4 Parametrische statistische Methoden

4.1 Datenniveau

Jede statistische Arbeit beginnt mit der Erhebung der Daten. Es ist heute üblich, die Daten in vier Gruppen einzuteilen.

1. Nominale Skala: Zum Beispiel Blutgruppe, Geschlecht, Partei, Postleitzahl.
2. Ordinale Skala: Noten, Sportranglisten, Windstärke.
3. Intervallskala: Temperatur, Zeitdauer.
4. Verhältnisskala: Länge, Gewicht, Fläche, Kosten.

Nominale und ordinale Daten gehören zum Bereich der *nichtparametrischen* Statistik, die wir im nächsten Kapitel behandeln, Intervall- und Verhältnisskala zur *parametrischen* Statistik.

Die parametrische Statistik stellt sich die Aufgabe, aus einer vorgegebenen Familie von möglichen Verteilungen Parameter zu bestimmen oder zumindest zu schätzen.

Die nichtparametrische Statistik versucht, aus den Daten allgemeine Strukturaussagen zu erzielen.

Von Skala 1 zu 4 erhalten wir einen Informationsgewinn, andererseits werden die Meßungenauigkeiten größer. Zum Beispiel ist die richtige Blutgruppe leicht anzugeben, während die exakte Flächenmessung deutlich schwieriger ist.

Die Datenerhebung erfolgt üblicherweise durch

- Befragung
- Beobachtung
- Experiment.

4.2 Ansatz der Statistik

Wir beginnen mit einem typischen Beispiel, der Qualitätskontrolle.

Ein Importeur erhält eine Lieferung von $N = 10.000$ Orangen. Er möchte schätzen, wieviele, sagen wir w, davon faul sind. Er macht eine Stichprobe von $n = 50$ Orangen und stellt fest, dass x davon faul sind.

1. Idee: Er sagt sich, $\frac{x}{n} \sim \frac{w}{N}$, also wird er als *Schätzer* von w den Ausdruck $T(x) = \frac{N}{n}x$ verwenden. $T(x)$ ist vom Zufall (der ausgesuchten Orangen) abhängig, also eine *Zufallsvariable* $T : \mathcal{X} \to \mathbb{R}$, wobei $\mathcal{X} = \{0, 1, 2, \ldots, n\}$ der *Stichprobenraum* ist.

2. Idee: Der Importeur sucht ein Intervall $C(x)$, in dem die richtige Anzahl w an faulen Orangen mit genügend hoher W-keit liegt. $C(x)$ heißt ein *Konfidenzintervall* und hängt von der beobachteten Größe x ab. $C(x)$ darf *nicht* von w abhängen, da w ja nicht bekannt ist. Die Forderung ist also

$$P_w(\{x \in \mathcal{X} : w \in C(x)\}) \geq 1 - \alpha \text{ für } \textit{alle } w \in \{0, 1, \ldots, N\}\,,$$

wobei P_w das Wahrscheinlichkeitsmaß ist, falls w die richtige Anzahl ist. Die Zahl α heißt *Konfidenzniveau* und wird üblicherweise als $\alpha = 0,05$ oder $0,025$ angenommen.

3. Idee: Angenommen, der Importeur muß den Preis nur zahlen, wenn höchstens 5% der Orangen faul sind. Er stellt folgende Hypothesen auf:

$$\begin{aligned} H_0 &: \; w \in \{0, 1, \ldots, 500\} \\ H_1 &: \; w \in \{501, \ldots, 10.000\}\,. \end{aligned}$$

H_0 heißt die *Nullhypothese*, H_1 die *Alternativhypothese*. Nun entwirft er ein *Entscheidungsverfahren*, um festzustellen, welche Hypothese zutrifft. Zum Beispiel bestimmt er eine Zahl c mit

$$\begin{aligned} x \leq c &\implies H_0 \;\; \text{(Importeur akzeptiert)}, \\ x > c &\implies H_1 \;\; \text{(Importeur lehnt ab)}. \end{aligned}$$

Die kritische Zahl c soll so bestimmt werden, dass

(I) $P_w(x > c)$ klein ist für $w \leq 500$

(II) $P_w(x > c)$ groß ist für $w > 500$.

Bei (I) sprechen wir von einem *Irrtum* $1.$*Art*. Es soll vermieden werden, dass der Importeur die Sendung ablehnt, obwohl sie in Ordnung ist.

Bei (II) sprechen wir von einem *Irrtum 2.Art*. Falls die Sendung nicht in Ordnung ist, so soll er sie nur mit kleiner W-keit akzeptieren.

In der Praxis ist H_0 oft die konservative Hypothese und der Irrtum 1.Art wichtiger. Soll zum Beispiel ein neues Medikament gegen ein am Markt bewährtes getestet werden, so wird als Nullhypothese H_0 angenommen, dass kein signifikanter Unterschied besteht, während die Alternativhypothese H_1 sagt, dass ein Unterschied besteht (zweiseitiger Test) oder dass das neue Medikament signifikant besser ist (einseitiger Test). Der Irrtum 1.Art besagt: Das neue Medikament wird eingeführt, obwohl es nichts bringt.

Wir haben also bei parametrischen Verfahren die folgende Struktur:

1. Die Beobachtungsergebnisse x bilden den Stichprobenraum $\mathcal{X}$, der mit einer σ-algebra $\mathcal{E}$ versehen wird.

2. Die Menge der möglichen Verteilungen ist $\{P_\vartheta : \vartheta \in \theta\}$, wobei θ der Parameterraum ist. Ein *parametrisches Modell* ist dann

$$\mathcal{M} = \{\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta\}, \ \theta \subseteq \mathbb{R} \text{ oder } \theta \subseteq \mathbb{R}^n \,.$$

Beispiel. Unser Orangenbeispiel entspricht offenbar den hypergeometrischen Verteilungen

$$\mathcal{X} = \{0, 1, \dots, n\}, \ \mathcal{E} = 2^{\mathcal{X}}, \ \theta = \{0, 1, \dots, N\}, \vartheta = \# \text{ faule Orangen;}$$

$$P_\vartheta(X = x) = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{N}{n}} \,.$$

Machen wir n unabhängige Stichproben, so haben wir das Modell

$$(\mathcal{X}^n, \mathcal{E}^{\otimes n}, P_\vartheta^n : \vartheta \in \theta) \,.$$

Wir halten fest:

$$\begin{aligned} X &= \text{ Zufallsvariable der Stichprobe,} \\ x &= \text{ beobachteter Wert.} \end{aligned}$$

4.3 Parameterschätzung

Es sei das Modell $\mathcal{M} = (\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$ gegeben, und $(W, \mathcal{F})$ sei eine σ-Algebra auf W. Sei $\tau : \theta \to W$ eine Abbildung, die jedem $\vartheta \in \theta$ einen gewissen Parameter $\tau(\vartheta) \in W$ zuordnet.

Zu einer Beobachtung $x \in \mathcal{X}$ wollen wir $T(x) \in W$ angeben, das den Parameter $\tau(\vartheta)$ schätzt. $T(x)$ heißt *Schätzer* von $\tau(\vartheta)$, und die Zufallsvariable $T : \mathcal{X} \to W$ heißt *Statistik*. Meist wird $\tau(\vartheta) = \vartheta$ sein.

Es seien $X_1, \ldots, X_n$ unabhängige identisch verteilte Stichprobenvariablen, $T_n = T(X_1, \ldots, X_n)$. T_n ist also eine Zufallsvariable

$$T_n : \mathcal{X}^n \to W \,.$$

Wir stellen folgende plausible Forderungen, wobei E_ϑ, Var_ϑ Erwartungswert und Varianz bedeuten, falls ϑ der richtige Parameter ist.

