

Politechnika Wrocławska  
Instytut Matematyki i Informatyki

**Testowanie zbioru hipotez z zadaną relacją hierarchii  
wraz z przykładami zastosowań w genetyce**

Przemysław Biecek

Rozprawa doktorska przygotowana pod kierunkiem  
prof. dra hab. Witolda Kloneckiego

Wrocław 2007

*dla Niny*

# Spis treści

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Wstęp</b>  | <b>3</b>  |
| <b>2</b> | <b>Testowanie hipotezy statystycznej w ujęciu klasycznym</b>                    | <b>5</b>  |
| 2.1      | Eksperyment losowy i hipoteza statystyczna . . . . .                            | 5         |
| 2.2      | Testowanie hipotezy statystycznej . . . . .                                     | 6         |
| 2.3      | Błąd I rodzaju oraz błąd II rodzaju . . . . .                                   | 7         |
| 2.4      | Wartość $p$ . . . . .   | 8         |
| <b>3</b> | <b>Testowanie zbioru hipotez</b>  | <b>9</b>  |
| 3.1      | Zbiór eksperymentów losowych i zbiór hipotez statystycznych . . . . .           | 9         |
| 3.2      | Zmienne losowe $V$ , $R$ , $S$ , $Q$ i współczynniki błędów . . . . .           | 10        |
| 3.3      | Procedury testowania zbioru hipotez . . . . .                                   | 12        |
| 3.4      | Procedury kontroli współczynnika PFER . . . . .                                 | 16        |
| 3.5      | Procedury kontroli współczynnika FWER . . . . .                                 | 16        |
| 3.6      | Procedury kontroli współczynnika FDR . . . . .                                  | 17        |
| <b>4</b> | <b>Testowanie zbioru hipotez z relacją hierarchii</b>                           | <b>18</b> |
| 4.1      | Zbiór hipotez i relacja hierarchii . . . . .                                    | 18        |
| 4.2      | Procedury testowania zbioru hipotez . . . . .                                   | 21        |
| 4.3      | Przykład testowania zbioru dwóch hipotez zerowych będących w relacji hierarchii | 23        |
| 4.4      | Procedury kontroli współczynnika PFER . . . . .                                 | 26        |
| 4.5      | Procedury kontroli współczynnika FWER . . . . .                                 | 29        |
| 4.6      | Procedury kontroli współczynnika FDR . . . . .                                  | 30        |
| <b>5</b> | <b>Zastosowanie uzyskanych wyników w badaniach genetycznych</b>                 | <b>33</b> |
| 5.1      | Wprowadzenie . . . . .  | 33        |
| 5.2      | Zagadnienie identyfikacji aktywnych procesów biologicznych . . . . .            | 36        |
| 5.3      | Predykcja funkcji białek . . . . .  | 39        |
| <b>A</b> | <b>Program wyznaczający poziomy istotności <math>\alpha^{(i)}</math></b>        | <b>44</b> |
| <b>B</b> | <b>Badania prowadzone podczas studiów doktoranckich</b>                         | <b>46</b> |
| <b>C</b> | <b>Tło historyczne zagadnienia testowania zbioru hipotez</b>                    | <b>48</b> |

# Rozdział 1

## Wstęp

W poniższej rozprawie doktorskiej przedstawione jest zagadnienie testowania zbioru hipotez z zadaną relacją hierarchii. Przedstawione są nowe procedury testowania, dla których wykazano kontrolę współczynników FWER, PFER i FDR. Rozprawę kończy przykład zastosowania opracowanych procedur w analizie danych genetycznych oraz studium symulacyjne porównujące opracowane procedury testowania. Wyniki te są częścią rezultatów uzyskanych podczas studiów doktoranckich w Instytucie Matematyki i Informatyki Politechniki Wrocławskiej. Kompletna lista prac przeprowadzonych podczas studiów doktoranckich znajduje się w Dodatku B.

Badania nad zagadnieniem testowania zbioru hipotez z zadaną relacją hierarchii rozpocząłem po wysłuchaniu referatu dra Adama Zagdańskiego poświęconego predykcji funkcji genów z uwzględnieniem struktury Gene Ontology [Harris i inni 2004]. Istniejące procedury testowania uwzględniające hierarchiczną relację pomiędzy hipotezami (np. wyniki [Marcus i inni 1976], [Finner, Strassburger 2002]) dotyczą tylko kontroli współczynnika FWER. Co więcej, nie mogą być wykorzystane w zagadnieniu predykcji funkcji genów z uwagi na trudność w określeniu przestrzeni parametrów  $\Theta$  tak, by była zachowana żądana hierarchia.

Ze względu na te trudności, zaproponowałem inne sformułowanie zagadnienia testowania zbioru hipotez z zadaną relacją hierarchii oraz przedstawiłem procedury testowania pozwalające na kontrolę najpopularniejszych współczynników błędów. Wyniki te zostały opisane w pracy [Biecek 2007] oraz są przedstawione w poniższej rozprawie. Prezentowane ujęcie zagadnienia testowania zbiorczego z zadaną relacją hierarchii może być zastosowane do szerszej klasy problemów niż ujęcie prezentowane w pracy [Finner, Strassburger 2002]. W rozdziale 5 przedstawione są przykłady dwóch zagadnień w których można wykorzystać wyniki teoretyczne przedstawione w tej pracy, a do których nie można zastosować procedur opisanych w pracach [Finner, Strassburger 2002] i [Marcus i inni 1976].

W całej pracy rozważany jest proces testowania hipotez w ujęciu częstościowym. W Rozdziale 2 zaprezentowane są podstawowe pojęcia Neymanowsko-Pearsonowskiej teorii testowania hipotez. W Rozdziale 3 wprowadzone jest zagadnienie testowania zbioru hipotez oraz przedstawione są najpopularniejsze procedury testowania, które będą porównywane z nowo proponowanymi. W Rozdziale 4 przedstawione są główne wyniki tej pracy dotyczące modyfikacji zagadnienia testowania zbioru hipotez uwzględniającej relację hierarchii pomiędzy rozważanymi hipotezami. Wprowadzone są dwie procedury testowania „wstępująca” i „zstępująca” oraz przedstawione są metody wyznaczania poziomów istotności dla tych procedur. Prezentowane twierdzenia dowodzą, iż te procedury kontrolują współczynniki PFER, FWER i FDR na zadanym poziomie  $\alpha$ .

W Rozdziale 5 przedstawione jest studium symulacyjne oraz podsumowanie zastosowania prezentowanych procedur w analizie danych rzeczywistych. Wskazane są też sytuacje, w których proponowane procedury prowadzą do znaczącej poprawy wyników testowania. W Dodatku A umieszczone są kody programów w języku R, wyznaczające poziomy istotności dla zaproponowanych procedur. W Dodatku B przedstawiona jest lista zagadnień nad którymi pracowałem w trakcie studiów doktoranckich. W Dodatku C znajduje się krótki opis tła historycznego dla zagadnienia testowania hipotez, testowania zbioru hipotez oraz testowania zbioru hipotez z relacją hierarchii.

Uzyskane wyniki stanowią wkład w prace nad zagadnieniem testowania zbioru hipotez. Kontynuacją tych badań będzie rozważanie innych, niż hierarchiczne, relacji na zbiorze hipotez (np. relacji wykluczenia) oraz uwzględnienie w procedurach testowania struktury korelacji pomiędzy statystykami testowymi.

Chciałbym podziękować promotorowi, profesorowi Witoldowi Kloneckiemu, za niezapomnianą i inspirującą współpracę oraz za pomoc w redagowaniu tej i innych prac. Chciałbym podziękować również profesor Teresie Ledwinie oraz obu recenzentom profesorowi Jarosławowi Bartoszewiczowi oraz profesorowi Tadeuszowi Calińskiemu za wiele cennych uwag, które pozwoliły ulepszyć tę pracę. Najbardziej jednak chciałbym podziękować mojej żonie Karolinie Bieck, za anielską cierpliwość, wsparcie i zrozumienie, oraz moim rodzicom, za wszystko czego mnie nauczyli.

## Rozdział 2

# Testowanie hipotezy statystycznej w ujęciu klasycznym

*Hipoteza (gr. hypóthesis 'podstawa, przypuszczenie' od hypotithénai 'podstawiać, przypuszczać')  
„Słownik Wyrazów Obcych” Wydawnictwa Europa*

### 2.1 Eksperyment losowy i hipoteza statystyczna

Zagadnienie testowania hipotezy statystycznej w ujęciu częstościowym jest nierozłącznie związane z pojęciem eksperymentu losowego.

**Definicja 1** *Eksperymentem losowym (lub mechanizmem losowym)  $\mathcal{E}$  nazywamy eksperyment, który ma następujące właściwości:*

- *wiemy jakie wyniki możemy zaobserwować (zbiór wyników oznaczamy przez  $\mathcal{X}$ ),*
- *nie wiemy, jaki wynik zaobserwujemy,*
- *eksperyment możemy wielokrotnie powtarzać w identycznych warunkach,*
- *wraz ze wzrostem liczby powtórzeń eksperymentu stabilizują się odpowiednie częstości.*

Eksperymentowi  $\mathcal{E}$  odpowiada pewien nieznan rozkład prawdopodobieństwa  $P(x)$  na  $\mathcal{X}$ , zgodnie z którym generowane są wyniki eksperymentu  $\mathcal{E}$ . Zakładamy, że należy on do przyjętej rodziny rozkładów  $\mathcal{P} = \{P_\theta(x) : \theta \in \Theta\}$  czyli, że  $P(x) = P_{\theta^*}(x)$  dla pewnego  $\theta^* \in \Theta$ . Badaczka interesuje weryfikacja przypuszczenia czy nieznan parametr  $\theta^*$  należy do zbioru  $\Theta_0 \subset \Theta$ .

**Definicja 2** *Hipotezę zerową  $H_0$  nazywamy przypuszczenie, że  $\theta^* \in \Theta_0$ .*

**Definicja 3** *Hipotezę alternatywną  $H_A$  nazywamy przypuszczenie, że  $\theta^* \notin \Theta_0$ .*

Wprowadzamy indeks  $H$ , określający faktyczny stan hipotezy zerowej (wartość tego indeksu jest nieznaną badaczowi),

$$H = \begin{cases} 0 & \text{gdy hipoteza zerowa jest prawdziwa, } \theta^* \in \Theta_0, \\ 1 & \text{gdy hipoteza zerowa jest fałszywa, } \theta^* \notin \Theta_0. \end{cases}$$

Równoważnie do powyższych definicji, hipotezą zerową  $H_0$  nazywamy przypuszczenie, że obserwowany wynik eksperymentu  $\mathcal{E}$  był wygenerowany zgodnie z rozkładem należącym do rodziny  $\{P_\theta(x) : \theta \in \Theta_0\}$ . Hipotezą alternatywną  $H_A$  nazywamy przypuszczenie, że obserwowany wynik został wygenerowany zgodnie z rozkładem z rodziny  $\{P_\theta(x) : \theta \in \Theta \setminus \Theta_0\}$ .

Jeżeli zbiór  $\Theta_0$  składa się z dokładnie jednego punktu, tj.  $\Theta_0 = \{\theta_0\}$ , to hipotezę zerową nazywamy hipotezą prostą, w przeciwnym przypadku będziemy ją nazywali hipotezą złożoną.

Bez straty ogólności dla prezentowanych twierdzeń w tej pracy rozważamy jedynie proste hipotezy zerowe oraz prawostronne hipotezy alternatywne. Tak więc, o ile nie zostanie napisane inaczej,  $\Theta = [\theta_0, \infty)$ , hipoteza zerowa to przypuszczenie, że  $\theta^* = \theta_0$ , a hipoteza alternatywna to przypuszczenie, że  $\theta^* > \theta_0$ .

## 2.2 Testowanie hipotezy statystycznej

Na podstawie obserwacji  $x \in \mathcal{X}$ , otrzymanej w wyniku przeprowadzenia eksperymentu  $\mathcal{E}$ , mamy podjąć decyzję czy hipotezę zerową przyjąć czy odrzucić.

**Definicja 4** *Testem statystycznym  $\psi : \mathcal{X} \rightarrow \{0, 1\}$  nazywamy funkcję, która określa, dla jakich wyników eksperymentu należy podjąć decyzję o przyjęciu  $H_0$  ( $\psi(x) = 0$ ), a dla jakich o odrzuceniu  $H_0$  i przyjęciu  $H_A$  ( $\psi(x) = 1$ ).*

**Uwaga 1** *W przyjętej nomenklaturze dla  $\psi(x) = 0$  stosuje się określenie, że „posiadane przesłanki nie dają podstaw do odrzucenia hipotezy zerowej na ustalonym poziomie istotności”. Dla uproszczenia zapisu w tej pracy będziemy pisać, że hipotezę zerową przyjmujemy, gdy  $\psi(x) = 0$  albo odrzucamy, gdy  $\psi(x) = 1$ .*

Przyjęcie hipotezy zerowej nie oznacza, że jesteśmy przekonani o jej prawdziwości. Oznacza jedynie, że posiadane przesłanki (czyli wynik eksperymentu  $\mathcal{E}$ ) nie dają podstaw do odrzucenia hipotezy zerowej i przyjęcia hipotezy alternatywnej na ustalonym poziomie istotności [Neyman, Pearson 1933]. Podobnie odrzucenie hipotezy zerowej nie oznacza, że jesteśmy przekonani o fałszywości hipotezy zerowej.

**Uwaga 2** *Funkcja  $\psi$  jest zmienną losową, podczas gdy  $\psi(x)$  oznacza jej realizację. W dalszej części pracy dla uproszczenia zapisu symbolem  $\psi$  będziemy oznaczać zarówno zmienną losową jak i jej realizację.*

Test statystyczny dzieli przestrzeń  $\mathcal{X}$  na dwa rozłączne podzbiory: obszar odrzucenia hipotezy zerowej (nazywany też obszarem krytycznym)  $B = \{x : \psi(x) = 1\} \subset \mathcal{X}$  oraz jego dopełnienie  $B^C$ , czyli obszar przyjęcia hipotezy zerowej.

Obszary odrzucenia i przyjęcia wygodnie jest opisywać w terminach statystyki testowej  $T : \mathcal{X} \rightarrow \mathcal{R}$  następująco:

$$B = \{x \in \mathcal{X} : T(x) \geq c\},$$

$$B^C = \{x \in \mathcal{X} : T(x) < c\}.$$

Liczbę  $c \in \mathcal{R}$  nazywamy wartością krytyczną testu. Statystykę testową wybiera się tak by dla hipotez alternatywnych jej rozkład był stochastycznie większy (patrz Definicja 5) od jej rozkładu dla hipotezy zerowej oraz by uzyskać możliwie dużą moc, czyli aby  $P_{\theta^*}(B)$  było możliwie duże dla  $\theta^* \in \Theta \setminus \Theta_0$ .

**Definicja 5** Mówimy, że rozkład zmiennej losowej  $Y_1$ , określonej na przestrzeni  $X$ , jest stochastycznie większy od rozkładu zmiennej losowej  $Y_2$ , określonej na przestrzeni  $X$ , jeżeli dla każdego  $c \in X$

$$Pr(Y_1 > c) \geq Pr(Y_2 > c)$$

oraz dla pewnego  $c' \in X$

$$Pr(Y_1 > c') > Pr(Y_2 > c').$$

## 2.3 Błąd I rodzaju oraz błąd II rodzaju

Testując hipotezę zerową możemy popełnić jeden z dwóch rodzajów błędów. Jeżeli odrzucimy hipotezę zerową, gdy jest ona prawdziwa ( $H = 0$ ), to mówimy, że popełniamy błąd I rodzaju. Jeżeli przyjmujemy hipotezę zerową, gdy jest ona fałszywa ( $H = 1$ ), to popełniamy błąd II rodzaju. W pozostałych przypadkach podejmujemy poprawną decyzję (patrz Tabela 2.1).

| Stan faktyczny | Decyzja          |                  |
|----------------|------------------|------------------|
|                | $\psi(x) = 0$    | $\psi(x) = 1$    |
| $H = 0$        | decyzja poprawna | błąd I rodzaju   |
| $H = 1$        | błąd II rodzaju  | decyzja poprawna |

Tabela 2.1: Możliwe scenariusze w procesie testowania hipotezy zerowej.

Prawdopodobieństwo popełniania błędu I rodzaju dla prostej hipotezy zerowej wyraża się wzorem

$$P_{\theta_0}(B) = P_{\theta_0}(\psi(x) = 1).$$

Prawdopodobieństwo popełnienia błędu I rodzaju będziemy też oznaczać  $P(B|H = 0)$  (nie jest to prawdopodobieństwo warunkowe, ponieważ  $H$  nie jest zmienną losową). Prawdopodobieństwo popełniania błędu II rodzaju zależy od wartości  $\theta^*$ . Maksymalne prawdopodobieństwo popełnienia błędu II rodzaju wynosi

$$\max Pr(\text{popełnimy błąd II rodzaju}) = \sup_{\theta \in \Theta \setminus \Theta_0} P_{\theta}(B^C) = 1 - \inf_{\theta \in \Theta \setminus \Theta_0} P_{\theta}(B).$$

Nie zawsze można minimalizować jednocześnie prawdopodobieństwa popełnienia błędu I i II rodzaju. J. Neyman i E. Pearson zaproponowali, by uznać błąd I rodzaju za istotniejszy. Jeżeli reguła testowania gwarantuje, że prawdopodobieństwo popełnienia błędu I rodzaju jest nie większe niż  $\alpha \in (0, 1)$ , to mówimy, że kontrolowany jest błąd pierwszego rodzaju na poziomie  $\alpha$ . W praktyce oznacza to, że wielokrotnie powtarzając eksperyment losowy średnio nie częściej niż z częstością  $\alpha$  odrzucimy prawdziwą hipotezę zerową.

Konstruując testy statystyczne wymagamy, by kontrolowany był błąd pierwszego rodzaju na poziomie  $\alpha$  oraz by prawdopodobieństwo popełnienia błędu II rodzaju było możliwie małe (lub równoważnie by moc testu była możliwie duża). Z grupy procedur kontrolujących błąd I rodzaju wybieramy tę, dla której uzyskuje się największą moc. W tej pracy rozważania teoretyczne dotyczą wyłącznie kontroli błędu I rodzaju. W rozdziale 4 przedstawione są wyniki symulacyjne dla pewnego odpowiednika mocy.

Wybór poziomu istotności zależy zarówno od natury problemu jak i badacza. W zastosowaniach, najczęściej wybiera się  $\alpha = 0.05$  lub  $\alpha = 0.01$ .



## 2.4 Wartość p

**Definicja 6** *Wartość p (lub p-wartość, ang. p-value) jest równa najmniejszemu poziomowi istotności, na którym dla obserwacji  $x \in \mathcal{X}$  przyjmujemy hipotezę  $H_0$ . Wartość p wyraża się wzorem*

$$p(x) = 1 - F_{\theta_0}(T(x)|H = 0), \quad (2.1)$$

gdzie  $F_{\theta_0}(T(x)|H = 0)$  jest wartością dystrybuanty rozkładu statystyki testowej  $T$  dla prawdziwej hipotezy zerowej, w punkcie  $T(x)$ .

Wartość p, jest zmienną losową określoną na odcinku  $[0, 1]$ . Ma ona następujące własności:

- jeżeli dystrybuanta  $F_{\theta_0}(\cdot|H = 0)$  jest ciągła, a hipoteza zerowa jest prawdziwa, to p-wartość ma rozkład jednostajny na odcinku  $[0, 1]$ , czyli  $p(x) \sim \mathcal{U}[0, 1]$ ,
- jeżeli hipoteza alternatywna jest prawdziwa, to p-wartość ma rozkład stochastycznie mniejszy od rozkładu jednostajnego  $\mathcal{U}[0, 1]$ .

Ze względu na te własności, wygodnie jest opisywać obszary przyjęcia i odrzucenia w terminach p-wartości. Wszystkie prezentowane w kolejnych rozdziałach procedury testowania opisywane będą w ten sposób, dzięki czemu uniezależniamy się od wyboru przestrzeni  $\mathcal{X}$ , rodziny rozkładów  $\mathcal{P}$  oraz statystyki testowej  $T$ .

Jeżeli dystrybuanta  $F_{\theta_0}(T(x)|H = 0)$  nie jest ciągła, to p-wartość można zastąpić przez zrandomizowaną p-wartość (ang. fuzzy p-value [Geyer, Meeden 2005]), która ma wszystkie opisane powyżej własności.

## Rozdział 3

# Testowanie zbioru hipotez

Technologia, wykorzystywana w badaniach genetycznych, pozwala dziś na pomiar 22 tys. cech w trwającym zaledwie kilka godzin eksperymencie (ekspresje tyłu genów można zmierzyć w jednym badaniu mikromacierzowym). Dostęp do tak otrzymanych zbiorów danych pozwala na postawienie i weryfikację tysięcy hipotez. Rozważając tak wiele hipotez, kontrolowanie współczynnika opisującego średnią liczbę błędów popełnionych dla wszystkich hipotez jest ważniejsze niż kontrolowanie prawdopodobieństwa popełnienia błędu I rodzaju dla poszczególnych hipotez. W tym rozdziale przedstawione są najpopularniejsze współczynniki błędów oraz procedury pozwalające na ich kontrolę.

### 3.1 Zbiór eksperymentów losowych i zbiór hipotez statystycznych

Rozważmy zbiór  $m$  eksperymentów losowych  $\mathcal{E} = \{\mathcal{E}^{(i)} : i \in I = \{1, 2, \dots, m\}\}$ , określonych tak jak w Rozdziale 2. Symbolem  $\mathcal{X}^{(i)}$  oznaczmy zbiór wszystkich możliwych wyników eksperymentu  $\mathcal{E}^{(i)}$ .

Na zbiorze  $\mathcal{X}^{(i)}$  określamy rodzinę rozkładów prawdopodobieństwa  $\mathcal{P}^{(i)} = \{P_{\theta^{(i)}}^{(i)} : \theta^{(i)} \in \Theta^{(i)}\}$  indeksowaną parametrem  $\theta^{(i)} \in \Theta^{(i)}$ . Eksperymentowi  $\mathcal{E}^{(i)}$  odpowiada pewien nieznan rozkład prawdopodobieństwa  $P^{(i)}$  na  $\mathcal{X}^{(i)}$ , zgodnie z którym w wyniku przeprowadzenia eksperymentu  $\mathcal{E}^{(i)}$  obserwujemy  $x^{(i)} \in \mathcal{X}^{(i)}$ . Zakładamy, że rozkład ten należy do rodziny  $\mathcal{P}^{(i)}$ , czyli że  $P^{(i)} = P_{\theta^{*(i)}}^{(i)}$  dla pewnego  $\theta^{*(i)} \in \Theta^{(i)}$ .

