}} \right) - K \cdot e^{-rT} \cdot \Phi\left(\frac{\ln(P(0)/K) + (r - \tfrac{1}{2}\sigma^2) \cdot T}{\sigma \cdot \sqrt{T}} \right),$$

wenn Φ hier die Vf zu N(0,1) ist.

Dies ist die **Black–Scholes–Formel** für die Bewertung einer Option in einem zeitstetigen Modell.

Dabei ist r die Zinsrate im zeitstetigen Modell,
σ^2 die sog. **Volatilität** als Maß für die Kursfluktationen,
P(0) der Kurs z.Zt. 0,
T der Fälligkeitskeitstermin der Option,
K der Wahrnehmungspreis der Option.

Unbekannt ist in der Regel die Volatilität σ^2; es ist bemerkenswert, daß der Preis nicht von dem i.a. ebenfalls unbekannten Drift a abhängt. Es gilt die interessante Beziehung:

(**) $\sigma^2 \to \pi(\sigma^2)$ ist isoton.

Beobachtet man die Kursentwicklung zu den Zeitpunkten t = 0,....,n, so erhält man die Realisierungen der Zva $X_1 = \ln P(1) - \ln P(0), \ldots, X_n = \ln P(n) - \ln P(n-1)$. Auf Grund obiger Verteilungsannahme gilt: $X_1, \ldots, X_n$ sind unabhängig, und $X_i \sim N(a, \sigma^2)$.

Gemäß 6.6 ist $\bar{\sigma}^2 := (n-1)S^2/\chi^2_{n-1,1-\alpha}$ eine obere Konfidenzschranke für σ^2. Nimmt die Bank nun $\pi(\bar{\sigma}^2)$ als Prämie für die Option, so weiß man von der fairen Prämie $\pi(\sigma^2)$:

$P_{a,\sigma^2}\left[\bar{\sigma}^2 \ge \sigma^2 \right] = 1 - \alpha \Rightarrow P_{a,\sigma^2}\left[\pi(\bar{\sigma}^2) \ge \pi(\sigma^2) \right] \ge 1 - \alpha$ wegen $\{\bar{\sigma}^2 \ge \sigma^2\} \subset \{\pi(\bar{\sigma}^2) \ge \pi(\sigma^2)\}$,

also: $P_{a,\sigma^2}\left[\pi(\bar{\sigma}^2) < \pi(\sigma^2) \right] \le \alpha$.

Die Bank nimmt also höchstens mit W. α einen zu niedrigen Preis. $[]$

§7 Exponentialfamilien und Familien mit monotonen Dichtequotienten.

Um den Binomialtest und den Gaußtest in einem gemeinsamen Rahmen behandeln zu können, soll zunächst eine gemeinsame Struktur der Verteilungsannahme herausgearbeitet werden.

Definition 1. Eine Verteilungsannahme $\mathfrak{P} = \{ P_\vartheta, \vartheta\in\Theta \}$ heißt (einparametrige) **Exponential–familie** auf $(S,\mathfrak{S})$ [in ζ und T], wenn gilt:

(i) Es existiert ein σ–endliches Maß μ auf $(S,\mathfrak{S})$ mit $\mathfrak{P} << \mu$, d.h. jedes $P_\vartheta \in \mathfrak{P}$ hat eine μ–Dichte p_ϑ .

(ii) Es existieren eine Normierungskonstante $C : \Theta \mapsto (0,\infty)$,
ein reeller Parameter $\zeta : \Theta \mapsto \mathbb{R}$,
eine reelle Statistik $T : S \mapsto \mathbb{R}$ und
eine mb Funktion $h : S \mapsto [0,\infty)$ mit
$p_\vartheta = C(\vartheta)\cdot\exp\{ \zeta(\vartheta)\cdot T \}\cdot h$.

7.1 Eigenschaften.

(a) Für ν mit $d\nu := h\, d\mu$ als dominierendes Maß kann $h \equiv 1$ gewählt werden..

(b) C ist durch ζ, T und ν bestimmt gemäß: $1/C(\vartheta) = \int \exp\{\zeta(\vartheta)\cdot T\}\, d\nu$.

(c) Θ kann so gewählt werden, daß $\zeta = id_\Theta$ gilt.
Aus $\zeta(\vartheta) = \zeta(\vartheta')$ folgt nämlich $P_\vartheta = P_{\vartheta'}$ und somit $\vartheta=\vartheta'$ per Def. von Θ. Also ist ζ injektiv. Nun kann ϑ und Θ durch ζ und $\zeta(\Theta)$ ersetzt werden.

(d) Als notwendige Bedingung für das Vorliegen einer Exponentialfamilie erhält man: $P_{\vartheta'}[B] = 0 \Leftrightarrow P_\vartheta[B] = 0 \ \forall\, \vartheta,\vartheta'\in\Theta$. Offenbar gilt: $P_\vartheta[B] = 0 \Leftrightarrow \nu[B] = 0$.

Beispiel 1. $\mathfrak{P} = \{ P_\vartheta = b(n,\vartheta), \vartheta \in \Theta = (0,1) \}$ mit dem Zählmaß μ auf $S = \{0,...,n\}$.

Dann gilt: $p_\vartheta(x) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} = \binom{n}{x} \left(\frac{\vartheta}{1-\vartheta}\right)^x (1-\vartheta)^n = \binom{n}{x} \exp\{x\cdot\ln(\frac{\vartheta}{1-\vartheta})\} (1-\vartheta)^n$
$=: h(x)\exp\{T(x)\cdot\zeta(\vartheta)\}\cdot C(\vartheta)$.

Hier ist $\zeta(\Theta) = (-\infty,\infty)$. Bei $\Theta = [0,1]$ wäre 7.1d verletzt.

Beispiel 2. $\mathfrak{P} = \{ P_\vartheta = N(a,\sigma_0^2), \vartheta=a \in \Theta \subset \mathbb{R} \}$ mit dem Lebesgue–Maß auf $(\mathbb{R},\mathfrak{B}_1)$ als μ.

Dann gilt: $p_\vartheta(x) = \sqrt{\frac{1}{2\pi\sigma_0^2}} \cdot \exp\{ -\frac{1}{2\sigma_0^2}(x-a)^2 \}$

$= \sqrt{\frac{1}{2\pi\sigma_0^2}} \exp\{ -\frac{a^2}{2\sigma_0^2} \}\cdot\exp\{ x\cdot\frac{a}{\sigma_0^2} \}\cdot\exp\{ -\frac{1}{2\sigma_0^2}\cdot x^2 \} =: C(\vartheta)\exp\{T(x)\cdot\zeta(\vartheta)\}\cdot h(x)$. []

Beispiel 3. Die Familie der Γ–Verteilungen $\mathfrak{P} = \{ P_\vartheta = \Gamma_{\alpha\nu}$, $\vartheta = \alpha \in (0,\infty) \}$ läßt sich leicht als Exponentialfamilie nachweisen.

Beispiel 4. Die Familie der Poissonverteilungen $\mathfrak{P} = \{ P_\vartheta = \pi(\vartheta)$, $\vartheta \in (0,\infty) \}$ ist ebenfalls eine Exponentialfamilie.

Betrachtet man zu den obigen Beispielen die zugehörigen Einstichprobenprobleme, so bilden auch deren Verteilungsannahmen wieder Exponentialfamilien nach dem folgenden

7.2 Satz. Ist $\mathfrak{P} = \{ P_\vartheta, \vartheta\in\Theta \}$ eine Exponentialfamilie auf $(S,\mathfrak{S})$ in ζ und T, so ist

$\mathfrak{P}^n := \{ \bigotimes_{i=1}^n P_\vartheta , \vartheta\in\Theta \}$ eine Exponentialfamilie auf $(\bigotimes_{i=1}^n S , \bigotimes_{i=1}^n \mathfrak{S})$ in ζ und

$T^{(n)} := \sum_{j=1}^n T\circ X_j$.

Beweis: Es gilt [vgl. W.th I 11.17]

$$d \bigotimes_{i=1}^n P_\vartheta / d \bigotimes_{i=1}^n \mu = C(\vartheta)^n \cdot \prod_{j=1}^n \exp\{ \zeta(\vartheta)\cdot T\circ X_j \} \, h\circ X_j \; d \bigotimes_{i=1}^n \mu$$
$$= C(\vartheta)^n \cdot \exp\{ \zeta(\vartheta)\cdot \sum_j T\circ X_j\} \cdot \prod_j h\circ X_j =: C^{(n)}(\vartheta)\cdot \exp\{ \zeta(\vartheta)\cdot T^{(n)}\} \cdot h^{(n)} \; . \; []$$

Für die Behandlung von einseitigen Testproblemen ist der Begriff der Exponentialfamilie noch zu eng. Er erfaßt nicht die nichtzentralen Studentverteilungen. Ein wichtiges Argument bei einseitigen Testproblemen [vgl. Beispiel 1 von §4] ist die Isotonie der Gütefunktionen. Diese läßt sich allgemein für die folgende Klasse von Verteilungsannahmen zeigen:

Definition 2. Eine Verteilungsannahme $\mathfrak{P} = \{ P_\vartheta, \vartheta\in\Theta \}$ mit $\Theta \subset \mathbb{R}$ hat **monotone Dichtequotienten** (DQ) in T, wenn gilt:

(i) μ ist ein σ–endliches Maß auf $(S,\mathfrak{S})$ sodaß gilt:
jedes $P_\vartheta \in \mathfrak{P}$ hat eine μ–Dichte p_ϑ mit $p_\vartheta > 0$ auf S;

(ii) $T : S \mapsto \mathbb{R}$ ist eine reellwertige Statistik, und zu $\vartheta_0, \vartheta_1 \in \Theta$ mit $\vartheta_0<\vartheta_1$ existiert eine isotone Funktion $f = f_{\vartheta_0\vartheta_1} : \mathbb{R} \mapsto (0,\infty)$ mit

$$p_{\vartheta_1} / p_{\vartheta_0} = f\circ T \; .$$

7.3 Bemerkung: Eigenschaft 7.1 (d) gilt auch hier.

7.4 Satz. Sei $\mathfrak{P} = \{ P_\vartheta, \vartheta\in\Theta \}$ eine einparametrige Exponentialfamilie mit

$$dP_\vartheta = C(\vartheta)\cdot \exp\{ \zeta(\vartheta)\cdot T \} \cdot h \; d\mu \; , \; \vartheta\in\Theta.$$

Ist dabei (o.E.) $h>0$ auf S und (o.E.) ζ isoton, so hat $\mathfrak{P}$ monotone DQ in T.

Beweis. Die Behauptung folgt aus

$$p_{\vartheta_1} / p_{\vartheta_0} = C(\vartheta_1)/C(\vartheta_0) \cdot \exp\{(\zeta(\vartheta_1)–\zeta(\vartheta_0))\cdot T\} \quad \text{mit} \quad \zeta(\vartheta_1)–\zeta(\vartheta_0) \geq 0. \; []$$

Für Tests φ der Gestalt $\varphi = \psi\circ T = \begin{cases} 1 & > \\ \gamma & \text{für } T = c \\ 0 & < \end{cases}$ ist die Gütefunktion isoton gemäß:

7.5 Satz. Hat $\mathfrak{P} = \{ P_\vartheta, \vartheta\in\Theta \}$ monotone DQ in T, so gilt:

Ist $\psi : \mathbb{R} \mapsto \mathbb{R}$ isoton und $E_\vartheta[\,|\psi\circ T|\,] < \infty \;\; \forall\, \vartheta\in\Theta$, so ist $\vartheta \mapsto E_\vartheta[\psi\circ T]$ isoton.

7.6 Bemerkung. Die Eigenschaft aus 7.5 impliziert [mit $\psi = \mathbf{1}_{(t,\infty)}$] und ist sogar äquivalent zu: $\vartheta \mapsto P_\vartheta[T>t]$ ist isoton $\forall\, t$ [vgl. Übung]. Man sagt:

"für $\vartheta_0<\vartheta_1$ ist T unter ϑ_1 stochastisch größer als ϑ_0" und schreibt: $P_{\vartheta_0} T^{-1} \preceq P_{\vartheta_1} T^{-1}$. []

Beweis von 7.5. Es sei $\vartheta_0<\vartheta_1$, p_ϑ und $f = f_{\vartheta_0\vartheta_1}$ wie in Def. 2. Setze:

$A := \{t\in\mathbb{R};\ f(t)\le 1\}$ und $B := \{t\in\mathbb{R};\ f(t)>1\}$ [dabei ist $A \neq \emptyset \neq B$],

Aufgrund der Isotonie von f und ψ existiert ein $b\in\mathbb{R}$ mit $\sup_A \psi \le b \le \inf_B \psi$. Es folgt:

$$E_{\vartheta_1}[\psi\circ T] - E_{\vartheta_0}[\psi\circ T] = \int \psi\circ T\,(p_{\vartheta_1} - p_{\vartheta_0})\,d\mu = \int \psi\circ T\,(f\circ T - 1)\,dP_{\vartheta_0}$$
$$= {}_{\{T\in A\}}\int \psi\circ T\,(f\circ T - 1)\,dP_{\vartheta_0} + {}_{\{T\in B\}}\int \psi\circ T\,(f\circ T - 1)\,dP_{\vartheta_0}$$
$$\ge {}_{\{T\in A\}}\int b\,(f\circ T - 1)\,dP_{\vartheta_0} + {}_{\{T\in B\}}\int b\,(f\circ T - 1)\,dP_{\vartheta_0} = E_{\vartheta_1}[b] - E_{\vartheta_0}[b] = 0.\ []$$

Der Nachweis für die Monotonie der DQ gestaltet sich manchmal schwierig. So folgt sie für die Klasse der nichtzentralen χ^2–Verteilungen mit folgendem

7.7 Lemma. Gegeben seien $c_{iz}>0$, $z\in\mathbb{N}_0$, mit $q_i(t) := \sum_{z=0}^{\infty} c_{iz}\, t^z < \infty$ für $0\le t<\infty$, i=0,1.

Ist dann $z \mapsto c_{1z}/c_{0z}$ isoton auf $\mathbb{N}_0$, so ist q_1/q_0 isoton auf $(0,\infty)$.

Eine mögliche Anwendung von 7.7 ist (vgl. den Beweis von 7.12 für n=1)

(7.8) Für $\zeta_0 < \zeta_1$ ist $t \mapsto \dfrac{e^{\zeta_1 t} + e^{-\zeta_1 t}}{e^{\zeta_0 t} + e^{-\zeta_0 t}}$ isoton auf $[0,\infty)$.

Für die Behandlung der nichtzentralen Student–Verteilung müssen Integralausdrücke studiert werden. Als Rahmen kann man folgendes Lemma wählen, das auch 7.7 miteinschließt.

7.9 Lemma. Seien μ ein σ–endliches Maß auf $[0,\infty)$, $c_i : [0,\infty) \mapsto (0,\infty)$ mb,

$g_t : [0,\infty) \mapsto (0,\infty)$ mb mit $0 < q_i(t) := \int c_i(z)\, g_t(z)\, \mu[dz] < \infty$ i = 0, 1, t>0.

Sind dann c_1/c_0 und g_{t_1}/g_{t_0} isoton für $t_0<t_1$, so ist q_1/q_0 isoton.

Beweis. Setze $p_t(z) := c_0(z)\cdot g_t(z) / q_0(t)$ und $\psi := c_1/c_0$; dann sind p_t μ–Dichten von W–Maßen auf $\mathbb{R}$ mit monotonen DQ in T(z)=z wegen

$$p_{t_1} / p_{t_0} = C \cdot g_{t_1} / g_{t_0} \quad \text{mit } C := q_0(t_1)/q_0(t_0) > 0$$

und ψ ist isoton. Also ist nach 7.5 $q_1(t)/q_0(t) = \int \psi(z)\, p_t(z)\, \mu[dz]$ isoton. []

Beweis von 7.7. Wegen der Stetigkeit in t=0 sei o.E. $t\in(0,\infty)$. Sei nun μ das Zählmaß auf $\mathbb{N}_0 \subset [0,\infty)$, $c_i(z) = c_{i\,[z]}$, $g_t(z)=t^z$. Dann kann 7.9 unmittelbar angewendet werden. []

Definition 3. Für $x\in\mathbb{R}^n$ sei

$x \succ 0$ (lexikographisch) $:\Leftrightarrow \exists\, i\in\{1,...,n\}$ mit $x_1=...=x_{i-1}=0$ und $x_i>0$.

$$\operatorname{sgn}(x) := \begin{cases} \pm 1 & \pm x \succ 0 \\ 0 & \text{für } x = 0 \end{cases}.$$

Offenbar gilt:

(7.10) $\operatorname{sgn}(-x) = -\operatorname{sgn}(x)$ und $\operatorname{sgn}(\cdot)$ ist mb.

Wir benötigen $U(x) := \operatorname{sgn}(x)\cdot x$, $x\in\mathbb{R}^n$, als n–dim. Verallgemeinerung von $|x|$, $x\in\mathbb{R}$.

7.11 Lemma. Sei P ein W–Maß auf $\mathbb{R}^n$ mit der μ–Dicht p; dabei sei $\mu = \lambda^n$ oder μ das Zählmaß auf $\mathbb{Z}^n$. Dann hat $U(x) := \operatorname{sgn}(x)\cdot x$ die Dichte

$t \mapsto p(t) + p(-t)$ bzgl. eines geeigneten Maßes μ_o.

Der **Beweis** mit $d\mu_o = (\mathbf{1}_{\{x\succ 0\}} + \frac{1}{2}\mathbf{1}_{\{x=0\}})\, d\mu$ ist eine Übungsaufgabe.

7.12 Satz. Die Familie der nicht–zentralen χ^2– Verteilungen $\{\chi^2_n(a^2),\ a^2\geq 0\}$ hat monotone DQ in T mit T als der Identität (auf $(0,\infty)$).

