APPLICATIONES MATHEMATICAE
41,1 (2014), pp. 81–92

Konrad Furmańczyk (Warszawa)

SELECTION IN PARAMETRIC MODELS VIA SOME STEPDOWN PROCEDURES

Abstract. The paper considers the problem of consistent variable selection in parametic models with the use of stepdown multiple hypothesis procedures. Our approach completes the results of Bunea et al. [J. Statist. Plann. Inference 136 (2006)]. A simulation study supports the results obtained.

1. Introduction. We consider the problem of variable selection via multiple hypothesis testing in parametric models. Bunea et al. [6] showed that the false discovery rate (FDR) procedure and the Bonferroni method applied to a linear regression and logistic regression selection problem lead to consistent variable selection. The procedure that controls the false discovery rate (FDR) (the expectation value of the proportion of false discoveries) has been developed in the context of multiple hypothesis testing by Benjamini and Hochberg [3]. This procedure belongs to the class of stepup procedures (for more discussion see [4], [18]). We used this approach for stepdown multitest procedures [4], [11]–[16], [18], [19] to obtain consistent variable selection in parametric models, especially for linear regression, logistic regression and for the Cox regression model.

We consider a class of models $\mathcal{M}_{\boldsymbol{\beta}}$ indexed by a parameter $\boldsymbol{\beta} \in \mathbb{R}^p$, where p is constant and independent of the sample size n. The true model, which is unknown, is specified as

$$\mathcal{M}_{\boldsymbol{\beta}}^{\text{true}} : \beta_i \neq 0 \text{ for } i \in I_0 \text{ and } \beta_i = 0 \text{ for } i \in I_1,$$

where $I_1 := \{1, \dots, p\} \setminus I_0$ ($|I_0| = p_0$), $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$. In particular, we consider the following special cases:

2010 *Mathematics Subject Classification*: Primary 62J15, 62F03; Secondary 62J05.
Key words and phrases: multiple hypothesis testing, stepdown procedure, consistent variable selection, linear regression, logistic regression, Cox model.

DOI: 10.4064/am41-1-7

(a) The *linear regression model*, where the vector of observations $\mathbb{Y} = (y_1, \ldots, y_n)^T$ is of the form

$$\mathbb{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbb{X}$ is a matrix with deterministic variables x_{ij}, $1 \le i \le n$, $1 \le j \le p$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ is a vector of iid random errors, $\mathbb{E}\varepsilon_1 = 0$, $\mathbb{E}\varepsilon_1^2 = \sigma^2$.

(b) The *logistic regression model*, where $y_1, \ldots, y_n$ are independent binomial observations, with y_i denoting the success count in n_i independent trials, and $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T$ is a vector of deterministic covariates related to y_i. The probability of success is given by

$$p_i(\mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}^T \mathbf{x}_i)}.$$

(c) The *Cox regression model*, where we observe the realization of independent continuously distributed positive random variables T_i, $i = 1, \ldots, n$, representing the times of death of n individuals. Each of the individuals can only be observed on a fixed time interval $[0, c_i]$ for certain censoring times c_i, $i = 1, \ldots, n$. Individual i has hazard rate

$$\lambda_i(t) = \lim_{h \to 0} \frac{1}{h} \mathbb{P}(T_i \le t + h \mid T_i \ge t)$$

of the special form

$$\lambda_i(t) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_i),$$

where $\mathbf{x}_i$ is a column vector of p covariates, and λ_0 is a fixed unknown baseline hazard rate for an individual with $\mathbf{x} = \mathbf{0}$.

The problem of model selection or variable selection is equivalent to the problem of estimating I_0. In a typical situation the distribution of the observed variables from those models depends on the model parameters β_i, $i = 1, \ldots, p$. Based on the statistics T_{ni} we test the multiple hypothesis:

$$(\mathrm{h}_0) \qquad H_i : \beta_i = 0 \quad \text{versus} \quad H_i' : \beta_i \neq 0, \qquad \text{for } i = 1, \ldots, p.$$