Do każdego eksperymentu  $\mathcal{E}^{(i)}$ ,  $i \in I$ , badacz formułuje przypuszczenie dotyczące parametru  $\theta^{*(i)}$ . Przypuszczenia te są związane z wyborem podzbiorów  $\Theta_0^{(i)} \subset \Theta^{(i)}$ . Przypuszczenie, że  $\theta^{*(i)} \in \Theta_0^{(i)}$  nazywamy  $i$ -tą hipotezą zerową oraz oznaczamy symbolem  $H_0^{(i)}$ . Przypuszczenie, że  $\theta^{*(i)} \notin \Theta_0^{(i)}$  nazywamy  $i$ -tą hipotezą alternatywną oraz oznaczamy symbolem  $H_A^{(i)}$ . Dla  $i \in I$  określamy indeksy

$$H^{(i)} = \begin{cases} 0, & \text{gdy hipoteza } H_0^{(i)} \text{ jest prawdziwa, } \theta^{*(i)} \in \Theta_0^{(i)}, \\ 1, & \text{gdy hipoteza } H_0^{(i)} \text{ jest fałszywa, } \theta^{*(i)} \notin \Theta_0^{(i)}. \end{cases}$$

Wprowadźmy ponadto następujące oznaczenia:  $\mathcal{X} = \mathcal{X}^{(1)} \times \dots \times \mathcal{X}^{(m)}$ ,  $x = (x^{(1)}, \dots, x^{(m)})$ ,  $\theta = (\theta^{(1)}, \dots, \theta^{(m)})$ .

Na podstawie wektora obserwacji  $x$  mamy zdecydować, które hipotezy zerowe  $H_0^{(i)}$ ,  $i \in I$  przyjąć, a które odrzucić.

**Definicja 7** *Procedurą testowania zbioru hipotez nazywamy regułę określającą, w jakiej kolejności oraz na jakim poziomie istotności należy testować poszczególne hipotezy zerowe.*

Określenie procedury testowania zbioru hipotez jest równoważne z określeniem rozbicia  $\mathcal{X}$  na dwa rozłączne podzbiory  $B^{(i)} \dot{\cup} (B^{(i)})^C$  ze względu na każdą hipotezę  $H_0^{(i)}$ . Odrzucamy hipotezę zerową  $H_0^{(i)}$ , jeżeli  $x \in B^{(i)} \subset \mathcal{X}$ , a przyjmujemy w przeciwnym wypadku.

**Definicja 8** *Symbolem  $\psi = (\psi^{(1)}, \dots, \psi^{(m)})$  oznaczamy decyzje podjęte w wyniku testowania zbioru  $m$  hipotez. Współrzędne  $\psi^{(i)} \in \{0, 1\}$  określają, czy hipoteza  $H_0^{(i)}$  została przyjęta ( $\psi^{(i)} = 0$ ), czy odrzucona ( $\psi^{(i)} = 1$ ).*

Procedury testowania oraz obszary przyjęcia i odrzucenia w kolejnych rozdziałach są przedstawione w terminach p-wartości. Symbolem  $p^{(i)}$  oznaczamy p-wartość wyznaczoną dla hipotezy  $i$ -tej ze wzoru (2.1). O ile nie zaznaczymy, że jest inaczej, w tej pracy ograniczamy się do przypadku, gdy p-wartości (statystyki testowe) dla różnych hipotez są niezależnymi zmiennymi losowymi.

Rozkłady zmiennych  $\psi^{(i)}$ , dla  $i \in I$ , zależą od obranych obszarów przyjęcia i odrzucenia  $(B^{(i)})^C$  i  $B^{(i)}$  oraz wartości  $\theta^{*(i)}$

$$\psi^{(i)}(x) = \begin{cases} 0, & \text{gdy } x \notin B^{(i)}, \\ 1, & \text{gdy } x \in B^{(i)}. \end{cases} \quad (3.1)$$

### 3.2 Zmienne losowe $V$ , $R$ , $S$ , $Q$ i współczynniki błędów

Określmy zmienne losowe  $V$ ,  $R$ ,  $S$ , i  $Q$  następująco (patrz Tabela 3.1).

- Zmienną losową  $V$  określamy liczbę odrzuconych prawdziwych hipotez zerowych,

$$V = \sum_{i \in I} \psi^{(i)}(1 - H^{(i)}). \quad (3.2)$$

- Zmienną losową  $S$  określamy liczbę odrzuconych fałszywych hipotez zerowych,

$$S = \sum_{i \in I} \psi^{(i)} H^{(i)}. \quad (3.3)$$

- Zmienną losową  $R$  określamy liczbę odrzuconych hipotez zerowych,

$$R = V + S = \sum_{i \in I} \psi^{(i)}. \quad (3.4)$$

- Zmienną losową  $Q$  określamy następująco

$$Q = \begin{cases} V/R, & \text{gdy } R > 0, \\ 0, & \text{gdy } R = 0. \end{cases} \quad (3.5)$$

|                               | #przyjętych<br>hipotez zerowych | #odrzuconych<br>hipotez zerowych | suma  |
|-------------------------------|---------------------------------|----------------------------------|-------|
| #prawdziwych hipotez zerowych | $m_0 - V$                       | $V$                              | $m_0$ |
| #fałszywych hipotez zerowych  | $m_1 - S$                       | $S$                              | $m_1$ |
| suma                          | $m - R$                         | $R$                              | $m$   |

Tabela 3.1: Zmienne losowe  $V$ ,  $S$  i  $R$  [Benjamini, Hochberg 1995].

|           | Stan hipotezy $H_0^{(i)}$ |               |               |               |
|-----------|---------------------------|---------------|---------------|---------------|
|           | $H^{(i)} = 0$             |               | $H^{(i)} = 1$ |               |
|           | Decyzja                   |               | Decyzja       |               |
|           | $\psi(i) = 1$             | $\psi(i) = 0$ | $\psi(i) = 1$ | $\psi(i) = 0$ |
| $V^{(i)}$ | 1                         | 0             | 0             | 0             |
| $S^{(i)}$ | 0                         | 0             | 1             | 0             |
| $R^{(i)}$ | 1                         | 0             | 1             | 0             |

Tabela 3.2: Zmienne losowe  $V^{(i)}$ ,  $S^{(i)}$  i  $R^{(i)}$ .

Rozkłady tych zmiennych zależą od rozkładów  $\psi^{(i)}$ ,  $i \in I$  (patrz wzór (3.1)). Symbolami  $V^{(i)}$ ,  $S^{(i)}$  i  $R^{(i)}$  oznaczmy odpowiedniki zmiennych  $V$ ,  $S$  i  $R$  wyznaczone dla jednej hipotezy zerowej  $H_0^{(i)}$  (patrz Tabela 3.2).

W literaturze poświęconej zagadnieniu testowania zbioru hipotez rozważa się wiele różnych współczynników błędów (przegląd można znaleźć w pracach [Dudoit i inni 2002], [Sarkar 2002]). W tej pracy jest rozważana kontrola trzech najpopularniejszych, tj. FWER, PFER i FDR. Poniżej przedstawiamy definicje tych współczynników, a w Sekcjach 3.4, 3.5 i 3.6 przedstawimy procedury kontrolujące te współczynniki.

**Współczynnik PFER (ang. per-family error rate)** równy jest oczekiwanej liczbie odrzuconych prawdziwych hipotez,

$$PFER = E(V).$$

**Współczynnik FWER (ang. family-wise error rate)** równy jest prawdopodobieństwu odrzucenia co najmniej jednej prawdziwej hipotezy

$$FWER = Pr(V \geq 1).$$

**Współczynnik FDR (ang. false discovery rate)**, zaproponowany przez Yoava Benjamini i Yosefa Hochberga [Benjamini, Hochberg 1995], równy jest wartości oczekiwanej frakcji fałszywie odrzuconych hipotez zerowych w zbiorze wszystkich odrzuconych hipotez zerowych

$$FDR = E(Q) = E(V/R | R > 0) Pr(R > 0).$$

Pomiędzy tymi współczynnikami zachodzą nierówności

$$PFER \geq FWER \geq FDR, \tag{3.6}$$

które wynikają z nierówności

$$V \geq 1_{\{V>0\}} \geq \frac{V}{R} 1_{\{R>0\}},$$

gdzie  $1_{\Omega}$  oznacza funkcję charakterystyczną.

**Definicja 9** *Mówimy, że procedura kontroluje w sensie mocnym współczynnik błędu na poziomie  $\alpha > 0$ , jeżeli bez względu na to ile oraz które hipotezy zerowe są prawdziwe, przyjęte obszary odrzucenia  $B^{(i)}, i \in I$  gwarantują, że powtarzając zbiór eksperymentów losowych dany współczynnik nie będzie średnio większy niż  $\alpha$ .*

W literaturze rozważana jest również kontrola w sensie słabym, jednak bez dodatkowych założeń jest ona nieprzydatna w praktycznych zastosowaniach i nie jest rozważana w tej pracy. Dziś na znaczeniu zyskuje współczynnik FDR, ponieważ kontrolujące go procedury charakteryzują się średnio większą liczbą poprawnych odrzuceń niż w przypadku kontroli innych współczynników.

Podkreślmy różnicę pomiędzy interpretacją zmiennych  $V$  i  $S$  w ujęciu częstościowym i w ujęciu Bayesowskim. W ujęciu Bayesowskim stan hipotezy zerowej  $H^{(i)}$  jest zmienną losową, podczas gdy w ujęciu częstościowym jest to nieznaną wartość stała dla eksperymentu. Zatem rozkłady zmiennych  $V$  i  $S$  są różne w zależności od przyjętego ujęcia (patrz wzory (3.2) i (3.3))

W literaturze prezentowane są procedury kontrolujące współczynniki PFER, FWER i FDR zarówno w ujęciu Bayesowskim jak i w ujęciu częstościowym.

W ujęciu Bayesowskim kontrolę danego współczynnika błędu wykazuje się dla ustalonego rozkładu a priori zmiennych  $H^{(i)}, i \in I$ . W ujęciu częstościowym kontrolę danego współczynnika wykazuje się dla każdego możliwego stanu hipotez zerowych, a więc dla każdego możliwego stanu wektora  $H = (H^{(1)}, \dots, H^{(m)})$ . Kontrola w ujęciu częstościowym oznacza, że średnia wartość danego współczynnika błędu uśredniona z wielu powtórzeń zbioru eksperymentów jest mniejsza lub równa ustalonej wielkości.

Procedura kontrolująca współczynnik błędu w ujęciu częstościowym kontroluje ten współczynnik również w ujęciu Bayesowskim, można więc powiedzieć, że ujęcie częstościowe jest „bardziej wymagające”. W tej pracy rozważamy kontrolę współczynników błędu w ujęciu częstościowym.

### 3.3 Procedury testowania zbioru hipotez

Opisywanie procedur testowania hipotez przez specyfikowanie obszarów przyjęcia  $(B^{(i)})^C$  i odrzucenia  $B^{(i)}$  jest niepraktyczne w użyciu. Procedury testowania są najczęściej opisywane jako jednokrokowe lub wielokrokowe reguły.

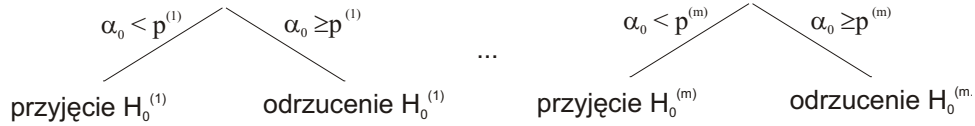
#### Procedury jednokrokowe

W procedurze jednokrokowej każda z  $m$  hipotez zerowych jest testowana niezależnie od wyniku testowania pozostałych. Wszystkie hipotezy zerowe są testowane na takim samym poziomie istotności  $\alpha_0$ . Schemat procedury jednokrokowej przedstawiamy na Rysunku 3.1. Odrzucane są hipotezy zerowe, dla których  $p$ -wartości  $p^{(i)}$  są mniejsze niż  $\alpha_0$ . W sekcji 3.5 przedstawiamy metody wyboru poziomu istotności  $\alpha_0$ .

Zauważmy, że obszary odrzuceń określone przez procedury jednokrokowe są postaci

$$B^{(i)} = \mathcal{X}^{(1)} \times \dots \times \mathcal{X}^{(i-1)} \times B_i \times \mathcal{X}^{(i+1)} \times \dots \times \mathcal{X}^{(m)},$$

gdzie  $B_i \subset \mathcal{X}^{(i)}$ , dla  $i \in I$ .



Rysunek 3.1: Schemat procedury jednokrokowej. Każda z  $m$  hipotez testowana jest niezależnie od pozostałych na tym samym poziomie istotności  $\alpha_0$ .

## Procedury wielokrokowe

Procedury jednokrokowe są konserwatywne, to znaczy, że dla dużych zbiorów hipotez prowadzą do odrzucenia bardzo małej liczby hipotez. Dlatego też opracowano procedury wielokrokowe, wykorzystujące informacje o łącznym rozkładzie  $p$ -wartości dla wszystkich hipotez. To podejście prowadzi do wyższej liczby odrzuconych hipotez, przy zapewnieniu kontroli danego współczynnika błędu.

Wprowadźmy następujące oznaczenia. Indeks górny  $(i:m)$  dla  $p$ -wartości oznacza  $i$ -tą statystykę pozycyjną, a więc  $p^{(1:m)}$  oznacza najmniejszą  $p$ -wartość, a  $p^{(m:m)}$  największą. Hipotezę zerową odpowiadającą  $p$ -wartości  $p^{(i:m)}$  oznaczamy  $H_0^{(i:m)}$ . Poziomy istotności dla hipotezy  $H_0^{(i:m)}$  oznaczamy  $\alpha^{(i:m)}$ . Zakładamy, że poziomy istotności tworzą niemalejący ciąg  $0 < \alpha^{(1:m)} \leq \alpha^{(2:m)} \leq \dots \leq \alpha^{(m:m)} < 1$ . W kolejnych rozdziałach przedstawimy metody wyznaczania poziomów istotności  $\alpha^{(i:m)}$ .

W procedurze wielokrokowej kolejność testowania hipotez zależy od kolejności odpowiadających im  $p$ -wartości.

### Procedura step-down

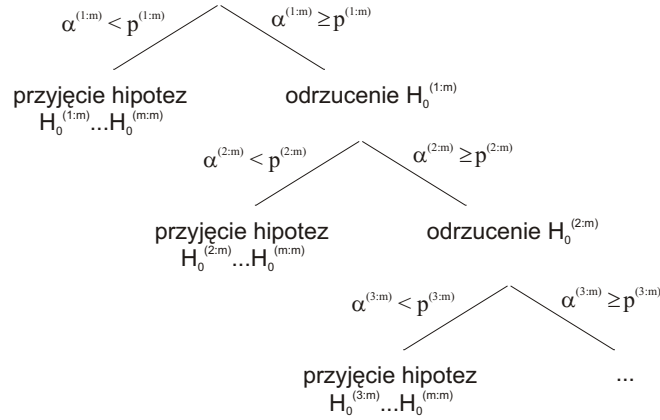
Testowanie rozpoczynamy od hipotezy  $H_0^{(1:m)}$ . Jeżeli  $p^{(1:m)} > \alpha^{(1:m)}$ , to przyjmujemy wszystkie  $m$  hipotez zerowych. W przeciwnym przypadku odrzucamy hipotezę  $H_0^{(1:m)}$  i powyższe postępowanie powtarzamy dla zbioru hipotez  $\{H_0^{(2:m)}, \dots, H_0^{(m:m)}\}$ . W  $i$ -tym kroku porównujemy  $p$ -wartość  $p^{(i:m)}$  z  $\alpha^{(i:m)}$ . Jeżeli  $p^{(i:m)} > \alpha^{(i:m)}$ , to kończymy testowanie i przyjmujemy hipotezy  $H_0^{(i:m)} \dots H_0^{(m:m)}$ . Jeżeli  $p^{(i:m)} \leq \alpha^{(i:m)}$ , to odrzucamy hipotezę  $H_0^{(i:m)}$  i przechodzimy do kroku  $i + 1$ .

Schemat tej procedury przedstawiamy na Rysunku 3.2.

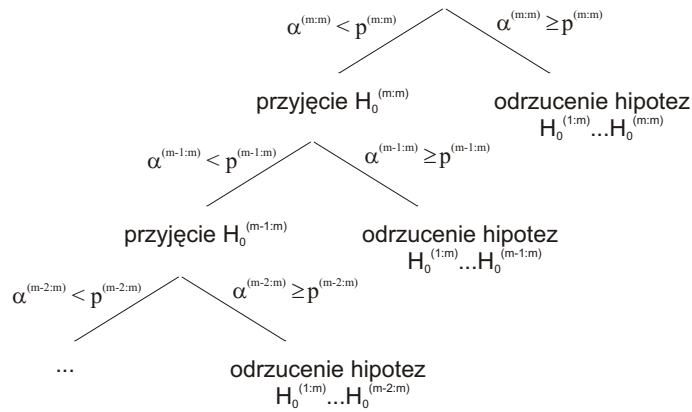
### Procedura step-up

Testowanie rozpoczynamy od hipotezy  $H_0^{(m:m)}$ . Jeżeli  $p^{(m:m)} \leq \alpha^{(m:m)}$ , to odrzucamy wszystkie  $m$  hipotez zerowych. W przeciwnym przypadku przyjmujemy hipotezę  $H_0^{(m:m)}$  i powyższe postępowanie powtarzamy dla zbioru hipotez  $\{H_0^{(1:m)}, \dots, H_0^{(m-1:m)}\}$ . W  $i$ -tym kroku porównujemy  $p$ -wartość  $p^{(m-i+1:m)}$  z  $\alpha^{(m-i+1:m)}$ . Jeżeli  $p^{(m-i+1:m)} \leq \alpha^{(m-i+1:m)}$ , to kończymy testowanie i odrzucamy hipotezy  $H_0^{(1:m)} \dots H_0^{(m-i+1:m)}$ . Jeżeli  $p^{(m-i+1:m)} > \alpha^{(m-i+1:m)}$ , to przyjmujemy hipotezę  $H_0^{(m-i+1:m)}$  i przechodzimy do kroku  $i + 1$ .

Schemat tej procedury przedstawiamy na Rysunku 3.3.



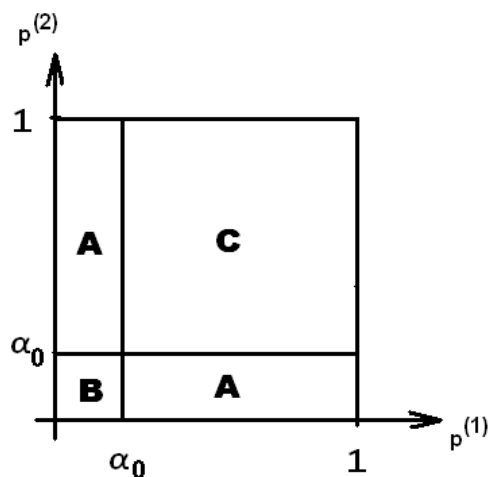
Rysunek 3.2: Schemat procedury step-down. W  $i$ -tym kroku podejmowana jest decyzja o odrzuceniu hipotezy  $H_0^{(i:m)}$  albo o przyjęciu hipotez  $H_0^{(i:m)}, \dots, H_0^{(m:m)}$ .



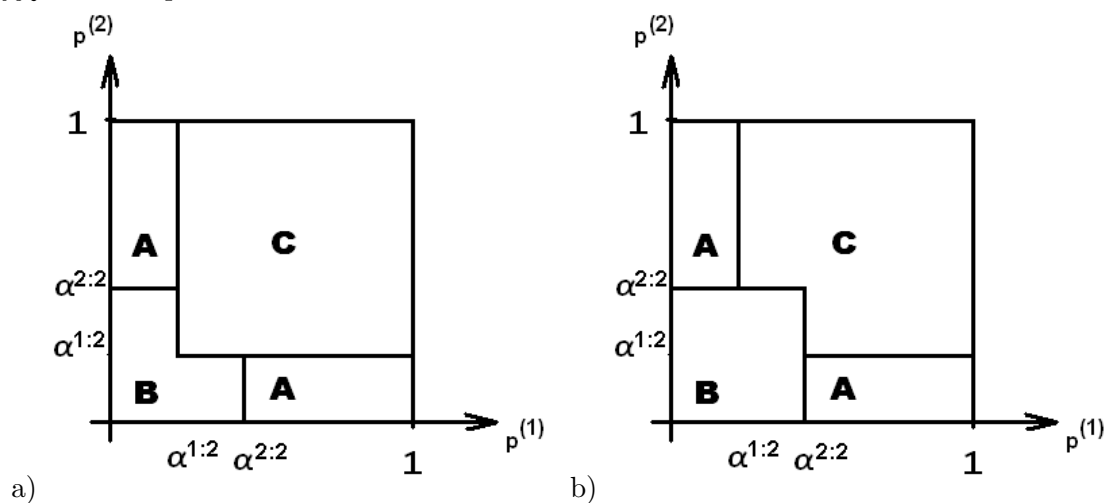
Rysunek 3.3: Schemat procedury step-up. W  $i$ -tym kroku podejmowana jest decyzja o przyjęciu hipotezy  $H_0^{(m-i+1:m)}$  albo o odrzuceniu hipotez  $H_0^{(1:m)}, \dots, H_0^{(m-i+1:m)}$ .

## Obszary odrzucenia

Na Rysunkach 3.4-3.5 przedstawiamy obszary odrzucenia i przyjęcia odpowiadające omówionym procedurom testowania dla zbioru dwóch hipotez zerowych. Przestrzeń  $\mathcal{X}$  została rozbita na trzy obszary:  $A$  - jedna hipoteza zerowa jest odrzucona,  $B$  - obie hipotezy zerowe są odrzucone i  $C$  - obie hipotezy zerowe są przyjęte. Obszar odrzucenia  $B^{(i)}$  to suma obszaru  $B$  z jednym z obszarów  $A$ .



