

Bayesianische FDR (Teil 1)

Mareile Große Ruse

08.11.2010

Aus der (frequentistischen) multiplen Testtheorie ist die *Family Wise Error Rate* (FWER) als eine Möglichkeit, den Fehler 1. Art zu kontrollieren, wohlbekannt (siehe Seminar). Die Kontrolle des multiplen Niveaus ist gerade in genetischen Fragestellungen, wo die Anzahl n der zu testenden Hypothesen sehr groß werden kann, häufig zu konservativ. Eine Alternative zur FWER ist die *False Discovery Rate*, die nicht die Wahrscheinlichkeit für einen Fehler 1. Art kontrolliert, sondern den (erwarteten) Anteil von Falsch-Verwerfungen an den Verwerfungen insgesamt.

Doch nicht nur in der frequentistischen Testtheorie findet die FDR ihre Anwendung. Gemäß eines Resultats von Storey lässt die FDR im Setting des sog. Zweiklassen-Mischmodells eine bayesianische Interpretation zu. Basierend auf dieser Aussage sollen hier zwei bayessche Testmethoden vorgestellt werden, die beide auf einer Form der FDR basieren.

1 Bayes-Theorem für bedingte Dichten

Im Folgenden werden häufig bedingte Dichten auftreten. Daher soll zu Beginn dieser Ausarbeitung an das Bayes-Theorem für bedingte Dichten erinnert werden. Zunächst jedoch eine Wiederholung zu bedingten Dichten allgemein:

Satz. *Es seien Zufallsvariablen $Z : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathfrak{B})$ und $\theta : (\Omega, \mathfrak{A}, \mathbb{P}) \rightarrow (S, \mathfrak{S})$ gegeben, wobei (S, \mathfrak{S}) einen Borelraum und μ ein σ -endliches Maß auf (S, \mathfrak{S}) bezeichnen. Ferner sei $f_{(Z,\theta)} : \mathbb{R} \times S \rightarrow \mathbb{R}_+$ die $\lambda \otimes \mu$ -Dichte von $\mathbb{P}^{(Z,\theta)}$. Dann*

$$f_{\theta}(\vartheta) = \int_{\mathbb{R}} f_{(Z,\theta)}(z, \vartheta) dz \quad \text{ist } \mu\text{-Dichte von } \mathbb{P}^{\theta}$$

$$f_Z(z) = \int_{\mathbb{R}} f_{(Z,\theta)}(z, \vartheta) d\vartheta \quad \text{ist } \lambda\text{-Dichte von } \mathbb{P}^Z$$

$$f_{Z|\theta=\vartheta}(z) = \frac{f_{(Z,\theta)}(z, \vartheta)}{f_{\theta}(\vartheta)} \quad \text{ist } \lambda\text{-Dichte von } \mathbb{P}^{Z|\theta=\vartheta}$$

$$f_{\theta|Z=z}(\vartheta) = \frac{f_{(Z,\theta)}(z, \vartheta)}{f_Z(z)} \quad \text{ist } \mu\text{-Dichte von } \mathbb{P}^{\theta|Z=z}$$

Satz (Bayes-Theorem für Dichten). *In der Situation des obigen Satzes gilt*

$$(i) \quad f_{\theta|Z=z}(\theta) = \frac{f_{Z|\theta=\theta}(z) f_{\theta}(\vartheta)}{f_Z(z)}$$

$$(ii) \quad f_{Z|\theta=\vartheta}(z) = \frac{f_{\theta|Z=z}(\vartheta) f_Z(z)}{f_{\theta}(\vartheta)}$$

Für die nächsten Abschnitte ist das unten stehende Anwendungsbeispiel sehr wichtig.

Beispiel. Es seien $Z : (\Omega, \mathfrak{A}) \rightarrow (\mathbb{R}, \mathfrak{B})$, $\theta : (\Omega, \mathfrak{A}) \rightarrow (\mathbb{N}, \mathfrak{P}(\mathbb{N}))$ und μ das Zählmaß auf $(\mathbb{N}, \mathfrak{P}(\mathbb{N}))$.
Es gelte:

$$\begin{aligned}\theta &\sim \text{Ber}(p_1), \quad p_0 := 1 - p_1 \in (0, 1) \\ Z &\sim \mathcal{N}(0, 1)(1 - \theta) + \mathcal{N}(100, 1)\theta\end{aligned}$$

Ziel ist die Bestimmung von $f_{\theta|Z=z}$.
Mit obigem Bayes-Theorem erhält man:

$$f_{\theta|Z=z}(\vartheta) = \frac{f_{Z|\theta=\vartheta}(z)f_{\theta}(\vartheta)}{f_Z(z)} = \begin{cases} \frac{p_0\varphi_{0,1}(z)}{p_0\varphi_{0,1}(z)+p_1\varphi_{100,1}(z)}, & \text{für } \vartheta = 0, \\ \frac{p_1\varphi_{100,1}(z)}{p_0\varphi_{0,1}(z)+p_1\varphi_{100,1}(z)}, & \text{für } \vartheta = 1 \end{cases}$$

2 False discovery rate

Definition. Es sei ein multiples Testproblem mit m zu testenden Hypothesen gegeben und $\varphi = (\varphi_i, i = 1, \dots, m)$ eine multiple Testprozedur.

- V : (zufällige) Anzahl der Typ-1-Fehler
 R : (zufällige) Anzahl von verworfenen Hypothesen insgesamt.

1. Die *False Discovery Rate (FDR)* von φ ist definiert als

$$FDR(\varphi) = \mathbb{E}\left(\frac{V}{R \vee 1}\right)$$

2. Die *positive False Discovery Rate (pFDR)* von φ ist gegeben durch

$$pFDR(\varphi) = \mathbb{E}\left[\frac{V}{R} \mid R > 0\right].$$

Ist nun ein Test φ durch einen Verwerfungsbereich $\{Z \in \Gamma_\alpha\}$ charakterisiert (in dem Sinne, dass $\varphi = 1 \Leftrightarrow Z \in \Gamma_\alpha$), so ist $pFDR(\varphi) = \mathbb{E}\left[\frac{V}{R} \mid Z \in \Gamma_\alpha\right]$. Beispielsweise könnte $\Gamma_\alpha = (z_\alpha, \infty)$ sein. In diesem Fall wird die Notation $pFDR(z_\alpha) := pFDR(\varphi)$ verwendet.

Die pFDR hat in einem gewissen Setting, dem Zweiklassen-Mischmodell, eine Bayesianische Interpretation. Dies wird durch einen Satz von Storey (2001) deutlich.

Definition (Zweiklassen-Mischmodell). Es seien $(Z_1, H_1), \dots, (Z_n, H_n)$ iid. Zufallsvariablen. Es gelte für $i = 1, \dots, n$

$$\begin{aligned}Z_i|H_i &\sim (1 - H_i)f_0 + H_i f_1 \\ H_i &\sim \text{Ber}(p_1)\end{aligned}$$

In dieser Ausarbeitung kann man H_i als Indikatorvariable mit der Interpretation $[H_i = 1 \Leftrightarrow i\text{-te Alternativhypothese ist wahr}]$ verstehen.

Satz (Storey (2001)). Gegeben sei ein Zweiklassen-Mischmodell. φ sei eine multiple Testprozedur mit obiger Charakterisierung durch den Ablehnbereich. Dann gilt

$$pFDR(\varphi) = pFDR(z_\alpha) = \mathbb{P}[H = 0 \mid Z > z_\alpha]$$

Das heißt, die pFDR kann als ein Bayesscher a posteriori Typ-1-Fehler angesehen werden.

Eine „lokale“ Version von $pFDR(z_\alpha) = \mathbb{P}[H = 0 \mid Z > z_\alpha]$ wäre dann $\mathbb{P}[H = 0 \mid Z = z]$ für $z \in \mathbb{R}$:

Definition. Im Setting des oberen Satzes ist für $z \in \mathbb{R}$ die *lokale fdr* definiert als

$$fdr(z) := \mathbb{P}[H = 0 \mid Z = z]$$

Mit der Bayes-Formel für Dichten erhält man:

$$fdr(z) = f_{H|Z=z}(0) = \frac{p_0 f_0(z)}{f(z)}$$

Es gilt: $pFDR(z) = \mathbb{E}_f(fdr(Z)|Z > z)$.

3 Die fdr in empirischen und nichtparametrischen Bayes-Methoden

Wir betrachten eine Anwendung des 2-Klassen-Mischmodells auf das Beispiel der Genexpression. Hier sei n die (meist sehr große) Anzahl zu untersuchender Gene. Jedes Gen habe entweder die Ausprägung „Tumor“ oder „normal“ und Ziel ist es, die n Gene in die Klassen „Tumor“/ „normal“ einzuteilen. Jedem Gen sei ein sogenannter z -Score zugeordnet, anhand dessen die Klassifikation vorgenommen werden soll. Es stehen also n Teststatistiken Z_1, \dots, Z_n zur Verfügung.

p_0 sei der Anteil der Gene, die der Kategorie „normal“ angehören und es sei $p_1 = 1 - p_0$.

Zur Durchführung der Klassifikation seien hier nun zwei Methoden vorgestellt: der empirische und der nichtparametrische Bayes-Ansatz. Anschließend sollen Schwächen des Erstgenannten im Vergleich zur nichtparametrischen Herangehensweise aufgezeigt werden. In beiden Ansätzen spielt die fdr als Entscheidungskriterium eine tragende Rolle (Erinnerung: $fdr(z_i)$ = Wahrscheinlichkeit, dass Gen i der Kategorie „normal“ angehört, gegeben, dass die zugehörige Teststatistik Z_i den Wert z_i annimmt).

Der empirische Bayes-Ansatz Zunächst werden a priori ein Wert für p_0 sowie eine Dichte f_0 gewählt. Diese Wahlen beruhen bestenfalls auf Erfahrungen oder Vorexperimenten. Anschließend wird die Dichte f aus den Daten durch \hat{f} geschätzt. Durch *plug-in* erhält man so

$$\widehat{fdr}(z_i) = \frac{p_0 \hat{f}_0(z_i)}{\hat{f}(z_i)} =: \mathbf{P}_{\text{emp}}(z_i)$$

Anhand dieser Größe wird nun die Einteilung eines jeden Gens in „Tumor“ oder „normal“ vorgenommen (z.B.: Ordne Gen i der Kategorie „Tumor“ zu, genau dann, wenn $Z_i > \gamma_\alpha$ mit einem Schwellenwert γ_α).

Problematisch ist hier die (durch Wahl von f_0, p_0 einfließende) Subjektivität der Entscheidung.

Der nichtparametrische Ansatz In diesem Ansatz wird zunächst ein Wahrscheinlichkeitsmodell für die unbekanntenen Größen p_0, f_0, f_1 festgelegt, d.h. für jede der Größen wird eine a priori Wahrscheinlichkeitsverteilung auf dem Raum $[0, 1]$ (für p_0) bzw. einem passenden Funktionenraum (für f_0, f_1) gewählt.

Ein Beispiel für a priori Verteilungen könnte sein: $p_0 \sim \text{unif}(0.05, 1)$ und für f_0, f_1 (unendliche) Mischungen von Normalverteilungen.

Gegeben alle Daten $Y := (Z_i)$ werden dann die a posteriori Verteilungen $\mathbb{P}^{f_0|Y}, \mathbb{P}^{f_1|Y}, \mathbb{P}^{p_0|Y}$ bestimmt (z.B. über MCMC). Als Entscheidungskriterium betrachtet man nun die Größe $\mathbb{E}[fdr(z) | Y]$ und verwendet

$$\mathbb{E}\left[\frac{p_1 f_1(z_i)}{f(z_i)} \mid Y\right] \approx \frac{\mathbb{E}[p_1 | Y] \mathbb{E}[f_1(z_i) | Y]}{\mathbb{E}[p_0 | Y] \mathbb{E}[f_0(z_i) | Y] + \mathbb{E}[p_1 | Y] \mathbb{E}[f_1(z_i) | Y]} =: \mathbf{P}(z_i)$$

Entscheidungskriterium: [Gen i habe Expression „Tumor“ $\Leftrightarrow \mathbf{P}(z_i) > \gamma$],

wobei γ ein zu wählender Schwellenwert ist.

Dann ist $\mathbf{P}(z_i) = 1 - \mathbb{E}[\widehat{fdr}(z_i) | Y]$.

Im Vergleich: $\mathbf{P}_{\text{emp}}(z_i) = 1 - \widehat{fdr}(z)$.

4 Simulationsstudie

In einer Simulationsstudie werden f_0, f_1, p_0, n wie folgt festgelegt:

$$f_0 = \mathcal{N}(0, 1), f_1 = 0.5\mathcal{N}(-2, 1) + 0.5\mathcal{N}(2, 1), p_0 = 0.8; n = 10\,000$$

Zunächst soll die Performance von $\mathbf{P}(z)$ untersucht werden. Man wird sehen, dass in dieser Studie für $\mathbf{P}(z)$ gute Ergebnisse erzielt werden.

In der unten stehenden Grafik wird deutlich, wie gut sich die Funktion $P(z)$ der Realität (der wahren Dichtefunktion f) tatsächlich anpasst.

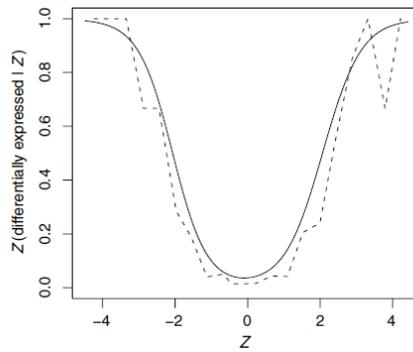


Abbildung 1: Vergleich von $P(z)$ mit der wahren Dichte

Nun kann man sich fragen, wie sich die auf der a posteriori Wahrscheinlichkeit für Tumor-Expression basierende Kategorisierung in „Tumor“ bzw. „normal“ in Abhängigkeit vom wahren zugrunde liegenden p_0 und dem z-Score verhält. Hierfür sei folgende Entscheidungsregel gewählt:

Entscheidungsregel: [Gen mit Score z hat Expression „Tumor“ $\Leftrightarrow P(z) > 0.4$].

Eine fett gedruckte Zahl bedeutet die Einordnung des zugehörigen Gens in Kategorie „Tumor“.

Es ist deutlich zu erkennen, dass mit fallendem Anteil an „normalen“ Genen in der Population auch die Anzahl der als „Tumor“ kategorisierten Gene wächst.

| p_0 | Z | | | | | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|-------------|-------------|-------------|-------------|-------------|
| | -5.00 | -4.00 | -3.00 | -2.00 | -1.00 | 0.00 | 1.00 | 2.00 | 3.00 | 4.00 | 5.00 |
| 0.4 | 1.00 | 1.00 | 0.98 | 0.87 | 0.46 | 0.19 | 0.43 | 0.85 | 0.98 | 1.00 | 1.00 |
| 0.8 | 0.94 | 0.90 | 0.75 | 0.41 | 0.14 | 0.07 | 0.13 | 0.44 | 0.81 | 0.93 | 0.96 |
| 0.95 | 0.46 | 0.42 | 0.27 | 0.11 | 0.05 | 0.03 | 0.04 | 0.10 | 0.28 | 0.43 | 0.50 |

Abschließend ein Vergleich zwischen dem Punktschätzer p_0 , der willkürlich auf den Wert 0.93

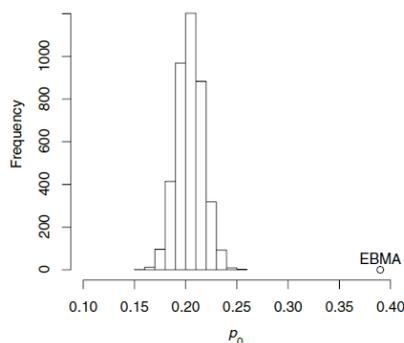


Abbildung 2: Vergleich von p_0 und $f_{p_0|Y}$

festgesetzt wurde, und dem Schätzer $f_{p_0|Y}$ der a posteriori Dichte von p_0 . Es ist schön zu sehen, wie Letzterer die ganze Unsicherheit, die über das wahre p_0 herrscht, beschreibt.

Quellenangaben

K.-A. Do, P. Müller, F. Tang: A Bayesian mixture model for differential gene expression. *App. Statist.* (2005) 54, Part 3, pp. 627-644

B. Efron: Microarrays, Empirical Bayes and the Two Groups Model. *Statistical Science* (2008) 23, No.1, pp. 1-22

J.D. Storey: The positive False Discovery Rate: A Bayesian interpretation and the q-value. *The Annals of Statistics* (2003), Vol. 31, No. 6, pp. 2013-2035

Bootstrap-Methoden zur Ermittlung kritischer Werte für asymptotische FWER-Kontrolle

Mathias Trabs

14.11.2010

1 Wiederholung

Sei $(\Omega, \mathcal{A}, \mathcal{M}, \mathcal{H})$ ein **multiples Testproblem** mit $P \in \mathcal{M}$, \mathcal{M} eine Familie von Wahrscheinlichkeitsmaßen auf (Ω, \mathcal{A}) , und einer Hypothesenmenge $\mathcal{H} = \{H_i : i \in I = \{1, \dots, m\}\}$. Sei weiter $\varphi = (\varphi_i : i \in I)$ ein **multipler Test**.