Beweis: Sei zunächst n=1. Ist f die Dichte zu N(0,1), so hat $\chi^2_1(a^2)$ gemäß 5.4 die Dichte:

$$h_{1,a^2}(t) = \tfrac{1}{2}\left[f(a-\sqrt{t}) + f(a+\sqrt{t})\right]/\sqrt{t} = \text{const}\cdot e^{-\frac{1}{2}a^2}\cdot e^{-\frac{1}{2}t}\cdot\tfrac{1}{2}\left[e^{a\sqrt{t}} + e^{-a\sqrt{t}}\right]/\sqrt{t}$$
$$= \text{const}\cdot e^{-\frac{1}{2}a^2}\cdot e^{-\frac{1}{2}t}\cdot t^{-\frac{1}{2}}\cdot\sum_{k=0}^{\infty}\frac{1}{(2k)!}a^{2k}\cdot t^k.$$

Für n>1 soll eine ähnliche Beziehung hergeleitet werden.

Ist $h_m(t) = \text{const}\cdot t^{\frac{1}{2}m-1}\cdot e^{-\frac{1}{2}t}$, t>0, die Lebesgue–Dichte der zentralen χ^2– Verteilungzu χ^2_m, so erhält man: $h_{1,a^2}(t) = e^{-\frac{1}{2}a^2}\cdot\sum_{k=0}^{\infty}\gamma_k\cdot a^{2k}\cdot h_{2k+1}(t)$ für gewisse Konstanten γ_k.

Für die Dichte h_{n,a^2} zu $\chi^2_n(a^2)$ ergibt sich per def. folgender Ausdruck, wobei die Vertauschung von $\int$ und $\sum$wegen monotoner Konvergenz gerechtfertigt ist:

$$h_{n,a^2} = h_{1,a^2} \star h_{n-1} = e^{-\frac{1}{2}a^2}\cdot\sum_{k=0}^{\infty}\gamma_k\cdot a^{2k}\cdot h_{2k+1}\star h_{n-1} = e^{-\frac{1}{2}a^2}\cdot\sum_{k=0}^{\infty}\gamma_k\cdot a^{2k}\cdot h_{2k+n}$$
$$\Rightarrow\ h_{n,a^2}(t) = e^{-\frac{1}{2}a^2}\cdot\sum_{k=0}^{\infty}\delta_k\cdot a^{2k}\cdot t^{k+\frac{1}{2}n-1}\cdot e^{-\frac{1}{2}t} = e^{-\frac{1}{2}a^2}\cdot e^{-\frac{1}{2}t}\cdot t^{\frac{1}{2}n-1}\sum_{k=0}^{\infty}\delta_k\cdot a^{2k}\cdot t^k$$

für gewisse Konstanten δ_k. Nun folgt die Behauptung aus 7.7; denn für $a_0<a_1$ ist $(a_1/a_0)^{2k}$ isoton in k. $[\,]$

Der zweiseitige t–Test (vgl. 5.10 (ii)) hängt nur über $|T|$ von der Stichprobe ab, wobei T unter der Hypothese t_{n-1}–verteilt ist. Für Aussagen über die Verteilung von $|T|$ unter der Alternativen brauchen wir Teil (b) des folgenden Satzes.

7.13 Satz.(a) Die Familie der nicht–zentralen t–Verteilungen $\{t_n(a),\ a\in\mathbb{R}\}$ hat monotone DQ in $T = \mathrm{id}_{\mathbb{R}}$, $n\in\mathbb{N}$.

(b) Für $U(t) := |t|$ hat $\{t_n(a)U^{-1},\ a\geq 0\}$ monotone DQ in $T = \mathrm{id}_{(0,\infty)}$, $n\in\mathbb{N}$.

Beweis: Nach 5.5c hat $t_n(a)$ die Lebesgue–Dichte:

$$q_a(t) := \text{const} \cdot \int_0^\infty y^n \exp\{-\frac{n}{2} y^2 (1 + \frac{1}{n} t^2) + tya\}\, dy .$$

a) Sei nun $a_0 < a_1$ und o.E. $t \in [0,\infty)$ wegen $q_a(-t) = q_{-a}(t)$ und $-a_1 < -a_0$.
Da q_a stetig in t ist [vgl. W.th. I 10.15], sei o.E. t>0. Mit ty=z ergibt sich:

$$q_a(t) = \text{const} \cdot t^{-(n+1)} \int_0^\infty e^{az} \exp\{-\tfrac{1}{2} z^2 (1 + n \cdot t^{-2})\}\, z^n\, dz.$$

Setzt man in 7.9 $c_i(z) = e^{a_i z}$, $g_t(z) = \exp\{-\frac{1}{2} z^2 (1 + n \cdot t^{-2})\}$ und $\mu[dz] = z^n\, dz$, so folgt die Behauptung.

b) $t_n(a)U^{-1}$ hat nach 7.11 die Dichte $q_a(t) + q_a(-t) = q_a(t) + q_{-a}(t)$ für t>0.
Sei nun a≥0 und t>0; dann gilt:

$$q_a(t) + q_{-a}(t) = \text{const} \cdot t^{-(n+1)} \int_0^\infty (e^{az} + e^{-az})\, g_t(z)\, \mu_o[dz] .$$

Setzt man in 7.9 $c_i(z) = e^{a_i z} + e^{-a_i z}$, so folgt mit (7.8) die Behauptung. []

§8 Das Testen zweier einfacher Hypothesen.

Sei $(S,\mathfrak{S},\mathfrak{P})$ ein statistischer Raum mit $\mathfrak{P} = \{P_0,P_1\}$, also $\Theta = \{0,1\}$, sowie

$$H := \{0\}\ ,\ K := \{1\}.$$

Dabei kann man sich etwa vorstellen, daß man etwa die Alternativen "in Ordnung"/"defekt" oder "weiblich"/"männlich" testen will. Im nächsten §9 kann ein allgemeineres einseitiges Testproblem auf den Fall zweier einfacher Hypothesen zurückgeführt werden.

Sei $E_i[...] = \int ...dP_i$ und μ ein dominierendes Maß, etwa $\mu = \frac{1}{2}(P_0+P_1)$; dabei seien

p_i die μ–Dichten von P_i , i = 0, 1.

Gesucht ist ein Test $\varphi^* \in \Phi_\alpha = \{\varphi\in\Phi;\ E_0[\varphi] \le \alpha\}$ mit

$$E_1[\varphi^*] = \sup_{\varphi\in\Phi_\alpha} E_1[\varphi] .$$

Es stellt sich also das folgende Optimierungsproblem:

Maximiere $\varphi \mapsto E_1[\varphi]$ unter der Nebenbedingung $\varphi\in\Phi,\ \alpha - E_0[\varphi] \ge 0$.

Dieses läßt sich auf ein Optimimierungsproblem ohne Nebenbedingung zurückführen mit der **Lagrange–Funktion**: $L(\varphi,r) := E_1[\varphi] + r\cdot(\alpha - E_0[\varphi])$;

dabei ist r≥0 der **Lagrangesche Multiplikator**.

Das Optimierungsproblem kann hier direkt ohne Anleihe an die Optimierungstheorie gelöst werden.

Mit $g_r := p_1 - r\cdot p_0$ gilt: $L(\varphi,r) = \alpha r + \int \varphi\cdot g_r\, d\mu$.

Es soll zunächst für $E_1[\varphi]$ eine von r abhängige obere Schranke hergeleitet und dann untersucht werden, ob für ein φ und ein r diese obere Schranke angenommen wird. Dann ist φ nämlich optimal.

Lemma 8.1. Für r>0 und $\varphi \in \Phi_\alpha$ gilt:

(a) $E_1[\varphi] \le \alpha r + \int g_r^+ \, d\mu$.

(b) Die folgenden Aussagen sind äquivalent:

(b1) $E_1[\varphi] = \alpha r + \int g_r^+ \, d\mu$;

(b2) $\varphi = \begin{cases} 1 & \text{für } p_1 > r \cdot p_0 \\ 0 & p_1 < r \cdot p_0 \end{cases}$ μ–f.s. und $E_0[\varphi] = \alpha$.

Beweis: a) $\int \varphi \, p_1 \, d\mu = \int \varphi \, g_r \, d\mu + r \cdot \int \varphi \, p_0 \, d\mu \le \int \varphi \, g_r^+ \, d\mu + r \cdot E_0[\varphi] \le \int g_r^+ \, d\mu + r \cdot \alpha$.
b) Offenbar gilt "=", also (b1) genau dann, wenn bei den obigen drei Abschätzungen stets "=" gilt, d.h. wenn die folgenden Beziehungen erfüllt sind:

(i) $\int \varphi \, g_r \, d\mu = \int \varphi \, g_r^+ \, d\mu$ (ii) $\int \varphi \, g_r^+ \, d\mu = \int g_r^+ \, d\mu$ (iii) $E_0[\varphi] = \alpha$.
Dabei gilt:

(i) $\Leftrightarrow$ $\int \varphi \, g_r^- \, d\mu = 0$ $\Leftrightarrow$ $\varphi = 0$ auf $\{g_r^- > 0\}$ μ–f.s. $\Leftrightarrow$ $\varphi = 0$ auf $\{g_r < 0\}$ μ–f.s. .

(ii) $\Leftrightarrow$ $\int (1-\varphi) \, g_r^+ \, d\mu = 0$ $\Leftrightarrow$ $1-\varphi = 0$ auf $\{g_r^+ > 0\}$ μ–f.s. $\Leftrightarrow$ $\varphi = 1$ auf $\{g_r > 0\}$ μ–f.s.
[]
Um eine einfachere Darstellung zu gewinnen, machen wir hier die folgende Annahme:

8.2 Annahme: $0 < p_i < \infty$ für i = 1,2.

Zur Konstruktion des optimalen φ^* machen wir in Hinblick auf 8.1(b2) den Ansatz $\varphi^* = \varphi_{r,\gamma}$ mit

$$\varphi_{r,\gamma} := \begin{cases} 1 & > \\ \gamma & \text{für } p_1/p_0 = r \\ 0 & < \end{cases}, \; r \ge 0, \; 0 \le \gamma \le 1 \quad \text{und}$$

$$E_0[\varphi_{r,\gamma}] = P_0[p_1/p_0 > r] + \gamma \, P_0[p_1/p_0 = r]$$

Ein passendes r wird sich als Fraktil der Verteilung von p_1/p_0 unter P_0 ergeben.

In Hinblick auf (b2) soll r so gewählt werden, daß gilt:

$$E_0[\varphi_{r,\gamma}] = P_0[p_1/p_0 > r] + \gamma \cdot P_0[p_1/p_0 = r] = \alpha .$$

Folgende Definition ist sinnvoll:

(8.3) $r_\alpha > 0$ sei das α–Fraktil der Verteilung der positiven Zva p_1/p_0 unter P_0.

8.4 Lemma. Sei Y eine f.s. reelle Zva auf $(\Omega, \mathfrak{F}, P)$ und $0 < \alpha < 1$.

Ist c das α–Fraktil von PY^{-1} und $\gamma = \dfrac{\alpha - P[Y>c]}{P[Y=c]}$, so gilt:

(i) $c \in \mathbb{R}$ bzw. $c \in (0,\infty)$ falls $Y > 0$ f.s. ;

(ii) $0 \le \gamma \le 1$;

(iii) $P[Y>c] + \gamma \cdot P[Y=c] = \alpha$;

(iv) $P[Y=c] = 0 \;\Rightarrow\; P[Y>c] = \alpha \;\Rightarrow\; \gamma = 0$.

Der **Beweis** folgt sofort aus 4.1. Dabei gilt im Falle $P[Y > 0] = 1$: $P[Y > c] \le \alpha < P[Y > 0]$.
Wählt man also $r = r_\alpha$ und $\gamma = \gamma_\alpha$ mit

$$\gamma_\alpha := (\alpha - P_0[p_1/p_0 > r_\alpha])/P_0[p_1/p_0 = r_\alpha] \quad \text{und}$$
$$\varphi^* := \varphi_{r_\alpha,\gamma_\alpha},$$

so erhält man mit 8.4: $r_\alpha > 0$, $0 \le \gamma_\alpha \le 1$ und somit: $\varphi^* \in \Phi$, ferner: $E_0[\varphi^*] = \alpha$.

8.5 Fundamentallemma von Neyman–Pearson. Seien r_α, γ_α und φ^* wie oben, $\varphi \in \Phi_\alpha$.

(a) Existiert ein $r>0$ mit $\varphi = \begin{cases} 1 & \text{für } p_1/p_0 > r \\ 0 & p_1/p_0 < r \end{cases}$ und gilt $E_0[\varphi] = \alpha$,
so ist φ schärfster Test in Φ_α.

(b) φ^* ist schärfster Test in Φ_α.

(c) Jeder schärfste Test φ in Φ_α hat die Gestalt $\varphi = \begin{cases} 1 & \text{für } p_1/p_0 > r_\alpha \\ 0 & p_1/p_0 < r_\alpha \end{cases}$
und es gilt: $E_0[\varphi] = \alpha$.

Beweis: Sei $\Lambda_\alpha(r) := \alpha \cdot r + \int g_r^+ \, d\mu$.
a) φ erfüllt (b2) von 8.1 und somit (b1), d.h. $E_1[\varphi] = \Lambda_\alpha(r)$.
Nach 8.1a gilt für jeden anderen Test $\varphi' \in \Phi_\alpha$: $E_1[\varphi'] \le \Lambda_\alpha(r) = E_1[\varphi]$.
b) folgt aus Teil (a) und den obigen Überlegungen. Insbesondere gilt: $E_1[\varphi^*] = \Lambda_\alpha(r_\alpha)$.
c) Sei φ schärfster Test in Φ_α. Mit dem Beweis von (b) und 8.1a gilt:
$\Lambda_\alpha(r_\alpha) = E_1[\varphi^*] \le E_1[\varphi] \le \Lambda_\alpha(r_\alpha)$. Also erfüllt φ (b1) von 8.1 und somit (b2) für $r = r_\alpha$. []

§9 Einseitige Tests bei monotonen Dichtequotienten.

Durch Monotoniebetrachtungen können die Ergebnisse von §8 auf einseitige Testprobleme übertragen werden.

9.1 Satz. Sei $0<\alpha<1$. $\mathfrak{P} = \{P_\vartheta, \vartheta \in \Theta\}$ habe monotone DQ in T. Zu $\vartheta_0 \in \Theta$ sei

$$H := \{ \vartheta \in \Theta; \vartheta \le \vartheta_0 \}, \quad K := \{ \vartheta \in \Theta; \vartheta > \vartheta_0 \} \ne \emptyset.$$

Für das Testen von H gegen K wählt man den Test

$$\varphi^* := \begin{cases} 1 & T > c \\ \gamma & \text{für } T = c, \\ 0 & T < c \end{cases}$$

dabei ist c das α–Fraktil zur Verteilung $P_{\vartheta_0} T^{-1}$ von T unter P_{ϑ_0} und $\gamma := (\alpha - P_{\vartheta_0}[T > c]) / P_{\vartheta_0}[T = c]$.

Dieser Test ist schärfster Test in Φ_α, insbesondere unverfälscht, und seine Gütefunktion ist isoton.

Beweis: Die Isotonie der Gütefunktion folgt sofort aus 7.5. Aus 8.4 ergibt sich: $0 \le \gamma \le 1$ und $E_{\vartheta_0}[\varphi^*] = \alpha$. Mit der Isotonie erhält man somit: $\varphi^* \in \tilde{\Phi}_\alpha$.

Nun wählen wir ein $\vartheta_1 \in K$ und setzen $H_0 := \{\vartheta_0\}$, $K_0 := \{\vartheta_1\}$.

Sei $f = f_{\vartheta_0\vartheta_1}$ die isotone Funktion mit $p_{\vartheta_1}/p_{\vartheta_0} = f \circ T$ und $r := f(c)$. Wir zeigen:

$$\varphi^* = \begin{cases} 1 & \text{für } p_{\vartheta_1}/p_{\vartheta_0} > r \\ 0 & p_{\vartheta_1}/p_{\vartheta_0} < r \end{cases} \quad \text{mit } r > 0 .$$

Aus der Isotonie von f folgt nämlich:

$\{p_{\vartheta_1}/p_{\vartheta_0} > r\} = \{ f \circ T > f(c) \} \subset \{ T > c \}$, $\{p_{\vartheta_1}/p_{\vartheta_0} < r\} = \{ f \circ T < f(c) \} \subset \{ T < c \}$.

Wäre $r \le 0$, so würde $\varphi^* \equiv 1$ und damit $E_{\vartheta_0}[\varphi^*] = 1 \ne \alpha$ gelten.

Mit $\Phi_{0\alpha} := \{\varphi; E_{\vartheta_0}[\varphi] \le \alpha\}$ folgt nun aus dem Fundamentallemma 8.5a: φ^* ist schärfster Test in $\Phi_{0\alpha}$ für das Testen von H_0 gegen K_0, also: $E_{\vartheta_1}[\varphi^*] = \sup_{\varphi \in \Phi_{0\alpha}} E_{\vartheta_1}[\varphi]$.

