Journal of Statistical Planning and Inference 78 (1999) 317–323

journal of statistical planning and inference

A Bayesian method for combining statistical tests [1]

James A. Koziol[a, *], Henry C. Tuckwell[b]

[a] *Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10666 N. Torrey Pines Road, La Jolla, CA 92037, USA*

[b] *Epidemiologie et Sciences de l'Information, Université Paris 6, INSERM U444, 27 rue Chaligny, 75012 Paris, France*

Abstract

Within the context of Lipták's (1958, Magyar. Tud. Akad. Mat. Kutato Int. Közl. 3, 171–197) formulation of the problem of combining independent p-values, a class of Bayes tests is constructed. Fisher's well-known combination procedure, based on the product of the p-values, is found to be a Bayes test in this setting with a noninformative prior. Good's weighted version of Fisher's procedure is shown to be an excellent approximation to other Bayes tests.

AMS classification: primary 62C10; secondary 62G10; 62A20

Keywords: Combining independent p-values; Fisher's combination procedure; Meta-analysis

Consider a set of n experiments in which the statistical hypotheses are interconnected: either the null hypothesis is true in each experiment or the alternative obtains in each case. Each experiment is evaluated by a test statistic, with the statistics being mutually independent random variables. It is desired to combine the levels of the individual statistics into one overall level summarizing the experiments.

In the ith experiment, $i = 1, \ldots, n$, let θ_i be an unknown parameter in some metric space Θ_i, and let T_i be the chosen statistic for testing the null hypothesis

$$\mathrm{H}_{i0}: \theta_i \in \Theta_{i0} \qquad (1)$$

versus

$$\mathrm{H}_{i1}: \theta_i \in \Theta_{i1},$$

where

$$\Theta_i = \Theta_{i0} \cup \Theta_{i1}, \quad \Theta_{i0} \cap \Theta_{i1} = \emptyset.$$

[1] Supported in part by Grant RR-00833 from the National Institutes of Health.

* Corresponding author. Tel.: +1-619-554-8247; e-mail: koziol@scripps.edu

0378-3758/99/$ – see front matter
PII: S0378-3758(98)00222-5

Assume that large values of T_i lead to rejection of H_{i0}, and that each T_i is an unbiased and similar test statistic. Let L_i denote the level attained by $T_i; L_i$ is sometimes called the p-value associated with T_i. It is well known that, in the null case, L_i is always stochastically larger than a uniform random variable on (0,1), that is,

$$P_{\theta_i}(L_i \leqslant \alpha) \leqslant \alpha, \quad \theta_i \in \Theta_{i0},\ 0 < \alpha < 1 \tag{2}$$

(Lipták, 1958). Inequality (2) can be taken as the starting point for determining exact slopes of test statistics (Bahadur, 1967, 1971). Exact slopes quantify the typical exponential rate of decrease to zero of L_i under the alternative: that is, the statistic T_i, based on a sample of size n_i, is said to have exact slope $c_i(\theta_i)$, $0 < c_i < \infty$, for a given non-null $\theta_i \in \Theta_{i1}$, if

$$n_i^{-1} \log L_i \to -c_i(\theta_i)/2 \tag{3}$$

with probability one (Bahadur, 1967, 1971; Lambert and Hall, 1982). (Note that the rate at which L_i tends to zero when a non-null θ_i obtains is considered by Bahadur (1960, 1967, 1971) as a measure of the asymptotic efficiency of the sequence of test statistics T_i against that θ_i.) In terms of the parameters, the overall null hypothesis may be formulated as

$$H_0 : (\theta_1, \ldots, \theta_n) \in \Theta_0 = \Theta_{10} \times \cdots \times \Theta_{n0} \tag{4}$$

versus the alternative

$$H_1 : (\theta_1, \ldots, \theta_n) \in \Theta_1 = \Theta_{11} \times \cdots \times \Theta_{n1}.$$

Given the levels L_i for the various experiments, a combination procedure is a method of combining the results to give an overall statistic which can be used to judge the diverse results in a unified way. The well-known non-parametric combination procedures (e.g., Fisher, Tippett) constitute methods for combining independent random variables on the unit interval. Optimality of the combination procedures can be approached through admissibility or monotonicity properties, derived from first principles (Birnbaum, 1954). An alternative approach is adopted here, using a Bayesian formulation in order to derive optimal combination procedures.

Let the levels L_i be represented as independent, (0,1)-valued observations X_i, and let $Y_i = 1 - X_i$. In order to derive a class of Bayes tests for the combination problem (4), we begin with specification of a family of distributions for the levels, and subsequently a prior distribution on that family, as suggested from both finite-sample and asymptotic considerations.