Die (zufälligen) Anzahlen von wahren / falschen Testentscheidungen können wir darstellen als:

| Hypothesen | Testentscheidung | | |
|------------|------------------|--------|-------|
| | 0 | 1 | |
| wahr | $m_0 - V(P)$ | $V(P)$ | m_0 |
| falsch | $m_1 - S(P)$ | $S(P)$ | m_1 |
| | $m - R(P)$ | $R(P)$ | m |

Def.: $FWER(P) = P(V(P) > 0) = P(\bigcup_{i \in I_0} \{\varphi_i = 1\})$

2 Problemstellung

2.1 Model

Seien X_1, \dots, X_n iid. Zufallsgrößen im \mathbb{R}^J , $X_i = (X_{ij} : j = 1, \dots, J) \sim P$ mit $P \in \mathcal{M}$ unbekannt. Dabei sind $(X_{ij})_{j=1, \dots, J}$, $i \in \{1, \dots, n\}$, **unspezifiziert korreliert**. Wir möchten beispielsweise **Lokationsparameter** der Form $\psi(P) = (\psi_i : i = 1, \dots, m)$ untersuchen.

Bsp.: Sei $X \sim P$ mit Werten in \mathbb{R}^J und $Y := g(X) : \mathbb{R}^J \rightarrow \mathbb{R}^m$. Dann wählen wir $\psi(P) = \mathbb{E}[Y]$, d.h. $\psi_i = \mathbb{E}[Y_i]$.

Wir haben **Teststatistiken** $T_n = (T_{ni} : i = 1, \dots, m) \in \mathbb{R}^m$ als Funktionen von X_1, \dots, X_n und bezeichnen deren wahre Verteilung mit $Q_n = Q_n(P)$.

Unsere **Testentscheidung** ist gegeben durch:

- H_i annehmen, falls $T_{ni} \leq c_i$,
- H_i ablehnen, falls $T_{ni} > c_i$,

mit den kritischen Werten $c \in \mathbb{R}^m$. Eine **multiple Testprozedur (MTP)** ist dann die (zufällige) Teilmenge $\mathcal{R}_n \subseteq I$ der abgelehnten Hypothesen.

Gilt $c_i = c$ für $i = 1, \dots, m$, heißt \mathcal{R}_n Simultantest.

Bsp.: Y und $\psi(P)$ wie oben. Hypothesen: $H_i = \{\psi_i(P) = \mathbb{E}[Y_i] \leq \psi_{0i}\}$, $i = 1, \dots, m$ mit einem Nullwert $\psi_0 \in \mathbb{R}^m$. Dann wählen wir die t-Statistiken:

$$T_{ni} = \frac{\text{Schätzer} - \text{Nullwert}}{\text{Standardfehler}} = \sqrt{n} \frac{\psi_{ni} - \psi_{0i}}{\sigma_{ni}}.$$

2.2 Typ-I-Fehlermaße

Die Theorie baut auf **Fehlermaßen** $\Theta(F_{V_n}) \in [0, 1]$ auf, die als Funktionen von der Verteilung der Anzahl der Typ-I-Fehler V_n definiert sind. Dabei ist F_{V_n} die Verteilungsfunktion von V_n auf $\{0, \dots, m\}$.

Insbesondere betrachten wir die **FWER**:

$$\Theta(F_{V_n}) = FWER(P) = P(V_n > 0) = 1 - F_{V_n}(0).$$

Seien F_1, F_2 zwei Verteilungsfunktionen auf $\{0, \dots, m\}$ und $d(F_1, F_2) := \max_{x \in I} |F_1(x) - F_2(x)|$ deren Abstand.

Wir machen folgende **Annahmen** an Θ :

- **(AMI)** Monotonie:

$$F_1 \geq F_2 \Rightarrow \Theta(F_1) \leq \Theta(F_2)$$

- **(ACI)** Stetigkeit bei (F_n) : Sei (F_n) eine Folge von Verteilungsfunktionen auf $\{0, \dots, m\}$ gegeben, dann soll für beliebige Verteilungsfunktionen (G_n) auf $\{0, \dots, m\}$ gelten:

$$\lim_{n \rightarrow \infty} d(F_n, G_n) = 0 \Rightarrow \lim_{n \rightarrow \infty} (\Theta(G_n) - \Theta(F_n)) = 0$$

In den meisten Fällen genügt in der (ACI)-Annahme $(F_n) = F$, für eine Verteilungsfunktion F .

3 Fehlerkontrolle und Wahl der Nullverteilung

Definition: Eine MTP $\mathcal{R}_n = \mathcal{R}(T_n, Q_0, \alpha)$ **kontrolliert das Niveau** $\alpha \in (0, 1)$ (**strikt**), falls

$$\Theta(F_{V_n}) \leq \alpha, \quad (FWER(P) \leq \alpha).$$

\mathcal{R}_n **kontrolliert das Niveau** $\alpha \in (0, 1)$ **asymptotisch**, falls

$$\limsup_{n \rightarrow \infty} \Theta(F_{V_n}) \leq \alpha.$$

V_n hängt von der wahren Verteilung $Q_n = Q_n(P)$ der Teststatistiken T_n ab, aber Q_n ist i.A. unbekannt und muss durch ein **Nullverteilung** Q_0 geschätzt werden (um kritische Werte zu ermitteln).

Seien Teststatistiken T_n mit wahrer Verteilung Q_n , einer m -dimensionalen Nullverteilung Q_0 zur Berechnung kritischer Werte, sowie eine Niveau α gegeben.

Für die gesamte Anzahl der abgelehnten Hypothesen R und die Anzahl der abgelehnten wahren Hypothesen V schreiben wir:

$$\begin{aligned}
R_n &= R(Q_0|Q_n) = |\mathcal{R}(T_n, Q_0, \alpha)|, & T_n &\sim Q_n, \\
R_0 &= R(Q_0|Q_0) = |\mathcal{R}(T_n, Q_0, \alpha)|, & T_n &\sim Q_0, \\
V_n &= V(Q_0|Q_n) = |\mathcal{R}(T_n, Q_0, \alpha) \cap I_0|, & T_n &\sim Q_n, \\
V_0 &= V(Q_0|Q_0) = |\mathcal{R}(T_n, Q_0, \alpha) \cap I_0|, & T_n &\sim Q_0.
\end{aligned}$$

Mit diesen Vorbereitungen können wir nun eine allgemeine Vorgehensweise angeben.

3.1 Road map

1. Null-Dominiertheit für das Typ-I-Fehlermaß $\Theta(F_{V_n})$:
Wähle eine Null-Verteilung Q_0 so, dass.

$$\begin{aligned}
\Theta(F_{V_n}) &\leq \Theta(F_{V_0}) && \text{[strikte Kontrolle]} \\
\limsup_{n \rightarrow \infty} \Theta(F_{V_n}) &\leq \Theta(F_{V_0}) && \text{[asymptotische Kontrolle]}. \quad (1)
\end{aligned}$$

2. Die Anzahl der Typ-I-Fehler ist nie größer als die gesamte Anzahl abgelehnter Hypothesen, damit

$$V_0 \leq R_0 \Rightarrow F_{V_0} \geq F_{R_0} \stackrel{\text{(AMI)}}{\Rightarrow} \Theta(F_{V_0}) \leq \Theta(F_{R_0})$$

3. Kontrolle des Parameters $\Theta(F_{R_0})$, bzgl. der beobachtbaren Anzahl von abgelehnten Hypothesen, unter der Null-Verteilung:

$$\Theta(F_{R_0}) \leq \alpha.$$

Hierbei ist (1) abhängig von Θ und gilt unter folgenden allgemeinen **Null-Dominiertheits-Bedingungen**:

- Q_0 dominiert die Verteilung $F_{V_n} : x \in \{0, \dots, m\}$:

$$\begin{aligned}
F_{V_n}(x) &\geq F_{V_0}(x), \\
\liminf_{n \rightarrow \infty} F_{V_n}(x) &\geq F_{V_0}(x),
\end{aligned}$$

Insbesondere gilt dies, falls

- Q_0 dominiert die gemeinsame Verteilung Q_{n, I_0} des I_0 -Vektors $(T_{ni} : i \in I_0)$:

$$\begin{aligned}
Q_{n, I_0} &\geq Q_{0, I_0}, \\
\liminf_{n \rightarrow \infty} Q_{n, I_0} &\geq Q_{0, I_0}.
\end{aligned}$$

Die erste Ungleichung in (1) folgt aus (AMI), für die zweite benötigen wir ebenfalls (ACI).

4 Umsetzung

4.1 Konstruktion einer MTP

Schreibe für einen kritischen Wert $c \in \mathbb{R}^m$ und eine Verteilung $Q \in \{Q_0, Q_n\}$

$$\begin{aligned} R(c|Q) &= \sum_{i \in I} \mathbf{1}_{\{T_{ni} > c_i\}}, & T_n &\sim Q, \\ V(c|Q) &= \sum_{i \in I_0} \mathbf{1}_{\{T_{ni} > c_i\}}, & T_n &\sim Q. \end{aligned}$$

Für die Null-Verteilung Q_0 auf dem \mathbb{R}^m mit Randverteilungen Q_{0i} und für ein $\delta \in [0, 1]$ definieren wir außerdem den Vektor $d(Q_0, \delta)$ der δ -Quantile:

$$d(Q_0, \delta)_i = Q_{0i}^{-1}(\delta) = \inf\{z : Q_{0i}(z) \geq \delta\}, \quad i = 1, \dots, m.$$

Methode 1: common-quantil

Gegeben eine Null-Verteilung Q_0 und ein Niveau $\alpha \in (0, 1)$, wähle

$$\delta_0(\alpha) = \inf\{\delta : \Theta(F_{R(d(Q_0, \delta)|Q_0)}) \leq \alpha\}.$$

Dann definieren wir die **Ein-Schritt common-quantil multiple Testprozedur** mittels der kritischen Werte

$$c(Q_0, \alpha) = d(Q_0, \delta_0(\alpha)) = (Q_{0i}^{-1}(\delta_0(\alpha)) : i = 1, \dots, m),$$

welche das Typ-I-Fehlermaß $\Theta(F_{V(c(Q_0, \alpha)|Q_n)})$ zum Niveau α kontrolliert:

$$\mathcal{R}(T_0, Q_0, \alpha) = \{i : T_{ni} > c(Q_0, \alpha)_i\}.$$

Theorem 1 (Asymptotische Kontrolle für die common-quantil Methode)

Es existiere eine \mathbb{R}^m -wertige Zufallsvariable $Z \sim Q_0$, so dass für alle $c \in \mathbb{R}^m$ und $x \in \{0, \dots, m\}$ gilt:

$$\liminf_{n \rightarrow \infty} P^{Q_n} \left(\sum_{i \in I_0} \mathbf{1}_{\{T_{ni} > c_i\}} \leq x \right) \geq P^{Q_0} \left(\sum_{i \in I_0} \mathbf{1}_{\{Z_i > c_i\}} \leq x \right) \quad (\text{AQ0})$$

Oder kurz: $\liminf_n F_{V(c|Q_n)}(x) \geq F_{V(c|Q_0)}(x), \forall x$. Weiterhin erfülle die Abb. Θ die Bedingungen (AMI) und (ACI) bei $F_{V(c|Q_0)}$.

Dann kontrolliert die common-quantil Methode mit kritischen Werten $c(Q_0, \alpha) = d(Q_0, \delta_0(\alpha))$ asymptotisch das Typ-I-Fehlermaßes $\Theta(F_{V(c|Q_n)})$ zum Niveau α , d.h.

$$\limsup_{n \rightarrow \infty} \Theta(F_{V(c|Q_n)}) \leq \alpha.$$

Methode 2: common-cut-off

Gegeben eine Null-Verteilung Q_0 und ein Niveau $\alpha \in (0, 1)$, wähle

$$e(Q_0, \alpha) = \inf\{c \in \mathbb{R} : \Theta(F_{R((c, \dots, c)|Q_0)}) \leq \alpha\}.$$

Dann definieren wir die **Ein-Schritt common-cut-off multiple Testprozedur** mittels des kritischen Wertes $e(Q_0, \alpha)$ durch

$$c(Q_0, \alpha) = (e(Q_0, \alpha), \dots, e(Q_0, \alpha)),$$

welche das Typ-I-Fehlermaß $\Theta(F_{V(c(Q_0, \alpha)|Q_n)})$ zum Niveau α kontrolliert:

$$\mathcal{R}(T_0, Q_0, \alpha) = \{i : T_{ni} > c(Q_0, \alpha)_i\}.$$

Vergleich von Common-quantil- und common-cut-off-Methode:

- Beide Methoden sind äquivalent, falls $(T_{ni})_{i=1, \dots, m}$ unter Q_0 identisch verteilt sind.
- Unterschiede in: Balance, Güte und technischer Umsetzbarkeit.
- Wird Q_0 durch Resampling geschätzt (bootstrap) tendiert die common-quantil Methode zur größerer Sensibilität gegenüber der Anzahl der Resampling-Schritte und der Diskretheit der geschätzten Null-Verteilung.

Theorem 2 (Allgemeine Konstruktion der Null-Verteilung) Es seien $\lambda_0 \in \mathbb{R}^m$ und $\tau_0 \in \mathbb{R}_0^m$ so gegeben, dass gilt

$$\limsup_{n \rightarrow \infty} \mathbb{E}[T_{ni}] \leq \lambda_0 \quad \text{und} \quad \limsup_{n \rightarrow \infty} \text{Var}(T_{ni}) \leq \tau_{0i}, \quad i \in I_0.$$

Definiere $\nu_i = \sqrt{\min\left(1, \frac{\tau_{0i}}{\text{Var}(T_{ni})}\right)}$ und einen Zufallsvektor verschobener und skaliertes Teststatistiken

$$Z_{ni} = \nu_i(T_{ni} + \lambda_{0i} - \mathbb{E}[T_{ni}]), \quad i = 1, \dots, m.$$

Falls $Z_n \xrightarrow{w} Z \sim Q_0 = Q_0(P)$, dann gilt für $c \in \mathbb{R}^m, x \in \{0, \dots, m\}$

$$\liminf_{n \rightarrow \infty} P^{Q_n} \left(\sum_{i \in I_0} \mathbf{1}_{\{T_{ni} > c_i\}} \leq x \right) \geq P^{Q_0} \left(\sum_{i \in I_0} \mathbf{1}_{\{Z_i > c_i\}} \leq x \right)$$

Damit gilt (AQ0) für die Nullverteilung Q_0 und Theorem 1 ist anwendbar.

Diskussion von Theorem 2

- Bei einer zusammengesetzten Hypothese H_i wird λ_{0i} am Schwellenwert bestimmt.
- $\lambda_0 \in \mathbb{R}^m$ zur Erzeugung von Statistiken $(Z_{ni})_{i \in I_0}$ die stochastisch größer sind als die $(T_{ni})_{i \in I_0}$ und daher gegen eine Verteilung konvergieren, die (AQ0) erfüllt.

- $\tau_0 \in \mathbb{R}_0^m$ zur Vermeidung einer degenerierter asymptotischer Nullverteilung und unendlicher kritischer Werte.
- λ_0, τ_0 hängen nur von den Randverteilungen der wahren Verteilung von T_n ab.
- hängen λ_0, τ_0 vom unbekanntem P ab, so können sie durch konsistente Schätzer ersetzt werden.
- τ_0 ist für FWER-Kontrolle nicht zwingend nötig.

4.2 Bootstrap-Schätzung der Nullverteilung

Wir schätzen die wahre Verteilung P aus den Daten X_1, \dots, X_n durch P_n^* . Hieraus wird das **bootstrap-sample** generiert: n iid. Realisierungen $X_1^\#, \dots, X_n^\# \sim P_n^*$.

Anschließend erzeugen wir aus dem bootstrap-sample die Teststatistik $(T_i^{\#n})_{i=1, \dots, m}$ und berechnen entsprechend Theorem 2

$$Z_i^{\#n} = \sqrt{\min\left(1, \frac{\tau_{0i}}{\text{Var}_{P_n^*}(T_i^{\#n})}\right)} (T_i^{\#n} + \lambda_{0i} - \mathbb{E}_{P_n^*}[T_i^{\#n}]), \quad i = 1, \dots, m.$$

Die Schätzung der Verteilung von $(Z_i^{\#n})_{i=1, \dots, m}$ erfolgt dann mittels der empirischen Verteilungsfunktion über B bootstrap-samples. Damit haben wir eine Approximation von $Q_0(P)$ (aus Theorem 2).

Dieses Vorgehen ist in folgenden drei Schätzmethoden umgesetzt.

Method 3: Bootstrap-Schätzung der Nullverteilung

1. Erzeuge B bootstrap samples $\{X_{1,b}^\#, \dots, X_{n,b}^\#\}$ für $b = 1, \dots, B$ mit $X_{i,b}^\# \sim P_n^*, i = 1, \dots, n, b = 1, \dots, B$
2. Berechne für jedes bootstrap sample die Teststatistiken $T_{\cdot,b}^{\#n} = (T_{i,b}^{\#n} : i = 1, \dots, m)$, so dass wir eine $m \times B$ -Matrix $T^{\#n} = (T_{i,b}^{\#n})$ erhalten.
3. Berechne zeilenweise Erwartungswerte und Varianzen in der Matrix $T^{\#n}$ um $\mathbb{E}[T_{ni}]$ und $\text{Var}(T_{ni}), i = 1, \dots, m$, zu schätzen.
4. Erzeuge $m \times B$ -Matrix $Z^{\#n} = (Z_{i,b}^{\#n})$ durch zeilenweises Verschieben und Skalieren von $T^{\#n}$
5. Die bootstrap Schätzung Q_{0n} der Nullverteilung Q_0 aus Theorem 2 erhalten wir als empirische Verteilung der Spalten $Z_{\cdot,b}^{\#n}$ der Matrix $Z^{\#n}$.