Nun war $\vartheta_1 \in K$ beliebig gewählt und φ^* unabhängig von ϑ_1; also ist φ^* schärfster Test in $\Phi_{0\alpha}$ für das Testen von H_0 gegen K. Wegen $\varphi^* \in \Phi_\alpha \subset \Phi_{0\alpha}$ folgt schließlich: φ^* ist schärster Test für das Testen von H gegen K. []

9.2 Bemerkung. φ^* ist determinisch mit $\gamma=0$, wenn $P_{\vartheta_0}T^{-1}$ stetig ist oder wenn allgemeiner der Wert $1-\alpha$ von der Vf zu $P_{\vartheta_0}T^{-1}$ angenommen wird, wenn also gilt:

$$1-\alpha \in \{ P_{\vartheta_0}[T \le t]; t \in \mathbb{R} \}.$$

Existiert nämlich ein t mit $P_{\vartheta_0}[T > t] = \alpha$, so gilt $c \le t$ und $P_{\vartheta_0}[T > c] = \alpha$ wegen $\alpha \ge P_{\vartheta_0}[T>c] \ge P_{\vartheta_0}[T>t] = \alpha$. Nach der Definition von γ folgt nun $\gamma=0$. []

Für den Fall, daß $P_{\vartheta_0}T^{-1}$ nicht vertafelt ist, die Werte für die zugehörige Vf oder die zugehörigen Fraktile also nicht einer Tabelle entnommen werden können, kann in manchen Fällen eine Transformation auf eine vertafelte Verteilung vorgenommen werden. Dies ist der Fall, wenn eine Funktion g existiert mit $T = g \circ \tilde{T}$ und $P_{\vartheta_0}\tilde{T}^{-1}$ vertafelt ist.Für diese Situation ist das folgende Lemma nützlich, dessen Beweis eine Übungsaufgabe ist.

9.3 Lemma. Sei $\tilde{T}$ eine reelle Zva auf $(\Omega,\mathfrak{F},P)$ und $T = g \circ \tilde{T}$ mit einer isotonen, linksstetigen Funktion $g : \mathbb{R} \mapsto \mathbb{R}$ (oder $g : \mathbb{R} \mapsto \bar{\mathbb{R}}$, aber $g \circ T$ f.s. reell).

Ist dann τ_α das α–Fraktil von $P\tilde{T}^{-1}$, so ist $g(\tau_\alpha)$ das α–Fraktil von PT^{-1}.

Nun können Beweise zu Sätzen aus §4 gegeben werden.

Beispiel. Einseitiger Binomialtest. Beweis von 4.6.

Zu $\mathfrak{P} = \{ \bigotimes_{i=1}^{n} b(1,\vartheta) ; 0<\vartheta<1 \}$ betrachten wir das einseitige Testproblem

$$H: \vartheta \le \vartheta_o \quad , \quad K: \vartheta>\vartheta_o \, .$$

$\mathfrak{P}$ ist nach Beispiel 1 in §7 und 7.2 eine einparametrige Exponentialfamilie in

$$\zeta(\vartheta) := \ln(\frac{\vartheta}{1-\vartheta}) \text{ und } T := \sum_{i=1}^{n} X_i .$$

Dabei ist ζ isoton, also hat $\mathfrak{P}$ gemäß 7.4 monotone DQ in T. Es gilt: $P_{\vartheta_o} T^{-1} = b(n,\vartheta_o)$.

Nun liefert Satz 9.1 direkt den Satz 4.6. $[\,]$

Beispiel. Einseitiger Gauß–Test. Beweis von 4.7.

Zu $\mathfrak{P} = \{ \bigotimes_{i=1}^{n} N(a.\sigma_o^2) ; a \in \mathbb{R}\}$ betrachten wir das einseitige Testproblem

$$H: a \le a_o \, , \quad K: a > a_o .$$

$\mathfrak{P}$ ist nach Beispiel 2 in §7 und 7.2 eine einparametrige Exponentialfamilie in $\zeta(a) = a/\sigma_o^2$ und $T = \sum_{i=1}^{n} X_i$ und somit auch in $\vartheta = \vartheta(a) = \sqrt{n}\, a/\sigma_o$ und $V = \sqrt{n}\, \frac{\bar{X} - a_0}{\sigma_o}$ mit $\bar{X} = \frac{1}{n} T$

wegen $\zeta \cdot T = \vartheta \cdot V + n\, a \cdot a_o/\sigma_o^2$. Wir haben also die folgende Situation mit $\vartheta_o = \sqrt{n}\, a_o/\sigma_o$:

(9.4) $$dP_\vartheta = C(\vartheta) \exp\{ \vartheta \cdot V\} \cdot h \, d\lambda^n \, ,$$

$$h = \exp\{ -\sum_{i=1}^{n} X_i^2 / 2\sigma_o^2 \} \quad \text{und} \quad C(\vartheta) := (2\pi\sigma_o^2)^{-n/2} \cdot \exp\{-\tfrac{1}{2}\vartheta^2 + \vartheta_o \cdot \vartheta\}$$

Dabei ist ϑ isoton, also hat $\mathfrak{P}$ gemäß 7.4 monotone DQ in V und $P_{a_o} \tilde{T}^{-1} = N(0,1)$ ist stetig. Hier konnte die Transformation auf eine Prüfgröße mit vertafelter Verteilung bereits vor Aufstellung des Tests vorgenommen werden. Nun liefern 9.1 und 9.2 direkt den Satz 4.7. $[\,]$

§ 10 Invarianz.

Bei der sogenannten Reduktion durch Invarianz wird eine Datenreduktion vorgenommen mit dem Bestreben, nicht mehr in die Daten hinein zu interpretieren, als diese hergeben.
Bei vielen statistischen Experimenten hängt das Beobachtungsmaterial von der willkürlichen Wahl des Maßsystems ab. So kann z.B. die Temperatur in Celsius, Fahrenheit, Kelvin und Reaumur gemessen werden. Die zugehörigen Meßskalen können dabei durch affine Transformationen ineinander überführt werden. Die Meßskala kann also willkürlich gewählt werden; es existiert eine Klasse von Maßsystemen , die alle gleichermaßen sinnvoll sind. Auch kann etwa die volle Struktur des $\mathbb{R}$ als zu reich empfunden werden. So ist etwa die Aussage "doppelt so warm" nicht sinnvoll.
Eine naheliegende Forderung wäre dann, daß die zu treffende Entscheidung unabhängig vom gewählten Maßsystem sein soll. Dies ist so zu verstehen, daß etwa beim Testen von Temperaturen die gemessenen Werte ohne Angabe der verwendeten Temperaturskala verwertet und in den Test eingegeben werden können. Eine Einschränkung auf in diesem Sinne invariante Tests ist aber nur dann sinnvoll, wenn auch das Testproblem in geeigneter Weise invariant ist.

Beispiel 1.
Liegen als Beobachtungsmaterial gewisse Temperaturwerte vor, so ist zu überlegen, ob Entscheidungen invariant gegenüber affinen Transformationen $\tau(x) = c \cdot x + d$, $c>0$, sein sollten. Das Einstichprobenproblem, etwa die Messung der Temperatur an einem Ort,

$$H': a = a_o\,, \quad K': a \neq a_o \quad \text{mit} \quad a := E_\vartheta[X_j]$$

hängt über a_o von der speziellen Skala ab. Dieses Einstichprobenproblem ist demnach nicht invariant gegenüber den obigen affinen Transformationen.
Das Zweistichprobenproblem, etwa die Messung der Temperatur an zwei Orten,

$$H: a_1 \le a_2\,, \quad K: a_1 > a_2 \quad \text{mit} \quad a_i := E_\vartheta[X_{ij}],$$

ist dagegen invariant in folgendem Sinne: Wegen $E_\vartheta[\tau(X_{ij})] = \tau(E_\vartheta[X_{ij}])$ gilt nämlich:

$$\vartheta \in H \;\Leftrightarrow\; E_\vartheta[\,X_{1j}\,] \le E_\vartheta[\,X_{2j}\,] \;\Leftrightarrow\; E_\vartheta[\,\tau(X_{1j})\,] \le E_\vartheta[\,\tau(X_{2j})\,]\,.$$

Definiert man $\pi = \pi_\tau : S \mapsto S$ zu τ gemäß:

$$\pi((x_{11},...,x_{1n_1},x_{21},...,x_{2n_2})) = (\tau(x_{11}),...,\tau(x_{1n_1}),\tau(x_{21}),...,\tau(x_{2n_2}))\,,$$

und $\mathfrak{P}_H := \{P_\vartheta\,;\, \vartheta \in H\}$, so gilt also $\tau(X_{ij}) = X_{ij} \circ \pi$ und damit:

$$P_\vartheta \in \mathfrak{P}_H \;\Leftrightarrow\; \int X_{1j}\, dP_\vartheta \le \int X_{2j}\, dP_\vartheta \;\Leftrightarrow\; \int X_{1j}\, dP_\vartheta\pi^{-1} \le \int X_{2j}\, dP_\vartheta\pi^{-1}$$

Gilt also $P_\vartheta \in \mathfrak{P}_H$ und $P_\vartheta\pi^{-1} \in \mathfrak{P}$, so hat man: $P_\vartheta\pi^{-1} \in \mathfrak{P}_H$. $[\,]$

Um Transformationen auch auf Θ zu vermeiden, wird es hier günstiger sein, anstelle von H mit $\mathfrak{P}_H$ zu argumentieren. Für diesen § 10 gelten die folgenden

Generalvoraussetzungen:

(1) $(S,\mathfrak{S},\mathfrak{P})$ sei ein statistischer Raum.

(2) Q sei eine Gruppe von mb Transformationen $\pi : S \mapsto S$; es gilt also:

(i) $\pi_1, \pi_2 \in Q \Rightarrow \pi_1 \circ \pi_2 \in Q$

(ii) $id_S \in Q$

(iii) $\forall\, \pi \in Q \;\exists\, \check{\pi} \in Q$ mit $\pi \circ \check{\pi} = id_S = \check{\pi} \circ \pi$.

(3) Es sei $\mathfrak{P} = \{P_\vartheta, \vartheta \in \Theta\}$ und $\Theta = H \cup K$ ein Testproblem. Dann setzen wir $\mathfrak{P}_H := \{P_\vartheta ; \vartheta \in H\}$, $\mathfrak{P}_K := \{P_\vartheta ; \vartheta \in K\}$, also: $\mathfrak{P} = \mathfrak{P}_H \cup \mathfrak{P}_K$.

Definition 1.

(a) Das Testproblem H gegen K heißt **invariant** (gegenüber Q), wenn gilt:

(I) $P\pi^{-1} \in \mathfrak{P} \quad \forall\, \pi \in Q, P \in \mathfrak{P}$;

(I_H) $P\pi^{-1} \in \mathfrak{P}_H \quad \forall\, \pi \in Q, P \in \mathfrak{P}_H$;

Diese beiden Eigenschaften implizieren bereits [vgl. 10.1]:

(I_K) $P\pi^{-1} \in \mathfrak{P}_K \quad \forall\, \pi \in Q, P \in \mathfrak{P}_K$.

(b) Ein Test φ heißt **invariant** (gegenüber Q), wenn gilt:

$\varphi \circ \pi = \varphi \quad \forall\, \pi \in Q \;\Leftrightarrow: \varphi \in \Phi_I$.

Wir setzen: $\Phi_{I,\alpha} := \Phi_I \cap \Phi_\alpha$.

10.1 Lemma. Folgende Aussagen sind äquivalent:

(1) Das Testproblem H gegen K ist invariant, d.h. (I) und (I_H) gelten.

(2) (I_H) und (I_K) gelten.

(3) $\{P\pi^{-1}, P \in \mathfrak{P}_H\} = \mathfrak{P}_H$, $\{P\pi^{-1}, P \in \mathfrak{P}_K\} = \mathfrak{P}_K \quad \forall\, \pi \in Q$.

Beweis: Aus (I_H) folgt für $\pi \in Q$ und damit $\check{\pi} \in Q$: $P \in \mathfrak{P}_H \Rightarrow P\check{\pi}^{-1} \in \mathfrak{P}_H \Rightarrow$

$P\,[= P(\pi\circ\check{\pi})^{-1}] = (P\check{\pi}^{-1})\pi^{-1} \in \mathfrak{P}_H$, somit "$\mathfrak{P}_H \subset \{P\pi^{-1}, P \in \mathfrak{P}_H\}$" also "$\{P\pi^{-1}, P \in \mathfrak{P}_H\} = \mathfrak{P}_H$" und in Verbindung mit (I) "$\{P\pi^{-1}, P \in \mathfrak{P}_K\} \subset \mathfrak{P}_K$" , also ($I_K$).

Die übrigen Aussagen folgen ebenso. []

Das nächste Ziel besteht im Nachweis der Existenz einer Statistik T mit:

(10.2) φ ist invariant $\Leftrightarrow \exists\, \psi$ mit $\varphi = \psi \circ T$.

Eine Beschränkung auf invariante Tests bedeutet dann gerade eine Beschränkung auf Tests, die nur über die Statistik T von der Stichprobe abhängen. Die historische Entwicklung verlief allerdings gerade umgekehrt. Man entwickelte eine naheliegende Prüfgröße T für ein Testproblem H: $\rho(\vartheta) \leq (=)\ \rho_o$, K: $\rho(\vartheta) > (\neq)\ \rho_o$, so daß die Verteilung von T und damit deren Fraktile für alle ϑ mit $\rho(\vartheta)=\rho_o$ übereinstimmen. Dann suchte man einen geigneten Test der Gestalt $\varphi = \psi \circ T$.

Hier soll jedoch die Beschränkung auf Test der Gestalt $\varphi = \psi \circ T$ durch eine Beschränkung auf invariante Test ersetzt werden. Wie (10.2) zeigen wird, sind beide Vorgehensweisen gleichwertig. Eine Beschränkung auf invariante Test läßt sich aber oft besser motivieren. Für die Implikation "$\Leftarrow$" in (10.2) ist offenbar die Eigenschaft $T \circ \pi = T \ \forall \ \pi \in Q$ hinreichend. Um auch die Implikation "$\Rightarrow$" zu erhalten, muß T so gewählt werden, daß T möglichst viel Information enthält.

Definition 2. (a) Eine Abbildung $T : S \mapsto D$ heißt **invariant** (gegenüber Q), wenn gilt:

$T \circ \pi = T \ \forall \ \pi \in Q$.

(b) Eine Abbildung $T : S \mapsto D$ heißt **maximal invariant**, wenn für alle $x, x' \in S$ gilt:

$T(x) = T(x') \Leftrightarrow \exists \ \pi \in Q$ mit $x' = \pi(x)$. [Dabei bedeutet "$\Leftarrow$" gerade die Invarianz.]

Von T wird zunächst keine Meßbarkeit gefordert. Meßbarkeitsüberlegungen sollen zunächst zurückgestellt werden.

10.3 Bemerkung. Die Gruppe Q zerlegt S in Äquivalenzklassen $Q(x) := \{\pi(x); \ \pi \in Q\}$. Die Invarianz von T bedeutet dann gerade, daß T auf jeder Äquivalenzklasse Q(x) konstant ist. Bei der maximalen Invarianz wird noch zusätzlich gefordert, daß T auf verschiedenen Äquivalenzklassen verschiedene Werte annimmt.

Es existiert immer eine maximale invariante Abbildung.

Die Abbildung $T(x) := Q(x)$ von S in die Potenzmenge von S ist eine solche.

Die Invarianz von T folgt aus der Beziehung: $\{\pi \circ \pi'; \ \pi \in Q\} = Q \ \ \forall \ \pi' \in Q$. []

10.4 Lemma. Sei φ ein Test und $T : S \mapsto D$ eine maximal invariante Abbildung. Dann sind die folgenden Aussagen äquivalent:

(i) φ ist invariant;

(ii) Es existiert eine Abbildung $\psi : D \mapsto [0,1]$ mit: $\varphi = \psi \circ T$.

Beweis. Die Folgerung "(ii) $\Rightarrow$ (i)" ergibt sich sofort aus der Invarianz von T.

Zum Beweis von "(i) $\Rightarrow$ (ii)" wählen wir zu $t \in T(S)$ ein $x_t \in T^{-1}(\{t\}) = \{x', T(x')=t\}$ und zeigen:

$\{x'; T(x')=t\} = \{\pi(x_t); \ \pi \in Q\}$ $[= Q(x_t)$ gemäß 10.3$]$ $\forall \ t \in T(S)$.

Dabei folgt die Relation "$\supset$" aus der Invarianz von T und die Relation "$\subset$" aus der maximalen Invarianz. Also ist $\varphi(x) = \varphi(x_t) \ \forall \ x$ mit $T(x)=t$. Somit ist ψ gemäß

$\psi(t) = \varphi(x_t)$ für $t \in T(S)$ und etwa $\psi(t) = 0$ für $t \notin T(S)$,

wohl definiert, und es folgt für $x \in S$ und $t=T(x)$: $\varphi(x) = \varphi(x_t) = \psi(t) = \psi(T(x))$. []

Nun sollen die Meßbarkeitsfragen untersucht werden. Es interessiert die folgende Frage:

Kann die Abbildung ψ aus 10.4 als meßbare Abbildung, also als Test gewählt werden, wenn T als meßbare Abbildung, also als Statistik vorausgesetzt wird.

Definition 3. T^{-} : $T(S) \mapsto S$ ist eine **meßbare schwache Inverse** zur Statistik $T : (S,\mathfrak{S}) \mapsto (D,\mathfrak{D})$, wenn gilt: $T(S) \in \mathfrak{D}$; T^{-} ist eine $T(S)\cap\mathfrak{D}$–$\mathfrak{S}$–mb Abbildung mit:

$T\circ T^{-} = id_{T(S)}$, d.h. $T^{-}(t) \in T^{-1}(\{t\}) \quad \forall\, t\in T(S)$.

Definition 4. Wir definieren zu einer Statistik $T : (S,\mathfrak{S}) \mapsto (D,\mathfrak{D})$:

$\Psi := \{\psi : D \mapsto [0,1]\,;\ \psi \text{ ist } \mathfrak{D}\text{–mb}\}$.

