- Passend zum Beispiel für
  - t-Test: Vergleich von μ_1, μ_2 bei unabhängigen Stichproben nur, falls $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ist.
  - Test auf Signifikanz von β_1 in linearer Einfachregression.
- Bereits nicht mehr anwendbar für
  - Vergleich von μ_1, μ_2 bei $\sigma_1^2 \neq \sigma_2^2$ (Behrens-Fisher-Problem).
  - Test auf Signifikanz von β_1 im Logit- oder Poisson-Regressionsmodell.

$\Rightarrow$ (asymptotische) Likelihood-Theorie, Bayes-Inferenz.

2.3 Bereichsschätzungen und Konfidenzintervalle

2.3.1 Definition und Beurteilung der Güte

Definition 2.39 (Bereichsschätzung)**.** *Eine* Bereichsschätzung *(ein* Konfidenzbereich*) C für $\tau(\theta)$, $\tau : \Theta \to \Sigma$, zum (vorgegebenen)* Vertrauensgrad *(*Konfidenzniveau*) $1 - \alpha$ ist eine Abbildung des Stichprobenraums $\mathcal{X}$ in die Potenzmenge $\mathcal{P}(\Sigma)$, also $x \to C(x) \in \mathcal{P}(\Sigma)$, mit $\{\tau(\theta) \in C(X)\}$ messbar und*

$$\mathbb{P}_\theta(\tau(\theta) \in C(X)) \geq 1 - \alpha \quad \textit{für alle } \theta \in \Theta.$$

$C(X)$ ist ein zufälliger Bereich in $\mathcal{P}(\Sigma)$. Nach Beobachtung der Stichprobe $X = x$ ist $C(x)$ gegeben. Der Aussage

$$\tau(\theta) \in C(x) \quad (\text{richtig} \quad \overset{!}{\text{oder}} \quad \text{falsch})$$

wird der Vertrauensgrad $1 - \alpha$ zugeordnet. Dabei gilt die bekannte Häufigkeitsinterpretation. Ist $C(x)$ für jedes x ein Intervall, so heißt $C(x)$ *Konfidenzintervall* und C eine *Intervallschätzung*.

Eine Wahrscheinlichkeitsaussage zu

$$\tau(\theta) \in C(x)$$

bei gegebenem x ist im Rahmen der Bayes-Inferenz (ohne logische Probleme) möglich.

Die „Präzision" von $C(X)$ wird gemessen durch die erwartete Größe des Bereichs bzw. durch die Länge des Konfidenzintervalls.

Beispiel 2.25. *Seien $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ und*

$$C(X) = \left[\bar{X} - t_{n-1}\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1}\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}}\right]$$

ein Konfidenzintervall für μ. Die Länge

$$L = 2\, t_{n-1}\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}}$$

von $C(X)$ ist zufällig mit Erwartungswert

$$\mathbb{E}(L) = 2\, t_{n-1}\Big(\frac{\alpha}{2}\Big)\frac{1}{\sqrt{n}}\ \mathbb{E}(S) = 2\, t_{n-1}\Big(\frac{\alpha}{2}\Big)\frac{\sigma}{\sqrt{n}}\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}.$$

Es gilt:

$$\begin{aligned} 1-\alpha \ \textit{größer} &\ \rightarrow\ \mathbb{E}(L)\ \textit{größer},\\ n\ \textit{größer} &\ \rightarrow\ \mathbb{E}(L)\ \textit{kleiner}. \end{aligned}$$

Bei der Beurteilung der Präzision eines Konfidenzintervalls durch die Länge ist ein Konfidenzintervall umso besser, je kürzer seine erwartete Länge ist. Allgemein wird ein Konfidenz*bereich* C durch die mittlere „Größe" beurteilt. Dazu sei π eine Verteilung (oder ein Maß) auf Θ. Dann ist

$$\pi(C(x))$$

die Größe von $C(x)$. Bei Konfidenzintervallen ergibt sich die Länge, wenn π das Lebesgue-Maß ist. Dann ist

$$\mathbb{E}_\theta(\pi(C(X)))$$

die zu erwartende Größe. Zur Beurteilung der Güte reicht die erwartete Länge bzw. Größe allein nicht aus.

Definition 2.40 (Kennfunktion eines Konfidenzbereichs)**.** *Eine* Kennfunktion *ist definiert als eine Funktion*

$$k_C(\theta,\theta') := \mathbb{P}_\theta(C(X) \ni \tau(\theta')).$$

Dabei ist θ der „wahre" Wert und θ' irgendein Wert in Θ.

