The empirical Bayes approach

Arises from the multiparameter estimation problem where known relationships among the coordinates of $\theta = (\theta_1, \ldots, \theta_k)'$ suggest pooling information across the similar experiments, to get a better estimate of each $\theta_i$. 
The empirical Bayes approach

Arises from the multiparameter estimation problem where known relationships among the coordinates of \( \theta = (\theta_1, \ldots, \theta_k)' \) suggest pooling information across the similar experiments, to get a better estimate of each \( \theta_i \).

Examples:

\[
\theta_i = \text{proportion of defectives in supplier’s lot } i \\
= \text{mean bushels of corn per acre for a random selection of farmers in a given county}
\]
The empirical Bayes approach

- Arises from the multiparameter estimation problem where known relationships among the coordinates of $\theta = (\theta_1, \ldots, \theta_k)'$ suggest pooling information across the similar experiments, to get a better estimate of each $\theta_i$.

  Examples:

  $\theta_i = \begin{align*}
  &\text{proportion of defectives in supplier’s lot } i \\
  = &\text{mean bushels of corn per acre for a random selection of farmers in a given county}
  \end{align*}$

- These problems have a long history:
  - “random effects models”
  - “mixed models”

  – the latter gave rise to \texttt{Proc MIXED} in SAS!
The empirical Bayes approach

Morris (’83 JASA) classified EB methods into two categories: **parametric** and **nonparametric**:

**Parametric EB (PEB):** We assume the prior for $\theta$ is in some parametric class $p(\theta|\eta)$, where only $\eta$ (the hyperparameter) is unknown.
The empirical Bayes approach

Morris (’83 *JASA*) classified EB methods into two categories: **parametric** and **nonparametric**:

- **Parametric EB (PEB):** We assume the prior for $\theta$ is in some parametric class $p(\theta | \eta)$, where only $\eta$ (the hyperparameter) is unknown.

  - estimate the hyperparameter by $\hat{\eta}$, and plug in to get the **estimated posterior**.
The empirical Bayes approach

Morris ('83 *JASA*) classified EB methods into two categories: **parametric** and **nonparametric**:

- **Parametric EB (PEB):** We assume the prior for $\theta$ is in some parametric class $p(\theta|\eta)$, where only $\eta$ (the hyperparameter) is unknown.
  - estimate the hyperparameter by $\hat{\eta}$, and plug in to get the **estimated posterior**.
  - well-developed in a series of papers by Efron and Morris in the mid-70’s.
The empirical Bayes approach

Morris ('83 *JASA*) classified EB methods into two categories: **parametric** and **nonparametric**:

- **Parametric EB (PEB)**: We assume the prior for \( \theta \) is in some parametric class \( p(\theta | \eta) \), where only \( \eta \) (the hyperparameter) is unknown.
  - estimate the hyperparameter by \( \hat{\eta} \), and plug in to get the estimated posterior.
  - well-developed in a series of papers by Efron and Morris in the mid-70’s.

- **Nonparametric EB (NPEB)**: We assume only that the \( \theta_i \) are *iid* from some distribution \( p \).
Morris ('83 JASA) classified EB methods into two categories: parametric and nonparametric:

- **Parametric EB (PEB):** We assume the prior for $\theta$ is in some parametric class $p(\theta|\eta)$, where only $\eta$ (the hyperparameter) is unknown.
  - estimate the hyperparameter by $\hat{\eta}$, and plug in to get the estimated posterior.
  - well-developed in a series of papers by Efron and Morris in the mid-70’s.

- **Nonparametric EB (NPEB):** We assume only that the $\theta_i$ are iid from some distribution $p$.
  - Use the data to estimate the prior or the marginal distribution directly.
The empirical Bayes approach

Morris (’83 *JASA*) classified EB methods into two categories: *parametric* and *nonparametric*:

- **Parametric EB (PEB):** We assume the prior for $\theta$ is in some parametric class $p(\theta | \eta)$, where only $\eta$ (the hyperparameter) is unknown.
  - estimate the hyperparameter by $\hat{\eta}$, and plug in to get the estimated posterior.
  - well-developed in a series of papers by Efron and Morris in the mid-70’s.