10.5 Satz. Sei $T : (S,\mathfrak{S}) \mapsto (D,\mathfrak{D})$ eine maximal invariante Statistik, die eine mb schwache Inverse besitzt. Dann sind die folgenden Aussagen äquivalent:

(i) φ ist ein invarianter Test;

(ii) Es existiert ein $\psi \in \Psi$ mit $\varphi = \psi\circ T$.

Beweis. Die Folgerung "(ii) $\Rightarrow$ (i)" ist wieder klar. Zum Beweis von "(i) $\Rightarrow$ (ii)" sehen wir uns den Beweis von 10.4 an. Dort ist nun $x_t = T^{-}(t)$ wählbar, also:

$$\psi = \begin{cases} \varphi\circ T^{-} & \text{auf } T(S) \\ 0 & \text{auf } S \setminus T(S) \end{cases}.$$

Diese Abbildung ψ ist offensichtlich meßbar [vgl. W–th. I 8.12]. []

Satz 10.5 ermöglicht nun die sogenannte

Reduktion durch Invarianz.

Liegt ein invariantes Testproblem H gegen K vor und beschränkt man sich auf die Klasse Φ_I der invarianten Tests, so kann nun in der Situation von 10.5 das Testproblem von $(S,\mathfrak{S},\mathfrak{P})$ auf den statistischen Raum $(D,\mathfrak{D},\mathfrak{P}T^{-1})$ mit $\mathfrak{P}T^{-1} := \{ P_\vartheta T^{-1};\ \vartheta\in\Theta \}$ verlagert werden. Die Suche nach einem schärfsten Test in $\Phi_{I,\alpha}$ wird dabei zurückgeführt auf das bekannte Problem der Suche nach einem schärfsten Test in der Menge Ψ_α von Tests z.N. α auf dem Raum D.

In der Regel wird der Fall "$P_{\vartheta_1}T^{-1} = P_{\vartheta_2}T^{-1}$ für gewisse $\vartheta_1\neq\vartheta_2$" eintreten. Dies führt zu einer erwünschten Reduktion des Parameterraums; man kann dann für $\mathfrak{P}T^{-1}$ eine andere Parametrisierung durch $\vartheta'\in\Theta' = H'\cup K'$ wählen. [Theoretisch kann auch die Situation "$P_{\vartheta_1}T^{-1} = P_{\vartheta_2}T^{-1}$ für gewisse $\vartheta_1\in H$, $\vartheta_2\in K$" eintreten. Dann wären H' und K' nicht mehr disjunkt wählbar. Diese Möglichkeit kann beim Aufbau der Theorie durchaus zugelassen werden. Allerdings wird die Situation in den Anwendungen nicht auftreten.]

Definition 5. Zu einer Statistik $T : (S,\mathfrak{S}) \mapsto (D,\mathfrak{D})$ setzen wir:

$$\mathfrak{P}T^{-1} := \{ P_\vartheta T^{-1};\ \vartheta\in\Theta \} = \{ P_{\vartheta'}T^{-1};\ \vartheta'\in\Theta'\}$$

$$\{ P_\vartheta T^{-1};\ \vartheta\in H \} = \{ P_{\vartheta'}T^{-1};\ \vartheta'\in H'\} \quad \text{, entprechend für K und K'}$$

$$\Psi_\alpha := \{ \psi\in\Psi\ ;\ \int \psi\circ T\, dP_\vartheta \le \alpha\ \ \forall\ \vartheta\in H \} = \{ \psi\in\Psi\ ;\ \int \psi\, dP_{\vartheta'}T^{-1} \le \alpha\ \ \forall\ \vartheta'\in H'\}.$$

10.6 Satz. Sei T eine maximal invariante Statistik mit einer mb schwachen Inversen.
Dann sind folgende Auusagen äquivalent:

(i) φ^* ist schärfster Test in $\Phi_{I,\alpha}$ für das Testen von H gegen K;

(ii) für das auf den statistischen Raum $(D,\mathfrak{D},\mathfrak{P}T^{-1})$ verlagerte Testproblem H' gegen K' existiert ein schärfster Test ψ^* in Ψ_α mit $\varphi^* = \psi^*\circ T$.

Beweis. Aus 10.5 und der Transformationsformel $\int \psi\circ T\, dP_\vartheta = \int \psi\, dP_\vartheta T^{-1}$ folgt:

(10.7) $\Phi_I = \{ \psi\circ T;\ \psi\in\Psi \}$, $\Phi_{I,\alpha} = \{ \psi\circ T;\ \Psi\in\Psi_\alpha \}$.

Mit der Transformationsformel folgt jetzt sofort die Behauptung. []

Im folgenden werden wir eine Verallgemeinerung der folgenden Situation behandeln:

$\mathfrak{P} = \{ P_\vartheta = \bigotimes_{i=1}^n N(a,\sigma_o^2)\ ;\ \vartheta := \sqrt{n}\cdot a/\sigma_o \in \mathbb{R} \}$, H: $a = 0$ ($\Leftrightarrow \vartheta = 0$), K: $a \neq 0$ ($\Leftrightarrow \vartheta \neq 0$),

Dann gilt also (vgl. (9.4)):

(10.8) $dP_\vartheta = C(\vartheta)\exp\{\vartheta\cdot V\}\cdot h\, d\lambda^n$ mit $V := \sqrt{n}\cdot\bar{X}/\sigma_o$,

$h = \exp\{-\sum_{i=1}^n X_i^2 / 2\sigma_o^2\}$ und $C(\vartheta) := (2\pi\sigma_o^2)^{-n/2}\cdot\exp\{-\frac{1}{2}\vartheta^2\}$.

10.9 Lemma. Sei $\mathfrak{P} = \{P_\vartheta;\ \vartheta\in\Theta\}$ eine 1–parametrige Exponentialfamilie auf $(S,\mathfrak{S}) := (\mathbb{R}^n,\mathfrak{B}_n)$ mit μ–Dichten $p_\vartheta = C(\vartheta)\exp\{\vartheta\cdot V\}\, h$. Betrachtet wird das zweiseitige Testproblem

H: $\vartheta = 0$, K: $\vartheta \neq 0$.

Es sei $Q = \{id_S, \pi\}$ mit $\pi(x) := -x$, $x\in S$. Ferner sei folgendes voraussgesetzt:

(1) μ sei das Lebesgue–Maß λ^n auf $\mathbb{R}^n$ oder das Zählmaß auf $\mathbb{Z}^n$;

(2) $V(-x) = -V(x)$, $x\in S$;

(3) $h(-x) = h(x)$, $x\in S$;

(4) $0\in\Theta$; $-\vartheta\in\Theta\ \ \forall\ \vartheta\in\Theta$.

Dann gilt:

(a) $P_\vartheta\pi^{-1} = P_{-\vartheta}$, $\vartheta\in\Theta$;

(b) das Testproblem H gegen K ist invariant gegenüber Q;

(c) $T(x) = \mathrm{sgn}(x)\cdot x$, $x\in S$, ist eine maximal invariante Statistik mit einer mb schwachen Inversen.

Beweis. a) Wegen (1) gilt $\mu\pi^{-1} = \mu$; mit dem Transformationssatz folgt:

$$P_\vartheta\pi^{-1}[B] = {}_{\pi^{-1}(B)}\int C(\vartheta)\exp\{\vartheta\cdot V\}\,h\,d\mu = {}_{\pi^{-1}(B)}\int C(\vartheta)\exp\{-\vartheta\cdot V\circ\pi\}\,h\circ\pi\,d\mu$$
$$= {}_B\int C(\vartheta)\exp\{-\vartheta\cdot V\}\,h\,d\mu = \frac{C(\vartheta)}{C(-\vartheta)}\,P_{-\vartheta}[B],$$

insbesondere mit $B = \mathbb{R}^n$: $1 = C(\vartheta) / C(-\vartheta)$, also

(10.9a') $C(-\vartheta) = C(\vartheta)$.

Damit ist $P_\vartheta\pi^{-1}[B] = P_{-\vartheta}[B]$ gezeigt.

b) folgt nun aus Teil a).

c) Die Invarianz von T folgt mit (7.10) aus:

$$T\circ\pi(x) = \mathrm{sgn}(-x)\cdot(-x) = -\mathrm{sgn}(x)\cdot(-x) = T(x).$$

Zum Nachweis der maximalen Invarianz sei nun $T(x) = T(x')$.

Im Fall 1 "x=0" gilt: $x = 0 = x' = id_S(x)$.

Im Fall 2 "$\mathrm{sgn}(x)\cdot\mathrm{sgn}(x') = 1$" gilt ebenfalls: $x = x' = id_S(x)$.

Im Fall 3 "$\mathrm{sgn}(x)\cdot\mathrm{sgn}(x') = -1$" gilt: $x = -x' = \pi(x)$.

Zur Konstruktion einer mb schwachen Inversen folgern wir zunächst aus (7.10):

$$T(S) = \{0\}\cup\{x;\ \mathrm{sgn}(x) = 1\} =: D.$$

Da $x \mapsto \mathrm{sgn}(x)$ mb ist, ist auch T(S) mb. Nun kann $T^- := id_{T(S)}$ gesetzt werden. []

10.10 Satz. Unter den Voraussetzungen von 10.9 ist folgender Test φ^* schärfster Test in $\Phi_{I,\alpha}$ für das Testen von H gegen K, dabei sei c das $\alpha/2$ – Fraktil von P_0V^{-1}:

$$\varphi^* := \begin{cases} 1 & |V| > c \\ \gamma \text{ für } & |V| = c \\ 0 & |V| < c \end{cases} \quad \text{mit } \gamma := \frac{\frac{1}{2}\alpha - P_0[V>c]}{P_0[V=c]} \quad (\text{mit } |V| = |V\circ T|).$$

Beweis. Sei $T(x) = \mathrm{sgn}(x)\cdot x$, $D = T(S)$. T hat gemäß 7.11 und den Voraussetzungen 10.9(2),(3) und (10.9a') die Dichte

$$p_\vartheta(t) + p_\vartheta(-t) = C(\vartheta)\cdot\left[e^{\vartheta\cdot V(t)} + e^{-\vartheta\cdot V(t)}\right]\cdot h(t)$$
$$= C(|\vartheta|)\cdot\left[e^{|\vartheta|\cdot|V(t)|} + e^{-|\vartheta|\cdot|V(t)|}\right]\cdot h(t).$$

Also gilt: $P_\vartheta T^{-1} = P_{|\vartheta|}T^{-1}$, und wir können folgende neue Indizierung vornehmen:

$\vartheta' = |\vartheta|$, $\Theta' = [0,\infty)\cap\Theta$, $H' = H = \{0\}$ und $K' = (0,\infty)\cap\Theta$.

Durch Wahl des dominierenden Maßes kann $h\equiv 1$ angenommen werden. Gemäß (7.8) hat nun $\mathfrak{P}T^{-1} = \{ P_\vartheta T^{-1}, \vartheta\ge 0 \}$ monotone DQ in $|V|$. Das Testproblem H' gegen K' ist einseitig. Gemäß 9.1 existiert ein schärfster Test in Ψ_α der Form:

$$\psi^*(t) := \begin{cases} 1 & > \\ \gamma \text{ für } & |V(t)| = c \\ 0 & < \end{cases}, \ t\in D\subset\mathbb{R}^n.$$

Dabei ist c das α–Fraktil von $(P_0T^{-1})|V|^{-1} = P_0|V\circ T|^{-1}$ und γ bestimmt sich aus:

$\alpha = \int \psi^* \, dP_0T^{-1} = E_0[\psi^*\circ T] = P_0[|V| > c] + \gamma \cdot P_0[|V| = c]$. Aus 10.9(2) und 10.9(a) folgt:

$|V\circ T| = |V|$

$P_0(-V)^{-1} = P_0(V\circ\pi)^{-1} = (P_0\pi^{-1})V^{-1} = P_0V^{-1}$.

Also hat man $P_0[|V| > (=)\ v] = 2 \cdot P_0[V > (=)\ v]$. Nun folgt leicht, daß das α–Fraktil von $P_0|V|^{-1}$ gleich dem $\alpha/2$–Fraktil von P_0V^{-1} ist und γ wie oben angegeben gewählt werden kann. Wegen $|V\circ T| = |V|$ hat $\varphi^* := \psi^*\circ T$ die obige Gestalt und ist nach 10.6 schärfster Test in $\Phi_{I,\alpha}$. []

10.11 Bemerkung. Will man Testprobleme der Form H: $\rho(\vartheta) = \rho_o$, K: $\rho(\vartheta) \neq \rho_o$ behandeln, so kann man sie oft mit Hilfe einer Transformation U auf den Fall $\rho_o=0$ zurückführen, z.B.
$U(x) = (x_1 - \rho_o,...,x_n - \rho_o)$. Dazu dient die folgende Überlegung:

Sei $(S,\mathfrak{S},\mathfrak{P})$ gegeben mit $\mathfrak{P} = \{P_\vartheta\ ,\ \vartheta\in\Theta\}$ und $\Theta=H\cup K$. Es sei $U : (S,\mathfrak{S}) \mapsto (\bar S,\bar{\mathfrak{S}})$ eine mb Abbildung mit einer mb Umkehrfunktion $\check U$. Wir nehmen nun an, daß das auf $(\bar S,\bar{\mathfrak{S}},\mathfrak{P}U^{-1})$ verlagerte Testproblem mit $\mathfrak{P}U^{-1} = \{P_\vartheta U^{-1},\ \vartheta\in\Theta=H\cup K\}$ im folgenden Sinne gelöst ist:

Das Testproblem auf $(\bar S,\bar{\mathfrak{S}},\mathfrak{P}U^{-1})$ sei invariant gegenüber einer Gruppe $\bar Q$ von mb Transformationen auf $\bar S$ und φ^* sei schärfster Test in $\Phi_{I,\alpha}$ bzgl. $\bar Q$.

Dann ist das Testproblem auf $(S,\mathfrak{S},\mathfrak{P})$ invariant gegenüber $Q := \check U\bar Q U := \{\check U\bar\pi U;\ \bar\pi\in\bar Q\}$, und $\varphi^*\circ U$ ist schärfster Test in $\Phi_{I,\alpha}$ bzgl. Q.
Der Beweis folgt sofort mit der Transformationsformel für Bildmaße. []

Beispiel 2. Zweiseitiger Gaußtest. Beweis von 4.13.
Seien $X_1,...,X_n$ iid Zva mit X_i~$N(a,\sigma_o^2)$. Für das Testproblem H: $a=a_o$, K: $a\neq a_o$ kann o.E. $a_o=0$ angenommen werden. Andernfalls betrachtet man die Zva $\tilde X_i := X_i - a_o$ und verwendet 10.11 mit $U(x) = (x_1-a_o,...,x_n-a_o)$ und $\bar Q$ wie in 10.9.

$\mathfrak{P} = \{ P_\vartheta = \bigotimes_{i=1}^n N(a,\sigma_o^2)\ ;\ \vartheta := \sqrt{n}\cdot a/\sigma_o \in \mathbb{R}\ \}$ ist gemäß (10.8) eine 1–parametrige Exponentialfamilie und die Voraussetzungen von 10.9 sind erfüllt. Außerdem gilt:

$P_0V^{-1} = N(0,1)$ für $V = \sqrt{n}\cdot\bar X/\sigma_o$.

Aus 10.10 folgt jetzt unmittelbar die Aussage von 4.13. Dabei ist $\gamma=0$ wegen 8.4(iv). []

10.12 Satz. Binomialtest im zweiseitigen symmetrischen Fall.

Es liege das Einstichprobenproblem $S = \{0,1\}^n$, $\mathfrak{P} = \{ \bigotimes_{i=1}^{n} b(1,\vartheta),\ 0<\vartheta<1 \}$ vor mit $T := \sum_1^n X_i$; sei $0<\alpha<1$. Für das zweiseitige Testproblem:

$H: \vartheta = \frac{1}{2}$, $K: \vartheta \neq \frac{1}{2}$.

verwendet man den Test

$$\varphi^* := \begin{cases} 1 & \text{sonst} \\ \gamma & \text{für } T = c \text{ oder } T = n-c \\ 0 & n-c < T < c \end{cases} = \begin{cases} 1 & > \\ \gamma & \text{für } \left|T - \frac{n}{2}\right| = c - \frac{n}{2} \\ 0 & < \end{cases}.$$

Dabei ist $c := b(n,\frac{1}{2})_{\alpha/2}$ und γ durch $P_{\frac{1}{2}}[T > c] + \gamma P_{\frac{1}{2}}[T = c] = \alpha/2$ bestimmt.

Dieser Test ist schärfster Test in $\Phi_{I,\alpha} = \{\varphi \in \Phi_\alpha;\ \varphi \circ \pi = \varphi\}$ mit

$\pi(x_1,...,x_n) = (1-x_1,...,1-x_n)$ und damit unverfälscht.

Die Transformation π aus 10.12 ersetzt in x Erfolge durch Mißerfolge und umgekehrt. Eine Beschränkung auf gegenüber π invariante Tests bedeutet, daß ein Überwiegen von Mißerfolgen in gleicher Weise zur Ablehnung von H führen sollte wie ein Überwiegen von Erfolgen.

Der **Beweis** ist eine Übungsaufgabe. []

Beispiel 3. Binomial–Test im zweiseitigen asymmetrischen Fall.

Sei S, $\mathfrak{P}$, T, α wie in 10.12. Dazu betrachten wir das Testproblem

$H: \vartheta = \vartheta_o$, $K: \vartheta \neq \vartheta_o$.

