University of Sydney, School of Mathematics and Statistics,
Research Report 99-9, March,1999.

A STEPWISE REJECTIVE TEST PROCEDURE WITH STRONG CONTROL OF FAMILYWISE ERROR RATE

By T. CHEN[1] and E. SENETA

Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A. and School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia

SUMMARY

In Holm's stepwise rejective procedure the critical values are replaced adaptively by larger ones using a degree 2 inequality in place of Boole's. The refinement retains strong control of familywise error rate. There is a cost in calculational simplicity; but a substantial improvement in actual error rate, according to simulations.
Some key words: Holm's procedure; Degree 2 inequalities; Dependent test statistics; Error rate; Power; Multivariate t-distribution; Strong control; Simes' test.

1. INTRODUCTION

We consider the multiple test problem where there are n hypotheses $H_1, H_2, \cdots, H_n$, and corresponding p-values $R_1, \cdots, R_n$, assuming the test statistics $X_1, \cdots, X_n$ are from a continuous distribution. Suppose that in a multiple test procedure the property

$$P(H_s, s\epsilon I\,,\ \text{are accepted}|H_s, s\epsilon I,\ \text{true}) \geq 1-\alpha \tag{1}$$

holds, for prespecified size of test (familywise error rate) α , where I is any non-null subset of $\{1, 2, \cdots n\}$, and thus contains m items, $1 \leq m \leq n$. Then the procedure is said to control strongly the familywise error rate (Hochberg and Tamhane, 1987, pp. 3, 7).

Let $R_{(1)}, R_{(2)}, \cdots, R_{(n)}$ be the ordered p-values, and $H_{(1)}, H_{(2)}, \cdots, H_{(n)}$ the corresponding hypotheses. The 'Bonferroni' multiple test procedure rejects the composite hypothesis $\{H_{(1)}, H_{(2)}, \cdots, H_{(n)}\}$ if $R_{(1)} \leq \alpha/n$, and accepts it otherwise. This procedure was refined by Holm (1979) as follows. Examine whether $R_{(1)} \leq \alpha/n$; if not, accept $H_{(i)}\,, i = 1, \cdots n$ as with Bonferroni; if so, reject $H_{(1)}$ and examine whether

[1] Work of this author done in part at the University of Sydney

$R_{(2)} \leq \alpha/(n-1)$. If the inequality is not satisfied accept $H_{(2)}, \cdots, H_{(n)}$; otherwise reject $H_{(2)}$. Continue in this way. To summarize: if $R_{(i)} \leq \alpha/(n-i+1),\ i \leq j-1$, then at step j the remaining hypotheses are $H_{(j)}, \cdots, H_{(n)}$ and the inequality next to check is $R_{(j)} \leq \alpha/(n-j+1)$. The process may run at most until a decision is made on the basis of whether $R_{(n)} \leq \alpha$ or not. Holm showed that his procedure strongly controls the familywise error rate. Inasmuch as it essentially depends on Boole's (first Bonferroni) inequality, which is a degree 1 bound (e.g. Seneta, 1988), Holm's procedure retains an elegant simplicity.

There have been a number of improvements on the Bonferroni-Holm degree 1 procedures, all of which are aimed at increasing power while retaining a simple structure of critical points (such as $\alpha/(n-j+1)$ above). Hochberg's (1988) and Hommel's (1988) (1989) procedures both strongly control the familywise error rate providing each Simes (1986) test of $H_I = \{H_i,\ i \epsilon I\}$ is a level α-test. Recently Sarkar (1998) has shown, that the Simes property holds if the random vector $(X_1, \cdots, X_n)$ is from an MTP_2 distribution. This class includes the central multivariate t with the associated correlation matrix having a common and non-negative correlation. Since they are based on the Simes procedure, the step-up procedures of Hochberg/Hommel, now validated in the MTP_2 situation, are more powerful than Holm's (1979) procedure.