- **Nonparametric EB (NPEB):** We assume only that the $\theta_i$ are *iid* from some distribution $p$.
  - Use the data to estimate the prior or the marginal distribution *directly*.
  - pioneered/championed by Robbins (1950’s; actually older than PEB)
Nonparametric EB basics

Start with compound sampling model:

\[ y_i | \theta_i \overset{iid}{\sim} f(y_i | \theta_i) = \text{Poisson}(\theta_i) \text{ and } \theta_i \overset{iid}{\sim} p(\cdot), \ i = 1, \ldots, k \]
Nonparametric EB basics

- Start with **compound sampling** model:

\[ y_i | \theta_i \overset{iid}{\sim} f(y_i | \theta_i) = \text{Poisson}(\theta_i) \text{ and } \theta_i \overset{iid}{\sim} p(\cdot), \ i = 1, \ldots, k \]

- Under squared error loss, the Bayes estimate is posterior mean:

\[
\hat{\theta}_i(y) = \frac{E_G(\theta_i | y)}{\int \frac{u^{y_i+1}}{y_i!} e^{-u} dG(u)} = \frac{(y_i+1)m_G(y_i+1)}{m_G(y_i)}. 
\]
Nonparametric EB basics

- Start with compound sampling model:

\[ y_i | \theta_i \overset{iid}{\sim} f(y_i | \theta_i) = \text{Poisson}(\theta_i) \quad \text{and} \quad \theta_i \overset{iid}{\sim} p(\cdot), \quad i = 1, \ldots, k \]

- Under squared error loss, the Bayes estimate is posterior mean:

\[
\hat{\theta}_i(y) = E_G(\theta_i | y) = \frac{\int u^{y_i+1} e^{-u} dG(u)}{\int u^{y_i} e^{-u} dG(u)} = \frac{(y_i + 1)m_G(y_i + 1)}{m_G(y_i)}.
\]

⇒ The “Robbins Miracle”: \( \hat{\theta}_i(y) \) is directly estimable as

\[
\hat{\theta}_i(y) = \frac{(y_i + 1)m_G(y_i + 1)}{m_G(y_i)} = \frac{(y_i + 1)[\#y's = y_i + 1]}{[\#y's = y_i]}.
\]
Nonparametric EB summary

Maritz and Lwin (1988) discuss “Simple EB,” a generalization of this idea for non-Poisson families. But can’t take it very far...
Nonparametric EB summary

Maritz and Lwin (1988) discuss “Simple EB,” a generalization of this idea for non-Poisson families. But can’t take it very far...

New idea: Use $m_{\hat{G}}(y_i + 1)$ instead of $\hat{m}_G(y_i + 1)$
Nonparametric EB summary

- Maritz and Lwin (1988) discuss “Simple EB,” a generalization of this idea for non-Poisson families. But can’t take it very far...

- **New idea:** Use \( \hat{m}_G(y_i + 1) \) instead of \( \hat{m}_G(y_i + 1) \)

- Computationally more feasible, and enables imposition of appropriate structure (monotonicity, convexity, etc.) that the empirical cdf doesn’t impose.
Nonparametric EB summary

Maritz and Lwin (1988) discuss “Simple EB,” a generalization of this idea for non-Poisson families. But can’t take it very far...

New idea: Use \( \hat{m}_G(y_i + 1) \) instead of \( \hat{m}_G(y_i + 1) \)

Computationally more feasible, and enables imposition of appropriate structure (monotonicity, convexity, etc.) that the empirical cdf doesn’t impose.