1. T_n heißt *erwartungstreu*, falls

$$E_\vartheta[T_n] = \tau(\vartheta) \text{ für alle } \vartheta \in \theta \,.$$

2. Etwas schwächer ist: T_n ist *asymptotisch erwartungstreu*, falls

$$\lim_{n\to\infty} E_\vartheta[T_n] = \tau(\vartheta) \text{ für alle } \vartheta \in \theta \,.$$

3. T_n heißt *konsistent*, wenn $T_n \xrightarrow{i.W.} \tau(\vartheta)$ für alle ϑ gilt, das heißt

$$P_\vartheta(|T_n - \tau(\vartheta)| \geq \varepsilon) \xrightarrow{n\to\infty} 0 \text{ für alle } \varepsilon > 0 \,.$$

Definition. T_n heißt *bester Schätzer*, falls gilt:

a. $E_\vartheta[T_n] = \tau(\vartheta)$ (T_n ist erwartungstreu),

b. für alle erwartungstreuen Schätzer U_n gilt

$$\mathrm{Var}_\vartheta[T_n] \leq \mathrm{Var}_\vartheta[U_n] \,.$$

Beispiel. Das Intervall $[0, \vartheta]$ sei gegeben, wobei $\vartheta > 0$ unbekannt ist. Es werden n unabhängige gleichverteilte Zufallszahlen aus $[0, \vartheta]$ gezogen. Es soll ϑ geschätzt werden. Hier ist

1. $\mathcal{X} = [0, \infty)$, $\mathcal{X}^n = [0, \infty)^n$, $\tau(\vartheta) = \vartheta$.

2. Gleichverteilung bedeutet für die Dichte $f_\vartheta(x) = \frac{1}{\vartheta}$ für alle $0 \le x \le \vartheta$, also $f_\vartheta(x_1, \ldots, x_n) = \frac{1}{\vartheta^n}$ für alle $(x_1, \ldots, x_n)$. Wir haben

$$E_\vartheta[X] = \frac{\vartheta}{2}, \mathrm{Var}_\vartheta[X] = \frac{\vartheta^2}{12}.$$

Idee I. Zu $X_1, \ldots, X_n$ nehmen wir als Schätzer

$$T_n(X_1, \ldots, X_n) = \frac{2}{n}(X_1 + \cdots + X_n),$$

da $2\frac{X_1+\cdots+X_n}{n} \approx \vartheta$ zu erwarten ist.

Wir haben

a. $E_\vartheta(T_n) = \frac{2}{n} n E_\vartheta[X] = 2\frac{\vartheta}{2} = \vartheta$, also ist T_n erwartungstreu.

b. Nach dem schwachen Gesetz für große Zahlen gilt $T_n = 2\frac{X_1+\cdots+X_n}{n} \xrightarrow{i.W.} 2E_\vartheta[X] = \vartheta$, also ist T_n konsistent.

c. $\mathrm{Var}_\vartheta[T_n] = (\frac{2}{n})^2 n \mathrm{Var}_\vartheta[X] = \frac{4}{n^2} n \frac{\vartheta^2}{12} = \frac{\vartheta^2}{3n}$.

Idee II. Wir nehmen als Schätzer

$$\widetilde{T}_n(X_1, \ldots, X_n) = \max(X_1, \ldots, X_n).$$

a. Wegen $\widetilde{T}_n \le \vartheta$ ist $\widetilde{T}_n$ sicher nicht erwartungstreu. Es ist aber

$$P_\vartheta(\widetilde{T}_n \le x) = P_\vartheta(X_1 \le x \wedge \ldots \wedge X_n \le x) = \prod_{i=1}^n P_\vartheta(X_i \le x) = (\frac{x}{\vartheta})^n \ (x \in [0, \vartheta]),$$

also ist die Dichte $\widetilde{f}(x) = \frac{nx^{n-1}}{\vartheta^n}$, und wir erhalten

$$E_\vartheta[\widetilde{T}_n] = \int_0^\vartheta x \frac{nx^{n-1}}{\vartheta^n} dx = \frac{nx^{n+1}}{(n+1)\vartheta^n}\Big|_0^\vartheta = \frac{n}{n+1}\vartheta \overset{n\to\infty}{\longrightarrow} \vartheta,$$

das heißt, $\widetilde{T}_n$ ist asymptotisch erwartungstreu.

b. Für die Varianz haben wir

$$E_\theta[\widetilde{T}_n^2] = \int_0^\vartheta x^2 \frac{nx^{n-1}}{\vartheta^n} dx = \frac{nx^{n+2}}{(n+2)\vartheta^n}\Big|_0^\vartheta = \frac{n}{n+2}\vartheta^2,$$

also
$$\mathrm{Var}_\vartheta[\widetilde{T}_n] = \frac{n}{n+2}\vartheta^2 - (\frac{n}{n+1}\vartheta)^2 = \frac{n}{(n+1)^2(n+2)}\vartheta^2\,.$$

c. Nun nehmen wir die Modifikation $T_n^* = \frac{n+1}{n}\widetilde{T}_n$. Dann ist $E_\vartheta[T_n^*] = \frac{n+1}{n}E_\vartheta[\widetilde{T}_n]$ $= \vartheta$, also ist T_n^* erwartungstreu. Für die Varianz gilt

$$\mathrm{Var}[T_n^*] = \frac{(n+1)^2}{n^2}\mathrm{Var}[\widetilde{T}_n] = \frac{\vartheta^2}{n(n+2)}\,,$$

also $\mathrm{Var}_\vartheta[T_n^*] < \mathrm{Var}_\vartheta[T_n]$ für $n \geq 2$ und alle $\vartheta \in \theta$.
T_n^* ist also ein besserer Schätzer als T_n.

Die nächste Idee führt zum wichtigsten allgemeinen Schätzer, dem Maximum Likelihood Schätzer. Wenn wir x beobachten, so ist im diskreten Fall $P_\vartheta(X = x)$ die Wahrscheinlichkeit, dass x eintritt, falls ϑ der richtige Parameter ist. Ein ϑ mit kleiner W-keit $P_\vartheta(X = x)$ wird also nicht der richtige Parameter sein.

Schätzregel. Man bestimme $T(x)$ zu x so, dass

$$P_{T(x)}(X = x) = \max_{\vartheta\in\theta} P_\vartheta(X = x)\,.$$

$T(x)$ heißt *Maximum Likelihood Schätzer*, kurz ML-Schätzer.

Beispiel. Analysieren wir das Beispiel mit der Orangenlieferung. Hier ist

$$P_\vartheta(X = x) = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{N}{n}} = \max$$

zu bilden.

Wir haben

$$\frac{P_\vartheta(X = x)}{P_{\vartheta-1}(X = x)} = \frac{\binom{\vartheta}{x}}{\binom{\vartheta-1}{x}}\frac{\binom{N-\vartheta}{n-x}}{\binom{N-\vartheta+1}{n-x}} = \frac{\vartheta}{\vartheta - x}\,\frac{N-\vartheta-n+x+1}{N-\vartheta+1} \geq 1$$

für
$$\vartheta N - \vartheta^2 - \vartheta n + \vartheta x + \vartheta \geq \vartheta N - \vartheta^2 + \vartheta - Nx + \vartheta x - x$$

also für
$$-\vartheta n \geq -Nx - x$$

das heißt für

$$\vartheta \leq \frac{x(N+1)}{n}\,.$$

Der ML-Schätzer ist daher

$$T(x) = \lfloor \frac{x(N+1)}{n} \rfloor \sim \frac{xN}{n}\,,$$

er entspricht also bis auf Rundung dem naiven Schätzer $\frac{xN}{n}$.