Rysunek 3.4: Obszary przyjęcia i odrzucenia dla procedury jednokrokowej dla  $m = 2$ .  $A$  to obszar przyjęcia dokładnie jednej z hipotez,  $B$  to obszar odrzucenia obu hipotez,  $C$  to obszar przyjęcia obu hipotez.



Rysunek 3.5: Obszary przyjęcia i odrzucenia (patrz oznaczenia przy Rysunku 3.4) dla procedury step-down dla  $m = 2$  (a) oraz dla procedury step-up (b).



### 3.4 Procedury kontroli współczynnika PFER

#### Procedura Bonferroniego (nazywana też korektą Bonferroniego)

**Twierdzenie 1** [Bonferroni 1936] Jednokrokowa procedura testowania z poziomem istotności  $\alpha_0$  wyrażonym wzorem

$$\alpha_0 = \alpha/m, \quad (3.7)$$

kontroluje współczynnik PFER na poziomie  $\alpha$ . Jeżeli  $m_0 = m$ , to współczynnik PFER jest równy  $\alpha$ , w pozostałych przypadkach współczynnik PFER jest mniejszy niż  $\alpha$ .

Procedura ta kontroluje współczynnik PFER, również w przypadku skorelowanych statystyk testowych. Więcej informacji o kontroli PFER można znaleźć w pracach [Sidak 1967], [Simes 1986], [Ge i inni 2003]. Zaletą i przyczyną popularności tej procedury jest jej prostota i łatwość stosowania.

### 3.5 Procedury kontroli współczynnika FWER

#### Procedura Bonferroniego

**Twierdzenie 2** Jednokrokowa procedura testowania z poziomem istotności  $\alpha_0$ , wyrażonym wzorem (3.7), kontroluje współczynnik FWER na poziomie  $\alpha$ .

Procedura ta kontroluje współczynnik FWER również w przypadku skorelowanych statystyk testowych. Twierdzenie 1 pokazuje, że ta procedura kontroluje współczynnik PFER, a ponieważ  $\text{FWER} \leq \text{PFER}$  (patrz wzór (3.6)), to również kontroluje współczynnik FWER.

#### Procedura Holma

W roku 1979 Holm zaproponował procedurę wielokrokową, kontrolującą współczynnik FWER na poziomie  $\alpha$ . Procedura ta odpowiada większym obszarom odrzucenia w porównaniu do procedury Bonferroniego, dzięki czemu średnio odrzucanych jest więcej fałszywych hipotez zerowych.

**Twierdzenie 3** [Holm 1979] Jeżeli statystyki testowe testowanych hipotez są niezależne, to procedura wielokrokowa step-down z poziomami istotności  $\alpha^{(i:m)}$  wyrażonymi wzorem

$$\alpha^{(i:m)} = \frac{\alpha}{m - i + 1}, \quad (3.8)$$

dla  $1 \leq i \leq m$ , kontroluje współczynnik FWER na poziomie  $\alpha$ .

#### Procedura Hochberga

**Twierdzenie 4** [Hochberg 1988] Jeżeli statystyki testowe weryfikowanych hipotez są niezależne, to procedura wielokrokowa step-up z poziomami istotności  $\alpha^{(i:m)}$  wyrażonymi wzorem

$$\alpha^{(i:m)} = \frac{\alpha}{m - i + 1}, \quad (3.9)$$

dla  $1 \leq i \leq m$ , kontroluje współczynnik FWER na poziomie  $\alpha$ .

W porównaniu do procedury Holma procedura Hochberga ma większe obszary odrzucenia, przez co odrzuca średnio więcej fałszywych hipotez zerowych.

## 3.6 Procedury kontroli współczynnika FDR

### Procedura Benjaminiego Hochberga

**Twierdzenie 5** [Benjamini, Hochberg 1995] *Jeżeli statystyki testowe weryfikowanych hipotez są pozytywnie zależne (macierz korelacji statystyk testowych jest dodatnio określona), to procedura wielokrokowa step-up z poziomami istotności  $\alpha^{(i:m)}$  wyrażonymi wzorem*

$$\alpha^{(i:m)} = \frac{i}{m}\alpha, \quad (3.10)$$

dla  $1 \leq i \leq m$ , kontroluje współczynnik FDR na poziomie  $\alpha$ .

Ta procedura została przedstawiona w pracy [Benjamini, Hochberg 1995], w której przedstawiono dowód kontroli współczynnika FDR, przy założeniu, że statystyki testowe są niezależne. W pracy [Benjamini, Yekutieli 2001] przedstawiono ogólniejszy dowód tego twierdzenia, przy założeniu, że statystyki testowe są pozytywnie zależne.

Kontroli współczynnika FDR poświęconych jest wiele prac, między innymi [Benjamini i inni 2005], [Kulinskaya, Lewin 2006], [Futschik, Posch 2005] [Owen 2005], [Storey 2002], [Storey i inni 2004].

### Inne procedury

Powyżej przedstawione procedury testowania nie uwzględniają informacji o strukturze korelacji statystyk testowych. Dla wybranych zastosowań (i związanych z nimi specyficznych struktur korelacji) opracowano procedury testowania pozwalające na wykorzystanie tej dodatkowej informacji. Wiele korekt zaproponowano dla zagadnienia lokalizacji genów sprzężonych z cechami ilościowymi [Siegmund 1985], [Rebai i inni 1994], [Cheverud 2001] i [Biecek 2005].

## Rozdział 4

# Testowanie zbioru hipotez z relacją hierarchii

### 4.1 Zbiór hipotez i relacja hierarchii

Rozważmy zbiór  $m$  eksperymentów losowych oraz zbiór  $m$  odpowiadających im hipotez, tak jak w Rozdziale 3. Na zbiorze rozważanych hipotez określamy relację hierarchii  $\mathcal{R}$ .

**Definicja 10** *Relacją hierarchii nazywamy antyzwrotną, przechodnią i antysymetryczną relację  $\mathcal{R} : I \times I \rightarrow \{0, 1\}$  określoną na iloczynie kartezjańskim zbioru indeksów  $I$ .*

Jeżeli  $R(i, j) = 1$  to hipoteza  $H_0^{(i)}$  jest w relacji z hipotezą  $H_0^{(j)}$ . W myśl Definicji 10, relacja hierarchii  $\mathcal{R}$  spełnia dla każdego  $i, j, k \in I$  następujące warunki:

- $(R(i, j) = 1) \wedge (R(j, k) = 1) \Rightarrow (R(i, k) = 1)$ ,
- $R(i, j) + R(j, i) \leq 1$ .

Każdą relację hierarchii można przedstawić jako acykliczny graf skierowany (ang. DAG, directed acyclic graph). Przykładowe relacje przedstawiamy w postaci grafowej na Rysunku 4.1 oraz w postaci funkcyjnej w Tabelach 4.1-4.3. Przykładem relacji hierarchii jest relacja liniowa.

**Definicja 11** *Relacją liniową  $\mathcal{R}_{lin}$  na zbiorze  $\mathcal{H}$  nazywamy każdą relację hierarchii spełniającą warunek*

$$\forall_{i, j \in I} [R_{lin}(i, j) + R_{lin}(j, i) = 1] \Leftrightarrow [i \neq j].$$

Przykładową relację liniową przedstawia Rysunek 4.1 i Tabela 4.3.

| $R_1(i, j)$ | j=1 | j=2 | j=3 | j=4 | j=5 | j=6 | j=7 |
|-------------|-----|-----|-----|-----|-----|-----|-----|
| i=1         | 0   | 1   | 1   | 1   | 1   | 1   | 1   |
| i=2         | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| i=3         | 0   | 0   | 0   | 1   | 1   | 1   | 1   |
| i=4         | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| i=5         | 0   | 0   | 0   | 0   | 0   | 1   | 1   |
| i=6         | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| i=7         | 0   | 0   | 0   | 0   | 0   | 0   | 0   |

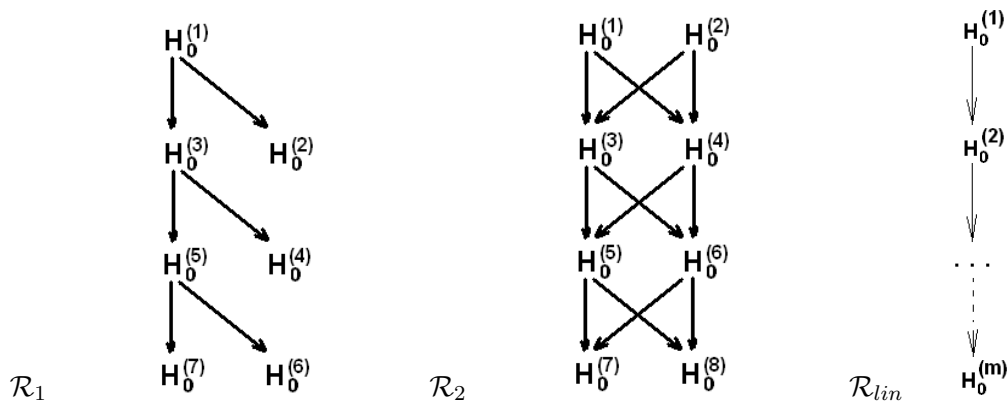
Tabela 4.1: Relacja hierarchii  $\mathcal{R}_1$ .

| $R_2(i, j)$ | j=1 | j=2 | j=3 | j=4 | j=5 | j=6 | j=7 | j=8 |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
| i=1         | 0   | 0   | 1   | 1   | 1   | 1   | 1   | 1   |
| i=2         | 0   | 0   | 1   | 1   | 1   | 1   | 1   | 1   |
| i=3         | 0   | 0   | 0   | 0   | 1   | 1   | 1   | 1   |
| i=4         | 0   | 0   | 0   | 0   | 1   | 1   | 1   | 1   |
| i=5         | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   |
| i=6         | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   |
| i=7         | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| i=8         | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |

Tabela 4.2: Relacja hierarchii  $\mathcal{R}_2$ .

| $R_{lin}(i, j)$ | j=1 | j=2 | j=3 | j=4 | j=5 | j=6 | j=7 | j=8 |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|
| i=1             | 0   | 1   | 1   | 1   | 1   | 1   | 1   | 1   |
| i=2             | 0   | 0   | 1   | 1   | 1   | 1   | 1   | 1   |
| i=3             | 0   | 0   | 0   | 1   | 1   | 1   | 1   | 1   |
| i=4             | 0   | 0   | 0   | 0   | 1   | 1   | 1   | 1   |
| i=5             | 0   | 0   | 0   | 0   | 0   | 1   | 1   | 1   |
| i=6             | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 1   |
| i=7             | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   |
| i=8             | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |

Tabela 4.3: Liniowa relacja hierarchii  $\mathcal{R}_{lin}$  dla  $m = 8$ .



Rysunek 4.1: Grafowa reprezentacja relacji hierarchii  $\mathcal{R}_1$ ,  $\mathcal{R}_2$  oraz liniowej relacji hierarchii  $\mathcal{R}_{lin}$ .

Relacja hierarchii opisuje semantyczną zależność pomiędzy stanami hipotez zerowych i wynika ona z rozważanego zagadnienia. Jeżeli hipoteza  $H_0^{(i)}$  jest w relacji do hipotezy  $H_0^{(j)}$ , to prawdziwość hipotezy  $H_0^{(i)}$  implikuje prawdziwość hipotezy  $H_0^{(j)}$ , czyli

$$\forall_{i,j \in I} [R(i,j) = 1] \Rightarrow [(H^{(i)} = 0) \Rightarrow (H^{(j)} = 0)].$$

Przykład relacji hierarchii dla zagadnienia związanego z analizą danych genetycznych jest przedstawiony w Rozdziale 5.

Wszystkich możliwych decyzji, które można podjąć w procesie testowania, jest  $2^m$  (w myśl Definicji 8). Jednak w procesie testowania rozsądnie jest podejmować tylko decyzje odpowiadające możliwym stanom hipotez zerowych. Dlatego decyzje z procesu testowania powinny być zgodne z relacją  $\mathcal{R}$  i tylko takie decyzje będziemy rozważać w dalszej części pracy.

**Definicja 12** *Decyzja  $\psi = (\psi^{(1)}, \dots, \psi^{(m)})$  jest zgodna z relacją  $\mathcal{R}$  wtedy i tylko wtedy, gdy*

$$\forall_{i,j \in I} [R(i,j) = 1] \Rightarrow [(\psi^{(i)} = 0) \Rightarrow (\psi^{(j)} = 0)].$$

Zauważmy, że każda decyzja jest zgodna z relacją pustą, tzn. taką, dla której  $\forall_{i,j} R(i,j) = 0$ .

Zagadnienie testowania zbioru hipotez, w którym decyzje muszą być zgodne z relacją  $\mathcal{R}$  nazywamy zagadnieniem testowania zbioru hipotez z relacją hierarchii  $\mathcal{R}$ . Nad kontrolą współczynnika FWER w zagadnieniach tego typu pracowało wielu statystyków od ponad 30 lat (taka struktura zależności pomiędzy hipotezami występuje między innymi w zagadnieniu „Maximum safe dose”). Pierwsze badania nad tym zagadnieniem prowadził E. Peritz (wyniki na ten temat znalazły się w jego rozprawie doktorskiej). Wiele istotnych prac pochodzi z lat 70-tych [Marcus i inni 1976], [Sonnemann, Finner 1988]. Przegląd wyników można znaleźć w pracach [Bauer 1997], [Finner, Strassburger 2002], [Tamhane i inni 1996] lub [Tamhane, Logan 2004].

W tych pracach relację hierarchii pomiędzy hipotezami wprowadza się w odmienny sposób niż prezentowany tutaj. Relacja hierarchii jest wywiedziona z relacji pomiędzy zbiorami  $\Theta_0^{(i)}, i \in I$  odpowiadającymi poszczególnym hipotezom zerowym. Autorzy określają dla zbioru wszystkich hipotez wspólną przestrzeń parametrów  $\Theta$ . Hipoteza zerowa  $H_0^{(i)}$  odpowiada przypuszczeniu, że  $\theta^* \in \Theta_0^{(i)} \subset \Theta$ . Ponieważ zbiory  $\Theta_0^{(i)}, i \in I$  są z tej samej przestrzeni  $\Theta$ , więc mogą zachodzić pomiędzy nimi relacje zawierania. Te relacje można przekształcić w relację hierarchii na zbiorze hipotez zerowych. Jeżeli zbiór  $\Theta_0^{(i)}$  jest zawarty w zbiorze  $\Theta_0^{(j)}$ , to przyjęcie hipotezy zerowej  $H_0^{(i)}$  wymaga przyjęcia hipotezy zerowej  $H_0^{(j)}$ .

W pewnych zagadnieniach (jak np. w omówionym w następnym rozdziale zagadnieniu dotyczącym hierarchicznej struktury Gene Ontology [Goeman 2004]) relacja hierarchii pomiędzy hipotezami nie wynika wprost z zależności pomiędzy testowanymi zbiorami parametrów, ale z interpretacji poszczególnych hipotez. Trudno jest określić w naturalny sposób zbiór  $\Theta$  i podzbiory  $\Theta_0^{(i)}, i \in I$  tak, by zachować pomiędzy hipotezami relację wynikającą ze struktury Gene Ontology (patrz Rozdział 5). W tego typu zagadnieniach nie można zastosować procedur wykorzystujących relacje pomiędzy zbiorami parametrów  $\Theta_0^{(i)}$ .

W tej pracy relacja hierarchii określona jest niezależnie od postaci hipotez zerowych. To podejście pozwala na stosowanie uzyskanych wyników dla szerszej klasy problemów. Poza wynikami dla kontroli współczynnika FWER w tej pracy są również przedstawione procedury testowania kontrolujące współczynniki FDR oraz PFER. W kontekście omówionych w Rozdziale 5 zastosowań szczególnie interesująca jest kontrola współczynnika FDR.

## 4.2 Procedury testowania zbioru hipotez

W tej sekcji przedstawiamy dwie nowe procedury testowania, których wyniki są zgodne z zadaną relacją  $\mathcal{R}$ .

### Wstępująca procedura testowania

W wielokrokowej wstępującej procedurze testowania w każdym kroku testowana jest hipoteza zerowa, nie będąca w relacji do żadnej innej nieprzetestowanej hipotezy zerowej. W jednym kroku przyjmowana jest jedna hipoteza zerowa albo odrzucany jest podzbiór hipotez zerowych.

Wstępującą procedurę testowania opisuje algorytm:

1. Przez  $K$  oznaczmy zbiór indeksów nie przetestowanych dotychczas hipotez. W pierwszym kroku  $K = I = \{1, \dots, m\}$ .
2. Ze zbioru  $K$  wybieramy indeks  $i$  hipotezy zerowej nie będącej w relacji do żadnej innej hipotezy zerowej o indeksie ze zbioru  $K$ , tzn. ze zbioru

$$\{i : i \in K \wedge \forall_{j \in K} R(i, j) = 0\}.$$

W przypadku, gdy zbiór ten zawiera więcej niż jeden element, wybieramy dowolny.

3. Jeżeli test odrzuci hipotezę  $H_0^{(i)}$  na poziomie istotności  $\alpha^{(i)}$ , to odrzucamy również wszystkie hipotezy będące w relacji do  $H_0^{(i)}$  i usuwamy ich indeksy ze zbioru  $K$ . W przeciwnym razie przyjmujemy tę hipotezę, a ze zbioru  $K$  usuwamy indeks  $i$ .
4. Jeżeli zbiór  $K$  nie jest pusty, to wracamy do punktu 2.

### Zstępująca procedura testowania

W wielokrokowej zstępującej procedurze testowania w każdym kroku testowana jest hipoteza zerowa, w stosunku do której żadna inna nieprzetestowana hipoteza nie jest w relacji. W jednym kroku odrzucana jest jedna hipoteza zerowa albo przyjmowany jest podzbiór hipotez zerowych. Zstępującą procedurę testowania opisuje algorytm:

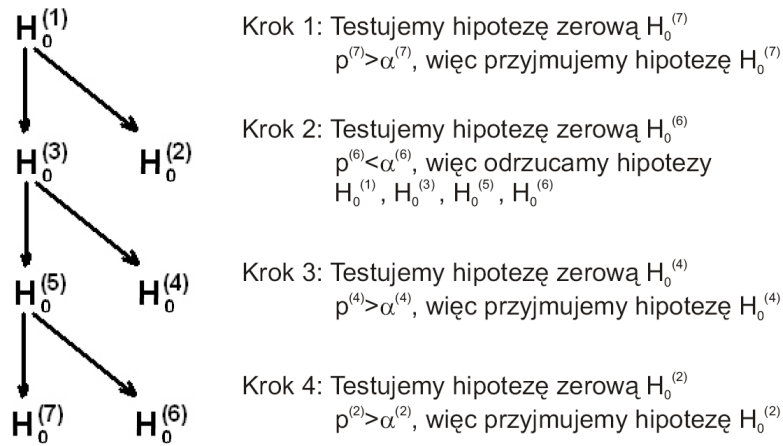
1. Przez  $K$  oznaczmy zbiór indeksów nie przetestowanych dotychczas hipotez. W pierwszym kroku  $K = I = \{1, \dots, m\}$ .
2. Ze zbioru  $K$  wybieramy indeks  $i$  hipotezy zerowej do której żadna inna hipoteza o indeksie ze zbioru  $K$  nie jest w relacji, tzn. ze zbioru

$$\{j : j \in K \wedge \forall_{i \in K} R(i, j) = 0\}.$$

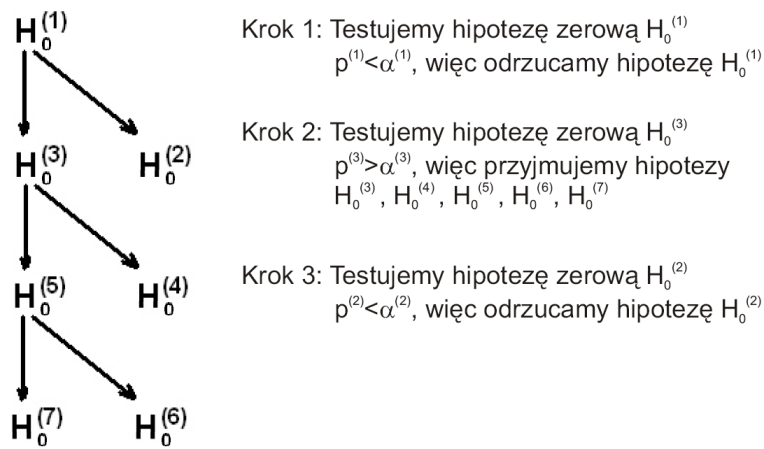
W przypadku, gdy zbiór ten zawiera więcej niż jeden element, wybieramy dowolny.

3. Jeżeli test przyjmie hipotezę  $H_0^{(i)}$  na poziomie istotności  $\alpha^{(i)}$ , to przyjmujemy również wszystkie hipotezy do których w relacji jest  $H_0^{(i)}$  i usuwamy ich indeksy ze zbioru  $K$ . W przeciwnym razie odrzucamy hipotezę  $H_0^{(i)}$ , a ze zbioru  $K$  wykreślamy indeks  $i$ .
4. Jeżeli zbiór  $K$  nie jest pusty, to wracamy do punktu 2.

Przykład obu procedur testowania są przedstawione na Rysunkach 4.2 i 4.3.



Rysunek 4.2: Przykład wstępującej procedury testowania dla relacji  $\mathcal{R}_1$ , gdy  $p^{(2)} > \alpha^{(2)}$ ,  $p^{(4)} > \alpha^{(4)}$ ,  $p^{(6)} < \alpha^{(6)}$ ,  $p^{(7)} > \alpha^{(7)}$ .



Rysunek 4.3: Przykład zstępującej procedury testowania dla relacji  $\mathcal{R}_1$ , gdy  $p^{(1)} < \alpha^{(1)}$ ,  $p^{(2)} < \alpha^{(2)}$ ,  $p^{(3)} > \alpha^{(3)}$ .