A selection procedure for the parametric model $\mathcal{M}_{\boldsymbol{\beta}}$ may be described by the set $\hat{I}$ of all indices $i \in I_1$ for which the null hypothesis H_i is rejected, and it is called *consistent* if

$$\mathbb{P}(\hat{I} = I_0) \to 1 \quad \text{as } n \to \infty. \tag{1.1}$$

We will assume that each statistic T_{ni}, whenever H_i is true, has asymptotic normal distribution. More formally, we assume that for some parameter estimator $\hat{\boldsymbol{\beta}}$ there exists an invertible matrix $\mathbb{M}_n$, dependent on the sample size n, such that

$$\mathbb{M}_n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d N_p(\mathbf{0}, \mathbf{I}) \quad \text{as } n \to \infty, \tag{1.2}$$

where $\mathbf{I}$ is the identity $p \times p$ matrix.

Moreover we assume

$$\mathbb{M}_n/n \to \mathbb{V} > 0. \tag{1.3}$$

The statistic T_{ni} has the form

$$T_{ni} = \frac{\hat{\boldsymbol{\beta}}_i}{\sqrt{m_{ii}}} \quad \text{for } i = 1, \ldots, p, \tag{1.4}$$

where $\mathbb{M}_n^{-1} = (m_{ij})_{1 \le i,i \le p}$.

Remark 1. In special cases under some technical conditions we have (1.2) for:

(i) linear regression

$$\mathbb{M}_n := \frac{1}{\sigma}\mathbb{X}^T\mathbb{X},$$

where $\hat{\beta}$ is the least squares estimator of $\boldsymbol{\beta}$ [17, p. 109];

(ii) logistic regression

$$\mathbb{M}_n := \mathbb{X}^T \operatorname{diag}(n_i\hat{p}_i(1 - \hat{p}_i))\mathbb{X},$$

where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$ [1, pp. 193–194];

(iii) the Cox model

$$\mathbb{M}_n := \int_0^t \left(\frac{\sum_{i=1}^n Y_i(s)\mathbf{x}_i^{\otimes 2} e^{\hat{\boldsymbol{\beta}}^T \mathbf{x}_i}}{\sum_{i=1}^n Y_i(s) e^{\hat{\boldsymbol{\beta}}^T \mathbf{x}_i}} - \left(\frac{\sum_{i=1}^n Y_i(s)\mathbf{x}_i e^{\hat{\boldsymbol{\beta}}^T \mathbf{x}_i}}{\sum_{i=1}^n Y_i(s) e^{\hat{\boldsymbol{\beta}}^T \mathbf{x}_i}} \right)^{\otimes 2} \right) d\bar{N}(s),$$

where $\hat{\boldsymbol{\beta}}$ is the partial maximum likelihood estimator of $\boldsymbol{\beta}$ [8], [9], $Y_i(t)$ is the at-risk process taking values 1 or 0 depending on whether the individual is under observation or not, N_i is the counting process which counts the observed events in the life of the ith individual over the time interval $[0, T]$, $\bar{N} = \sum_{i=1}^n N_i$ and $a^{\otimes 2}$ denotes the matrix aa^T [2].

Remark 2. The condition (1.3) holds for:

(i) linear regression if

$$\frac{1}{n}\mathbb{X}^T\mathbb{X} \to \mathbb{W} > 0;$$

(ii) logistic regression if

$$\frac{1}{n}\mathbb{X}^T \operatorname{diag}(n_i\hat{p}_i(1 - \hat{p}_i))\mathbb{X} \to \mathbb{V} > 0;$$

(iii) the Cox model under regular conditions [2, p. 1106] for $\mathbb{V} = \mathbb{I}(\beta)$, where $\mathbb{I}(\beta)$ is the Fisher information matrix.

Due to (1.2), we assume that the p-values of the statistic T_{ni} are

$$\pi_i(t_{ni}) = 2(1 - \Phi(|t_{ni}|)),$$

where Φ is the cdf of the standard normal distribution.