Allgemein gehen wir folgendermaßen vor. Für diskrete Zufallsvariablen wollen wir bei gegebenem x die Funktion $P_\vartheta(X = x)$ in ϑ maximieren, und für stetige Variable die Dichte $f_\vartheta(x)$. Die *Likelihood Funktion* ist

$$\rho : \mathcal{X} \times \theta \to [0, \infty) \text{ mit } \rho(x, \vartheta) = \begin{cases} P_\vartheta(X = x) & X \text{ diskret} \\ f_\vartheta(x) & X \text{ stetig.} \end{cases}$$

Definition. $T(x)$ heißt *ML-Schätzer*, falls

$$\rho(x, T(x)) = \max_{\vartheta \in \theta} \rho(x, \vartheta).$$

Beispiele. 1. Ein Bernoulli Experiment mit $P(X = 1) = \vartheta$, $P(X = 0) = 1 - \vartheta$, $0 < \vartheta < 1$, wird n Mal wiederholt. Gesucht ist ein Schätzer für ϑ. Sei x die Anzahl der Erfolge. Wir haben

$$\mathcal{X} = \{0, 1, \ldots, n\},\ \theta = (0, 1),\ P_\vartheta(X = x) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x}.$$

Um $\max\limits_{\vartheta \in \theta} P_\vartheta(X = x)$ zu berechnen, maximieren wir den Logarithmus und berechnen

$$\begin{aligned} \frac{d}{d\vartheta} \log P_\vartheta(X = x) &= \frac{d}{d\vartheta}\left[\log \binom{n}{x} + x \log \vartheta + (n - x) \log(1 - \vartheta)\right] \\ &= \frac{x}{\vartheta} - \frac{n - x}{1 - \vartheta}\,, \end{aligned}$$

und diese Funktion ist monoton fallend in ϑ. Das Maximum ergibt sich also für $\frac{d}{d\vartheta} \log P_\vartheta(X = x) = 0$, und wir erhalten

$$\frac{x}{\vartheta} = \frac{n - x}{1 - \vartheta} \Leftrightarrow x - x\vartheta = n\vartheta - x\vartheta \Leftrightarrow \vartheta = \frac{x}{n}\,.$$

Der naive Schätzer $T_n(x) = \frac{x}{n}$ ist also auch der ML-Schätzer. Wir haben ferner

$$\begin{aligned} E_\vartheta(T_n) &= E_\vartheta(\frac{X_1 + \cdots + X_n)}{n}) = \vartheta \\ \mathrm{Var}_\vartheta(T_n) &= \mathrm{Var}_\vartheta(\frac{X_1 + \cdots + X_n}{n}) = \frac{\vartheta(1-\vartheta)}{n} . \end{aligned}$$

2. Betrachten wir nochmals das Beispiel der Ziehung von n Zahlen aus $[0, \vartheta]$, $\vartheta > 0$ mit Gleichverteilung. Hier ist

$$f_\vartheta(x_1, \ldots, x_n) = \begin{cases} \frac{1}{\vartheta^n} & \text{falls } x_1, \ldots, x_n \leq \vartheta \\ 0 & \text{sonst} . \end{cases}$$

Der ML-Schätzer ist daher $\widetilde{T}_n(x) = \max(x_1, \ldots, x_n)$, da $\vartheta \geq x_1, \ldots, x_n$ möglichst klein sein soll, um $f_\vartheta(x_1, \ldots, x_n)$ zu maximieren.

3. Ausfallswahrscheinlichkeit von Geräten
Es wird angenommen, dass die Lebensdauer von Glühbirnen exponential verteilt ist mit $f_\vartheta(x) = \vartheta e^{-\vartheta x}$, ϑ unbekannt. Es werden n Stichproben gezogen. Wir haben also

$$\mathcal{X} = [0, \infty), \ f_\vartheta(x) = \vartheta e^{-\vartheta x}, f_\vartheta(x_1, \ldots, x_n) = \vartheta^n e^{-\vartheta(x_1 + \cdots + x_n)} .$$

Um das Maximum zu berechnen, logarithmieren wir wie eben und erhalten

$$\begin{aligned} \frac{d}{d\vartheta} \log f_\vartheta(x_1, \ldots, x_n) &= \frac{d}{d\vartheta}(n \log \vartheta - \vartheta(x_1 + \cdots + x_n)) \\ &= \frac{n}{\vartheta} - (x_1 + \cdots + x_n) . \end{aligned}$$

Der ML-Schätzer ist daher

$$T_n(x_1, \ldots, x_n) = \frac{n}{x_1 + \cdots + x_n} = \frac{1}{\bar{x}} ,$$

wobei $\bar{x} = \frac{x_1 + \cdots + x_n}{n}$ der Durchschnittswert ist.
Ist $\tau(\vartheta)$ die W-keit für den Ausfall der Glühbirne bis zur Zeit t, so haben wir

$$P_\vartheta(X \leq t) = \int_0^t \vartheta e^{-\vartheta} dx = 1 - e^{-\vartheta t} .$$

Der ML-Schätzer dafür ist also $T_n(x_1, \ldots, x_n) = 1 - e^{-\frac{t}{\bar{x}}}$.

4. Sei $X \sim N(\mu, \sigma^2)$ normalverteilt. Wir wollen μ, σ^2 schätzen, und führen dazu n Messungen durch. Also

$$\mathcal{X} = \mathbb{R},\ f_{\mu,\sigma^2}(x_1, \dots, x_n) = (\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}.$$

Logarithmieren ergibt

$$\log f_{\mu,\sigma^2}(x_1, \dots, x_n) = -n\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2.$$

Fall 1. σ^2 bekannt, μ unbekannt.
Mit

$$\frac{d}{d\mu}\log f_{\mu,\sigma^2}(x_1, \dots, x_n) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu)\ (= 0)$$

erhalten wir den ML-Schätzer $\widehat{\mu} = \overline{x} = \frac{x_1 + \cdots + x_n}{n}$.
Fall 2. μ bekannt, σ^2 unbekannt.
Mit

$$\frac{d}{d\sigma}\log f_{\mu,\sigma^2} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n (x_i - \mu)^2\ (= 0)$$

erhalten wir $n = \frac{1}{\sigma^2}\sum\limits_{i=1}^n (x_i - \mu)^2$, also den ML-Schätzer

$$\widehat{\sigma^2} = \frac{\sum\limits_{i=1}^n (x_i - \mu)^2}{n}.$$

Fall 3. μ, σ unbekannt.
Dann nehmen wir $\widehat{\mu} = \overline{x}$, $\widehat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n}$ und haben

$$E[\widehat{\sigma}^2] = \frac{E[\sum_{i=1}^n (X_i - \overline{X})^2]}{n} = \frac{\sum_{i=1}^n E[(X_i - \overline{X})^2]}{n}$$

und berechnet leicht (beachte $\sigma^2 = \mathrm{Var}[X]$)

$$E[\widehat{\sigma}^2] = \frac{n-1}{n}\sigma^2.$$

Der normierte Schätzer $\frac{\sum (x_i - \overline{x})^2}{n-1}$ ist also erwartungstreuer Schätzer für σ^2.