### 4.3 Przykład testowania zbioru dwóch hipotez zerowych będących w relacji hierarchii

Rozważmy dwie hipotezy zerowe  $H_0^{(1)}$  i  $H_0^{(2)}$ , z których druga jest fałszywa tylko wtedy, jeżeli pierwsza jest fałszywa. Odpowiada to relacji liniowej na zbiorze dwóch hipotez (patrz Rysunek 4.4 oraz Tabela 4.4).



Rysunek 4.4: Reprezentacja grafowa relacji  $\mathcal{R}_3$ .

| $R_3(i, j)$ | j=1 | j=2 |
|-------------|-----|-----|
| i=1         | 0   | 1   |
| i=2         | 0   | 0   |

Tabela 4.4: Relacja  $\mathcal{R}_3$ . Symbole  $i, j$  odpowiadają indeksom hipotez.

W tym wypadku możliwe są trzy następujące stany rozważanych hipotez  $H_0^{(1)} = 0$  i  $H_0^{(2)} = 0$  albo  $H_0^{(1)} = 1$  i  $H_0^{(2)} = 0$  albo  $H_0^{(1)} = 1$  i  $H_0^{(2)} = 1$ .

Decyzje zgodne z relacją  $\mathcal{R}_3$  to:

- $\psi_1 = (0, 0)$ : przyjąć  $H_0^{(1)}$  i przyjąć  $H_0^{(2)}$ ,
- $\psi_2 = (1, 0)$ : odrzucić  $H_0^{(1)}$  i przyjąć  $H_0^{(2)}$ ,
- $\psi_3 = (1, 1)$ : odrzucić  $H_0^{(1)}$  i odrzucić  $H_0^{(2)}$ .

W kolejnych sekcjach przedstawiamy procedury testowania umożliwiające kontrolę współczynników zdefiniowanych w Sekcjach 3.4-3.6. Obszary odrzucenia określone przez te procedury przedstawiamy w terminach p-wartości (patrz Rysunki 4.5-4.7).

Jeżeli hipoteza zerowa jest prawdziwa, to odpowiadająca jej p-wartość ma rozkład jednostajny na odcinku  $[0, 1]$ . Jeżeli obie hipotezy zerowe są prawdziwe, a statystyki testowe są niezależne, to p-wartości mają rozkład jednostajny na kwadracie  $[0, 1] \times [0, 1]$ .

| Stan faktyczny             | Decyzja           |                   |                   | $\Sigma$ |
|----------------------------|-------------------|-------------------|-------------------|----------|
|                            | $\psi_1 = (0, 0)$ | $\psi_2 = (1, 0)$ | $\psi_3 = (1, 1)$ |          |
| $H^{(1)} = 0, H^{(2)} = 0$ | $P_{00}^{00}$     | $P_{00}^{10}$     | $P_{00}^{11}$     | 1        |
| $H^{(1)} = 1, H^{(2)} = 0$ | $P_{10}^{00}$     | $P_{10}^{10}$     | $P_{10}^{11}$     | 1        |
| $H^{(1)} = 1, H^{(2)} = 1$ | $P_{11}^{00}$     | $P_{11}^{10}$     | $P_{11}^{11}$     | 1        |

Tabela 4.5: Oznaczenia prawdopodobieństw podjęcia decyzji  $\psi$  dla ustalonego stanu hipotez.



Symbolem  $P_{ij}^{kl}$  oznaczamy prawdopodobieństwo podjęcia decyzji  $\psi = (k, l)$ , gdy stan hipotez to  $H^{(1)} = i, H^{(2)} = j$ , dla  $i, j, k, l \in \{0, 1\}$  (patrz Tabela (4.5)).

Procedura testowania gwarantuje kontrolę współczynnika FWER na poziomie  $\alpha$ , jeżeli

$$\begin{aligned} P_{00}^{10} + P_{00}^{11} &\leq \alpha, \\ P_{10}^{11} &\leq \alpha. \end{aligned}$$

Dwie przedstawione poniżej procedury kontrolują współczynnik FWER na poziomie  $\alpha$ .

1. Procedura jednokrokowa. Jeżeli  $p^{(2)} < \alpha/2$ , to odrzucamy obie hipotezy zerowe. Jeżeli  $p^{(1)} < \alpha/2$ , to odrzucamy hipotezę  $H_0^{(1)}$ . W przeciwnym wypadku przyjmujemy obie hipotezy zerowe. Obszary odrzucenia są przedstawione na Rysunku 4.5a.
2. Procedura step-down. Jeżeli  $p^{(2)} < \alpha/2$  lub  $\max(p^{(1)}, p^{(2)}) < \alpha$ , to odrzucamy obie hipotezy zerowe. Jeżeli  $p^{(1)} < \alpha/2$ , to odrzucamy  $H_0^{(1)}$ . W przeciwnym wypadku przyjmujemy obie hipotezy zerowe. Obszary odrzucenia są przedstawione na Rysunku 4.5b.

Procedura testowania gwarantuje kontrolę współczynnika PFER na poziomie  $\alpha$ , jeżeli

$$\begin{aligned} P_{00}^{10} + 2 * P_{00}^{11} &\leq \alpha, \\ P_{10}^{11} &\leq \alpha. \end{aligned}$$

Dwie przedstawione poniżej procedury gwarantują kontrolę współczynnika PFER na poziomie  $\alpha$ .

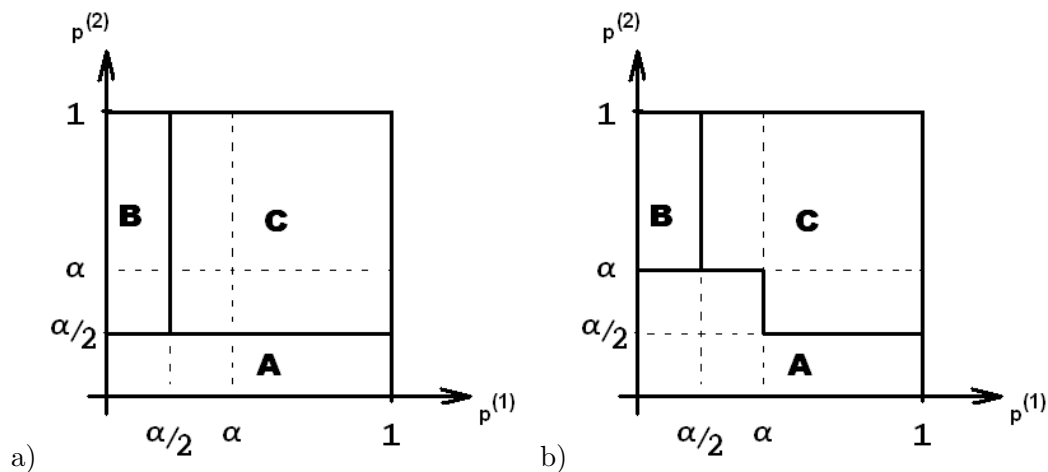
1. Procedura wstępująca. Jeżeli  $p^{(2)} < \alpha/4$ , to odrzucamy obie hipotezy zerowe. Jeżeli  $p^{(1)} < \alpha/2$ , to odrzucamy hipotezę  $H_0^{(1)}$ . W przeciwnym wypadku przyjmujemy obie hipotezy zerowe. Obszary odrzucenia dla tej procedury zostały przedstawione na Rysunku 4.6a.
2. Procedura zstępująca. Jeżeli  $p^{(1)} > \alpha/(1 + \alpha)$ , to przyjmujemy obie hipotezy zerowe. W przeciwnym wypadku, jeżeli  $p^{(2)} > \alpha/(1 + \alpha)$ , to przyjmujemy tylko hipotezę  $H_0^{(2)}$ , a jeżeli  $\max(p^{(1)}, p^{(2)}) \leq \alpha/(1 + \alpha)$ , to odrzucamy obie hipotezy zerowe. Obszary odrzucenia dla tej procedury zostały przedstawione na Rysunku 4.6b.

Procedura testowania gwarantuje kontrolę współczynnika FDR na poziomie  $\alpha$ , jeżeli

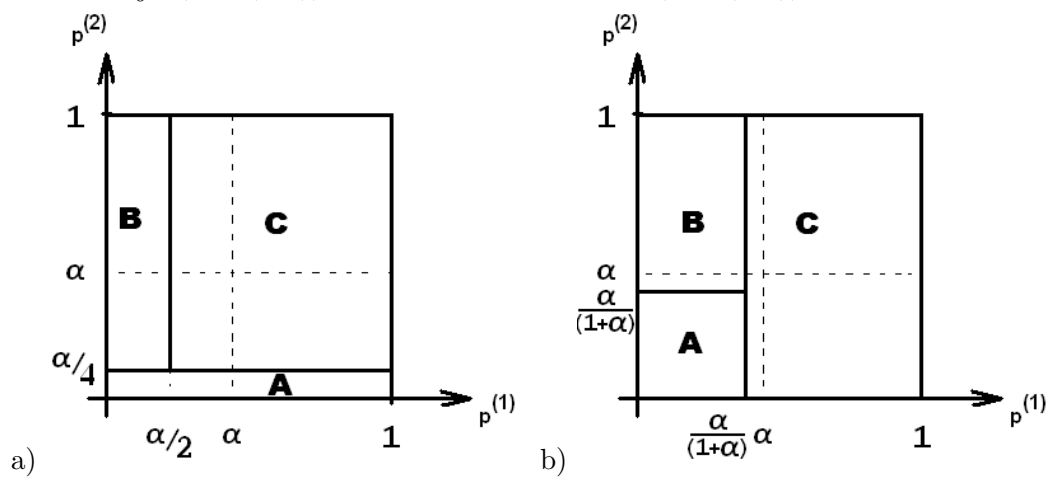
$$\begin{aligned} P_{00}^{10} + P_{00}^{11} &\leq \alpha, \\ \frac{1}{2}P_{10}^{11} &\leq \alpha. \end{aligned}$$

Dwie poniżej przedstawione procedury gwarantują kontrolę współczynnika FDR na poziomie  $\alpha$ .

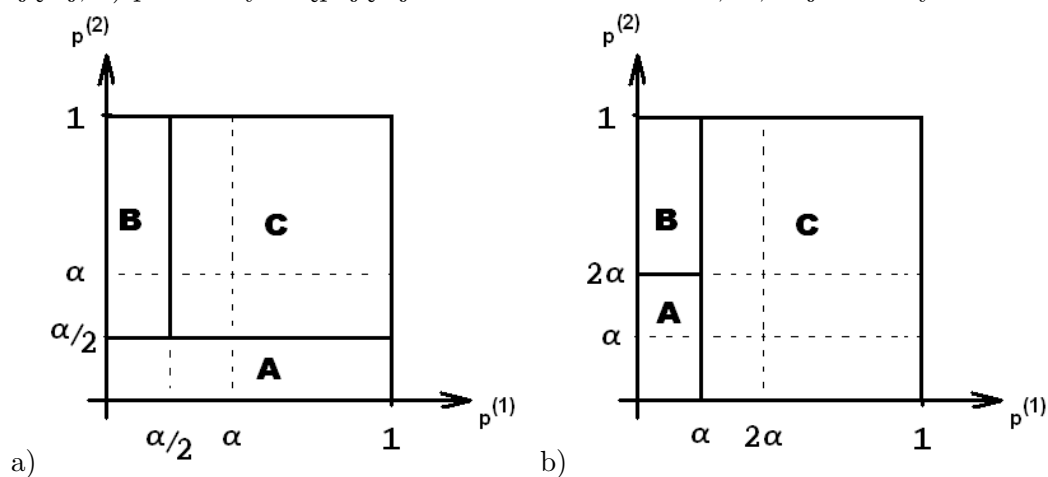
1. Procedura wstępująca. Jeżeli  $p^{(2)} < \alpha/2$ , to odrzucamy obie hipotezy zerowe. Jeżeli  $p^{(1)} < \alpha/2$ , to odrzucamy hipotezę  $H_0^{(1)}$ . W przeciwnym wypadku przyjmujemy obie hipotezy zerowe. Obszary odrzucenia dla tej procedury zostały przedstawione na Rysunku 4.7a.
2. Procedura zstępująca. Jeżeli  $p^{(1)} > \alpha$ , to przyjmujemy obie hipotezy zerowe. W przeciwnym wypadku, jeżeli  $p^{(2)} > 2\alpha$ , to przyjmujemy tylko hipotezę  $H_0^{(2)}$ , a jeżeli  $p^{(2)} \leq 2\alpha$ , to odrzucamy obie hipotezy zerowe. Obszary odrzucenia dla tej procedury zostały przedstawione na Rysunku 4.7b.



Rysunek 4.5: Obszary odrzucenia procedur kontrolujących współczynnik FWER: a) procedury jednokrokowej, b) procedury step-down. *A* to obszar odrzucenia obu hipotez ( $\psi = (1, 1)$ ), *B* odrzucenia  $H_0^{(1)}$  ( $\psi = (1, 0)$ ), *C* przyjęcia obu hipotez ( $\psi = (0, 0)$ ).



Rysunek 4.6: Obszary odrzucenia procedur kontrolujących współczynnik PFER: a) procedury wstępującej, b) procedury zstępującej. Określenie obszarów *A*, *B*, *C* jak na Rysunku 4.5.



Rysunek 4.7: Obszary odrzucenia procedur kontrolujących współczynnik FDR: a) procedury wstępującej, b) procedury zstępującej. Określenie obszarów *A*, *B*, *C* jak na Rysunku 4.5.

## 4.4 Procedury kontroli współczynnika PFER

Wybór poziomów istotności  $\alpha^{(i)}, i \in I$  zależy od liczebności cięciw relacji  $\mathcal{R}$  w punkcie  $i$ .

### Definicja 13 Cięciwa relacji.

Cięciwą relacji  $\mathcal{R}$  nazywamy każdy podzbiór  $J \subset \{1..m\}$ , którego żadne dwa elementy nie są ze sobą w relacji  $\mathcal{R}$ .

Symbolem  $\phi(i)$  oznaczamy liczebność największej cięciwy relacji  $\mathcal{R}$  zawierającej punkt  $i$ . Wyraża się ona wzorem

$$\phi(i) = \max\{\#J : i \in J, \forall l \in J R(i, l) = 0\}.$$

Cięciwami relacji  $\mathcal{R}_1$  z Rysunku 4.1 są między innymi podzbiory  $\{2, 4, 6, 7\}$ ,  $\{2, 4, 5\}$ ,  $\{2\}$ ,  $\{2, 6, 7\}$ , a więc  $\phi(2) = 4$ . Cięciwami relacji  $\mathcal{R}_2$  z Rysunku 4.1 są między innymi podzbiory  $\{1, 2\}$ ,  $\{2\}$ , a więc  $\phi(2) = 2$ .

### Zstępująca procedura testowania

Zstępująca procedura testowania odrzuca w jednym kroku co najwyżej jedną hipotezę zerową. Nie ma zatem możliwości odrzucenia wielu prawdziwych hipotez zerowych w jednym kroku, tak jak to może się zdarzyć w przypadku procedury wstępującej.

**Twierdzenie 6** *Jeżeli statystyki testowe są niezależne, to zstępująca procedura testowania z poziomami istotności wyrażonymi wzorem*

$$\alpha^{(i)} = \frac{\alpha}{(1 + \alpha)\phi(i)},$$

kontroluje współczynnik PFER na poziomie  $\alpha$ .

### Dowód

Niech  $I_t = \{i : H^{(i)} = 0\}$  będzie zbiorem indeksów prawdziwych hipotez zerowych, a  $I_f = I \setminus I_t$  zbiorem indeksów fałszywych hipotez zerowych. Dla  $i \in I_f$  mamy  $V^{(i)} = 0$ , więc

$$PFER = E(V) = E\left(\sum_{i \in \{1..m\}} V^{(i)}\right) = E\left(\sum_{i \in I_t} V^{(i)}\right) = PFER(I_t), \quad (4.1)$$

gdzie  $PFER(I_t)$  oznacza wartość współczynnika PFER dla hipotez ze zbioru  $I_t$ .

Niech relacja  $R_t : I_t \times I_t \rightarrow \{0, 1\}$  będzie obcięciem relacji  $R$  do podzbioru  $I_t$ , a więc  $R(i, j) = R_t(i, j)$  dla  $i, j \in I_t$ . Rozważmy testowanie hipotez zerowych o indeksach ze zbioru  $I_t$ . Symbolem  $V_t^{(i)}$  oznaczamy odpowiednik zmiennej  $V^{(i)}$  w procesie testowania hipotez o indeksach ze zbioru  $I_t$ .

Ponieważ

$$\begin{aligned} E(V^{(i)} | H^{(i)} = 0) &= Pr(\psi(i) = 1 \wedge \forall_{j \in \{1..m\}: R(j, i) = 1} \psi(j) = 1 | H^{(i)} = 0), \\ E(V_t^{(i)} | H^{(i)} = 0) &= Pr(\psi(i) = 1 \wedge \forall_{j \in I_t: R_t(j, i) = 1} \psi(j) = 1 | H^{(i)} = 0), \end{aligned}$$

to

$$E(V^{(i)}) \leq E(V_t^{(i)}).$$

Ze wzoru (4.1) otrzymujemy

$$E(V) \leq \sum_{i \in I_t} E(V_t^{(i)}) = PFER_t(I_t).$$

Przez  $L_1$  oznaczamy podzbiór  $I_t$  odpowiadający hipotezom, z których żadne dwie nie są ze sobą w relacji (jest to podzbiór hipotez maksymalnych w zbiorze  $L_1$ )

$$L_1 = \{i \in I_t : \forall j \in I_t R_t(i, j) = 0\}.$$

Niech  $n_1 = \#L_1$ . Ponieważ  $\forall i, j \in L_1 R(i, j) = 0$ , więc dla każdego  $i \in L_1$  zachodzi nierówność

$$\phi(i) \geq n_1.$$

Zatem, dla  $i \in L_1$

$$\alpha^{(i)} = \alpha / (\phi(i) * (1 + \alpha)) \leq \alpha / (n_1 * (1 + \alpha))$$

oraz

$$PFER_t(L_1) = E\left(\sum_{i \in L_1} V_t^{(i)}\right) = \sum_{i \in L_1} \alpha^{(i)} \leq \alpha / (1 + \alpha). \quad (4.2)$$

W procesie testowania może zdarzyć się jedna z dwóch następujących sytuacji:

- wszystkie hipotezy zerowe o indeksach ze zbioru  $L_1$  zostaną przyjęte;
- przynajmniej jedna z hipotez zerowych o indeksach ze zbioru  $L_1$  zostanie odrzucona.

Pierwsza sytuacja kończy proces testowania. Zgodnie z procedurą zstępującą przyjęte zostaną wszystkie hipotezy o indeksach ze zbioru  $I_t$ . Druga sytuacja zdarza się z prawdopodobieństwem nie większym niż  $\alpha / (1 + \alpha)$ , ponieważ

$$Pr(\exists i \in L_1 V_t^{(i)} = 1) = Pr\left(\sum_{i \in L_1} V_t^{(i)} \geq 1\right) \leq E\left(\sum_{i \in L_1} V_t^{(i)}\right) \leq \alpha / (1 + \alpha).$$

Zauważmy, że

$$E\left(\sum_{i \in I_t \setminus L_1} V_t^{(i)} \mid \sum_{j \in L_1} V_t^{(j)} > 1\right) \leq E\left(\sum_{i \in I_t \setminus L_1} V_t^{(i)} \mid \forall j \in L_1 V_t^{(j)} = 1\right). \quad (4.3)$$

Oznaczmy symbolem  $PFER_t(I_t \setminus L_1 | L_1)$  oczekiwaną wartość współczynnika PFER, gdy testujemy zbiór hipotez zerowych o indeksach ze zbioru  $I_t \setminus L_1$  przy założeniu, że hipotez z  $L_1$  zostały odrzucone. Po wykonaniu testów dla hipotez ze zbioru  $L_1$  mamy

$$PFER_t(I_t \setminus L_1 | L_1) = E\left(\sum_{i \in I_t \setminus L_1} V_t^{(i)} \mid \forall j \in L_1 V_t^{(j)} = 1\right).$$

Zatem, jeżeli zstępująca procedura testowania gwarantuje, że

$$PFER_t(I_t \setminus L_1 | L_1) \leq \alpha,$$

to

$$\begin{aligned}
PFER &= E(\sum_{i \in I_t} V^{(i)}) \leq E(\sum_{i \in I_t} V_t^{(i)}) = E(\sum_{i \in L_1} V_t^{(i)}) + E(\sum_{i \in I_t \setminus L_1} V_t^{(i)}) \\
&\leq \alpha/(1 + \alpha) + Pr(\sum_{i \in L_1} V_t^{(i)} \geq 1) E(\sum_{i \in I_t \setminus L_1} V_t^{(i)} | \sum_{j \in L_1} V_t^{(j)} \geq 1) \\
&\leq \alpha/(1 + \alpha) + \alpha/(1 + \alpha) PFER_t(I_t \setminus L_1 | L_1) \\
&\leq \alpha/(1 + \alpha) + \alpha^2/(1 + \alpha) = \alpha.
\end{aligned}$$

Pokazaliśmy, że jeżeli kontrolujemy PFER dla zbioru  $I_t \setminus L_1$ , to kontrolujemy PFER również dla zbioru  $I_t$ . Zauważmy, że dla zbioru pustego współczynnik  $PFER(\emptyset) = 0$ , a więc jest kontrolowany. A zatem, z zależności rekurencyjnej, jest on również kontrolowany dla każdego zbioru  $I_t$ .

■

W procedurze wstępującej błąd popełniony podczas błędnego przyjęcia lub odrzucenia hipotezy  $H_0^{(i)}$  zależy od stopnia oddziaływania tej hipotezy.

**Definicja 14** *Stopień oddziaływania hipotezy  $H_0^{(i)}$  w relacji  $\mathcal{R}$ . Stopień oddziaływania hipotezy  $H_0^{(i)}$  w relacji  $\mathcal{R}$  oznaczamy symbolem  $\rho(i)$  i określamy jako 1 + liczba hipotez zerowych w relacji do  $H_0^{(i)}$ . Wyraża się on wzorem*

$$\rho(i) = 1 + \#\{j : R(j, i) = 1\}.$$

Stopień oddziaływania hipotezy  $H_0^{(7)}$  w relacji  $\mathcal{R}_1$  przedstawionej na Rysunku 4.1 wynosi 4, a w relacji  $\mathcal{R}_2$  wynosi 7.