The problem of variable selection with parametric model can be viewed as multiple testing (h_0). In this testing, we use stepdown procedures [15], which we describe as follows. Let $\pi_1, \ldots, \pi_p$ be the p-values for individual tests, let $\pi_{(1)} \leq \cdots \leq \pi_{(p)}$ denote these p-values ordered, and let $H_{(1)}, \ldots, H_{(p)}$ stand for the corresponding null hypotheses. Let in addition $\alpha_1 \leq \cdots \leq \alpha_p$ be given thresholds dependent on n. We proceed according to the following scheme. If $\pi_{(1)} > \alpha_1$, we reject no null hypotheses. Otherwise, if

$$\pi_{(1)} \leq \alpha_1, \ \ldots, \ \pi_{(r)} \leq \alpha_r, \tag{h_1}$$

we reject the hypotheses $H_{(1)}, \ldots, H_{(r)}$, where the largest r satisfying (h_1) is used. Suppose that R is the total number of rejections, and V denotes the number of false rejections for the multitesting problem (h_0), (h_1). It is easy to check (see [6]) that the selection procedure is consistent if

$$\mathbb{P}(R = p_0, V = 0) \to 1 \quad \text{as } n \to \infty. \tag{1.5}$$

The paper is organized as follows. In Section 2 we formulate our main result, Theorem 1, and some technical lemmas and conditions implying consistency of some class of stepdown model selection procedures. The consistency of those procedures for the linear model with independent and dependent errors when p may depend on n has been presented in [10]. The results of our simulation study are displayed in Section 3. All the necessary proofs are contained in the Appendix.

2. Consistency of some stepdown procedures. Denote by $q_1, \ldots, q_{p-p_0}$ the p-values corresponding to the true null hypotheses (h_0), and by $r_1, \ldots, r_{p_0}$ the p-values corresponding to the false null hypotheses. Let F_j be the distribution function of the random variable r_j for $j = 1, \ldots, p$.

Before we formulate our main result, we introduce the following assumptions:

(A1) $\alpha_p \to 0$ as $n \to \infty$.

(A1a) $\log(2/\alpha_p)/n \to 0$ as $n \to \infty$.

(A2) $\max_{j \in I_0}(1 - F_j(\alpha_p)) \to 0$ as $n \to \infty$.

Theorem 1. *The stepdown procedure satisfying* (A1)–(A2) *under conditions* (1.2)–(1.3) *is consistent for the selection problem in the parametric model.*

By (1.5), it is clear that the model selection procedure is consistent if the following conditions are fulfilled:

(i) $\mathbb{P}(V \geq 1) \to 0$ as $n \to \infty$,
(ii) $\mathbb{P}(R \neq p_0) \to 0$ as $n \to \infty$.

We introduce the condition:

(iii) $\sum_{j=1}^{p_0} \mathbb{P}(\pi_{(j)} > \alpha_j) + \mathbb{P}(\pi_{(p_0+1)} \leq \alpha_{p_0+1}) \to 0$ as $n \to \infty$.

Theorem 1 follows from

Lemma 1. *Under conditions* (i) *and* (iii), *any stepdown model selection procedure is consistent.*

Condition (i) is a consequence of Lemma 2. Condition (iii) follows from Lemmas 2–3 and [7, Proposition 3].

Lemma 2. *For any* $0 \leq x \leq 1$,

$$\mathbb{P}(q_i \leq x) = x + o(1) \quad \textit{as } n \to \infty. \tag{2.1}$$

Our next lemma is very similar to Lemma 2.4 in Bunea et al. [6] where consistency was shown for the Bonferroni and the FDR procedures [5]. We consider some stepdown procedure satisfying (A1)–(A2).

Lemma 3. *Suppose that* (A1)–(A2) *and* (1.2)–(1.3) *hold. Then*

$$\mathbb{P}(r_{(p_0)} \leq q_{(1)}) \to 1 \quad \textit{as } n \to \infty, \tag{2.2}$$

where

$$r_{(p_0)} = \max\{r_1, \ldots, r_{p_0}\} \quad \textit{and} \quad q_{(1)} = \min\{q_1, \ldots, q_{p-p_0}\}.$$

Since the most restrictive assumption among (A1)–(A2) is (A2), we will establish some conditions implying (A2).

Lemma 4. *Assume that* (A1), (A1a), *and* (1.2)–(1.3) *hold. Then* (A2) *is fulfilled.*

Now, we give some example when (A1a) holds.

