

Excerpt from Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS 2002 *Truncated product method for combining p-values*. Genetic Epidemiology 22: 170–185

## Independence case

As a procedure that combines the features of Fisher’s product method and Wilkinson’s truncation method, we suggest the use of the product  $W$  of all those  $p_i$  values that do not exceed some fixed value  $\tau$ :

$$W = \prod_{i=1}^L p_i^{I(p_i \leq \tau)} \quad (1)$$

where  $I(\cdot)$  is the indicator function and  $L$  is the number of tests.

Useful aspects of the TPM test include the following:

1. Experience shows that the ordinary Fisher product test loses power in cases where there are a few large p-values. This can happen when tests are one-sided, with noncentrality in the “wrong” direction, or when there are a predominance of near-null effects. By truncating, these large components are removed, thereby providing more power, much like a “trimmed mean” gains efficiency in the presence of outliers [Huber, 1977].
2. A natural, although arbitrary choice for  $\tau$  is  $\alpha$  (commonly 0.05). This allows easy use of the test in cases where only p-values for “statistically significant” results are given; it also allows estimation of “file drawer effects” in meta-analysis as shown below.
3. The truncated combination emphasizes smaller p-values, somewhat like the Simes and Šidák methods. However, Simes and Šidák p-values can never be smaller than  $p_{(1)}$ , the smallest p-value, whereas the TPM p-value will be smaller than  $p_{(1)}$  when there are several small and reinforcing p-values in the set. In genome scans, this case is likely to occur in a local neighborhood of a susceptibility gene.
4. One can incorporate weights into the analysis as in Good [1955], as  $W = \prod_{i=1}^L p_i^{w_i I(p_i \leq \tau)}$ , thereby allowing studies or tests with more precision to play a larger role.

5. Isolation of individual significances is possible (and computationally feasible even for large numbers of tests as we show below in “Closed testing with Truncated Fisher Test”), through the closure principle of Marcus et al, [1976] when using the TPM test.

Consider the case when all  $p$ -values are independent. Under the null hypothesis  $H_T$ , the distribution of  $W$  for  $w < 1$  can be evaluated by conditioning on the number,  $k$ , of the  $p_i$ 's less than  $\tau$ :

$$\begin{aligned} \Pr(W \leq w) &= \sum_{k=1}^L \Pr(W \leq w \mid k) \Pr(k) \\ &= \sum_{k=1}^L \binom{L}{k} (1 - \tau)^{L-k} \\ &\quad \times \left( w \sum_{s=0}^{k-1} \frac{(k \ln \tau - \ln w)^s}{s!} I(w \leq \tau^k) + \tau^k I(w > \tau^k) \right) \end{aligned} \quad (2)$$

When  $L$  is large ( $> 1,000$  tests, assuming double precision calculations), the probability in (2) should be computed through a Monte Carlo algorithm described in the next section. The derivation of equation (2) is given below (“Distribution of  $W$ ” section). At one extreme, setting  $\tau = \min p$  results in Šidák’s correction. At the other extreme, when  $\tau = 1$ , equation (2) provides Fisher’s combined  $p$ -value. Thus, the method we describe here is “intermediate” between combination and individual adjustment techniques. Note that setting  $\tau = 1$  provides a way of calculating Fisher’s combined  $p$ -value directly. Instead of looking up the cumulative probability from the tail of a chi-square distribution, it can be obtained as

$$\Pr(W \leq w) = w \sum_{s=0}^{L-1} \frac{(-\ln w)^s}{s!} \quad (3)$$

C++ code for calculating the TPM  $p$ -value is available at

<http://statgen.ncsu.edu/zaykin/tpm/> . An executable for a specific OS can be requested from Dmitri Zaykin (zaykind@niehs.nih.gov).

## Distribution of $W$

When  $H_T$  is true and  $\tau < 1$ , the number of small  $p$ -values ( $k$ ) has a binomial distribution, and  $p_i$ 's are observations from the uniform  $(0, 1)$  distribution, truncated at  $\tau$  (i.e. the distribution of  $p_i$ 's is uniform on  $(0, \tau)$ ).