The maximizing \( G \) (the NPML) is a finite mixture of degenerate distributions with no more than \( k \) mass points, computable via the Expectation-Maximization (EM) algorithm (C&L Sec 5.2.2).
Nonparametric EB summary

Maritz and Lwin (1988) discuss “Simple EB,” a generalization of this idea for non-Poisson families. But can’t take it very far...

New idea: Use $\hat{m}_G (y_i + 1)$ instead of $\hat{m}_G (y_i + 1)$

Computationally more feasible, and enables imposition of appropriate structure (monotonicity, convexity, etc.) that the empirical cdf doesn’t impose.

The maximizing $G$ (the NPML) is a finite mixture of degenerate distributions with no more than $k$ mass points, computable via the Expectation-Maximization (EM) algorithm (C&L Sec 5.2.2).

On the whole NPEB, can do quite well in wide variety of scenarios (i.e., when true prior is bimodal), and has spawned research into fully Bayesian nonparametric approaches (C&L Sec 2.6).
Parametric EB basics

Stage 1: $Y_i | \theta_i \sim \text{indep} f_i(y_i | \theta_i), i = 1, \ldots, k$
Parametric EB basics

Stage 1: \[ Y_i | \theta_i \overset{\text{indep}}{\sim} f_i(y_i | \theta_i), \ i = 1, \ldots, k \]

Stage 2: \[ \theta_i \overset{iid}{\sim} p(\theta_i | \eta), \ i = 1, \ldots, k \]
Parametric EB basics

- Stage 1: $Y_i | \theta_i \overset{indep}{\sim} f_i(y_i | \theta_i), i = 1, \ldots, k$
- Stage 2: $\theta_i \overset{iid}{\sim} p(\theta_i | \eta), i = 1, \ldots, k$

Suppose we seek point estimates for the $\theta_i$. The marginal distribution of $y = (y_1, \ldots, y_k)$ is

$$m(y | \eta) = \int f(y | \theta)p(\theta | \eta)d\theta$$

$$= \int \left[ \prod_{i=1}^{k} f_i(y_i | \theta_i) \right] \left[ \prod_{i=1}^{k} p(\theta_i | \eta) \right] d\theta$$

$$= \prod_{i=1}^{k} \int f_i(y_i | \theta_i)p(\theta_i | \eta)d\theta_i = \prod_{i=1}^{k} m_i(y_i | \eta)$$
Parametric EB basics

• Stage 1: $Y_i | \theta_i \overset{\text{indep}}{\sim} f_i(y_i | \theta_i), i = 1, \ldots, k$

• Stage 2: $\theta_i \overset{iid}{\sim} p(\theta_i | \eta), i = 1, \ldots, k$

• Suppose we seek point estimates for the $\theta_i$. The marginal distribution of $y = (y_1, \ldots, y_k)$ is

$$m(y | \eta) = \int f(y | \theta)p(\theta | \eta)d\theta$$

$$= \int \left[ \prod_{i=1}^{k} f_i(y_i | \theta_i) \right] \left[ \prod_{i=1}^{k} p(\theta_i | \eta) \right] d\theta$$

$$= \prod_{i=1}^{k} \int f_i(y_i | \theta_i)p(\theta_i | \eta)d\theta_i = \prod_{i=1}^{k} m_i(y_i | \eta)$$

$\Rightarrow y_i$ are marginally independent (and iid if $f_i = f$ for all $i$)
Similarly, the posterior for $\theta_i$ depends on the data only through $y_i$, namely

$$p(\theta_i|y_i, \eta) = \frac{f_i(y_i|\theta_i)p(\theta_i|\eta)}{m_i(y_i|\eta)}$$
Similarly, the posterior for $\theta_i$ depends on the data only through $y_i$, namely

$$p(\theta_i | y_i, \eta) = \frac{f_i(y_i | \theta_i) p(\theta_i | \eta)}{m_i(y_i | \eta)}$$