Methode 4: Bootstrap-Schätzung der common-quantil kritischen Werte

1. Wende Methode 3 an um die Matrix $Z^{\sharp n}$ und die geschätzte Nullverteilung Q_{0n} zu ermitteln.
2. Die bootstrap common-quantil cut-offs sind die Zeilenquantile der Matrix $Z^{\sharp n}$, also die δ -Quantile des B -Vektors $Z_{i,\cdot}^{\sharp n}$:

$$d(Q_{0n,i}, \delta) = Q_{0n,i}^{-1}(\delta) = \inf \left\{ z : \frac{1}{B} \sum_{b=1}^B \mathbf{1}_{\{Z_{i,b}^{\sharp n} \leq z\}} \geq \delta \right\},$$

$$i = 1, \dots, m.$$

3. Für einen Test zum Niveau $\alpha \in (0, 1)$, wird δ gewählt als

$$\delta_{0n}(\alpha) = \inf \{ \delta : \Theta(F_{R(d(Q_{n0}, \delta) | Q_{0n})}) \leq \alpha \}.$$

4. FWER: (min-P)
 - (a) p-Wert-Matrix $P^{\sharp n}$ bestimmen durch Ersetzen der Einträge in $Z^{\sharp n}$ durch deren zeilenweise Ordnungszahlen (groß zu klein).
 - (b) Wähle in jeder Spalte von $P^{\sharp n}$ den kleinsten p-Wert.
 - (c) $(1 - \delta_{0n}(\alpha))$ ist das α -Quantil dieses B-Vektors der kleinsten p-Werte.

Methode 5: Bootstrap-Schätzung der common-cut-offs

1. Wende Methode 3 an um die Matrix $Z^{\sharp n}$ und die geschätzte Nullverteilung Q_{0n} zu ermitteln.
2. Berechne den gemeinsamen kritischen Werte aus Q_{0n} entsprechend

$$c(Q_{0n}, \alpha) = e(Q_{0n}, \alpha) = \inf \{ c \in \mathbb{R} : \Theta(F_{R((c, \dots, c) | Q_{0n})}) \leq \alpha \}$$

3. FWER: (max-T)
 - (a) Bestimme in jeder Spalte von $Z^{\sharp n}$ den größten Wert.
 - (b) $e(Q_{0n}, \alpha)$ ist das $(1 - \alpha)$ -Quantil des B-Vektors der größten Werte.

5 Quelle

Dudoit, van der Laan, Pollard: Multiple Testing. Part I Single-Step Procedures for Control of General Type-I-Error Rates.

1. Einleitung: Problematik der Variablenselektion

Modell: $Y = g(X) + \varepsilon = X\beta + \varepsilon$,

$Y \in \mathbb{R}^n$... Response, Zielgröße, $X \in \mathbb{R}^{n \times m}$... Versuchsplan/Designmatrix

$X_i = (X_{ij})_{j=1, \dots, m} \in \mathbb{R}^m$... Prädiktorenvektor, $\beta = (\beta_1, \dots, \beta_m)'$... Regressionskoeffizienten
 $\varepsilon_i \sim N(0, \sigma^2)$ iid, σ^2 unbekannt.

Aufgabe: Wähle aus $\{1, \dots, m\}$ k Zahlen, belasse $X_{ij_1}, \dots, X_{ij_k}$ im Modell

Als multiples Testproblem aufgefasst: Teste $H_i : \beta_i = 0, i = 1, \dots, m$.

2. Multiplizitätskorrektur

Multiplizitätsproblem: möglich, wenn mehrere simultane Tests am Datenmaterial einer Studie durchgeführt werden (Z.B. bei klinischen Studien)

Beispiel: $\varphi = (\varphi_i : i = 1, 2)$ ein multipler Test, $\varphi_i, i = 1, 2$, stochastisch unabhängig,

$\mathbb{P}(\varphi_j = 1) = 0.05, j = 1, 2$

\Rightarrow : $\text{FWER}(\varphi) > 0.05 = \alpha$

2.1. automatische Multiplizitätskorrektur

empirischer Ansatz:

a priori Wahrscheinlichkeit eines Modells: $\mathbb{P}(M_\gamma | p) = p^{k_\gamma} (1 - p)^{m - k_\gamma}$

-schätze p (z.B. EM-Algorithmus, siehe George und Foster (2000))

-dann wähle Modell mit maximaler a posteriori Wahrscheinlichkeit:

$\mathbb{P}(M_\gamma | Y) \sim \hat{p}^{k_\gamma} (1 - \hat{p})^{m - k_\gamma} \cdot f(Y | M_\gamma)$

Multiplizitäts-Korrektur: Annahme: ex. k wahre Parameter $\beta_i: m \rightarrow \infty \Rightarrow \hat{p} \rightarrow 0$,

mit $\hat{p} = \arg \max_{p \in [0, 1]} \sum_{\gamma} \mathbb{P}(M_\gamma | p) \cdot f(Y | M_\gamma) \Rightarrow$ Modelle mit vielen Parametern unwahrscheinlich

vollständiger Bayes Ansatz:

Annahme: $p \sim \text{Beta}(a, b)$ für a priori Wahrscheinlichkeit für Modell M_γ

$\Rightarrow \mathbb{P}(M_\gamma) = \int_0^1 f(M_\gamma | p) \pi(p) dp = \frac{\beta(a + k_\gamma, b + m - k_\gamma)}{\beta(a, b)}$.

Fall $a = b = 1$: a posteriori Wahrscheinlichkeit $\mathbb{P}(M_\gamma | Y) \sim \frac{1}{m+1} \binom{m}{k_\gamma}^{-1} \cdot f(Y | M_\gamma)$

\Rightarrow : Korrektur, da Wahrscheinlichkeit p nicht in Formel

2.2. Problematik beim empirischen Bayes-Ansatz:

Lemma 1. :

Beim Problem der Variablenselektion gilt:

Hat das Nullmodell die echt größte likelihood, so ist der MLE-Schätzer \hat{p} von p gleich 0.

Hat das vollständige Modell die echt größte likelihood, so ist der MLE-Schätzer \hat{p} von p gleich 1.

Beweis:

-die Wahrscheinlichkeiten $\mathbb{P}(M_\gamma)$ summieren sich über γ zu 1

-likelihood von p erfüllt $f(Y) = \sum_{\gamma} \mathbb{P}(M_\gamma) f(Y | M_\gamma) \leq \max_{\gamma \in \Gamma} f(Y | M_\gamma)$

-Ungleichung ist echt, wenn die Bedingungen des Lemmas gelten, wenn $\mathbb{P}(M_\gamma)$ nicht schon 1 ist

- $\mathbb{P}(M_\gamma | p) = p^{k_\gamma} (1 - p)^{m - k_\gamma} = 1$ ist nur für $p = 1$ oder $p = 0$ möglich,

also dann, wenn das Nullmodell oder das vollst. Modell gelten

-damit gilt Gleichheit und Rest folgt

□

3. Median probability model: - Wahl des optimalen Vorhersagemodells -

Definition 1. :

Existiert die Gesamt-a-posteriori-Inklusionswahrscheinlichkeit

$$p_i \equiv \sum_{\gamma: \gamma_i = 1} \mathbb{P}(M_\gamma | y)$$

für Variable i , so ist das Median probability model M_{γ^*} als das Modell definiert, welches aus denjenigen Variablen besteht, deren a-posteriori-Inklusions-Wahrscheinlichkeit mindestens $\frac{1}{2}$ beträgt.

$$\text{formal: } \gamma_i^* = \begin{cases} 1 & , \quad p_i \geq 1/2 \\ 0 & , \quad \text{sonst} \end{cases}$$

Existenz gewährleistet, wenn:

- 1) alle möglichen Modelle H_i betrachtet werden
- 2) betrachtete Modelle gehören zur Klasse der graphischen Modelle

Definition 2. :

Sei $I(i)$ die zu einer Variablen x_i gehörenden Menge von Parameter-Indizes. Eine Teilklasse linearer Modelle besitzt graphische Modellstruktur, wenn es aus allen Modellen besteht, die die folgende Bedingung erfüllen:

„ Wenn für jedes i die Variable x_i zu dem Modell gehört, dann gehören die Variablen $x_j, j \in I(i)$ ebenfalls dazu.“

- Median probability model erfüllt diese Bedingung

4.Optimalität:

-Bayesianischen Ansatz: Modell mit höchster a-posteriori-Wahrscheinlichkeit ermöglicht beste Datenprädiktion

-diese Wahrscheinlichkeiten sind oft klein \Rightarrow Median probability model ermöglicht Alternative (Barbieri und Berger (2002))

-dazu: Kovarianzmatrix $Q = \mathbb{E}[(X'X)]$ betrachten; Diagonalgetalt \Rightarrow a posteriori Mittelwerte $\tilde{\beta}_\gamma$ erfüllen $\tilde{\beta}_\gamma = H'_\gamma \tilde{\beta}$, wobei $(H)_{ij} = \begin{cases} 1 & , \quad \gamma_i = 1 \text{ und } j = \sum_{r=1}^i \gamma_r \\ 0 & , \quad \text{sonst} \end{cases}$

Theorem 1. :

Sei Q Diagonalmatrix mit Diagonaleinträgen $q_i > 0$, es gelte $\tilde{\beta}_\gamma = H'_\gamma \tilde{\beta}$ und die betrachteten Modellhypothesen besitzen graphische Modellstruktur.

Dann ist das Median probability model das beste Vorhersagemodell.

Beweis:

-Minimiere das quadratische Risiko $R(M_\gamma) = \sum_{i=1}^k \tilde{\beta}_i^2 q_i (\gamma_i - p_i)^2$ über alle möglichen γ

-dies ist genau dann der Fall, wenn $\gamma_i^* = \begin{cases} 1 & , \quad p_i \geq 1/2 \\ 0 & , \quad \text{sonst} \end{cases}$ gilt

-es liegt graphische Modellstruktur vor, also gehört das Median probability model zu den betrachteten

□

-Fall aller möglichen Modelle: Median probability model besitzt Optimalitätseigenschaft:

Korollar 1. : (Kurzform)

Sei Q diagonal mit $q_i > 0$, es gelte $\tilde{\beta}_\gamma = H'_\gamma \tilde{\beta}$ und alle Modelle M_γ werden betrachtet.

Dann ist das beste Vorhersagemodell das Median probability model.

Ist σ^2 bekannt und

$$\mathbb{P}(M_\gamma) = \prod_{i=1}^k (p_i^0)^{\gamma_i} (1 - p_i^0)^{1-\gamma_i}$$

, dann ist das Median probability model das Modell mit der größten a posteriori Wahrscheinlichkeit.

Theorie für ANOVA-Modell im 2-Faktor-Modell:

$y_{ijk} = \mu + a_i + b_j + ab_{ij} + \varepsilon_{ijk}, i = 1, 2, j = 1, 2, k = 1, 2, \dots, K, \varepsilon \sim N(0, \sigma^2)$ iid, σ^2 unbekannt. Kurz:
 $y = X\beta + \varepsilon$

$\Rightarrow Q = 4 * K * I_4$ und Korollar lässt sich anwenden

Literatur

-Scott, James G., Berger, James O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem, The Annals of Statistics

-Barbieri, Maria M., Berger, James O.(2004). Optimal predictive model selection. The Annals of Statistics, Vol. 32, 870-897

”Higher Criticism Thresholding” zur Merkmalsauswahl

(Donoho & Jin, PNAS Vol. 105, No. 39, 14790-14795)

Thorsten Dickhaus

29.11.2010

1. MOTIVATION UND MODELLBILDUNG

Wir betrachten zur Motivation des vorzustellenden Verfahrens Klassifikationsprobleme als Teildisziplin des statistischen Lernens. Gegeben ist ein Trainingsdatensatz der Länge n , modelliert als Realisierung von (Y_i, X_i) , $i = 1, \dots, n$ mit $Y_i \in \{+1, -1\}$ (das Label) und $X_i \in \mathbb{R}^p$ (der Merkmalsvektor) für alle i . Die p -dimensionale Verteilung von einem jeden X_i sei gegeben durch $X_i \sim \mathcal{N}(Y_i\mu, \Sigma)$, wobei $\mu \in \mathbb{R}^p$ Kontrastvektor und Σ Merkmals-Kovarianzmatrix heißt. Aufgabe der Klassifikation ist nun, einen funktionalen Zusammenhang von Y und X anhand der bivariaten mathematischen Stichprobe $(Y_i, X_i)_{i=1, \dots, n}$ zu ”erlernen”. Eine Klasse von Verfahren sind die linearen Klassifikatoren. Sie sind von der Form $L(X) = \sum_{j=1}^p w_j X_j + b$, wobei $w = (w_1, \dots, w_p)^t$ Gewichtsvektor heißt. Für eine neue Realisierung x bestimmt das Vorzeichen von $L(x)$ dann, welches Label ihr angeheftet wird.

Ein sinnvolles Kriterium zur Bestimmung von w ist, dass die Klasse mit höherer a posteriori-Wahrscheinlichkeit ausgewählt werden soll, d. h., $\hat{\ell} = \arg \max_{\ell=1,2} \{p(\ell)f(x|\ell)\}$ mit $p(\ell)$ als einer geeignet gewählten a priori Wahrscheinlichkeit für Klasse ℓ (z. B. könnte $p(\ell)$ die relative Häufigkeit von Klasse ℓ im Trainingsdatensatz sein oder es könnte quasi-objektiv $p(1) = p(2) = 1/2$ gewählt werden). Da nach der Modellbildung von oben $f(x|\ell) = (2\pi)^{-p/2}(\det \Sigma)^{-1/2} \times \exp\{-1/2(x - \mu_\ell)^t \Sigma^{-1}(x - \mu_\ell)\}$ gilt ($\mu_1 = -\mu_2 = \mu$ in unserem Beispiel), folgt sofort (wir lassen in ℓ invariante Terme weg und transformieren streng isoton), dass $\hat{\ell} = \arg \max_{\ell=1,2} \{\ln(p(\ell) - 1/2(x - \mu_\ell)^t \Sigma^{-1}(x - \mu_\ell))\}$ die zu wählende Klasse ist. Wir erhalten also den linearen Klassifikator durch die Diskriminanzfunktion $d(x)$. Dazu sei $d_\ell(x) = \ln(p(\ell) - 1/2(x - \mu_\ell)^t \Sigma^{-1}(x - \mu_\ell))$, $\ell = 1, 2$, und $d(x) = d_1(x) - d_2(x) = [x - 1/2(\mu_1 + \mu_2)]^t \Sigma^{-1}(\mu_1 - \mu_2) - \ln(p(2)/p(1))$. Damit ist der Bayes-optimale Gewichtsvektor w proportional zu $\Sigma^{-1}\mu$. Unsere Rechnung zeigt zudem, dass dies auch für den Maximum Likelihood-basierten linearen Klassifikator gilt. Das Problem ist nun, dass Σ singulär (nicht-invertierbar) wird, falls $p > n$ ist. In einem solchen Fall (viele Merkmale, kleiner Umfang des Trainingsdatensatzes) ist also eine Merkmalsauswahl vonnöten.

2. HIGHER CRITICISM THRESHOLDING

Das ”Higher Criticism” Thresholding ist ein Merkmalsauswahlverfahren. Es hat gute Eigenschaften unter dem sogenannten ”rare/weak feature model”, also in Situationen, in denen es nur wenige informative Merkmale gibt und μ klein ist. Es gebe also $k \ll p$ Elemente von μ ungleich Null, wobei $\varepsilon = k/p$ ein kleiner Anteil, z. B. 1%, ist. Zusätzlich sollen alle nicht-null Elemente von μ einen konstanten Wert μ_0 haben. Mit $\tau = \sqrt{n}\mu_0$ ist damit $\text{RW}(\varepsilon, \tau)$ ein zweiparametriges

Modell. Wir betrachten nun $Z_j = n^{-1/2} \sum_{i=1}^n Y_i X_{i,j}$, $j = 1, \dots, p$, die Teststatistik für die Hypothese $H_j : \text{Cov}(Y, X_j) = 0$. Drei intuitive Strategien zur Merkmalsauswahl sind dann gegeben durch

- (i) Clipping: $\eta^{\text{clip}}(z) = \text{sgn}(z)$,
- (ii) Hard thresholding: $\eta_t^{\text{hard}}(z) = z \mathbf{1}_{\{|z| > t\}}$,
- (iii) Soft thresholding: $\eta_t^{\text{soft}}(z) = \text{sgn}(z)(|z| - t)_+$.