Given  $k$ , the conditional distribution of the product ( $W$ ) can be calculated directly. Let  $X_1, \dots, X_k$  be independent uniform  $(0, \tau)$  random variables. Consider the transformation:

$$\begin{aligned} Z_1 &= X_1 \\ Z_2 &= X_1 X_2 \\ &\dots \\ Z_k &= X_1 X_2 \dots X_k \end{aligned}$$

with inverse

$$\begin{aligned} X_1 &= Z_1 \\ X_2 &= Z_2/Z_1 \\ &\dots \\ X_k &= Z_k/Z_{k-1} \end{aligned}$$

The Jacobian of the transformation ( $\mathbf{J}$ ) has the following structure:

$$\partial x_i / \partial z_j = \begin{cases} 1 & i = j = 1 \\ 1/z_{i-1} & i = j; \geq 2 \\ -z_i/z_{i-1}^2 & j = i - 1 \\ 0 & \text{otherwise} \end{cases}$$

Therefore

$$|\mathbf{J}| = \prod_{i=1}^{k-1} 1/z_i$$

and the joint density is

$$f(\mathbf{Z}) = \frac{1}{\tau^k \prod_{i=1}^{k-1} z_i}$$

Integrating out  $z_1$  through  $z_{k-1}$  from the joint density gives the conditional probability,  $P(W \leq w | k)$ :

$$\begin{aligned} \Pr(W \leq w | k) &= \int_0^w \left[ \int_t^{\tau^k} \int_{z_{k-1}}^{\tau^k} \dots \int_{z_2}^{\tau^k} \frac{\prod_{i=1}^{k-1} dz_i}{\tau^k \prod_{i=1}^{k-1} z_i} \right] dt \\ &= \int_0^w \frac{(\ln \tau^k - \ln t)^{k-1}}{(k-1)! \tau^k} dt \end{aligned} \tag{4}$$

Then the unconditional distribution is found as follows:

$$\begin{aligned} \Pr(W \leq w) &= \int_0^w \sum_{k=1}^L \frac{(\ln \tau^k - \ln t)^{k-1}}{(k-1)! \tau^k} \\ &\quad \times I(\ln \tau^k > \ln t) \binom{L}{k} \tau^k (1-\tau)^{L-k} dt \end{aligned} \quad (5)$$

The probability calculated in (5) corresponds to the combined  $p$ -value. After  $\tau^k$  in (5) is canceled, this probability is

$$\begin{aligned} \Pr(W \leq w) &= \int_0^w \sum_{k=1}^L \frac{(k \ln \tau - \ln t)^{k-1}}{(k-1)!} \\ &\quad \times I(\tau^k > t) \binom{L}{k} (1-\tau)^{L-k} dt \end{aligned} \quad (6)$$

or equivalently

$$\begin{aligned} \Pr(W \leq w) &= \sum_{k=1}^L \binom{L}{k} \frac{(1-\tau)^{L-k}}{(k-1)!} \left[ \int_0^w (k \ln \tau - \ln t)^{k-1} \right. \\ &\quad \left. \times I(\tau^k > t) dt \right] \end{aligned} \quad (7)$$

Provided  $\tau^k > t$ , the integral in (7), which we denote by  $I_k$  is:

$$I_k = \int_0^w (\ln \tau^k - \ln t)^{k-1} dt \quad (8)$$

$$\begin{aligned} &= (\ln \tau^k - \ln t)^{k-1} t \Big|_0^w \\ &\quad - \int_0^w t d[(\ln \tau^k - \ln t)^{k-1}] \end{aligned} \quad (9)$$

$$\begin{aligned} &= w(\ln \tau^k - \ln w)^{k-1} \\ &\quad - (k-1) \int_0^w t \left(-\frac{1}{t}\right) (\ln \tau^k - \ln t)^{k-2} dt \end{aligned} \quad (10)$$