In Section 6 of Seneta and Chen (1997), a degree 2 step-down procedure is proposed which retains familywise control of error rate. This procedure is adaptive in that calculation at each step is determined by the joint outcome of all pairs of statistics in the experiment involved until the procedure stops. In view of the continuing interest in a general procedure with familywise control of error rate, we present here a substantial refinement of this procedure, in a form which resembles Holm's. Specifically, the values $\alpha/(n-j+1), j \geq 1$, are replaced by larger ones, thus increasing the power. We present, using simulation, a crude power comparison with the Bonferroni/Holm procedure and with the Hochberg procedure in the setting of multivariate t . A more extensive comparison with Hochberg/Hommel, is indicated, but outside our present intention of presenting briefly the streamlined procedure. By comparison with Bonferroni/Holm and Hochberg/Hommel, there is a loss of computational simplicity. The first step in particular requires the calculation of $n(n-1)/2$ two dimensional probabilities (in specific situations such as Dunnett's test, fewer two dimensional probabilities need to be computed) ; and our procedure may be useful only if n is small. It may, however, be possible to obtain approximations to two dimensional probabilities by bootstrapping in specific situations (these are being investigated), thus eliminating a need to know two dimensional distributions.

PROCEDURE CS

Write for the moment $R_{(i)} = R_{t_i}$, so t_i is a random variable from the set $\{1, 2, \cdots, n\}$. Using the ordered p-values R_{t_i}, $i = 1, \cdots, n$ observed, define the index sets $K(\cdot)$ by $K(p) = \{t_p, t_{p+1}, \cdots, t_n\}$, $p = 1, 2, \cdots, n$ (these are random sets for $p \geq 2$) and write

$$\gamma(p) = \max(j \epsilon K(p)) \sum_{i \epsilon K(p)-\{j\}} P(R_i \leq \frac{\alpha}{n-p+1},\ R_j \leq \frac{\alpha}{n-p+1} | H_s, s \epsilon K(p), \text{true}) \tag{2}$$

for $1 \leq p \leq n-1$, with $\gamma(n) = 0$. These may be calculated for successive p as far as required in what follows.

Step 1 :

$$R_{(1)} \leq \min(\frac{\alpha}{n-1}, \frac{\alpha + \gamma(1)}{n}) \ ?$$

If yes, reject $H_{(1)}$ and go to Step 2. If no, accept $H_{(1)}, H_{(2)}, \cdots, H_{(n)}$ and stop. Continue in this way. If the i-th step is reached:

Step i :

$$R_{(i)} \leq \min(\frac{\alpha}{n-i}, \frac{\alpha + \gamma(i)}{n-i+1}) \ ?$$

If yes, reject $H_{(i)}$ and go to Step $i+1$. If no, accept $H_{(i)}, H_{(i+1)}, \cdots, H_{(n)}$ and stop. If the n-th Step is reached:

Step n :

$$R_{(n)} \leq \alpha \ ?$$

If yes, reject $H_{(n)}$ and stop . If no, accept $H_{(n)}$ and stop. (Note that since $\gamma(n) = 0$, Step n is consistent with the others. Also, at the $(n-1)^{th}$ step, $\min\left(\alpha, \frac{\alpha + \gamma(n-1)}{2}\right) = \frac{\alpha + \gamma(n-1)}{2}$, since $\gamma(n-1) \leq \frac{\alpha}{2}$). $\square$

The procedure is therefore an adaptive one, and the $\gamma(p)'s$ for $2 \leq p \leq n-1$ are random variables.

2. STRONG CONTROL OF FAMILYWISE ERROR RATE

A key feature of the proof of the theorem is the use of the inequality (from which (2) derives) of Kounias (1968)

$$P(\bigcup_{i=1}^{k} A_i) \le \sum_{i=1}^{k} P(A_i) - \max(j=1,\cdots,k) \sum_{i \neq j} P(A_i \bigcap A_j),$$

a second-degree inequality.

