3.2 Gleichmäßig beste Verfahren

Def. 3.30 (Gleichmäßig beste Verfahren)

Gegeben sei das auf eine Menge $\mathcal{D}_0$ eingeschränkte Auswertungsproblem $(\mathcal{D}_0, \Theta, R(\cdot))$ eines datengestütztes Entscheidungsproblem Gegeben sei das auf eine Menge $\mathcal{D}_0$ eingeschränkte Auswertungsproblem $(\mathcal{D}_0, \Theta, R(\cdot))$ eines datengestütztes Entscheidungsproblem $((\mathbb{A}, \Theta, l(\cdot)); (\mathcal{X}, \sigma(\mathcal{X}), (p_\vartheta)_{\vartheta\in\Theta}))$

Eine Entscheidungsfunktion $d^* \in \mathcal{D}_0$ heißt *gleichmäßig bestes Verfahren* aus $\mathcal{D}_0$, wenn für die Risikofunktion gilt:

$$R(d^*, \vartheta) \leq R(d, \vartheta) \quad \text{für alle } \vartheta \in \Theta \text{ und alle } d \in \mathcal{D}_0. \tag{3.21}$$

Bem. 3.31 (zu Def. 3.30)

i) Man beachte, dass die Risikofunktion von d^* für *alle* $\vartheta \in \Theta$ nicht größer sein soll; d^* soll also im zugehörigen Auswertungsproblem $(\mathcal{D}_0, \Theta, R(\cdot))$ alle Elemente von $\mathcal{D}_0$ dominieren.

ii) Dies ist für „großes $\mathcal{D}_0$“ eine extrem starke Forderung – insbesondere im Lichte der elementaren Beispiele aus Kapitel 2 – aber v.a. bei Exponentialfamilie und geeigneter Einschränkung der Menge der betrachteten Entscheidungsfunktionen möglich (UMVU-Schätzer, gleichmäßig bester Test, siehe später).

Satz 3.32 (Lehmann und Scheffé (1950))

Gegeben sei ein Schätzproblem im Sinne Beispiel 1.59. Ferner sei

- $(p_\vartheta)_{\vartheta \in \Theta}$ eine strikt q-parametrige Exponentialfamilie in $T(\vec{x}) = (T_1(\vec{x}), ..., T_q(\vec{x}))$.

- $\ell(\hat{\vartheta}, \vartheta)$ konvex in $\hat{\vartheta}$ für alle ϑ.

Betrachtet man die Schätzung einer Transformation $\gamma(\vartheta)$ von ϑ und die Klasse $\mathcal{D}_\gamma$ aller für $\gamma(\vartheta)$ erwartungstreuen Schätzer, so gilt:
Ist $\mathcal{D}_\gamma$ nicht leer, so gibt es eine nichtrandomisierte Schätzfunktion der Form $\eta(T(\vec{X}))$, die gleichmäßig beste in der Klasse $\mathcal{D}_\gamma$ ist.

Ist umgekehrt $\eta(T(\vec{X}))$ eine erwartungstreue Schätzfunktion für $\gamma(\vartheta)$, so ist $\eta(T(\vec{X}))$ gleichmäßig bestes Verfahren in D_γ.

Bem. 3.33 (Zur Interpretation des Satzes)

- „Informationsdeutung“:

- Konstruktive Anwendung:

Korollar 3.34

Gegeben sei eine strikt q-parametrige Exponentialfamilie in $T(\vec{x}) = (T_1(\vec{x}), ..., T_q(\vec{x}))$.
Dann ist für jede (messbare) Funktion $\eta(\cdot)$ der Schätzer $\eta(T_1, ..., T_q)$ UMVU-Schätzer für $\mathbb{E}_\vartheta(\eta(T_1, ..., T_q))$.

Bem. 3.35 (Zum Satz von Lehmann-Scheffé)

Die Beschränkung auf erwartungstreue Schätzer ist wesentlich.