$$= (k-1) I_{k-1} + w A(\tau, k, w)^{k-1}. \quad (11)$$

where  $A(\tau, k, w) = k \ln \tau - \ln w$ . Since  $I_1 = w$ , then

$$I_k = (k-1)! \left[ w + w \sum_{s=1}^{k-1} \frac{A(\tau, k, w)^s}{s!} \right] \quad (12)$$

$$= w(k-1)! \sum_{s=0}^{k-1} \frac{A(\tau, k, w)^s}{s!} \quad (13)$$

Therefore,

$$\Pr(W \leq w) = w \sum_{k=1}^L \binom{L}{k} \frac{(1-\tau)^{L-k}}{(k-1)!} (k-1)! \sum_{s=0}^{k-1} \frac{A(\tau, k, w)^s}{s!} \quad (14)$$

$$= w \sum_{k=1}^L \binom{L}{k} (1-\tau)^{L-k} \sum_{s=0}^{k-1} \frac{A(\tau, k, w)^s}{s!} \quad (15)$$

$$= w \sum_{k=1}^L \sum_{s=0}^{k-1} \binom{L}{k} (1-\tau)^{L-k} \frac{A(\tau, k, w)^s}{s!} \quad (16)$$

### Closed testing with Truncated Fisher Test

Adjustments for subsets of hypotheses and individual adjustments are available through the application of the closure principle of Marcus et al. (1976). Generally, the procedure considers all possible combination hypotheses obtained via the intersection of the set of individual hypotheses of interest. If an individual hypothesis and all intersections that contain it as a component are rejected by an appropriate  $\alpha$ -level test, then the closure principle states that the given hypothesis can be also rejected, at the level  $\alpha$ . The closure procedure controls the family-wise error rate (FWER) strongly, meaning that  $\text{FWER} \leq \alpha$  regardless of which subset of null hypotheses happens to be true (Hochberg and Tamhane, 1987). The total number of combination hypotheses ( $N_h$ ) is

$$N_h = \sum_{i=1}^L \binom{L}{i} = 2^L - 1 \quad (17)$$

which grows quickly with  $L$  and often limits applicability of the method.

Fortunately, this is not the case for the TPM test. Noting that (2) is an increasing function of  $L$  and a decreasing function of  $W$ , we see that, among all intersections of a given size  $s$  (where  $s \leq L$ ) that include  $H_i$ , only the combination that includes  $H_i$  and the remaining  $s - 1$  largest  $p$ -values needs to be tested. Thus, significance for any given hypothesis can be determined using  $L$  tests; and when all  $L$  component tests are considered, the maximum number of evaluations is  $L^2$ . However, many of these evaluations are redundant, and in practice the number is less than  $L^2$ . In some cases, e.g., with the Šidák combined tests, the number of evaluations is as small as  $L$ . To illustrate this argument, consider the case when  $\tau = 1$ . Then for the ordered set of  $p_j$ 's ( $j = 1, \dots, L$ ) the adjusting procedure for any subset of  $p$ -values,  $\mathcal{P}_i$ , such that  $p_{(i)}$  is the largest

$p$ -value in the set, is as follows. Compute (3) for most stringent subsets:

$$\{\mathcal{P}_i, p_{(L)}\}$$

$$\{\mathcal{P}_i, p_{(L)}, p_{(L-1)}\}$$

$$\{\mathcal{P}_i, p_{(L)}, p_{(L-1)}, p_{(L-2)}\}$$

...

$$\{\mathcal{P}_i, p_{(L)}, p_{(L-1)}, p_{(L-2)}, \dots, p_{(i+1)}\}$$

The adjusted  $p$ -value for the subset  $\mathcal{P}_i$  is given by the maximum of these values. Many subsets, such as, for example,  $\{\mathcal{P}_i, p_{(L-1)}\}$  do not need to be considered, because they will yield  $p$ -values smaller than the one for  $\{\mathcal{P}_i, p_{(L)}\}$ .

