ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 8, 2004

New parametric and nonparametric multiple test procedures for high-dimensional data

SIEGFRIED KROPF AND GERHARD HOMMEL

ABSTRACT. Modern techniques in biomedical research as microarrays or computer based imaging techniques often yield extremely high-dimensional data for a patient. We propose several procedures for separate tests with all variables controlling the experimentwise type I error in a parametric as well as in a nonparametric setup. These procedures utilise the idea that all variables should have a similar scale. Otherwise the procedures are less powerful but the type I error is still under strong control.

Various modifications of the basic procedures weaken the power-dependence on the assumption of equal variances. All procedures are very simple to implement. They are demonstrated here in a microarray data set, comparing their performance with standard techniques.

1. Introduction

Many new biomedical investigation techniques deliver lots of data per patient to characterize complex biological mechanisms. Examples are gene expression analyses by microarrays, long time series of sets of physiological parameters or imaging techniques. But still the sample sizes in studies are restricted by the number of available patients, costs or others.

Statistical comparisons can be based on the whole multivariate observations, where usually special methods as those of Läuter, Glimm and Kropf (1996,1998) are necessary to overcome problems of the extreme relation of number of parameters and sample sizes. However, these multivariate tests answer only global questions.

A deep understanding of the biological problem often requires the parallel investigation of the single parameters. Then special multiple comparison

Received October 24, 2003.
2000 *Mathematics Subject Classification.* 62G10, 62G30, 62P10, 62J15.
Key words and phrases. Multiple tests, familywise error rate, sequential procedure, parametric tests, nonparametric tests, microarrays.

procedures are necessary to avoid an unacceptable high number of false positive results caused by the cumulation of stochastic errors in the multitude of tests. Here, we focus on methods controlling the familywise error rate in the strong sense. They guarantee that in all the tests a type I error occurs with a pre-specified probability α at most, not only when all null hypotheses are true but also when some may be wrong.

When the number of variables is high and the sample sizes are small then classical procedures ensuring the familywise type I error in the strong sense have problems. Testing with a-priori ordered hypotheses (Bauer et al., 1998) is usually not possible because of missing a-priori information. The Bonferroni-Holm method (Holm, 1979) has extremely small critical levels for the smallest p-values because of the large number of hypotheses. The Westfall-Young permutation procedure (Westfall, Young, 1993) has only a small number of possible permutations in small samples. A closed test procedure (Marcus, Peritz, Gabriel, 1976) cannot be applied in a basic form with explicit consideration of all intersection hypotheses because the number of these becomes astronomical.

Therefore, we apply another principle here. We assume that all the variables have a similar scale. This is given in many applications or it can be reached by suitable transformations. In this case, variance-like terms or similar nonparametric measures can be considered as additional information which is used for ordering or weighting hypotheses. However, this additional assumption is not necessary to ensure the type I error. Thus in the case of large differences of the scale of the variables, the tests lose power but they are still valid.

In the next section, we specify the considered parametric and nonparametric test problems. Test procedures with data-driven ordering of hypotheses (Kropf, 2000; Kropf, Läuter, 2002; Kropf et al., 2004) are described in Section 3, procedures with weighted hypotheses (Westfall, Kropf, Finos, 2004; Kropf et al. 2004) in Section 4. In Section 5, a modified procedure with weighted hypotheses based on a procedure by Benjamini and Hochberg (1997) with fixed weights is proposed and discussed. A proposal by Hommel using the closed test procedure (Marcus, Peritz, Gabriel, 1976) is considered in the last section.

We only treat the one-sample case here. The two-sample case is described in the above papers or is straight-forward in the new proposals.

All procedures are demonstrated with gene expression data from the University of Leipzig, Germany (continued research work from Eszlinger, Krohn, Paschke, 2001, deadline September 2002). Here, 15 patients with cold nodules in the thyroids have been investigated and issue samples have been taken from the nodule as well as from the surrounding tissue. The extracted mRNA has been analyzed in Affymetrics GeneChips® yielding expression

activities for 12,625 genes in parallel. We use the differences of the logarithmic expression values from nodule and surrounding tissue.