Example 1. For $a_n \sim n^{-\eta}$ for some $\eta > 0$ the following stepdown procedures satisfy (A1a):

(a) the Holm procedure with

$$\alpha_j = \frac{a_n}{p+1-j},$$

(b) a generalization of the Holm procedure [15] with

$$\alpha_j = \frac{([\gamma j]+1)a_n}{p+[\gamma j]+1-j} \quad \text{for some } 0 < \gamma < 1,$$

(c) the Bonferroni procedure with

$$\alpha_j = \frac{a_n}{p}, \quad j = 1, \ldots, p,$$

(d) the Benjamini–Hochberg stepdown version with

$$\alpha_j = \frac{j a_n}{p}, \qquad j = 1, \dots, p,$$

(e) the Benjamini–Yekutieli stepdown version with

$$\alpha_j = \frac{j a_n}{p \sum_{i=1}^{p} 1/i}, \qquad j = 1, \dots, p.$$

3. Simulation study

3.1. Models

(a) *Linear models.* We generate p independent vectors X_j from the standard normal distribution, $j = 1, \dots, p$, the design matrix $\mathbb{X}$ consists of the vectors $X_1, \dots, X_p$ as columns, and the true model has the form $\mathbb{Y} = \sum_{j=1}^{p_0} X_j + \boldsymbol{\varepsilon}$, $p_0 \le p$, where $\boldsymbol{\varepsilon}$ is a vector generated from the normal distribution with mean zero and $\sigma = 2$. We consider three linear models: for $p_0 = 3$, $p = 5$; $p_0 = 10$, $p = 20$; $p_0 = 50$, $p = 100$, for $n = 200$ and $n = 500$.

(b) *Logistic model.* As above, we generate p independent vectors X_j from the standard normal distribution, $j = 1, \dots, p$. Then we take $\eta_i = \sum_{j=1}^{p_0} x_{ij} + \varepsilon_i$, where $\boldsymbol{\varepsilon}$ is a vector generated from the normal distribution with mean zero and $\sigma = 2$. Next, we compute

$$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}, \qquad i = 1, \dots, n,$$

and draw a response y_i from the Bernoulli distribution with probability of success p_i. The true model has coefficients $\beta_1 = \dots = \beta_{p_0} = 1$, $\beta_i = 0$ for $i = p_0 + 1, \dots, p$.

We consider four models: for $p_0 = 3$, $p = 5$; $p_0 = 10$, $p = 20$; $p_0 = 20$, $p = 30$; $p_0 = 30$, $p = 50$, for $n = 200$ and $n = 500$.

(c) *Cox regression model.* Based on the fact that the times of death T_i in the Cox model have the form $T_i = \Lambda_0^{-1}(-\log(U_i)\exp(\boldsymbol{\beta}^T \mathbf{x}_i))$, where $\Lambda_0(t) = \int_0^t \lambda_0(s)\,ds$, and U_i is a random variable with uniform distribution $U[0,1]$, we generate p independent vectors $\mathbf{x}_j$ from the standard normal distribution, $j = 1, \dots, p$, and simulate the times of death

$$T_i = -\log(U_i)\exp\Big(-\Big(\sum_{j=1}^{p_0} \beta_j x_{ij} + \sum_{j=p_0+1}^{p} \beta_j x_{ij}\Big)\Big),$$

where $\Lambda_0^{-1}(t) = t$, U_i is generated from the uniform distribution $U[0,1]$ for $i = 1, \dots, n$, for $n = 200, 500$. We consider nine Cox models for $\beta_j = 1$ for $j = 1, \dots, p_0$ and $\beta_j = 0$ for $j = p_0 + 1, \dots, p$, where, respectively, $p_0 = 5$, $p = 10$; $p_0 = 10$, $p = 20$; $p_0 = 30$, $p = 50$ with 20%, 30% and 40% censored observations.

We simulated samples of sizes $n = 200, 500$ from the above models and we recorded the numbers of true models selected from each of $N = 1000$ MC replications with the use of the following stepdown procedures: Holm's, a generalization of Holm's (UHolm for $\gamma = 0.01,\ 0.1,\ 0.5,\ 0.9$) and Bonferroni's (Bonf), Benjamini–Hochberg (BH), Benjamini–Yekutieli (BY) for $a_n = 1/\sqrt{n}$. The results of the simulations are presented in Tables 1–3.