Schließlich wollen wir noch *beste* Schätzer analysieren. Wir machen die folgenden Annahmen:

1. Es sei $\theta \subseteq \mathbb{R}$ ein offenes Intervall.

2. Die Likelihood Funktion $\rho(x, \vartheta)$ ist auf $\mathcal{X} \times \theta$ positiv und nach ϑ stetig differenzierbar.

3. Es gilt die Vertauschungsrelation
$$\int\limits_{\mathcal{X}} \frac{d}{d\vartheta} \rho(x, \vartheta) dx = \frac{d}{d\vartheta} \int\limits_{\mathcal{X}} \rho(x, \vartheta) dx \,.$$
Wenn $\mathcal{X}$ diskret ist, so wird das Integral durch die Summe $\sum\limits_{\mathcal{X}}$ ersetzt.

4. Sei $U_\vartheta(x) = \frac{d}{d\vartheta} \log \rho(x, \vartheta) = \frac{\frac{d}{d\vartheta} \rho(x,\vartheta)}{\rho(x,\vartheta)}$. Für jedes $\vartheta \in \theta$ existiert $\mathrm{Var}_\vartheta[U_\vartheta(X)]$ und ist $\neq 0$.

Ein Modell $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$, das diese Bedingungen erfüllt, heißt *regulär*. Es soll $\tau(\vartheta)$ geschätzt werden. Ein Schätzer $T(x)$ heißt *regulär*, falls für alle ϑ
$$\int\limits_{\mathcal{X}} T(x) \frac{d}{d\vartheta} \rho(x, \vartheta) dx = \frac{d}{d\vartheta} \int\limits_{\mathcal{X}} T(x) \rho(x, \vartheta) dx$$
gilt.

Satz 4.1. *Gegeben ein reguläres Modell $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$ und ein erwartungstreuer regulärer Schätzer T für $\tau(\vartheta)$. Dann gilt*
$$\mathrm{Var}_\vartheta[T] \geq \frac{\tau'(\vartheta)^2}{\mathrm{Var}_\vartheta[U_\vartheta]} \quad \textit{für alle } \vartheta \in \theta \,.$$
Gleichheit für alle ϑ gilt genau dann, wenn
$$T = \tau(\vartheta) + \frac{\tau'(\vartheta) U_\vartheta}{\mathrm{Var}[U_\vartheta]} \quad \textit{für alle } \vartheta$$
ist.

Beweis. Sei T regulärer erwartungstreuer Schätzer. Wir haben

$$\begin{aligned} E_\vartheta[U_\vartheta] &= \int \frac{\frac{d}{d\vartheta}\rho(x,\vartheta)}{\rho(x,\vartheta)}\rho(x,\vartheta)dx = \int \frac{d}{d\vartheta}\rho(x,\vartheta)dx \\ &= \frac{d}{d\vartheta}\int \rho(x,\vartheta)dx = \frac{d}{d\vartheta}1 = 0\,. \end{aligned}$$

Für die Covarianz $\mathrm{cov}[T, U_\vartheta]$ erhalten wir mit $E_\vartheta[T] = \tau(\vartheta)$

$$\begin{aligned} \mathrm{cov}_\vartheta[T,U_\vartheta] &= E_\vartheta[TU_\vartheta] - E_\vartheta[T]E_\vartheta[U_\vartheta] \\ &= \int T(x)\frac{\frac{d}{d\vartheta}\rho(x,\vartheta)}{\rho(x,\vartheta)}\rho(x,\vartheta)dx \\ &= \frac{d}{d\vartheta}\int T(x)\rho(x,\vartheta)dx = \frac{d}{d\vartheta}E_\vartheta[T] \\ &= \frac{d}{d\vartheta}\tau(\vartheta) = \tau'(\vartheta)\,. \end{aligned}$$

Mit Lemma 2.8 folgt daraus

$$\begin{aligned} 0 \le \mathrm{Var}_\vartheta\left[T - \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}\right] &= \mathrm{Var}_\vartheta[T] + \frac{\tau'(\vartheta)^2\mathrm{Var}[U_\vartheta]}{(\mathrm{Var}_\vartheta[U_\vartheta])^2} - \frac{2\tau'(\vartheta)}{\mathrm{Var}_\vartheta[U_\vartheta]}\underbrace{\mathrm{cov}_\vartheta[T,U_\vartheta]}_{\tau'(\vartheta)} \\ &= \mathrm{Var}_\vartheta[T] - \frac{\tau'(\vartheta)^2}{\mathrm{Var}_\vartheta[U_\vartheta]}\,, \end{aligned}$$

also

$$\mathrm{Var}_\vartheta[T] \ge \frac{\tau'(\vartheta)^2}{\mathrm{Var}_\vartheta[U_\vartheta]}\,.$$

Gleichheit gilt genau dann, wenn $\mathrm{Var}_\theta[T - \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}[U_\vartheta]}] = 0$ ist. Nach Lemma 2.9 bedeutet dies

$$P_\vartheta\left(T - \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]} = E_\vartheta\left[T - \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}\right]\right) = 1\,.$$

Nun ist $E_\vartheta[T] = \tau(\vartheta)$, $E[\frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}] = 0$. Gleichheit gilt also genau dann, wenn

$$P_\vartheta\left(T = \tau(\vartheta) + \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}\right) = 1$$

ist. Da $\rho(x,\vartheta) > 0$ ist, folgt daraus sofort $T = \tau(\vartheta) + \frac{\tau'(\vartheta)U_\vartheta}{\mathrm{Var}_\vartheta[U_\vartheta]}$ im diskreten Fall. Im stetigen Fall kommt man aus Stetigkeitsgründen auf dasselbe Ergebnis. $\square$

Beispiele. 1. Eine Münze mit $P(K) = \vartheta$, $P(Z) = 1-\vartheta$, $0 < \vartheta < 1$, wird n Mal geworfen, x sei die Anzahl von Kopf, ϑ soll geschätzt werden. Hier haben wir $\mathcal{X} = \{0, 1, \ldots, n\}$, $\vartheta = (0, 1)$,

$$\begin{aligned}\rho(x,\vartheta) &= \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x} > 0\,,\\ U_\vartheta(X) &= \frac{X}{\vartheta} - \frac{n-X}{1-\vartheta} = \frac{X}{\vartheta(1-\vartheta)} - \frac{n}{1-\vartheta}\,,\end{aligned}$$

$$\mathrm{Var}_\vartheta[U_\vartheta(X)] = \frac{1}{\vartheta^2(1-\vartheta)^2}\mathrm{Var}[X] = \frac{n\vartheta(1-\vartheta)}{\vartheta^2(1-\vartheta)^2} = \frac{n}{\vartheta(1-\vartheta)} \neq 0\,.$$

Da $\mathcal{X}$ endlich ist, ist die Vertauschungsrelation trivialerweise erfüllt, $(\mathcal{X}, 2^{\mathcal{X}}, P_\vartheta : \vartheta \in \theta)$ ist also ein reguläres Modell. Es folgt für jeden erwartungstreuen Schätzer T (die Regularität ist wiederum erfüllt)

$$\mathrm{Var}_\vartheta(T) \geq \frac{1}{\mathrm{Var}_\vartheta[U_\vartheta]} = \frac{\vartheta(1-\vartheta)}{n}\,.$$

Der ML-Schätzer $\frac{x}{n}$ ist also bester Schätzer.