2. The test problems

In this paper, we consider parametric and nonparametric versions of the one-sample case. The parametric one-sample case is characterized by a sample of n iid multivariate normal observation vectors of dimension p

$$\mathbf{x}_j = \begin{pmatrix} x_{j1} \\ \vdots \\ x_{jp} \end{pmatrix} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{with} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix} \quad (1)$$

$(j = 1, \ldots, n)$, where we want to test the p local hypotheses $H_i : \mu_i = 0 \quad (i = 1, \ldots, p)$ under strong control of the familywise type I error. The covariance matrix $\boldsymbol{\Sigma}$ is arbitrary positive semidefinite.

For a nonparametric version, we restrict to observation vectors with continuous density function. Thus, we assume that the n iid p-dimensional sample vectors $\mathbf{x}_j$ $(j = 1, \ldots, n)$ have a density $f_p(\mathbf{x})$ which is symmetrical to the location vector $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 & \cdots & \mu_p \end{pmatrix}'$,

$$f_p(\boldsymbol{\mu} + \mathbf{x}) = f_p(\boldsymbol{\mu} - \mathbf{x}) \,. \quad (2)$$

The local null hypotheses of interest are again $H_i : \mu_i = 0 \quad (i = 1, \ldots, p)$.

3. Procedures with data-driven ordering of hypotheses

For the parametric one-sample problem (1), Kropf (2000) and Kropf, Läuter (2002) proposed the following

Procedure I:

- Sort the p variables $x_1, \ldots, x_p$ for decreasing values $w_i = \sum_{j=1}^{n} x_{ji}^2$,
- in this order carry out the usual one-sample t tests for the variables at the unadjusted level α as long as significance is attained. Stop at the first non-significant test result.

The proof that this procedure controls the familywise type I error is based on the multivariate theorems in Läuter et al. (1996, 1998). It follows from the decomposition $w_i = \sum_{j=1}^{n} x_{ji}^2 = n \cdot \overline{x}^2 + \sum_{j=1}^{n} (x - \overline{x})^2$ that the ordering by w_i is essentially an ordering by the absolute values of the means if the variances of all variables are equal or similar.

A nonparametric counterpart for the model (2) is described in Kropf et al. (2004) utilizing the independence of rank and order statistics under the null hypothesis:

Procedure I*:

- Sort the p variables $x_1, \ldots, x_p$ for decreasing values of the median m_i of the absolute sample values,
- in this order carry out the usual one-sample Wilcoxon tests for the variables at the unadjusted level α as long as significance is attained. Stop at the first non-significant test result.

In the example, sorting of the variables for decreasing w_i gives the following series of corresponding p-values in the one-sample t tests: .0002, .0006, .0032, .0871, .0254, .2227, .5001, .1334, .9365, .1066, .0858, .0006, .0000, ... Analogously, the nonparametric procedure brings the p-values .0006, .0034, .5245, .0256, .0730, .4887, .2524, .1070, .0009, ... Thus, we have to stop after the third variable in the parametric case and after the second variable in the nonparametric procedure though in both cases very small p-values are in later positions. Nevertheless, the three or two significant variables are remarkable insofar as neither parametric or nonparametric Bonferroni nor the Westfall-Young permutation procedure could show significance in any variable. The principle of utilizing the quadratic forms or the medians as additional information appears to be useful, but the stringent use of ordering makes the procedures accessible for disturbances.

4. A weighted test procedure

In his basic paper, Holm (1979) already described the possibility of applying (fixed) weights for the local hypotheses. Westfall et al. (2004) now choose data dependent weights based on the above proposals. In the parametric case this gives

Procedure II:

- Determine the unadjusted p-values p_i from the one-sample t tests and the sums of squares w_i for all p variables,
- with a fixed $\eta \geq 0$, calculate weights $g_i = w_i^\eta$ and weighted p-values $q_i = p_i/g_i$ $(i = 1, \ldots, p)$,
- sort the variables for increasing weighted p-values $q_{(1)} \leq q_{(2)} \leq \cdots \leq q_{(p)}$ and denote the corresponding weights by $g_{(1)}, g_{(2)}, \ldots, g_{(p)}$,
- reject $H_{(j)}$ as long as

$$q_{(j)} \leq \frac{\alpha}{\sum_{i=j}^{p} g_{(i)}}, \tag{3}$$

stop at the first non-significant result.

