What is Empirical Bayes? A brief history, well-known examples & connections.

Berkeley Stanford Joint Colloquium (BSJC) in Statistics

Leonid Pekelis

Apr 24, 2012

A brief (incomplete, biased) history.

- von Mises (1940s)
- Robbins (1955) - “An Empirical Bayes Approach to Statistics”

A brief (incomplete, biased) history.

- von Mises (1940s)
- Robbins (1955) - "An Empirical Bayes Approach to Statistics"
- Efron & Morris (1977) - "Stein's Paradox in Statistics"
  - Stein (1956), James & Stein (1961)
- Morris (1983) - "Parametric Empirical Bayes Inference: Theory & Applications"
- Casella (1985) - "An Introduction to Empirical Bayes Data Analysis"

A brief (incomplete, biased) history.

- von Mises (1940s)
- Robbins (1955) - “An Empirical Bayes Approach to Statistics”
- Efron & Morris (1977) - “Stein’s Paradox in Statistics”
  - Stein (1956), James & Stein (1961)
- Morris (1983) - “Parametric Empirical Bayes Inference: Theory & Applications”
- Casella (1985) - “An Introduction to Empirical Bayes Data Analysis”
- Efron, Tibshirani, Storey & Tusher (2001) - “Empirical Bayes Analysis of a Microarray Experiment”
  - Benjamini & Hochberg (1995)
- Efron (2010) - “Large Scale Inference”

A brief history.

Robbins - "Our reason for doing this ... is that w hope for large N to be able to extract information about the [prior] from the values which have been observed, hopefully in such a way that [EBayes] will be close to the optimal but unknown [Bayes] ..."

A brief history.

Robbins - “Our reason for doing this ... is that w hope for large N to be able to extract information about the [prior] from the values which have been observed, hopefully in such a way that [EBayes] will be close to the optimal but unknown [Bayes] ...”

Efron - “Empirical Bayes blurs the line between testing and estimation as well as between frequentism Bayesianism.”

A brief history.

Robbins - "Our reason for doing this ... is that w hope for large N to be able to extract information about the [prior] from the values which have been observed, hopefully in such a way that [EBayes] will be close to the optimal but unknown [Bayes] ..."

Efron - "Empirical Bayes blurs the line between testing and estimation as well as between frequentism Bayesianism."

Morris - "Should statisticians use empirical Bayes modeling in most multiparameter inference problems? Probably not"

A brief history.

- Non-Parametric Empirical Bayes
- Parametric Empirical Bayes

A brief history.

- Non-Parametric Empirical Bayes
  - leaves prior completely unspecified (Robbins - 1955, Tweedie's formula)
- Parametric Empirical Bayes
  - Specifies family of priors, estimate hyperparameters (James-Stein)

Robbins Original Formulation

$X \in \mathcal{X}$, and a σ-finite $\nu(x)$

$\mu \in \mathbb{Q}$, $\mu \sim G(\mu)$

when parameter is μ, $X|\mu \sim f_\mu$

for any loss $L(t, \mu)$, t a decision function

Robbins Original Formulation

$X \in \mathcal{X}$, and a σ-finite $\nu(x)$

$\mu \in \mathbb{Q}$, $\mu \sim G(\mu)$

when parameter is μ, $X|\mu \sim f_\mu$

for any loss $L(t, \mu)$, t a decision function

$$R(t, G) = \boldsymbol{E}_G \boldsymbol{E}_{f_\mu}[L(t(X), \mu)]$$
$$R(G) = \min_t R(t, G)$$

Robbins Original Formulation

$$R(G) = \min_t R(t, G)$$

What if G is not known?

Robbins Original Formulation

$$R(G) = \min_t R(t, G)$$

What if G is not known? But you have samples (x_i, μ_i), $i = 1, \ldots, N$ iid with $\mu_i \sim G$, $x_i \sim f_\mu$.

Robbins Original Formulation

$$R(G) = \min_t R(t, G)$$

What if G is not known? But you have samples (x_i, μ_i), $i = 1, \dots, N$ iid with $\mu_i \sim G$, $x_i \sim f_\mu$. Then you can estimate $t_n(x) = t_n(x; x_1, \dots, x_n)$ and get

$$R_N(\{t_N\}, G) = \boldsymbol{E}_{\mathbf{x}} R(t_N, G) \geq R(G)$$

Robbins Original Formulation

$$R(G) = \min_t R(t, G)$$

What if G is not known? But you have samples (x_i, μ_i), $i = 1, \ldots, N$ iid with $\mu_i \sim G$, $x_i \sim f_\mu$. Then you can estimate $t_n(x) = t_n(x; x_1, \ldots, x_n)$ and get

$$R_N(\{t_N\}, G) = \boldsymbol{E}_{\mathbf{x}} R(t_N, G) \geq R(G)$$

Definition

$T = \{t_N\}$ is asymptotically optimal relative to G if

$$\lim_N R_N(T, G) = R(G)$$

We want the above to hold on a class $\mathcal{G}$ containing the true G.