2. Es sei $(\mathcal{X}, 2^{\mathcal{X}}, P_\vartheta : \vartheta \in \theta)$, wobei $\mathcal{X} = \{0, 1, 2, \ldots\}$ und P_ϑ die Familie der Poissonverteilungen mit $\rho(x,\vartheta) = e^{-\vartheta}\frac{\vartheta^x}{x!}$ ist. Es soll ϑ geschätzt werden. Wir haben $\theta = (0,\infty)$, $\rho(x,\vartheta) > 0$. Wir verwenden den Schätzer $\widetilde{T}(x) = x$. Da $E_\vartheta[X] = \vartheta$ ist, so ist $\widetilde{T}$ erwartungstreu, außerdem ist $\mathrm{Var}_\vartheta[\widetilde{T}] = \vartheta$.

Für U_ϑ berechnen wir $U_\vartheta(X) = -\vartheta + \frac{X}{\vartheta} \neq 0$, $\mathrm{Var}_\vartheta[U_\vartheta(X)] = \frac{\mathrm{Var}[X]}{\vartheta^2} = \frac{1}{\vartheta} \neq 0$, und die Vertauschungsrelation ist leicht verifiziert. Nach dem Satz gilt für jeden erwartungstreuen Schätzer T

$$\mathrm{Var}_\vartheta[T] \geq \frac{1}{1/\vartheta} = \vartheta\,,$$

$\widetilde{T}(x) = x$ ist also ein bester Schätzer.

4.4 Konfidenzintervalle

Sei das Modell $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$ gegeben, ϑ oder allgemein $\tau(\vartheta)$ soll geschätzt werden.

Definition. Die Abbildung $C : \mathcal{X} \to 2^{\mathbb{R}}$, $x \mapsto C(x)$ Intervall $\subseteq \mathbb{R}$, heißt *Konfidenzabbildung zum Irrtumsniveau* α, $0 \leq \alpha \leq 1$, falls

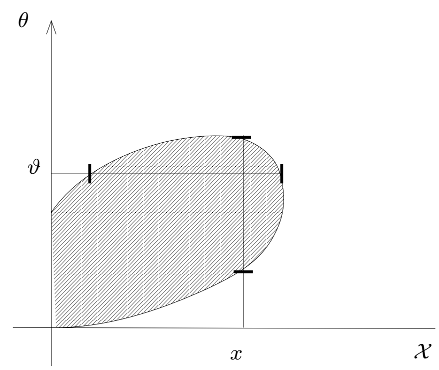
$$P_\vartheta(\{x \in \mathcal{X} : \vartheta \in C(x)\}) \geq 1 - \alpha \text{ für alle } \vartheta$$

gilt, also

$$\inf_\vartheta P_\vartheta(\{x \in \mathcal{X} : \vartheta \in C(x)\}) \geq 1 - \alpha\,.$$

Wir wollen nun eine Konstruktion solcher Konfidenzintervalle angeben. Natürlich soll $C(x)$ möglichst klein sein.

Sei $C = \{(x, \vartheta) \in \mathcal{X} \times \theta : \vartheta \in C(x)\}$, etwa



Für $x \in \mathcal{X}$ ist $C(x)$ der vertikale Schnitt. Für $\vartheta \in \theta$ ist

$$C_\vartheta = \{x \in \mathcal{X} : (x, \vartheta) \in C\}$$

der horizontale Schnitt, mit

$$P_\vartheta(C_\vartheta) = P_\vartheta(\{x \in \mathcal{X} : \vartheta \in C(x)\})\,.$$

Wir verlangen also

$$\inf_\vartheta P_\vartheta(C_\vartheta) \geq 1 - \alpha\,.$$

Die Konstruktion erfolgt in zwei Schritten:

A. Zu $\vartheta \in \theta$ bestimme ein möglichst kleines C_ϑ mit $P_\vartheta(C_\vartheta) \geq 1 - \alpha$, z.B. $C_\vartheta = \{x \in \mathcal{X} : \rho(x, \vartheta) \geq c_\vartheta\}$, wobei c_ϑ so bestimmt ist, dass $P_\vartheta(C_\vartheta) \geq 1 - \alpha$ erfüllt ist.

B. Setze $C = \{(x, \vartheta) \in \mathcal{X} \times \theta : x \in C_\vartheta\}$, und dann

$$C(x) = \{\vartheta \in \theta : x \in C_\vartheta\}.$$

Dann gilt $\inf\limits_\vartheta P_\vartheta(C_\vartheta) \geq 1 - \alpha$.

Beispiele. 1. Wir werfen eine Münze n Mal mit $P(K) = \vartheta$, $P(Z) = 1 - \vartheta$, $0 < \vartheta < 1$, $x =$Anzahl Kopf. Wir wissen, dass $\frac{x}{n}$ ein bester Schätzer für ϑ ist. Wir wählen $C(x) = (\frac{x}{n} - \varepsilon, \frac{x}{n} + \varepsilon)$ mit $\varepsilon > 0$, so dass

$$P_\vartheta(\{x : |\frac{x}{n} - \vartheta| < \varepsilon\}) \geq 1 - \alpha \Leftrightarrow P_\vartheta(\{x : |\frac{x}{n} - \vartheta| \geq \varepsilon) \leq \alpha\,.$$

Wie erreichen wir das? Nach der Ungleichung von Tschebyschev haben wir

$$P_\vartheta(\{x : |\frac{x}{n} - \vartheta| \geq \varepsilon) \leq \frac{\vartheta(1-\vartheta)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}\,.$$

Ein $\varepsilon > 0$ funktioniert also, sobald

$$\frac{1}{4n\varepsilon^2} \leq \alpha \Longleftrightarrow \varepsilon \geq \frac{1}{2\sqrt{n\alpha}}$$

gilt.

Ist zum Beispiel $n = 1000$, $\alpha = 0.025$, so genügt $\varepsilon = 0,1$. Wollen wir ein Konfidenzintervall mit Radius $\varepsilon = 0,01$ erhalten, so muß $\sqrt{n} \geq \frac{1}{2\varepsilon\sqrt{\alpha}}$ sein, also $n \geq \frac{1}{4\varepsilon^2\alpha} = 100.000$.

Mit de Moivre-Laplace haben wir

$$\begin{aligned} P_\vartheta(\{x : |\frac{x}{n} - \vartheta| < \varepsilon\}) &= P_\vartheta\left(\left\{x : |\frac{x - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}| < \varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right\}\right) \\ &\approx \phi\left(\varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) - \phi\left(-\varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) \\ &= 2\phi\left(\varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) - 1\,, \end{aligned}$$

da $\phi(-x) = 1 - \phi(x)$ ist.
Für $\alpha = 0,025$ ergibt dies $\varepsilon \geq 0,0446$ bei $n = 1000$.

2. Wir wollen den Prozentsatz der Wähler einer Partei A schätzen. Werden n Wähler befragt und x sind Wähler von A, so nehmen wir als Schätzer $\frac{x}{n}$ für die W-keit ϑ, dass ein zufällig ausgewählter Wähler für A stimmt. Wie groß muß n sein, dass die W-keit eines Irrtums von höchstens 1% nicht mehr als $0,05$ beträgt?
Wir verlangen also

$$P_\vartheta(|\frac{x}{n} - \vartheta| \le 0,01) \ge 0,95.$$

Durch Zentrieren ist dies äquivalent zu

$$P_\vartheta\left(|\frac{x - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}| \le 0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) \ge 0,95\,,$$

also

$$\phi\left(0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) - \phi\left(-0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) = 2\phi\left(0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) - 1 \ge 0,95\,,$$

somit

$$\phi\left(0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) \ge 0,975\,.$$

Aus der Tabelle ergibt sich

$$0,01\sqrt{\frac{n}{\vartheta(1-\vartheta)}} \ge 1,96\,,$$

das heißt

$$n \ge 38416 \cdot \vartheta(1-\vartheta)\,.$$

Da $\vartheta(1-\vartheta) \le \frac{1}{4}$ ist, genügt jedenfalls $n \ge 9604$. Falls wir von vorneherein wissen, dass $\vartheta \le 0,1$ ist, so genügt $n \ge 38416 \cdot 0,09 \approx 3458$.