The nonparametric counterpart (Kropf et al., 2004) uses the medians m_i of the absolute values instead of the w_i and the p-values from the one-sample Wilcoxon tests instead of those from the t tests.

It is easy to see that we have the usual unweighted Bonferroni-Holm method with $\eta = 0$. Furthermore, Westfall and Krishen (2001) have shown

that this procedure converges to Procedure I for $\eta \to \infty$. The procedure with intermediate values of η can be considered as smoothed version of procedure I, where the smoothing effect becomes larger for smaller η.

The results for the example depending on the parameter η are shown in Figure 1 in the next section. One can see a real improvement for intermediate values of the smoothing parameter. Furthermore, the parametric procedure finds more significant genes than the nonparametric one in this example.

5. An alternative weighted procedure

The linkage of Procedure I with the Bonferroni-Holm principle reduces the problem that extremely small p-values in the local tests have to be disregarded in the procedure because of previous non-significant results. It does not prevent the problem, however. Therefore, we consider an alternative proposal by Benjamini and Hochberg (1997). They suggested this for fixed weights but we will show that it can be used with the above data-dependent weights as well.

Procedure III:

- Determine the unadjusted p-values p_i and the sums of squares w_i for all p variables,
- with a fixed $\eta \geq 0$, calculate weights $g_i = w_i^\eta$,
- sort the variables for increasing *unweighted* p-values $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(p)}$ and denote the corresponding weights by $g_{(1)}, g_{(2)}, \ldots, g_{(p)}$,
- reject $H_{(j)}$ as long as

$$p_{(j)} \leq \frac{g_{(j)}\,\alpha}{\sum_{i=j}^{p} g_{(i)}}\,, \tag{4}$$

stop at the first non-significant result.

Please note that the subscripts in parantheses now characterize a different ordering as in Procedure II. Again with $\eta = 0$, we have the usual unweighted Bonferroni-Holm method but the procedure does not converge to Procedure I for $\eta \to \infty$.

For the proof that the procedure keeps the familywise type I error in the strong sense, we denote the set of all variables with true null hypotheses by M_0, the subscript of that variable from M_0 with the smallest p-value by i_0 and the set of variables containing this variable and all others which follow it in the sorted procedure (according to the denominator of (4)) by S_0. Then $M_0 \subseteq S_0$ and as in the proof in Benjamini, Hochberg (1997),

$$P\left(p_{i_0} \leq \frac{g_{i_0}\,\alpha}{\sum_{i\in S_0} g_i}\right) \leq P\left(p_{i_0} \leq \frac{g_{i_0}\,\alpha}{\sum_{i\in M_0} g_i}\right) \leq \sum_{l\in M_0} P\left(p_l \leq \frac{g_l\,\alpha}{\sum_{i\in M_0} g_i}\right).$$

Let now $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ denote the whole sample matrix and $\mathbf{X}_0$ its submatrix containing only the variables from M_0. Then the theory of spherically

distributed matrices (Fang, Zhang, 1990) states that $\mathbf{X}_0$ is left-spherically distributed and also the conditional distribution of $\mathbf{X}_0$ for fixed $\mathbf{X}_0'\mathbf{X}_0$ is left-spherical. Hence, each of the columns of $\mathbf{X}_0$ (corresponding the single variables from M_0) is also conditionally left-spherically distributed and the one-sample t tests exactly maintain their type I error. Furthermore, for fixed $\mathbf{X}_0'\mathbf{X}_0$, all w_i and hence g_i are fixed ($i \in M_0$), such that the above chain of inequalities can be continued

$$P\left(p_{i_0} \le \frac{g_{i_0}\,\alpha}{\sum_{i\in S_0} g_i}\right) \le \sum_{l\in M_0} P\left(p_l \le \frac{g_l\,\alpha}{\sum_{i\in M_0} g_i}\right) \le \sum_{l\in M_0} \frac{g_l\,\alpha}{\sum_{i\in M_0} g_i} = \alpha\,.$$

As this is valid for each fixed value of $\mathbf{X}_0'\mathbf{X}_0$, it is valid unconditionally, too. Thus, the first true null hypothesis in the ordered sequence of variables is accepted with probability $1-\alpha$, if the procedure did not even stop before. This completes the proof.

