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.

One-way random effects ANOVA setup

(Treat $\theta$, $\sigma^2$ and $\tau^2$ as known.)

$\mathbf{\mu}' = (\mu_1, \ldots, \mu_m)$, $\mathbf{y}' = (y_1, \ldots, y_m)$

$\mu_j \sim N(\theta, \tau^2)$, $y_j | \mu_j \sim N(\mu_j, \sigma^2/n)$

$y_j \sim N(\theta, \tau^2 + \sigma^2/n)$, $\mu_j | y_j \sim N(\mu_j, \nu)$.

$\hat{\mu}_j = \frac{n\tau^2 y_j + \sigma^2 \theta}{n\tau^2 + \sigma^2}$, $\nu = \tau^2 \sigma^2 / (n\tau^2 + \sigma^2)$. 
Consider all pairwise comparisons

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

\[ \psi_i \sim N(0, 2\tau^2), \quad \bar{y}_j - \bar{y}_{j'} \mid \psi_i \sim N(\psi_i, 2\sigma^2/n) \]

\[ \psi_i \mid \hat{\psi}_i \sim N(\hat{\psi}_i, 2\nu), \quad \bar{y}_j - \bar{y}_{j'} \sim N(0, 2\tau^2 + 2\sigma^2/n) \]

for \( i = 1, \ldots, m(m-1)/2 = m^* \).
Decision theory framework

(Based on early work of Lehmann)

For each $\psi_i$, take action $a_i$.

- $a_i = +1$: declare $\psi_i$ to be positive,
- $a_i = -1$: declare $\psi_i$ to be negative,
- $a_i = 0$: unable to determine sign of $\psi_i$. 
Two components for loss functions

\[ L_1 (\psi_i, a_i) = 1 \text{ if the signs of } \psi_i \text{ and } a_i \text{ disagree and 0 otherwise;} \]

\[ L_2 (\psi_i, a_i) = 1 \text{ if } a_i = 0 \text{ and 0 otherwise;} \]

*used to indicate wrong sign declarations*

*used to indicate signs not determined.*
Per comparison loss function for declaring sign of $\psi_i$

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

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

$$E_{\psi_i | \hat{\psi}_i} \left[ L_{PCI} (\psi_i, \delta_{PCI} (\hat{\psi}_i)) | \hat{\psi}_i \right]$$

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 $\rho_i = \Pr(\psi_i < 0 | \hat{\psi}_i)$; if $\Pr(\psi_i > 0 | \hat{\psi}_i) \leq 0.5$, define $a_i^* = -1$ and $\rho_i = \Pr(\psi_i > 0 | \hat{\psi}_i)$. It then follows that

$$E_{\psi_i | \hat{\psi}_i} \left[ L_{PCI} (\psi_i, a_i^*) | \hat{\psi}_i \right] = E_{\psi_i | \hat{\psi}_i} \left[ L_1 (\psi_i, a_i^*) | \hat{\psi}_i \right] = \rho_i.$$
If \( a_i = 0 \), we have

\[
L_{PCI}(\psi_i, 0) = L_1(\psi_i, 0) + (a/2)L_2(\psi_i, 0) = a/2 .
\]

Therefore, the Bayes rule declares the sign of \( \psi_i \), namely \( \delta_{PCI}(\psi_i) = a^*_i \), iff

\[
p_i < a/2; \text{ otherwise it takes } \delta_{PCI}(\psi_i) = 0.
\]
Since the posterior expected loss for $\delta_{PCI}$ is always less than or equal to $\alpha/2$, it follows that the Bayes risk for $\delta_{PCI}$ is also less than or equal to $\alpha/2$:

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

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

This expectation is the (random effects) probability of incorrectly declaring the sign of \( \psi_i \) using the decision rule \( \delta_{PCI} \): the per comparison wrong sign rate for \( \psi_i \).
Explicit expression for $p_i$

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

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

$$= \Phi \left[ \frac{-|y_j - y_j'|}{\sqrt{2\sigma^2/n}} \sqrt{\frac{n\tau^2}{n\tau^2 + \sigma^2}} \right] .$$
For the usual (fixed effects) per comparison test, 
\[ p_{fi} = \Phi \left[ \frac{-|\bar{y}_j - \bar{y}_{j'}|}{\sqrt{2\sigma^2/n}} \right], \]

so \( p_i > p_{fi} \).

