Undesirable Optimality Results in Multiple Testing?

Charles Lewis
Dorothy T. Thayer

Intuitions about multiple testing:

- Multiple tests should be more conservative than individual tests.
- Controlling per comparison error rate is not enough. Need control of a familywise error rate or, better, *FDR*.

Multiple testing for multilevel models

- applying Bayesian ideas in a sampling theory context.

Examples: Shaffer (1999), Gelman & Tuerlincks (2000), Lewis & Thayer (2004), and Sarkar & Zhou (2008).

One-way random effects ANOVA setup

(Treat θ, σ^2 and τ^2 as known.)

$$\boldsymbol{\mu}' = (\mu_1, \cdots, \mu_m), \quad \bar{\mathbf{y}}' = (\bar{y}_1, \cdots, \bar{y}_m)$$

$$\mu_j \sim \mathrm{N}(\theta, \tau^2), \quad \bar{y}_j \,|\, \mu_j \sim \mathrm{N}(\mu_j, \sigma^2/n)$$

$$\bar{y}_j \sim \mathrm{N}(\theta, \tau^2 + \sigma^2/n), \; \mu_j \,|\, \bar{y}_j \sim \mathrm{N}(\hat{\mu}_j, v).$$

$$\hat{\mu}_j = \frac{n\tau^2 \bar{y}_j + \sigma^2 \theta}{n\tau^2 + \sigma^2}, \; v = \tau^2 \sigma^2 / (n\tau^2 + \sigma^2).$$

Consider all pairwise comparisons

$$\psi_i = \mu_j - \mu_{j'},\ \ \hat{\psi}_i = \hat{\mu}_j - \hat{\mu}_{j'} = \frac{n\tau^2\left(\bar{y}_j - \bar{y}_{j'}\right)}{n\tau^2 + \sigma^2}$$

$$\psi_i \sim \mathrm{N}\left(0, 2\tau^2\right),\ \bar{y}_j - \bar{y}_{j'} \,|\, \psi_i \sim \mathrm{N}\left(\psi_i, 2\sigma^2/n\right)$$

$$\psi_i \,|\, \hat{\psi}_i \sim \mathrm{N}\left(\hat{\psi}_i, 2v\right),\ \bar{y}_j - \bar{y}_{j'} \sim \mathrm{N}\left(0, 2\tau^2 + 2\sigma^2/n\right)$$

for $i = 1, \cdots, m(m-1)/2 = m^*$.

Decision theory framework

(Based on early work of Lehmann)

For each ψ_i, take action a_i.

$a_i = +1$: declare ψ_i to be positive,

$a_i = -1$: declare ψ_i to be negative,

$a_i = 0$: unable to determine sign of ψ_i.

Two components for loss functions

$L_1(\psi_i, a_i) = 1$ if the signs of ψ_i and a_i disagree and 0 otherwise;

used to indicate wrong sign declarations

$L_2(\psi_i, a_i) = 1$ if $a_i = 0$ and 0 otherwise;

used to indicate signs not determined.

Per comparison loss function for declaring sign of ψ_i

$$L_{PCi}(\psi_i, a_i) = L_1(\psi_i, a_i) + (\alpha/2) L_2(\psi_i, a_i).$$

Bayesian decision theory identifies the optimal decision rule, $\delta_{PCi}(\hat{\psi}_i)$ such that

$E_{\psi_i|\hat{\psi}_i}\left[L_{PCi}(\psi_i, \delta_{PCi}(\hat{\psi}_i)) \middle| \hat{\psi}_i\right]$ is minimized.

Finding the posterior expected loss

(Some helpful notation)

If $\Pr(\psi_i > 0 | \hat{\psi}_i) > 0.5$, define $a_i^* = +1$ and

$p_i = \Pr(\psi_i < 0 | \hat{\psi}_i)$;

if $\Pr(\psi_i > 0 | \hat{\psi}_i) \le 0.5$, define $a_i^* = -1$ and

$p_i = \Pr(\psi_i > 0 | \hat{\psi}_i)$. It then follows that

$$E_{\psi_i | \hat{\psi}_i}\left[L_{PCi}\left(\psi_i, a_i^*\right) | \hat{\psi}_i\right] = E_{\psi_i | \hat{\psi}_i}\left[L_1\left(\psi_i, a_i^*\right) | \hat{\psi}_i\right] = p_i \ .$$