In the nonparametric version, we use the weights $g_i = m_i^\eta$ and the p-values of the one-sample Wilcoxon test. The rest is the same as in the parametric procedure. The proof for the familywise type I error control is quite analogous to that for Procedure II in Kropf et al. (2004).

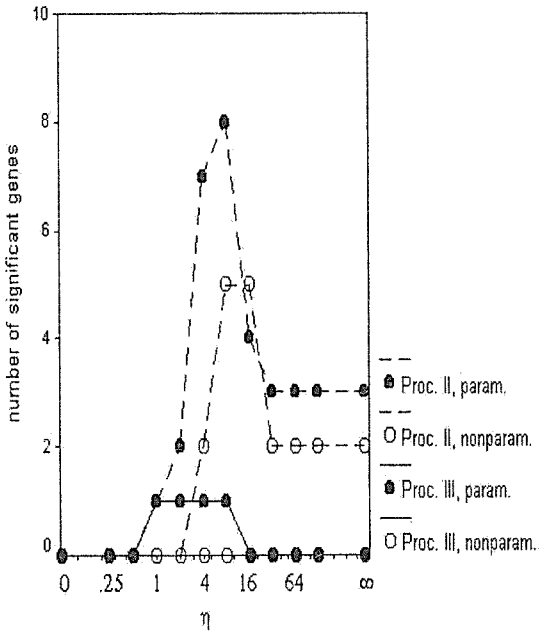

FIGURE 1. Results of the parametric and nonparametric versions of Procedures II and III in the example, depending on the parameter η

Figure 1 shows the number of significant variables for the example with the parametric and nonparametric versions of Procedures II and III. Obviously, Procedure III finds less significant genes in both set-ups. In the parametric version, at least one significant gene is found with intermediate values of η, none is found in the nonparametric version. The differences have not been so drastical in other examples. However, already Benjamini and Hochberg (1997) have noted a reduced power in many cases compared to the weighted procedure as in Procedure II with fixed weights. They also pointed out that Procedure III has a monotonicity-problem: it may happen that a decrease in some or all local p-values can lead to reduced number of significances.

Both Procedures II and III have the common problem that the smoothing parameter has to be fixed in advance. Here, some general experience may help. Large η should be used in very small samples when additionally the variances of the variables can be assumed to be similar. The larger the sample size or the more heterogeneity of variances is expected, the smaller η should be choosen. Furthermore, simulation experiments for the special size of the data matrix may help to fix a suitable η.

6. Smoothed procedure with ordered hypotheses

A very recent proposal by Hommel again utilizes the ordered sequence of hypotheses as in Procedure I but gives the possibility to skip $m-1$ non-significant hypotheses, with a fixed number m.

Procedure IV:

- Sort the p variables $x_1, \ldots, x_p$ for decreasing values $w_i = \sum_{j=1}^{n} x_{ji}^2$,
- in this order carry out the usual one-sample t tests for the variables at the level α/m, where up to $m-1$ non-significant tests may be skipped before the procedure definitely stops.

It can easily be seen that the procedure is identical to Procedure I for $m = 1$ and to the Bonferroni procedure for $m = p$. Intermediate values give a varying level of smoothing. Again, the procedure is based on the theory of spherical distributions in the parametric case and on the independence of rank and order statistics in the nonparametric case. Applying the closure testing principle, a slightly more powerful version can be derived. The detailled description and the proof for the control of the familywise type I error in the strong sense will be given in a forthcoming paper (Hommel, Kropf, 2004).

As an example, we consider again the sorted p-values for the parametric version of Procedure I: .0002, .0006, .0032, .0871, .0254, .2227, .5001, .1334, .9365, .1066, .0858, .0006, .0000, ...With $m = 1$ we could find only 3 significances in Procedure I. If we choose $m = 10$, for example, then the first three p-values are also below the reduced critical value 0.005. Then there is a gap of 8 values which are larger. But now we can utilize also some p-values after