Our principal motivation devolves from the work of Pearson (1939), who provides a finite-sample approach to the distribution of levels. Now, the Y_i can be determined directly from the probability integral transformations of the test statistics T_i. Hence, we can write the densities f_i of the Y_i as

$$f_i(y|\theta_{i1}) = \left.\frac{g(t_i|\theta_{i1})}{g(t_i|\theta_{i0})}\right|_{t=G_{i0}^{-1}(y)}, \tag{5}$$

where $g(\cdot|\theta_{ij})$ is the density of T_i for $\theta_{ij} \in \Theta_{ij}$, $j = 0, 1$, and G_{i0} is the cumulative distribution function of T_i under $\theta_{i0} \in \Theta_{i0}$. In the normal case, at least (that is, T_i remains normally distributed under both H_{i0} and H_{i1}), Pearson noted that the distributions of Y_i can be closely approximated by Pearson Type I (beta) distributions.

The remarkable accuracy of beta approximations to the alternative distributions of the levels in the normal case is readily illustrated. Let $T_i \sim T$, where T is normally distributed with mean μ and standard deviation σ (that is, $T \sim \mathrm{N}(\mu, \sigma)$). Pearson (1939) considers the following testing problems:

Case 1: $H_0{:}T \sim \mathrm{N}(0,1)$ vs. $H_1{:}T \sim \mathrm{N}(0.5,1)$.
Case 2: $H_0{:}T \sim \mathrm{N}(0,1)$ vs. $H_1{:}T \sim \mathrm{N}(0,\frac{3}{2})$.
Case 3: $H_0{:}T \sim \mathrm{N}(0,1)$ vs. $H_1{:}T \sim \mathrm{N}(0,\frac{2}{3})$.

For each of these cases, $Y = \Phi(T), \Phi$ denoting the cumulative distribution function of the N(0,1) distribution. In Fig. 1 we plot the exact density of Y under H_1 as determined from Eq. (5), as well as an approximating beta density for Pearson's three cases. Clearly, the beta approximations here are quite good.

Note that Lipták's formulation of the combination problem is akin to Pearson's case I, testing normality against a one-sided alternative of a shift in mean. The prior distribution for Y_i from Eq. (6) below is identical to Pearson's Eq. (19), his approximation to Eq. (5) in the case I setting.

We thus specify the density functions of the X_i as

$$f_i(x|\theta_i) = (1-\theta_i)x^{-\theta_i}, \quad 0 < x < 1,\ \theta_i \in [0,1),\ 1 \leqslant i \leqslant n. \tag{6}$$

Under H_{i0}, $\theta_i \in \Theta_{i0} = \{0\}$ and under the alternative, $\theta_i \in \Theta_{i1} = (0,1)$. The reduction to a simple null hypothesis follows from Eq. (2), since the least favorable distribution of the levels under H_0 would be the uniform distribution on (0,1), and the T_i are similar. The beta class of priors (6) forms a class of Lehmann alternatives to uniformity, thereby preserving the unbiasedness. Let us also mention an asymptotic consideration, namely, that class (6) is suggested from Eq. (3) to represent the exponential convergence of the levels to 0 under the alternative (Bahadur, 1960).

Next, consider the family of beta priors

$$g(\theta_i; a_i, b_i) = \frac{1}{B(a_i, b_i)} \theta_i^{a_i - 1}(1-\theta_i)^{b_i - 1}, \quad a_i > 0,\ b_i > 0,\ i = 1, \ldots, n \tag{7}$$

for $\theta_i \in \Theta_{i1}$. Now,

$$\begin{aligned} &\int_{\Theta_1} \prod_{i=1}^{n} f_i(x_i|\theta_i) g(\theta_i; a_i b_i)\,\mathrm{d}\theta_i \\ &\quad = \prod_{i=1}^{n} \int_0^1 \frac{1}{B(a_i, b_i)} \theta_i^{a_i - 1}(1-\theta_i)^{b_i} \exp(-\theta_i \log x_i)\,\mathrm{d}\theta_i \\ &\quad = \prod_{i=1}^{n} \frac{1}{B(a_i, b_i)} B(a_i, b_i + 1) M(a_i, a_i + b_i + 1; -\log x_i), \end{aligned} \tag{8}$$

here, $M(a, b; x)$ is Kummer's function, a confluent hypergeometric function with series expansion

$$M(a, b; x) = 1 + \frac{ax}{b} + \frac{(a)_2}{(b)_2}\frac{x^2}{2!} + \cdots + \frac{(a)_n}{(b)_n}\frac{x^n}{n!} + \cdots, \tag{9}$$