In allen drei Fällen benutzen wir den linearen Klassifikator $\hat{L}_t^*(X) = \sum_{j=1}^p w_t^*(j) X_j$ mit $* \in \{\text{clip}, \text{hard}, \text{soft}\}$ und $w_t^*(j) = \eta_t^*(Z_j)$. Es werden also nur Merkmale in die Klassifikationsfunktion aufgenommen, denen eine hinreichend große Korrelation mit dem Label zugemessen wird, die also genug Label-relevante Information tragen.

Offensichtlich bleibt nun nur noch die Frage offen, wie t (der Schwellenwert, englisch: threshold) zu wählen ist. Wir möchten dabei ein Verfahren benutzen, dass nicht nur jede Komponente X_j für sich genommen bewertet, sondern die Gesamtheit der Information in X würdigt, also der Multiplizität der Fragestellung adäquat Rechnung trägt. Eine Lösung ist die Verwendung der "Higher Criticism" Statistik (siehe unten). John W. Tukey hat das Schlagwort "Higher Criticism" aus der Bibelexegese entlehnt, wo es im Englischen die historisch-kritische Methode, also die kontextbezogene Bibelauslegung, bezeichnet. Beschreiben wir die "Higher Criticism" Methode zunächst abstrakt.

Seien unter unserem generellen Rahmen des multiplen Testens m marginale p -Werte p_i , $i = 1, \dots, m$, gegeben mit p_i iid $\sim \text{UNI}[0, 1]$ unter der Globalhypothese $H_0 = \bigcap_{i=1}^m H_i$. Bezeichne $p_{[1]} \leq p_{[2]} \leq \dots \leq p_{[m]}$ die geordneten p -Werte. Die Theorie für Orderstatistiken gleichverteilter, unabhängiger Zufallsgrößen (vgl. z. B. [Shorack and Wellner, 1986]) liefert, dass

$$\forall i = 1, \dots, m : p_{[i]} \underset{\text{approx.}}{\sim} \mathcal{N} \left(\frac{i}{m}, \frac{i}{m} \left(1 - \frac{i}{m} \right) \right).$$

Um nun einen "Higher Criticism" Test für H_0 zu konstruieren, versuchen wir, Abweichungen von dieser Grenzverteilung (m sei groß) zu erkennen und definieren dazu

$$\begin{aligned} HC(i, p_{[i]}) &= \sqrt{m} \frac{i/m - p_{[i]}}{\sqrt{i/m(1 - i/m)}}, \\ HC^* &= \max_{1 \leq i \leq \alpha_0 m} HC(i, p_{[i]}), \end{aligned}$$

wobei α_0 ein Tuningparameter aus $(0, 1]$ ist. HC^* kann dann als Z-Score zum Prüfen der Globalhypothese H_0 verwendet werden.

Für das Merkmalsauswahlproblem adaptieren wir diese Methodik wie folgt (HC Thresholding). Dazu seien Z_j , $j = 1, \dots, p$ wie zuvor.

- (a) Transformiere die realisierten Werte z_j zu p -Werten, also bilde $p_j = \mathbb{P}(|\mathcal{N}(0, 1)| > |z_j|)$, $j = 1, \dots, p$.
- (b) Maximiere die HC-Funktion über den Index j . Sei \hat{j} der Index mit maximalem HC-Wert. Dann definiere den Schwellenwert für HC Thresholding als $HCT := \hat{t}^{\text{HC}} = |z|_{\hat{j}}$.

Ohne Beweis präsentieren wir den folgenden Satz, der zeigt, dass HCT eine vernünftige Wahl ist.

Theorem 2.1. Sei $F_p(t) = p^{-1} \sum_{j=1}^p \mathbf{1}_{\{Z_j \leq t\}}$. Dann strebt für alle t die Zufallsvariable $F_p(t)$ für $p \rightarrow \infty$ gegen $F_{\varepsilon, \tau}(t) = (1 - \varepsilon)\Phi(t) + \varepsilon\Phi(t - \tau)$, den Wert der wahren Verteilungsfunktion des Zweiklassen-Mischmodells. Definiere

$$\begin{aligned} TPR(t) \equiv TPR(t; \varepsilon, \tau) &= \Phi(t - \tau) + \Phi(-t - \tau), \\ FPR(t) \equiv FPR(t; \varepsilon, \tau) &= 2\Phi(-t), \\ PR(t) \equiv PR(t; \varepsilon, \tau) &= (1 - \varepsilon)FPR(t) + \varepsilon TPR(t) \end{aligned}$$

und betrachte

$$\widetilde{HC}(t; \varepsilon, \tau) := \frac{\varepsilon(TPR(t) - FPR(t))}{\sqrt{PR(t)(1 - PR(t))}} = \frac{PR(t) - FPR(t)}{\sqrt{PR(t)(1 - PR(t))}}.$$

Sei t_0 so, dass $PR(t_0) = \alpha_0$ gilt und sei $T_{HC}(F_{\varepsilon, \tau}) = \arg \max_{t \geq t_0} \widetilde{HC}(t; \varepsilon, \tau)$. Dann gilt

$$\hat{t}^{\text{HC}} \xrightarrow{w} T_{HC}(F_{\varepsilon, \tau}), \quad p \rightarrow \infty.$$

In Worten maximiert \hat{t}^{HC} also im $\text{RW}(\varepsilon, \tau)$ -Modell für gegen unendlich strebende Dimensionalität des Merkmalsraums die gewichtete Anzahl informativer Merkmale, die ausgewählt werden.

References

[Shorack and Wellner, 1986] Shorack, G. P. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

k-FWER

Betrachte ein Testproblem endlicher Anzahl der Nullhypothesen mit $H_i, i = 1, \dots, s$.

Wir nehmen an, dass die Tests für einzelne Hypothese verfügbar sind und wir müssen sie in einem Verfahren kombinieren. Der einfachste Ansatz für dieses Problem ist, dass wir jede Hypothese auf einem Niveau α testen. Aber bei solchem Verfahren steigt die Wahrscheinlichkeit, dass ein oder mehrere wahren Hypothesen verworfen werden, mit s . Wenn die Anzahl der wahren Hypothesen groß ist, sind wir sicher, dass einige von denen verworfen werden. Deswegen lautet der klassische Ansatz, dass wir die Wahrscheinlichkeit von einer oder mehreren Verwerfungen kontrollieren müssen. Diese Wahrscheinlichkeit wird als FWER (in Englisch: familywise error rate) genannt. Der Begriff „family“ bezieht sich auf eine Sammlung der Hypothesen H_1, \dots, H_s und sie werden zusammen getestet. Für die Kontrolle der FWER auf einem Niveau α brauchen wir die Bedingung:

$$FWER \leq \alpha$$

für $\alpha \in (0,1)$ und alle mögliche Kombinationen von falschen und richtigen Hypothesen.

Aber wenn die Anzahl der Tests in die Zehntausend oder Hunderttausend geht, wird die Kontrolle der FWER stärker, so dass die einzelnen Abweichungen der Hypothese mit kleiner Chance entdeckt werden, wenn sie auftreten. Deswegen betrachten wir eine Alternative, damit die Kontrolle der Verwerfungen weniger stark wird und somit bessere Ergebnisse liefern.

Wir betrachten k-FWER, für $k > 1$ ist diese Fehlerquote sehr geeignet, wenn man bereit ist, eine oder mehrere falsche Verwerfungen zu tolerieren, sofern die Anzahl der Verwerfungen kontrolliert ist.

Def.1: Die Daten X seien verfügbar, $\omega \in \Omega$, betrachte $H_i : P \in \omega_i, i = 1, \dots, s$. Sei $I(P)$ die Indexmenge der wahren Nullhypothesen wenn P die wahre Wahrscheinlichkeitsverteilung ist, d.h. $i \in I(P) \Leftrightarrow P \in \omega_i$. Dann heißt die Wahrscheinlichkeit von mindestens k falsche Verwerfungen k-FWER.

$$\text{k-FWER} = P\{\text{Verwerfen von mindestens } k \text{ Hypothesen } H_i \text{ mit } i \in I(P)\}$$

Für die Kontrolle der k-FWER brauchen wir die Bedingung: $\text{k-FWER} \leq \alpha$
Für $k=1$ kommen wir wieder zurück zu der Bedingung der FWER.

Def.2: Betrachte eine Nullhypothese $H : P \in \omega$. Sei S_α der Ablehnbereich auf dem Niveau α . Es gelten

- i) $P\{X \in S_\alpha\} \leq \alpha \quad \forall \alpha \in (0,1), P \in \omega$
- ii) $S_\alpha \subset S_{\alpha'}$ wenn $\alpha < \alpha'$

Dann $p = p(X) = \inf\{\alpha : X \in S_\alpha\}$.

Lemma1: Sei p wie oben definiert

- i) Falls $P \in \omega$, dann $P\{p \leq u\} \leq u$
- ii) $P\{p \leq u\} \geq P\{X \in S_u\}$

Bem.: Wenn S_α so ist, dass $P\{X \in S_\alpha\} = \alpha$, dann ist p gleichverteilt in $(0,1)$, wenn $P \in \omega$

Beweis Lemma1:

Sei $P \in \omega$. $\{p \leq u\}$ impliziert $\{X \in S_{u+\varepsilon}\}$ für ein kleines $\varepsilon > 0$. Also

$P\{p \leq u\} \leq P\{X \in S_{u+\varepsilon}\} \leq u + \varepsilon$. Jetzt lassen wir $\varepsilon \rightarrow 0$. Damit folgt i).

$\{X \in S_u\}$ impliziert $\{p \leq u\}$ und damit folgt ii).

2 klassische Verfahren, die KWER kontrolliert, sind Bonferroni- und Holm-Verfahren. Beim Bonferroni Verfahren wird jede Hypothese H_i verworfen, wenn $p_i \leq \alpha/s$, wobei p_i die Bedingung $P\{p_i \leq u\} \leq u$ für ein kleines beliebiges $u \in (0,1)$ erfüllt.

Holm Verfahren: Seien $p_1 \leq \dots \leq p_s$. Dann ist das Holm Verfahren wie folgt definiert:

Setze $k=0$

Schritt 1: Wenn $p_{k+1} > \alpha / (s - k)$, dann geht zum Schritt 2, andernfalls $k=k+1$ und wiederholen Schritt 1.

Schritt 2: Verwerfen H_j wenn $j \leq k$, nicht verwerfen H_j wenn $j > k$

Das Bonferroni Verfahren ist ein Beispiel für single-step Prozedur, d.h. jede Nullhypothese wird verworfen, wenn deren entsprechenden p-Werten kleiner, gleich dem gemeinsamen Wert (bei Bonferroni ist α / s).

Das Holm Verfahren ist ein Spezialfall der stepdown Prozedur, die wie folgt aussieht:

Seien $\alpha_1 \leq \dots \leq \alpha_s$.

Wenn $p_1 > \alpha_1$, dann wird keine Hypothese verworfen. Andernfalls wenn $p_1 \leq \alpha_1, \dots, p_r \leq \alpha_r$, dann werden die Hypothesen $H_1 \leq \dots \leq H_r$ verworfen.

Das übliche Bonferroni Verfahren vergleicht jeden p-Wert p_i mit α / s . Die Kontrolle der k-FWER erlaubt uns eine Erhöhung von α / s zu $k\alpha / s$ und damit steigt die Wahrscheinlichkeit, falsche Hypothese zu erkennen.

Theorem1: $H_i : P \in \omega_i, i = 1, \dots, s$ und p_i erfüllt die Bedingung $P\{p_i \leq u\} \leq u$. Betrachte das single-step Verfahren, bei dem jede H_i verworfen wird wenn $p_i \leq k\alpha / s$.

- i) Das Verfahren kontrolliert k-FWER, so dass $k\text{-FWER} \leq \alpha$. Äquivalenz: Wenn jede Hypothese auf dem Niveau $k\alpha / s$ getestet wird, dann ist k-FWER kontrolliert.
- ii) $k\text{-FWER} \leq \alpha$ ist scharf in dem Sinne, dass es eine gemeinsame Verteilung für p_1, \dots, p_s , so dass $k\text{-FWER} = \alpha$

Beweis: Seien die Nullhypothesen H_i mit $i \in I = I(P)$ wahr und der Rest falsch. $|I|$ ist die Kardinalität von I. Sei N die Anzahl der falschen Verwerfungen. Dann

$$P\{N \geq k\} \leq \frac{E(N)}{k} = \frac{E\left[\sum_{i \in I(P)} I\{p_i \leq k\alpha / s\}\right]}{k} \leq \sum_{i \in I(P)} \frac{k\alpha / s}{k} = |I(P)| \frac{\alpha}{s} \leq \alpha \Rightarrow \text{i)}$$

Um ii) zu beweisen, betrachten wir die folgende Konstruktion:

Wir nennen J die Indexmenge der k zufällig genommenen ohne Zurücklegen Indizes aus der Indexmenge $\{1, \dots, s\}$

Wenn $i \in J$, dann $p_i = U_1$, wobei U_1 auf dem $(0, k/s)$ gleichverteilt ist, d.h. $U_1 \sim U(0, k/s)$

Wenn $i \notin J$, dann $p_i = U_2$, wobei U_2 unabhängig von U_1 und $U_2 \sim U(k/s, 1)$.

Dann

$$p_i \sim \frac{k}{s}U(0, k/s) + (1 - \frac{k}{s})U(k/s, 1) \sim U(0, 1)$$

$$\text{Wenn } u \leq k/s, \text{ dann } P\{p_i \leq u\} = P\{i \in J\} \cdot P\{U_1 \leq u\} = \frac{k}{s} \frac{u}{k/s} = u$$

$$\text{Wenn } u > k/s, \text{ dann } P\{p_i \leq u\} = P\{i \in J\} \cdot 1 + P\{i \notin J\} \cdot P\{U_2 \leq u\} = \frac{k}{s} + (1 - \frac{k}{s}) \frac{u - k/s}{1 - k/s} = u$$

Also, wenn es exakt k-mal $p_i \leq k\alpha / s$ gibt, dann

$$P\{U_1 \leq k\alpha / s\} = \frac{\alpha k / s}{k / s} = \alpha \Rightarrow \text{ii)}$$

Die oben beschriebene single-step Prozedur kann verbessert werden, in dem wir stepdown Prozedur benutzen.

Theorem2: $H_i : P \in \omega_i, i = 1, \dots, s$ und p_i erfüllt die Bedingung $P\{p_i \leq u\} \leq u$. Die stepdown Prozedur mit $\alpha_i = k\alpha / s$ wenn $i \leq k$ und $\alpha_i = k\alpha / (s + k - i)$ wenn $i > k$ kontrolliert k-FWER, d.h.

$k\text{-FWER} \leq \alpha$ gilt.

Beweis: Sei $I(P)$ die Indexmenge der wahren Nullhypothesen. $|I(P)| \geq k$

Nennen wir die entsprechenden p-Werten zu $|I(P)|$ wahren Hypothesen $q_1 \leq \dots \leq q_{|I(P)|}$.

Sei j der kleinste Index, der $p_j = q_k$ erfüllt, also $k \leq j \leq s - |I(P)| + k$

Dann hat die stepdown Prozedur mindestens k falsche Verwerfungen genau dann wenn

$$p_1 \leq \alpha_1, \dots, p_j \leq \alpha_j \Rightarrow q_k = p_j \leq \alpha_j = k\alpha / (s + k - j)$$

Aber $\frac{k\alpha}{s + k - j} \leq \frac{k\alpha}{|I(P)|}$

Also ist die Wahrscheinlichkeit von mindestens k falschen Verwerfungen nach oben beschränkt

durch $P\{q_k \leq \frac{k\alpha}{|I(P)|}\}$. Nach Theorem 1 i) $P\{q_k \leq \frac{k\alpha}{|I(P)|}\} \leq \alpha$

Stepup Prozedur : Seien $p_1 \leq \dots \leq p_s$, $\alpha_1 \leq \dots \leq \alpha_s$

Wenn $p_s \leq \alpha_s$, dann werden alle Hypothesen verworfen, andernfalls werden die Hypothesen

H_1, \dots, H_r verworfen, wenn r der kleinste Index ist, der $p_s > \alpha_s, \dots, p_{r+1} > \alpha_{r+1}$ erfüllt. Wenn für alle r , $p_r > \alpha_r$, dann verwerfen keine Hypothese.

Wenn wir beide stepup und stepdown Prozedure mit den gleichen kritischen Werten betrachten, stellen wir fest, dass stepup Prozedur mindestens soviel Hypothesen verwirft wie stepdown Prozedur. Und wenn wir für beide stepup und stepdown Prozeduren die gleichen Maßstäbe für die Kontrolle der k -FWER haben, ist stepup Prozedur besser in dem Sinne, dass die falschen Verwerfungen besser erkannt werden.