If $a_i = 0$, we have

$$
\begin{aligned}
L_{PCi}(\psi_i, 0) &= L_1(\psi_i, 0) + (\alpha/2) L_2(\psi_i, 0) \\
&= \alpha/2 \ .
\end{aligned}
$$

Therefore, the Bayes rule declares the sign of ψ_i, namely $\delta_{PCi}(\hat{\psi}_i) = a_i^*$, iff $p_i < \alpha/2$; otherwise it takes $\delta_{PCi}(\hat{\psi}_i) = 0$.

Since the posterior expected loss for δ_{PCi} is always less than or equal to $\alpha/2$, it follows that the Bayes risk for δ_{PCi} is also less than or equal to $\alpha/2$:

$$E_{\psi_i,\hat{\psi}_i}\left[L_{PCi}\left(\psi_i,\delta_{PCi}\left(\hat{\psi}_i\right)\right)\right] \leq \alpha/2 \ .$$

Consequently,

$$E_{\psi_i,\hat{\psi}_i}\left[L_1(\psi_i,\delta_{PC}(\hat{\psi}_i))\right] \le \alpha/2 \ .$$

This expectation is the (random effects) probability of incorrectly declaring the sign of ψ_i using the decision rule δ_{PCi}: the per comparison wrong sign rate for ψ_i.

Explicit expression for p_i

$$p_i = \min\{\Pr(\psi_i > 0 \mid \hat{\psi}_i), \Pr(\psi_i < 0 \mid \hat{\psi}_i)\}$$

$$= \Phi\left[\frac{-|\hat{\psi}_i|}{\sqrt{2v}}\right]$$

$$= \Phi\left[\frac{-|\bar{y}_j - \bar{y}_{j'}|}{\sqrt{2\sigma^2/n}}\sqrt{\frac{n\tau^2}{n\tau^2+\sigma^2}}\right].$$

For the usual (fixed effects) per

comparison test, $p_{fi} = \Phi\left[\frac{-\left|\overline{y}_j - \overline{y}_{j'}\right|}{\sqrt{2\sigma^2/n}}\right]$,

so $p_i > p_{fi}$.

Define a fixed effects decision rule by

$\delta_{fi}(\hat{\psi}_i) = a_i^*$, iff $p_{fi} < \alpha/2$;

otherwise we have $\delta_{fi}(\hat{\psi}_i) = 0$.

Since p_{fi} is based on the distribution of

$\hat{\psi}_i \mid \psi_i$, we may write

$$E_{\hat{\psi}_i \mid \psi_i}\left[L_1\left(\psi_i, \delta_{fi}\left(\hat{\psi}_i\right)\right) \mid \psi_i\right] \leq p_{fi} \leq \alpha/2 \text{ ,}$$

and so

$$E_{\hat{\psi}_i, \psi_i}\left[L_1\left(\psi_i, \delta_{fi}\left(\hat{\psi}_i\right)\right)\right] \leq \alpha/2 \text{ .}$$

Conclusion: the Bayesian random effects rule and the fixed effects rule both control the random effects per comparison wrong sign rate at $\alpha/2$, but the Bayesian rule is more conservative than the fixed effects rule.

Extend definition of the per comparison loss function to the set of comparisons

$$L_{PC}(\boldsymbol{\psi},\mathbf{a}) = \frac{\sum_{i=1}^{m^*} L_1(\psi_i, a_i)}{m^*} + \left(\frac{\alpha}{2}\right)\frac{\sum_{i=1}^{m^*} L_2(\psi_i, a_i)}{m^*}$$

$$= \frac{\sum_{i=1}^{m^*} L_{PCi}(\psi_i, a_i)}{m^*} .$$

Interpretation of L_{PC}

This new loss function equals the proportion of comparisons whose signs are incorrectly declared using **a**, plus $\alpha/2$ times the proportion of comparisons whose signs are not determined using **a**.

Family of optimal action vectors $\mathbf{a}^{(k)}$

Order the p_i so that $p_{(1)} \leq \cdots \leq p_{(m^*)}$.