Es existiert ein schärfster Test φ^* in $\tilde{\Phi}_\alpha$ der Form $\varphi^* = \varphi_2$ mit

$$\varphi_2 = \begin{cases} 1 & T \notin [c_1, c_2] \\ \gamma_i & \text{für } T = c_1 \text{ oder } c_2 \\ 0 & T \in (c_1, c_2) \end{cases}.$$

Für die Werte c_i,γ_i können dabei lediglich Bestimmungsgleichungen angegeben werden [vgl. Witting (66) S.102, Witting (85) S. 263]. Man erhält zumindest einen Test z.N α von der Form φ_2, wenn man die c_i, γ_i aus den folgenden Gleichungen bestimmt:

(i) $P_{\vartheta_o}[T < c_1] + \gamma_1 P[T = c_1] = \alpha/2 \Leftrightarrow P_{\vartheta_o}[n - T > n - c_1] + \gamma_1 P[n - T = n - c_1] = \alpha/2$

(ii) $P_{\vartheta_o}[T > c_2] + \gamma_2 P[T = c_2] = \alpha/2$.

Als Lösungen für c_i erhält man

$$c_2 = b(n,\vartheta_o)_{\alpha/2}\ ,\ c_1 = n - b(n,1-\vartheta_o)_{\alpha/2}$$

wegen $P_{\vartheta_o}(n-T)^{-1} = b(n,1-\vartheta_o)$.

Im symmetrischen Fall $\vartheta_o = \frac{1}{2}$ stimmt der zugehörige Test mit dem aus 10.12 überein. []

Bei der Bestimmung eines schärfsten invarianten Test ist es oft angenehm, wenn man eine Nullmenge abspalten kann. Dabei muß eine Verträglichkeit mit Q gewährleistet sein.

10.13 Lemma. Sei $N \in \mathfrak{S}$ mit $P_\vartheta[N] = 0 \ \forall \vartheta\in\Theta$ und $\pi(N) \subset N \ \forall \pi\in Q$. Setzt man nun $S' := S\setminus N$, $P'_\vartheta := P_\vartheta\big|_{S'\cap\mathfrak{S}}$, $\pi' := \pi\big|_{S'}$, $Q' := \{\pi';\ \pi\in Q\}$ und $\Phi'_{I,\alpha} := \{\varphi'\colon S' \mapsto [0,1];\ \varphi' \text{ mb}, \varphi'\circ\pi' = \varphi' \ \forall \pi'\in Q', \int \varphi'\, dP'_\vartheta \le \alpha \ \forall \vartheta\in H\}$, so gilt:

(a) Q' ist eine Gruppe von mb Transformationen auf S'.

(b) Ist das Testproblem auf S invariant gegenüber Q, so ist das auf S' verlagerte Testproblem invariant gegenüber Q'.

(c) Ist φ'^* schärfster Test in $\Phi'_{I,\alpha}$, so ist folgende Fortsetzung auf S schärfster Test in $\Phi_{I,\alpha}$:

$$\varphi^* := \begin{cases} \varphi'^* & \text{auf } S' \\ 0 & \text{auf } N \end{cases}.$$

Beweis: a) Es gilt: $\pi(N) = N$ und $\pi(S') = S'$ wegen $\pi(N) \subset N$.

b) Es gilt offensichtlich: $P'_\vartheta\pi'^{-1} = (P_\vartheta\pi^{-1})'$.

c) (i): Es gilt: $\varphi^* \in \Phi_I$; es genügt $x\in N$ zu betrachten:

$x \in N \Rightarrow \pi(x) \in N \Rightarrow \varphi^*(\pi(x)) = 0 = \varphi^*(x)$.

(ii) Es gilt: $\int \varphi^*\, dP_\vartheta = \int \varphi'^*\, dP'_\vartheta \ \forall \vartheta\in\Theta$, insbesondere $\varphi^* \in \Phi_\alpha$.

(iii) Für $\varphi \in \Phi_{I,\alpha}$ ist offenbar $\varphi' := \varphi\big|_{S'} \in \Phi'_I$.

Aus $\int \varphi\, dP_\vartheta = \int \varphi'\, dP'_\vartheta \ \forall \vartheta\in\Theta$, folgt nun sofort: $\varphi' \in \Phi'_{I,\alpha}$ und

$\int \varphi\, dP_\vartheta \le \int \varphi^*\, dP_\vartheta \ \forall \vartheta\in K$. [ ]

§ 11 (ausführliche Variante) **Invariante Tests in linearen Modellen.**

11 A Das Testproblem.

Es sei $(S,\mathfrak{S},\mathfrak{P})$ ein lineares Modell mit Normalverteilungsannahme und unbekannter Varianz gemäß Bemerkung 3.3; es liege also folgende Situation vor:

Generalvoraussetzung: $(S,\mathfrak{S}) = (\mathbb{R}^n,\mathfrak{B}_n)$, L_k ein k–dim. linearer Teilraum von $\mathbb{R}^n$ mit k<n,

$$\mathfrak{P} = \{ P_\vartheta = \bigotimes_{i=1}^{n} N(a_i,\sigma^2) \,;\, \vartheta=(a,\sigma^2) \in \Theta = L_k\times(0.\infty) \text{ mit } a = (a_1,...,a_n)^\top \}.$$

Definition 1. H heißt **lineare Hypothese**, falls es einen h–dim. linearen Teilraum $L_h \subset L_k$, $0\le h<k$ gibt mit $H = L_h\times(0,\infty)$.

Im weiteren Verlauf von § 11 gelte ebenfalls folgende

Generalvoraussetzung: Es sei h+1=k. Sei $b_1,...,b_k$ eine Basis des L_k, also

$L_k = \mathfrak{L}(b_1,...,b_h,b_k)$, mit $L_h = \mathfrak{L}(b_1,...,b_h)$.

Die Parameter $\alpha_j(\vartheta)$, $1\le j\le k$, seien definiert gemäß $a = \sum_1^k \alpha_j\, b_j$.

$\hat{X}_h$ und $\hat{X}_k = \hat{a}$ seien die Projektionen auf L_h bzw. L_k und $\hat{X}_k =: \sum_1^k \hat{\alpha}_j\, b_j$.

$L_h^\perp$ und $L_k^\perp$ seien das orthogonale Komplement zu L_h bzw. L_k.

Gemäß 3.4 ist $\hat{a} = \hat{X}_k$ KQS für a unter der Verteilungsannahme und $\hat{X}_h$ KQS für a im Fall $\Theta = H$.

Nach 3.9(c) ist $\hat{\alpha}_k$ GMS für den Parameter α_k, insbesondere linear und erwartungstreu.

$X - \hat{X}_h$ und $X - \hat{X}_k$ sind die Projektionen auf $L_h^\perp$ bzw. $L_k^\perp$. Dann gilt für $x\in S$:

(11.0) $\quad x = x_h + \hat{\alpha}_k(x)\, b_k + x_{n-k}$ mit $x_h := \sum_1^h \hat{\alpha}_j(x)\, b_j \in L_h$, $x_{n-k} := (X - \hat{X}_k)(x) \in L_k^\perp$.

Wir betrachten das Testproblem

(11.1) $\quad H: a \in L_h \ [\Leftrightarrow \alpha_k = 0] \qquad K: a \in L_k \setminus L_h \ [\Leftrightarrow \alpha_k \ne 0]$.

Auch das einseitige Testproblem $H: \alpha_k \le 0$ $K: \alpha_k > 0$ kann mit den hier verwendeten Methoden behandelt werden. Dies soll aber nicht geschehen. Zunächst zwei Beispiele, zu denen die Lösungen später im Abschnitt "Anwendungen" gegeben werden.

Beispiel 1. Das Testen von Mittelwerten im Einstichprobenfall.

Sei $\mathfrak{P} := \{ \bigotimes_{i=1}^{n} N(\mu,\sigma^2) \,;\, \mu\in\mathbb{R},\, \sigma^2>0\}$ und $e = (1,...,1)^\top$. Dann gilt offenbar

$a = \mu\, e$ und somit $L_k = \mathfrak{L}(e)$, k=1, $b_k = e$, $\alpha_k = \mu$, h=0. $L_h = \{0\}$.

Damit ergibt (11.1) folgendes Testproblem:

$H: \mu = 0 \qquad K: \mu \ne 0$.

Der Fall, wo in dem Testproblem μ_o $[\in \mathbb{R}]$ anstelle von 0 steht, läßt sich mit 10.11 auf den obigen Fall zurückführen. $[]$

Beispiel 2. Das Testen von Mittelwerten im 2–Stichprobenfall.

Sei $\mathfrak{P} := \{ \bigotimes_{i=1}^{n_1} N(\mu_1,\sigma^2) \times \bigotimes_{i=1}^{n_2} N(\mu_2,\sigma^2);\ \mu_1, \mu_2 \in \mathbb{R},\ \sigma^2>0 \}$, $n := n_1 + n_2$,

$e_1 := (1,...,1,0,...,0)^\top \in \mathbb{R}^n$, $e_2 := (0,...,0,1,...,1)^\top \in \mathbb{R}^n$, $e := e_1 + e_2$.

Dann ergibt sich:

$a = \mu_1 e_1 + \mu_2 e_2$, also $L_k = \mathfrak{L}(e_1,e_2)$.

Es sei $L_h = \mathfrak{L}(e)$, also h=1, dann gilt:

$a = \mu_2 e + (\mu_1-\mu_2) e_1 =: \alpha_1 b_1 + \alpha_2 b_2$.

Es ergibt sich also aus (11.1) folgendes Testproblem:

$H: \mu_1 = \mu_2$, $K: \mu_1 \neq \mu_2$. []

11 B Invarianz des Testproblems.

Es sei

(10.2) $Q := \{ \pi(x) = c\,x + d \ ;\ c\neq 0,\ d \in L_h \}$.

11.3 Lemma. (a) Q ist eine Gruppe von mb Transformationen.

(b) Das Testproblem H gegen K ist invariant gegenüber Q.

Beweis: Teil a) ist offensichtlich.

b) Sind X_i unabhängig und $N(a_i,\sigma^2)$–verteilt, $1\le i\le n$, so sind $c\cdot X_i + d_i$ wieder unabhängig und $N(c\cdot a_i+d_i,c^2\sigma^2)$–verteilt. Also gilt: $P_\vartheta\pi^{-1} = P_{(\pi(a),c^2\sigma^2)}$ für $\vartheta = (a,\sigma^2)$ mit $\pi(a) = c\cdot a+d$. Wegen "$\pi(L_k) \subset L_k$" und $\pi(L_h) \subset L_h$" folgt die Behauptung. []

11 C Abspaltung einer Nullmenge.

Es sei (unter Berücksichtigung von (11.0))

$N := L_h \oplus L_k^\perp = \{ y := y_h + y_{n-k} \ ;\ y_h \in L_h,\ y_{n-k} \in L_k^\perp \} = \{\hat\alpha_k = 0\}$.

Dann ist N also ein (n–1)–dim. Teilraum von $\mathbb{R}^n$ und hat somit das Lebesgue–Maß 0. Daher gilt auch:

$P_\vartheta[N] = 0 \ \forall\, \vartheta\in\Theta$.

Es soll nun $\pi(N) \subset N$ gezeigt werden für $\pi \in Q$. Sei dazu $y := y_h + y_{n-k} \in N$.

Die Behauptung folgt nun aus:

$\pi(y_h + y_{n-k}) = c\,y_h + d + c\,y_{n-k}$ mit $c\,y_h + d \in L_h$, $c\,y_{n-k} \in L_k^\perp$.

Wie in 10.13 sei jetzt

$S' = S \setminus N$, $Q' = \{ \pi' = \pi\big|_{S'} \ ;\ \pi\in Q \}$.

11 D Reduktion durch Invarianz

Es soll nun eine maximal invariante Statistik gesucht werden. Nun bedeutet $T(x+d) = T(x)$ $\forall\, d \in L_h$, daß T von x (in Hinblick auf (11.0)) nicht über $\hat{\alpha}_1(x),...,\hat{\alpha}_h(x)$ abhängen kann. Denn für $x_h, x_h' \in L_h$ gilt: $x_h' = x_h + d$ mit $d = x_h' - x_h \in L_h$. Also kann T nur von $\hat{\alpha}_k$ und der Projektion $X - \hat{X}_k$ auf $L_k^{\perp}$ abhängen. Eine gegenüber $x \mapsto c\,x$ invariante Statistik ist nun

(11.4) $\qquad T := 1/\hat{\alpha}_k\,(X - \hat{X}_k)$.

Daß $\hat{\alpha}_k$ als Schätzer für den zu testenden Parameter α_k bei T im Nenner stehen muß, liegt daran, daß man den Vektor $(X - \hat{X}_k)$ nicht in den Nenner nehmen kann. Die endgültige Prüfgröße wird aber wieder die gewohnte Gestalt haben: Im Zähler steht $\hat{\alpha}_k$ als Schätzer für den zu testenden Parameter α_k; im Nenner steht ein Schätzer für die Varianz; dazu kommt noch ein deterministischer Normierungsfaktor.

11.5 Lemma. T ist maximal invariant auf S' gegenüber Q' und hat eine mb schwache Inverse.

Beweis: (i) Zum Nachweis der Invarianz von T sei $\pi(x) = c\,x + d$ mit $c \neq 0$, $d \in L_h$.

Dann folgt aus der Linearität von $\hat{\alpha}_k$ und $\hat{X}_k$ die Invarianz mit:

$\hat{\alpha}_k(cx+d) = c\,\hat{\alpha}_k(x) + \hat{\alpha}_k(d) = c\,\hat{\alpha}_k(x)$; $(X - \hat{X}_k)(cx+d) = c\,(X - \hat{X}_k)(x)$.

(ii) Zum Nachweis der maximalen Invarianz sei $T(x) = T(x')$, d.h.

$(X - \hat{X}_k)(x') = \hat{\alpha}_k(x')/\hat{\alpha}_k(x)\,(X - \hat{X}_k)(x')$, $x, x' \in S'$. Dann gilt mit $c := \hat{\alpha}_k(x')/\hat{\alpha}_k(x) \neq 0$:

$x' - \hat{X}_k(x') = c\,(x - \hat{X}_k(x))$, $\hat{\alpha}_k(x') = c \cdot \hat{\alpha}_k(x)$. Es folgt:

$x' = x_h' + \hat{\alpha}_k(x')\,b_k + x' - \hat{X}_k(x') = x_h' - c\,x_h + c\,x_h + c\,\hat{\alpha}_k(x)\,b_k + c\,(x - \hat{X}_k(x))$

$x_h' - c\,x_h + c\,x = c\,x + d$ mit $d := x_h' - c\,x_h \in L_h$.

(iii) Zur Bestimmung einer mb schwachen Inversen überlegt man sich zunächst:

(11.6) $\qquad T(S') \subset L_k^{\perp}$.

Definiert man nun $T^{-} : L_k^{\perp} \mapsto \mathbb{R}^n$ gemäß $T^{-}(y) := b_k + y$, $y \in L_k^{\perp}$, so erhält man: $\hat{\alpha}_k(b_k+y) = 1$ und $(X - \hat{X}_k)(b_k+y) = y$, also $T(b_k+y) = y$. Damit gilt auch "=" in (11.6) und T^{-} ist als meßbare Inverse nachgewiesen. □

11 E Übergang zu einem anderen Koordinatensystem

Zur Berechnung der Verteilung der maximal invarianten Statistiken ist es günstig, eine Koordinatentransformation zu vorzunehmen.

Es sei $e_1,...,e_n$ eine Orthonormalbasis des $\mathbb{R}^n$ mit $L_i = \mathfrak{L}(e_1,...,e_i)$, i = h,k.

Dabei kann $e_k \in L_h^{\perp} \cap L_k$ o.E. parallel zum Anteil von b_k im $L_h^{\perp} \cap L_k$ gewählt werden gemäß

$$b_k - \hat{X}_h(b_k) \in L_h^{\perp} \cap L_k \ , \ \beta_k := \|b_k - \hat{X}_h(b_k)\| \neq 0 \ , \ e_k := \frac{1}{\beta_k}\,(b_k - \hat{X}_h(b_k)).$$

Sei nun A die orthogonale Matrix

$$A := \begin{pmatrix} e_1^{\top} \\ \vdots \\ e_n^{\top} \end{pmatrix} \quad \text{und} \quad Y := A\,X \ , \text{ d.h. } \ Y_j := e_j^{\top} \cdot X \ , \ 1 \le j \le n, \text{ (Anteil von X in Richtung } e_j).$$

Dann gilt: $X = A^{\top}Y = \sum_1^n Y_j\, e_j$ und $\hat{X}_k = \sum_1^k Y_j\, e_j = \hat{X}_h + Y_k\, e_k$, $X - \hat{X}_k = \sum_{k+1}^n Y_j\, e_j$,

insbesondere $a = \hat{X}_k(a) = \sum_1^k e_j^{\top} \cdot a\, e_j$. Andererseits haben wir:

$$\hat{X}_k = \sum_1^k \hat{\alpha}_j\, b_j \ = \sum_1^h \hat{\alpha}_j\, b_j + \hat{\alpha}_k\, \hat{X}_h(b_k) + \hat{\alpha}_k \cdot (b_k - \hat{X}_h(b_k)) \ = \ ... + \hat{\alpha}_k \cdot \beta_k\, e_k \ ,$$

$$a = \sum_1^k \alpha_j\, b_j \ = \sum_1^h \alpha_j\, b_j + \alpha_k\, \hat{X}_h(b_k) + \alpha_k \cdot (b_k - \hat{X}_h(b_k)) \ = \ ... + \alpha_k \cdot \beta_k\, e_k$$

(Die zweite Gleichung folgt aus der ersten auch aus $E_{\vartheta}[\hat{\alpha}_k] = \alpha_k(\vartheta)$). Damit ergibt sich:

(11.7) $\quad Y_k = \beta_k \cdot \hat{\alpha}_k$ und $e_k^{\top} \cdot a = \beta_k \cdot \alpha_k$ mit $\beta_k = \|b_k - \hat{X}_h(b_k)\|$.