3. Die Familie der Verteilungen sei $N(\mu, \sigma^2)$, wobei μ der unbekannte Parameter ist, $\sigma^2 > 0$ ist bekannt. Wir machen n Beobachtungen $x_1, \ldots, x_n$, dann ist $X_1 + \cdots + X_n$ nach $N(n\mu, n\sigma^2)$ verteilt, also

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} \sim N(\mu, \frac{\sigma^2}{n}),\ Z = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim N(0,1)\,.$$

Sei t_α so gewählt, dass gilt

$$P_\mu(|Z| \le t_\alpha) \ge 1 - \alpha \Longleftrightarrow P_\mu(|Z| \ge t_\alpha) \le \alpha\,.$$

Wir nehmen den Schätzer $\overline{x} = \frac{x_1+\cdots+x_n}{n}$ und wählen das Intervall

$$C(x) = (\overline{x} - \frac{t_\alpha \sigma}{\sqrt{n}},\ \overline{x} + \frac{t_\alpha \sigma}{\sqrt{n}})\,,$$

das heißt

$$\mu \in C(x) \Longleftrightarrow |\frac{\sqrt{n}}{\sigma}(\overline{x} - \mu)| \leq t_\alpha\,.$$

Es folgt $P_\mu(\{x : \mu \in C(x)\}) \geq 1 - \alpha$, also ist $C(x)$ Konfidenzintervall zum Niveau α.

Wählen wir zum Beispiel $\alpha = 0,05$, dann berechnet man $t_\alpha = 1,96$, also ist $C(x) = (\overline{x} - \frac{2\sigma}{\sqrt{n}}, \overline{x} + \frac{2\sigma}{\sqrt{n}})$ Konfidenzintervall für den Mittelwert μ.

Bei unbekannten μ und σ^2 nimmt man als Konfidenzintervall

$$C(x) = (\overline{x} - \frac{2s}{\sqrt{n}},\ \overline{x} + \frac{2s}{\sqrt{n}})$$

wobei $s^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n}$ ist.

4.5 Testen von Hypothesen

Wir wiederholen nochmals die Situation in den folgenden fünf Schritten:

1. Das Modell $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$ wird festgelegt.

2. Die Hypothesen werden formuliert:

   $\theta = \theta_0 \dot{\cup} \theta_1$,
   $H_0 : \vartheta \in \theta_0$ Nullhypothese,
   $H_1 : \vartheta \in \theta_1$ Alternativhypothese.

3. Das Irrtumsniveau α wird gewählt, $0 < \alpha < 1$ meist $\alpha = 0,1$ oder $0,05$ oder $0,025$.

   Irrtum 1. Art: $\vartheta \in \theta_0$, aber H_1 wird angenommen

   Irrtum 2. Art: $\vartheta \in \theta_1$, aber H_0 wird angenommen.

   Ein Irrtum 1. Art soll höchstens mit W-keit α vorkommen.

4. Eine Entscheidungsregel wird festgelegt. Man wählt einen Test $\varphi : \mathcal{X} \to [0,1]$, φ ist Zufallsvariable.

   Deterministischer Test:

$$\varphi(x) = \begin{cases} 0 & H_0 \text{ wird angenommen} \\ 1 & H_1 \text{ wird angenommen.} \end{cases}$$

   Randomisierter Test: $\varphi(x) \in [0,1]$ ist die W-keit, mit der man sich für H_1 entscheidet.

5. Jetzt erst wird das Experiment durchgeführt!

Definition. Sei $(\mathcal{X}, \mathcal{E}, P_\vartheta : \vartheta \in \theta)$, $\theta = \theta_0 \mathbin{\dot\cup} \theta_1$, $\varphi : \mathcal{X} \to [0,1]$ gegeben. $A = \{x \in \mathcal{X} : \varphi(x) = 0\}$ heißt *Annahmebereich*, $R = \{x \in \mathcal{X} : \varphi(x) = 1\}$ *Ablehnungsbereich* (von H_0). Falls φ deterministischer Test ist, so heißt

$$G_\varphi : \theta \to [0,1],\ G_\varphi(\vartheta) = P_\vartheta(\varphi \in R) \ \textit{Gütefunktion.}$$

Falls φ randomisierter Test ist, so ist die Gütefunktion

$$G_\varphi(\vartheta) = E_\vartheta[\varphi]\,.$$

Natürlich ist $G_\varphi(\vartheta) = E_\vartheta[\varphi]$ auch für deterministische Tests.

Die Forderungen an die Gütefunktion sind demnach:

$$\begin{aligned} \vartheta \in \theta_0 &\implies G_\varphi(\vartheta) \le \alpha \quad \text{Irrtum 1. Art,} \\ \vartheta \in \theta_1 &\implies G_\varphi(\vartheta) \text{ möglichst groß } (1 - G_\varphi(\vartheta) \text{ Irrtum 2. Art),} \end{aligned}$$

und wir nennen dann φ einen Test zum *Niveau* α.

Definition. $\varphi : \mathcal{X} \to [0,1]$ heißt *bester Test* zum Niveau α, wenn für alle Tests ψ zum Niveau α gilt

$$G_\varphi(\vartheta) \ge G_\psi(\vartheta) \text{ für alle } \vartheta \in \theta_1\,.$$

Beispiel. Tea tasting lady. Die Dame behauptet, sie könne feststellen, ob zuerst Tee und dann Milch in die Tasse gegeben wurde, oder umgekehrt. Es seien $n = 8$ Tassen und 4 von jedem Typ, x sei die Anzahl der Treffer. Wir haben also folgende Situation:

1. $\mathcal{X} = \{0, 1, \ldots, 8\}$, $P_\vartheta(X = x) = \binom{8}{x}\vartheta^x(1-\vartheta)^{8-x}$,

2. $\theta = [\frac{1}{2}, 1]$,

   $H_0 : \theta_0 = \{\frac{1}{2}\}$, sie hat nicht die Fähigkeit

   $H_1 : \theta_1 = (\frac{1}{2}, 1]$, sie kann tatsächlich die Zusammensetzung richtig bestimmen.

3. $\alpha = 0,05$.

4. $\varphi(x) = \begin{cases} 0 & x < c \Rightarrow H_0 \\ 1 & x \geq c \Rightarrow H_1 \end{cases}$ deterministisch.