Define $\mathbf{a}^{(k)}$ for $k = 1, \cdots, m^*$ as

$a_{(i)}^{(k)} = a_{(i)}^{*}$, for $i = 1, \cdots, k$, and

$a_{(i)}^{(k)} = 0$, for $i = k+1, \cdots, m^*$.

Take $a_{(i)}^{(0)} = 0$, for $i = 1, \cdots, m^*$.

The Bayesian decision rule for the loss function L_{PC}

$$\boldsymbol{\delta}_{PC}\left(\hat{\boldsymbol{\psi}}\right) = \mathbf{a}^{(k_{PC})},$$

where k_{PC} is the largest value of k such that $p_{(k)} < \alpha/2$, or $k_{PC} = 0$ if $p_{(1)} \geq \alpha/2$.

Posterior expected loss for $\boldsymbol{\delta}_{PC}$

$$E_{\boldsymbol{\psi}|\hat{\boldsymbol{\psi}}}\left[L_{PC}\left(\boldsymbol{\psi},\boldsymbol{\delta}_{PC}\left(\hat{\boldsymbol{\psi}}\right)\right)\middle|\hat{\boldsymbol{\psi}}\right]=\frac{\sum_{i=1}^{k_{PC}} p_{(i)}}{m^*}+\left(\frac{\alpha}{2}\right)\left(\frac{m^*-k_{PC}}{m^*}\right)$$

if $k_{PC} > 0$, and

$$E_{\boldsymbol{\psi}|\hat{\boldsymbol{\psi}}}\left[L_{PC}\left(\boldsymbol{\psi},\boldsymbol{\delta}_{PC}\left(\hat{\boldsymbol{\psi}}\right)\right)\middle|\hat{\boldsymbol{\psi}}\right]=\alpha/2 \text{ if } k_{PC}=0.$$

Since $L_{PC}\left(\boldsymbol{\psi}, \mathbf{a}^{(0)}\right) = \alpha/2$, the posterior expected loss for the Bayesian decision function must be less than or equal to $\alpha/2$, and the Bayes risk for $\boldsymbol{\delta}_{PC}$ must also be less than or equal to $\alpha/2$:

$$r\left(\boldsymbol{\delta}_{PC}\right) = E_{\boldsymbol{\psi},\hat{\boldsymbol{\psi}}}\left[L_{PC}\left(\boldsymbol{\psi}, \boldsymbol{\delta}_{PC}\left(\hat{\boldsymbol{\psi}}\right)\right)\right] \leq \alpha/2 \; .$$

Consequently,

$$E_{\boldsymbol{\psi},\hat{\boldsymbol{\psi}}}\left[\left(\frac{1}{m^*}\right)\sum_{i=1}^{m^*} L_1(\psi_i, \delta_{PCi}(\hat{\psi}_i))\right] \le \frac{\alpha}{2}.$$

This expectation is the (random effects)

per comparison wrong sign rate for $\boldsymbol{\psi}$

using the Bayes rule $\boldsymbol{\delta}_{PC}$.

Rewriting the bound on the posterior expected loss for $\boldsymbol{\delta}_{PC}$ given $\hat{\boldsymbol{\psi}}$, we have

$$\frac{\sum_{i=1}^{k_{PC}} p_{(i)}}{m^*} + \left(\frac{\alpha}{2}\right)\left(\frac{m^* - k_{PC}}{m^*}\right) \leq \frac{\alpha}{2}.$$

Consequently, we may write

$$\frac{\sum_{i=1}^{k_{PC}} p_{(i)}}{m^*} \leq \left(\frac{\alpha}{2}\right)\left(\frac{k_{PC}}{m^*}\right), \text{ so}$$

$$\frac{\sum_{i=1}^{k_{PC}} p_{(i)}}{\min\{1, k_{PC}\}} \leq \frac{\alpha}{2}.$$

Since this inequality gives an upper bound on the posterior expectation, a corresponding upper bound holds for the unconditional expectation:

$$E_{\boldsymbol{\psi},\hat{\boldsymbol{\psi}}}\left[\frac{\sum_{i=1}^{m^*} L_1\left(\psi_i, \delta_{PCi}\left(\hat{\psi}_i\right)\right)}{\min\{1, k_{PC}\}}\right] \leq \frac{\alpha}{2}.$$

This quantity (evaluated for any decision rule $\boldsymbol{\delta}$) is referred to by Sarkar and Zhou (2008) as the Bayesian directional false discovery rate, or BDFDR, for $\boldsymbol{\delta}$.