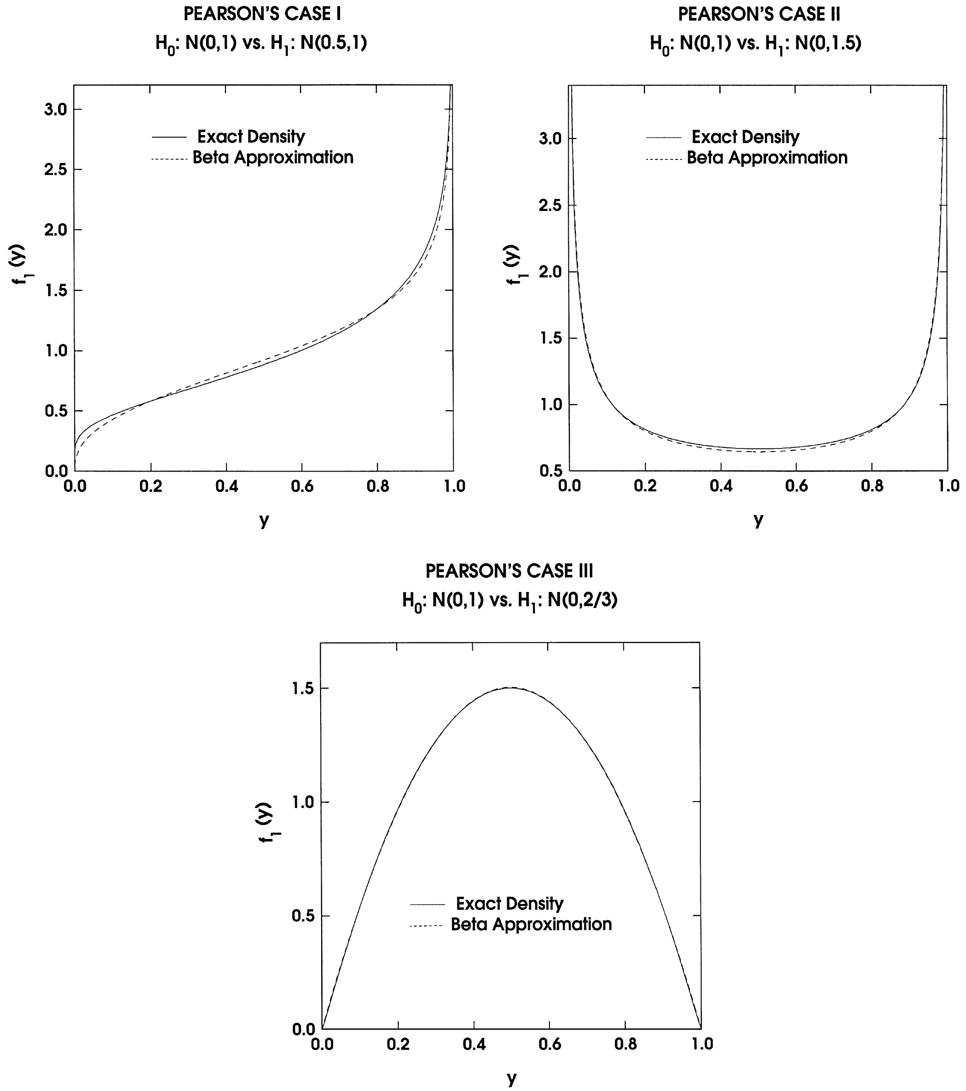

Fig. 1. Consider the testing problem $H_0: T \sim N(0,1)$ vs. $H_i: T \sim N(\mu,\sigma)$, $(\mu,\sigma) \neq (0,1)$. Let $Y = \Phi(T)$, $\Phi(.)$ denoting the standard normal cumulative distribution function. The exact probability density function of Y is illustrated for: Case 1, $T \sim N(0.5, 1)$; Case 2, $T \sim N(0, \frac{3}{2})$; Case 3, $T \sim N(0, \frac{2}{3})$. Approximating beta densities are also depicted. See Pearson (1939) and the text for further details.

where

$$(a)_n = a(a+1)\cdots(a+n-1), \quad (a)_0 = 1$$

and similarly for $(b)_n$ (Abramowtiz and Stegun, 1964, Eq. (13.1.2)).