Zur richtigen Wahl von c berechnen wir

$$G_\varphi(\frac{1}{2}) = P_{\frac{1}{2}}(X \geq c) = \sum_{k=c}^{8} \binom{8}{k} \frac{1}{2^8} \leq 0,05$$

und das gilt für $c \geq 7$. Also ist die Entscheidungsregel

$$\begin{aligned} x < 7 &\Rightarrow H_0 \\ x \geq 7 &\Rightarrow H_1 . \end{aligned}$$

Für $\vartheta \in \theta_1$, das heißt $\vartheta > \frac{1}{2}$, haben wir

$$G_\varphi(\vartheta) = \sum_{k=7}^{8} \binom{8}{k} \vartheta^k (1-\vartheta)^{8-k} = 8\vartheta^7(1-\vartheta) + \vartheta^8 = \vartheta^7(8 - 7\vartheta),$$

$G_\varphi(\vartheta)$ ist monoton steigend. Der Irrtum 2. Art beträgt $1 - \vartheta^7(8 - 7\vartheta)$. Zum Beispiel erhalten wir für $\vartheta = \frac{3}{4}$ (die Lady bestimmt im Mittel 6 von 8 Tassen richtig), $G_\varphi(\frac{3}{4}) \sim 0,367$, also ist der Irrtum 2. Art $0,633$. Die Lady hat keine Chance, weil die Stichprobenzahl n zu klein ist.

Definition. Ein Test φ heißt *unverfälscht* (unbiased) zum Niveau α, falls

$$G_\varphi(\vartheta_0) \leq \alpha \leq G_\varphi(\vartheta_1) \text{ für alle } \vartheta_0 \in \theta_0, \vartheta_1 \in \theta_1$$

gilt.

Das Verfahren bei der Tea tasting lady ist biased, da

$$G_\varphi(\frac{1}{2}) = \frac{9}{2^8} < 0,05$$

ist, also $G_\varphi(\frac{1}{2} + \varepsilon) < \alpha = 0,05$ für kleine $\varepsilon > 0$.

Wir studieren nun im Detail sogenannte *Alternativtests*. Es gibt nur zwei Verteilungen P_0, P_1,

$$(\mathcal{X}, \mathcal{E}, P_0 \cup P_1), \ \theta = \{0, 1\},$$

$$H_0 : \theta_0 = \{0\}, \ H_1 : \theta_1 = \{1\} \,.$$

Wir setzen für die Likelihood Funktion $\rho(x, \vartheta)$

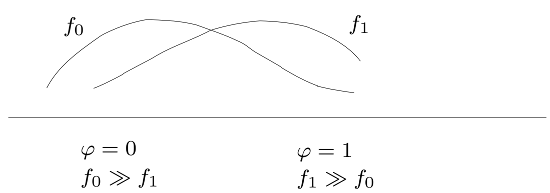
$$\rho_0(x) = \rho(x, 0), \rho_1(x) = \rho(x, 1)$$

und setzen voraus $\rho_0(x) + \rho_1(x) > 0$ für alle $x \in \mathcal{X}$, das heißt:

$P_0(X = x) + P_1(X = x) > 0$	falls X diskrete Variable ist
$f_0(x) + f_1(x) > 0$	falls X stetige Variable ist, mit Dichten $f_0(x)$, $f_1(x)$.

Dies ist keine Einschränkung, da wir ansonsten $\mathcal{X}$ kleiner machen können.

Die Idee ist, dass wir H_0 beibehalten, falls f_0 größer als f_1 ist bzw. H_1 nehmen, falls f_1 größer als f_0 ist:



Definition. Der *Likelihood Quotient* ist

$$R(x) = \begin{cases} \frac{\rho_1(x)}{\rho_0(x)} & (\rho_0(x) > 0) \\ \infty & (\rho_0(x) = 0) \end{cases} .$$

Wenn also $R(x) > c$ für geeignetes c ist, so ist die Tendenz zu H_1 “hinreichend stark”.

Definition. Sei $(\mathcal{X}, \mathcal{E}, P_0 \cup P_1)$, $\theta = \{0, 1\}$ gegeben. $\varphi^* : \mathcal{X} \to [0, 1]$ heißt *Neyman-Pearson Test*, wenn es eine Konstante $c^* > 0$ gibt mit

$$\varphi^*(x) = \begin{cases} 0 & R(x) < c^* \\ 1 & R(x) > c^* \\ \gamma(x) & R(x) = c^*, \ \gamma(x) \text{ beliebig in } [0, 1] \,. \end{cases}$$

Satz 4.2 (Neyman-Pearson). *Gegeben* $(\mathcal{X}, \mathcal{E}, P_0 \cup P_1)$, $\theta = \{0,1\}$, $0 < \alpha < 1$, $\rho_0(x) + \rho_1(x) > 0$ *für alle* $x \in \mathcal{X}$. *Dann gilt:*

a. *Jeder Neyman-Pearson Test* φ^* *mit* $E_0[\varphi^*] = \alpha$ *ist bester Test.*

b. *Es gibt einen Neyman-Pearson Test* φ^* *mit* $E_0[\varphi^*] = \alpha$.

c. *Jeder beste Test zum Niveau* α *ist Neyman-Pearson Test, bis auf eine Menge vom Maß 0.*

Beweis. a. Sei $\varphi*$ Neyman-Pearson Test mit $c^* > 0$ und φ ein beliebiger Test mit $E_0[\varphi^*] = \alpha$, $E_0[\varphi] \leq \alpha$. Zu zeigen ist $E_1[\varphi] \leq E_1[\varphi^*]$.

Sei $g = (\rho_1 - c^*\rho_0)(\varphi^* - \varphi)$. Für $R(x) = \frac{\rho_1(x)}{\rho_0(x)} < c^*$ haben wir $\rho_1(x) - c^*\rho_0(x) < 0$, $\varphi^*(x) = 0$, also $\varphi^*(x) - \varphi(x) \leq 0$, das heißt $g(x) \geq 0$.
Für $R(x) > c^*$ ist $\rho_1(x) - c^*\rho_0(x) > 0$ (wegen $\rho_1(x) + \rho_0(x) > 0$) und $\varphi^*(x) = 1$, $\varphi^*(x) - \varphi(x) \geq 0$, also wiederum $g(x) \geq 0$. Für $R(x) = c^*$ ist $\rho_1(x) - c^*\rho_0(x) = 0$, also $g(x) = 0$. Daraus folgt (für diskrete Variablen nehme man die Summe)

$$\begin{aligned}
0 \leq \int g(x)dx &= \int (\varphi^*(x) - \varphi(x))(\rho_1(x) - c^*\rho_0(x))dx \\
&= E_1[\varphi^* - \varphi] - c^* E_0[\varphi^* - \varphi] \\
&= E_1[\varphi*] - E_1[\varphi] - c^*(E_0 \underbrace{[\varphi^*] - E_0[}_{\geq 0}\varphi]) \\
&\leq E_1[\varphi^*] - E_1[\varphi]\,,
\end{aligned}$$

also $E_1[\varphi^*] \geq E_1[\varphi]$.

b. Für $c \geq 0$ setzen wir

$$\alpha(c) = P_0(R(X) > c),\ \overline{\alpha}(c) = P_0(R(X) \geq c)\,.$$

Dann ist $\alpha(c) \leq \overline{\alpha}(c)$, $\alpha(c)$ ist monoton fallend für steigendes c, und $\overline{\alpha}(0) = 1$. Ferner ist

$$\overline{\alpha}(c) - \alpha(c) = P_0(R(X) = c).$$

Sei (c_n) eine strikt monoton steigende Folge gegen ∞, und $A_n = \{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) > c_n\}$. Wir haben

$$A_1 \supseteq A_2 \supseteq \dots\,,$$

$$P_0(\bigcap A_n) = P_0(\{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) = \infty\}) = P_0(\varnothing) = 0\,,$$

somit $\lim P(A_n) = 0$, das heißt $\lim\limits_{c_n \to \infty} \alpha(c_n) = 0$.