Lemma 2: Sei $|I|$ Indexmenge der wahren Hypothesen, $\alpha_1 \leq \dots \leq \alpha_s$, $q_1 \leq \dots \leq q_{|I|}$ die p-Werten der wahren Hypothesen. Dann gilt für eine stepup Prozedur die Ungleichung:

$$k\text{-FWER} \leq P\left\{ \bigcup_{k \leq j \leq |I|} [q_j \leq \alpha_{s-|I|+j}] \right\}$$

Wenn $\alpha_1 \leq \dots \leq \alpha_s$ gegeben ist, benutzen wir das Lemma 2, um eine stepup Prozedur zu kontrollieren, welche wiederum k -FWER kontrolliert und anschließend definieren wir

$$S_1 = S_1(k, s, |I|) = |I| \frac{\alpha_s - |I| + k}{k} + |I| \sum_{k \leq j \leq |I|} \frac{\alpha_{s-|I|+j} - \alpha_{s-|I|+j-1}}{j} \text{ und}$$

$$D_1 = D_1(k, s) = \max_{k \leq |I| \leq s} S_1(k, s, |I|).$$

Theorem 3: Seien $\alpha_1 \leq \dots \leq \alpha_s$, $H_i : P \in \omega_i, i = 1, \dots, s$ und p_i erfüllt die Bedingung $P\{p_i \leq u\} \leq u$,

$p_1 \leq \dots \leq p_s$. Sei $\alpha'_i = \alpha\alpha_i / D_1(k, s)$, wobei D_1 wie oben definiert. Dann

- i) $k\text{-FWER} \leq \alpha$ für eine stepup Prozedur
- ii) Für jede stepup Prozedur mit kritischen Werten $\hat{\alpha}_i = \alpha\alpha_i / D'(k, s)$, wobei D' Konstante, die $k\text{-FWER} \leq \alpha$ erfüllt, haben wir für jede i $\alpha'_i \geq \hat{\alpha}_i$.

Bevor wir das Theorem beweisen können, brauchen wir noch ein Lemma.

Lemma3: p_i erfüllt die Bedingung $P\{p_i \leq u\} \leq u$ und seien $p_1 \leq \dots \leq p_s$. Für $m \leq s$ setze

$$0 = \beta_0 \leq \beta_1 \leq \dots \leq \beta_m \leq 1. \text{ Dann}$$

i) $P\{\{p_1 \leq \beta_1\} \cup \{p_2 \leq \beta_2\} \cup \dots \cup \{p_m \leq \beta_m\}\} \leq s \sum_{i=1}^m (\beta_i - \beta_{i-1}) / i$

ii) Solange die rechte Seite der obigen Ungleichung ≤ 1 , dann ist diese Grenze scharf in dem Sinne, dass es eine gemeinsame Verteilung für p-Werten gibt, so dass aus der Ungleichung eine Gleichung wird.

Beweis zum Theorem 3: i) Nach Lemma 2 und 3 gilt :

$$\begin{aligned} \text{k-FWER} &\leq P\left\{\bigcup_{k \leq j \leq |I|} \{q_j \leq \alpha'_{s-|I|+j}\}\right\} \leq |I| \frac{\alpha'_{s-|I|+k}}{k} + |I| \sum_{k \leq j \leq |I|} \frac{\alpha'_{s-|I|+j} - \alpha'_{s-|I|+j-1}}{j} = \frac{\alpha}{D_1(k,s)} S_1(k,s | I) \\ &\leq \alpha \end{aligned}$$

ii) Nenne $|I|^*$ die Anzahl der $|I|$ maximalen $S_1(k,s | I)$. Die p-Werten von $s-|I|^*$ falschen Hypothesen ist gleich 0 und die p-Werten der $|I|^*$ wahren Hypothesen werden nach Lemma 3 so konstruiert, dass

$$P\left\{\bigcup_{k \leq j \leq |I|^*} \{q_j \leq \hat{\alpha}_{s-|I|^*+j}\}\right\} = \frac{\alpha}{D'(k,s)} S_1(k,s | I|^*) = \frac{D_1}{D'} \alpha$$

Für solche gemeinsame Verteilung der p-Werten ist die Wahrscheinlichkeit von mindestens k falschen Verwerfungen äquivalent zu der Wahrscheinlichkeit von mindestens $s-|I|^*+k$ falschen Verwerfungen.

$$\text{Also, k-FWER} = P\left\{\bigcup_{k \leq j \leq |I|^*} \{q_j \leq \hat{\alpha}_{s-|I|^*+j}\}\right\} = \frac{D_1}{D'} \alpha$$

So, um die Kontrolle der k-FWER sicherzustellen, muss es der Fall sein, dass $D_1 \leq D'$. Daraus folgt $\alpha'_i \geq \hat{\alpha}_i$ für jeden i.

Quelle:

-Artikel "Generalizations of the familywise Error Rate" von E.L.Lehmann und Joseph P.Romano

-Artikel "Stepup procedures for control of generalizations of the familywise Error Rate" von Joseph P.Romano und Azeem M.Shaikh

Bayesianische FDR -Teil2-

Anna Ferrari

Ziel. Suche nach einer Entscheidungsregel, die uns hilft, die False Discovery Rate und die False Negative Rate abzunehmen im Kontext der Gene und ihre Veränderung.

Daten und Modell.

ω model Parameter, Y die Daten, n zu testende Nullhypothese;

eine Regulationsindikator $H_i = \begin{cases} 1, & \text{falls Gen mit Veränderung} \\ 0, & \text{sonst} \end{cases}$

eine Indikatorvariable $d_i = \begin{cases} 1, & \text{falls Discovery} \\ 0, & \text{ansonsten} \end{cases}$

Sei auch die Variable $v_i = \mathbb{P}(H_i|Y)$, d.h. die Wahrscheinlichkeit, dass Gen i eine Veränderung hat gegeben die Daten (die i -te a-posteriori-Randverteilung).

Setting.

Wir konstruieren wieder die False Discovery und False Negative Rate mit den Daten.

Wir erhalten also.

$$FDR = \frac{\sum d_i(1 - H_i)}{D}$$

Und für die FNR erhalten wir

$$FNR = \frac{\sum (1 - d_i)H_i}{n - D}$$

Da wir H_i aus den Daten schätzen müssen, berechnen wir die Erwartete a-posteriori FDR und FNR wie folgt: wir marginalisieren nach Y und bedingen mit H_i , und zwar:

$$\begin{aligned} \overline{FDR}(d, Y) &= \int FDR(d, H) d\mathbb{P}(H|Y) \\ &= \frac{\sum d_i(1 - v_i)}{D} = \frac{\overline{FD}}{D} \end{aligned}$$

Und für die FNR erhalten wir

$$\begin{aligned} \overline{FNR}(d, Y) &= \int FNR(d, H) d\mathbb{P}(H|Y) \\ &= \frac{\sum (1 - d_i)v_i}{n - D} = \frac{\overline{FN}}{n - D} \end{aligned}$$

Um eine optimale Entscheidung zu treffen, benötigt man eine zu minimierende Verlustfunktion.

Wir haben vier Vorschläge, die uns leiten die FNR und FRD zu minimieren. Um diese Fehler zu minimieren, reicht es auch nur die Nenner zu betrachten, d.h. die \overline{FN} und \overline{FD} .

Wir erhalten folgenden Verlustfunktionen:

1. $L_N(d, Y) = c\overline{FD} + \overline{FN}$
2. $L_R(d, Y) = c\overline{FDR} + \overline{FNR}$

Sie hängen von der gesamten Anzahl von abgeleiteten Hypothesen und von den Negatives. c ist eine Konstante, die die Anzahl der fälschlicherweise abgeleiteten Hypothesen bestrafen soll.

Und eine bivariate Version:

3. $L_{2N}(d, Y) = (\overline{FD}, \overline{FN})$
4. $L_{2R}(d, Y) = (\overline{FDR}, \overline{FNR})$

Wo die Minimierung wird nach \overline{FN} gemacht (bzw. nach \overline{FNR}), während \overline{FD} (bzw. \overline{FND}) durch eine konstante α beschränkt wird: $\overline{FD} \leq \alpha$.

Theorem: Unter die vier Verlustfunktionen ist die optimale Entscheidung gegeben durch:

$$d_i = I(v_i \geq t^*)$$

Wobei t^* die optimal cutoff ist und ist fuer je Verlustfunktion wie folgt bestimmt:

$$t^*_N = \frac{c}{c + 1}$$

$$t^*_R(Y) = v_{(n-D^*)}$$

$$t^*_{2N}(Y) = \min \{s: \overline{FD}(s, Y) = \alpha_N\}$$

$$t^*_{2R}(Y) = \min \{s: \overline{FDR}(s, Y) = \alpha_R\}$$

Wobei $v_{(i)}$ die i -te Orderstatistik ist und D^* die optimale Anzahl der abgelehnte Nullhypothese ist.

Kritik: mit (1) bis (4) Verlustfunktionen sind sowohl die Negatives als auch die Discovery Rates gleichmäßig unerwünscht. Daher ist das obige Modell häufig unanwendbar.

Mit anderen Worten haben wir angenommen, dass falsche Negative (bzw. falsche Positive) immer gleich unerwünscht sind. Aber es ist schlechter, wenn man ein Gen als falsch Negativ klassifiziert hat, das eine kleine Veränderung hat, als die Klassifikation eines Genes als falsch Negative die eine grosse Veränderung hat.

Daher moechten wir also unsere Analysis verfeinern, d.h. wir wollen mit einem Parameter ins unsere Modell diese entsprechende neue Charakteristik einfuehren, die wir Differential Expression eines Genes nennen und mit dem Parametern γ_i bezeichnen.

Es gilt: $\gamma_i = 0$, falls $H_i = 0$ und $\gamma_i > 0$, falls $H_i = 1$

Und die neue Verlustfunktion wird also die folgende.

$$L_\gamma(\gamma, d, H) = - \sum_{i=1}^n d_i \gamma_i + k \sum_{i=1}^n (1 - d_i) \gamma_i + cD$$

Wie oben schaezten wir von den Daten unseren neuen Parameter und nennen den $\bar{\gamma}_i = \mathbb{E}(\gamma_i | Y)$ (das a-posteriori erwartete Expressionsniveau von Gene i ggb. die Daten). Die Konstante k will die falsch Negetive bestreifen und die Konstante c bestreift eine grosse Anzahl von abgelente Nullhypothese.

Analog mit den ersten Velustfunktion, trifft man die optimale Entscheidung:

$$\mathbf{d}_i^* = I \left\{ \bar{\gamma}_i \geq \frac{c}{1+k} \right\},$$

die Eins ist, falls unserer Parameter groesser eine feste cutoff ist. Die feste cutoff ist gegeben durch die Kostanten, die wir vorher gewaehlt haben.

Die Entscheidung bleibt im Wesentlichen die selbe, wenn man den Parameter der Differentialexpression in einer Funktion $f(\gamma)$ transformiert.

Kontrolle der False Discovery Exceedance Rate FDX

Zusammenfassung

Bei multiplen Testproblemen mit hoher Anzahl m zu testender Hypothesen (im 6-stelligen Bereich z.B. bei genetischen Problemstellungen), kann es hilfreich sein, den Anteil fälschlich abgelehnter Hypothesen an insgesamt abgelehnten Hypothesen zu untersuchen, um die Zahl zurecht abgelehnter Hypothesen (Güte) zu erhöhen.

Betrachte die False Discovery Proportion $FDP = FDP(\varphi)$ als Funktion in den von einer multiplen Testprozedur φ abgelehnten (rejected) Hypothesen (bzw. deren Anzahl R):

$$FDP = \begin{cases} \frac{V}{R}, & \text{falls } R > 0 \\ 0, & \text{sonst} \end{cases}$$

(V sei die nicht beobachtbare zufällige Anzahl der fälschlich abgelehnten Hypothesen)

Ziel ist es, diese Zufallsvariable durch ihre Verteilung oder aus der Verteilung abgeleitete Funktionen zu beschreiben, um damit ein Typ-I-Fehlermaß zu definieren, wie zum Beispiel:

False Discovery Exceedance Rate $FDX := FDX_{\mathbb{P}}(c) = \mathbb{P}[FDP > c] \quad c \in (0, 1)$

(auch „survival distribution function“, „tail probability of FDP“, ...)

Im folgenden sind nun 4 Methoden skizziert, mit der man dieses Fehlermaß zu beliebigem Signifikanzniveau α kontrollieren kann.

1 Augmentierungsmethode

1. Führe multiplen Test $\tilde{\varphi} = (\tilde{\varphi}_i)_{i=1, \dots, m}$ durch, der die $FWER$ zum Niveau α kontrolliert.
Seien $(\tilde{p}_{[i]})_{i=1, \dots, m}$ die geordneten (adjustierten) p -Werte, so dass ein Ablehnbereich $\mathcal{R}_0 = \{\tilde{p}_{[1]}, \dots, \tilde{p}_{[R_0]}\}$ vorliegt.
2. Mit $R_0 = |\mathcal{R}_0|$ der Zahl abgelehnten Hypothesen, bestimme zu gegebenem $c \in (0, 1)$

$$k^* = \max \left\{ k \in \{1, \dots, m - R_0\} \mid \text{so dass } \frac{k}{R_0 + k} \leq c \right\}$$

sowie eine Indexmenge

$$K = \{j \in \{1, \dots, m\} \mid \tilde{p}_j = \tilde{p}_{[i]} \text{ für } R_0 < i \leq R_0 + k^*\}$$

dann augmentiere: $\mathcal{R}^+ = \mathcal{R}_0 \cup K$

Der Test $\varphi = (\varphi_j = \mathbb{I}_{\{j \in \mathcal{R}^+\}})_{j=1, \dots, m}$ kontrolliert die FDX .

2 Step-Down-Prozedur

Seien $(p_{[j]})_{j=1,\dots,m}$ die geordneten marginalen p -Werte.
 Folgende Step-Down-Prozedur φ_{SD} kontrolliert die FDX bei c zum Niveau α :

Beginne mit $j = 1$:

1. Überprüfe:

$$p_{[j]} \leq \alpha_j := \frac{(\lceil cj \rceil + 1)\alpha}{m + \lceil cj \rceil + 1 - j}$$

2. JA Lehne die Hypothese zu $p_{[j]}$ ab, gehe zurück zu Schritt 1 mit $j = j + 1$

NEIN Lehne die entsprechenden Hypothesen zu den p -Werten $\{p_{[j]}, \dots, p_{[m]}\}$ nicht ab und beende die Prozedur.

3 Inversionsmethode

Allgemeine Vorgehensweise:

1. Zu jeder Teilmenge $W \subseteq \{1, \dots, m\}$ führe Test φ_W zum Niveau α auf die Hypothese $\{\mathcal{P}_W \sim UNI(0, 1)\}$ durch.

(Abkürzung: $\mathcal{P}_U \sim UNI(0, 1)$ für $(p_i)_{i \in U} \stackrel{\text{i.i.d.}}{\sim} UNI(0, 1)$)

2. Bestimme die Indexmengen

$$\mathcal{U} = \{U \subseteq \{1, \dots, m\} \mid \text{Test: } \varphi_U = 0\}$$

3. Definiere:

$$\overline{FDP}(C) = \begin{cases} \max_{\{U \in \mathcal{U}\}} \frac{|U \cap C|}{|C|}, & \text{falls } C \neq \emptyset \\ 0, & \text{sonst} \end{cases}$$

4. Finde Ablehnbereich \mathcal{R} , so dass $\overline{FDP}(\mathcal{R}) \leq c$

Vorschlag um „short-cut“ zu bekommen:

Zum Testen kann man den $p_{[k]}$ -Test verwenden:

Zu $W = \{w(1), \dots, w(r)\} \subseteq \{1, \dots, m\}$ teste $\mathcal{P}_W \sim UNI(0, 1)$ mit:

$$\varphi_W(p_{w(1)}, \dots, p_{w(r)}) = \begin{cases} 0, & \text{falls } p_{w([k])} \geq q_{B(k, r-k+1)}(\alpha) \text{ oder } r < k \\ 1, & \text{sonst} \end{cases}$$

$q_{B(a,b)}(\alpha)$ sei α -Quantil der Beta-Verteilung.

Damit:

- (a) Bestimme $J(k) = \min_{j=1,\dots,m} \{p_{[j]} \geq q_{B(k, m-j+1)}(\alpha)\}$

- (b) Definiere: $\overline{FDP}(C) = \frac{|\{i_k, \dots, i_{J(k)-1}\}^c \cap C|}{|C|}$
 wobei $\{i_k, \dots, i_{J(k)-1}\}$ die Indizes der p -Werte $\{p_{[k]}, \dots, p_{[J(k)-1]}\}$

(c) Die Schwellwertfunktion $\overline{fdp}(t)$ ist hier:

$$\overline{fdp}(t) = \begin{cases} 1, & \text{falls } t \leq p_{[k-1]} \\ \frac{k-1}{m\hat{F}(t)}, & \text{falls } p_{[k-1]} < t \leq p_{[J(k)]} \\ \frac{m\hat{F}(t) - (J(k) - k)}{m\hat{F}(t)}, & \text{sonst} \end{cases}$$

(\hat{F} ist empirische Verteilungsfunktion der p -Werte)

(d) zu $c \in (0, 1)$ bestimme $T = \sup_t \{\overline{fdp}(t) \leq c\}$, und den Bereich

$$\mathcal{R} = \{j \in \{1, \dots, m\} | p_j \leq T\}$$

\implies Der Test $\varphi = (\varphi_j = \mathbb{I}_{\{j \in \mathcal{R}\}})_{j=1, \dots, m}$ kontrolliert die FDX zum Niveau α .