The result that $\boldsymbol{\delta}_{PC}$ controls the BDFDR was given by Lewis and Thayer (2004).

Having a per comparison rule control a version of the FDR is counterintuitive!

Sarkar and Zhou (2008) propose another decision rule (here labeled $\boldsymbol{\delta}_{SZ}$) that also controls the BDFDR and maximizes the posterior per comparison power rate.

Specifically, $\boldsymbol{\delta}_{SZ}(\hat{\boldsymbol{\psi}}) = \mathbf{a}^{(k_{SZ})}$, where k_{SZ} is the largest value of k such that

$$\frac{\sum_{i=1}^{k} p_{(i)}}{k} \leq \frac{\alpha}{2}, \text{ or } k_{SZ} = 0 \text{ if } p_{(1)} \geq \alpha/2 .$$

Thus $\boldsymbol{\delta}_{SZ}$ controls the BDFDR at $\alpha/2$.

Sarkar and Zhou (2008) also proved that, among (non-randomized) rules that control the BDFDR, $\boldsymbol{\delta}_{SZ}$ maximizes the posterior per comparison power rate:

$$E_{\boldsymbol{\psi}|\hat{\boldsymbol{\psi}}}\left[\frac{\sum_{i=1}^{m^*}\left[1-L_1(\psi_i,\delta_i(\hat{\psi}_i))-L_2(\psi_i,\delta_i(\hat{\psi}_i))\right]}{m^*}\middle|\hat{\boldsymbol{\psi}}\right]$$

Too much power?

Not only does $\boldsymbol{\delta}_{SZ}$ have more power than the Bayes rule $\boldsymbol{\delta}_{PC}$, it may also have more power than the fixed effects rule $\boldsymbol{\delta}_{f}$. In other words, $\boldsymbol{\delta}_{SZ}$ will sometimes declare a sign for ψ_i even when $p_{fi} > \alpha/2$. This is counterintuitive!

To summarize, in a multilevel model like random effects ANOVA, Bayesian ideas have sampling interpretations. In particular, we may define a Bayesian (or random effects) version of the FDR: The average (over both levels) proportion of declared signs for a set of comparisons that are incorrectly declared.

1. A Bayesian per comparison decision rule turns out to provide control of this FDR, even though it was only designed to minimize an expected per comparison loss function.

2. And a rule *designed* to control this FDR may have more power than a conventional per comparison rule.

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.

Gelman, A., & Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics*, *15*, 373-390.

Jones, L. V., & Tukey, J. W. (2000). A sensible formulation of the significance test. *Psychological Methods, 5*, 411-414.

Lehmann, E. L. (1950). Some principles of the theory of testing hypotheses. *The Annals of Mathematical Statistics, 21,* 1-26.

Lehmann, E. L. (1957a). A theory of some multiple decision problems. I. *The Annals of Mathematical Statistics, 28,* 1-25.

Lehmann, E. L. (1957b). A theory of some multiple decision problems. II. *The Annals of Mathematical Statistics, 28,* 547-572.

Lewis, C., & Thayer, D. T. (2004). A loss function related to the FDR for random effects multiple comparisons. *Journal of Statistical Planning and Inference*, *125*, 49-58.

Sakar, S. K., & Zhou, T. (2008). Controlling directional Bayesian false discovery rate in random effects model. *Journal of Statistical Planning and Inference*, *138*, 682-693.

Shaffer, J. P. (1999). A semi-Bayesian study of Duncan’s Bayesian multiple comparison procedure. *Journal of Statistical Planning and Inference*, *82*, 197-213.

Shaffer, J. P. (2002). Multiplicity, directional (Type III) errors and the null hypothesis. *Psychological Methods, 7*, 356-369.

Williams, V. S. L., Jones, L. V., & Tukey, J. W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics, 24*, 42-69.