Table 1. Frequencies of the true model being selected by multiple procedures in 1000 simulations for linear models

	$n = 200$ $p_0 = 3$ $p = 5$	$n = 500$ $p_0 = 3$ $p = 5$	$n = 200$ $p_0 = 10$ $p = 20$	$n = 500$ $p_0 = 10$ $p = 20$	$n = 200$ $p_0 = 50$ $p = 100$	$n = 500$ $p_0 = 50$ $p = 100$
Bonf	968	980	966	975	69	974
Holm	938	945	939	949	124	951
UHolm_0.01	938	945	939	949	124	951
UHolm_0.1	938	945	892	920	335	782
UHolm_0.5	903	919	767	832	237	472
UHolm_0.9	897	907	705	787	183	340
BH	905	930	669	755	165	330
BY	957	965	892	925	271	796

Table 2. Frequencies of the true model being selected by multiple procedures in 1000 simulations for logistic models

	$n = 200$ $p_0 = 3$ $p = 5$	$n = 500$ $p_0 = 3$ $p = 5$	$n = 200$ $p_0 = 10$ $p = 20$	$n = 500$ $p_0 = 10$ $p = 20$	$n = 200$ $p_0 = 20$ $p = 30$	$n = 500$ $p_0 = 20$ $p = 30$
Bonf	739	982	12	883	0	352
Holm	768	966	40	900	0	557
UHolm_0.01	768	966	40	900	0	557
UHolm_0.1	768	966	60	888	0	684
UHolm_0.5	800	939	148	833	9	673
UHolm_0.9	804	934	177	769	18	647
BH	772	933	177	767	14	640
BY	769	965	38	899	0	674

3.2. Conclusions. For large sample size all the procedures applied have very good power. When the sample size is not very large, the Bonferroni method is the best and the generalized Holm method for $\gamma = 0.9$ is the worst. In logistic regression with p large, Holm and UHolm procedures work better than other stepdown methods. In Cox models with greater percentage of censored observations our procedures have weaker power than in models with lower percentage of censored observations.

Table 3. Frequencies of the true model being selected by multiple procedures in 1000 simulations for Cox models

	$n = 200$ $p_0 = 3$ $p = 5$	$n = 500$ $p_0 = 3$ $p = 5$	$n = 200$ $p_0 = 10$ $p = 20$	$n = 500$ $p_0 = 10$ $p = 20$	$n = 200$ $p_0 = 30$ $p = 50$	$n = 500$ $p_0 = 30$ $p = 50$
Cen 20%/Bonf	953	971	928	963	867	973
Holm	927	945	877	933	770	918
UHolm_0.01	927	945	877	933	770	918
UHolm_0.1	927	945	816	891	519	791
UHolm_0.5	836	896	663	784	286	551
UHolm_0.9	804	881	593	741	198	468
BH	792	838	572	733	178	494
BY	923	948	840	920	573	831
Cen 30%/Bonf	941	973	940	969	870	968
Holm	885	955	884	948	757	937
UHolm_0.01	885	955	884	948	757	937
UHolm_0.1	885	955	819	914	482	777
UHolm_0.5	783	899	674	827	230	577
UHolm_0.9	746	881	605	776	160	476
BH	767	868	577	731	158	459
BY	907	956	856	905	562	829
Cen 40%/Bonf	961	980	950	974	831	968
Holm	921	955	902	953	724	915
UHolm_0.01	921	955	902	953	724	915
UHolm_0.1	921	955	833	910	479	775
UHolm_0.5	800	877	694	809	228	566
UHolm_0.9	771	859	613	763	163	470
BH	796	879	601	761	151	480
BY	908	955	844	925	538	796

4. Appendix

Proof of Lemma 1. Observe that

$$\begin{aligned}\mathbb{P}(R < p_0) &= \mathbb{P}(\pi_{(1)} > \alpha_1) + \mathbb{P}(\pi_{(1)} \leq \alpha_1, \pi_{(2)} > \alpha_2) + \cdots \\ &\quad + \mathbb{P}(\pi_{(1)} \leq \alpha_1, \pi_{(2)} \leq \alpha_2, \ldots, \pi_{(p_0-1)} \leq \alpha_{p_0-1}, \pi_{(p_0)} > \alpha_{p_0}) \\ &\leq \sum_{j=1}^{p_0} \mathbb{P}(\pi_{(j)} > \alpha_j),\end{aligned}$$

and

$$\mathbb{P}(R > p_0) \leq \mathbb{P}(\pi_{(p_0+1)} \leq \alpha_{p_0+1}),$$

as $\{R > p_0\} \subseteq \{\pi_{(p_0+1)} \leq \alpha_{p_0+1}\}$.