Example: James-Stein Estimator

Suppose $G = N(0, A)$ and $f_\mu = N(\mu, 1)$.

Then $\boldsymbol{E}(\mu|x) = \left(1 - \frac{1}{A+1}\right) x$

And $E(\frac{N-2}{S}) = \frac{1}{A+1}$ where $S = \|\mathbf{x}\|^2$.

So $\hat{\mu}^{JS}(x) = \left(1 - \frac{N-2}{S}\right) x$

Example: James-Stein Estimator

Suppose $G = N(0, A)$ and $f_\mu = N(\mu, 1)$.

Then $\boldsymbol{E}(\mu|x) = \left(1 - \frac{1}{A+1}\right) x$

And $E(\frac{N-2}{S}) = \frac{1}{A+1}$ where $S = \|\mathbf{x}\|^2$.

So $\hat{\mu}^{JS}(x) = \left(1 - \frac{N-2}{S}\right) x$

Theorem

For $N \geq 3$, $\hat{\mu}^{JS}$ everywhere dominates $\hat{\mu}^{MLE}$ for $L(t, \mu) = \|t - \mu\|^2$.

Example: Benjamini-Hochberg / FDR

BH procedure: Have $p_1, \ldots, p_n$ p-values

Fix some $q \in (0,1)$, and take i_{max} as the largest index such that $p_{(i)} \leq \frac{i}{N} q$.

Reject $H_{0(i)}$ for $i \leq i_{max}$.

Example: Benjamini-Hochberg / FDR

BH procedure: Have $p_1, \ldots, p_n$ p-values

Fix some $q \in (0, 1)$, and take i_{max} as the largest index such that $p_{(i)} \leq \frac{i}{N} q$.

Reject $H_{0(i)}$ for $i \leq i_{max}$.

Theorem

If $p_i \sim U(0, 1)$ under H_{0i}, and independent for all i, then

$$\boldsymbol{E}\left(\frac{\#FR}{\#R}\right) \leq q$$

Example: BH / FDR - Empirical Bayes

Suppose p_i correspond to z_i test statistics.

Specifically $\mu \sim \pi_0 \mathbf{1}(\mu = 0) + \pi_1 g(\mu)$, and $x|\mu \sim N(\mu, 1)$. And test $H_{0i} : \mu_i > 0$.

Define $\bar{F}$ as empirical distribution of $\mathbf{z}$, F_0 as common null distribution, and $\hat{Fdr}(z) = \frac{\pi_0 F_0(z)}{\bar{F}(z)}$.

Example: BH / FDR - Empirical Bayes

Suppose p_i correspond to z_i test statistics.

Specifically $\mu \sim \pi_0 \mathbf{1}(\mu = 0) + \pi_1 g(\mu)$, and $x|\mu \sim N(\mu, 1)$. And test $H_{0i} : \mu_i > 0$.

Define $\bar{F}$ as empirical distribution of $\mathbf{z}$, F_0 as common null distribution, and $\hat{Fdr}(z) = \frac{\pi_0 F_0(z)}{\bar{F}(z)}$.

Corollary

Then the procedure rejecting for $i < i_{max}$, with i_{max} largest such that $\hat{Fdr}(z_{(i)}) < q$, controls false discovery propostion at level q as before.

Example: Tweedie's formula

$$\mu \sim g(\mu),\ z|\mu \sim f_\mu(z) = e^{\mu z - \psi(\mu)} f_0(z)$$

Example: Tweedie's formula

$$\mu \sim g(\mu),\ z|\mu \sim f_\mu(z) = e^{\mu z - \psi(\mu)} f_0(z)$$

Bayes rule: $g(\mu|z) = \frac{f_\mu(z) g(\mu)}{f(z)}$, where $f(z) = \int f_\mu(z) g(\mu) d\mu$

$$\Rightarrow g(\mu|z) = e^{z\mu - \lambda(z)} \left(g(\mu) e^{-\psi(\mu)} \right) \quad , \quad \lambda(z) = \log\left(\frac{f(z)}{f_0(z)}\right)$$

$$\Rightarrow \boldsymbol{E}(\mu|z) = \lambda'(z) \quad , \quad Var(\mu|z) = \lambda''(z)$$

$$\Rightarrow \mu|z \sim (l'(z) - l_0'(z), l''(z) - l_0''(z)) \quad , \quad l(z) = \log f(z)$$