LEMMA. *Let* I, *of fixed size* m , *be as in Section 1. Define*

$$\gamma = \max(I) \sum_{j \epsilon I - \{i\}} P(R_i \le \frac{\alpha}{m},\ R_j \le \frac{\alpha}{m} | H_s, s \epsilon I\,, \text{true}),$$

where $\gamma = 0$ *when* $m = 1$. *Then*

$$\bigcap_{i \epsilon I} \left\{ R_i > \ \min \left(\frac{(\alpha+\gamma)}{m}, \frac{\alpha}{m-1} \right) \right\} \subseteq \{H_s, s\epsilon I\,, \text{are all accepted}\}.$$

Proof. For any fixed realization (sample point) of the experiment within

$$\bigcap_{i \epsilon I} \left\{ R_i > \ \min \left(\frac{(\alpha+\gamma)}{m}, \frac{\alpha}{m-1} \right) \right\}, \tag{3}$$

if the $R_i, i\epsilon I$, are in fact the largest m values $R_{(i)}, i = n-m+1, \cdots, n$, then $I = K(n-m+1)$, and $\gamma = \gamma(n-m+1)$, so $H_i, i\epsilon I$, will be accepted by the procedure, by at most the $(n-m+1)^{th}$ Step.

If the $R_i, i\epsilon I$, are not the largest m values, then the smallest value of $R_i, i\epsilon I$, is $R_{(n-m+1-k)}$ for some k, $1 \le k \le n-m$, and on account of (3) either

(a) $R_{(n-m+1-k)} > (\alpha+\gamma)/m$; or (b) $R_{(n-m+1-k)} > \alpha/(m-1)$.

If (a), then $R_{(n-m+1-k)} > \alpha/m \ge \alpha/(m+k-1) = \alpha/(n-(n-m-k+1))$, so $H_{(i)}, n-m-k+1 \le i \le n$, and so $H_i, i\epsilon I$, are all accepted by at most the $(n-m-k+1)^{th}$ Step. If (b), then $R_{(n-m+1-k)} > \alpha/(m-1) > \alpha/m \ge \alpha/(m+k-1)$, with the same conclusion. □

THEOREM: *If the set* $\{H_i, i\epsilon I\}$ *is the set of true hypotheses (where* I *is any non-null subset of* $\{1, \cdots, n\}$), *then* (1) *holds for* PROCEDURE CS.

Proof. For notational convenience put $A = \{H_s,\ s\epsilon I,\ \text{true}\}$.

$$\begin{aligned} & P(H_s, s\epsilon I, \text{are accepted } |A) \\ & \quad \ge P\left(\bigcap_{i\epsilon I} \left\{ R_i > \ \min\left(\frac{(\alpha+\gamma)}{m}, \frac{\alpha}{m-1} \right) \right\} |A \right) \end{aligned}$$

by the Lemma;

$$= 1 - P\left(\bigcup_{i\epsilon I}\left\{R_i \leq \min\left(\frac{(\alpha+\gamma)}{m}, \frac{\alpha}{m-1}\right)\right\}|A\right)$$
$$\geq 1 - \sum_{i\epsilon I} P\left(R_i \leq \min\left(\frac{(\alpha+\gamma)}{m}, \frac{\alpha}{m-1}\right)|A\right)$$
$$+ \max(i\epsilon I) \sum_{j\epsilon I-\{i\}} P\left(R_i \leq \min\left(\frac{(\alpha+\gamma)}{m}, \frac{\alpha}{m-1}\right), R_j \leq \min\left(\frac{(\alpha+\gamma)}{m}, \frac{\alpha}{m-1}\right)|A\right)$$

by Kounias' inequality;

$$\geq 1 - \min((\alpha+\gamma), m\alpha/(m-1)) + \max(i\epsilon I) \sum_{j\epsilon I-\{i\}} P(R_i \leq \frac{\alpha}{m}, R_j \leq \frac{\alpha}{m}|A)$$

by the uniformity of distribution of each $R_i \epsilon I$, since $\min((\alpha+\gamma)/m,\ \alpha/(m-1)) < 1$;

$$= 1 - \min((\alpha+\gamma), m\alpha/(m-1)) + \gamma$$
$$\geq 1 - (\alpha+\gamma) + \gamma = 1 - \alpha\ .$$

COROLLARY. *For any given* I , *if* $\gamma \geq \alpha/(m-1)$, *then*

$P(H_s, s\epsilon I,$ are accepted $|A) \geq 1 - m\alpha/(m-1) + \gamma$. □

It will be seen that the test procedure could have been constructed, with increase of complexity at (2), round the sharper inequality of Hunter (1976), of which there are non-graph theoretic formulations (Margaritescu, 1986; Seneta, 1988). Possible use of degree 2 inequalities such as ours is mentioned in this context by Simes (1986, Section 5) and Shaffer (1986), and use of them for confidence intervals is made by Stoline (1983) (see Seneta, 1993).