Aus 5.1c folgt:

(11.8) $\quad$ Unter $\vartheta = (a,\sigma^2)$ gilt: $Y_1,...,Y_n$ unabh., $Y_j \sim N(e_j^{\top} \cdot a\ ,\ \sigma^2)$ mit

(i) $\quad e_j^{\top} \cdot a = 0 \ , \ j > k;$

(ii) $\quad e_k^{\top} \cdot a = \beta_k\, \alpha_k \ .$

Für die in Abschnitt D definierte maximal invariante Statistik erhält man somit:

(11.9) $\quad T = (\beta_k / Y_k) \sum_{k+1}^n Y_j\, e_j \ = (\beta_k / (\frac{Y_k}{\sigma})) \sum_{k+1}^n \frac{Y_j}{\sigma}\, e_j.$

11 F Reduktion durch Suffizienz.

Die Reduktion durch Invarianz gestattet nun eine Beschränkung auf Tests, die nur von T abhängen. In einer weiteren Reduktion soll eine Beschränkung auf reellwertige Statistiken gerechtfertigt werden. Als Motivation für das folgende Lemma bemerken wir:

$\frac{Y_j}{\sigma}$, $k < j \le n$, sind iid mit $\frac{Y_j}{\sigma} \sim N(0,1)$ unter $\vartheta = (a,\sigma^2)$.

11.10 Seien $Y_1,...,Y_n$ iid mit $Y_j \sim N(0,1)$, $Z := \sqrt{\sum_1^n Y_j^2}$, $W := \frac{1}{Z}\,(Y_1,...,Y_n)$; dann sind W und Z unabhängig. [Dabei ist die Verteilung von W die Gleichverteilung $U(S^{n-1})$ auf der Sphäre S^{n-1}, und $Z^2 \sim \chi_n^2$.]

Beweis: Sei $\tilde{p}(t) := (2\pi)^{-n/2} \exp\{-\frac{1}{2}t\}$, $t \in \mathbb{R}$.

Dann ist $y = (y_1,...,y_n) \mapsto \tilde{p}(\sum_1^n y_j^2)$ eine Dichte von $Y := (Y_1,...,Y_n)$ und es gilt:

$P[\, W \in B, Z \in D] = E[\mathbf{1}_B(\frac{1}{\|Y\|}\, Y) \cdot \mathbf{1}_D(\|Z\|)] = \int ... \int \mathbf{1}_B(\frac{1}{\|y\|}\, y) \cdot \mathbf{1}_D(\|y\|)\, \tilde{p}(\|y\|^2)\, dy$.

Wir gehen nun zu Polarkoordinaten über [Barner/Flohr (1983) S. 43] und setzen n>1 voraus. [Der Fall n=1 kann ohne Polarkoordinaten gelöst werden und ist eine Übungsaufgabe.]

Es ist $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = r \cdot f(\alpha,\vartheta_3,...,\vartheta_n)$ mit $r = \|y\| > 0$, $0 \le \alpha < 2\pi$, $-\frac{1}{2}\pi < \vartheta_j < \frac{1}{2}\pi$.

Für die Funktionaldeterminante Γ ergibt sich

$$\Gamma = \left| \partial_r(r\cdot f), \partial_\alpha(r\cdot f), \dots, \partial_{\vartheta_n}(r\cdot f) \right| = \left| f, r\cdot\partial_\alpha f, \dots, r\cdot\partial_{\vartheta_n}(r\cdot f) \right| .$$

Eine Entwicklung nach der ersten Spalte zeigt:

$\Gamma = r^{n-1} g(\alpha,\vartheta_3,...)$ für eine gewisse Funktion g.

(Es gilt sogar $g(\alpha,\vartheta_3,...) = \sin^{n-1}\alpha \cdot \sin^{n-2}\vartheta_3 \cdot \dots \cdot \sin\vartheta_n$ [vgl. Elstrodt(1996) S. 206]). Also:

$$P[W \in B, Z \in D] = \int\dots\int \mathbf{1}_B(f(\alpha,\vartheta_3,...,\vartheta_n)) \cdot \mathbf{1}_D(r)\ \tilde{p}(r^2)\ |\Gamma|\ dr\, d\alpha\, d\vartheta_3\dots$$

$$= \int\dots\int \mathbf{1}_B(f(\alpha,\vartheta_3,...,\vartheta_n)) \cdot \mathbf{1}_D(r)\ \tilde{p}(r^2)\, r^{n-1}\ |g(\alpha,...)|\ dr\, d\alpha\, d\vartheta_3\dots$$

$$= \int_D \tilde{p}(r^2)\, r^{n-1}\, dr \int_0^{2\pi} \int\dots\int \mathbf{1}_B(f(\alpha,\vartheta_3,...,\vartheta_n))\ |g(\alpha,...)|\ d\alpha\, d\vartheta_3\dots =: F[D]\cdot G[B].$$

Mit $D = (0,\infty)$ bzw. $B = \mathbb{R}^n$ und $C := F[(0,\infty)]$ folgt: $P[W \in B] = C \cdot G[B]$,

$P[Z \in D] = \frac{1}{C} \cdot F[D]$, also: $P[W \in B, Z \in D] = P[W \in B] \cdot P[Z \in D]$.

[Für eine orthogonale Matrix $A \in \mathbb{R}^{n\times n}$ gilt: $P[AW \in B] = P[A(Y_1,...,Y_n)/\|(Y_1,...,Y_n)\| \in B]$

$= P[A(Y_1,...,Y_n)/\|A(Y_1,...,Y_n)\| \in B] = (PW^{-1})[B]$,

da $A(Y_1,...,Y_n)$ nach 5.1c die gleiche Verteilung wie $(Y_1,...,Y_n)$ hat. Also ist PW^{-1} rotationsinvariant und muß somit die Gleichverteilung auf der Sphäre S^{n-1} sein. Die Eigenschaft läßt sich aus der Charakterisierung des Lebesgue–Maßes als einziges translationsinvariantes und auf dem Einheitsintervall normiertes Maß herleiten.] []

11.11 Satz. Sei $T^* := Y_k / \sqrt{\frac{1}{n-k} \sum_{k+1}^{n} Y_j^2}$. Dann gilt: Zu einem Test $\varphi : \mathbb{R}^n \mapsto [0,1]$ existiert ein Test $\psi : \mathbb{R} \mapsto [0,1]$ mit: $E_\vartheta[\varphi\circ T] = E_\vartheta[\psi\circ|T^*|]$.

Beweis: In Hinblick auf (11.9) kann man nach Fubini $E_\vartheta[\varphi\circ T]$ iteriert berechnen (denn $\frac{Y_k}{\sigma}$ und $(\frac{Y_j}{\sigma}, k<j\le n)$ sind unter ϑ unabhängig), und man erhält:

$$E_\vartheta[\varphi\circ T] = \int_{(0,\infty)} P_\vartheta(\tfrac{Y_k}{\sigma})^{-1}[dt]\ \ E_\vartheta[\varphi(\tfrac{\beta_k}{t}\cdot\textstyle\sum_{k+1}^{n}\frac{Y_j}{\sigma}e_j)].$$

Ist $P = P_{(0,1)}$ $(0\in\mathbb{R}^n)$, so können wir weiter schreiben:

$E_\vartheta[\varphi(\frac{\beta_k}{t}\cdot\sum_{k+1}^{n}\frac{Y_j}{\sigma}e_j)] =: E_\vartheta[\tilde{\varphi}(\frac{1}{t}(Y_{k+1}/\sigma,...,Y_n/\sigma)] = E[\tilde{\varphi}(\frac{1}{t}(Y_{k+1},...,Y_n)]$

$=E[\tilde{\varphi}(\frac{1}{|t|}(Y_{k+1},...,Y_n)] = E[\tilde{\varphi}(\frac{1}{|t|}W\cdot Z)]$,

da die Verteilung von $(Y_{k+1},...,Y_n)$ symmetrisch ist, mit $Z := \sqrt{\sum_{k+1}^{n} Y_j^2}$,

$W := \frac{1}{Z}(Y_{k+1},...,Y_n)$. Gemäß Lemma 11.10 sind W und $Z^* := \sqrt{n-k}\cdot|t|/Z$ unabhängig. Mit Fubini und $\psi(z) := E[\tilde{\varphi}(\sqrt{n-k}\cdot W/z)]\cdot 1_{(0,\infty)}(z)$ ergibt sich:

$E[\tilde{\varphi}(\frac{1}{|t|}Z\cdot W)] = E[\tilde{\varphi}(\sqrt{n-k}\cdot\frac{1}{Z^*}\cdot W)\cdot 1_{(0,\infty)}(Z^*)] = E[\psi(Z^*)]$

$= E[\,\psi(|t|/\sqrt{\frac{1}{n-k}\sum_{k+1}^{n} Y_j^2}\,)\,] = E_\vartheta[\,\psi(|t|/\sqrt{\frac{1}{n-k}\sum_{k+1}^{n}(Y_j/\sigma)^2}\,)\,]$.

Es folgt: $E_\vartheta[\varphi\circ T] = \int P_\vartheta(\frac{Y_k}{\sigma})^{-1}[dt]\; E_\vartheta[\,\psi(|t|/\sqrt{\frac{1}{n-k}\sum_{k+1}^{n}(Y_j/\sigma)^2}\,)\,]$

$= E_\vartheta[\,\psi(|Y_k/\sigma|/\sqrt{\frac{1}{n-k}\sum_{k+1}^{n}(Y_j/\sigma)^2}\,)\,] = E_\vartheta[\,\psi(|T^*|)\,].[]$

11.12 Bemerkung. Schreibt man $|T^*| = \beta_k (n-k)^{-\frac{1}{2}}/|T| =: V(T)$ und $\tilde{P}_\vartheta := P_\vartheta T^{-1}$, so zeigt der Beweis von 11.11 die Gültigkeit folgender Aussagen. [Für $B = \mathbb{R}^n$ hat man dabei genau 11.11.]

Zu einem Test $\varphi : \mathbb{R}^n \mapsto [0,1]$ existiert ein Test $\psi : \mathbb{R} \mapsto [0,1]$ mit:

$$\{V\in B\}\int \varphi\, d\tilde{P}_\vartheta = \{V\in B\}\int \psi\circ V\, d\tilde{P}_\vartheta \quad \text{für } B\in\mathfrak{B}_1.$$

Die Statistik V heißt dann **suffizient** für $\{\tilde{P}_\vartheta;\ \vartheta\in\Theta\}$. $\psi(v)$ ist dabei gerade eine von ϑ unabhängige Version des bedingten Erwartungswertes $\tilde{E}[\varphi|V=v]$ von φ unter der Bedingung V=v bez. des W–Maßes $\tilde{P}_\vartheta$. []

11.13 Korollar. Ist ψ^* ein schärfster Test unter den Tests z.N. α für das auf den Raum $(\mathbb{R},\mathfrak{B}_1,\mathfrak{P}|T^*|^{-1})$ verlagerte Testproblem H gegen K, so ist $\varphi^* = \psi^*\circ|T^*|$ schärfster Test in $\Phi_{I,\alpha}$ für das Testen von H gegen K.

Beweis: Offenbar ist φ^* ($= \psi^*\circ V\circ T$) invariant. Zu $\varphi\in\Phi_{I,\alpha}$ existiert nach 10.5 und 11.5 ein Test φ' mit $\varphi = \varphi'\circ T$. Nach 11.11 existiert ein Test ψ, sodaß $\psi\circ|T^*|$ die gleiche Gütefunktion wie $\varphi'\circ T$ hat. Nach Voraussetzung gilt einerseits: $\int \psi\, dP_\vartheta|T^*|^{-1} = E_\vartheta[\psi\circ|T^*|] = E_\vartheta[\varphi] \le \alpha$ $\forall\ \vartheta\in H$ und andererseits $\forall\ \vartheta\in K$:

$E_\vartheta[\varphi] = E_\vartheta[\psi\circ|T^*|] = \int \psi\, dP_\vartheta|T^*|^{-1} \le \int \psi^*\, dP_\vartheta|T^*|^{-1} = E_\vartheta[\psi^*\circ|T^*|] = E_\vartheta[\varphi^*]$. []

Die Reduktion durch Suffizienz kann auch durch eine weitere Reduktion durch Invarianz ersetzt werden; dabei wird aber die Gruppe Q vergrößert und damit die zur Konkurrenz zugelassenen Menge von invarianten Test verkleinert [vgl. Witting/Nölle (1970), Witting (1985)].

11 G Die optimalen Tests.

Nach 11 E gilt $X - \hat{X}_k = \sum_{k+1}^{n} Y_j e_j$ und $Y_k e_k = \hat{X}_k - \hat{X}_h$, somit $\sum_{k+1}^{n} Y_j^2 = \|X - \hat{X}_k\|^2$. Mit (11.7), also $Y_k = \beta_k \cdot \hat{\alpha}_k$, erhält man also folgende Darstellung:

(11.14) $$|T^*| = \sqrt{n-k}\ \frac{\|\hat{X}_k - \hat{X}_h\|}{\|X - \hat{X}_k\|} = \frac{\|\hat{X}_k - \hat{X}_h\|}{\hat{\sigma}} \text{ mit } \hat{\sigma}^2 := \frac{1}{n-k}\|X - \hat{X}_k\|^2.$$

Dabei ist $\hat{\sigma}^2$ der Residualschätzer von 3.7 für die Varianz σ^2. Die **Normierungskonstante** β_k wurde in (11.7) erklärt.

11.15 Lemma. Unter ϑ ist die Verteilung von T* die nichtzentrale Student–Verteilung $t_{n-k}(\delta)$ mit dem Nichtzentralitätsparameter $\delta := \beta_k\alpha_k/\sigma = E_\vartheta[Y_k/\sigma]$.

Beweis. Die Aussage folgt mit (11.8) (insbesondere $E_\vartheta[Y_k] = \beta_k \cdot \alpha_k$) aus der Darstellung

$$T^* = (Y_k/\sigma)/\sqrt{\frac{1}{n-k}\ \sum_{k+1}^{n}\ (Y_j/\sigma)^2}\ . \ []$$

11.16 Satz. Für das Testen von H gegen K ist der Test

$$\varphi^* := \begin{cases} 1 & \text{für } |T^*| > t_{n-k,\alpha/2} \\ 0 & |T^*| \le t_{n-k,\alpha/2} \end{cases} \text{ mit } |T^*| \text{ wie in (11.14)}$$

schärfster Test unter den bzgl. Q invarianten Tests z.N. α.

Beweis: Mit $U(x) := |x|$ gilt nach 11.15: $|T^*| = U\circ T^* \sim t_{n-k}(\delta)U^{-1}$ unter ϑ. Nach 7.13(b) und dem Beweis dazu wissen wir:

(11.17) $$t_{n-k}(\delta)U^{-1} = t_{n-k}(|\delta|)U^{-1},$$
$$\{t_{n-k}(|\delta|)U^{-1},\ |\delta| \ge 0\} \text{ hat monotone DQ in id}_{[0,\infty)}.$$

Das auf den Raum $([0,\infty), [0,\infty)\cap\mathfrak{B}_1, \mathfrak{P}|T^*|^{-1})$ verlagerte Testproblem H gegen K hat die Gestalt

$$H: |\delta| = 0\ ,\quad K: |\delta| > 0\ ,$$

ist also einseitig in Bezug auf den Parameter $|\delta|$. Nach 9.1 und 9.2 sowie (11.17) ergibt sich dafür als schärfster Test in der Menge der Tests z.N. α :

$$\psi^*(t) := \begin{cases} 1 & t > c \\ 0 & t \le c \end{cases} \text{ für } .$$

Dabei ist c das α–Fraktil von $t_{n-k}U^{-1}$. Da t_{n-k} symmetrisch zum Ursprung ist, gilt: $c = t_{n-k,\alpha/2}$. Gemäß 11.13 ist nun $\psi^*\circ|T^*|$ der gesuchte Test auf S'. Wegen $\varphi^* = 0$ auf $N = \{\hat{\alpha}_k = 0\} = \{Y_k = 0\}$ ist φ^* in Hinblick auf 10.13 auch der gesuchte Test auf S. []

11.18 Anmerkung. Für den allgemeinen Fall $0 < h < k < n$ des Testen einer linearen Hypothese H: $a \in L_h$ gegen K: $a \in L_k \setminus L_h$ ergibt sich als schärfster Test in $\Phi_{I,\alpha}$ bzgl einer geeigneten Gruppe von Transformationen der folgende F–Test :

$$\varphi^* := \begin{cases} 1 \\ 0 \end{cases} \text{ für } T^{*2} \begin{matrix} > \\ \leq \end{matrix} F_{k-h,n-k,\alpha} \text{ mit } T^{*2} := \frac{\frac{1}{k-h} \|\hat{X}_k - \hat{X}_h\|^2}{\frac{1}{n-k} \|X - \hat{X}_k\|^2}$$

[Witting/Nölle (1970) S. 37, Witting (1985) S. 464]. Dabei ist $F_{k-h,n-k,\alpha}$ das α–Fraktil der Fisher–Verteilung. Wegen (11.14) und einer einfachen Beziehung zwischen $F_{1,n-k,\alpha}$ und $t_{n-k,\alpha}$ stimmt dieser Test im Falle k–h=1 mit dem Test aus 11.17 überein. []

Durch den Fall "h=k–1" werden zum einen bereits wichtige Beispiele erfaßt. Zum anderen können in diesem Rahmen auch einseitige Testprobleme mit nichtlinearen Hypothesen H behandelt werden, was hier nicht näher ausgeführt wurde. Drittens läßt sich in diesem Rahmen, wo Q möglichst klein gewählt wurde, die Beschränkung auf invariante Test in Anwendungsbeispielen wie etwa dem t–Test aus 5.7 leichter motivieren.