## Wstępująca procedura testowania

W procedurze wstępującej odrzucenie hipotezy prawdziwej, do której wiele innych jest w relacji, powoduje znaczne zwiększenie całkowitej liczby błędnych odrzuceń. Dlatego też hipotezy, do których wiele innych jest w relacji, powinny być testowane na niższym poziomie istotności.

**Twierdzenie 7** *Wstępująca procedura testowania z poziomami istotności wyrażonymi wzorem*

$$\alpha^{(i)} = \frac{\alpha}{m * \rho(i)},$$

*kontroluje współczynnik PFER na poziomie  $\alpha$ .*

**Dowód** Dowód tego twierdzenia opiera się na addytywności wartości oczekiwanej i nie wymaga założenia o niezależności statystyk testowych.

$$\begin{aligned}
PFER &= E(\sum_{i \in I} V^{(i)}) = \sum_{i \in I} E(V^{(i)}) = \sum_{i \in I} Pr(V^{(i)} = 1) = \\
&= \sum_{i \in I} Pr(\max_{j: R(i, j) = 1} (\psi^{(j)}(x)) H^{(j)} = 1) \leq \\
&\leq \sum_{i \in I} \sum_{j: R(i, j) = 1} Pr(\psi^{(j)}(x) H^{(j)} = 1) \leq \\
&\leq \sum_{i \in I} \alpha/m = \alpha.
\end{aligned}$$

■

## Relacja liniowa

Jeżeli  $\mathcal{R}_{lin}$  jest relacją liniową na zbiorze  $m$  hipotez (patrz Definicja 11), to kontrolę współczynnika PFER na poziomie  $\alpha$  zapewnia:

- wstępująca procedura testowania z parametrami  $\alpha^{(i)} = \alpha/(m * i)$ ,
- zstępująca procedura testowania z parametrami  $\alpha^{(i)} = \alpha/(1 + \alpha)$ .

## 4.5 Procedury kontroli współczynnika FWER

W Rozdziale 3.5 przedstawiliśmy różne procedury testowania kontrolujące współczynnik FWER. W tej sekcji pokażemy jak transformować wyniki tych procedur do zgodnych z relacją  $\mathcal{R}$ .

### Domknięcie procedury testowania

Wyniki każdej procedury testowania można przekształcić do wyników zgodnych z zadaną relacją  $\mathcal{R}$ . Funkcję przekształcającą wyniki testowania na zgodne z relacją  $\mathcal{R}$  nazywamy domknięciem procedury testowania.

Poniżej prezentujemy dwa przykładowe domknięcia. Takich domknięć można zaproponować wiele, ale z praktycznego punktu widzenia interesujące jest wyłącznie domknięcie w dół.

#### Definicja 15 *Domknięcie w dół.*

Domknięciem w dół decyzji  $\psi = (\psi^{(i)} : i \in \mathcal{I})$  nazywamy decyzję  $\check{\psi} = (\check{\psi}^{(i)} : i \in \mathcal{I})$ , gdzie

$$\check{\psi}^{(i)} = \max\{\psi^{(i)}, \max_j\{\psi^{(j)} : R(i, j) = 1\}\}.$$

Domknięcie w dół decyzji  $\psi$  polega na odrzuceniu również tych hipotez zerowych  $H_0^{(i)}$ , dla których przynajmniej jedna hipoteza, do której  $H_0^{(i)}$  jest w relacji, została odrzucona.

#### Definicja 16 *Domknięcie w górę.*

Domknięciem w górę decyzji  $\psi = (\psi^{(i)} : i \in \mathcal{I})$  nazywamy decyzję  $\hat{\psi} = \{\hat{\psi}^{(i)} : i \in \mathcal{I}\}$ , gdzie

$$\hat{\psi}^{(i)} = \min\{\psi^{(i)}, \min_j\{\psi^{(j)} : R(j, i) = 1\}\}.$$

Domknięcie w górę decyzji  $\psi$  polega na przyjęciu również tych hipotez zerowych, których hipotezy nadrzędne zostały przyjęte. W dalszej części pracy nie będziemy rozważać tego domknięcia, ponieważ prowadzi ono do bardzo niewielu odrzuceń. Przykład domknięcia w górę i w dół decyzji dla relacji  $\mathcal{R}_1$  został przedstawiony w Tabeli 4.6.

|                                       | $\psi^{(1)}$ | $\psi^{(2)}$ | $\psi^{(3)}$ | $\psi^{(4)}$ | $\psi^{(5)}$ | $\psi^{(6)}$ | $\psi^{(7)}$ |
|---------------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| decyzja $\psi^{(i)}$                  | 1            | 1            | 0            | 1            | 0            | 1            | 0            |
| domknięcie w dół $\check{\psi}^{(i)}$ | 1            | 1            | 1            | 1            | 1            | 1            | 0            |
| domknięcie w górę $\hat{\psi}^{(i)}$  | 1            | 1            | 0            | 0            | 0            | 0            | 0            |

Tabela 4.6: Przykład domknięcia w górę i w dół przykładowej decyzji dla Relacji  $\mathcal{R}_1$ .

#### Definicja 17 *Domknięcie w dół (w górę) procedury testowania.*

Domknięcie w dół (w górę) procedury testowania, to procedura, której decyzją jest domknięcie w dół (w górę) decyzji procedury wyjściowej.

Domknięcie procedury testowania pozwala na uzyskanie wyników zgodnych z zadaną relacją.

**Twierdzenie 8** *Domknięcie (zarówno w dół jak i w górę) procedury testowania kontrolującej współczynnik FWER na poziomie  $\alpha$  gwarantuje zgodność wyników testowania z zadaną relacją hierarchiczną  $\mathcal{R}$  oraz kontrolę współczynnika FWER na poziomie  $\alpha$ .*

Twierdzenie to przedstawili [Marcus i inni 1976]. Wynika ono ze spostrzeżenia, iż domknięcie w dół decyzji  $\psi$  powoduje odrzucenie prawdziwych hipotez jedynie wtedy, gdy decyzja  $\psi$  odrzuca choć jedną prawdziwą hipotezę. Więc

$$FWER = Pr(V \geq 1) = Pr\left(\sum_i \psi^{(i)}(x)H^{(i)} \geq 1\right) = Pr\left(\sum_i \psi^{(i)}(x)H^{(i)} \geq 1\right) \leq \alpha.$$

Z twierdzenia 8 wynika, że do kontroli współczynnika FWER dla zbioru hipotez o zadanej strukturze hierarchicznej, można wykorzystać domknięcie w dół procedury step-up Hochberga (patrz Twierdzenie 4) lub procedury step-down Holma (patrz Twierdzenie 3).

Domknięcie w dół procedury testowania zwiększa współczynnik PFER i FDR. Tym samym, procedura kontrolująca współczynnik PFER lub FDR, po domknięciu nie zapewnia już kontroli na tym samym poziomie. Dlatego też potrzebne są nowe procedury kontrolujące te współczynniki.

## 4.6 Procedury kontroli współczynnika FDR

Najpopularniejszą procedurą kontroli współczynnika FDR w zagadnieniu testowania zbioru hipotez jest procedura step-up Benjaminiego Hochberga (patrz Twierdzenie 5). Podobnie jak w przypadku współczynnika PFER, domknięcie w dół tej procedury nie gwarantuje kontroli współczynnika FDR, choć nie jest to tak oczywiste jak dla współczynnika PFER.

**Przykład, gdy domknięcie w dół procedury step-up Benjaminiego Hochberga, nie kontroluje współczynnika FDR.**

Rozważmy zbiór 100 hipotez zerowych z zadaną relacją liniową. Niech  $H^{(1)} = 1$  oraz  $H^{(i)} = 0$  dla  $2 \leq i \leq 100$ . Przyjmijmy, że p-wartość  $p^{(1)}$  odpowiadająca hipotezie  $H_0^{(1)}$  jest tak mała, że fałszywa hipoteza  $H_0^{(1)}$  jest zawsze odrzucana. Z uwagi na jednostajny rozkład p-wartości przy prawdziwości hipotezy zerowej, z niezerowym prawdopodobieństwem któraś prawdziwa hipoteza zerowa może mieć mniejszą p-wartość niż fałszywa hipoteza zerowa. Możemy jednak dobrać hipotezę alternatywną tak, że to prawdopodobieństwo będzie dowolnie małe.

W przedostatnim kroku procedury step-up p-wartość  $p^{(2:100)} = \min\{p^{(i)} : 2 \leq i \leq 100\}$  jest porównywana z  $\alpha^{(2:100)} = \frac{2}{100}\alpha$ . Prawdopodobieństwo, że najmniejsza z 99 p-wartości, odpowiadających prawdziwym hipotezom, każda o rozkładzie jednostajnym, jest mniejsza od  $\frac{2}{100}\alpha$  wynosi

$$Pr(p^{(2:100)} \leq \frac{2}{100}\alpha) = 1 - \left(1 - \frac{2}{100}\alpha\right)^{99}.$$

Jeżeli  $p^{(2:100)} \leq \alpha^{(2:100)}$ , to odrzucana jest hipoteza  $H_0^{(1)}$  i jedna z pozostałych 99 hipotez zerowych, które oznaczmy symbolem  $H_0^{(j)}$ . W tym wypadku domknięcie w dół procedury step-up odrzuca wszystkie hipotezy  $H_0^{(i)}$  takie, że  $i \leq j$  (z czego  $j - 1$  to prawdziwe hipotezy zerowe). Współczynnik FDR można więc oszacować z dołu

$$FDR \geq \left(1 - \left(1 - \frac{2}{100}\alpha\right)^{99}\right) \frac{1}{99} \sum_{i=1}^{99} \frac{i}{i+1} \approx \frac{2 * 99}{100} \alpha \frac{1}{99} (99 - \ln(99)) > \alpha,$$

co oznacza, że domknięcie w dół procedury step-up nie gwarantuje kontroli współczynnika FDR.

## Kontrola współczynnika FDR dla relacji liniowej

**Twierdzenie 9** *Zstępująca procedura testowania z poziomami istotności wyrażonymi wzorami*

$$\begin{aligned}
 \alpha_{lin}^{(1)} &= \alpha, \\
 \alpha_{lin}^{(i)} &= \min \left( \alpha_{max}, \alpha \left[ \sum_{k=i}^{m-1} \frac{k-i+1}{k} (1 - \alpha_{lin}^{(k+1)}) \prod_{l=i+1}^k \alpha_{lin}^{(l)} + \frac{m-i+1}{m} \prod_{l=i+1}^m \alpha_{lin}^{(l)} \right]^{-1} \right), \quad \text{dla } 1 < i < m, \\
 \alpha_{lin}^{(m)} &= \min(\alpha_{max}, m\alpha),
 \end{aligned} \tag{4.4}$$

kontroluje współczynnik FDR na poziomie  $\alpha$  dla relacji liniowej.

Parametr  $\alpha_{max}$  jest dowolną liczbą z przedziału  $(0, 1]$ . Prezentowane wyniki symulacyjne są wyznaczone dla  $\alpha_{max} = 0.5$ . Przedstawiona procedura gwarantuje kontrolę współczynnika FDR dla każdej wartości  $\alpha_{max}$ , jednak wybór  $\alpha_{max}$  powinien zależeć od spodziewanej liczby fałszywych hipotez zerowych oraz rozkładu statystyki testowej dla prawdziwej hipotezy alternatywnej. Generalnie, im mniej spodziewamy się fałszywych hipotez zerowych, tym parametr  $\alpha_{max}$  powinien być mniejszy. Odpowiednio dobierając ten parametr możemy uzyskać wyższą liczbę poprawnie odrzuconych hipotez zerowych. Problem doboru tego współczynnika nie jest rozważany w tej pracy.

### Dowód

Niech  $m_1$  oznacza liczbę fałszywych hipotez zerowych.

Dla  $m_1 = 0$  zachodzi  $FDR \leq \alpha^{(1)}$ , gdyż  $Q = 0$ , jeżeli nie odrzucimy żadnej hipotezy zerowej i  $Q = 1$ , jeżeli odrzucimy przynajmniej jedną hipotezę zerową. Odrzucenie przynajmniej jednej hipotezy zerowej wymaga odrzucenia hipotezy  $H_0^{(1)}$  (będącej w relacji do wszystkich pozostałych) na poziomie istotności  $\alpha^{(1)}$ .

Dla  $m_1 > 0$  w procedurze zstępującej dla liniowej relacji odrzucenie prawdziwej hipotezy zerowej jest możliwe po odrzuceniu wszystkich fałszywych hipotez zerowych. Więc

$$\begin{aligned}
 FDR(\mathcal{H}) &= E(Q|R > 0)Pr(R > 0) \leq E(Q|R \geq m_1) = \\
 &FDR(\mathcal{H}_t | \forall_{i \in I} \psi(i) \geq H^{(i)}),
 \end{aligned}$$

gdzie  $FDR(\mathcal{H})$  oznacza współczynnik FDR wyznaczony dla zbioru hipotez  $\mathcal{H}$  a  $FDR(\mathcal{H}_t | \forall_{i \in I} \psi(i) \geq H^{(i)})$  oznacza współczynnik FDR wyznaczony dla zbioru prawdziwych hipotez zerowych, przy założeniu, że wszystkie fałszywe zostały odrzucone.

Wartość oczekiwana  $E(Q|R \geq m_1)$  jest nie większa od

$$\begin{aligned}
 E(Q|R \geq m_1) &= \sum_{i=m_1}^m \frac{i-m_1}{i} Pr(R = i | R \geq m_1) \leq \\
 &\leq \left( \sum_{i=m_1+1}^{m-1} \frac{i-m_1}{i} (1 - \alpha_{lin}^{(i+1)}) \prod_{j=m_1+1}^i \alpha_{lin}^{(j)} \right) + \frac{m-m_1}{m} \prod_{j=m_1+1}^m \alpha_{lin}^{(j)} = \\
 &= \alpha_{lin}^{(m_1+1)} \left[ \left( \sum_{i=m_1+1}^{m-1} \frac{i-m_1}{i} (1 - \alpha_{lin}^{(i+1)}) \prod_{j=m_1+2}^i \alpha_{lin}^{(j)} \right) + \frac{m-m_1}{m} \prod_{j=m_1+2}^m \alpha_{lin}^{(j)} \right].
 \end{aligned}$$

Podstawiając parametry (4.4) z Twierdzenia 9 otrzymujemy

$$\alpha \geq E(Q|R \geq m_1) \geq E(Q|R > 0) \geq FDR.$$

■



## Kontrola współczynnika FDR dla hierarchicznej relacji zależności

Twierdzenie 9 można uogólnić na dowolną relację hierarchii  $\mathcal{R}$  odpowiednio modyfikując poziomy istotności w procedurze zstępującej.

### Definicja 18 *Stopień oddziaływania zbioru hipotez.*

Niech  $K \subset I$  będzie podzbiorem indeksów zbioru hipotez. Stopień oddziaływania zbioru hipotez  $\{H_0^{(i)} : i \in K\}$ , oznaczamy symbolem  $\rho(K)$  i wyrażamy wzorem

$$\rho(K) = 1 + \#\{j : R(j, i) = 1, i \in K\}.$$

**Twierdzenie 10** Niech  $\mathcal{K}(i)$  będzie zbiorem cięciw dla relacji  $\mathcal{R}$  zawierających indeks  $i$ . Zstępująca procedura testowania z poziomami istotności wyrażonymi wzorem

$$\alpha^{(i)} = \min_{K \in \mathcal{K}(i)} \{\alpha_{lin}^{(\rho(K))} / \#K\}, \quad (4.5)$$

gdzie  $\alpha_{lin}^{(\rho(K))}$  wyrażają się wzorem (4.4), kontroluje współczynnik FDR na poziomie  $\alpha$ .

W załączniku został umieszczony program w języku R wyznaczający parametry  $\alpha^{(i)}$ .

### Dowód

Niech  $I_f$  oznacza zbiór indeksów fałszywych hipotez,  $I_f^R$  oznacza podzbiór zbioru  $I_f$  z indeksami odrzuconych fałszywych hipotez. Niech  $L_1$  będzie podzbiorem maksymalnych hipotez ze zbioru  $I_t$  (tzn. takich, że żadne dwie z tego podzbioru nie są ze sobą w relacji).

Ponieważ poziom istotności dla testowania hipotez ze zbioru  $L_1$  jest mniejszy lub równy  $\alpha_{lin}^{(\rho(L_1))} / \#L_1$ , to z prawdopodobieństwem nie mniejszym od  $1 - \alpha_{lin}^{(\rho(L_1))}$  testowanie kończy się i wszystkie hipotezy ze zbioru  $L_1$  są przyjmowane. Z prawdopodobieństwem  $\alpha_{lin}^{(\rho(L_1))}$  przynajmniej jedna hipoteza jest odrzucona (zauważmy, że  $\rho(L_1) \leq \#I_f^R$ ).

Symbolem  $L_2$  oznaczmy podzbiór hipotez zerowych, dla których wszystkie hipotezy nadrzędne zostały odrzucone. Zachodzi nierówność  $\rho(L_1) + 1 \geq \rho(L_2)$ , a więc prawdopodobieństwo odrzucenia hipotezy ze zbioru  $L_2$  spełnia nierówność

$$Pr(\exists_{i \in L_2} \psi(i) = 1 | \exists_{j \in L_1} \psi(j) = 1) \leq \alpha_{lin}^{(\rho(L_2))} \leq \alpha_{lin}^{(\rho(L_1)+1)}.$$

Iterując tę procedurę otrzymujemy

$$Pr(\exists_{i \in L_{k+1}} \psi(i) = 1 | \exists_{j \in L_k} \psi(j) = 1) \leq \alpha_{lin}^{(\rho(L_{k+1}))} \leq \alpha_{lin}^{(\rho(L_k)+1)} \leq \alpha_{lin}^{(\rho(L_1)+k)} \leq \alpha_{lin}^{(\#I_f^R+k)}.$$

Poziomy istotności  $\alpha_{lin}^{(\#I_f^R+k)}$  wyrażone wzorem (4.4) kontrolują współczynnik FDR na poziomie  $\alpha$  dla liniowej relacji, więc współczynniki z Twierdzenia 10 pozwalają na kontrolę FDR na poziomie  $\alpha$  dla dowolnej relacji hierarchicznej. ■

### Uwaga

Jeżeli liczebność zbioru hipotez jest duża, to wyznaczenie wszystkich cięciw zawierających punkt  $i$  może być bardzo czasochłonne. W takim przypadku zamiast parametrów wyrażonych wzorem (4.5) można stosować parametry  $\alpha'^{(i)}$

$$\alpha'^{(i)} = \alpha_{lin}^{(\rho(i))} / \max_{i \in I} \phi(i). \quad (4.6)$$

Parametry  $\alpha'^{(i)}$  gwarantują kontrolę współczynnika FDR, ponieważ  $\alpha'^{(i)} \leq \alpha^{(i)}$ . Jednak ich stosowanie prowadzi do mniejszej średniej liczby odrzuconych hipotez. Różnica ta jest tym większa, im większe jest zróżnicowanie współczynników  $\phi(i)$  dla danej relacji.

## Rozdział 5

# Zastosowanie uzyskanych wyników w badaniach genetycznych

„Statistics is the servant to all sciences”  
Jerzy Sława-Neyman.

### 5.1 Wprowadzenie

W tym rozdziale przedstawiamy zastosowanie wyników opisanych w poprzednich rozdziałach w zagadnieniu identyfikacji aktywnych procesów biologicznych (z wykorzystaniem danych mikromacierzowych) oraz w zagadnieniu predykcji funkcji genu (z wykorzystaniem danych o interakcjach pomiędzy białkami). Na początku wprowadzimy kilka terminów genetycznych, które będą wykorzystywane w kolejnych sekcjach.

Genetyka to nauka o dziedziczności i zmienności organizmów wynikającej z informacji zawartej w genach. Każdy żywy organizm ma ustalony zestaw genów.

**Definicja 19** *Gen* [gr. *génos* ‘ród’, ‘pochodzenie’, ‘gatunek’], to odcinek łańcucha kwasu deoksyrybonukleinowego (DNA), w którym kolejność ułożenia nukleotydów stanowi informację genetyczną o zdolności organizmu do syntezy określonych białek.

Pomimo, iż wszystkie komórki w żywym organizmie mają ten sam zestaw genów, komórki te pełnią różne funkcje. Dzieje się tak, ponieważ pewne geny są aktywne we wszystkich komórkach (są to geny niezbędne do życia komórki), inne są aktywowane tylko w niektórych komórkach. Aby opisać aktywność genu w komórce korzysta się z terminu „ekspresja genu”.

**Definicja 20** *Ekspresja genu* to miara ilości genu w komórce, odpowiada aktywności tego genu. Ekspresję można mierzyć stężeniem pikomolarnym lub średnią liczbą transkryptów.

Istnieje wiele metod pozwalających na pomiar ekspresji zbioru genów jednocześnie. Obecnie najpopularniejsze są badania z wykorzystaniem techniki mikromacierzowej, które umożliwiają pomiar ekspresji wielu tysięcy genów jednocześnie w jednym eksperymencie mikromacierzowym. Mikromacierz U-133A firmy Affymetrix mierzy ekspresję 22 tys. ludzkich genów jednocześnie.

Analizując fizjologię komórki nie wystarczy badać ekspresji pojedynczych genów ponieważ współdziałają one ze sobą. Dla genetyka, bardziej interesujące od badania aktywności poszczególnych genów jest badanie aktywności procesów biologicznych.

**Definicja 21** *Proces biologiczny*, to proces występujący w żywych organizmach. Jest regulowany przez geny, a w jego wyniku mogą powstawać, być naprawiane lub ulegać degradacji różne związki i struktury.

Przykładami procesów biologicznych są: „transkrypcja DNA”, „denaturalizacja białek”, „transport cząsteczek energetycznych”, „transport cząsteczek ATP”. Procesy te mogą być określone na różnym poziomie szczegółowości.

## Gene Ontology [GO]

W 1998 roku powstało Konsorcjum Gene Ontology ([Harris i inni 2004] [www.geneontology.org](http://www.geneontology.org)). Organizacja ta postawiła sobie za zadanie uporządkowanie istniejącej wiedzy o procesach biologicznych, opracowała i aktualizuje trzy ontologie opisujące różne aspekty funkcjonowania genów:

- procesy biologiczne (ang. biological process);
- komponenty komórkowe (ang. cellular component);
- funkcje molekularne (ang. molecular function).