With a uniform distribution for X_i on (0,1) under H_0, the posterior odds in favor of H_1 when the sample levels are $(x_1,\ldots,x_n)$ is the product of the prior odds and factor (8). Since the distribution of the X_i is $U(0,1)$ under H_0, the Bayes factor is the reciprocal of Eq. (8). (Recall that the Bayes factor can be interpreted as the likelihood ratio of H_0 to H_1, where the likelihood under H_1 is calculated under the 'weighting' g, Berger and Delampady, 1987.)

Initially, suppose that the X_i are identically distributed on (0,1), so that the single index parameter $\theta \in \Theta_0 = \{0\}$ under H_0 and $\theta \in \Theta_1 = (0,1)$ under the alternative. With the beta prior (7),

$$\int_{\Theta_1} \prod_{i=1}^{n} f(x_i|\theta) g(\theta; a, b)\, \mathrm{d}\theta = \frac{1}{B(a,b)} B(a, b+n) \\ M(a, a+b+n; T(x_1,\ldots,x_n)), \tag{10}$$

where

$$T(x_1,\ldots,X_n) = -\sum_{i=1}^{n} \log x_i.$$

In particular, one may adopt a non-informative prior for θ by selecting all $a_i = 1$ and $b_i = 1$ in Eq. (7). Then, it is clear from the series expansion (9) of $M(1, n+2; x)$ that the integral in Eq. (10) is a monotone function of $T(x_1,\ldots,x_n)$. It follows that the Bayes test of H_0 vs. the alternative H_1 here will be based on $T(x_1,\ldots x_n)$, which is equivalent to Fisher's procedure (Fisher, 1950).

More generally, the beta prior density $g(\theta_i; a_i, b_i)$ might reasonably be chosen with $b_i = 1$ and a_i proportional to the information in the ith experiment. In this setting, it is possible to characterize Kummer's functions in Eq. (8) more precisely. First, note that

$$M(a, a+2; z) = a(a+1) \sum_{i=0}^{\infty} \frac{z^i}{(a+i)(a+i+1)i!} \\ = a(a+1) \sum_{i=0}^{\infty} \left(\frac{1}{a+i} - \frac{1}{a+i+1} \right) \frac{z^i}{i!}.$$

Recall that the incomplete gamma function,

$$\gamma(a,x) = \int_0^x \mathrm{e}^{-t} t^{a-1}\, \mathrm{d}t, \quad \mathrm{Re}(a) > 0$$

may be expressed through the series representation (Gradshteyn and Ryzhik, 1980, p. 941),

$$\gamma(a,x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{a+k}}{k!(a+k)}.$$

It follows that

$$M(a, a+2; z) = a(a+1) z^{-a} \left[\gamma(a, -z) - \frac{1}{z} \gamma(a+1, -z) \right].$$

In the special case where a is a positive integer, one may obtain a closed-form expression for $M(a, a+2; z)$ using the general result (loc.cit., p. 940)

$$\gamma(1+n, x) = n! \left[1 - \mathrm{e}^{-x} \left(\sum_{k=0}^{n} \frac{x^k}{k!} \right) \right], \quad n = 0, 1, 2, \ldots .$$

This gives

$$M(a,a+2;z)=(a+1)!z^{-a}\left[1-\frac{a}{z}+\frac{(-z)^a}{a!}+(a-1)e^z\left(\sum_{k=0}^{a-1}\frac{(-z)^k}{k!}\right)\right].$$

For a large,

$$M(a,a+2;-\log x)\sim\frac{1}{x}$$

(since $\lim_{a\to\infty} M(a,a+2;z)=\exp(z)$); and for moderate values of a, the function

$$m(x)=a_1x^{-a_2}$$

provides a reasonable approximation to

$$M(a,a+2;-\log x),$$

where the a_i may be determined numerically. In other words, the Bayes test of H_0 vs. H_1 can here be based on the approximation

$$\prod_{i=1}^{n} x_i^{c_i},$$

which is Good's (1955) weighted version of Fisher's combination procedure. Closed-form formulas for the null distribution of the weighted product of levels are discussed by Good, and moment approximations have recently been described by Bhoj (1992).

Lastly, it is also of interest to consider conjugate prior distributions to Eq. (6). A conjugate prior density for

$$f(x)=(1-\theta)x^{-\theta},\quad 0<x<1,\ 0<\theta<1$$

takes the form

$$g(\theta;a,b)=\frac{b+1}{M(1,b+2;-a)}e^{-a\theta}(1-\theta)^b,\quad 0<\theta<1,\ a,b>0 \tag{11}$$

(Cox and Hinkley, 1974). If the beta priors (7) are replaced by Eq. (11), factor (8) becomes

$$\prod_{i=1}^{n}\left(\frac{b_i+1}{b_i+2}\right)\frac{M(1,b_i+3;-(a_i+\log x_i))}{M(1,b_i+2;-a_i)}.$$

Further simplifications and approximations are possible, as previously described.