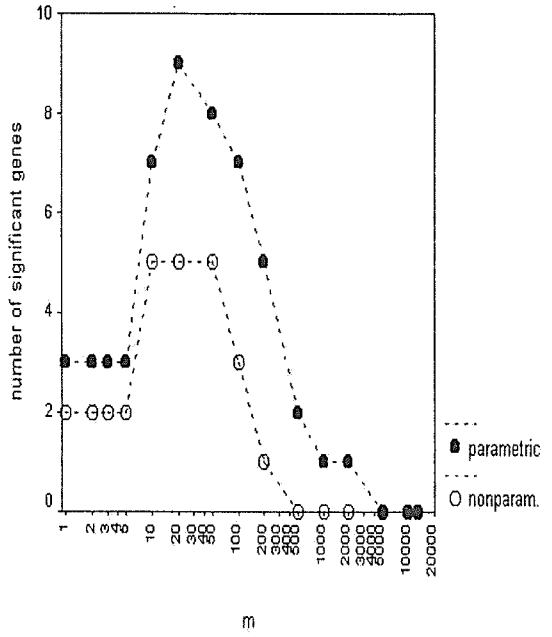

FIGURE 2. Results of the parametric and nonparametric version of Procedure IV for varying m

this gap, giving 7 genes with significant changes. Figure 2 shows the results of Procedure IV in the example for varying values of m. The shape of the curve is similar to that of Procedure II but it is mirrored because large m have similar effects as small η. In the example it seems that Procedure IV has slight advantages compared to Procedure II. The problem to determine m in advance is analogous to Procedures II and III and can be handled in the same way.

References

Bauer, P., Röhmel, J., Maurer, W. and Hothorn, L. A. (1998). Testing strategies in multiple-dose experiments including active control. *Statistics in Medicine* **17**, 2133–2146.

Benjamini, Y. and Hochberg, Y. (1997). Multiple hypothesis testing with weights. *Scand. J. Statist.* **24**, 407–418.

Eszlinger, M., Krohn, K. and Paschke, R. (2001). cDNA expression array analysis suggests a lower expression of signal transduction proteins and receptors in cold and hot nodules. *Journal of Clinical Endocrinology and Metabolism* **86**, 4834–4842.

Fang, K. T. and Zhang, Y. T. (1990). *General Multivariate Analysis*. Science Press, Beijing and Springer-Verlag, Berlin Heidelberg.

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

Hommel, G. and Kropf, S. (2004). Tests for differentiation in gene expression using a data-driven order or weights for hypotheses. (Submitted)

Kropf, S. (2000). *Hochdimensionale multivariate Verfahren in der medizinischen Statistik.* Shaker Verlag, Aachen.

Kropf, S. and Läuter, J. (2002). Multiple tests for different sets of variables using a data-driven ordering of hypotheses, with an application to gene expression data. *Biometrical J.* **44**, 789–800.

Kropf, S., Läuter, J., Eszlinger, M., Krohn, K. and Paschke, R. (2004). Multiple test procedures with data-driven order of hypotheses and with weighted hypotheses. *J. Statist. Plann. Inference* **125**, 31–47.

Läuter, J., Glimm, E. and Kropf, S. (1996). New multivariate tests for data with an inherent structure. *Biometrical J.* **38**, 5–23. Erratum: *Biometrical J.* **40**, 1015.

Läuter, J., Glimm, E. and Kropf, S. (1998). Multivariate tests based on left-spherically distributed linear scores. *Ann. Statist.* **26**, 1972–1988. Correction: *Ann. Statist.* **27**, 1441.

Marcus, R., Peritz, E. and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63**, 655–660.

Westfall, P. H. and Krishen, A. (2001). Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures. *J. Statist. Plann. Inference* **99**, 25–40.

Westfall, P. H., Kropf, S. and Finos, L. (2004). Weighted FWE-controlling methods in high-dimensional situations. In: *Recent developments in multiple comparison procedures*, Eds: Benyamini, Y., Bretz, F. and Sakar, S. K., IMS Lecture Notes and Monograph Series. (In press)

Westfall, P. H. and Young, S. S. (1993). *Resampling Based Multiple Testing.* John Wiley & Sons, New York.

Institute of Biometry and Medical Informatics, Otto von Guericke University Magdeburg, Leipziger Str. 44, 39120 Magdeburg, Germany
E-mail address: Siegfried.Kropf@medizin.uni-magdeburg.de

Institute for Medical Biometry, Epidiomology and Informatics, Johannes Gutenberg University Mainz, D-55101 Mainz, Germany
E-mail address: Hommel@imbei.uni-mainz.de