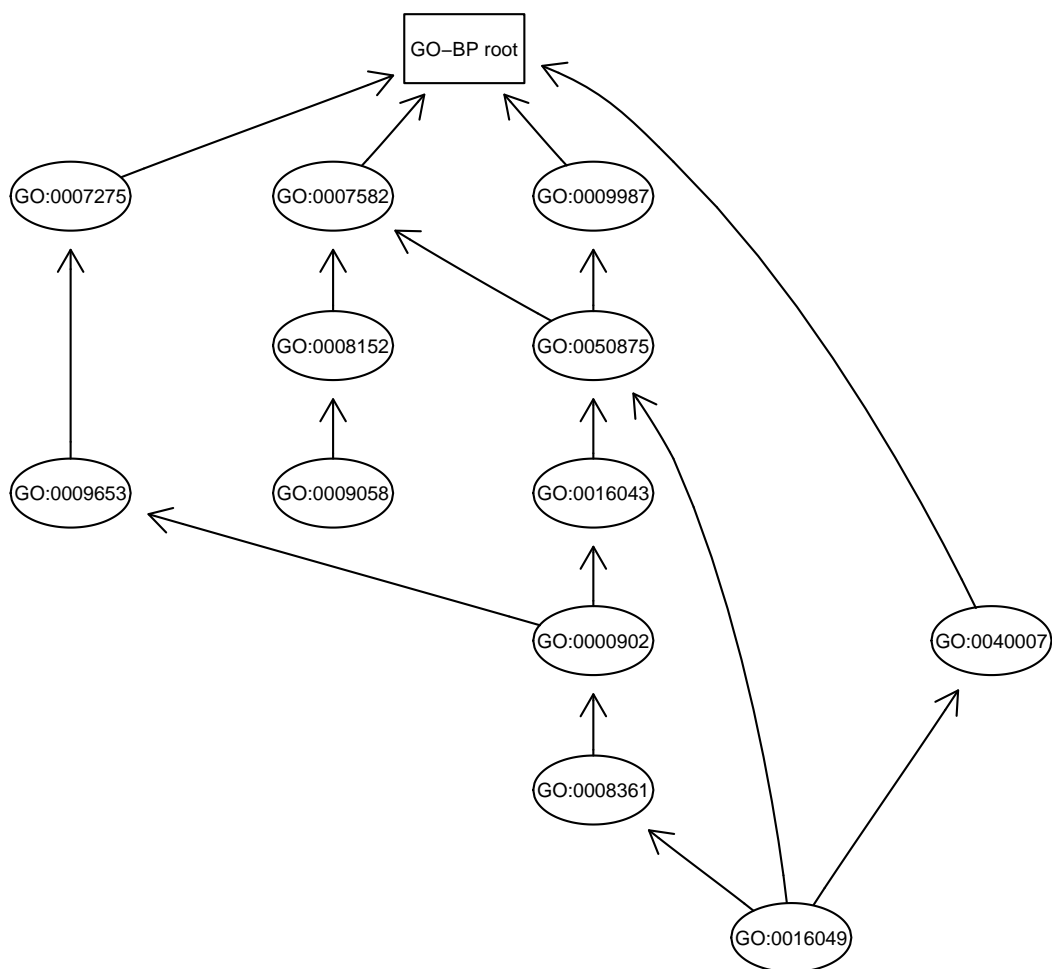
Każda ontologia opisana jest przez acykliczny graf skierowany. Wyniki przedstawione w tym rozdziale dotyczą ontologii „procesy biologiczne”. Wierzchołki odpowiadające tej ontologii grafu reprezentują procesy biologiczne, a krawędzie odpowiadają relacjom pomiędzy poszczególnymi procesami. Ontologia procesów biologicznych, to nieustannie aktualizowany graf, na dzień dzisiejszy o 18834 węzłach. Fragment tego grafu przedstawiony jest na Rysunku 5.1. Zamiast nazw procesów stosuje się etykiety w postaci „GO : xxxxxx”, gdzie xxxxxx to identyfikator procesu biologicznego.

Wyróżnia się dwa rodzaje relacji pomiędzy procesami biologicznymi. Relacje „is-a” oraz relacje „is-a-part-of”. Relacja „is-a” pomiędzy procesami, oznacza, że jeden proces jest przykładem drugiego, bardziej ogólnego, procesu. Np. proces „rozkład cukrów” jest specyficzną wersją procesu „metabolizm”, jest też ogólniejszy niż proces „rozkład glukozy”. Relacja „is-part-of” pomiędzy procesami, oznacza, że jeden proces jest częścią składową drugiego procesu. Np. proces „translacja”, składa się z wielu podprocesów, między innymi inicjującego podprocesu „przyłączenie kompleksu inicjacji translacji” oraz kończącego podprocesu „zatrzymanie translacji”.

W prowadzonych analizach najczęściej zaniedbuje się różnice pomiędzy relacjami „is a” i „is a part of”. Suma obu relacji jest relacją antyzwrotną, antysymetryczną i przechodnią, jest więc relacją hierarchii w sensie Definicji 10 (z Rozdziału 4).

## Baza adnotacji

Informacje o tym, który gen uczestniczy w danym procesie znaleźć można w bazach adnotacji genów. Obecnie takie bazy, nazywane również bazami funkcji genów, są intensywnie rozwijane. Najpopularniejszą bazą funkcji jest Entrez Gene (<http://www.ncbi.nlm.nih.gov/entrez/>). Zgromadzono w niej dane o funkcjach ponad 2.075.132 genów występujących w 3.572 organizmach (informacje z końca roku 2006). Wpisy w tej bazie danych określają w których procesach biologicznych udział poszczególnych genów jest potwierdzony. Informacje te pochodzą z publikacji, w których autorzy pracowali nad identyfikacją pojedynczych genów w procesach biologicznych. Baza ta jest nieustannie uaktualniana.



Rysunek 5.1: Fragment grafu reprezentującego ontologie „procesy biologiczne”. Węzeł oznaczony etykietką  $GO:0009058$  reprezentuje proces „biosynteza”, węzeł o etykiecie  $GO:0016049$  reprezentuje proces „wzrost komórki”, a o etykiecie  $GO:0007582$  reprezentuje „procesy fizjologiczne”. Krawędzie grafu określają hierarchię pomiędzy procesami.

W dalszej części tego rozdziału będziemy korzystać ze sformułowania „gen (białko) pełni funkcję  $f$ ”, co oznacza, że dany gen (białko) uczestniczy w procesie biologicznym  $f$ .

## 5.2 Zagadnienie identyfikacji aktywnych procesów biologicznych

Celem identyfikacji aktywnych procesów biologicznych jest wskazanie aktywnych procesów na podstawie mikromacierzowych pomiarów ekspresji genów.

### Metoda

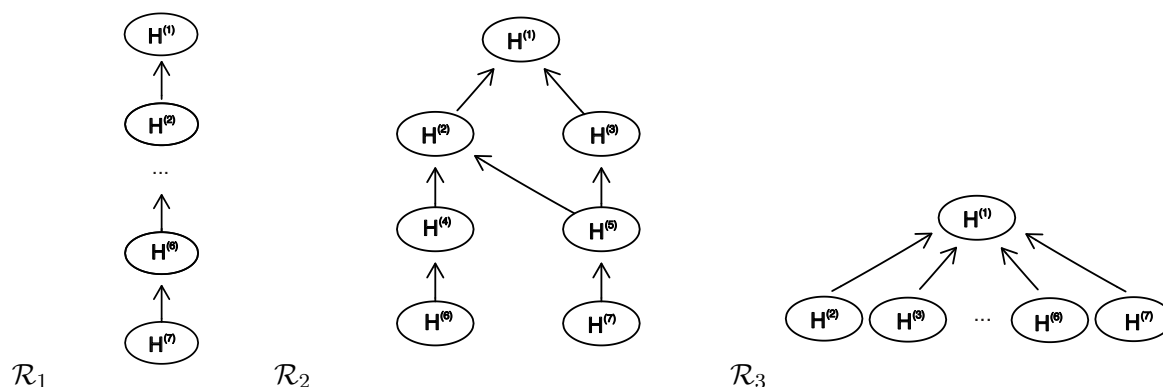
Przyjmujemy, że jeżeli pewien proces biologiczny jest aktywny, to ekspresja genów uczestniczących w tym procesie jest wyższa niż średnia ekspresja wszystkich genów.

Określamy zbiór badanych procesów biologicznych  $I$ . Dla każdego z nich stawiamy hipotezę zerową  $H_0^{(i)}$ ,  $i \in I$ , że ekspresja genów uczestniczących w tym procesie jest równa 0, przeciw hipotezie alternatywnej, że ekspresja genów uczestniczących w tym procesie jest większa od zera.

Do testowania każdej z tych hipotez wykorzystywany jest test  $t$ -Studenta lub jego nieparametryczny odpowiednik, test Wilcoxon.

### Studium symulacyjne

Podczas studium symulacyjnego będziemy rozważać 7-elementowe zbiory procesów biologicznych. Ze zbiorem procesów skojarzony jest zbiór hipotez zerowych, z których każda dotyczy pewnego procesu biologicznego. Dla każdego zbioru hipotez zerowych będziemy rozważali relacje  $\mathcal{R}_1$ ,  $\mathcal{R}_2$  i  $\mathcal{R}_3$ , przedstawione na Rysunku 5.2. Relacje te różnią się współczynnikami  $\phi(i)$  i  $\rho(i)$  dla poszczególnych hipotez (patrz Tabela 5.1).



Rysunek 5.2: Grafowa reprezentacja relacji hierarchicznych  $\mathcal{R}_1$ ,  $\mathcal{R}_2$  i  $\mathcal{R}_3$ .

W Tabeli 5.1 przedstawiamy poziomy istotności wyznaczone dla różnych procedur dla  $\alpha = 0.05$ . Dolny indeks przy poziomach istotności określa, dla której procedury testowania są one wyliczone ( $FU$  dla procedury wstępującej oraz  $FD$  dla procedury zstępującej). Górny indeks wskazuje, który współczynnik jest kontrolowany. Warto zauważyć, że poziomy istotności  $\alpha_{FD}^{FDR}(i)$  (a więc kontrolujące współczynnik  $FDR$  dla procedury zstępującej) są większe od  $\alpha$ , a pomimo to przyjęta procedura testowania gwarantuje kontrolę współczynnika  $FDR$  na poziomie  $\alpha$ .

W prezentowanym studium symulacyjnym losowaliśmy wartości statystyki testowej z rozkładu  $t$ -Studenta z 9 stopniami swobody. Niech  $\mu = (\mu_1, \dots, \mu_7)$ , gdzie  $\mu_i$  oznacza średnią wartość statystyki testowej dla  $i$ -tej hipotezy. A zatem  $\mu_i = 0$ , gdy hipoteza  $H_0^{(i)}$  jest prawdziwa i  $\mu_i > 0$  w przeciwnym wypadku.

|     | Relacja $\mathcal{R}_1$ |           |                         |                         |                        | Relacja $\mathcal{R}_2$ |           |                         |                         |                        |
|-----|-------------------------|-----------|-------------------------|-------------------------|------------------------|-------------------------|-----------|-------------------------|-------------------------|------------------------|
|     | $\rho(i)$               | $\phi(i)$ | $\alpha_{FU}^{PFER}(i)$ | $\alpha_{FD}^{PFER}(i)$ | $\alpha_{FD}^{FDR}(i)$ | $\rho(i)$               | $\phi(i)$ | $\alpha_{FU}^{PFER}(i)$ | $\alpha_{FD}^{PFER}(i)$ | $\alpha_{FD}^{FDR}(i)$ |
| i=1 | 1                       | 1         | 0.0071                  | 0.0476                  | 0.0500                 | 1                       | 1         | 0.0071                  | 0.0476                  | 0.0500                 |
| i=2 | 2                       | 1         | 0.0036                  | 0.0476                  | 0.0952                 | 2                       | 2         | 0.0036                  | 0.0238                  | 0.0476                 |
| i=3 | 3                       | 1         | 0.0024                  | 0.0476                  | 0.1364                 | 2                       | 2         | 0.0036                  | 0.0238                  | 0.0476                 |
| i=4 | 4                       | 1         | 0.0018                  | 0.0476                  | 0.1739                 | 3                       | 2         | 0.0024                  | 0.0238                  | 0.0682                 |
| i=5 | 5                       | 1         | 0.0014                  | 0.0476                  | 0.2083                 | 4                       | 2         | 0.0018                  | 0.0238                  | 0.0870                 |
| i=6 | 6                       | 1         | 0.0012                  | 0.0476                  | 0.2400                 | 4                       | 2         | 0.0018                  | 0.0238                  | 0.1042                 |
| i=7 | 7                       | 1         | 0.0010                  | 0.0476                  | 0.3500                 | 5                       | 2         | 0.0014                  | 0.0238                  | 0.1042                 |

Tabela 5.1: Współczynniki  $\rho(i)$ ,  $\phi(i)$  oraz poziomy istotności  $\alpha$  dla relacji  $\mathcal{R}_1$  i  $\mathcal{R}_2$ .

W symulacjach rozważaliśmy następujące przypadki:

- $\mu^{(1)} = (0, 0, 0, 0, 0, 0, 0)$ , żaden proces nie jest aktywny,
- $\mu^{(2)} = (2, 2, 2, 2, 0, 0, 0)$ , aktywne są pierwsze cztery procesy,
- $\mu^{(3)} = (3, 2.75, 2.5, 2.25, 0, 0, 0)$ , aktywne są pierwsze cztery procesy, ale w różnym stopniu,
- $\mu^{(4)} = (2, 2, 2, 2, 2, 2, 2)$ , wszystkie procesy są jednakowo aktywne,
- $\mu^{(5)} = (3, 2.75, 2.5, 2.25, 2, 1.75, 1.5)$ , wszystkie procesy są aktywne, ale w różnym stopniu,
- $\mu^{(6)} = (3, 0, 0, 0, 0, 0, 0)$ , aktywny jest tylko pierwszy proces.

Procedurę losowania wartości statystyki testowej oraz proces testowania powtórzyliśmy  $10^6$  razy. Dla takiej liczby powtórzeń wariancja oceny współczynnika błędu jest pomijalnie mała. W Tabeli 5.2 prezentujemy uśrednione wyniki z symulacji. Potwierdzają one, że zaproponowane procedury kontrolują rozważane współczynniki na poziomie  $\alpha$ . W kolumnie oznaczonej  $E(S)$  przedstawiono średnią liczbę odrzuconych fałszywych hipotez zerowych, dla każdego rozważanego scenariusza. W każdym scenariuszu empiryczne współczynniki błędów są niższe od postulowanego poziomu  $\alpha = 0.05$ . Gdy frakcja prawdziwych hipotez zerowych jest niewielka, empiryczny współczynnik błędu jest znacznie mniejszy od  $\alpha$ . W większości scenariuszy największą oczekiwaną liczbę odrzuconych hipotez uzyskuje się stosując procedurę zstępującą z współczynnikami  $\alpha_{FD}^{FDR}$ . Ta procedura prowadzi do najlepszych wyników w sytuacji, gdy dla najbardziej specjalistycznych hipotez moc testowania jest najniższa (co jest częstą sytuacją w tym zagadnieniu). W pewnych scenariuszach, szczególnie dla relacji typu  $\mathcal{R}_3$  najlepsze wyniki uzyskuje się dla procedury zstępującej. Wybór procedury (pomiędzy wstępującą a zstępującą) powinien być uzależniony od struktury relacji oraz od różnicy w średnich mocach pomiędzy hipotezami specjalistycznymi a ogólnymi. Gdy hipotezy specjalistyczne mają średnio niższą moc, lepsze wyniki otrzymuje się korzystając z procedury zstępującej.

| Relacja         | Stan hipotez | Procedura  |  |  |  |  |  |
|-----------------|--------------|--|--|--|--|--|--|
|                 |              | wstępująca<br>z $\alpha_{FU}^{PFER}(i)$<br>z Tw. (7) | zstępująca<br>z $\alpha_{FD}^{PFER}(i)$<br>z Tw. (6) | zstępująca<br>z $\alpha_{FD}^{FDR}(i)$<br>z Tw. (10) | wstępująca<br>z $\alpha_{FU}^{PFER}(i)$<br>z Tw. (7) | zstępująca<br>z $\alpha_{FD}^{PFER}(i)$<br>z Tw. (6) | zstępująca<br>z $\alpha_{FD}^{FDR}(i)$<br>z Tw. (10) |
|                 |              | Współczynnik błędu                                   |  |  | E(S)   |  |  |
|                 |              | PFER   | PFER   | FDR  |  |  |  |
| $\mathcal{R}_1$ | $\mu^{(1)}$  | 0.050  | 0.050  | 0.050  | —  | —  | —  |
| $\mathcal{R}_1$ | $\mu^{(2)}$  | 0.007  | 0.008  | 0.017  | 1.712  | 1.433  | 1.852  |
| $\mathcal{R}_1$ | $\mu^{(3)}$  | 0.007  | 0.022  | 0.035  | 2.503  | 2.761  | 3.236  |
| $\mathcal{R}_1$ | $\mu^{(4)}$  | —  | —  | —  | 3.384  | 1.634  | 2.670  |
| $\mathcal{R}_1$ | $\mu^{(5)}$  | —  | —  | —  | 3.425  | 3.258  | 4.842  |
| $\mathcal{R}_1$ | $\mu^{(6)}$  | 0.031  | 0.045  | 0.045  | 0.712  | 0.909  | 0.912  |
| $\mathcal{R}_2$ | $\mu^{(1)}$  | 0.050  | 0.050  | 0.050  | —  | —  | —  |
| $\mathcal{R}_2$ | $\mu^{(2)}$  | 0.006  | 0.008  | 0.010  | 1.552  | 1.432  | 1.722  |
| $\mathcal{R}_2$ | $\mu^{(3)}$  | 0.006  | 0.022  | 0.023  | 2.314  | 2.679  | 3.034  |
| $\mathcal{R}_2$ | $\mu^{(4)}$  | —  | —  | —  | 2.975  | 1.638  | 2.265  |
| $\mathcal{R}_2$ | $\mu^{(5)}$  | —  | —  | —  | 3.092  | 3.184  | 4.186  |
| $\mathcal{R}_2$ | $\mu^{(6)}$  | 0.028  | 0.044  | 0.043  | 0.712  | 0.909  | 0.912  |
| $\mathcal{R}_3$ | $\mu^{(1)}$  | 0.050  | 0.050  | 0.050  | —  | —  | —  |
| $\mathcal{R}_3$ | $\mu^{(2)}$  | 0.011  | 0.015  | 0.006  | 1.448  | 1.272  | 1.303  |
| $\mathcal{R}_3$ | $\mu^{(3)}$  | 0.011  | 0.022  | 0.007  | 2.225  | 2.365  | 2.394  |
| $\mathcal{R}_3$ | $\mu^{(4)}$  | —  | —  | —  | 2.345  | 1.915  | 1.967  |
| $\mathcal{R}_3$ | $\mu^{(5)}$  | —  | —  | —  | 2.784  | 3.069  | 3.116  |
| $\mathcal{R}_3$ | $\mu^{(6)}$  | 0.021  | 0.043  | 0.022  | 0.714  | 0.909  | 0.912  |

Tabela 5.2: W kolumnach 3, 4 i 5 są przedstawione empirycznie wyznaczone współczynniki PFER i FDR dla trzech prezentowanych procedur testowania. W kolumnach 6, 7 i 8 są przedstawione średnie liczby odrzuconych fałszywych hipotez zerowych.

| Relacja         | Wektor $\mu$ | Domknięcie                             |   |
|-----------------|--------------|--|---|
|                 |              | procedury single-step<br>Bonferroniego | procedury step-up<br>Benjaminiego-Hochberga |
|                 |              | Współczynnik PFER                      | Współczynnik FDR                            |
| $\mathcal{R}_1$ | $\mu^{(1)}$  | 0.201                                  | 0.050                                       |
| $\mathcal{R}_1$ | $\mu^{(2)}$  | 0.073                                  | 0.046                                       |
| $\mathcal{R}_1$ | $\mu^{(6)}$  | 0.151                                  | 0.061                                       |
| $\mathcal{R}_2$ | $\mu^{(1)}$  | 0.151                                  | 0.050                                       |
| $\mathcal{R}_2$ | $\mu^{(2)}$  | 0.045                                  | 0.035                                       |
| $\mathcal{R}_2$ | $\mu^{(6)}$  | 0.102                                  | 0.056                                       |
| $\mathcal{R}_3$ | $\mu^{(1)}$  | 0.093                                  | 0.050                                       |
| $\mathcal{R}_3$ | $\mu^{(2)}$  | 0.294                                  | 0.028                                       |
| $\mathcal{R}_3$ | $\mu^{(6)}$  | 0.044                                  | 0.042                                       |

Tabela 5.3: Wyniki dla domknięcia procedury jednokrokowej Bonferroniego oraz procedury step-up Benjaminiego-Hochberga.

W Tabeli 5.3 prezentujemy wyniki domknięcia w dół procedur testowania „single step” i „step-up” dla zagadnienia testowania zbioru hipotez. Potwierdzają one, że domknięcie procedury kontrolującej współczynnik FDR lub PFER już nie kontroluje tych współczynników. Uzasadnia to potrzebę opracowania nowych procedur testowania.

### 5.3 Predykcja funkcji białek

W tej sekcji przedstawiamy przykład zastosowania procedur wprowadzonych w Rozdziale 4, w zagadnieniu predykcji funkcji białek dla danych rzeczywistych. Przedstawione poniżej zagadnienie dotyczy prognozy funkcji biologicznych dla pojedynczych białek. Dla każdego z rozważanych białek oraz dla każdej z rozważanych funkcji należy zdecydować, czy są przesłanki, że dane białko pełni daną funkcję. Wykonanie takiej wstępnej prognozy pozwala genetykom określić, które białka warto poddać dalszym badaniom laboratoryjnym.

Do wykonania takiej prognozy stosuje się wiele różnych metod wnioskowania, np. sieci bayesowskie [Troyanskaya i inni 2003], uczenie maszynowe [Vinayagam 2004], analizę czynnikową [Kustra i inni 2006] oraz wiele innych. Poniżej przedstawiona jest metoda związana z badaniem częstości wystąpienia interesującej funkcji w zbiorze białek współdziałających z interesującym nas białkiem.