The quantity γ defined in the LEMMA above is required for the proof of the THEOREM following it, which establishes strong control of the adaptive test procedure, but is not needed in the adaptive test procedure itself. Note that $\gamma \leq (m-1)\alpha/m \leq \alpha$.

3. EXAMPLE AND SIMULATIONS

We shall measure power by

$$P(\text{Reject at least one} \quad H_i,\ i = 1, \cdots, n).$$

This has the advantage that when all of the $H_i,\ i = 1, \cdots, n$ hold, from (1) this value will be $\leq \alpha$, and its closeness to the nominal error α will measure the actual conservativeness of the error rate. According to the COROLLARY above, if we take $I = \{1, 2, \cdots, n\}$ then $\gamma > \alpha/(n-1)$ results in a bound $< \alpha$. This suggests that the degree of conservativeness of the PROCEDURE CS is related to the strength of positive association between the R_i's (and hence of X_i's) from the definition of γ . This is confirmed by Table 1 below.

We take the test statistics to be exchangeable under corresponding null hypotheses, so from (2)

$$\gamma(p) = (n-p)P(R_1 \leq \frac{\alpha}{n-p+1},\ R_2 \leq \frac{\alpha}{n-p+1}), \tag{4}$$

which is thus non-random in this special setting. More specifically we consider upper-tail tests where $X_i = |T_i|,\ i = 1, 2, \cdots, n$ with $T_1, T_2, \cdots, T_n$ defined by $T_i = W_i/\sqrt{\chi^2_\nu/\nu},\ i = 1, 2, \cdots, n$ where the W_i's are multivariate normal with $EW_i = \mu_i$, $\mathrm{Var}(W_i) = 1,\ i = 1, 2, \cdots, n$, $\mathrm{Corr}(W_i, W_j) = \rho, \quad i \neq j$, and are independently distributed of χ^2_ν . Thus under $H_i : \mu_i = 0,\ i = 1, \cdots, n$, the $T_i,\ i = 1, \cdots, n$ have jointly a multivariate exchangeable t distribution with parameters $n,\ \rho(\rho \geq -1/(n-1)), \nu$ as in Dunnett's tests (Miller, 1981). We take $n = 3, \nu = 16, \alpha = 0.05$, and consider $0 \leq \rho \leq 1$. We can calculate from tables giving upper-tail values $P(T_1 \leq a,\ T_2 \leq a)$ for various a and $\rho = 0,\ \pm 0.1, \cdots, \pm 0.9$ (Krishnaiah, Armitage and Breiter, 1969) our values of $\Delta(1) = (\alpha + \gamma(1))/n$. Some of these are shown in Table 1. Notice that in our setting $\Delta(1) > \alpha/2 = 0.025$ at $\rho = 1$, but is < 0.025 for $\rho \leq 0.9$. Our measure of power when $\rho \leq 0.9$ (in fact for ρ upto approximately 0.95) is thus $P(R_{(1)} \leq \Delta(1))$, whereas for the Bonferroni-Holm procedure the measure of power is $P(R_{(1)} \leq \alpha/n)$, which is smaller; and remains smaller than our measure of power for ρ very close to 1 viz. $P(R_{(1)} \leq \alpha/(n-1))$. (A more sensitive measure of power would separate out Holm from Bonferroni.) Table 2 displays the power at $\rho = 0.9$ when $\mu_1 = \delta, \mu_2 = 2|\delta|, \mu_3 = 3|\delta|$ for the Bonferroni-Holm, Hochberg and CS procedures.

Table 1. Values of $\Delta(1)$ and Error Rate (E.R.)