But if we assume $\eta$ is unknown and estimate it from the marginal distribution of all the data, $m(y | \eta)$, we get the estimated posterior,

$$p(\theta_i | y_i, \hat{\eta})$$

where $\hat{\eta} = \hat{\eta}(y)$, usually obtained as a MLE or method of moments (MOM) estimate from $m(y | \eta)$. 
Parametric EB basics (cont’d)

- Similarly, the posterior for \( \theta_i \) depends on the data only through \( y_i \), namely

\[
p(\theta_i | y_i, \eta) = \frac{f_i(y_i | \theta_i) p(\theta_i | \eta)}{m_i(y_i | \eta)}
\]

- But if we assume \( \eta \) is unknown and estimate it from the marginal distribution of all the data, \( m(y | \eta) \), we get the estimated posterior,

\[
p(\theta_i | y_i, \hat{\eta})
\]

where \( \hat{\eta} = \hat{\eta}(y) \), usually obtained as a MLE or method of moments (MOM) estimate from \( m(y | \eta) \).

- Now take \( \hat{\theta}_i \) to be the mean of the estimated posterior. Note that \( \hat{\theta}_i \) depends on all the data through \( \hat{\eta} \).
Example: Normal/Normal model

\[ y_i \mid \theta_i \overset{ind}{\sim} N(\theta_i, \sigma^2), \ i = 1, \ldots, k, \ \sigma^2 \text{ known}; \]
\[ \theta_i \overset{iid}{\sim} N(\mu, \tau^2), \ i = 1, \ldots, k, \ (\mu, \tau^2) \text{ both unknown}. \]

- We know \( m(y_i \mid \mu, \tau^2) = N(\mu, \sigma^2 + \tau^2) \), so

\[
m(y \mid \mu, \tau^2) = \prod_{i=1}^{k} \left\{ \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)^{1/2}}} \exp \left[ -\frac{(y_i - \mu)^2}{2(\sigma^2 + \tau^2)} \right] \right\}.
\]
Example: Normal/Normal model

\[ y_i \mid \theta_i \sim \text{ind} N(\theta_i, \sigma^2), \quad i = 1, \ldots, k, \quad \sigma^2 \text{ known} ; \]
\[ \theta_i \sim \text{iid} N(\mu, \tau^2), \quad i = 1, \ldots, k, \quad (\mu, \tau^2) \text{ both unknown}. \]

We know \( m(y_i \mid \mu, \tau^2) = N(\mu, \sigma^2 + \tau^2) \), so

\[
m(y \mid \mu, \tau^2) = \prod_{i=1}^{k} \left\{ \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)^{1/2}}} \exp \left[ -\frac{(y_i - \mu)^2}{2(\sigma^2 + \tau^2)} \right] \right\}.
\]

Maximizing this as a function of \((\mu, \tau^2)\), we get

\[
\hat{\mu} = \bar{y} \quad \text{and} \quad \hat{\tau}^2 = (s^2 - \sigma^2)^+ \equiv \max \{0, s^2 - \sigma^2\},
\]

where \( \bar{y} = \frac{1}{k} \sum y_i \) and \( s^2 = \frac{1}{k} \sum (y_i - \bar{y})^2 \).
Example: Normal/Normal model

Thus the estimated posterior is

\[
p(\theta_i | y_i, \hat{\mu}, \hat{\tau}^2) = N \left( \hat{B} \hat{\mu} + (1 - \hat{B}) y_i, (1 - \hat{B}) \sigma^2 \right),
\]

where \( \hat{\mu} = \bar{y} \) and \( \hat{B} = \frac{\sigma^2}{\sigma^2 + \hat{\tau}^2} = \frac{\sigma^2}{\sigma^2 + (s^2 - \sigma^2)^+} \in [0, 1]. \)
Example: Normal/Normal model