4 Resampling-basierte Methode

Gegeben: Daten $X = (X_1, \dots, X_n)$ i.i.d. \mathbb{P}

Betrachte Teststatistiken $(T_j(X))_{j=1, \dots, m} \mathbb{Q}$

für multiplen Test $\varphi = (\varphi_j = \mathbb{I}_{\{T_j(X) > K_j\}})_{j=1, \dots, m}$.

$$\text{Sei } K^* = \inf \left\{ K \geq 0 \mid \mathbb{P} \left[\frac{\sum_{i \in I_0} \mathbb{I}_{\{T_i(X) > K\}}}{\sum_{i=1}^m \mathbb{I}_{\{T_i(X) > K\}}} > c \right] \leq \alpha \right\}$$

Mit Bootstrap-samples $(X_b^\#)_{b=1, \dots, B}$ lassen sich

- die Nullverteilung \mathbb{Q}_0 ,
- die Verteilung \mathbb{Q} ,
- eine Indexmenge $\hat{\mathcal{I}}_0 \subseteq \{1, \dots, m\}$ von wahren Hypothesen

schätzen.

Für $b = 1, \dots, B$ (d.h. für jeden Bootstrap-Schritt) definiere:

$$r_b(K) = \frac{\sum_{i \in \hat{\mathcal{I}}_0} \mathbb{I}_{\{T_{i,b}^\# > K\}}}{\sum_{j=1}^m \mathbb{I}_{\{T_{j,b}^\# > K\}}} \quad K \geq 0$$

\implies Schätzung für K^* :

$$\hat{K} = \inf \left\{ K \geq 0 \mid \frac{1}{B} \sum_{b=1}^B r_b(K) \leq \alpha \right\}$$

Die daraus resultierende (common-cut-off)-Testprozedur: $(\varphi_j = \mathbb{I}_{\{T_j(X) > \hat{K}\}})_{j=1, \dots, m}$ ist FDX -kontrollierend.

Compatible Confidence Regions

Calvin Seward

January 17, 2011

1 Motivation and Definitions

Review

In classic test theory, we have:

- A model $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$
- A confidence level $\alpha \in (0, 1)$

Then a confidence region is a set $\mathcal{C} = (C(x) : x \in \mathcal{X})$ with

- $C(x) \subseteq \Theta$
- $\{x : C(x) \ni \vartheta\}$ measurable for all $\vartheta \in \Theta$
- $\mathbb{P}_\vartheta(\{x : C(x) \ni \vartheta\}) \geq 1 - \alpha$

How can one translate this idea to a multiple testing framework in a way that makes sense?

Just as a reminder, $\Phi_\alpha(\mathcal{H})$ is the set of multiple tests on \mathcal{H} with strong control of the family wise error rate, i.e.,

$$\forall \vartheta \in \Theta : \mathbb{P}_\vartheta \left(\bigcup_{i \in I_0(\vartheta)} \{\varphi_i(x) = 1\} \right) \leq \alpha$$

Say that $\varphi \in \Phi_\alpha(\mathcal{H})$ and $C \in \mathcal{C}_{1-\alpha}$, then we per definition have

$$\forall \vartheta \in \Theta : \mathbb{P}_\vartheta \left(\bigcup_{i \in I_0(\vartheta)} \{\varphi_i(x) = 1\} \right) \leq \alpha$$

and

$$\forall \vartheta \in \Theta \quad \mathbb{P}_\vartheta(\{x : C(x) \ni \vartheta\}) \geq 1 - \alpha.$$

Then a natural requirement on a confidence region $C(x)$ is that

$$\forall \vartheta \in \Theta : \mathbb{P}_\vartheta \left(\bigcup_{i \in I_0(\vartheta)} \{\varphi_i(x) = 1\} \cup \{x : C(x) \not\ni \vartheta\} \right) \leq \alpha$$

Intuition: $\bigcup_{i \in I_0(\vartheta)} \{x : \varphi_i(x) = 1\}$ and $\{x : C(x) \not\ni \vartheta\}$ should more or less overlap. This requirement makes sense, because one would hope that rejection a hypothesis and that hypothesis not being in the confidence region are equivalent.

Compatible

Thus we get the following definition

Definition 1. A confidence region $C \in \mathcal{C}_{1-\alpha}$ and a multiple test $\varphi \in \Phi_\alpha(\mathcal{H})$ are said to be **compatible** if $C \subseteq C(\varphi)$, i.e. $C(x) \subseteq \bigcap_{i:\varphi_i(x)=1} K_i$ for all $x \in \mathcal{X}$.

Notice that compatibility implies

$$\forall \vartheta \in \Theta : \mathbb{P}_\vartheta \left(\bigcup_{i \in I_0(\vartheta)} \{x : \varphi_i(x) = 1\} \cup \{x : C(x) \not\subseteq \vartheta\} \right) \leq \alpha,$$

the condition from the motivational slide.

Example 2. Let $X_i \sim N(\vartheta_i, 1)$, $i = 1, 2$ iid, $\vartheta_i \in \mathbb{R}$, $\mathcal{H} = \{H_1, H_2\}$ with $H_i : \vartheta_i = 0$. Define c_1 and c_2 by:

$$\mathbb{P}_0(|X_1| \leq c_1) = 1 - \alpha \quad \text{and} \quad \mathbb{P}_{(0,0)}\left(\max_{i=1,2} |X_i| \leq c_2\right) = 1 - \alpha,$$

note that $c_2 \geq c_1$. Now define a multiple test φ by:

$$\begin{aligned} \varphi_1(x) = 0 &\Leftrightarrow |x_1| \leq c_1 \quad \text{or} \quad |x_2| \leq c_2 \\ \varphi_2(x) = 0 &\Leftrightarrow |x_2| \leq c_1 \quad \text{or} \quad |x_1| \leq c_2 \end{aligned}$$

and define a confidence region $C(x)$ by:

$$C(x) := \{ \vartheta \in \mathbb{R}^2 : \max_{i=1,2} |x_i - \vartheta_i| \leq c_2 \}.$$

Since

$$\mathbb{P}_\vartheta(\{x : C(x) \ni \vartheta\}) = \mathbb{P}_\vartheta(\max_{i=1,2} |X_i - \vartheta_i| \leq c_2) = \mathbb{P}_{(0,0)}(\max_{i=1,2} |X_i| \leq c_2) \geq 1 - \alpha$$

and

$$\begin{aligned} \mathbb{P}_\vartheta \left(\bigcup_{i \in I_0(\vartheta)} \{x_i : \varphi_i(x) = 1\} \right) &\leq 1 - \mathbb{P}_{(0,0)} \left(\exists i : \varphi_i(X) = 0 \right) \\ &= 1 - \mathbb{P}_{(0,0)} \left(\max_{i=1,2} |X_i| \leq c_2 \text{ or } \max_{i=1,2} |X_i| \leq c_1 \right) \\ &= 1 - \mathbb{P}_{(0,0)} \left(\max_{i=1,2} |X_i| \leq c_2 \right) \\ &= \alpha. \end{aligned}$$

All together this gives us $\varphi \in \Phi_\alpha(\mathcal{H})$ and $C \in \mathcal{C}_{1-\alpha}(\mathcal{H})$. So the question is, are φ and C compatible? Consider the case that $\vartheta_1 = 0$ and $\vartheta_2 < -4c_2$:

$$\begin{aligned} &\mathbb{P}_\vartheta(\{x : \varphi_1(x) = 1\} \cup \{x : C(x) \not\subseteq \vartheta\}) \\ &= 1 - \mathbb{P}_\vartheta(\{x : \varphi_1(x) = 0\} \cap \{x : C(x) \ni \vartheta\}) \\ &\geq 1 - \mathbb{P}_{(0,0)}(|X_1| \leq c_1, \max_{i=1,2} |X_i| \leq c_2) - \mathbb{P}_{-4c_2}(|X_2| \leq c_2) \\ &= 1 - \mathbb{P}_{(0,0)}(|X_1| \leq c_1, |X_2| \leq c_2) - \mathbb{P}_{-4c_2}(|X_2| \leq c_2) \\ &= 1 - (1 - \alpha)(1 - \alpha)^{1/2} - \mathbb{P}_{-4c_2}(c_2 \leq X_2 \leq c_2) \\ &> \alpha \end{aligned}$$

so no, they are not compatible.

2 Examples

Extended Correspondence Theorem

With this theorem we can easily construct confidence regions for single step procedures:

Theorem 3. Let $\mathcal{H} = \{H_i : i \in I\}$ be a family of hypothesis. If

- $\varphi \in \Phi_\alpha(\mathcal{H})$
- $C(x) = \bigcap_{i: \varphi_i(x)=1} K_i \quad \forall x \in \mathcal{X}$ where the convention $\bigcap_{i \in \emptyset} K_i = \Theta$ is used

then

$$C = (C(x) : x \in \mathcal{X}) \in \mathcal{C}_{1-\alpha}.$$

and with this, C and φ are compatible.

Model for Tukey Test

Let $X_{i1}, \dots, X_{in} \sim N(\mu_i, \sigma^2)$ be iid observations, $\mu = (\mu_1, \dots, \mu_k) \in \mathbb{R}^k$ unknown. We want find confidence regions for $\vartheta_{ij} := \mu_i - \mu_j$, $1 \leq i < j \leq k$. Let $\bar{Y}_i := \bar{X}_i - \mu_i$, $1 \leq i \leq k$, then it is clear that

$$\bar{X}_i - \bar{X}_j - \vartheta_{ij} = \bar{Y}_i - \bar{Y}_j \quad \forall 1 \leq i < j \leq k$$

and that for the statistic

$$\tilde{T}_{ij}(x) = \sqrt{n} \frac{|\bar{Y}_i - \bar{Y}_j|}{S(x)} \quad \text{with} \quad S(x) = \frac{1}{k(n-1)} \sum_{i=1}^k \sum_{l=1}^n (x_{il} - \bar{x}_i)^2$$

$\max_{1 \leq i < j \leq k} \tilde{T}_{ij}(x)$ is $q_{k,k(n-1)}$ distributed.

Now we can test a candidate μ' to level α using the Tukey principle, because μ' will be rejected if and only if

$$\sqrt{n} \frac{|\bar{y}'_i - \bar{y}'_j|}{S(x)} > q_{k,k(n-1);\alpha}$$

where $\bar{y}'_i = \bar{x}_i - \mu'_i$. If we rewrite this, we see that μ' will be accepted if and only if for all $1 \leq i < j \leq k$:

$$\begin{aligned} |(\bar{x}_i - \bar{x}_j) - \vartheta'_{ij}| &\leq \frac{S(x)}{\sqrt{n}} q_{k,k(n-1);\alpha} \\ \Leftrightarrow \vartheta'_{ij} &\in \left[(\bar{x}_i - \bar{x}_j) - \frac{S(x)}{\sqrt{n}} q_{k,k(n-1);\alpha}, (\bar{x}_i - \bar{x}_j) + \frac{S(x)}{\sqrt{n}} q_{k,k(n-1);\alpha} \right] \end{aligned}$$

Because of the extended correspondence theorem we get a compatible confidence region for our ϑ'_{ij} .

Scheffé test

In the ANOVA model, let $q \leq k$, $a_1, \dots, a_q \in \mathbb{R}^k$ be linear independent and define $\mathcal{L} = \text{span}(a_1, \dots, a_q)$. In the lecture on Multiple Tests we learned that $\forall \mu \in \mathbb{R}^k, \sigma^2 > 0$ we have:

$$\mathbb{P}\left(c^\top \mu \in [c^\top \hat{\mu} - K, c^\top \hat{\mu} + K] \quad \forall c \in \mathcal{L}\right) = 1 - \alpha$$

where $\hat{\mu} := (\bar{X}_1, \dots, \bar{X}_k)^\top$, and

$$K := \sqrt{q \cdot s^2 \sum_{i=1}^k \frac{c_i^2}{n} F_{q, k(n-1); \alpha}}$$

We can, for example, choose $c = (0, \dots, 1, \dots, -1, \dots, 0)$ where the 1, -1 are at i, j to test $H_{ij} : \{\mu_i = \mu_j\}$.

3 Confidence regions for stepwise tests

Why Bother??

- Stepwise procedures are typically more powerful
- But in the past single step procedures were used when confidence regions were required (medical experiments...)
- So if one wants the best of both worlds, a better theory needs to be developed

With the extended correspondence theorem it was very easy to create confidence regions for multiple tests such as the Tukey test. But when one tries to create confidence regions for stepwise tests like the popular Holm test, the extended correspondence theorem is no longer usable, so to construct a meaningful confidence region, a different approach is needed.

General Strategy

We can construct (in theory) a confidence region $C \in \mathcal{C}_{1-\alpha}$ which is compatible with $\psi \in \Phi_\alpha(\mathcal{H})$ using the following two steps:

1. Determine $\varphi \in \Phi_\alpha(\Theta)$ induced by ψ . Then calculate $\mathbb{P}_\vartheta(\varphi_\vartheta = 1)$ for all $\vartheta \in \Theta$. If $\mathbb{P}_\vartheta(\varphi_\vartheta = 1) < \alpha$ try to find a $\tilde{\varphi}_\vartheta \geq \varphi_\vartheta$ where
 - $\alpha \geq \mathbb{P}_\vartheta(\tilde{\varphi}_\vartheta = 1) > \mathbb{P}_\vartheta(\varphi_\vartheta = 1)$
 - ideally we want $\mathbb{P}_\vartheta(\tilde{\varphi}_\vartheta = 1) = \alpha$
2. Then construct $\tilde{C} := C(\tilde{\varphi})$
 - $\forall x \in \mathcal{X} : \tilde{C}(x) \subseteq C(x)$ where $C = C(\varphi)$.
 - \tilde{C} and ψ are compatible.

Problems with stepwise procedures

There are two main problems with the practical implementation of this strategy:

1. Finding a good $\tilde{\varphi}_\vartheta$ for all $\vartheta \in \Theta$.
2. Inverting the set of $\tilde{\varphi}_\vartheta$ into a multiple confidence region \tilde{C} .

To better illustrate how difficult this can be, let us consider the following example.

Example 4.

- $Z_i, i \in I = \{1, \dots, k\}$ permutation-symmetric random variables, with Lebesgue density $f(x) > 0 \forall x \in \mathbb{R}$
- $X_i := Z_i + \vartheta_i, \vartheta_i \in \mathbb{R}$
- $\mathcal{H} = \{H_i : i \in I\}$ with $H_i : \vartheta_i \leq 0$ vs. $K_i : \vartheta_i > 0$

We want to find lower confidence bounds for $\underline{\vartheta}_i$ which are compatible with step-up and step-down procedures. Thus set:

- $\alpha \in (0, 1)$
- $c_1 \leq \dots \leq c_k$ defined by $\mathbb{P}(\max_{1 \leq i \leq j} Z_i \leq c_j) = 1 - \alpha, j \in I$

Here $Z_{1:j}, \dots, Z_{j:j}$ are the order statistics of Z_1, \dots, Z_j and the existence of such d_1, \dots, d_j is discussed in Section 3.2 of Finner (1994, Habilitationsschrift).

We define a step-down test with the rule:

$$\psi_i(x) = 1 \Leftrightarrow \exists r \in I : x_i \geq c_r \text{ and} \\ \forall j \in \{r, \dots, k\} : x_{(j)} \geq c_j$$

Now we must choose a test $\varphi = (\varphi_\vartheta : \vartheta \in \Theta)$ with $C(\varphi) \subset C(\psi)$. To do this, define:

$$\kappa(\vartheta) := \{i \in I : \vartheta_i \leq 0\}, \vartheta \in \mathbb{R}$$

and consider the following two cases:

Case 1: $\kappa(\vartheta) \neq \emptyset$ In this case, define a level- α test by:

$$\varphi_\vartheta(x) = 1 \Leftrightarrow \max_{i \in \kappa(\vartheta)} (x_i - \vartheta_i) \geq c_{|\kappa(\vartheta)|}.$$

This is indeed a level- α test because:

$$\begin{aligned} \mathbb{P}_\vartheta(\varphi_\vartheta(X) = 1) &= \mathbb{P}_\vartheta\left(\max_{i \in \kappa(\vartheta)} Z_i \geq c_{|\kappa(\vartheta)|}\right) \\ &= \mathbb{P}_\vartheta\left(\max_{1 \leq i \leq |\kappa(\vartheta)|} Z_i \geq c_{|\kappa(\vartheta)|}\right) \leq \alpha \end{aligned}$$

Here we used the fact that Z_i is permutations symmetric to swap $\max_{i \in \kappa(\vartheta)}$ for $\max_{1 \leq i \leq |\kappa(\vartheta)|}$.

Case 2: $\kappa(\vartheta) = 0$ Here we define a level- α test by:

$$\varphi_{\vartheta}(x) = 1 \Leftrightarrow \max_{i \in I} (x_i - \vartheta_i) \geq c_k.$$

Again we see that

$$\mathbb{P}_{\vartheta}(\varphi_{\vartheta}(X) = 1) = \mathbb{P}_{\vartheta}(\max_{i \in I} Z_i \geq c_k) = \mathbb{P}_{\vartheta}(\max_{1 \leq i \leq k} Z_i \geq c_k) \leq \alpha$$

Theorem 5. For the Bonferroni-Holm test we have the following lower bounds:

$$\underline{\vartheta}_i(x) = \begin{cases} 0 & \psi_i(x) = 1, m(x) < k \\ L_i(x) & \psi_i(x) = 0 \\ K_i(x) & m(x) = k \end{cases}$$

where $m(x) := |\{i \in I : \psi_i(x) = 1\}|$ and $K_i(x)$ is such that:

$$\forall \vartheta \in \Theta : \mathbb{P}_{\vartheta}(\vartheta_i \geq K_i(x)) \geq 1 - \alpha_k$$

and $L_i(x)$ such that:

$$\forall \vartheta \in \Theta : \mathbb{P}_{\vartheta}(\vartheta_i \geq L_i(x)) \geq 1 - \alpha_{k - \kappa(\vartheta) + 1}$$

Example 4 (cont.). Returning to our example, we can define

$$K_i(x) = x_i - c_k \quad \text{and} \quad L_i(x) = x_i - c_{k - m(x)}.$$

This gives us then

$$\underline{\vartheta}_i(x) = \begin{cases} 0 & \psi_i(x) = 1, m(x) < k \\ x_i - c_{k - m(x)} & \psi_i(x) = 0 \\ x_i - c_k & m(x) = k \end{cases}$$

or equivalently

$$\underline{\vartheta}_i(x) = \begin{cases} \min\{0, x_i - c_{k - m(x)}\} & m(x) < k \\ \max\{0, x_i - c_k\} & m(x) = k. \end{cases}$$

This equivalence is because:

- If $m(x) = k$ obviously $x_i - c_i$ is a lower bound, and since $\forall i : \psi_i(x) = 1$ it follows that $x_i - c_k \geq 0$ so we can use 0 and $x_i - c_k$ as a lower bound.
- If $m(x) < k$ and
 - $\psi_i(x) = 1$ then 0 is a lower bound
 - $\psi_i(x) = 0$ then $x_i - c_{k - m(x)}$ is a lower bound

Fazit

- Compatible confidence regions are easy to create for single step procedures
- For multiple step procedures things are a good bit harder, but not impossible
- But all told it is an interesting question deserving more consideration.

1 Einführung

Wenn man eine große Anzahl von Hypothesen testen will, dann ergeben sich drei zentrale Problemstellungen:

- **Detection Problem:** Existieren wahre Alternativen, also kann die Globalhypothese verworfen werden (Binärer Test)?
- **Selection Problem:** Welche Hypothesen sind Nullhypothesen? Prüfen jedes Einzelhypothesenpaars $H_i = 0$ vs. $H_i = 1$, $i \in I$ (Multipler Test)
- **Estimation Problem:** Schätzung des Anteils wahrer Nullhypothesen aus dem *Selection Problem*

2 Mixture-Model

Wir nehmen an, dass m Tests durchgeführt wurden. Für $i \in \{1, \dots, m\}$ seien P_i die zur jeweiligen Hypothese gehörenden p-Werte. Sei $H_i = 0$, wenn die i-te Hypothese wahr ist und $H_i = 1$, wenn die i-te Hypothese falsch ist. Sei $\pi := \frac{m_1}{m} \in (0, 1)$ der unbekannte Anteil falscher Nullhypothesen zu allen Hypothesen. Dann erhält man mit $(P_i, H_i)_{1 \leq i \leq m}$ iid, $P(H_i = 1) = \pi$, $P_i | H_i = 0 \sim U(0, 1)$ und $P_i | H_i = 1 \sim G_1$ als Verteilung der p-Werte:

$$G(u) = (1 - \pi)u + \pi G_1(u),$$

mit der zugehörigen Dichte:

$$g(u) = 1 - \pi + \pi g_1(u),$$

wobei wir nachfolgend voraussetzen, dass G konkav, $G \in C([0, 1])$ mit $G(0) = 0$ ist.

3 BH95-Prozedur und Criticality

3.1 Benjamini-Hochberg-Prozedur (BH95-Prozedur)

Seien $P_{(1)} \leq \dots \leq P_{(m)}$ die geordneten p-Werte zu den m Tests und sei $P_{(0)} := 0$, sowie $P_{(m+1)} := 1$, sodass $P_{(0)} \leq P_{(1)} \leq \dots \leq P_{(m)} \leq P_{(m+1)}$. Der Schwellenwert t_{BH} der Benjamini-Hochberg Prozedur für vorgegebenes Level α ist definiert durch

$$t_{BH} = \frac{\alpha \hat{t}_m}{m}, \text{ wobei } \hat{t}_m = \max\{i : P_{(i)} \leq \alpha \frac{i}{m}, 0 \leq i \leq m\}.$$

Existiert solch ein Index \hat{t}_m , dann werden alle p-Werte kleiner als t_{BH} abgelehnt.

Sei nun $G_m(t) = \frac{1}{m} \sum_{i=1}^m 1_{\{P_i \leq t\}}$ die empirische Verteilungsfunktion der p-Werte. Dann gilt im stetigen Fall mit $G_m(P_i) \approx \frac{i}{m}$:

$$\hat{t}_m = \sup\{t : P_{(t)} \leq \alpha G_m(t)\} = \sup\{t : G_m(t) \leq \frac{P_{(t)}}{\alpha}\} = \sup\{t : \frac{P_{(t)}}{G_m(t)} \leq \alpha\}.$$

Eine wesentliche Interpretation davon ist, dass \hat{t}_m ein Plug-In-Schätzer, gemäß

$$u_*(\pi, G) = \max\{u : G(u) \leq \frac{u}{\alpha}\} \leq \max\{u : \frac{u}{G(u)} \leq \alpha\}, \text{ denn } \lim_{m \rightarrow \infty} G_m(u) = G(u)$$

ist. u_* ist somit der größte abgelehnte p-Wert für $m \rightarrow \infty$.

Die False Discovery Rate ist definiert durch $FDR(t) = E(FDP(t)) := E(\frac{V(t)}{R(t)})$. Sie ist also der Erwartungswert der False Discovery Proportion (FDP), welche den Anteil der (unbeobachteten) fälschlichen Ablehnungen V zu allen Ablehnungen R angibt. Somit ergibt sich mit $V := \{i \leq m : P_i \leq P_{\hat{t}_m}, H_i = 0 \text{ (i-te Nullhypothese ist wahr)}\}$, $R = \hat{t}_m$ und dem Obigen die FDR als $FDR = (1 - \pi)\alpha$.

3.2 Criticality

Die FDR ist eng verknüpft mit der pFDR durch

$$FDR(t) = pFDR(t)P(R(t) > 0) .$$

Storey zeigte 2004, dass die FDR und pFDR asymptotisch gleich sind für fixierte Ablehnungsbereiche, nämlich das $P(R(t) > 0) \rightarrow 1$ für $m \rightarrow \infty$. Dennoch entsteht eine Situationen für $\alpha < \alpha_* \in (0, 1)$ in der das sogenannte Criticality-Phänomen auftritt. Ist dies der Fall, so sind insbesondere die FDR und pFDR asymptotisch nicht mehr äquivalent, d.h. $P(R(t) = 0) > 0$ für $m \rightarrow \infty$.

Der **kritische Wert** für die BH95-Prozedur ist definiert durch:

$$\alpha_* := \inf_{u>0} \frac{u}{G(u)} \leq 1, \text{ wobei } \beta_* := (1 - \pi)\alpha_*$$

Da nach Voraussetzung G konkav, sowie $G(0) = 0$ und $u \mapsto \frac{u}{G(u)}$ nicht wachsend auf $[0, 1]$ ist

$$\implies \alpha_* := \inf_{u>0} \frac{u}{G(u)} = \lim_{u \rightarrow 0} \frac{u}{G(u)} = \lim_{u \rightarrow 0} \frac{1}{g(u)},$$

sowie

$$\beta_* = \frac{1 - \pi}{g(u)} \implies \alpha_* = \frac{\beta_*}{1 - \pi} \implies \alpha_* = 0 \iff \beta_* = 0$$

Bemerkung: Man kann also die Betrachtungen auf den Fall $\alpha_* = 0$ oder nicht reduzieren.

Für die BH95-Prozedur hatten wir im asymptotischen Fall: $u_*(\pi, G) = \max\{u : \frac{u}{\alpha} \leq G(u)\}$ als Grenzwert des größten abgelehnten p-Wert und somit $p_* = G(u_*) = \frac{u_*}{\alpha}$ als Grenzwert des Anteils der p-Werte.

Satz 1: $\alpha < \alpha_*$ (Criticality-Phänomen)

Sei $\alpha < \alpha_*$. Wenn die Menge m der Hypothesen gegen unendlich strebt, dann ist \hat{t}_m beschränkt im Sinne von:

$$\hat{t}_m \xrightarrow{d} \tau > 0, \tau \in \mathbb{N}$$

Mit anderen Worten: bezeichne $p_*(\alpha) = \frac{\hat{t}_m}{m}$ den Anteil der Ablehnungen durch die Benajmini-Hochberg-Prozedur, dann konvergiert $p_*(\alpha)$ gegen 0 für $m \rightarrow +\infty$. Es gilt außerdem:

$$\text{pFDR} \rightarrow (1 - \pi)\alpha_* =: \beta_*$$

Es kann also keine Prozedur eine pFDR kleiner als $\beta_* = \inf_{t>0} \text{pFDR}(t) = \inf_{t>0} \frac{(1-\pi)t}{G(t)}$ erreichen. Die FDR und pFDR sind also asymptotisch nicht mehr äquivalent, denn es folgt für ein multiples Testproblem mit $\beta_* > 0$, dass jede Prozedur zum Level $\alpha \leq \beta_*$:

$$P(R(t) = 0) = 1 - \frac{FDR(t)}{\text{pFDR}(t)} \geq 1 - \frac{\alpha}{\beta_*} > 0$$

Satz 2: $\alpha > \alpha_*$

Sei $\alpha > \alpha_*$ und $\delta := 1 - \alpha g(u_*) > 0$ und sei $q_* = 1 - p_*$. Dann gilt:

$$\limsup_m \pm \frac{\hat{t}_m - mp_*}{\sqrt{m \log(\log(m))}} = \frac{\sqrt{2p_*q_*}}{\delta} := p_\infty, \text{ f.s.}$$

und $p_*(\alpha)$ ist asymptotisch proportional zur Macht:

$$M_m = \frac{\hat{t}_m}{m} \left(\frac{1 - \alpha}{\pi} + \alpha \right) + o_p(1) \rightarrow G_1(u_*)$$

Mit anderen Worten heißt das, dass in diesem Fall $p_*(\alpha)$ gegen einen positiven Wert p_∞ für $m \rightarrow \infty$ konvergiert.

Bemerkung:

Wählt man $\alpha = \alpha_*$, dann ist das optimal in dem Sinn, dass für $m \rightarrow \infty$ die pFDR die untere Schranke β_* erreicht und gleichzeitig die Anzahl der wahren Entdeckungen unbeschränkt ist.

Man kann Criticality auch charakterisieren, indem man den Likelihoodquotienten der Dichten betrachtet.

Lemma: Criticality und Likelihood-Quotienten

Sei G_1 konkav und die Verteilungsfunktion der Teststatistik symmetrisch. Dann gilt:

1. Wenn $\frac{f_1}{f_0}(t)$ beschränkt ist für $t \rightarrow \infty$, dann hat die Dichte g_1 der p -Werte unter der Alternative eine endlichen Grenzwert bei 0 (Notation: $g_1(0)$). In diesem Fall entsteht Criticality. Der kritische Wert is gegeben durch:

$$\alpha_* = \frac{1}{1 + \pi + \pi g_1(0)}.$$

2. Wenn $\frac{f_1}{f_0}(t) = +\infty$ für $t \rightarrow \infty$, dann gilt $\lim_{u \rightarrow 0} \frac{G(u)}{u} = +\infty$ und $\alpha_* = 0$. Dann entsteht keine Criticality und alle gewünschten FDR-Level sind erreichbar.

(Dieses Lemma gilt sowohl für einseitige, als auch zweiseitige p -Werte)

3.3 Beispiele

Gauss-Test:

Seien m Statistiken gemäß $N(0, 1)$ verteilt unter der Nullhypothese und gemäß $N(\mu, 1)$ mit $\mu \neq 0$ unter der Alternative. Dann ist der Likelihoodquotient gegeben durch:

$$\frac{f_1}{f_0}(t) = \exp\left(-\frac{1}{2}(t - \mu)^2 + \frac{1}{2}t^2\right) = \exp\left(-\frac{\mu^2}{2} + \mu t\right).$$

Da der Likelihood-Quotient nicht beschränkt ist für $t \rightarrow \infty$, folgt aus dem Lemma, dass keine *Criticality* entsteht, also $\alpha_* = 0$.

Laplace-Test:

Sei die Dichte der Teststatistik gegeben durch $f_0 : t \rightarrow \frac{1}{2}e^{-|t|}$ unter der Nullhypothese und $f_1 : t \rightarrow \frac{1}{2}e^{-|t-\theta|}$ unter der Alternative mit $\theta > 0$. Dann ist der Likelihood-Quotient gegeben durch

$$\begin{aligned} \frac{f_1}{f_0}(t) &= e^{2t-\theta}, \text{ falls } t \leq \theta \\ \frac{f_1}{f_0}(t) &= e^\theta, \text{ falls } t > \theta. \end{aligned}$$

Da der Likelihood-Quotient beschränkt ist, entsteht *Criticality* und der kritische Wert ist

$$\alpha_* = \frac{1}{\pi e^\theta + (1 - \theta)}$$

und die BH95-Prozedur hat asymptotisch Macht Null für $\alpha < \alpha_*$.

4 Higher-Criticism

Im Rahmen des Detection Problems, also der Möglichkeit zu entscheiden ob, ein Signal *Sparse-Data* (seltene Signale) enthält oder nicht untersucht man die **Detection Boundary**, welche die asymptotische Grenze ist, über der wir Signale entdecken und unter der es unmöglich ist Signale auszumachen. Dazu betrachten wir folgendes Modell.

4.1 Sparse-Mixture-Model

Wir betrachten das Gauss-Modell. Seien m unabhängige Beobachtungen $X_i \sim N(\mu_i, 1)$ gegeben und nur ein kleiner Anteil der μ 's ist merklich von Null verschieden.

Nach Voraussetzung ist der Vektor $\mu = (\mu_1, \dots, \mu_m)$ sparse i.S.v.

$$\frac{\#\{i : \mu_i \neq 0\}}{m} \leq \pi \approx 0$$

Im Gegensatz zum Mixture-Model von oben und der Frage: welche X_i sind gemäß $N(\mu, 1)$ verteilt, also **Simultanes Testen von m Hypothesen**

$$H_i = 0 : X_i \sim N(0, 1)$$

$$H_i = 1 : X_i \sim N(\mu, 1), \mu > 0$$

untersuchen wir hier die Frage ob π gleich Null ist, betrachten also das **Binäres-Test-Problem**

$$H_0 : X_i \sim N(0, 1), 1 \leq i \leq m \text{ iid}$$

$$H_1^m : X_i \sim (1 - \pi)N(0, 1) + \pi N(\mu, 1) := F, 1 \leq i \leq m$$

H_0 bezeichne hierbei die Globalhypothese: alle X_i sind nach $N(0, 1)$ verteilt und H^m die Alternative: es existiert ein i , so dass X_i nach F verteilt ist.

Um *Sparsity* und eine schwache Signalstärke zu quantifizieren betrachtet man die Parameter r und β und parametrisiert wie folgt.

$$\pi = \pi_m = m^{-\beta}, \mu = \mu_m = \sqrt{2r \log(m)}, \frac{1}{2} < \beta < 1, 0 < r < 1$$

und definiert dann die **Detection-Boundary** durch:

$$\rho_*(\beta) = \beta - \frac{1}{2}, \text{ für } \frac{1}{2} < \beta \leq \frac{3}{4}$$

$$\rho_*(\beta) = (1 - \sqrt{1 - \beta})^2$$

Der nachfolgende Satz charakterisiert die Detection-Boundary. Sei $LR_{m,i} = \log(1 - \pi + \pi e^{\mu_m x_i - \mu_m^2/2})$. Dann ist der Log-Likelihood-Quotient für das obige Binäre Testproblem $LR_m = \sum_{i=1}^m LR_{m,i}$ und somit gilt:

Satz 3:

1. Sei $r > \rho_*(\beta)$ und betrachte den Likelihood-Quotienten-Test der H_0 ablehnt, wenn

$$LR_m > 0, \text{ dann:}$$

$$P_{H_0}(\{\text{lehne } H_0 \text{ ab}\}) + P_{H_1^m}(\{\text{akzeptiere } H_0\}) \xrightarrow{m \rightarrow \infty} 0$$

2. Sei $r < \rho_*(\beta)$, dann:

$$P_{H_0}(\{\text{lehne } H_0 \text{ ab}\}) + P_{H_1^m}(\{\text{akzeptiere } H_0\}) \xrightarrow{m \rightarrow \infty} 1$$

Bemerkung:

Es gibt also eine Schwellenwert-Effekt für den Likelihood-Quotienten-Test: Die Summe der Typ I- und Typ II-Fehler konvergiert gegen Null oder hängt davon ab, ob μ die Detection-Boundary überschreitet oder nicht. Das heißt, es gibt eine Funktion $\rho_*(\beta)$, so dass wenn

1. $r > \rho_*(\beta)$: H_0 and H_1^m trennen sich asymptotisch,
2. $r < \rho_*(\beta)$: H_0 and H_1^m gehen asymptotisch zusammen.

Die Detection-Boundary $r = \rho_*(\beta)$, als Funktion der Signalstärke, definiert also eine Grenze, welche die *entdeckbare* von der *nicht-entdeckbaren* Region trennt.

Handout

1 Einführung

Modell

Wir betrachten eine mathematische Stichprobe X_1, \dots, X_n iid im \mathbb{R}^J , $X_i = (X_i(1), \dots, X_i(J)) \sim P$, unbekannt; P liegt in einer Menge von Verteilungen \mathcal{M} , dem statistischen Modell.

M Nullhypothesen: $H_0(1), \dots, H_0(M)$
 $H_0(m) := \mathbb{1}_{(P \in \mathcal{M}(m))}$ mit $\mathcal{M}(m) \subseteq \mathcal{M}$

M Alternativen: $H_1(1), \dots, H_1(M)$
 $H_1(m) := \mathbb{1}_{(P \notin \mathcal{M}(m))}$

$\mathcal{H}_0 := \{m \mid H_0(m) = 1\}$ Menge der wahren Nullhypothesen

$\mathcal{H}_1 := \{m \mid H_1(m) = 1\}$ Menge der falschen Nullhypothesen

$T_n(1), \dots, T_n(M)$ Teststatistiken, $(T_n(1), \dots, T_n(M)) \sim Q_n(P)$ unbekannte Verteilung im \mathbb{R}^M
Die Nullverteilung Q_0 schätzt $Q_n(P)$.

$V := V(n, M) :=$ Anzahl der Fehler 1. Art (wahre Nullhypothese verworfen), $FWER := \mathbb{P}(V \geq 1)$

Ziel:

\mathcal{H}_0 schätzen und dabei asymptotische FWER-Kontrolle gewährleisten, also $\limsup_{n \rightarrow \infty} \mathbb{P}(V_n \geq 1) \leq \alpha$.

Schrittweise Step-Down Verfahren

Für einseitige Tests ist folgende Form üblich: $T_m(X_1, \dots, X_n) > c(m) \Rightarrow H_0(m)$ wird abgelehnt

Einschrittverfahren: kritische Werte $c(m) = c(Q_0, \alpha)(m)$ sind unabhängig von den anderen Tests (siehe Vortrag von Mathias Trabs)

Schrittweise Verfahren: kritische Werte $c(m) = c(T_n, Q_0, \alpha)(m)$ dürfen von Statistiken (Daten) abhängen (z.B. die im Folgenden behandelten Step-Down Verfahren)

Hypothesen zu den *signifikantesten* Teststatistiken werden sukzessive geprüft. Sobald eine Nullhypothese akzeptiert wird, werden alle weiteren Hypothesen ungeprüft ebenfalls akzeptiert.

Signifikanz ist z.B. die Größe der absoluten Teststatistik (\rightsquigarrow maxT-Verfahren) oder die geringste Größe der unbereinigten p -Werte (\rightsquigarrow minP-Verfahren).

2 Step-Down minP

Verfahren

Seien $(Q_{0,1}, \dots, Q_{0,M})$ die stetigen Randverteilungen der Nullverteilung Q_0 .

Wir definieren unadjustierte p -Werten als:

$$P_{0n}(m) := 1 - Q_{0m}(T_n(m)), \quad P_0(m) := 1 - Q_{0m}(Z(m)) \quad \text{mit } Z = (Z(1), \dots, Z(M)) \sim Q_0$$

Offensichtlich gilt $P_{0n}(m) \in [0, 1]$. Außerdem nehmen wir an, dass die Werte bereits sortiert vorliegen, also o.B.d.A.: $P_{0n}(1) \leq \dots \leq P_{0n}(M)$.

Für eine Teil-Indexmenge $\mathcal{A} \subseteq \{1, \dots, M\}$ und ein multiples Niveau α definieren wir α -Quantile

$$c(\mathcal{A}) := c(\mathcal{A}, Q_0, \alpha) := F_{\mathcal{A}, Q_0}^{-1}(\alpha) := \inf \{z \mid F_{\mathcal{A}, Q_0}(z) \geq \alpha\}, \quad \text{mit } F_{\mathcal{A}, Q_0}(z) := \mathbb{P}_{Q_0} \left(\min_{m \in \mathcal{A}} P_0(m) \leq z \right)$$

(Wir sehen: Mit wachsendem \mathcal{A} ist $c(\mathcal{A})$ monoton fallend.)

Für Teilmengen $\mathcal{A}_m := \{m, \dots, M\}$ definieren wir $C_n(m) := c(\mathcal{A}_m, Q_0, \alpha) := F_{\mathcal{A}_m, Q_0}^{-1}(\alpha)$

Schließlich konstruieren wir die kritischen Werte als:

$$c_1 := C_n(1)$$

$$c_m := \begin{cases} C_n(m) & , \text{ falls } P_{0n}(m-1) < c_{m-1} \\ 0 & , \text{ sonst} \end{cases}$$

Entscheidungsregel:

Lehne die Nullhypothese $H_0(m)$ zum m -ten signifikantesten (d.h. kleinsten) unadjustierten p -Wert ab, falls $P_{0n}(m) < c_m$. Kurz: $\mathcal{R}(T_n, Q_0, \alpha) = \{m \mid P_{0n}(m) < c_m\}$.

asymptotische Kontrolle der FWER

Annahme AP1 - asymptotische Null-Dominanz

$$\exists \text{ Nullverteilung } Q_0: \quad \limsup_{n \rightarrow \infty} \mathbb{P}_{Q_n} \left(\min_{m \in \mathcal{H}_0} P_{0n}(m) < x \right) \leq \mathbb{P}_{Q_0} \left(\min_{m \in \mathcal{H}_0} P_0(m) < x \right) \quad \forall x \in \mathbb{R}$$

Theorem 1 - asymptotische Kontrolle der FWER

Es sei Annahme AP1 erfüllt. Dann gilt für das Step-Down minP Verfahren: $\limsup_{n \rightarrow \infty} \mathbb{P}(V_n \geq 1) \leq \alpha$

[Beweis:

$m_n := \min_{m \in \mathcal{H}_0} (m \in \{1, \dots, M\})$ Index der wahren Nullhypothesen mit dem kleinsten unber. p -Wert.

$$\Rightarrow P_{0n}(m_n) = \min_{m \in \mathcal{H}_0} P_{0n}(m), \quad \{1, \dots, m_n - 1\} \subseteq \mathcal{H}_1$$

Somit:

$$\begin{aligned} \mathbb{P}(V_n \geq 1) &= \mathbb{P}(m_n \in \mathcal{R}_n) && \text{(sonst alle folgenden auch angenommen)} \\ &\leq \mathbb{P} \left(\min_{m \in \mathcal{H}_0} P_{0n}(m) < c(\mathcal{A}_{m_n}, Q_0, \alpha) \right) && \text{(Es könnte auch schon eine Hypothese zuvor} \\ &&& \text{angenommen worden sein.)} \\ &\leq \mathbb{P} \left(\min_{m \in \mathcal{H}_0} P_{0n}(m) < c(\mathcal{H}_0, Q_0, \alpha) \right) && \text{denn: } \mathcal{H}_0 \subseteq \mathcal{A} = \{m_n, \dots, M\} \\ &&& \text{und } c(\mathcal{A}, Q_0, \alpha) \text{ fallend mit wachsendem } \mathcal{A} \\ &&& \text{also } c(\mathcal{A}_{m_n}, Q_0, \alpha) \leq c(\mathcal{H}_0, Q_0, \alpha) \end{aligned}$$

Also:

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} \mathbb{P}(V_n \geq 1) &\leq \limsup_{n \rightarrow \infty} \mathbb{P} \left(\min_{m \in \mathcal{H}_0} P_{0n}(m) < c(\mathcal{H}_0, Q_0, \alpha) \right) \\
 &\stackrel{(AP1)}{\leq} \mathbb{P}_{Q_0} \left(\min_{m \in \mathcal{H}_0} P_0(m) < c(\mathcal{H}_0, Q_0, \alpha) \right) \\
 &= \mathbb{P}_{Q_0} \left(\min_{m \in \mathcal{H}_0} P_0(m) < \inf \left\{ z \mid \mathbb{P}_{Q_0} \left(\min_{m \in \mathcal{H}_0} P_0(m) \leq z \right) \geq \alpha \right\} \right) \\
 &\leq \alpha
 \end{aligned}$$

└

exakte asymptotische Kontrolle der FWER

Annahme AP2 - asymptotische Trennbarkeit von wahren und falschen Nullhypothesen

- $\forall \epsilon > 0 : \lim_{n \rightarrow \infty} \mathbb{P}_{Q_n} \left(\max_{m \in \mathcal{H}_1} P_{0n}(m) \leq \epsilon \right) = 1$
- $\lim_{\epsilon \downarrow 0} \lim_{n \rightarrow \infty} \mathbb{P}_{Q_n} \left(\min_{m \in \mathcal{H}_0} P_{0n}(m) \leq \epsilon \right) = 0$
- $\forall \alpha \in (0, 1) : \min_{\mathcal{A} \subseteq \{1, \dots, M\}} c(\mathcal{A}, Q_0, \alpha) > 0$

Bemerkung

Wenn die Annahmen AP1 und AP2 erfüllt sind, gilt nach dem ersten Theorem erst recht:

$$\limsup_{n \rightarrow \infty} \mathbb{P}(V_n \geq 1) \leq \alpha$$

Theorem 2 - exakte asymptotische Kontrolle der FWER

Es seien Annahmen AP1 und AP2 erfüllt. Falls außerdem Annahme AP1 strikt gilt („=“ statt „≤“) und Q_0 stetig ist, gilt:

$$\limsup_{n \rightarrow \infty} \mathbb{P}(V_n \geq 1) = \alpha$$

┌ Beweis:

Das Verfahren lässt sich äquivalent umformulieren:

$$\begin{aligned}
 P_{0n}^*(1) &:= P_{0n}(1) \\
 P_{0n}^*(m) &:= \begin{cases} P_{0n}(m) & , \text{ falls } P_{0n}^*(m-1) < C_n(m-1) \\ 1 & , \text{ sonst} \end{cases}
 \end{aligned}$$

Entscheidungsregel: Wir lehnen $H_0(m)$ ab, falls $P_{0n}^*(m) < C_n(m)$.

Für $h_1 := |\mathcal{H}_1|$ definieren wir eine Bernoulli-Zufallsvariable:

$$\begin{aligned}
 B_n &:= \mathbb{1}_{\{\{1, \dots, h_1\} = \mathcal{H}_1, P_{0n}^*(1) < C_n(1), \dots, P_{0n}^*(h_1) < C_n(h_1)\}} \\
 \mathbb{P}(B_n = 1) &= \mathbb{P}(\{1, \dots, h_1\} = \mathcal{H}_1, P_{0n}^*(1) < C_n(1), \dots, P_{0n}^*(h_1) < C_n(h_1)) \\
 &\geq \mathbb{P} \left(\max_{m \in \mathcal{H}_1} P_{0n}^*(m) \leq \min_{m \in \mathcal{H}_1} C_n(m) \right) \xrightarrow{n \rightarrow \infty} 1, \text{ nach AP2}
 \end{aligned}$$

Dann:

$$\begin{aligned}
\mathbb{P}(V_n \geq 1) &= \mathbb{P}\left(\bigcup_{m \in \mathcal{H}_0} \{P_{0n}^*(m) < C_n(m)\}\right) \\
&= \mathbb{P}\left(\left(\bigcup_{m \in \mathcal{H}_0} \{P_{0n}^*(m) < C_n(m)\}\right) \cap \{B_n = 1\}\right) \\
&\quad + \mathbb{P}\left(\left(\bigcup_{m \in \mathcal{H}_0} \{P_{0n}^*(m) < C_n(m)\}\right) \cap \{B_n \neq 1\}\right) \\
&= \mathbb{P}\left(\bigcup_{m \in \mathcal{H}_0} \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n = 1\}\right) \cdot \mathbb{P}(B_n = 1) \\
&\quad + \mathbb{P}\left(\bigcup_{m \in \mathcal{H}_0} \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n \neq 1\}\right) \cdot \mathbb{P}(B_n \neq 1) \\
&= \mathbb{P}\left(\bigcup_{m \in \mathcal{H}_0} \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n = 1\}\right) \cdot (1 - \mathbb{P}(B_n \neq 1)) \\
&\quad + \mathbb{P}\left(\bigcup_{m \in \mathcal{H}_0} \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n \neq 1\}\right) \cdot \mathbb{P}(B_n \neq 1) \\
&= \mathbb{P}\left(\bigcup_{m \in \mathcal{H}_0} \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n = 1\}\right) + \mathbb{P}(B_n \neq 1) \cdot [\dots] \\
&= \mathbb{P}\left(\bigcup_{m \in \mathcal{H}_0} \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n = 1\}\right) + o(1) \\
&= \mathbb{P}\left(\bigcup_{m=h_1+1}^M \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n = 1\}\right) + o(1) \\
&\quad \left(\begin{aligned} \mathbb{P}(A \mid B) &= \mathbb{E}[\mathbf{1}_A \mid B] \\ &= \mathbb{E}[\mathbb{E}[\mathbf{1}_A \mid B] \mid B] \quad (\text{Turmeigenschaft}) \\ &= \mathbb{E}[\mathbb{P}(A \mid B) \mid B] \end{aligned} \right) \\
&= \mathbb{E}\left[\mathbb{P}\left(\bigcup_{m=h_1+1}^M \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n = 1\}\right) \mid \{B_n = 1\}\right] + o(1)
\end{aligned}$$

Wobei:

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_{m=h_1+1}^M \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n = 1\}\right) \\
&= \mathbb{P}\left(\underbrace{\bigcup_{m=h_1+1}^M \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n = 1\}, \{P_{0n}(h_1+1) < C_n(h_1+1)\}}_{= 1, \text{ nach Konstruktion des Verfahrens}}\right) \\
&\quad \cdot \mathbb{P}(P_{0n}(h_1+1) < C_n(h_1+1) \mid \{B_n = 1\}) \\
&+ \mathbb{P}\left(\underbrace{\bigcup_{m=h_1+1}^M \{P_{0n}^*(m) < C_n(m)\} \mid \{B_n = 1\}, \{P_{0n}(h_1+1) \not< C_n(h_1+1)\}}_{= 0, \text{ nach Konstruktion des Verfahrens}}\right) \\
&\quad \cdot (1 - \mathbb{P}(P_{0n}(h_1+1) < C_n(h_1+1) \mid \{B_n = 1\}))
\end{aligned}$$

Somit:

$$\begin{aligned}
\mathbb{P}(V_n \geq 1) &= \mathbb{E} [\mathbb{P}(P_{0n}(h_1 + 1) < C_n(h_1 + 1) \mid \{B_n = 1\}) \mid \{B_n = 1\}] + o(1) \\
&= \mathbb{E} \left[\mathbb{P} \left(\min_{m \in \mathcal{H}_0} P_{0n}(m) < C_n(h_1 + 1) \mid \{B_n = 1\} \right) \mid \{B_n = 1\} \right] + o(1) \\
&= \mathbb{P} \left(\min_{m \in \mathcal{H}_0} P_{0n}(m) < C_n(h_1 + 1) \mid \{B_n = 1\} \right) + o(1) \\
&= \mathbb{P} \left(\min_{m \in \mathcal{H}_0} P_{0n}(m) < C_n(h_1 + 1) \right) + o(1)
\end{aligned}$$

Asymptotisch ergibt sich:

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \mathbb{P}(V_n \geq 1) &= \limsup_{n \rightarrow \infty} \mathbb{P} \left(\min_{m \in \mathcal{H}_0} P_{0n}(m) < C_n(h_1 + 1) \right) \\
&\stackrel{\text{(AP1, strikt)}}{=} \mathbb{P} \left(\min_{m \in \mathcal{H}_0} P_0(m) < C_n(h_1 + 1) \right) \\
&= \mathbb{P} \left(\min_{m \in \mathcal{H}_0} P_0(m) < \inf \left\{ z \mid \mathbb{P}_{Q_0} \left(\min_{m \in \mathcal{H}_0} P_0(m) \leq z \right) \geq \alpha \right\} \right) \\
&\stackrel{\text{(} Q_0 \text{ stetig)}}{=} \alpha
\end{aligned}$$

⌊

3 Bemerkungen

Ein ganz ähnliches Verfahren ist das *Step-Down maxT* Verfahren.

Hier werden anstatt der unadjustierten p -Werte Maxima der Teststatistiken betrachtet. Theorem 1 und 2 gelten hier analog.

Wenn die Teststatistiken identisch verteilt sind, sind beide Verfahren äquivalent. Allgemein ist minP besser ausbalanciert (bewegt sich auf dem Einheitsintervall).

Zur Nullverteilung:

- Es lassen sich Nullverteilungen konstruieren, die unsere Annahme AP1 erfüllen (siehe Vortrag von Mathias Trabs).
- In der Praxis ist die Verteilung P der Daten unbekannt, also auch die Nullverteilung $Q_0(P)$. Ein konsistenter Schätzer für Q_0 lässt sich mittels Bootstrap-Verfahren konstruieren (siehe Mathias Trabs). Die Step-Down Verfahren funktionieren für diesen analog.

(Details siehe van der Laan et al., 2004)