Bem. 3.36

Korollar 3.34 wird typischerweise „andersherum“ angewendet. Will man eine Funktion $\gamma(\vartheta)$ schätzen, so sucht man eine Funktion $g(T)$, so dass $\mathbb{E}g(T) = \gamma(\vartheta)$. Gemäß Korollar 3.34 weiß man dann, dass $g(T)$ UMVU für $\gamma(\vartheta)$ ist.
Üblicherweise wird man versuchen einen Ansatz zu wählen, bei dem sich $g(\cdot)$ aus „einfachen Grundfunktionen“ zusammensetzt.

Bsp. 3.37

Gegeben sei eine i.i.d. Stichprobe eines normalverteilten Merkmals mit unbekanntem Mittelwert μ, aber bekannter Varianz σ^2. Man bestimme einen Schätzer UMVU
a) für μ und
b) für $\exp(\mu)$.

Satz 3.39 (Optimale Tests)

Betrachtet werde das Testproblem als Entscheidungsproblem gemäß Beispiel 1.60 mit

a_0 für H_0 entscheiden
a_1 für H_1 entscheiden

	ϑ	
a_0	0	1
a_1	1	0

$$l(a_0, \vartheta) = \begin{cases} 0 & \vartheta \in \Theta_0 \\ 1 & \vartheta \in \Theta_1 \end{cases}$$

$$l(a_1, \vartheta) = \begin{cases} 1 & \vartheta \in \Theta_0 \\ 0 & \vartheta \in \Theta_1 \end{cases}$$

Ferner sei $\Theta_0 = \{\vartheta | \vartheta \leq \vartheta_0\}, \Theta_1 = \{\vartheta | \vartheta \geq \vartheta_1\}, \vartheta_1 > \vartheta_0$, und ein Signifikanzniveau $\alpha \in (0,1)$ vorgegeben.
Bildet $(P_\vartheta^{\oplus n})$ eine strikt einparametrige Exponentialfamilie in T (mit dem natürlichen Parameter $c(\vartheta)$, der in eineindeutiger Beziehung zu ϑ stehe), so gibt es ein $\kappa \in \mathbb{R}$, so dass der Test

$$\varphi^*(\vec{x}) = \begin{cases} 1 & \mathrm{T}(\vec{x}) > \kappa \\ \gamma & \mathrm{T}(\vec{x}) = \kappa \\ 0 & \mathrm{T}(\vec{x}) < \kappa \end{cases} \tag{3.22}$$

mit

$$\mathbb{E}_{\vartheta_0} \varphi^\star = \alpha \tag{3.23}$$

gleichmäßig bester Test (UMP) ist, d.h es gilt

$$\mathbb{E}_\vartheta \varphi^\star \geq \mathbb{E}_\vartheta \varphi, \quad \forall \vartheta \in \Theta_1,$$

für alle φ mit $\sup\limits_{\vartheta \in \Theta_0} \mathbb{E}_{\vartheta_0} \varphi \leq \alpha$.

Bem. 3.40 (Zur Interpretation von Satz 3.39)

- Informationsdeutung:

- konstruktive Anwendung:

Bem. 3.41

- Die Aussage gilt allgemeiner für Verteilungen mit *monotonen Dichtequotienten*; bei denen also für alle $\vartheta_1 \in \Theta_1$, $\vartheta_0 \in \Theta_0$ und eine geeignete Funktion $T(\vec{X})$ der Quotient $\dfrac{f_{T||\vartheta_1}(t)}{f_{T||\vartheta_0}(t)}$ monoton in t ist.

- Auch bei der Fragestellung $H_0 : \vartheta = \vartheta_0$ gegen $H_1 : \vartheta \neq \vartheta_0$, gibt es bei Exponentialfamilien einen gleichmäßig besten Test, wenn man sich auf *unverfälschte* Tests (d.h. Gütefunktion $\geq \alpha$ für alle Elemente der Alternative) beschränkt.
  Ähnliches gilt in der Situation $H_0 : \vartheta \in [\underline{\vartheta}_0, \bar{\vartheta}_0]$ gegen $H_1 : \vartheta \in [\underline{\vartheta}_1, \bar{\vartheta}_1]$, wenn man nur unverfälschte und ähnliche Tests (Gütefunktion am Rand der Nullhypothese $= \alpha$) betrachtet.