## REFERENCES

- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society series B.* 57:289–300.
- Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. 1994. *Genetics.* 138:963-971
- Doerge RW, Churchill GA. Permutation tests for multiple loci affecting a quantitative character. 1996. *Genetics.* 142:285-294.
- Edgington ES. 1972. An additive method for combining probability values from independent experiments. *J Psychology* 80:351–363.
- Feller W. 1966. *An Introduction to Probability Theory and its Applications.* Wiley and Sons, New York.
- Fisher RA. 1932. *Statistical Methods for Research Workers.* Oliver and Boyd, London.
- Givens GH, Smith DD, Tweedie RL. 1997. Publication bias in meta-analysis: A Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate. *Statistical Science* 12: 221–240.
- Goldin LR, Chase GA, Wilson AF. 1999. Regional inference with averaged  $p$ -values increases the power to detect linkage. *Genetic Epidemiology* 17:157–164.
- Good IJ. 1955. On the weighted combination of significance tests. *Journal of the Royal Statistical Society series B* 17:264–265.
- Goodman SN. 1999. Toward evidence-based medical statistics. I: The  $P$ -value fallacy. *Annals of Internal Medicine* 130:995–1004.
- Hedges LV, Olkin I 1985. *Statistical Methods for Meta-Analysis.* Academic Press, New York.
- Hochberg Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75:800–802.

- Hochberg Y, Tamhane AC 1987. Multiple Comparison Procedures. Wiley and Sons, NY.
- Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6:65–70.
- Huber P. 1977. Robust statistical procedures. SIAM Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Juo SH, Beaty TH, Duffy DL, Maestri NE, Prenger VL, Zeiger J, Lei HH, Coresh J. 1997. A comprehensive analysis of complex traits in problem 2A. *Genetic Epidemiology* 14:815–820.
- Kaplan, NL, Weir BS. 1995. Are moment bounds on the recombination fraction between a marker and a disease locus too good to be true? Allelic association mapping revisited for simple genetic diseases in the Finnish population. *Am J Human Genetics* 57:1486–1498.
- Knapp M. 1998. Discriminating between true and false-positive peaks in a genomewide linkage scan, by use of the peak length. *Am J Hum Genetics*, 62:1561–1562.
- Lander ES, Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199.
- Marcus R, Peritz E, Gabriel KR. 1976. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 63:655–660.
- Pallini A. 1994. Bahadur exact slopes for a class of combinations of dependent tests. *Metron* 52: 53–65.
- Rawlings JO. 1988. *Applied Regression Analysis*. Wadsworth Inc., California.
- Rice WR. 1990. A consensus combined  $p$ -value and the family-wide significance of component tests. *Biometrics* 46:303–308.
- Rosenthal R. 1978. Combining results of independent studies. *Psychological Bull* 85:185–193.
- Shaffer JP. 1995. Multiple hypothesis testing: A review. *Ann Rev Psychology* 46:561–584.
- Simes RJ. 1986. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*



73:751–754.

Šidák, Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* 78:626–633.

Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RM, Jr. 1949. *The American Soldier, Vol. 1. Adjustment During Army Life*. Princeton Univ. Press, Princeton.

Terwilliger JD, Shannon WD, Lathrop GM, Nolan JP, Goldin LR, Chase GA, Weeks DE. 1997. True and false positive peaks in genomewide scans: applications of length-biased sampling to genome mapping. *Am J Hum Genetics* 61:430–438.

Westfall PH, Young SS. 1993. *Resampling-Based Multiple Testing*. Wiley, New York.

Wilkinson B. 1951. A statistical consideration in psychological research. *Psychological Bull.* 48:156–158.

Wjst M, Immervoll T. 1998. An internet linkage and mutation database for the complex phenotype asthma. *Bioinformatics* 14:827–828.

Zaykin D, Zhivotovsky L, Weir BS. 1995. Exact tests for association between alleles at arbitrary numbers of loci. *Genetica* 96:169–178.