Für $\theta = \theta'$ ist „$C(X) \ni \tau(\theta')$" eine Aussage, deren Wahrscheinlichkeit möglichst groß sein soll. Für $\theta \neq \theta'$ mit $\tau(\theta') \neq \tau(\theta)$ ist „$C(X) \ni \tau(\theta')$" eine Aussage, deren Wahrscheinlichkeit möglichst klein gehalten werden soll.

Im Weiteren betrachten wir den Spezialfall $\tau(\theta) = \theta$ mit skalarem θ. Dann ist

$$k_C(\theta,\theta') = \mathbb{P}_\theta(C(X) \ni \theta').$$

Definition 2.41.

1. *Ein Konfidenzbereich besitzt den* Vertrauensgrad $1-\alpha :\overset{def}{\Leftrightarrow}$

$$k_C(\theta,\theta') \geq 1-\alpha \quad \textit{für alle } \theta' = \theta.$$

2. *Ein Konfidenzbereich zum Vertrauensgrad $1-\alpha$ heißt* unverfälscht $:\overset{def}{\Leftrightarrow}$

$$k_C(\theta,\theta') \leq 1-\alpha \quad \textit{für } \theta' \neq \theta.$$

3. *Ein [unverfälschter] Konfidenzbereich C_0 zum Vertrauensgrad $1-\alpha$ heißt* gleichmäßig bester (trennscharfer) *[bzw. gleichmäßig bester unverfälschter]* Konfidenzbereich $:\overset{def}{\Leftrightarrow}$ *für alle $\theta' \neq \theta$ und alle [unverfälschten] Konfidenzbereiche C zum Vertrauensgrad $1-\alpha$ gilt*

$$k_{C_0}(\theta,\theta') \leq k_C(\theta,\theta').$$

Lemma 2.42. *Jeder gleichmäßig beste Konfidenzbereich besitzt auch die kleinste zu erwartende Größe (aber nicht umgekehrt).*

Beweis.

$$\begin{aligned}\int_{\mathcal{X}} \pi(C(x))d\mathbb{P}_\theta(x) &= \int_{\mathcal{X}}\int_{\Theta} I_{C(x)}(\theta')d\pi(\theta')d\mathbb{P}_\theta(x) \\ &= \int_{\Theta}\int_{\mathcal{X}} I_{C(x)}(\theta')d\mathbb{P}_\theta(x)d\pi(\theta') \qquad \text{(Fubini)} \\ &= \int_{\Theta} \underbrace{\mathbb{P}_\theta(\{x : C(x) \ni \theta'\})}_{k_C(\theta,\theta')} d\pi(\theta').\end{aligned}$$

Für jedes „wahre” θ gilt also

$$\underbrace{\int_{\mathcal{X}} \pi(C(x))d\mathbb{P}_\theta(x)}_{\text{erwartete Größe}} = \underbrace{\int_{\Theta} k_C(\theta,\theta')d\pi(\theta')}_{\substack{\text{erwarteter Wert der Kenn-}\\ \text{funktion des Konfidenzbereichs}}} .$$

□

2.3.2 Dualität zwischen Konfidenzbereichen und Tests

Wir legen den oben beschriebenen Spezialfall $\tau(\theta) = \theta$ mit skalarem θ zugrunde.

Zu jedem festen θ betrachten wir einen Niveau-α-Test $\phi_\theta(x)$ für die Nullhypothese $H_0 = \{\theta\}$ gegen die Alternative $H_1 = \Theta \backslash H_0$. Die Tests sollen nicht randomisiert sein, so dass sie durch die Festlegung einer Prüfgröße $T_\theta = T_\theta(x)$ und eines kritischen Bereichs (Ablehnbereichs) K_θ bestimmt werden:

$$\phi_\theta(x) = \begin{cases} 1 & \text{für } T_\theta(x) \in K_\theta, \\ 0 & \text{sonst.} \end{cases}$$

Die Nullhypothese „Der unbekannte Parameter hat den Wert θ” wird nach Beobachtung von $X = x$ genau dann nicht abgelehnt — durch die Beobachtung „bestätigt” — wenn

$$T_\theta\ (x)\ \in\ \bar{K}_\theta\ =\ \text{Annahmebereich des Tests } \phi_\theta$$

gilt. Daher ist es naheliegend, als einen Konfidenzbereich nach der Beobachtung $X = x$ den Bereich

$$C(x) := \{\theta \in \Theta : T_\theta(x) \in \bar{K}_\theta\}$$

zu definieren; dem entspricht vor der Beobachtung der zufällige Bereich

$$C(X) = \{\theta \in \Theta : T_\theta(X) \in \bar{K}_\theta\}$$

bzw.

$$C(X) = \{\theta \in \Theta : \phi_\theta(X) = 0\}$$

Eine Bestätigung dieser Vorgangsweise ist der folgende Satz.