Example: Tweedie's formula

$$\mu \sim g(\mu),\ z|\mu \sim f_\mu(z) = e^{\mu z - \psi(\mu)} f_0(z)$$

<u>Bayes rule</u>: $g(\mu|z) = \frac{f_\mu(z)g(\mu)}{f(z)}$, where $f(z) = \int f_\mu(z)g(\mu)d\mu$

$$\Rightarrow g(\mu|z) = e^{z\mu - \lambda(z)} \left(g(\mu)e^{-\psi(\mu)} \right) \quad , \quad \lambda(z) = \log\left(\frac{f(z)}{f_0(z)}\right)$$

$$\Rightarrow \boldsymbol{E}(\mu|z) = \lambda'(z) \quad , \quad Var(\mu|z) = \lambda''(z)$$

$$\Rightarrow \mu|z \sim (l'(z) - l_0'(z), l''(z) - l_0''(z)) \quad , \quad l(z) = \log f(z)$$

<u>For:</u> $z|\mu \sim N(\mu, 1)$, $\mu|z \sim (z + l'(z), 1 + l''(z))$.

Example: Tweedie's formula

The Empirical Bayes part comes from assuming

$$f(z) = e^{\sum_{j=0}^{J} \beta_j z^j}$$

Then can estimate $\hat{\beta}$ by GLM (Lindsey's Method).

And finally $\hat{l}'(z) = \sum_{j=1}^{J} j\hat{\beta}_j z^{j-1}$.

Example: Tweedie's formula

The Empirical Bayes part comes from assuming

$$f(z) = e^{\sum_{j=0}^{J} \beta_j z^j}$$

Then can estimate $\hat{\beta}$ by GLM (Lindsey's Method).

And finally $\hat{l}'(z) = \sum_{j=1}^{J} j \hat{\beta}_j z^{j-1}$.

Theorem

Under above assumptions, Lindsey's Method produces nearly unbiased estimates of f.

Tweedie & James-Stein

Suppose $\mu \sim N(0, A)$, and take $J = 2$.

Tweedie & James-Stein

Suppose $\mu \sim N(0, A)$, and take $J = 2$.

$$\hat{\mu}^{Tweedie}(z) = \left(1 - \frac{N}{S}\right) z$$
$$\hat{\mu}^{JS}(z) = \left(1 - \frac{N-2}{S}\right) z$$

Other Connections

- under 2 groups model, $fdr(z) = \frac{\pi_0 f_0(z)}{f_(z)}$

  $$-\frac{\partial}{\partial z}\log(fdr(z)) = l'(z) - l_0'(z) = \boldsymbol{E}(\mu|z)$$

  - $\hat{Fdr}$ similar, and more general than Tweedie's as it only requires specification of null density $f_0(z)$.

Other Connections

- under 2 groups model, $fdr(z) = \frac{\pi_0 f_0(z)}{f_(z)}$

  $$-\frac{\partial}{\partial z} \log(fdr(z)) = l'(z) - l_0'(z) = \boldsymbol{E}(\mu|z)$$

  - $\hat{Fdr}$ similar, and more general than Tweedie's as it only requires specification of null density $f_0(z)$.

- $g_{BH}(\mu) = \pi_0 \mathbf{1}(\mu = 0) + \pi_1 g(\mu)$, and $g_{JS}(\mu) = N(0, A)$

- JS can have $N = 10$, but Fdr or Tweedie with $J > 2$ need $N >> 10$

Other Connections

- under 2 groups model, $fdr(z) = \frac{\pi_0 f_0(z)}{f_(z)}$

  $$-\frac{\partial}{\partial z}\log(fdr(z)) = l'(z) - l_0'(z) = \boldsymbol{E}(\mu|z)$$

  - $\hat{Fdr}$ similar, and more general than Tweedie's as it only requires specification of null density $f_0(z)$.

- $g_{BH}(\mu) = \pi_0 \mathbf{1}(\mu = 0) + \pi_1 g(\mu)$, and $g_{JS}(\mu) = N(0, A)$

- JS can have $N = 10$, but Fdr or Tweedie with $J > 2$ need $N >> 10$

- Suggests Bias / Variance tradeoff in empirical Bayes estimates.

The Last Slide

Robbins - "... it is obvious that any theory of statistical inference will find itself in and out of fashion as the winds of doctrine blow. Here, then, are some remarks and references for further reading which I hope will interest my audience in thinking the matter through for themselves."

The Last Slide

Robbins - "... it is obvious that any theory of statistical inference will find itself in and out of fashion as the winds of doctrine blow. Here, then, are some remarks and references for further reading which I hope will interest my audience in thinking the matter through for themselves."

Me - "Thanks!"