$(n=3,\ \nu=16,\ \alpha=0.05)$

ρ	0	0.5	0.8	0.9	1
$\Delta(1)$	0.0171	0.0183	0.0211	0.0228	0.0278
E.R. (α/n)	0.049	0.042	0.033	0.028	
E.R. $(\Delta(1))$	0.050	0.046	0.042	0.040	

Table 2. Power at $\rho=0.9\,(n=3,\ \nu=16,\ \alpha=0.05)$

δ	-1	-0.5	0	0.5	1
α/n	0.685	0.184	0.028	0.159	0.633
Hochberg	0.689	0.187	0.034	0.164	0.634
$\Delta(1)$	0.749	0.227	0.040	0.193	0.685

The error rate (E.R.) entries in Table 1 were produced from a simulation of 20,000 independent sets of values of the triple $T_i,\ i=1,2,3$ at each ρ. These values for $\rho=.9$ are given again in Table 2 at $\delta=0$. The other values of Table 2 were also produced from 20,000 triples.

Overall, the simulations support a conclusion that our proposed procedure is most effective as regards power when test statistics are strongly positively dependent. The error rate is closer to the nominal value α irrespective of degree of dependence, and is not much affected by it. The indication is that Procedure CS controls error rate well, and has significantly better power than Hochberg.

While our procedure may be useful only for small n (1) holds without any restriction on the continuous joint distribution of test statistics.

Finally, our computational results on E.R. are consistent with those of Sarkar and Chang (1997, Table 2), inasmuch as

$$P(R_{(i)} \geq \frac{\alpha}{n-i+1},\ 1 \leq i \leq n) \geq P(R_{(i)} \geq \frac{i\alpha}{n},\ 1 \leq i \leq n)$$

ACKNOWLEDGEMENT

E.S. wishes to thank Harry Khamis and Barbara Mann for helping to motivate this work. The simulations were mostly done by Greta Lai in the course of an MA project during 1997.

REFERENCES

HOCHBERG, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**, 800-802.

HOCHBERG, Y. and ROM, D. (1995). Extensions of multiple testing procedures based on Simes' test. *J. Statist. Plan. Inf.* **48**, 141-152.

HOCHBERG, Y. and TAMHANE, A.C. (1987). *Multiple Comparison Procedures.* New York: Wiley.

HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65-70.

HOMMEL, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75**, 383-386.

HOMMEL, G. (1989). A comparison of two modified Bonferroni procedures. *Biometrika* **76**, 624-625.

HUNTER, D. (1976). An upper bound for the probability of a union. *Journal of Applied Probability* **13**, 597-603.

KOUNIAS, E.G. (1968). Bounds for the probability of a union of events, with applications. *Ann. Math. Statist.* **39**, 2154-2158.

KRISHNAIAH, P.R., ARMITAGE, J.V. and BREITER, M.C. (1969). *Tables for the Probability Integrals of the Bivariate t Distribution.* ARL 69-0060. Applied Mathematics Research laboratory, Aerospace Research Laboratories, Ohio: Wright Patterson AFB.

MARGARITESCU, E. (1986). A note on Bonferroni's inequalities. *Biometrical Journal* **28**, 937-943.

MILLER, R.G. (1981). *Simultaneous Statistical Inference*, 2nd Edition, New York: Springer.

ROM, D.M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika* **77**, 663-666.

SAMUEL-CAHN, E. (1996). Is Simes improved Bonferroni procedure conservative? *Biometrika* **83**, 928-933.

SARKAR, S.K., (1998). Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *Ann. Statist.*, **26**, 494-504.

SARKAR, S.K. and CHANG,C.K. (1997) The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Amer. Statist. Assoc.*, **92**,1601-1608.

SENETA, E. (1988). Degree, iteration and permutation in improving Bonferroni-type bounds. *Australian J. Statist.* **30A**, 27-38.

SENETA, E. (1993). Probability Inequalities and Dunnett's Test. In F.M. Hoppe, Ed: *Multiple Comparisons, Selection, and Applications in Biometry* Chapter 3, pp 29-45. New York: Marcel Dekker.

SENETA, E. and CHEN, T.H. (1997). A sequentially rejective test procedure. *Theory of Stochastic Processes* **4** (to appear).

SHAFFER, J.P. (1986). Modified sequentially rejective multiple test procedure. *J. American Statistical Association* **81**, 826-831.

SIMES, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751-754.

STOLINE, M.R. (1983). The Hunter method of simultaneous inference and its recommended use for applications having large known correlation structures, *Journal of the American Statistical Association* **78**, 366-370.

WRIGHT, S.P. (1992). Adjusted P-Values for simultaneous inference. *Biometrics* **48**, 1005-1013.