Hence $\mathbb{P}(R \neq p_0) \leq \sum_{j=1}^{p_0} \mathbb{P}(\pi_{(j)} > \alpha_j) + \mathbb{P}(\pi_{(p_0+1)} \leq \alpha_{p_0+1})$ and (ii) follows directly from (iii). Then, by conditions (i), (ii) from Section 2, our model selection procedure is consistent. ∎

Proof of Lemma 2. If $\beta_i = 0$, then

$$T_{ni} = \frac{\hat{\beta}_i}{\sqrt{m_{ii}}} = \frac{\hat{\beta}_i - \beta_i}{\sqrt{m_{ii}}} =: \tilde{T}_{ni}.$$

Note that

$$\begin{aligned}\mathbb{P}(q_i \leq x) &= \mathbb{P}\big(2(1 - \Phi(|T_{ni}|)) \leq x\big) = \mathbb{P}\big(2(1 - \Phi(|\tilde{T}_{ni}|)) \leq x\big) \\ &= \mathbb{P}(|\tilde{T}_{ni}| \geq \xi_x) = 1 - (\Phi(\xi_x) - \Phi(-\xi_x)) + o(1),\end{aligned}$$

where $\xi_x := \Phi^{-1}(1 - x/2)$. Therefore, from the asymptotic distribution of $\tilde{T}_{ni}$ (1.2), we have

$$\begin{aligned}\mathbb{P}(q_i \leq x) &= 1 - (\Phi(\xi_x) - \Phi(-\xi_x)) + o(1) \\ &= x + o(1). \ ∎\end{aligned}$$

Proof of Lemma 3. It is clear that

$$\mathbb{P}(r_{(p_0)} > q_{(1)}) \leq \mathbb{P}(r_{(p_0)} > \alpha_p) + \mathbb{P}(q_{(1)} \leq \alpha_p).$$

It follows from (A2) that

$$\mathbb{P}(r_{(p_0)} > \alpha_p) \leq \sum_{i=1}^{p_0} \mathbb{P}(r_j > \alpha_p) \leq p_0 \max_{j \in I_0}(1 - F_j(\alpha_p)) = o(1).$$

On the other hand, by Lemma 2,

$$\mathbb{P}(q_{(1)} \leq \alpha_p) \leq \sum_{i=1}^{p-p_0} \mathbb{P}(q_i \leq \alpha_p) \leq p\alpha_p + o(1).$$

Hence from (A1) we get (2.2), as claimed. ∎

Proof of Theorem 1. By Lemma 1, in order to prove Theorem 1, it is enough to show that conditions (i) and (iii) hold.

First, we will prove (i). Let j be the smallest (random) index such that $\pi_{(j)} = q_{(1)}$. Then

$$\begin{aligned}\mathbb{P}(V \geq 1) &\leq \mathbb{P}(\pi_{(1)} \leq \alpha_1, \ldots, \pi_{(j)} \leq \alpha_j) \leq \mathbb{P}(q_{(1)} \leq \alpha_j) \\ &\leq \mathbb{P}(q_{(1)} \leq \alpha_p) \leq \sum_{j \in I_1} \mathbb{P}(q_j \leq \alpha_p).\end{aligned}$$

By Lemma 2 (see (2.1)), we have

$$\mathbb{P}(V \geq 1) \leq p\alpha_p + o(1).$$

Therefore, (i) holds.