#### Dane

Do poniższych analiz został wykorzystany zbiór danych opisany w pracy [Wu i inni 2004]. Zawiera on dane o 3224 białkach oraz 1829 procesach biologicznych. Symbolem  $I_g = \{1 \dots 3224\}$  oznaczamy zbiór indeksów białek, symbolem  $I_f = \{1 \dots 1829\}$  zbiór indeksów procesów biologicznych. Przez  $\mathcal{G} = \{g_i : i \in I_g\}$  oznaczamy zbiór wszystkich białek, gdzie  $g_i$  to białko o indeksie  $i$ . Przez  $\mathcal{F} = \{f_j : j \in I_f\}$  oznaczamy zbiór wszystkich procesów biologicznych, gdzie  $f_j$  to proces biologiczny o indeksie  $j$ .

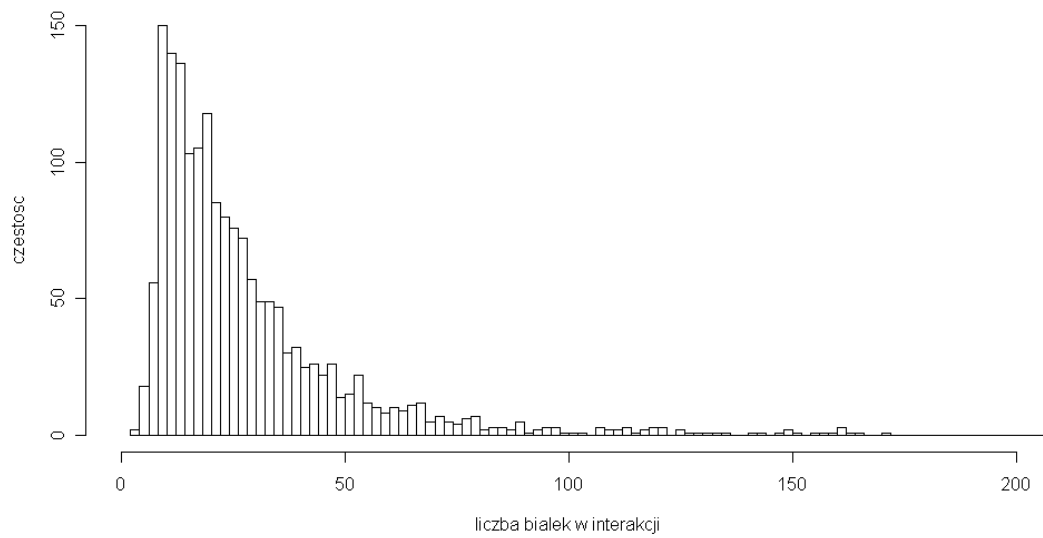
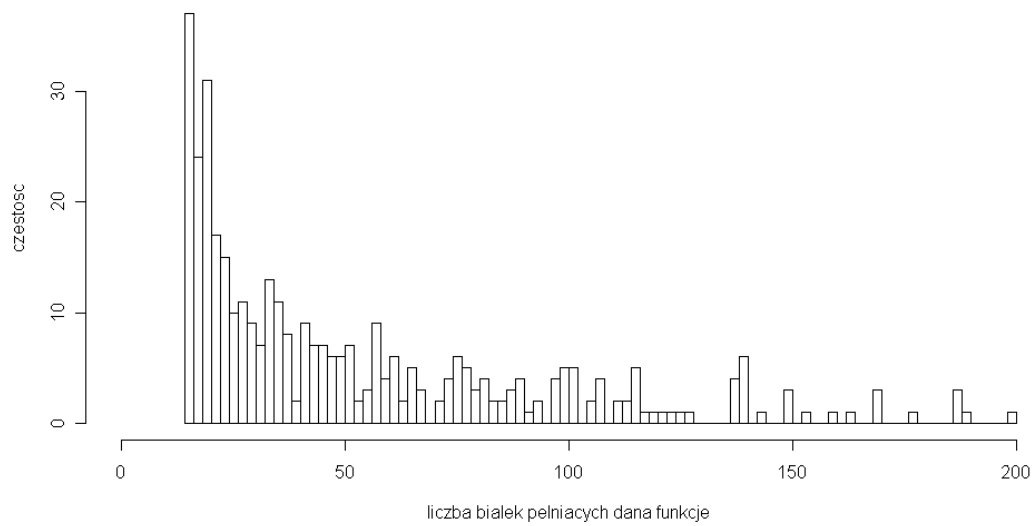
Informacje o białkach przedstawione są w postaci macierzy interakcji PPI (ang. protein-protein interaction), o wymiarach  $3224 \times 3224$ . Symbolem  $PPI(i, j)$  oznaczamy komórkę o współrzędnych  $i, j \in I_g$  tej macierzy. Jeżeli  $PPI(i, j) = 1$ , to w bazie danych znaleziono informacje o tym, że białko  $i$ -te współdziała z białkiem  $j$ -tym, a jeżeli  $PPI(i, j) = 0$ , to takiej informacji nie znaleziono (nie oznacza to, że białka te nie współdziałają ze sobą, ponieważ najczęściej w publikacjach nie są odnotowane, tzw. „przykłady negatywne”).

Informacje o odnotowanych funkcjach poszczególnych białek przedstawione są w postaci macierzy adnotacji o wymiarach  $3224 \times 1829$ . Symbolem  $Adnot(i, j)$  będziemy oznaczali komórkę o współrzędnych  $(i, j)$  tej macierzy, gdzie  $i \in I_g, j \in I_f$ . Jeżeli  $Adnot(i, j) = 1$ , to w bazie danych znaleziono informacje o tym, że białko  $i$ -te pełni funkcję  $j$ -tą, a jeżeli  $Adnot(i, j) = 0$ , to takiej informacji nie znaleziono (nie oznacza to, że dane białko tej funkcji nie pełni, ponieważ nie wszystkie funkcje białek są jeszcze poznane). Na zbiorze funkcji określona jest relacja hierarchii określona przez strukturę Gene Ontology.

Dla analizowanego zbioru danych wykazano  $\sum_{i \in I_g} \sum_{j \in I_g} PPI(i, j) = 61020$  par interakcji (opisanych w bazie BioGrid) oraz wykazano  $\sum_{i \in I_g} \sum_{j \in I_f} Adnot(i, j) = 78841$  udokumentowanych przypadków pełnienia danej funkcji przez dane białko (opisanych w bazie danych Entrez).

Na Rysunku 5.3 przedstawione są histogramy opisujące liczby białek pełniących poszczególne funkcje oraz liczby białek w interakcji z którymi jest badane białko.





Rysunek 5.3: Statystyki opisowe dla danych [Wu i inni 2004].

## Metoda

Zakłada się, że jeżeli badane białko współdziała z grupą białek pełniących pewną funkcję, to najprawdopodobniej również pełni daną funkcję (z aktualnego stanu wiedzy wynika, że około 70 – 80% współdziałających białek pełni również takie same funkcje). Przyjmując to założenie, predykcje funkcji białek można wykonać stosując następujący algorytm:

- Określamy podzbiór białek  $G \subset \mathcal{G}$ , dla których chcemy przeprowadzić prognozę;
- Dla każdego białka  $g_i \in G$  określamy zbiór białek będących w interakcji z białkiem  $g_i$ , oznaczmy ten zbiór przez  $G_i \in \mathcal{G}$ ;
- Określamy podzbiór procesów biologicznych  $F \subset \mathcal{F}$ , dla których chcemy przeprowadzić prognozę;
- Dla każdego białka  $g_i \in G$  i każdego procesu biologicznego  $f_j \in F$  weryfikujemy hipotezę  $H_0^{(i,j)} : \rho(f_j|G_i) = \rho(f_j|\mathcal{G})$ , przeciwko  $H_A^{(i,j)} : \rho(f_j|G_i) > \rho(f_j|\mathcal{G})$ , gdzie  $\rho(f_j|G_i)$  oznacza częstość występowania procesu biologicznego  $f_j$  w zbiorze białek  $G_i$ . Jeżeli test odrzuci hipotezę zerową na poziomie istotności  $\alpha^{(i,j)}$ , to uznajemy, że białko  $g_i$  może uczestniczyć w procesie biologicznym  $f_j$ , ponieważ białka będące z nim w interakcji, uczestniczą w tym procesie istotnie częściej niż białka nie będące z nim w interakcji.

Aby zweryfikować hipotezę, czy dana funkcja występuje istotnie częściej wykorzystywany jest dokładny test Fishera dla tablic kontyngencji  $2 \times 2$  (patrz np. [Xiao, Pan 2005]). Wartość p dla tego testu wyraża się wzorem

$$p^{(i,j)} = 1 - \sum_{l=0}^{k_{i,j}-1} \frac{\binom{k_j}{l} \binom{n-k_j}{n_i-l}}{\binom{n}{n_i}}, \quad (5.1)$$

gdzie:

- $n = \#G$ , to liczba wszystkich białek,
- $k_j = \#G|f_j = \sum_{i \in I_g} Adnot(i, j)$ , to liczba białek ze zbioru  $G$  uczestniczących w procesie  $f_j$ ,
- $n_i = \#G_i = \sum_{j \in I_g} PPI(i, j)$ , to liczba białek w interakcji z białkiem  $g_i$ ,
- $k_{i,j} = \#G_i|f_j = \sum_{l \in I_g} PPI(i, l) Adnot(l, j)$ , to liczba białek w zbiorze  $G_i$  uczestniczących w procesie  $f_j$ .

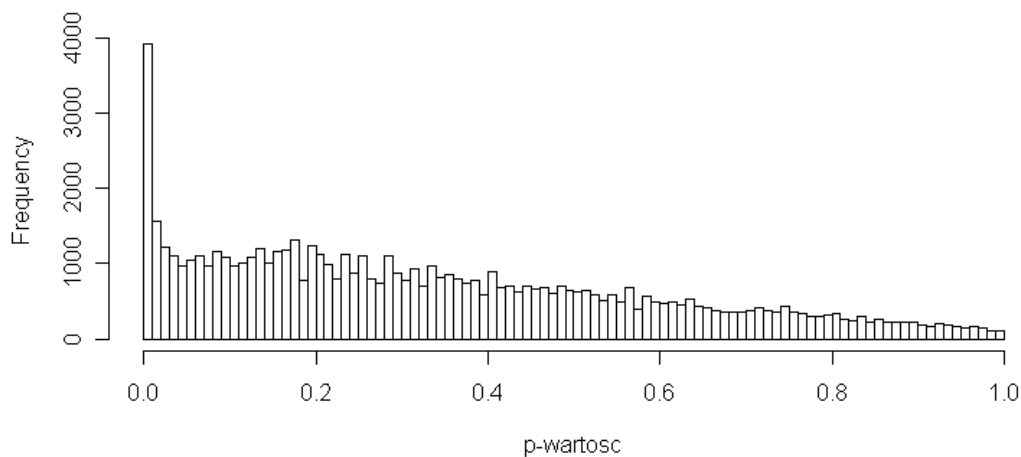
Z przyczyn praktycznych do zbioru  $F$  wybiera się procesy biologiczne pełnione przez przynajmniej 10 białek, czyli  $F = \{f_j \in \mathcal{F} : k_j \geq 10\}$ . Do zbioru  $G$  wybiera się białka, z którymi przynajmniej 10 innych jest w interakcji, czyli  $G = \{g_i \in \mathcal{G} : n_i \geq 10\}$ . Uwzględniając te wymagania, zbiór  $G$  liczy 1099 białek, a zbiór  $F$  liczy 271 funkcji.

Wartości p wyznaczone ze wzoru (5.1) nie mają rozkładu ciągłego, można więc zamiast nich stosować „fuzzy p-values” [Kulinskaya, Lewin 2006]. Wyniki testowania, uzyskane z użyciem zwykłych p-wartości oraz fuzzy p-wartości nie różnią się istotnie.

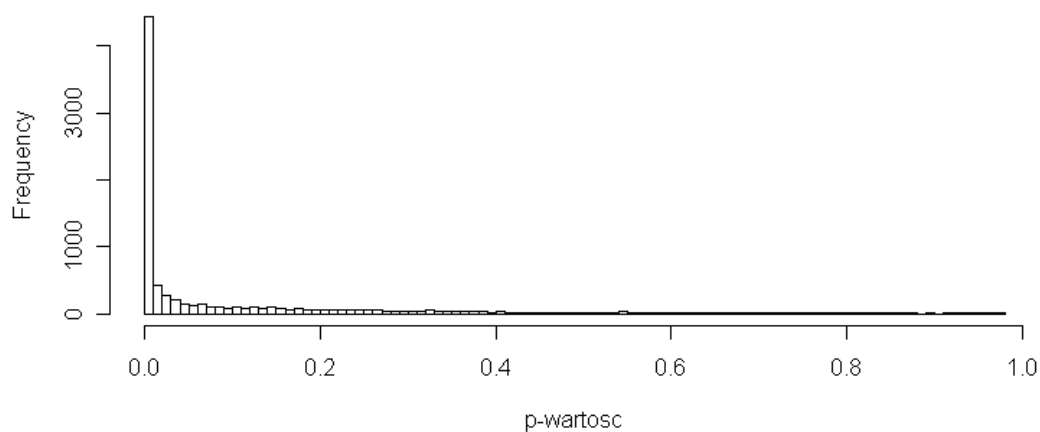
Aby ocenić współczynniki błędów wykorzystano procedurę *one leave out cross validation*. A więc, wyznaczając predykcje dla białka  $g_i$  wykorzystane są informacje o pełniących funkcjach

przez wszystkie białka za wyjątkiem białka  $g_i$ . Wyniki predykcji porównywane są ze znaną adnotacją danej funkcji dla danego białka. Współczynnik błędu wyznacza się uśredniając wyniki dla wszystkich funkcji ze zbioru  $F$  i wszystkich białek ze zbioru  $G$ .

Na Rysunku 5.4 przedstawiony jest histogram dla p-wartości wyznaczonych dla prawdziwych i fałszywych hipotez zerowych. Dla danych rzeczywistych rozkład p-wartości dla prawdziwych hipotez zerowych odbiega od rozkładu jednostajnego. Może to być spowodowane brakiem odnotowanej informacji o genach, które nie pełnią rozpatrywanej funkcji.



a)



b)

Rysunek 5.4: Histogram p-wartości dla prawdziwych hipotez zerowych (a) oraz fałszywych hipotez zerowych (b).

## Wyniki

Dla każdego z 1099 białek rozważano zbiór 271 hipotez. Każda hipoteza odpowiada przypuszczeniu, że białko  $g_i$  uczestniczy w procesie biologicznym  $f_j$ . Wszystkie zbiory hipotez zostały przetestowane czterema różnymi procedurami testowania, dwoma kontrolującymi współczynnik PFER, jedną kontrolującą współczynnik FWER i jedną kontrolującą współczynnik FDR.

Metodą *one leave out cross validation* dla zbioru hipotez dotyczących każdego białka wyznaczyliśmy empirycznie oceny zmiennych  $V$ ,  $S$  i  $R$ , oraz oceny współczynników PFER, FWER i FDR. W Tabeli 5.4 przedstawiliśmy uśrednione oceny współczynników uzyskanych dla różnych białek oraz dla różnych procedur testowania.

Otrzymane wyniki potwierdzają, że każda z zastosowanych procedur kontroluje odpowiedni współczynnik błędu na postulowanym poziomie 0.05. Najmniejszą średnią liczbę odrzuceń fałszywych hipotez zanotowano dla procedur kontroli współczynnika PFER (podobnie jak w studium symulacyjnym). Największą średnią liczbę odrzuceń fałszywych hipotez zanotowano dla procedury zstępującej kontrolującej współczynnik FDR.

Wyniki przedstawione w Tabeli 5.4 obrazują, jak istotny wpływ na wyniki testowania ma wybór procedury testowania. W tym przypadku zdecydowanie lepsze wyniki uzyskuje się dla procedury zstępującej. Może to być spowodowane dużą liczbą węzłów w relacji opisującej strukturę Gene Ontology.

| Procedura testowania                                | #odrzuconych hipotez zero-<br>wych | ocena współczyn-<br>nika błędu | #odrzuconych fałszywych hipo-<br>tez zerowych |
|---|------------------------------------|--------------------------------|---|
| Procedura zstępująca kontro-<br>lująca PFER (Tw. 6) | 1706                               | PFER = 0.0209                  | 1683  |
| Procedura wstępująca kontro-<br>lująca PFER (Tw. 7) | 1191                               | PFER = 0.0164                  | 1173  |
| Domknięcie w dół procedury<br>step-up Hochberga     | 2300                               | FWER = 0.0382                  | 2255  |
| Procedura zstępująca kontro-<br>lująca FDR (Tw. 10) | 3892                               | FDR = 0.01725                  | 3823  |

Tabela 5.4: Wyniki zastosowania proponowanych procedur w zagadnieniu predykcji funkcji białek.

## Dodatek A

# Program wyznaczający poziomy istotności $\alpha^{(i)}$

Programy w języku *R*, wyliczające poziomy istotności  $\alpha^{(i)}$  dla procedur przedstawionych w tej pracy.

```
#
# poziomy istotności dla procedury wstępującej dla współczynnika PFER
# alpha - żądany poziom kontroli współczynnika PFER
# phi   - wektor ze współczynnikami średnich

get.parametry.PFER.follow.up <- function(alpha, phi) {
  alpha/((1+alpha)*phi)
}

#
# poziomy istotności dla procedury zstępującej dla współczynnika PFER
# alpha - żądany poziom kontroli współczynnika PFER
# rho   - wektor ze współczynnikami stopni oddziaływania

get.parametry.PFER.follow.down <- function(alpha, rho) {
  alpha/(length(rho)*rho)
}

pprod <- function(parametry,m,n) {
  if (m<=n)
    return (prod(parametry[m:n]))
  return (1)
}
```

```

#
# poziomy istotności dla procedury zstępującej dla współczynnika FDR w relacji liniowej
# alpha - żądany poziom kontroli współczynnika FDR
# N      - liczba hipotez

get.parametry.FDR.lin <- function(alpha, N) {
  parametry = NULL
  parametry[N] = min(1,N * alpha)
  for (m in (N-1):2) {
    sum = 0
    for (n in m:(N-1)) {
      sum = sum + ((n-m+1)/n)*(1 - parametry[n+1])*pprod(parametry,m+1,n)
    }
    parametry[m] = min(1,alpha/(sum + ((N-m+1)/N) * prod(parametry[(m+1):N])))
  }
  parametry[1] = alpha

  parametry
}

#
# poziomy istotności dla procedury zstępującej dla współczynnika FDR
# alpha - żądany poziom kontroli współczynnika FDR
# rho    - wektor stopni oddziaływania średnich

get.parametry.FDR <- function(alpha, rho) {
  parametry.lin = get.parametry.FDR.lin(alpha, length(rho))
  parametry.lin/rho
}

```

## Dodatek B

# Badania prowadzone podczas studiów doktoranckich

Podczas studiów doktoranckich prowadziłem badania nad zagadnieniem testowania zbioru hipotez. Pierwsze wyniki, zawarte w pracy magisterskiej [Biecek 2003], dotyczą metody doboru poziomów istotności dla uogólnionej procedury step-up step-down oraz zastosowania tej metody do analiz danych mikromacierzowych. Prace nad metodami analiz danych mikromacierzowych były kontynuowane i w ich rezultacie, wspólnie z dr Anną Jasińską, opracowaliśmy metodę estymacji stężenia transkryptu genu na podstawie mikromacierzowych pomiarów jasności [Biecek, Jasińska 2004]. Mój wkład polegał na wyborze modelu zależności pomiędzy stężeniem transkryptu a jasnością luminescencji oraz estymacją współczynników w tym modelu.

Kolejne badania dotyczyły zastosowania metod testowania wielokrotnego w zagadnieniu detekcji genów wpływających na cechy ilościowe (ang. Quantitative Trait Loci, QTL). W tym zagadnieniu mamy do czynienia z wysoką korelacją pomiędzy statystykami testowymi, która prowadzi, w przypadku większości procedur testowania, do małej mocy. Podczas stażu w Purdue University pracowałem nad doбором wartości krytycznej dla kryterium mBIC, dla silnie skorelowanych statystyk testowych. Wyniki tych badań są przedstawione w pracach [Biecek i inni 2004], [Szyda i inni 2004]. Mój wkład polegał na opracowaniu procedury wyznaczającej współczynniki kary w modelu mBIC uwzględniając korelacje pomiędzy statystykami testowymi.

Ogólniejsze wyniki, dla dowolnej znanej struktury korelacji pomiędzy statystykami testowymi, zostały przedstawione w pracy [Biecek 2005]. W tej pracy wykorzystano koncepcję efektywnej liczby hipotez [Cheverud 2001] oraz zaproponowano korektę poziomu istotności kontrolującą współczynnik FWER. Zaproponowana korekta prowadzi do znacznie wyższej mocy niż korekta Cheveruda. Najistotniejsze zalety zaproponowanej korekty, to duża ogólność (kontrolę współczynnika FWER uzyskujemy dla wielu struktur korelacji statystyk testowych) oraz niższe wymagania obliczeniowe niż w przypadku procedur permutacyjnych.

Z kryterium mBIC związane są również wyniki uzyskane we współpracy z Instytutem Statystyki Uniwersytetu Wiedeńskiego [Baierl i inni 2006], dotyczące analizy wpływu niespełnienia założenia o normalności rozkładu badanej cechy na efektywność detekcji QTLi. Mój wkład w tę pracę polegał na przygotowaniu programów komputerowych implementujących estymatory odporne (m-estymatory: Hubera, Bisquare i Hampela) w kryterium mBIC oraz na wykonaniu studium symulacyjnego porównującego wyniki uzyskane z wykorzystaniem m-estymatorów, z wynikami uzyskanymi z użyciem estymatorów maksymalnej wiarygodności.

Zagadnieniu modelowania efektów QTLi poświęcona jest praca [Biecek, Klonecki 2006]. Badany jest w niej wpływ przyjętej parametryzacji efektów genetycznych na obciążenie ocen efektów statystycznych w sytuacjach, gdy rozważany jest niepełny lub niezrównoważony model. Mój wkład w tę pracę polegał na wyznaczeniu obciążenia ocen dla parametryzacji Cockerhama, parametryzacji  $F_2$  i  $F_\infty$  w modelu z dwoma QTLami i dwoma markerami oraz w modelu z jednym QTLem i  $n$  markerami.