Bsp. 3.43 Anwendung von Satz 3.39 auf den Mittelwertstest bei der Normalverteilung und auf das Testen bei der Binomialverteilung

a) $X_1, \ldots, X_n \quad i.i.d. \qquad \sim N(\mu, \sigma^2)$

$H_0 : \mu \leq \mu_0 \quad$ gegen $\quad H_1 : \mu \geq \mu_1 > \mu_0 \quad \sigma^2$ bekannt

$$f_{X_1,\ldots,X_n||\mu}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2\sigma}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

$$= \underbrace{\left(\frac{1}{\sqrt{2\pi}}\right)^n}_{\substack{\text{konstant} \\ \text{(bei bek. } \sigma^2)}} \cdot \frac{1}{\sigma^n} \cdot \underbrace{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right)}_{\substack{\text{nur von } X \text{ abhängig} \\ \text{(bei bekanntem } \sigma^2)}} \cdot \underbrace{\exp\left(-\frac{1}{2\sigma^2} n\mu^2\right)}_{\substack{\text{nur von } \mu \text{ abh.} \\ \text{nicht von } x}} \cdot \exp\left(\frac{\mu}{\sigma^2} \cdot \sum_{i=1}^{n} x_i\right)$$

natürlicher Parameter $\frac{\mu}{\sigma^2}$

Suffiziente Statistik: $T = \sum_{i=1}^{n} X_i$

Die kritische Region $K(\kappa)$, d.h. der Bereich aller $\vec{x}$ mit $\varphi(\vec{x}) = 1$, ist wegen (3.5) von der Form

$$K(\kappa) := \{\vec{x} | T(\vec{x}) > \kappa\}$$

mit κ so, dass

$$P_{\mu_0}(K(\kappa)) \stackrel{!}{=} \alpha.$$

(Man weiß ja wegen (3.23), dass die maximale Fehlerwahrscheinlichkeit an der Grenze der Nullhypothese angenommen wird.)

Das heißt, es soll gelten

$$P_{\mu_0}(K(\kappa)) = P\mu_0(\{\vec{x}|T(\vec{x}) > \kappa\}) = P\mu_0(\{\vec{x}|\sum_{i=1}^{n} X_i > \kappa\}) = \alpha.$$

Zur Berechnung dieser Wahrscheinlichkeit benutzt man

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2),$$

d.h.

$$\frac{\sum_{i=1}^{n} X_i - n\mu_0}{\sigma\sqrt{n}} \sim N(0,1).$$

Also führt der Ansatz $P_{\mu_0}(K(\kappa)) \stackrel{!}{=} \alpha$, d.h.

$$P\left(\frac{\sum_{i=1}^{n} X_i - n\mu_0}{\sigma \cdot \sqrt{n}} \stackrel{!}{>} \underbrace{\frac{\kappa - n\mu}{\sigma \cdot \sqrt{n}}}_{\kappa' = \tau_\alpha}\right) \stackrel{!}{=} \alpha,$$

Fraktil der Normalverteilung

dazu, dass die kritische Region so zu wählen ist, dass

$$
\begin{aligned}
& \frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sigma \cdot \sqrt{n}} > \tau_\alpha \\
\Longleftrightarrow\ & \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} > \tau_\alpha \\
\Longleftrightarrow\ & \bar{X} > \mu_0 + \frac{\tau_\alpha \cdot \sigma}{\sqrt{n}}.
\end{aligned}
$$

Da ferner $\mu_1 > \mu_0$ bei dieser Konstruktion beliebig gewählt werden konnte, gilt die Aussage für *alle* $\mu > \mu_0$.