It remains to prove (iii). It is obvious that for $1 \leq j \leq p_0$,

$$\mathbb{P}(\{\pi_{(j)} > \alpha_j\}) \leq \mathbb{P}(\{\pi_{(j)} > \alpha_j\} \cap A_n) + \mathbb{P}(A_n^c) \leq \mathbb{P}(r_{(j)} > \alpha_j) + \mathbb{P}(A_n^c),$$

where $A_n = \{r_{(p_0)} \le q_{(1)}\}$. Similarly

$$\mathbb{P}(\pi_{(p_0+1)} \le \alpha_{p_0+1}) \le \mathbb{P}(q_{(1)} \le \alpha_{p_0+1}) + \mathbb{P}(A_n^c).$$

Therefore, the l.h.s. of (iii) is not greater than

$$\sum_{j=1}^{p_0} \mathbb{P}(r_{(j)} > \alpha_j) + \mathbb{P}(q_{(1)} \le \alpha_{p_0+1}) + (p_0+1)\mathbb{P}(A_n^c). \tag{4.1}$$

It follows from [7, Proposition 3] that

$$\mathbb{P}(r_{(j)} > \alpha_j) \le \frac{\sum_{i\in I_0}(1-F_i(\alpha_j))}{p_0 - j + 1},$$

and

$$\mathbb{P}(q_{(1)} \le \alpha_{p_0+1}) \le \sum_{i\in I_1} \mathbb{P}(q_i \le \alpha_{p_0+1}) \le \sum_{i\in I_1} \mathbb{P}(q_i \le \alpha_p). \tag{4.2}$$

Reasoning as previously, we have $\sum_{i\in I_1} \mathbb{P}(q_i \le \alpha_p) \to 0$ as $n \to \infty$. Hence

$$\sum_{j=1}^{p_0} \mathbb{P}(r_{(j)} > \alpha_j) \le \sum_{j=1}^{p_0} \frac{\sum_{i\in I_0}(1-F_i(\alpha_j))}{p_0 - j + 1} = \mathcal{O}\Big(\max_{j\in I_0}(1 - F_j(\alpha_j))\Big). \tag{4.3}$$

By Lemma 3 and (4.2)–(4.3), the r.h.s. of (4.1) tends to zero as $n \to \infty$. Consequently, (iii) holds. ∎

Proof of Lemma 4. Observe that

$$1 - F_j(\alpha_p) = 1 - \mathbb{P}(\pi_j(T_{nj}) \le \alpha_p) = \mathbb{P}(|T_{nj}| \le \xi_{\alpha_p}),$$

where $\xi_{\alpha_p} = \Phi^{-1}(1-\alpha_p/2)$. Hence

$$\begin{aligned} 1 - F_j(\alpha_p) &= \tilde{G}_{nj}(\xi_{\alpha_p}) - \tilde{G}_{nj}(-\xi_{\alpha_p}) \\ &= G_{nj}\left(\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}}\right) - G_{nj}\left(-\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}}\right), \end{aligned}$$

where $\tilde{G}_{nj}$ is the distribution function of the statistic $T_{nj} = \hat{\beta}_j/\sqrt{m_{jj}}$, and G_{nj} is the distribution function of $(\hat{\beta}_j - \beta_j)/\sqrt{m_{jj}}$. From (1.2) we obtain

$$1 - F_j(\alpha_p) = \Phi\left(\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}}\right) - \Phi\left(-\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}}\right) + o(1).$$

It remains to prove

$$\Phi\left(\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}}\right) - \Phi\left(-\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}}\right) = o(1) \quad \text{for all } j \in I_0. \tag{4.4}$$

Since $\Phi^{-1}(1-\alpha_p/2) < \sqrt{2\log(2/\alpha_p)}$ for large n, from (A1a) we have

$$\frac{\Phi^{-1}(1-\alpha_p/2)}{\sqrt{n}} = \frac{\xi_{\alpha_p}}{\sqrt{n}} \to 0.$$

From (1.3) we have $\sqrt{n}\sqrt{m_{jj}} = \mathcal{O}(1)$ as $n \to \infty$, so

$$\frac{\xi_{\alpha_p}\sqrt{m_{jj}}}{|\beta_j|} = \frac{\xi_{\alpha_p}\sqrt{n}\sqrt{m_{jj}}}{\sqrt{n}\,|\beta_j|} \to 0 \quad \text{as } n \to \infty,$$

and

$$\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}} = \frac{\beta_j}{\sqrt{m_{jj}}}\left(\frac{\xi_{\alpha_p}\sqrt{m_{jj}}}{|\beta_j|} - \frac{\beta_j}{|\beta_j|}\right) = \frac{\beta_j}{\sqrt{m_{jj}}}\left(o(1) - \frac{\beta_j}{|\beta_j|}\right)$$

and

$$-\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}} = \frac{\beta_j}{\sqrt{m_{jj}}}\left(o(1) - \frac{\beta_j}{|\beta_j|}\right).$$