14 H Anwendungen.

Beispiel 1. (Fortsetzung). **Einstichprobenproblem**.

Hier ist h=0, k=1 und $b_k = e = (1,...,1)^\top$, $\alpha_k = \mu$. Nach Beispiel 1 aus §3 gilt:

$\hat{X}_k = \hat{a} = \hat{\mu}\, e$, $\hat{\mu} = \bar{X}$, also $\|\hat{X}_k\| = \sqrt{n}\, |\bar{X}|$ und $\hat{\sigma}^2 = S^2$.

Es ergibt sich somit:

$$|T^*| = \sqrt{n}\, \frac{|\bar{X}|}{S} .$$

Wegen h=0 hat man: $Q = \{ \pi(x) = c\, x \,;\, c \neq 0 \}$. Aus 11.16 folgt nun 5.7. []

Beispiel 2. (Fortsetzung) **Zweistichprobenfall.**

Hier ist h=1, k=2, $b_h = e = (1,...,1)^\top$, $b_k = e_1 = (1,...,1,0,...,0)^\top$, $\alpha_k = \mu_1 - \mu_2$.

Aus Beispiel 1 folgt: $\hat{X}_h = \bar{X}\, e$ und aus Beispiel 2 in §3: $\hat{X}_k = \bar{X}_1\, e_1 + \bar{X}_2\, e_2$ und

$$\hat{\sigma}^2 = \frac{1}{n-2} \left[\sum_1^{n_1} (X_{1j} - \bar{X}_1)^2 + \sum_1^{n_2} (X_{2j} - \bar{X}_2)^2 \right] = \frac{n_1-1}{n-2} S_1^2 + \frac{n_2-1}{n-2} S_2^2.$$

Es ergibt sich wegen $\bar{X} = \frac{1}{n}(n_1\bar{X}_1 + n_2\bar{X}_2)$ und $\bar{X}_i = \frac{1}{n}(n_1\bar{X}_i + n_2\bar{X}_i)$:

$$\hat{X}_k - \hat{X}_h = (\bar{X}_1 - \bar{X})\, e_1 + (\bar{X}_2 - \bar{X})\, e_2 = (n_2/n) \cdot (\bar{X}_1 - \bar{X}_2)\, e_1 + (n_1/n) \cdot (\bar{X}_2 - \bar{X}_1)$$

und somit $\|\hat{X}_k - \hat{X}_h\|^2 = (\bar{X}_1 - \bar{X}_2)^2 \cdot (n_1+n_2) \cdot n_1 \cdot n_2/n^2$ also

$$\|\hat{X}_k - \hat{X}_h\|^2 = \frac{1}{\frac{1}{n_1} + \frac{1}{n_2}} (\bar{X}_1 - \bar{X}_2)^2 = \frac{n_1 \cdot n_2}{n_1+n_2} (\bar{X}_1 - \bar{X}_2)^2.$$

Aus 11.16 folgt nun

11.19 Satz. Zweistichproben–t–Test.

Seien $X_{11},...,X_{1n_1},X_{21},...,X_{2n_2}$ unabhängige Zva mit $X_{ij} \sim N(\mu_i,\sigma^2)$, $\mu_i \in \mathbb{R}$, $\sigma^2>0$,

$$|T^*| := \frac{1}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \cdot \frac{|\bar{X}_1 - \bar{X}_2|}{\hat{\sigma}}.$$

Für das Testproblem $H : \mu_1 = \mu_2$ gegen $K : \mu_1 \neq \mu_2$ verwendet man den Test

$$\varphi^* := \begin{cases} 1 \\ 0 \end{cases} \text{ für } |T^*| \begin{matrix} > \\ \leq \end{matrix} t_{n-2,\alpha/2}.$$

Mit $Q = \{ \pi(x) = c\,x + d\,e;\ c \neq 0,\ d \in \mathbb{R} \}$ ist φ^* schärfster Test für das Testen von H gegen K unter den bzgl. Q invarianten Tests z.N. α.

Satz 11.19 ist ein Beispiel einer **Varianzanalyse**. Wir wollen hier die Ausdruck $\sum_i (x_i - \bar{x})^2$ mit $\bar{x} := \sum_i x_i$ als "Stichprobenvarianz ohne Normierung" der Stichprobe x bezeichnen. Nach Pythagoras ergibt sich

$a^2 + b^2 = c^2$ mit $a^2 := \|\hat{X}_k - \hat{X}_h\|^2$, $b^2 := \|X - \hat{X}_k\|^2$, $c^2 := \|X - \hat{X}_h\|^2$. Dabei ist

(i) $a^2 := \frac{1}{\frac{1}{n_1}+\frac{1}{n_2}} (\bar{X}_1 - \bar{X}_2)^2$ die "Stichprobenvarianz ohne Normierung" von $(\bar{X}_1,...,\bar{X}_1,\bar{X}_2,..)$, wobei die Schwankungen der Werte X_{ij} unter gleichen Versuchsbedingung i herausgemittelt sind und nur die Unterschiedlichkeit der Werte unter verschiedenen Versuchbedingugen eingeht;

(ii) $b^2 := (n-2)\,\hat{\sigma}^2$ die Summe der "Stichprobenvarianzen ohne Normierung" innerhalb der Werte, die unter gleichen Versuchbedingungen erhalten werden;

(iii) $c^2 = \sum_{i,j} (X_{ij} - \bar{X})^2$ die "Stichprobenvarianz ohne Normierung" der Gesamtstichprobe, die also zerlegt wird in die Anteile a^2 und b^2.

Es gilt nun $|T^*| = \sqrt{n-2}\ |a| / |b|$. Man trifft eine Entscheidung für K, falls a^2 groß ist gegenüber b^2 (oder c^2). Diese Zerlegung der Stichprobenvarianz, auf der die Entscheidung basiert, rechtfertigt den Namen Varianzanalyse, obgleich die Mittelwerte getestet werden.

Der Fall, wo im Zweistichprobenproblem die Varianz τ^2 von X_{2j} von der Varianz σ^2 von X_{ij} verschieden sein kann, also $\vartheta = (\mu_1,\mu_2,\sigma^2.\tau^2)$ gewählt werden kann,ist unter dem Namen **Behrends–Fisher–Problem** bekannt. Der Fall verschiedener, aber bekannter Varianzen kann auf den Fall gleicher bekannter Varianzen zurückgeführt werden und ist eine Übungsaufgabe.

Beispiel 3. Lineare Regression in einer Variablen.

Es wird das Beispiel 3 aus §3 aufgegriffen. Seien wieder $s_1,...,s_n \in \mathbb{R}$, sodaß nicht alle Werte übereinstimmen, $a_i := \mu + \nu \cdot s_i$, $e := (1,...,1)^\top$, $s := (s_1,...,s_n)^\top$, $a = \mu\, e + \nu\, s$, $L_k := \mathfrak{L}(e,s)$, also k=2. Es interessiert, ob eine Abhängigkeit von den durch die Werte s_i gegebenen Versuchsbedingungen vorliegt, d.h. ob $\nu = 0$ oder $\nu \neq 0$ gilt. Man wählt $L_h = \mathfrak{L}(e)$, also h=1.

Wir können $b_h = e$ und $b_k = s$ wählen; dann folgt $\alpha_k = \nu$. Aus §3 ist folgendes bekannt:

$$\hat{\nu} = \hat{\alpha}_k = (\overline{s\bar{X}} - \bar{s}\bar{X})/(\overline{s^2} - \bar{s}^2) \text{ und } \hat{\sigma}^2 = \frac{1}{n-2} \sum [X_i - \bar{X} - \hat{\nu} \cdot (s_i - \bar{s})]^2,$$

$$\hat{X}_k = \hat{\mu}\, e + \hat{\nu}\, s,\ \hat{\mu} = \bar{X} - \bar{s}\, \hat{\nu} \quad (\text{z.B. ist } \overline{s\bar{X}} = \frac{1}{n} \sum s_i \cdot X_i).$$

Wie in den vorangehenden Beispielen ergibt sich: $\hat{X}_h = \bar{X}\, e$, also:

$$\hat{X}_k - \hat{X}_h = -\bar{s} \cdot \hat{\nu}\, e + \hat{\nu}\, s = \hat{\nu}\, (s - \bar{s}\, e),$$

$$\|\hat{X}_k - \hat{X}_h\|^2 = \hat{\nu}^2 \sum (s_i - \bar{s})^2 = \hat{\nu}^2 \cdot n \cdot (\overline{s^2} - \bar{s}^2).$$

Es ergibt sich nun der folgende Satz mit

$$|T^*| := \sqrt{n \cdot (\overline{s^2} - \bar{s}^2)}\ |\hat{\nu}|\ /\ \hat{\sigma} = \sqrt{n}\ |\overline{s\bar{X}} - \bar{s}\bar{X}|\ /\ \sqrt{\hat{\sigma}^2 (\overline{s^2} - \bar{s}^2)}.$$

$|T^*|$ hat also bis auf den Faktor $\sqrt{n}$ die Gestalt des Betrags einer Art von "Korrelationskoeffizienten" von $(s_1,...,s_n)$ und $(X_1,...,X_n)$.

11.20 Satz. Seien $X_1,...,X_n$ unabhängige Zva mit $X_i \sim N(\mu + \nu s_i, \sigma^2)$, $\mu,\nu \in \mathbb{R}$, $\sigma^2 > 0$.

Für das Testproblem $H : \nu = 0$ gegen $K : \nu \neq 0$ verwendet man den Test

$$\varphi^* := \begin{cases} 1 & \text{für } |T^*| \ > \\ 0 & \text{für } |T^*| \ \leq \end{cases} t_{n-2,\alpha/2}\,.$$

Mit $Q = \{\, \pi(x) = c\, x + d\, e;\ c \neq 0,\ d \in \mathbb{R} \,\}$ ist φ^* schärfster Test für das Testen von H gegen K unter den bzgl. Q invarianten Tests z.N.α.

§ 11 (verkürzte Variante) **Der zweiseitige t–Test als invarianter Test in einem linearen Modell.**

11 A Das Testproblem.

Es sei $(S,\mathcal{S},\mathfrak{P})$ ein lineares Modell mit Normalverteilungsannahme und unbekannter Varianz gemäß Bemerkung 3.3:

Generalvoraussetzung: $(S,\mathcal{S}) = (\mathbb{R}^n,\mathfrak{B}_n)$, L_k ein k–dim. linearer Teilraum von $\mathbb{R}^n$ mit k<n,

$\mathfrak{P} = \{ P_\vartheta = \bigotimes_{i=1}^n N(a_i,\sigma^2) ; \vartheta=(a,\sigma^2) \in \Theta = L_k \times (0.\infty)$ mit $a = (a_1,...,a_n)^\top \}$

$= \{ P_{\mu,\sigma^2} = \bigotimes_{i=1}^n N(\mu,\sigma^2) ; \mu\in\mathbb{R}, \sigma^2>0\}$

und $e = (1,...,1)^\top$. Dann gilt offenbar $a = \mu\, e$ und somit $L_k = \mathfrak{L}(e)$, k=1, $\alpha_k := \mu$.

Wir interessieren uns für folgendes Testproblem:

(11.1) H: $\mu = 0$ K: $\mu \neq 0$, d.h. H: $a \in L_h := \{0\}$, K: $a \in L_k \setminus L_h$.

Der Fall, wo in dem Testproblemen μ_o $[\in \mathbb{R}]$ anstelle von 0 steht, läßt sich mit 10.11 auf den obigen Fall zurückführen. Im weiteren Verlauf von § 11 gelte ebenfalls folgende

Generalvoraussetzung: $\hat{X}_k =: \hat{\mu}\, e$ sei die Projektion auf L_k mit $\hat{\mu} = \bar{X}$.

$L_k^\perp$ seien das orthogonale Komplement zu L_k.

Gemäß §3 (Beispiel 1 und3.4) ist $\hat{X}_k$ KQS für a. Nach 3.9(c) ist $\hat{\mu} = \bar{X}$ GMS für den Parameter μ, insbesondere linear und erwartungstreu.

$X - \hat{X}_k = X - \bar{X}\, e$ ist die Projektion auf $L_k^\perp$.

Auch das einseitige Testproblem H: $\mu \le 0$ K: $\mu > 0$ kann mit den hier verwendeten Methoden behandelt werden. Dies soll aber nicht geschehen.

11 B Invarianz des Testproblems.

Es sei

(11.2) $Q := \{ \pi(x) = c\, x \ ; c\neq 0 \}$.

11.3 Lemma. (a) Q ist eine Gruppe von mb Transformationen.

(b) Das Testproblem H gegen K ist invariant gegenüber Q.

Beweis: Teil a) ist offensichtlich.

b) Sind X_i unabhängig und $N(\mu,\sigma^2)$–verteilt, $1\le i\le n$, so sind $c\cdot X_i$ wieder unabhängig und $N(c\cdot\mu,c^2\sigma^2)$–verteilt. Also gilt: $P_{\mu,\sigma^2}\,\pi^{-1} = P_{c\,\mu,c^2\sigma^2}$.

Wegen "$\mu = 0 \Leftrightarrow c\cdot\mu = 0$" folgt die Behauptung. $[]$

11 C Abspaltung einer Nullmenge.

Es sei

$$N := L_k^{\perp} = \{\bar{X} = 0\}.$$

Dann ist N also ein (n–1)–dim. Teilraum von $\mathbb{R}^n$ und hat somit das Lebesgue–Maß 0. Daher gilt auch:

$$P_\vartheta[N] = 0 \quad \forall\, \vartheta \in \Theta .$$

Es gilt offenbar $\pi(N) = N \;\forall\, \pi$. Wie in 10.13 sei jetzt

$$S' = S \setminus N \,,\; Q' = \{\, \pi' = \pi|_{S'} \,;\, \pi \in Q \,\},$$

und damit auch $\pi(S') = \pi(S') \;\forall\, \pi$.

11 D Reduktion durch Invarianz

Es soll nun eine maximal invariante Statistik gesucht werden. Jede Statistik T hängt nur von $\hat{\alpha}_k$ und der Projektion $X–\hat{X}_k = X - \bar{X}\, e$ auf $L_k^{\perp}$ ab. Eine gegenüber $x \mapsto c\, x$ invariante Statistik ist

(11.4) $$T := (X - \hat{X}_k)/\bar{X} = (X - \bar{X}\, e)/\bar{X}.$$

Daß $\bar{X}$ als Schätzer für den zu testenden Parameter μ bei T im Nenner stehen muß, liegt daran, daß man den Vektor $(X - \hat{X}_k)$ nicht in den Nenner nehmen kann. Die entgültige Prüfgröße wird aber wieder die gewohnte Gestalt haben: Im Zähler steht $\bar{X}$ als Schätzer für den zu testenden Parameter μ; im Nenner steht ein Schätzer für die Varianz; dazu kommt noch ein deterministischer Normierungsfaktor.

11.5 Lemma. T ist maximal invariant auf S' gegenüber Q' und hat eine mb schwache Inverse.

Beweis: (i) Zum Nachweis der Invarianz von T sei $\pi(x) = c\, x$ mit $c \neq 0$.

Dann folgt aus der Linearität von $\bar{X}$ und $\hat{X}_k$ die Invarianz mit:

$\bar{X}(cx) = c\, \bar{X}(x)$; $(X–\hat{X}_k)(cx) = c\, (X–\hat{X}_k)(x)$.

(ii) Zum Nachweis der maximalen Invarianz sei $T(x) = T(x')$, $x \in S'$. Dann gilt mit

$c := \bar{X}(x')/\bar{X}(x) \neq 0$ also $\hat{X}_k(x') = c\, \hat{X}_k(x)$ auch: $x' - \hat{X}_k(x') = c \cdot (x - \hat{X}_k(x))$. Nun folgt:

$x' = \hat{X}_k(x') + x' - \hat{X}_k(x') = c\, \hat{X}_k(x) + c \cdot (x - \hat{X}_k(x)) = c \cdot x =: \pi(x)$.

(iii) Zur Bestimmung einer mb schwachen Inversen überlegt man sich zunächst:

(11.6) $$T(S') \subset L_k^{\perp} .$$

Setzt man nun $T^{-}(y) := e + y$, $y \in L_k^{\perp}$, so erhält man: $\hat{X}_k(e+y) = e$ und

$(X - \hat{X}_k)(e+y) = y$, also $T(e+y) = y$. Damit gilt auch "=" in (11.6). $[\,]$

11 E Übergang zu einem anderen Koordinatensystem

Zur Berechnung der Verteilung der maximal invarianten Statistiken ist es günstig, eine Koordinatentransformation zu machen.

Es sei $e_1,...,e_n$ eine Orthonormalbasis des $\mathbb{R}^n$ mit $e_1 = (1/\sqrt{n})\, e$ (vgl. Beweis von Satz 5.5). Sei nun A die orthogonale Matrix

$$A := \begin{pmatrix} e_1^\top \\ \vdots \\ e_n^\top \end{pmatrix} \text{ und } Y := A\,X \text{ , d.h. } Y_j := e_j^\top \cdot X \text{ , } 1 \le j \le n, \text{ (Anteil von X in Richtung } e_j).$$

Dann gilt: $X = A^\top Y = \sum_1^n Y_j\, e_j$ und $\hat{X}_k = Y_1\, e_1$, $X - \hat{X}_k = \sum_2^n Y_j\, e_j$.
Andererseits haben wir:

$$\hat{X}_k = \bar{X}\, e = \sqrt{n} \cdot \bar{X}\, e_1 , \qquad a = \mu\, e = \mu \cdot \sqrt{n}\, e_1 .$$

Damit ergibt sich:

(11.7) $$Y_1 = \sqrt{n}\, \bar{X} \quad .$$

Aus 5.1c folgt:

(11.8) $Y_1,...,Y_n$ unabh., $Y_j \sim N(b_j,\sigma^2)$ mit $b_j = e_j^\top \cdot a$ und

(i) $b_j = 0$, $j > 1$;

(ii) $b_1 = \sqrt{n}\, \mu$.