Define a fixed effects decision rule by
\[ \delta_{fi}(\hat{\psi}_i) = a_i^*, \text{ iff } p_{fi} < \alpha/2; \]
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

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

and so

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

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

\[
= \sum_{i=1}^{m^*} L_{PCI}(\psi_i, a_i) \frac{m^*}{m^*}.
\]
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 $a^{(k)}$

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

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

\[
    a^{(k)}_{(i)} = a^*_{(i)}, \text{ for } i = 1, \cdots, k, \text{ and }
\]

\[
    a^{(k)}_{(i)} = 0, \text{ for } i = k + 1, \cdots, m^*.
\]

Take $a^{(0)}_{(i)} = 0$, for $i = 1, \cdots, m^*$. 
The Bayesian decision rule for the loss function $L_{PC}$

$$\delta_{PC}(\hat{\psi}) = a^{(k_{PC})},$$

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 $\delta_{PC}$

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

if $k_{PC} > 0$, and

$$E_{\psi|\hat{\psi}} \left[ L_{PC} (\psi, \delta_{PC} (\hat{\psi})) | \hat{\psi} \right] = \frac{\alpha}{2} \text{ if } k_{PC} = 0.$$
Since $L_{PC} \left( \psi, a^{(0)} \right) = \alpha/2$, the posterior expected loss for the Bayesian decision function must be less than or equal to $\alpha/2$, and the Bayes risk for $\delta_{PC}$ must also be less than or equal to $\alpha/2$:

$$r \left( \delta_{PC} \right) = E_{\psi, \hat{\psi}} \left[ L_{PC} \left( \psi, \delta_{PC} \left( \hat{\psi} \right) \right) \right] \leq \alpha/2.$$
Consequently,

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

This expectation is the (random effects) per comparison wrong sign rate for \( \psi \) using the Bayes rule \( \delta_{PC} \).
Rewriting the bound on the posterior expected loss for $\mathbf{\delta}_{PC}$ given $\hat{\psi}$, we have

\[
\sum_{i=1}^{k_{PC}} \frac{\rho_{(i)}}{m^*} + \left( \frac{\alpha}{2} \right) \left( \frac{m^* - k_{PC}}{m^*} \right) \leq \frac{\alpha}{2}.
\]
Consequently, we may write

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

\[
\frac{k_{PC}}{\min \{1, k_{PC} \}} \sum_{i=1}^{k_{PC}} p_i \leq \frac{\alpha}{2}.
\]
Since this inequality gives an upper bound on the posterior expectation, a corresponding upper bound holds for the unconditional expectation:

\[
E_{\psi, \hat{\psi}} \left[ \sum_{i=1}^{m^*} \frac{L_1(\psi_i, \delta_{PCI}(\hat{\psi}_i))}{\min\{1, k_{PC}\}} \right] \leq \frac{\alpha}{2}.
\]
This quantity (evaluated for any decision rule $\delta$) is referred to by Sarkar and Zhou (2008) as the Bayesian directional false discovery rate, or BDFDR, for $\delta$. The result that $\delta_{PC}$ 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 $\delta_{sz}$) that also controls the BDFDR and maximizes the posterior per comparison power rate.
Specifically, $\delta_{sz}(\hat{\psi}) = a^{(k_{sz})}$, where $k_{sz}$ is the largest value of $k$ such that

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

Thus $\delta_{sz}$ controls the BDFDR at $\alpha/2$. 
Sarkar and Zhou (2008) also proved that, among (non-randomized) rules that control the BDFDR, $\delta_{sz}$ maximizes the posterior per comparison power rate:

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

$$m^*$$
Too much power?

Not only does $\delta_{SZ}$ have more power than the Bayes rule $\delta_{PC}$, it may also have more power than the fixed effects rule $\delta_f$. In other words, $\delta_{SZ}$ will sometimes declare a sign for $\psi_i$ 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