- Thus the estimated posterior is

\[
p(\theta_i|y_i, \hat{\mu}, \tau^2) = N \left( \hat{B}\hat{\mu} + (1 - \hat{B})y_i, (1 - \hat{B})\sigma^2 \right),
\]

where \( \hat{\mu} = \bar{y} \) and \( \hat{B} = \frac{\sigma^2}{\sigma^2 + \tau^2} = \frac{\sigma^2}{\sigma^2 + (s^2 - \sigma^2)} \in [0, 1]. \)

- The PEB point estimator is the mean of this dist’n:

\[
\hat{\theta}_{i \text{PEB}} = \hat{B}\hat{\mu} + (1 - \hat{B})y_i = \hat{B}\bar{y} + (1 - \hat{B})y_i
\]

This is sometimes called a “shrinkage” estimator, since every point estimate will be “shrunk back” toward the grand mean \( \bar{y} \) from its original estimate \( y_i \). Also, \( \hat{B} \) is sometimes called a “shrinkage factor.”
Thus the estimated posterior is

\[ p(\theta_i|y_i, \hat{\mu}, \hat{\tau}^2) = N \left( \hat{B}\hat{\mu} + (1 - \hat{B})y_i, (1 - \hat{B})\sigma^2 \right) , \]

where \( \hat{\mu} = \bar{y} \) and

\[ \hat{B} = \frac{\sigma^2}{\sigma^2 + \hat{\tau}^2} = \frac{\sigma^2}{\sigma^2 + (s^2 - \sigma^2)} + \in [0, 1]. \]

The PEB point estimator is the mean of this dist’n:

\[ \hat{\theta}_i^{PEB} = \hat{B}\hat{\mu} + (1 - \hat{B})y_i = \hat{B}\bar{y} + (1 - \hat{B})y_i \]

This is sometimes called a “shrinkage” estimator, since every point estimate will be “shrunk back” toward the grand mean \( \bar{y} \) from its original estimate \( y_i \). Also, \( \hat{B} \) is sometimes called a “shrinkage factor.”

Intuitively, shrinkage makes sense here: problems are independent, but similar.
**Illustration: Morris’ Baseball Data**

<table>
<thead>
<tr>
<th>$i$</th>
<th>player</th>
<th>$y_i$</th>
<th>$\theta_i$</th>
<th>$i$</th>
<th>player</th>
<th>$y_i$</th>
<th>$\theta_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Clemente</td>
<td>.400</td>
<td>.346</td>
<td>10</td>
<td>Swoboda</td>
<td>.244</td>
<td>.230</td>
</tr>
<tr>
<td>2</td>
<td>F. Robinson</td>
<td>.378</td>
<td>.298</td>
<td>11</td>
<td>Unser</td>
<td>.222</td>
<td>.264</td>
</tr>
<tr>
<td>3</td>
<td>F. Howard</td>
<td>.356</td>
<td>.276</td>
<td>12</td>
<td>Williams</td>
<td>.222</td>
<td>.256</td>
</tr>
<tr>
<td>4</td>
<td>Johnstone</td>
<td>.333</td>
<td>.222</td>
<td>13</td>
<td>Scott</td>
<td>.222</td>
<td>.303</td>
</tr>
<tr>
<td>5</td>
<td>Berry</td>
<td>.311</td>
<td>.273</td>
<td>14</td>
<td>Petrocelli</td>
<td>.222</td>
<td>.264</td>
</tr>
<tr>
<td>6</td>
<td>Spencer</td>
<td>.311</td>
<td>.270</td>
<td>15</td>
<td>E. Rodriguez</td>
<td>.222</td>
<td>.226</td>
</tr>
<tr>
<td>7</td>
<td>Kessinger</td>
<td>.289</td>
<td>.263</td>
<td>16</td>
<td>Campaneris</td>
<td>.200</td>
<td>.285</td>
</tr>
<tr>
<td>8</td>
<td>L. Alvarado</td>
<td>.267</td>
<td>.210</td>
<td>17</td>
<td>Munson</td>
<td>.178</td>
<td>.316</td>
</tr>
<tr>
<td>9</td>
<td>Santo</td>
<td>.244</td>
<td>.269</td>
<td>18</td>
<td>Alvis</td>
<td>.156</td>
<td>.200</td>
</tr>
</tbody>
</table>