Für die in Abschnitt D definierte Statistik erhält man somit:

(11.9) $$T = (\sqrt{n}/Y_1) \sum_2^n Y_j\, e_j .$$

Insbesondere hat man $\|T\| = (\sqrt{n}/|Y_1|) \sqrt{\sum_2^n Y_j^2}$; diese Statistik $\|T\|$ ist schon nahe an unserer endgültigen Prüfgröße.

11 F Reduktion durch Suffizienz.

Die Reduktion durch Invarianz gestattet nun eine Beschränkung auf Tests, die nur von T abhängen. In einer weiteren Reduktion soll eine Beschränkung auf reellwertige Statistiken gerechtfertigt werden. Als Motivation für das folgende Lemma schreiben wir mit Fubini:

$$E_\vartheta[\varphi \circ T] = \int P_\vartheta(Y_1/\sigma)^{-1}[dt]\; E_\vartheta[\varphi((\sqrt{n}/t) \cdot \sum_2^n (Y_j/\sigma)\, e_j)]$$
$$=: \int P_\vartheta(Y_1/\sigma)^{-1}[dt]\; E_\vartheta[\tilde{\varphi}(\tfrac{1}{t}(Y_2/\sigma,...,Y_n/\sigma)] .$$

Im folgenden Lemma schreiben wir $Y_1,...,Y_n$ statt $Y_2,...,Y_n$.

11.10 Lemma. Seien $Y_1,...,Y_n$ iid mit $Y_j \sim N(0,1)$, $Z := \sqrt{\sum_1^n Y_j^2}$, $W := \frac{1}{Z}(Y_1,...,Y_n)$; dann sind W und Z unabhängig. [Dabei ist die Verteilung von W die Gleichverteilung $U(S^{n-1})$ auf der Sphäre S^{n-1} und $Z^2 \sim \chi_n^2$.]

Beweis: Sei $\tilde{p}(t) := (2\pi)^{-n/2} \exp\{-\frac{1}{2}t\}$, $t \in \mathbb{R}$.

Dann ist $y = (y_1,...,y_n) \mapsto \tilde{p}(\sum_1^n y_j^2)$ eine Dichte von $(Y_1,...,Y_n)$ und es gilt:

$P[\, W \in B, Z \in D] = \int...\int \mathbf{1}_B(\frac{1}{\|y\|}\, y) \cdot \mathbf{1}_D(\|y\|)\, \tilde{p}(\|y\|^2)\, dy$.

Wir gehen nun zu Polarkoordinaten über [Barner/Flohr (1983) S. 43] und setzen n>1 voraus.
[Der Fall n=1 kann ohne Polarkoordinaten gelöst werden und ist eine Übungsaufgabe.]

Es ist $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = r \cdot f(\alpha,\vartheta_3,...,\vartheta_n)$ mit $r>0$, $0 \le \alpha < 2\pi$, $-\frac{1}{2}\pi < \vartheta_j < \frac{1}{2}\pi$.

Für die Funktionaldeterminante Γ ergibt sich

$$\Gamma = \left| \partial_r(r\cdot f), \partial_\alpha(r\cdot f), \dots, \partial_{\vartheta_n}(r\cdot f) \right| = \left| f, r\cdot\partial_\alpha f, \dots, r\cdot\partial_{\vartheta_n}(r\cdot f) \right| .$$

Eine Entwicklung nach der ersten Spalte zeigt:

$\Gamma = r^{n-1} g(\alpha,\vartheta_3,...)$ für eine gewisse Funktion g.

(Es gilt sogar $g(\alpha,\vartheta_3,...) = \sin^{n-1}\alpha \cdot \sin^{n-2}\vartheta_3 \cdot ... \cdot \sin\vartheta_n$ [vgl. Elstrodt S. 206]). Also ergibt sich

$$\begin{aligned} P[\, W \in B, Z \in D] &= \int...\int \mathbf{1}_B(\, f(\alpha,\vartheta_3,...,\vartheta_n)\,) \cdot \mathbf{1}_D(\, r\,)\ \tilde{p}(r^2)\ |\Gamma|\ dr\, d\alpha\, d\vartheta_3... \\ &= \int...\int \mathbf{1}_B(\, f(\alpha,\vartheta_3,...,\vartheta_n)\,) \cdot \mathbf{1}_D(\, r\,)\ \tilde{p}(r^2)\, r^{n-1}\ |g(\alpha,...)|\ dr\, d\alpha\, d\vartheta_3... \\ &= \int_D \tilde{p}(r^2)\, r^{n-1}\, dr \int_0^{2\pi} \int...\int \mathbf{1}_B(\, f(\alpha,\vartheta_3,...,\vartheta_n)\,)\ |g(\alpha,...)|\ d\alpha\, d\vartheta_3... \\ &=: F[D] \cdot G[B] . \end{aligned}$$

Mit $D = (0,\infty)$ bzw. $B = \mathbb{R}^n$ und $C := F[(0,\infty)]$ folgt: $P[\, W \in B] = C \cdot G[B]$,

$P[\, Z \in D] = \frac{1}{C} \cdot F[D]$, also: $P[\, W \in B, Z \in D] = P[\, W \in B] \cdot P[\, Z \in D]$.

[Für eine orthogonale Matrix $A \in \mathbb{R}^{n\times n}$ gilt: $P[AW \in B] = P[A(Y_1,...,Y_n)/\|(Y_1,...,Y_n)\| \in B]$

$= P[A(Y_1,...,Y_n)/\|A(Y_1,...,Y_n)\| \in B] = (PW^{-1})[B]$,

da $A(Y_1,...,Y_n)$ nach 5.1c die gleiche Verteilung wie $(Y_1,...,Y_n)$ hat. Also ist PW^{-1} rotationsinvariant und muß somit die Gleichverteilung auf der Sphäre S^{n-1} sein. Die Eigenschaft läßt sich aus der Charakterisierung des Lebesgue–Maßes als einziges translationsinvariantes und auf dem Einheitsintervall normiertes Maß herleiten.] []

11.11 Satz. Sei $T^* := Y_1/\sqrt{\frac{1}{n-1}\ \sum_2^n\ Y_j^2}$. Dann gilt: Zu einem Test $\varphi : \mathbb{R}^n \mapsto [0,1]$ existiert ein Test $\psi : \mathbb{R} \mapsto [0,1]$ mit: $E_\vartheta[\varphi\circ T] = E_\vartheta[\psi\circ|T^*|]\quad \forall\, \vartheta$.

Wir haben also: $(1/\sqrt{n})\ \|T\| = \frac{\sqrt{n-1}}{|T^*|} = \frac{1}{|Y_1|}\sqrt{\sum_2^n\ Y_j^2}$ sowie

$$|T^*| = \sqrt{n\cdot(n-1)}\ /\|T\| =: V(T).$$

Beweis: Gemäß (11.9) gilt:

$E_\vartheta[\varphi\circ T] = \int E_\vartheta[\varphi((\sqrt{n}/t)\cdot\Sigma_2^n \frac{Y_j}{\sigma} e_j)]\ P_\vartheta(\frac{Y_1}{\sigma})^{-1}[dt]$
nach Fubini. Ist $P := P_{0,1}$, so können wir schreiben:

$E_\vartheta[\varphi((\sqrt{n}/t)\cdot\Sigma_2^n \frac{Y_j}{\sigma} e_j)] =: E_\vartheta[\tilde{\varphi}(\frac{1}{t}(Y_2/\sigma,...,Y_n/\sigma)] = E[\tilde{\varphi}(\frac{1}{t}(Y_2,...,Y_n)]$

$= E[\tilde{\varphi}(\frac{1}{|t|}(Y_2,...,Y_n)] = E[\tilde{\varphi}(\frac{1}{|t|} W\cdot Z)]$

mit $Z := \sqrt{\Sigma_2^n Y_j^2}$, $W := \frac{1}{Z}(Y_2,...,Y_n)$, da die Verteilung von $(Y_2,...,Y_n)$ symmetrisch ist. Gemäß Lemma 11.10 sind W und $Z^* := \sqrt{n-1}\cdot|t|/Z$ unabhängig. Mit Fubini und

$\psi(z) := E[\tilde{\varphi}(\sqrt{n-1}\cdot W/z)]\cdot 1_{(0,\infty)}(z)$ ergibt sich:

$E[\tilde{\varphi}(\frac{1}{|t|} Z\cdot W)] = E[\tilde{\varphi}(\sqrt{n-1}\cdot\frac{1}{Z^*}\cdot W)] = E[\psi(Z^*)]$

$= E[\psi(|t|/\sqrt{\frac{1}{n-1}\Sigma_2^n Y_j^2})] = E_\vartheta[\psi(|t|/\sqrt{\frac{1}{n-1}\Sigma_2^n (Y_j/\sigma)^2})]$.

Es folgt: $E_\vartheta[\varphi\circ T] = \int E_\vartheta[\psi(|t|/\sqrt{\frac{1}{n-1}\Sigma_2^n (Y_j/\sigma)^2})]\ P_\vartheta(\frac{Y_1}{\sigma})^{-1}[dt]$

$= E_\vartheta[\psi(|Y_1/\sigma|/\sqrt{\frac{1}{n-1}\Sigma_2^n (Y_j/\sigma)^2})] = E_\vartheta[\psi(|T^*|)]$. []

11.12 Bemerkung. Schreibt man $|T^*| = V(T)$, so zeigt der Beweis von 11.11 die Gültigkeit folgender Aussagen. [Für $B = \mathbb{R}^n$ hat man dabei genau 11.11.]

Zu einem Test $\varphi : \mathbb{R}^n \mapsto [0,1]$ existiert ein Test $\psi : \mathbb{R} \mapsto [0,1]$ mit:

$$\{V\in B\}\int \varphi\, dP_\vartheta T^{-1} = \{V\in B\}\int \psi\circ V\, dP_\vartheta T^{-1} \text{ für } B\in\mathfrak{B}_1.$$

Die Statistik V heißt dann **suffizient** für $\{P_\vartheta T^{-1};\ \vartheta\in\Theta\}$. $\psi(v)$ ist dabei gerade eine von ϑ unabhängige Version des bedingten Erwartungswertes von φ unter dem W–Maß $P_\vartheta T^{-1}$ unter der Bedingung V=v . []

11.13 Korollar. Ist ψ^* ein schärfster Test unter den Tests z.N. α für das auf den Raum $(\mathbb{R},\mathfrak{B}_1,\mathfrak{P}|T^*|^{-1} := \{P_\vartheta|T^*|^{-1},\vartheta\in\Theta\})$ verlagerte Testproblem H gegen K, so ist $\varphi^* = \psi^*\circ|T^*|$ schärfster Test in $\Phi_{I,\alpha}$ für das Testen von H gegen K.

Beweis: Zu $\varphi\in\Phi_{I,\alpha}$ existiert nach 10.5 und 11.5 ein Test φ' mit $\varphi = \varphi'\circ T$. Nach 11.11 existiert ein Test ψ, sodaß $\psi\circ|T^*|$ die gleiche Gütefunktion wie $\varphi'\circ T = \varphi$ hat. Nach Voraussetzung gilt einerseits: $\int\psi\, dP_\vartheta|T^*|^{-1} = E_\vartheta[\psi\circ|T^*|] = E_\vartheta[\varphi] \le \alpha\ \forall\ \vartheta\in H$ und andererseits $\forall\ \vartheta\in K$:

$E_\vartheta[\varphi] = E_\vartheta[\psi\circ|T^*|] = \int\psi\, dP_\vartheta|T^*|^{-1} \le \int\psi^*\, dP_\vartheta|T^*|^{-1} = E_\vartheta[\psi^*\circ|T^*|] = E_\vartheta[\varphi^*]$. []

Die Reduktion durch Suffizienz kann auch durch eine weitere Reduktion durch Invarianz ersetzt werden; dabei wird aber die Gruppe Q der Transformationen vergrößert und damit die zur Konkurrenz zugelassenen Menge von invarianten Test verkleinert [vgl. Witting/Nölle (1970), Witting (1985)].

11 G Der optimale Test.

Nach 11 E gilt $X - \hat{X}_k = \sum_2^n Y_j\, e_j$ und $\hat{X}_k = Y_1\, e_1 = \sqrt{n}\,\bar{X}\, e_1 = \bar{X}\, e$, somit $\sum_2^n Y_j^2 = \|X - \hat{X}_k\|^2 = \|X - \bar{X}\, e\|^2$. Also erhält man folgende Darstellung:

(11.14) $$|T^*| = \sqrt{n-1}\ \frac{\|\hat{X}_k\|}{\|X - \hat{X}_k\|} = \sqrt{n}\,\frac{|\bar{X}|}{S}$$

mit $S^2 = \frac{1}{n-1}\,\|X - \bar{X}\, e\|^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - \bar{X})^2$.

Dabei ist S^2 die Stichprobenstreuung und der Residualschätzer von 3.7 für die Varianz σ^2.

11.15 Lemma. (a) Unter ϑ ist die Verteilung von T* die nichtzentrale Student–Verteilung $t_{n-1}(\delta)$ mit dem Nichtzentralitätsparameter $\delta := \sqrt{n}\,\mu/\sigma = E_\vartheta[Y_1/\sigma]$.

(b) Der Test $\psi^*(t) := \begin{cases} 1 & t > t_{n-1,\alpha/2} \\ 0 & t \le t_{n-1,\alpha/2} \end{cases}$ für ist schärfster Test unter den Tests z.N. α für das auf den Raum $(\mathbb{R}, \mathfrak{B}_1, \mathfrak{P}|T^*|^{-1} := \{P_\vartheta|T^*|^{-1}, \vartheta\in\Theta\})$ verlagerte Testproblem H gegen K.

Beweis. a) Die Aussage folgt mit (11.8) aus der Darstellung

$$T^* = (Y_1/\sigma)/\sqrt{\frac{1}{n-1}\ \sum_2^n\ (Y_j/\sigma)^2}\ .$$

Im Beweis von Satz 5.5 wurde der Fall $\mu = 0$ behandelt.

b) Mit $U(x) := |x|$ gilt nach (a): $|T^*| \sim t_{n-1}(\delta)U^{-1}$ unter ϑ. Nach 7.13(b) und dem Beweis dazu wissen wir:

(11.17) $$t_{n-1}(\delta)U^{-1} = t_{n-1}(|\delta|)U^{-1},$$
$$\{t_{n-1}(|\delta|)U^{-1},\ |\delta| \ge 0\} \text{ hat monotone DQ in } \mathrm{id}_{[0,\infty)}.$$

Das auf den Raum $([0,\infty), [0,\infty)\cap\mathfrak{B}_1, \mathfrak{P}|T^*|^{-1})$ verlagerte Testproblem H gegen K hat die Gestalt

$$H: |\delta| = 0\,, \quad K: |\delta| > 0\,,$$

ist also einseitig in Bezug auf den Parameter $|\delta|$.

Nach 9.1 und 9.2 ergibt sich dafür als schärfster Test in der Menge der Tests z.N. α :

$$\psi^*(t) := \begin{cases} 1 & t > c \\ \text{für} & \\ 0 & t \le c \end{cases}.$$

Dabei ist c das α–Fraktil von $t_{n-1}U^{-1}$. Da t_{n-1} symmetrisch zum Ursprung ist, gilt: $c = t_{n-1,\alpha/2}$. $[]$

11.16 Satz. Für das Testen von H gegen K ist der Test

$$\varphi^* := \begin{cases} 1 & \text{für } |T^*| > t_{n-1,\alpha/2} \\ 0 & |T^*| \le t_{n-1,\alpha/2} \end{cases} \quad \text{mit } |T^*| = \sqrt{n}\,\frac{|\bar{X}|}{S}$$

schärfster Test unter den bzgl. Q invarianten Tests z.N. α.

Beweis: Gemäß 11.13 ist nun $\psi^* \circ |T^*|$ der gesuchte Test auf S'. Wegen $\varphi^* = 0$ auf $N = \{\bar{X} = 0\}$ und $\pi(N) = N$, $\pi(S') = S'$ $\forall\, \pi \in Q$ ist φ^* auch invariant auf S. In Hinblick auf 10.13 ist φ^* der gesuchte Test auf S. $[]$

11.18 Anmerkung. Für den allgemeinen Fall $0 < h < k < n$ des Testen einer linearen Hypothese H: $a \in L_h$ gegen K: $a \in L_k \backslash L_h$ mit einem h–dimensionalen linearen Teilraum L_h ergibt sich als schärfster Test in $\Phi_{I,\alpha}$ bzgl einer geeigneten Gruppe von Transformationen der folgende F–Test :

$$\varphi^* := \begin{cases} 1 & \text{für } T^{*2} > \\ 0 & \le \end{cases} F_{k-h,n-k,\alpha} \quad \text{mit } T^{*2} := \frac{\frac{1}{k-h}\,\|\hat{X}_k - \hat{X}_h\|^2}{\frac{1}{n-k}\,\|X - \hat{X}_k\|^2}$$

[Witting/Nölle (1970) S. 37, Witting (1985) S. 464]. Dabei ist $\hat{X}_h$ die Projektion auf L_h und $F_{k-h,n-k,\alpha}$ das α–Fraktil der Fisher–Verteilung $F_{k-h,n-k}$. Wegen (11.14) und einer einfachen Beziehung zwischen $F_{1,n-1,\alpha}$ und $t_{n-1,\alpha}$ stimmt dieser Test im Falle k–h=1, n–k=n–1 mit dem Test aus 11.17 überein. $[]$