Sei (c_n) eine strikt monoton steigende Folge gegen $c > 0$. Mit der Definition der A_n wie oben haben wir

$$P_0(\bigcap A_n) = P_0(\{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) \geq c\}) = \overline{\alpha}(c)\,,$$

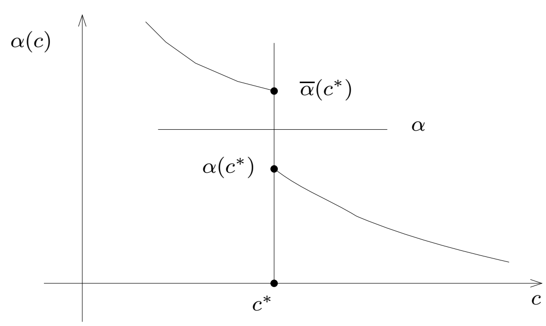
also $\lim\limits_{c_n \to c} \alpha(c_n) = \overline{\alpha}(c)$.
Schließlich sei (b_n) eine strikt fallende Folge gegen b,

$$B_n = \{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) > b_n\}\,.$$

Hier ist

$$\begin{aligned}
&B_1 \subseteq B_2 \subseteq \dots\,, \ \bigcup B_n = \{x \in \mathcal{X} : \rho_0(x) > 0 \wedge R(x) > b\}\,,\\
&P_0(\bigcup B_n) = \alpha(b)\,,
\end{aligned}$$

also $\lim\limits_{b_n \to b} \alpha(b_n) = \alpha(b)\,.$
Wir sehen, dass $\alpha(c)$ eine rechtsstetige Funktion in c ist. Sei $c^* = \inf\{c : \alpha(c) \leq \alpha\}$, dann ist $\overline{\alpha}(c^*) \geq \alpha \geq \alpha(c^*)$.



Falls $\overline{\alpha}(c^*) = \alpha(c^*)$ ist, so setzen wir $\gamma^* = 0$. Falls $\overline{\alpha}(c^*) > \alpha(c^*)$ ist, so setzen wir

$$\gamma^* = \frac{\alpha - \alpha(c^*)}{\overline{\alpha}(c^*) - \alpha(c^*)}\,,$$

und erklären in beiden Fällen den Test φ^* durch

$$\varphi^*(x) = \begin{cases} 0 & R(x) < c^* \\ 1 & R(x) > c^* \\ \gamma^* & R(x) = c^*\,. \end{cases}$$

Daraus ergibt sich (im Fall einer diskreten Variablen nehmen wir wieder die Summe)

$$\begin{aligned}
E_0[\varphi^*] = \int \varphi^*(x)\rho_0(x)dx &= \int\limits_{R(x)>c^*} \rho_0(x)dx + \int\limits_{R(x)=c^*} \gamma^*\rho_0(x)dx \\
&= P_0(R(X) > c^*) + \gamma^* P_0(R(X) = c^*) \\
&= \alpha(c^*) + \begin{cases} 0 \qquad \text{falls} \quad \overline{\alpha}(c^*) = \alpha(c^*) = \alpha \\ \frac{\alpha-\alpha(c^*)}{\overline{\alpha}(c^*)-\alpha(c^*)}(\overline{\alpha}(c^*) - \alpha(c^*)) \text{ falls } \overline{\alpha}(c^*) > \alpha(c^*). \end{cases}
\end{aligned}$$

In beiden Fällen ergibt sich $E_0[\varphi^*] = \alpha$, also ist das Niveau α ausgeschöpft.

c. Es sei φ ein beliebiger bester Test mit $E_0[\varphi] = \alpha$, und φ^* mit $c^* > 0$ ein Neyman-Pearson Test mit $E_1[\varphi] = E_1[\varphi^*]$. Mit der Funktion $g(x)$ wie in Teil a) haben wir $0 = \int g(x)dx$, also ist wegen $g(x) \geq 0$, $g(x) = 0$ bis auf eine Menge mit Maß 0. Da $\{x : \rho_1(x) - c^*\rho_0 = 0\}$ Maß 0 hat, muß also $\varphi(x) = \varphi^*(x)$ sein bis auf eine Menge vom Maß 0. $\square$

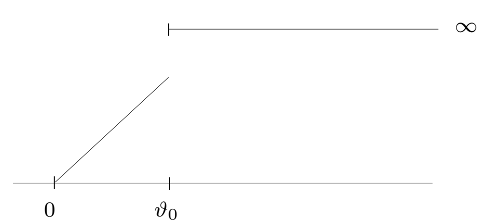
Beispiel. Betrachten wir noch einmal das Orangen Beispiel mit Parametern N und n. Wir wollen

$$\theta_0 = \{0, 1, \ldots, \vartheta_0\} \text{ gegen } \theta_1 = \{\vartheta_0 + 1, \ldots, N\}$$

testen. Für irgendein $\vartheta_1 \in \theta_1$ ist (wir setzen $P_\vartheta(x) = P_\vartheta(X = x)$)

$$\begin{aligned}
R(x) = \frac{P_{\vartheta_1}(x)}{P_{\vartheta_0}(x)} &= \frac{P_{\vartheta_0+1}(x)P_{\vartheta_0+2}(x)\cdots P_{\vartheta_1}(x)}{P_{\vartheta_0}(x)P_{\vartheta_0+1}(x)\cdots P_{\vartheta_1-1(x)}} = \prod_{k=\vartheta_0}^{\vartheta_1-1} \frac{P_{k+1}(x)}{P_k(x)} \\
&= \prod_{k=\vartheta_0}^{\vartheta_1-1} \frac{\binom{k+1}{x}\binom{N-k-1}{n-x}}{\binom{k}{x}\binom{N-k}{n-x}} = \prod_{k=\vartheta_0}^{\vartheta_1-1} \frac{k+1}{k+1-x} \cdot \frac{N-k-n+x}{N-k}\,,
\end{aligned}$$

und diese Funktion ist für $x \leq \vartheta_0$ monoton steigend, und $R(x) = \infty$ für $x > \vartheta_0$:



Wir setzen den Neyman-Pearson an:

$$\varphi^*(x) = \begin{cases} 0 & x < c^* \\ 1 & x > c^* \\ \gamma^* & x = c^* \,. \end{cases}$$

Dabei werden die Konstanten c^*, γ^* aus

$$E_{\vartheta_0}[\varphi^*] = P_{\vartheta_0}(X > c^*) + \gamma^* P_{\vartheta_0}(X = c) = \alpha$$

ermittelt. Der Test φ^* hängt nicht von ϑ_1 ab, und daher ist ϑ_0 gegen jedes ϑ_1 testbar.

Nehmen wir $\vartheta < \vartheta_0$, so sagt der Satz, dass φ_{ϑ_0} besser als φ_ϑ ist, also $\alpha = E_0[\varphi^*_{\vartheta_0}] \geq E_0[\varphi^*_\vartheta]$. Unsere Intuition ist also richtig. Man suche ϑ_0 mit $E_{\vartheta_0}[\varphi^*] = \alpha$.

Unser Eingangsbeispiel mit $N = 10.000$, $n = 50$ führt bei $\alpha = 0,025$ zu $c^* = 6$, $\gamma^* = 0,52$, $\vartheta_0 = 500$ $(= 5\%)$, und daher zum Test

$$\begin{aligned} x < 6 &\;\Rightarrow\; H_0 \\ x > 6 &\;\Rightarrow\; H_1 \\ x = 6 &\;\Rightarrow\; H_1 \text{ mit } W\text{-keit } 0,52\,. \end{aligned}$$