Przygotowywana praca [Biecek 2007] poświęcona jest tematyce omówionej w tej rozprawie doktorskiej. Zaproponowano w niej inne sformułowanie zagadnienia testowania zbioru hipotez z zadaną relacją hierarchii oraz przedstawiono procedury testowania pozwalające na kontrolę najpopularniejszych współczynników błędów. Prezentowane ujęcie zagadnienia testowania zbiorczego z zadaną relacją hierarchii może być zastosowane do szerszej klasy problemów niż ujęcie prezentowane w pracy [Finner, Strassburger 2002].



## Dodatek C

# Tło historyczne zagadnienia testowania zbioru hipotez

Pojęcie hipotezy statystycznej ewoluowało przez setki lat. Pierwsze zachowane wzmianki o terminie *hipoteza* można znaleźć w pracy „Teoria Matematyki” greckiego filozofa Geminusa (pierwsze dziesięciolecie naszej ery). Termin ten był przez kolejne wieki używany w astrologii oraz fizyce. Przykładami są prace Gottfrieda Wilhelma Leibniza („Nowe hipotezy fizyczne”, 1671) oraz Isaaca Newtona („Hipotezy o świetle”, 1675).

Wzmianki o pierwszej hipotezie zweryfikowanej na gruncie analizy statystycznej dotyczą pracy medyka Johna Arbuthnota (1667 – 1735), który w roku 1710 opublikował w Royal Society pracę „An argument for Divine Providence, taken from the constant regularity observ'd in the births of both sexes”. W pracy tej przedstawił roczne liczby urodzeń chłopców oraz dziewcząt w Londynie w latach 1625-1710 oraz zauważył, że w każdym roku rodziło się więcej chłopców niż dziewcząt. Przyjmując, że częstość urodzin chłopców jest równa  $1/2$ , prawdopodobieństwo, że przez 86 lat co roku rodziło się więcej chłopców niż dziewcząt jest równe  $1/2^{86} < 10^{-24}$  czyli jest niezmiernie małe. Było to dla niego dowodem na to, że częstość urodzin chłopców jest statystycznie istotnie większa niż częstość urodzin dziewcząt.

Wnioskowanie Johna Arbuthnota zostało skrytykowane między innymi przez Nicholasa Bernoulliego, co prawdopodobnie spowodowało, że pierwszeństwo w konstrukcji pierwszego statystycznego testu często przypisuje się Pierreowi-Simonowi Laplaceowi (1749 – 1827). Przedstawił on w roku 1796 fizykalno-matematyczne uzasadnienie „nebular hypothesis” (hipotezy mgławicowej, nazywanej też hipotezą Kanta-Laplacea), opisującej genezę powstania Układu Słonecznego. Hipoteza ta zakłada, że planety powstały w wyniku stopniowego odrywania się od wirującego Słońca pierścieni gazowej materii, przekształcających się z czasem w zwarte kule. Jako uzasadnienie tej hipotezy Laplace wykazał, że ekliptyki planet nie są losowe, lecz leżą blisko jednej pierwotnej ekliptyki. Laplace był zwolennikiem subiektywnej interpretacji prawdopodobieństwa, dlatego przeprowadzone przez niego wnioskowanie było bliskie wnioskowaniu Bayesowskiemu.

Przez kolejny wiek uczeni stawiając i weryfikując hipotezy statystyczne kierowali się intuicją. Dopiero w latach dwudziestych XX wieku aksjomatyczne podstawy dla zagadnienia testowania opracowali Jerzy Sława-Neyman (matematyk polskiego pochodzenia) i Egon Pearson (syn znakomitego matematyka Karla Pearsona). Matematyczne wykształcenie Jerzego Neymana, wspólnie z intuicjami jego współpracownika Egona Pearsona, pozwoliło na spójne przedstawienie teorii testowania hipotez. W latach 1928–1933 napisali wspólnie wiele istotnych prac o procesie testo-

wania hipotez, testach statystycznych, testach najefektywniejszych, rozmiarach testu, poziomach istotności itp. Jako pierwsi do procesu testowania wprowadzili pojęcie hipotezy alternatywnej (dziś oczywiste i niekwestionowane), przez co byli długo krytykowani przez współczesne im autorytety, między innymi Ronalda Aylmera Fishera.

Jerzy Spława-Neyman podczas studiów wyższych w Charkowie poznał, wywodzącą się z Rosji, częstościową interpretację prawdopodobieństwa, której aksjomaty zostały później opisane przez Andrieja Kołomogorowa. Na częstościowej interpretacji prawdopodobieństwa oparł on swoją aksjomatyczną teorię testowania hipotez, przez co jest ona nazywana również częstościowym, lub klasycznym ujęciem testowania hipotez. Na cześć jej twórców jest również nazywana Neymanowsko-Pearsonowską teorią testowania hipotez. Takie ujęcie procesu testowania hipotez dominowało od lat 30-40 do końca XX wieku.

Metody testowania były i są rozwijane nie tylko przez matematyków, ale również przez biologów, fizyków oraz chemików. Przykładowo znaczący wkład do statystyki wniosły prace podpisywane pseudonimem „Student”, których autorem był pracownik browaru Guinness, chemik William Gosset.

W teorii testowania hipotez ważny jest wybór poziomu istotności, czyli dopuszczalnego prawdopodobieństwa odrzucenia prawdziwej hipotezy zerowej. Najczęściej przyjmowany poziom istotności to 0.05, co oznacza, że średnio błędnie odrzucimy hipotezę zerową nie częściej niż raz na 20 razy. Rozważmy jednak zbiór 100 hipotez zerowych, każda orzekająca, że pewien lek nie wpływa na stan pacjenta. Wykonajmy 100 testów, każdy na poziomie istotności  $\alpha = 0.05$ . Nawet, jeżeli żaden z rozważanych leków nie wpływa na zdrowie pacjenta, to w około 5 przypadkach test odrzuci błędnie hipotezę zerową, a przyjmie hipotezę alternatywną. Czy to oznacza, że te 5 leków istotnie wpływa na stan pacjenta? Oczywiście, że nie. Jednak ze wzrostem liczby przeprowadzonych testów na ustalonym poziomie istotności, rośnie prawdopodobieństwo błędnego odrzucenia przynajmniej jednej hipotezy zerowej.

Zwrócił na to uwagę biostatystyk Graham Martin, który w grudniu 1984 roku w *The Lancet* opublikował list zatytułowany „Munchausen Statistical Grid, that makes all trials look significant”. W liście tym Graham Martin oskarżał badaczy stosujących metody statystyczne, szczególnie w medycynie, o powtarzanie eksperymentu wielokrotnie lub też wykorzystywanie różnych testów, tak długo, aż któryś test odrzuci hipotezę zerową i „potwierdzi” słuszność ich przypuszczeń. Taka procedura nazwana została statystyczną siatką Munchausena. Nazwa pochodzi od nazwiska Barona von Munchausena, który znany był z opowiadania niesamowitych i nierzeczywistych historii (historia o tym, jak to Baron von Munchausen wyciągnął się z bagna za własne paski od butów pojawia się często w genezie nazwy metody bootstrap). Zarzuty tego typu sprawiły, że przy publikacji wyników zaczęto stawiać wymóg stosowania korekty poziomu istotności uwzględniającej liczbę weryfikowanych hipotez. Historycznie pierwszą i wciąż najpopularniejszą korektą jest korekta Bonferroniego [Bonferroni 1936]. Korekta ta była z powodzeniem wykorzystywana w przypadku, gdy testowano kilka lub kilkadziesiąt hipotez.

Od lat 80. XX wieku trwa rewolucja bioinformatyczna. Coraz częściej w eksperymentach wykorzystywane są techniki wysokoprzepustowe, dostarczające badaczom olbrzymią ilość danych. Ogrom pozyskanych danych pozwala na testowanie wielu tysięcy (a często setek tysięcy) hipotez. Dla tak dużej liczby hipotez korekta Bonferroniego prowadzi do znaczącego zmniejszenia poziomu istotności w pojedynczym teście, co prowadzi do dramatycznego spadku mocy. Konieczne stało się opracowanie nowych, mniej konserwatywnych, procedur testowania. Do dziś powstało wiele prac poświęconych *Multiple Hypothesis Testing* (polskie tłumaczenia to „testo-

wanie jednoczesne”, „testowanie wielokrotne”, „testowanie zbiorcze” oraz stosowane w tej pracy „testowanie zbioru hipotez”). Prace nad procedurami testowania zbioru hipotez prowadzi wielu statystyków między innymi Terry Speed, Sandrine Dudoit, Yoseph Hochberg, John D. Storey i Brad Efron ([Sarkar 2002], [Dudoit i inni 2002], Storey2002). Teoria testowania zbioru hipotez stała się popularnym zagadnieniem, rozwijanym tak w ujęciu klasycznym jak i Bayesowskim. Zdarza się jednak, że badacze korzystają z Bayesowskiego lub częstościowego ujęcia zagadnienia testowania bez względu na charakter eksperymentu, zapominając, że oba te ujęcia zagadnienia testowania różnią się znacznie zarówno w interpretacji wyników, jak i w możliwości stosowania różnych metod statystycznych ([Efron 2005]).

Procedury wykorzystywane w zagadnieniu testowania zbioru hipotez zakładają, że decyzje o przyjęciu lub odrzuceniu każdej z hipotez mogą być podejmowane niezależnie. Przykładem zagadnienia, w którym decyzja podjęta dla jednej hipotezy wpływa na decyzje podjęte dla innych hipotez jest problem określenia maksymalnej bezpiecznej dawki lekarstwa (ang. „Maximum safe dose”). Przegląd wyników dla tego zagadnienia znaleźć można w pracach [Bauer 1997], [Liu 1997]. Przypuśćmy, że przeprowadziliśmy eksperymenty, w wyniku których zmierzylśmy jak na stan pacjenta wpływają różne dawki pewnego leku. Pytanie, które stawia badacz, to jaka maksymalna dawka leku nie ma szkodliwego wpływu na stan pacjenta. Stawiamy więc zbiór hipotez zerowych postaci „dawka mniejsza równa  $x_i$  nie ma szkodliwego wpływu na stan pacjenta”. Tak postawione hipotezy są zależne semantycznie. Statystyki testowe dla poszczególnych hipotez, mogą być niezależne, jednak decyzja o przyjęciu lub odrzuceniu jednej hipotezy wpływa na decyzje podjęte dla innych hipotez. Wynik testowania musi być logicznie spójny, a więc jeżeli odrzucimy hipotezę, że  $x$  jest bezpieczną dawką, to musimy też odrzucić hipotezę zerową dla każdej dawki  $y \geq x$ . Prace nad tym zagadnieniem rozpoczął E. Peritz (z tematem tym związana była jego praca doktorska). Badania nad takimi zależnościami prowadzili między innymi [Marcus i inni 1976], [Dror 1990], [Hommel 1988], [Tamhane, Logan 2004].

W literaturze rozważane są również bardziej złożone relacje. Wiele prac poświęconych hierarchicznej relacji pomiędzy hipotezami opublikowali E. Sonnemann i H. Finner (np. [Sonnemann, Finner 1988], [Finner, Strassburger 2002]). Hierarchiczna zależność pomiędzy hipotezami jest w tych pracach wywiedziona z postaci hipotez zerowych. Autorzy określają dla zbioru wszystkich hipotez wspólną przestrzeń parametrów  $\Theta$ , a rzeczywisty stan wszystkich hipotez opisują parametrem  $\theta^* \in \Theta$ . Hipotezy zerowe  $H_0^{(i)}$  odpowiadają przypuszczeniu, że  $\theta^* \in \Theta_0^{(i)} \subset \Theta$ . Ponieważ wszystkie zbiory  $\Theta_0^{(i)}$  są podzbiórami w tej samej przestrzeni, więc mogą zachodzić pomiędzy nimi relacje zawierania, co odpowiada hierarchicznej relacji pomiędzy hipotezami. Jeżeli zbiór  $\Theta_0^{(i)}$  jest zawarty w zbiorze  $\Theta_0^{(j)}$ , to przyjęcie hipotezy zerowej  $H_0^{(i)}$  wymaga przyjęcia hipotezy zerowej  $H_0^{(j)}$ .

W tej pracy również rozważamy relację hierarchii pomiędzy hipotezami, jednak wprowadzoną w odmienny sposób. Badania nad tym zagadnieniem zostały zainspirowane hierarchiczną strukturą funkcji biologicznych Gene Ontology wykorzystywaną między innymi w analizach mikromacierzowych [Goeman 2004]. W tym przypadku trudno jest opisać hipotezy zerowe określając wspólną przestrzeń parametrów  $\Theta$  tak by zachować relacje pomiędzy hipotezami odpowiadające relacjom pomiędzy funkcjami biologicznymi. Wyniki zaprezentowane w tej pracy mogą być wykorzystane w analizach mikromacierzowych, przedstawione procedury pozwalają na kontrolę błędów PFER, FDR i FWER.

Przedstawione w tym rozdziale informacje powstały głównie na podstawie pozycji [Fenton 1921], [Beattie 1967], [Numbers 1978], [Reid 1982], [Klonecki 1995], [Ledwina 2005].

# Bibliografia

- [Baierl i inni 2006] A. Baierl, A. Futschik, M. Bogdan, P. Biecek. Locating multiple interacting quantitative trait loci using robust model selection. *Computational Statistic and Data Analysis* (2006).
- [Beattie 1967] L. M. Beattie. *John Arbuthnot - mathematician and satirist*. Russell&Russell Pub (1967) ISBN: 978-0846209157.
- [Bauer 1997] P. Bauer. A note on multiple testing procedures in dose finding. *Biometrics* (1997), Vol. 53, No. 3, pp. 1125-1128.
- [Benjamini, Hochberg 1995] Y. Benjamini, Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* (1995) 57, No. 1, pp. 289-300.
- [Benjamini, Yekutieli 2001] Y. Benjamini, D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* (2001) 29, 1165-1188.
- [Benjamini i inni 2005] Y. Benjamini, A. M. Krieger, D. Yekutieli. Adaptive linear step-up procedures that control the false discovery rate (2005).
- [Biecek 2003] P. Biecek. *Zastosowanie zagadnienia jednoczesnego testowania wielu hipotez w badaniach genetycznych*. Praca magisterska (2003).
- [Biecek, Jasińska 2004] P. Biecek, A. Jasinska. New algorithm to estimate absolute mRNA concentrations from GeneChip data. *Raporty Inst. Mat. PWroc.* (2005) Ser. PRE nr 42, 11 poz. maszyn. I18/05/P-042.
- [Biecek i inni 2004] P. Biecek, M. Bogdan, R. Doerge, J. Ghosh. Locating multiple interacting quantitative trait loci with the modified version of Bayesian Information Criterion. *Raporty Inst. Mat. PWroc.* (2005) Ser. PRE nr 11, 29 poz. maszyn. XN I18/05/P-011.
- [Biecek 2005] P. Biecek. A modified Bonferroni correction for strongly correlated test statistics. *Proceedings of the 11th National Conference on Application of Mathematics in Biology and Medicine* (2005) p. 67-72. ISBN: 83-903893-3-9.
- [Biecek, Klonecki 2006] P. Biecek, W. Klonecki. Relations between parameters in full models and submodels for QTL mapping. *Zgłoszone do Genetics* (2006).
- [Biecek 2007] P. Biecek. Multiple testing procedures for hierarchically related hypotheses. *Zgłoszone do Statistical Applications in Genetics and Molecular Biology* (2007).
- [Bonferroni 1936] C.E. Bonferroni. *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze (1936) 8:3-62.
- [Cheverud 2001] J. Cheverud. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* (2001) 87 Issue 1 Page 52.
- [Dudoit i inni 2002] S. Dudoit, Y.H. Yang, M.J. Callow, T.P. Speed. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* (2002), 12 111-139.

- [Dror 1990] M.R. Dror. A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika* (1990) 77(3):663-665.
- [Efron 2005] B. Efron. Modern Science and the Bayesian-Frequentist controversy (2005).
- [Fenton 1921] C. Fenton. The origin and development of the earth. *American Midland Naturalist* (1921). Vol. 7, No. 4/5 (1921), pp. 101-126.
- [Futschik, Posch 2005] A. Futschik, M. Posch. On the optimum number of hypotheses to test when the number of observations is limited. *Statistica Sinica* (2005), 15: 841-855.
- [Finner, Strassburger 2002] H. Finner, K. Strassburger. The partitioning principle: A powerful tool in multiple decision theory. *The Annals of Statistics* (2002). Vol. 30, No. 4, 1194–1213.
- [Ge i inni 2003] Y. Ge, S. Dudoit, T.P. Speed. Resampling-based multiple testing for microarray data analysis. Technical Report (2003) <http://www.stat.berkeley.edu/~gyc/633.pdf>.
- [Geyer, Meeden 2005] C.J. Geyer, G.D. Meeden. Fuzzy and randomized confidence intervals and p-values (2005).
- [Goeman 2004] J. Goeman. Global testing as an alternative to single gene testing in genomic microarray data with some open multiple testing problems (2004). CIRM Seminar Luminy.
- [Harris i inni 2004] M.A. Harris, i inni. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* (2004) 32: D258–D261.
- [Hochberg 1988] Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* (1988), 75: 800-803.
- [Hochberg, Tamhane 1987] Y. Hochberg, A. Tamhane. *Multiple Comparison Procedures*. New York Wyd. John Wiley & Sons 1987.
- [Holm 1979] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* (1979) 6: 65-70.
- [Hommel 1988] G. Hommel. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* (1988) 75(2):383-386.
- [Klonecki 1995] W. Klonecki. Jerzy Neyman. *Probability and Mathematical Statistics* (1995), vol. 15.
- [Kulinskaya, Lewin 2006] E. Kulinskaya, A. Lewin. On randomised FWER and FDR procedures for discrete distributions. (2006). <http://www.bgx.org.uk/alex/randFDRn4.pdf>.
- [Kustra i inni 2006] R. Kustra, R. Shioda, M. Zhu. A factor analysis model for functional genomics. *BMC Bioinformatics* (2006), 7:216.
- [Ledwina 2005] T. Ledwina. O pewnych testach statystycznych i ich porównywaniu. *Wiadomości Matematyczne XLI* 2005.
- [Liu 1997] W. Liu. On step-up tests for comparing several treatments. *Statistica Sinica* (1997), 7 957-972.
- [Marcus i inni 1976] R. Marcus, E. Peritz, K.R. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* (1976) 63(3):655-660.
- [Numbers 1978] R. Numbers. Creation by natural law: Laplace's nebular hypothesis in American Thought. *Isis* (1978), Vol. 69, No. 2 p. 312.
- [Neyman, Pearson 1933] J. Neyman, E. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society* (1933) p. 289-337
- [Owen 2005] A. Owen. Variance of the number of false discoveries. *J. R. Statist. Soc. B* (2005) 67, 411-26.

- [Rebai i inni 1994] A. Rebai, B. Goffinet, B. Mangin. Approximate thresholds of interval mapping test for QTL detection. *Genetics* (1994) 138: 235-240.
- [Reid 1982] C. Reid. *Neyman from life*. Springer (1982). ISBN: 978-0387907475.
- [Sarkar 2002] S.K. Sarkar. Some results on false discovery rate in stepwise multiple testing procedures. *Annals of Statistics*, (2002) 30, 239-257.
- [Sidak 1967] Z. Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* (1967) 62: 626–633.
- [Siegmund 1985] D. Siegmund. *Sequential analysis: tests and confidence intervals*. Springer-Verlag, New York (1985).
- [Simes 1986] R.J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* (1986) 73, 751-754.
- [Storey 2002] J. Storey. A direct approach to false discovery rates. *J. R. Statist. Soc. B* (2002) 64, Part 3, pp 479-498.
- [Storey i inni 2004] J. Storey, J. Taylor, D. Siegmund. Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Statist. Soc. B* (2004) 66: 187-205.
- [Sonnemann, Finner 1988] E. Sonnemann, H. Finner. Vollständigkeitsätze für multiple Testprobleme. In *Multiple Hypothesenprüfung* 121-135. Springer, Berlin (1988).
- [Szyda i inni 2004] J. Szyda, P. Biecek, F. Frommlet, J.K. Ghosh, M. Bogdan. Analysis of genetic background of quantitative traits related to alcoholism by mixed inheritance and oligogenic models. *Raporty Inst. Mat. PWroc* (2005) Ser. PRE nr 50, 7 poz. maszyn. I10/05/P-050.
- [Tamhane i inni 1996] A.C. Tamhane, C. W. Dunnett, Y. Hochberg. Multiple test procedures for dose finding. *Biometrics* (1996), 52, 21-37.
- [Tamhane, Logan 2004] A.C. Tamhane, B.R. Logan. Multiple comparison procedures in dose response studies. Working Paper (2004) No. 04-017.
- [Troyanskaya i inni 2003] O.G. Troyanskaya, K. Dolinski, A.B. Owen, R.B. Altman, D. Botstein. A Bayesian framework for combining heterogeneous data sources for gene function prediction. *PNAS* (2003) Vol. 100 no. 14 p. 8348–8353.
- [Westfall, Young 1993] P.H. Westfall, S.S. Young. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley and Sons (1993).
- [Wu i inni 2004] L.F. Wu, T.R. Hughes, A.P. Davierwala, M.D. Robinson, R. Stoughton, S.J. Altschuler. Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. *Nature Genetics* (2002) 31(3): 137-142.
- [Xiao, Pan 2005] G. Xiao, W. Pan. Gene function prediction by a combined analysis of gene expression data and protein-protein interaction data. *J Bioinform Comput Biol.* (2005) 3(6):1371-89.
- [Vinayagam 2004] A. Vinayagam, R. König, J. Moormann, F. Schubert, R. Eils, K.H. Glatting, S. Suhai. Applying Support Vector Machines for Gene Ontology based gene function prediction. *BMC Bioinformatics* (2004) 5:116.
- [Zhong i inni 2004] S. Zhong, L. Tian, C. Li, F.K. Storch, W.H. Wong. Comparative analysis of gene sets in the Gene Ontology space under the multiple hypothesis testing framework. *Proc IEEE Computational Systems Bioinformatics* (2004) 425-435.