For players $i = 1, \ldots, 18$, 

\[
y_i = \text{batting average after first 45 at bats in 1970},
\]

\[
\theta_i = \text{true 1970 batting ability}
\]

(pretend the final 1970 averages measure this)
Data: $\bar{y} = .265$, $\hat{B} = .788$. 
Illustration: Morris’ Baseball Data

- Data: \( \bar{y} = .265, \; \hat{B} = .788. \)
- Use our normal/normal EB model, so that

\[
\hat{\theta}^{PEB}_i = \hat{B}\bar{y} + (1 - \hat{B})y_i = .788(.265) + .212 y_i
\]
Illustration: Morris’ Baseball Data

- **Data:** \( \bar{y} = .265, \ \hat{B} = .788. \)

- **Use our normal/normal EB model, so that**

  \[
  \hat{\theta}_{i}^{PEB} = \hat{B}\bar{y} + (1 - \hat{B})y_i = .788(.265) + .212y_i
  \]

- **Results** show that the PEB point estimates work well:
Illustration: Morris’ Baseball Data

- **Data:** \( \bar{y} = 0.265, \; \hat{B} = 0.788. \)
- **Use our normal/normal EB model, so that**
  \[
  \hat{\theta}_i^{PEB} = \hat{B}\bar{y} + (1 - \hat{B})y_i = 0.788(0.265) + 0.212y_i
  \]
- **Results** show that the PEB point estimates work well:
  - **individually:** in 16 of the 18 cases,
    \[
    (\hat{\theta}_i^{PEB} - \theta_i)^2 < (y_i - \theta_i)^2 \text{ (smaller individual risk)}
    \]
Data: \( \bar{y} = .265, \ B = .788. \)

Use our normal/normal EB model, so that

\[
\hat{\theta}_i^{PEB} = B\bar{y} + (1 - B)y_i = .788(.265) + .212y_i
\]

Results show that the PEB point estimates work well:

- individually: in 16 of the 18 cases,
  \[
  (\hat{\theta}_i^{PEB} - \theta_i)^2 < (y_i - \theta_i)^2
  \]
  (smaller individual risk)

- overall: aggregate MSE numbers are:
  
  \[
  MSE(y) = \sum_{i=1}^{18} (y_i - \theta_i)^2 = .077
  \]
  
  \[
  MSE(\hat{\theta}^{PEB}) = \sum_{i=1}^{18} (\hat{\theta}_i^{PEB} - \theta_i)^2 = .022
  \]
  (PEB has smaller ensemble risk)
Theoretical support for PEB

It turns out that the PEB estimate will always have lower ensemble risk in this setting provided $k \geq 3$!
Theoretical support for PEB

- It turns out that the PEB estimate will always have lower ensemble risk in this setting provided $k \geq 3$!

- Surprising! The PEB estimate has better frequentist risk (MSE) properties than the usual, “unshrunken” estimate, which itself is the MLE, UMVUE, etc.
Theoretical support for PEB

- It turns out that the PEB estimate will always have lower ensemble risk in this setting provided $k \geq 3$!

- Surprising! The PEB estimate has better frequentist risk (MSE) properties than the usual, “unshrunk” estimate, which itself is the MLE, UMVUE, etc.

- This general area is called Stein Estimation (Stein, 1955; James and Stein, 1961)
Theoretical support for PEB

- It turns out that the PEB estimate will always have lower ensemble risk in this setting provided $k \geq 3$!

- Surprising! The PEB estimate has better frequentist risk (MSE) properties than the usual, “unshrunk” estimate, which itself is the MLE, UMVUE, etc.