Under certain regularity conditions (Bahadur, 1971; Raghavachari, 1970; Bahadur and Raghavachari, 1972), likelihood ratio statistics are optimal test statistics in the sense of exact slopes. Koziol (1993) recently showed that Bayes tests are also optimal in the sense of exact slopes. It follows immediately that Fisher's test is optimal for Lipták's formulation of the combination problem. See Littell and Folks (1973) for an alternative demonstration of the asymptotic optimality of Fisher's method, derived from first principles.

In closing, it should perhaps be noted that the Fisher–Lipták formulation of the combination problem is in some sense reductionist. More recent work in this area (much of it under the heading of meta-analysis) starts with a more structured, parametric framework. Within this context, see Hedges and Olkin (1985) for descriptions of more conventional statistical methods for meta-analysis, Dickersin and Berlin (1992) for practical considerations when undertaking meta-analyses, and, in particular, DuMouchel and Harris (1983) and DuMouchel (1990) for an introduction to the literature on Bayesian meta-analysis.

References

Abramowtiz, M., Stegun, I.A. (Eds.), 1964. Handbook of Mathematical Functions. U.S. Government Printing Office, Washington, DC.

Bahadur, R.R., 1960. Stochastic comparison of tests. Ann. Math. Statist. 31, 276–295.

Bahadur, R.R., 1967. Rates of convergence of estimates and test statistics. Ann. Math. Statist. 38, 303–324.

Bahadur, R.R., 1971. Some limit theorems in statistics. Regional Conf. Series in Applied Mathematics, vol. 4. SIAM, Philadelphia.

Bahadur, R.R., Raghavachari, M., 1972. Some asymptotic properties of likelihood ratios on general sample spaces. Proc. Sixth Berkley Symp. Math. Statist. Prob. vol. 1. University of California Press, Berkeley, pp. 129–152.

Berger, J.O., Delampady, M., 1987. Testing precise hypotheses (with discussion). Statist. Sci. 2, 317–348.

Bhoj, D.S., 1992. On the distribution of the weighted combination of independent probabilities. Statist. Probab. Lett. 15, 37–40.

Birnbaum, A., 1954. Combining independent tests of significance. J. Amer. Statist. Assoc. 49, 559–575.

Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman & Hall, London.

Dickersin, K., Berlin, J.A., 1992. Meta-analysis: state-of-the-science. Epidemiol. Rev. 14, 154–176.

DuMouchel, W.H., Harris, J.T., 1983. Bayes methods for combining the results of cancer studies in humans and other species. J. Amer. Statist. Assoc. 78, 293–308.

DuMouchel, W., 1990. Bayesian meta-analysis. In: Berry, D.A. (Ed.), Statistical Methodology in the Pharmaceutical Sciences. Marcel Dekker, New York, pp. 509–529.

Fisher, R.A., 1950. Statistical Methods for Research Workers, 11th ed. Oliver and Boyd, London, pp. 99–101.

Good, I.J., 1955. On the weighted combination of significance tests. J. Roy. Statist. Soc. Ser. B 17, 264–265.

Gradshteyn, L.S., Ryzhik, I.M., 1980. Table of Integrals, Series, and Products. Academic Press, New York.

Hedges, L.V., Olkin, I., 1985. Statistical Methods for Meta-Analysis. Academic Press, Orlando.

Koziol, J.A., 1993. On the optimality of Bayes test statistics and Fisher's combination procedure. In: Ghosh, J.K., Mitra, S.K., Parthasarathy, K.R., Prakasa Rao, B.L.S. (Eds.), Statistics and Probability: A Raghu Raj Bahadur Festschrift. Wiley Eastern Limited, New Delhi, pp. 365–370.

Lambert, D., Hall, W.J., 1982. Asymptotic lognormality of *p*-values. Ann. Statist. 10, 44–64.

Lipták, T., 1958. On the combination of independent tests. Magyar. Tud. Akad. Mat. Kutato Int. Közl. 3, 171–197.

Littell, R.C., Folks, J.L., 1973. Asymptotic optimality of Fisher's method of combining independent tests II. J. Amer. Statist. Assoc. 68, 193–194.

Pearson, E.S., 1939. The probability integral transformation for testing goodness of fit and combining independent tests of significance. Biometrika 30, 134–148.

Raghavachari, M., 1970. On a theorem of Bahadur on the rate of convergence of test statistics. Ann. Math. Statist. 41, 1695–1699.