Since $j \in I_0$, $\xi_{\alpha_p} - \beta_j/\sqrt{m_{jj}}$ and $-\xi_{\alpha_p} - \beta_j/\sqrt{m_{jj}}$ each tend to $-\infty$ or ∞. Hence

$$\Phi\left(\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}}\right) - \Phi\left(-\xi_{\alpha_p} - \frac{\beta_j}{\sqrt{m_{jj}}}\right) \to 0 \quad \text{as } n \to \infty. \ ■$$

References

[1] A. Agresti, *Categorical Data Analysis*, 2nd ed., New York, 2002.

[2] P. K. Andersen and R. D. Gill, *Cox's regression model for counting processes: a large sample study*, Ann. Statist. 10 (1982), 1100–1120.

[3] Y. Benjamini and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, J. Roy. Statist. Soc. Ser. B 57 (1995), 289–300.

[4] Y. Benjamini and W. Liu, *A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence*, J. Statist. Plann. Inference 82 (1999), 163–170.

[5] Y. Benjamini and D. Yekutieli, *The control of the false discovery rate in multiple testing under dependency*, Ann. Statist. 29 (2001), 1165–1188.

[6] F. Bunea, M. H. Wegkamp and A. Auguste, *Consistent variable selection in high dimensional regression via multiple testing*, J. Statist. Plann. Inference 136 (2006), 4349–4364.

[7] G. Caraux and O. Gascuel, *Bounds on distribution functions of order statistics for dependent variates*, Statist. Probab. Lett. 14 (1992), 103–105.

[8] D. R. Cox, *Regression models and life tables (with discussion)*, J. Roy. Statist. Soc. Ser. B 34 (1972), 187–220.

[9] D. R. Cox, *Partial likelihood*, Biometrika 62 (1975), 269–276.

[10] K. Furmańczyk, *On some stepdown procedures with application to consistent variable selection in linear regression*, Statistics (2013) (online).

[11] Y. Ge, S. C. Sealfon and T. P. Speed, *Some step-down procedures controlling the false discovery rate under dependence*, Statistica Sinica 18 (2008), 881–904.

[12] D. Haughton, *On the choice of a model to fit data from an exponential*, Ann. Statist. 16 (1988), 342–355.

[13] S. Holm, *A simple sequentially rejective multiple test procedure*, Scand. J. Statist. 6 (1979), 65–70.

[14] G. Hommel, *A stagewise rejective multiple test procedure based on a modifed Bonferroni test*, Biometrika 75 (1988), 383–386.

[15] E. L. Lehmann and J. P. Romano, *Generalizations of the familywise error rate*, Ann. Statist. 28 (2005), 1–25.
[16] J. P. Romano and M. Wolf, *Exact and approximate stepdown methods for multiple hypothesis testing*, J. Amer. Statist. Assoc. 100 (2005), 94–108.
[17] A. Sen and M. Srivastava, *Regression Analysis*, Springer, New York, 1990.
[18] P. N. Somerville, *FDR step-down and step-up procedures for the correlated case*, in: Recent Developments in Multiple Comparison Procedures, IMS Lecture Notes Monogr. Ser. 47, Inst. Math. Statist., Beachwood, OH, 2004, 100–118.
[19] M. J. Van der Laan, S. Dudoit and K. S. Pollard, *Multiple testing. II. Step-down procedures for control of the family-wise error rate*, Statist. Appl. Genet. Molec. Biol. 3 (2004), art 15, 27 pp.

Konrad Furmańczyk
Department of Applied Mathematics
Warsaw University of Life Sciences (SGGW)
Nowoursynowska 159
02-776 Warszawa, Poland
E-mail: konfur@wp.pl

Received on 8.7.2013;
revised version on 28.3.2014 (2189)