- This general area is called **Stein Estimation** (Stein, 1955; James and Stein, 1961)

- In summary, PEB point estimates have excellent ensemble risk performance, with respect to either:
Theoretical support for PEB

- It turns out that the PEB estimate will always have lower ensemble risk in this setting provided $k \geq 3$!
- Surprising! The PEB estimate has better frequentist risk (MSE) properties than the usual, “unshrunk” estimate, which itself is the MLE, UMVUE, etc.
- This general area is called Stein Estimation (Stein, 1955; James and Stein, 1961)
- In summary, PEB point estimates have excellent ensemble risk performance, with respect to either:
  - frequentist risk: $E_{Y|\theta} L(\theta, \hat{\theta}(y))$
Theoretical support for PEB

- It turns out that the PEB estimate will **always** have lower ensemble risk in this setting **provided** $k \geq 3$!

- **Surprising!** The PEB estimate has better frequentist risk (MSE) properties than the usual, “unshrunk” estimate, which itself is the MLE, UMVUE, etc.

- This general area is called Stein Estimation (Stein, 1955; James and Stein, 1961)

- In summary, PEB point estimates have excellent ensemble risk performance, with respect to either:
  - frequentist risk: $E_{Y|\theta} L(\theta, \hat{\theta}(y))$
  - preposterior (or “EB”) risk:

$$E_{\theta,Y} L(\theta, \hat{\theta}(y)) = E_{\theta} E_{Y|\theta} L(\theta, \hat{\theta}(y)) = E_{Y} E_{\theta|Y} L(\theta, \hat{\theta}(y))$$
What about EB interval estimation?

Taking the upper and lower $\alpha/2$-points of the estimated posterior $p(\theta_i | y, \hat{\eta})$ gives a $100 \times (1 - \alpha)$% credible set for $\theta_i$:

$$P (\theta_i \leq q_{\alpha}(y_i, \eta) \mid \theta_i \sim p(\theta_i | y_i, \eta)) = \alpha,$$

then the naive EBCI is $(q_{\alpha/2}(y_i, \hat{\eta}), q_{1-(\alpha/2)}(y_i, \hat{\eta}))$. 

What about EB *interval* estimation?

- Taking the upper and lower $\alpha/2$-points of the estimated posterior $p(\theta_i | y, \hat{\eta})$ gives a $100 \times (1 - \alpha)$% credible set for $\theta_i$:

  \[ P(\theta_i \leq q_\alpha(y_i, \eta) \mid \theta_i \sim p(\theta_i | y_i, \eta)) = \alpha, \]

  then the *naive EBCI* is $(q_{\alpha/2}(y_i, \hat{\eta}), q_{1-(\alpha/2)}(y_i, \hat{\eta}))$.

- In the normal/normal, the 95% naive EBCI is

  \[ E(\theta_i | y_i, \hat{\eta}) \pm 1.96 \sqrt{\text{Var}(\theta_i | y_i, \hat{\eta})}. \]
What about EB *interval* estimation?

- Taking the upper and lower $\alpha/2$-points of the estimated posterior $p(\theta_i|y, \hat{\eta})$ gives a $100 \times (1 - \alpha)$% credible set for $\theta_i$:

  $$P(\theta_i \leq q_\alpha(y_i, \eta) \mid \theta_i \sim p(\theta_i|y_i, \eta)) = \alpha,$$

  then the naive EBCI is $(q_{\alpha/2}(y_i, \hat{\eta}), q_{1-(\alpha/2)}(y_i, \hat{\eta}))$.