Satz 2.43 (Korrespondenzsatz).

1. *Ist* $\{\phi_\theta\}$ *eine Menge von Tests* ϕ_θ *für* $H_0 = \{\theta\}$ *gegen* $H_1 = \Theta\backslash\{\theta\}$ *zum Niveau* α*, so ist* $C(X) := \{\theta \in \Theta : \phi_\theta(X) = 0\}$ *ein Konfidenzbereich zum Vertrauensgrad* $\gamma = 1 - \alpha$*.*

2. *Ist* $\{\phi_\theta\}$ *eine Menge gleichmäßig bester [unverfälschter] Tests, so ist auch* $C(X)$ *ein gleichmäßig bester [unverfälschter] Konfidenzbereich.*

Beweis. Der Beweis zu 1. ergibt sich aus

$$\mathbb{P}_\theta(C(X) \ni \theta) = \mathbb{P}_\theta(\phi_\theta(X) = 0) = 1 - \alpha \quad \text{für alle } \theta \in \Theta,$$

derjenige für 2. aus der Beziehung

$$\begin{aligned} k_C(\theta, \theta^{'}) &= \mathbb{P}_\theta(C(X) \ni \theta^{'}) = \mathbb{P}_\theta(\phi_{\theta'}(X) = 0) \\ &= 1 - \mathbb{P}_\theta(\phi_{\theta'}(X) = 1) = 1 - g_{\phi_{\theta'}}(\theta) \end{aligned}$$

für alle $\theta, \theta^{'} \in \Theta$. Dabei bezeichnet $g_{\phi_{\theta'}}$ die Gütefunktion des Tests $\phi_{\theta'}$. □

Der Korrespondenzsatz lässt sich verallgemeinern auf die Situation, in der man gegenüber bestimmten Fehlschätzungen besonders empfindlich ist; man hat dazu eine Testfamilie solcher Tests zugrunde zu legen, die die entsprechenden Hypothesen testen, also nicht mehr Tests mit zweiseitiger Fragestellung. Darüber hinaus gilt der im Korrespondenzsatz enthaltene Zusammenhang zwischen Tests und einem Konfidenzbereich auch dann, wenn randomisierte Tests zugelassen werden, so dass man auf diese Weise zu einem randomisierten Konfidenzbereich kommt: $C(x)$ ist die Menge aller θ, die bei der Beobachtung x von dem Test ϕ_θ (auch nach Randomisierung) nicht abgelehnt werden.

Auf diese Weise lässt sich die Theorie der Bereichsschätzungen auf die Testtheorie zurückführen bis auf das folgende Problem: Damit ein „vernünftiger" Konfidenzbereich (vernünftig im topologischen Sinn, also zum Beispiel ein Konfidenzintervall) aus der Testfamilie konstruierbar ist, muss die Testfunktion $\phi_\theta(x)$, besser noch die Prüfgröße $T_\theta(x)$ *als Funktion in* θ (für jedes feste θ) „gutartig" sein (im Idealfall monoton in θ); außerdem darf die Verteilung von $T_\theta(X)$ nicht von θ abhängen, zusammen bedeutet dies: $T_\theta(X)$ muss eine *Pivotgröße* sein, die auf „einfache" (zum Beispiel monotone) Weise von θ abhängt: Gesucht sind einfach strukturierte Pivotgrößen.

2.4 Multiples Testen

Literatur:

- Lehmann & Romano, Kapitel 9
- Dudoit, Shaffer & Boldrick (2003): *Multiple Hypothesis Testing in Microarray Experiments*, Statistical Science (**18**), Seiten 71-103

Problem: Eine endliche Menge von (Null-) Hypothesen $H_1, ..., H_m$ soll mit Hilfe *eines* Datensatzes simultan getestet werden.

Beispiele: