Bayesianische FDR (Teil2)

Theorie des Multiplen Testens

Anna Ferrari

13.12.2010

Inhalt

Inhalt

Bedingte Wahrscheinlichkeit

$(\Omega, \mathfrak{F}, \mathbb{P})$ W-Raum, $A_1, A_2, \ldots A_n \in \mathfrak{F}$, $\sum A_i = \Omega$, $\mathbb{P}(A_i) > 0$.

Theorem (Bayes Theorem)

Die bedingte Wahrscheinlichkeit von A_i gegeben B ist definiert als

$$\underbrace{\mathbb{P}(A_i \mid B)}_{\text{a posteriori Wkt von } \mathrm{A_i}} = \frac{\mathbb{P}(B \mid A_i) \overbrace{\mathbb{P}(A_i)}^{\text{a priori Wkt von} \mathrm{A_i}}}{\sum_{k=1}^{n} \mathbb{P}(B \mid A_k)\mathbb{P}(A_k)}$$

Bayes-Theorem für Dichten

$\theta : (\Omega, \mathfrak{F}, \mathbb{P}) \longrightarrow (\mathcal{S}, \mathfrak{S})$, μ σ-endl. Maß auf dem Borelraum $(\mathcal{S}, \mathfrak{S})$,

Z:$(\Omega, \mathfrak{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathfrak{B})$ Zufallsvariable

$f_{(Z,\theta)} : \mathbb{R}\, x\, \mathcal{S} \longrightarrow \mathbb{R}_+$ $(\lambda \bigotimes \mu)$-Dichte von $\mathbb{P}^{(Z,\theta)}$.

Theorem (Bayes-Theorem für Dichten)

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

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

Beispiel

$\theta : (\Omega, \mathfrak{F}) \longrightarrow (\mathbb{N}, \mathfrak{P})$, μ Zählmaß auf $(\mathbb{N}),\mathfrak{P}(\mathbb{N})$
Z:$(\Omega, \mathfrak{F}) \longrightarrow (\mathbb{R}, \mathfrak{B})$ Es gelte:

$$\theta \sim Ber(p_1),\ p_0 := 1 - p_1$$
$$Z \sim (1-\theta)f_0 + \theta f_1$$

Dann mit Bayes-Theorem:

$$f_{\theta|Z=z}(\vartheta) = \frac{f_{Z|\theta=\vartheta}(z) f_\theta(\vartheta)}{f_Z(z)} = \begin{cases} \frac{p_0 f_0(z)}{p_0 f_0(z)+p_1 f_1(z)} & \text{für } \vartheta\text{=0} \\ \frac{p_1 f_1(z)}{p_0 f_0(z)+p_1 f_1(z)} & \text{für } \vartheta\text{=1} \end{cases}$$

Inhalt

False Discovery Rate

Multiples Testproblem, n zu testende Hypothesen,
$\varphi = (\varphi_i, i = 1, \ldots, n)$ multiple Testprozedur.
V : (zufällige) Anzahl der Typ-I Fehler
D : (zufällige) Anzahl der verworfenen Hypothesen insgesamt:

Definition

Der erwartete Anteil der Typ-I Fehler von φ ist

$$FDR(\varphi) = \mathbb{E}[\frac{V}{D \vee 1}]$$

und heißt False Discovery Rate (FDR) von φ;
Die positive False Discovery Rate (pFDR) von φ ist

$$pFDR(\varphi) = \mathbb{E}[\frac{V}{D} \mid D > 0]$$

False Negative Rate

S: (zufällige) Anzahl der Typ-II Fehler
n-D: (zufällige) Anzahl der nichtverworfenen Hypothesen

Definition (False Negative Rate)

Der erwartete Anteil der Typ-II Fehler von φ ist

$$FNR(\varphi) = \mathbb{E}[\frac{S}{n-D}]$$

und heißt False Negative Rate von φ.

Inhalt

Zweiklassen-Mischmodell

Gegeben seien $\varphi = (\varphi_i, i = 1, \ldots, n)$ multiple Testprozedur, ein Zweiklassen-Mischmodell $(Z_1, H_1), \ldots, (Z_n, H_n)$ iid. Für $i = 1, \ldots, n$: Zi Teststatistik, Hi Indikatorvariable und

$H_i \sim Be(p_1)$ *mit* $p_1 \equiv 1 - p_o$

$Z_i \mid H_i \sim (1 - H_i)f_0 + H_i f_1$

Theorem (Theorem (Storey))

Sei φ_i basierend auf Teststatistik Zi mit Verwerfungsbereich C.

$$pFDR(C) \equiv pFDR(\varphi) = \mathbb{P}[H_1 = 0 \mid Z_1 \in C]$$

Lokale fdr

Analog zu $pFDR(C) = \mathbb{P}[H = 0 \mid Z \in C]$

Definition

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

Bayes-Formel für Dichten:

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

Empirische vs Nichtparametrische Ansatz mit Zweiklassen-Mischmodell

- Empirischer Bayes-Ansatz: $P_{Emp}(z_i) = \frac{p_0 f_0(z_i)}{\widehat{f_Z}(z_i)} = 1 - \widehat{fdr}(z_i)$ $=\mathbb{P}(H_i = 1 \mid Z = z_i) \rightsquigarrow$ ziehe Schlüsse aus dem Wert
- Nichtparametrischer Bayes-Ansatz: $\mathbb{P}[H = 1 \mid Z, f_0, f_1, p_0] = \frac{p_1 f_1(Z)}{p_0 f_0(Z) + p_1 f_1(Z)} = \frac{p_1 f_1(Z)}{f(Z)}$ Festlegen eines a priori Wahrscheinlichkeitsmodells für f_0, f_1, p_0. Bestimmung der a-posteriori-Verteilungen $\mathbb{P}^{f|Y}, \mathbb{P}^{f|Y}, \mathbb{P}^{p0|Y}$ gegeben alle Daten Y. $P(z_i) \equiv \mathbb{E}[\frac{p_1 f_1}{f(Z)} \mid Y] = \frac{p_1 \widehat{f_1}(z_i)}{\widehat{f_1}(z_i)} = 1 - \mathbb{E}[fdr(z_i) \mid Y]$ $\rightsquigarrow$ ziehe Schlüsse aus dem Wert

Empirische vs Nichtparametrische Ansatz mit Zweiklassen-Mischmodell

- Empirischer Bayes-Ansatz: $P_{Emp}(z_i) = \frac{p_0 f_0(z_i)}{\widehat{f_Z}(z_i)} = 1 - \widehat{fdr}(z_i)$ $=\mathbb{P}(H_i = 1 \mid Z = z_i) \rightsquigarrow$ ziehe Schlüsse aus dem Wert
- Nichtparametrischer Bayes-Ansatz:
  $\mathbb{P}[H = 1 \mid Z, f_0, f_1, p_0] = \frac{p_1 f_1(Z)}{p_0 f_0(Z) + p_1 f_1(Z)} = \frac{p_1 f_1(Z)}{f(Z)}$
  Festlegen eines a priori Wahrscheinlichkeitsmodells für f_0, f_1, p_0.
  Bestimmung der a-posteriori-Verteilungen $\mathbb{P}^{f|Y}, \mathbb{P}^{f|Y}, \mathbb{P}^{p0|Y}$ gegeben alle Daten Y.
  $P(z_i) \equiv \mathbb{E}[\frac{p_1 f_1}{f(Z)} \mid Y] = \frac{p_1 \widehat{f_1}(z_i)}{\widehat{f_1}(z_i)} = 1 - \mathbb{E}[fdr(z_i) \mid Y]$
  $\rightsquigarrow$ ziehe Schlüsse aus dem Wert

Kritik: Entscheidungsregel des Typs $\mathbb{I}\{z_i \leqslant z\}$ sind intuitiv aber nicht immer optimal.

Inhalt

1. Wiederholung Schwerpunkte des ersten Teils
   - Bayes-Theorem
   - FDR, FNR, pFDR, fdr
   - Empirische vs. Nichtparametrische Bayes Ansatz
2. Teil 2: Entscheidungsregel und Verlustfunktion
   - Theorie
   - FDR, FNR und Bayesregel
   - Abschluss Bayesianische FDR

Entscheidungsregel

$(\mathfrak{X}, \mathfrak{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ *Statistisches Experiment*

Definition

Eine Entscheidungsregel ρ ist eine messbare Abbildung

$$\rho : \mathfrak{X} \longrightarrow A$$

wobei der Messraum $(A, \mathfrak{A})$ der sog. Aktionsraum ist.

Verlustfunktion und Risiko

Dazu ist es hilfreich die Verlustfunktion und das Risiko zu definieren.

Definition (Verlustfunktion)

Jede Funktion $\ell : (\Theta, A) \longrightarrow [0, \infty) =: \mathbb{R}_+$,
die messbar im zweiten Argument ist, heißt Verlustfunktion.

Definition (Risiko)

Das Risiko einer Entscheidungsregel ρ bei Vorliegen des Parameters $\vartheta \in \Theta$ ist
$\mathcal{R}(\vartheta, \rho) := \mathbb{E}_\vartheta[\ell(\vartheta, \rho)] = \int_\mathfrak{X} \ell\,(\theta, \rho(x))\mathbb{P}_\vartheta(dx)$

ρ besser, ρ zulässig

Definition

Die Entscheidungsregel ρ heißt *besser* als eine Entscheidungsregel $\check{\rho}$, falls

$$\mathcal{R}(\vartheta, \rho) \leqslant \mathcal{R}(\vartheta, \check{\rho}) \, \textit{für alle} \; \vartheta \in \Theta$$

gilt und falls es ein

$$\vartheta_0 \in \Theta \; \textit{mit} \; \mathcal{R}(\vartheta_0, \rho) < \mathcal{R}(\vartheta_0, \check{\rho})$$

existiert.
Eine Entscheigungsragel heißt *zulässig*, wenn es keine bessere Entscheidungsregel gibt.

Bayesrisiko

Der Parameterraum Θ trage die σ-Algebra $\mathfrak{F}_\Theta$,die Verlustfunktion ℓ sei produktmessbar und $\vartheta \mapsto \mathbb{P}_\vartheta(B)$ sei messbar für alle B $\in \mathfrak{F}$. Die a-priori-Verteilung π des Parameters ϑ ist gegeben durch ein Wahrscheinlichkeitsmaß auf $(\Theta, \mathfrak{F}_\Theta)$. Das zu π-assoziierte Bayesrisiko einer Entscheidungsregel ρ ist

$$\mathcal{R}_\pi := \mathbb{E}_\pi[\mathcal{R}(\vartheta, \rho)]$$

$$= \int_\Theta \mathcal{R}(\vartheta, \rho)\, \pi(d\vartheta)$$

$$= \int_\Theta \int_{\mathfrak{X}} \ell(\vartheta, \rho(x)) \mathbb{P}_\vartheta(dx)\pi(d\vartheta)$$

Bayesrisiko

Man kann das Bayesrisiko als erwartete Verlust in folgendem Sinne verstehen:
Sei $\Omega := \mathfrak{X} \times \Theta$ und $\tilde{\mathbb{P}}$ auf $(\Omega, \mathfrak{F} \otimes \mathfrak{F}_{\Theta})$,
$\tilde{\mathbb{P}}(dx, d\vartheta) = \mathbb{P}_{\vartheta}(dx)\,\pi(d\vartheta)$ (gemeinsame Verteilung von Beobachtung und Parameter). Bezeichne mit X und T die Koordinatenprojektionen von Ω auf $\mathfrak{X}$ *bzw*. Θ.
Dann gilt

$$R_{\pi}(\rho) = \mathbb{E}_{\tilde{\mathbb{P}}}[\ell(T, \rho(X))]$$

Bayes-optimal Regel

Die Verteilung von T unter der regulären bedingten Wahrscheinlichkeit $\tilde{\mathbb{P}}(\bullet \mid X = x)$ von $\tilde{\mathbb{P}}$ heißt a-posteriori-Verteilung des Parameters gegeben die Beobachtung $X = x$

Theorem

Eine Regel ρ ist Bayes-optimal, falls

$$\rho(x) = arg\min_{a \in A} \mathbb{E}_{\tilde{\mathbb{P}}}[\ell(T,a) \mid X = x] \;\; \tilde{\mathbb{P}}\text{-}f.s.$$

d.h. $\mathbb{E}_{\tilde{\mathbb{P}}}[\ell(T,\rho(X)) \mid X = x] \leqslant \mathbb{E}_{\tilde{\mathbb{P}}}[\ell(T,a) \mid X = x]$ *für alle* $a \in A$ *und* $\tilde{\mathbb{P}}^X$*- fast alle* $x \in \mathfrak{X}$.

Inhalt

Setting

- Modell: ω Parameter, y die Daten, H_i Regulationsindikator des Gens i
- Indikatorvariabel

$$H_i = \begin{cases} 1 & \text{falls Gen i mit Veränderung} \\ 0 & \text{falls Gen i ohne Veränderung} \end{cases}$$

- Indikatorvariabel

$$d_i = \begin{cases} 1 & \text{falls Discovery} \\ 0 & \text{falls Negative} \end{cases}$$

mit $D = \sum_{i=1}^{n} d_i$ = Anzahl der verworfenen Hyposthesen;

- und sei $v_i = \mathbb{P}(H_i = 1 \mid Y)$ die i-te a-posteriori-Randverteilung (d.h Wkt Gen i mit Veränderung gegeben die Daten $:= Y$).

Setting

- Modell: ω Parameter, y die Daten, H_i Regulationsindikator des Gens i
- Indikatorvariabel

$$H_i = \begin{cases} 1 & \text{falls Gen i mit Veränderung} \\ 0 & \text{falls Gen i ohne Veränderung} \end{cases}$$

- Indikatorvariabel

$$d_i = \begin{cases} 1 & \text{falls Discovery} \\ 0 & \text{falls Negative} \end{cases}$$

mit $D = \sum_{i=1}^{n} d_i$ = Anzahl der verworfenen Hyposthesen;

- und sei $v_i = \mathbb{P}(H_i = 1 \mid Y)$ die i-te a-posteriori-Randverteilung (d.h Wkt Gen i mit Veränderung gegeben die Daten := Y).

Setting

- Modell: ω Parameter, y die Daten, H_i Regulationsindikator des Gens i
- Indikatorvariabel

$$H_i = \begin{cases} 1 & \text{falls Gen i mit Veränderung} \\ 0 & \text{falls Gen i ohne Veränderung} \end{cases}$$

- Indikatorvariabel

$$d_i = \begin{cases} 1 & \text{falls Discovery} \\ 0 & \text{falls Negative} \end{cases}$$

mit $D = \sum_{i=1}^{n} d_i$ = Anzahl der verworfenen Hyposthesen;

- und sei $v_i = \mathbb{P}(H_i = 1 \mid Y)$ die i-te a-posteriori-Randverteilung (d.h Wkt Gen i mit Veränderung gegeben die Daten $:= Y$).

Setting

- Modell: ω Parameter, y die Daten, H_i Regulationsindikator des Gens i
- Indikatorvariabel

$$H_i = \begin{cases} 1 & \text{falls Gen i mit Veränderung} \\ 0 & \text{falls Gen i ohne Veränderung} \end{cases}$$

- Indikatorvariabel

$$d_i = \begin{cases} 1 & \text{falls Discovery} \\ 0 & \text{falls Negative} \end{cases}$$

mit $D = \sum_{i=1}^{n} d_i$ = Anzahl der verworfenen Hyposthesen;

- und sei $v_i = \mathbb{P}(H_i = 1 \mid Y)$ die i-te a-posteriori-Randverteilung (d.h Wkt Gen i mit Veränderung gegeben die Daten := Y).

FDR

Konstruiere:

$$FDR(d,H) = \frac{\sum_{i=1}^{n} d_i(1-H_i)}{D}$$

Die Summe summiert nur falls $d_i = 1$ (Discovery) und $H_i = 0$ (Gen i ohne veränderung) $\rightsquigarrow$ wieder Typ-I Fehler

FNR

Konstruiere:

$$FNR(d, H) = \frac{\sum_{i=1}^{n}(1 - d_i)H_i}{n - D}$$

Die Summe summiert nur falls $d_i = 0$ (Negative) und $H_i = 1$ (Gen i mit Veränderung) $\rightsquigarrow$ wieder Typ-II Fehler

A-Posteriori erwartete FDR

Aus FDR bekommt man die a-posteriori erwartete FDR:

$$\overline{FDR}(d,y) = \mathbb{E}[FDR \mid Y = y]$$

$$= \int FDR(d,H) d\tilde{\mathbb{P}}(H_i \mid Y = y)$$

$$= \frac{\sum_{i=1}^n d_i(1 - v_i)}{D}$$

A-Posteriori erwartete FNR

Aus FNR bekommt man die a-posteriori erwartete FNR:

$$\overline{FNR}(d, y) = \mathbb{E}[FNR \mid Y = y]$$

$$= \int FNR(d, H) d\tilde{\mathbb{P}}(H \mid Y = y)$$

$$= \frac{\sum_{i=1}^{n}(1 - d_i)v_i}{n - D}$$

Idee

Idee: Reduziere die Fehler unserer Testentscheidungen, d.h *minimiere* den negativen und positiven a-posteriori erwarteten Zähler:

Erinnerung: $v_i = \mathbb{P}(H_i = 1 \mid Y)$

$$\overline{FN} = \sum_{k=1}^{n} (1 - d_i) v_i$$

$$\overline{FD} = \sum_{i=1}^{n} d_i (1 - v_i)$$

Gegebenen Verlustfunktionen

Vorschläge für die Verlustfunktionen:

$$\ell_N(d, Y) = c\overline{FD} + \overline{FN} \tag{1}$$

hängt von der gesamten Anzahl von abgelehnten Hypothesen und von den Negatives. c ist eine Konstante , die die Anzahl der fälschlicherweise abgelehnten Hypothesen bestrafen soll.

$$\ell_R(d, Y) = c\overline{FDR} + \overline{FNR} \tag{2}$$

Gegebenen Verlustfunktionen

Weitere Vorschläge, Bivariate Verlustfunktionen:

$$\boldsymbol{\ell_{2N}(d,Y) = (\overline{FN}, \overline{FD})} \tag{3}$$

$$\boldsymbol{\ell_{2R}(d,Y) = (\overline{FDR}, \overline{FNR})} \tag{4}$$

deren Minimeirung nach $\overline{FN}$ (bzw. $\overline{FNR}$) gemacht wird, während $\overline{FD}$ $(bzw.\overline{FDR})$ von einem α beschränkt wird ($\overline{FD} \leqslant \alpha$).

Entscheidungsregel

Theorem

Unter (1)-(4) die optimale Enscheidung ist gegeben durch

$$d_i = \mathbb{I}\{v_i \geqslant t^\star\}$$

Die optimalen Cut-offs sind

$$\begin{aligned} t_N^\star &= \frac{c}{c+1} \\ t_R^\star &= v_{(n-D^\star)} \\ t_{2N}^\star &= \min\left\{s : \overline{FD}(s,y) \leqslant \alpha_N\right\} \\ t_{2R}^\star &= \min\left\{s : \overline{FDR}(s,y) \leqslant \alpha_R\right\} \end{aligned}$$

mit $D^\star$-optimale Anzahl der Discoveries, und $v_{(i)}$ der i-te Statistikorder.

Beweis:

Kritik

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

Differential Expression

Einführung eines zusäzlichen Parameters γ_i ins Modell, der als differentielles Expressionsniveaus des Gens i interpretiert wird, so dass

$$\gamma_i = 0,\ falls\ H_i = 0\ und\ \gamma_i > 0,\ falls\ H_i = 1$$

Dann ist

$$\mathcal{L}_\gamma(\gamma, d, z) = -\sum_{i=1}^{n} d_i\gamma_i + \kappa \sum_{i=1}^{n}(1-d_i)\gamma_i + cD$$

die "γ -modifizierte" Verlustfunktion.

Differential Expression und Entscheidungsregel

Sei $\overline{\gamma}_i = \mathbb{E}[\gamma_i \mid Y]$ das a-posteriori erwartete Expressionsniveau von Gen i ggb. die Daten. Analog mit den ersten Velustfunktionen, trifft man die optimale Entscheidung:

$$d_i^\star = \mathbb{I}\{\overline{\gamma}_i \geqslant \underbrace{\frac{c}{(1+\kappa)}}_{\text{feste Cutoff}}\}$$

Bemerkung

- Die Entscheidung bleibt im Wesentlichen die selbe wenn man den Parameter des Expressionsniveus γ in eine Funktion $f(\gamma)$ transformiert.
- Wir haben angenommen, dass der Untersucher, d.h. derjeniger, der dei Entscheidung trifft, *rationell* ist.

Bemerkung

- Die Entscheidung bleibt im Wesentlichen die selbe wenn man den Parameter des Expressionsniveus γ in eine Funktion $f(\gamma)$ transformiert.
- Wir haben angenommen, dass der Untersucher, d.h. derjeniger, der dei Entscheidung trifft, *rationell* ist.

Inhalt

Zusammenfassung

1. Begriffe der FDR FNR pFDR
2. Zweiklassen-Mischmodell
3. Anwendung der empirischen und nichtparametrischen Bayes-Ansätze
4. Verlustfunktion und Risiko
5. Optimale Bayesregel mit ggb. Verlustfunktion

- mit gleichmäßigen unerwünschten FDR und FNR
- mit differentiellen Expressionsniveaus der FDR und FNR.

Zusammenfassung

1. Begriffe der FDR FNR pFDR
2. Zweiklassen-Mischmodell
3. Anwendung der empirischen und nichtparametrischen Bayes-Ansätze
4. Verlustfunktion und Risiko
5. Optimale Bayesregel mit ggb. Verlustfunktion

- mit gleichmäßigen unerwünschten FDR und FNR
- mit differentiellen Expressionsniveaus der FDR und FNR.

Zusammenfassung

1. Begriffe der FDR FNR pFDR
2. Zweiklassen-Mischmodell
3. Anwendung der empirischen und nichtparametrischen Bayes-Ansätze
4. Verlustfunktion und Risiko
5. Optimale Bayesregel mit ggb. Verlustfunktion

- mit gleichmäßigen unerwünschten FDR und FNR
- mit differentiellen Expressionsniveaus der FDR und FNR.

Zusammenfassung

1. Begriffe der FDR FNR pFDR
2. Zweiklassen-Mischmodell
3. Anwendung der empirischen und nichtparametrischen Bayes-Ansätze
4. Verlustfunktion und Risiko
5. Optimale Bayesregel mit ggb. Verlustfunktion
   - mit gleichmäßigen unerwünschten FDR und FNR
   - mit differentiellen Expressionsniveaus der FDR und FNR.

Zusammenfassung

1. Begriffe der FDR FNR pFDR
2. Zweiklassen-Mischmodell
3. Anwendung der empirischen und nichtparametrischen Bayes-Ansätze
4. Verlustfunktion und Risiko
5. Optimale Bayesregel mit ggb. Verlustfunktion

- mit gleichmäßigen unerwünschten FDR und FNR
- mit differentiellen Expressionsniveaus der FDR und FNR.

Zusammenfassung

1. Begriffe der FDR FNR pFDR
2. Zweiklassen-Mischmodell
3. Anwendung der empirischen und nichtparametrischen Bayes-Ansätze
4. Verlustfunktion und Risiko
5. Optimale Bayesregel mit ggb. Verlustfunktion

- mit gleichmäßigen unerwünschten FDR und FNR
- mit differentiellen Expressionsniveaus der FDR und FNR.

Zusammenfassung

1. Begriffe der FDR FNR pFDR
2. Zweiklassen-Mischmodell
3. Anwendung der empirischen und nichtparametrischen Bayes-Ansätze
4. Verlustfunktion und Risiko
5. Optimale Bayesregel mit ggb. Verlustfunktion

- mit gleichmäßigen unerwünschten FDR und FNR
- mit differentiellen Expressionsniveaus der FDR und FNR.

$$\mathfrak{D}\forall\mathcal{N}\Bbbk\exists\int\mathsf{C}H\ddot{o}\aleph$$