Beachte: es ergibt sich die übliche kritische Region des Gauss Tests, dieser ist also UMP.

b) Bei der Bernoulliverteilung sei zu einem konkreten Beispiel übergegangen.
Man testet $H_0 : p \leq 0.5$ gegen $H_1 : p \geq 0.6$, wobei bei der i.i.d. Stichprobe $X_1, \dots X_n$ der Stichprobenumfang $n = 5$ sei und das Signifikanzniveau auf $\alpha = 0.1$ gesetzt sei.
Zur Anwendung von Satz 3.39 bringt man zunächst die Wahrscheinlichkeitsfunktion $f_p(\cdot)$ der Bernoulliverteilung auf

„Exponentialfamilien-Gestalt“.

$$
\begin{aligned}
f_p(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \cdot p^{x_i} \cdot (1-p)^{1-x_i} && = \\
&= (1-p)^n \cdot \left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i} && = \\
&= (1-p)^n \cdot \exp\left(\ln\left(\left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i}\right)\right) && = \\
&= (1-p)^n \cdot \exp\left(\sum_{i=1}^{n} x_i \cdot \ln\left(\frac{p}{1-p}\right)\right)
\end{aligned}
$$

Man erhält $T = \displaystyle\sum_{i=1}^{n} X_i$, und $\ln(\frac{p}{1-p})$ ist der natürliche Parameter.

Der Ansatz

$$K(\kappa) := \{\vec{x} | T(\vec{x}) > \kappa\}$$

für die kritische Region führt auf ein Problem. Da $T = \sum_{i=1}^{n} X_i$ binomialverteilt ist, ergibt sich [18]

$$P_{p_0}(K(\kappa)) = P_{p_0}(\{T > \kappa\}) = \sum_{\substack{j>\kappa \\ j\in\mathbb{N}}}^{n} \binom{5}{j} \underbrace{0.5^j \cdot 0.5^{5-j}}_{0.5^5}$$

[18]Wegen (3.23) setzt man hier wieder den oberen Randwert der Nullhypothese ein, also denjenigen Wert der Nullhypothese, der am schwersten von H_1 zu unterscheiden ist.

Allerdings gibt es kein κ, so dass diese Gleichung erfüllt ist:

$$\begin{aligned}
\kappa &\in (4,5] \quad P(K(\kappa)) = 0 \\
\kappa &= 4 \Rightarrow P(K(\kappa)) = \binom{5}{5} \cdot (\tfrac{1}{2})^5 = \tfrac{1}{32} < 0.1 \\
\kappa &\in (3,4] \Rightarrow P(K(\kappa)) = P(K(4)) = \tfrac{1}{32} \\
\kappa &= 3 \Rightarrow P(K(\kappa)) = \tfrac{1}{32} + \binom{5}{4} \cdot 0.5^5 = \tfrac{1}{32} + \tfrac{5}{32} = \tfrac{6}{32} > 0.1
\end{aligned}$$

Was tun? Klar ist

$\varphi(\vec{x}) = 1$, d.h. H_0 ablehnen, für alle $\vec{x}$ mit $T(\vec{x}) = 5$ und
$\varphi(\vec{x}) = 0$, d.h. H_0 nicht ablehnen, für alle $\vec{x}$ mit $T(\vec{x}) \leq 3$.

Ist $T(\vec{x}) = 4$, dann hat man so zu randomisieren, dass $\mathbb{E}_{0.5}\varphi = \alpha$. Man setzt hierzu $\varphi(\vec{x}) = \gamma$, falls $T(\vec{x}) = 4$, und erhält:

$$\begin{aligned}\mathbb{E}_{0.5}\varphi &= 1 \cdot P(\varphi(x) = 1) + \gamma \cdot P(\varphi(x) = \gamma) + \\ &\quad +0 \cdot P(\varphi(x) = 0) \overset{!}{=} \alpha\end{aligned}$$

also

$$\begin{aligned}\gamma &= \frac{\alpha - P(\varphi(x) = 1)}{P(\varphi(x) = \gamma)} = \\ &= \frac{\alpha - P(T = 5)}{P(T = 4)} = \frac{0.1 - \frac{1}{32}}{\frac{5}{32}} = 0.44\ .\end{aligned}$$