- In the normal/normal, the 95% naive EBCI is

  $$E(\theta_i \mid y_i, \hat{\eta}) \pm 1.96 \sqrt{\text{Var}(\theta_i \mid y_i, \hat{\eta})}.$$

- “Naive” since the variance approximates only the first term in the true posterior variance,

  $$\text{Var}(\theta_i|y) = E_{\eta|y} [\text{Var}(\theta_i|y_i, \eta)] + \text{Var}_{\eta|y} [E(\theta_i|y_i, \eta)].$$

The naive EBCI is ignoring the posterior uncertainty about $\eta \Rightarrow$ naive interval may be too short.
Possible remedies for EBCIs

- Morris: get a “plug in” estimate for $\text{Var}_{\eta|y}[E(\theta_i|y_i, \eta)]$
Possible remedies for EBCIs

- **Morris**: get a “plug in” estimate for $\text{Var}_{\eta|y}[E(\theta_i|y_i, \eta)]$

- **bias corrected naive method**: solve

  $$E_{\hat{\eta},y_i|\eta}P(\theta_i \leq q_{\alpha}(y_i, \hat{\eta}) \mid \theta_i \sim p(\theta_i|y_i, \eta)) = \alpha$$

  for $\alpha' = \alpha'(\hat{\eta}, \alpha)$, and take the naive interval with $\alpha$ replaced by $\alpha'$. 
Possible remedies for EBCIs

- **Morris:** get a “plug in” estimate for $\text{Var}_{\eta|y}[E(\theta_i|y_i, \eta)]$

- **bias corrected naive method:** solve

\[
E_{\hat{\eta},y_i|\eta}P(\theta_i \leq q_\alpha(y_i, \hat{\eta}) \mid \theta_i \sim p(\theta_i|y_i, \eta)) = \alpha
\]

for $\alpha' = \alpha'(\hat{\eta}, \alpha)$, and take the naive interval with $\alpha$ replaced by $\alpha'$.

- **marginal posterior approach:** place a hyperprior $\psi(\eta)$ on $\eta$, and base EBCI for $\theta_i$ on the marginal posterior,

\[
l_h(\theta_i|y) = \int p(\theta_i|y, \eta)h(\eta|y)d\eta ,
\]

where $h(\eta|y) \propto m(y|\eta)\psi(\eta) \leftarrow \text{“Bayes empirical Bayes”}$. 

Chapter 5: The Empirical Bayes Approach – p. 13/14
Possible remedies for EBCIs

- **Morris**: get a “plug in” estimate for \( \text{Var}_{\eta|y} [E(\theta_i|y_i, \eta)] \)

- **bias corrected naive method**: solve

\[
E_{\hat{\eta}, y_i|\eta} P(\theta_i \leq q_\alpha(y_i, \hat{\eta}) \mid \theta_i \sim p(\theta_i|y_i, \eta)) = \alpha
\]

for \( \alpha' = \alpha'(\hat{\eta}, \alpha) \), and take the naive interval with \( \alpha \) replaced by \( \alpha' \).

- **marginal posterior approach**: place a hyperprior \( \psi(\eta) \) on \( \eta \), and base EBCI for \( \theta_i \) on the marginal posterior,

\[
l_h(\theta_i|y) = \int p(\theta_i|y, \eta) h(\eta|y) d\eta ,
\]

where \( h(\eta|y) \propto m(y|\eta)\psi(\eta) \leftarrow \text{“Bayes empirical Bayes”} \)

- **Trouble**: All of these are hard to do outside of low-dimensional, conjugate, two-stage models!
So dump EB and return to full Bayes?

- EB still easier for non-Bayesians to accept
So dump EB and return to full Bayes?

- EB still easier for non-Bayesians to accept
- Bayes requires **MCMC** computing, which can be tricky and may tempt us to fit models larger than the data can support – or we can understand...
So dump EB and return to full Bayes?

- EB still easier for non-Bayesians to accept
- Bayes requires MCMC computing, which can be tricky and may tempt us to fit models larger than the data can support – or we can understand...
- **Example:** Choice of “vague” hyperprior for a variance component $\tau^2$. Most widespread current choice:

$$
\tau^2 \sim \text{Gamma}(\epsilon, \epsilon); \quad E(\tau^2) = 1, \ Var(\tau^2) = 1/\epsilon.
$$
So dump EB and return to full Bayes?

- EB still easier for non-Bayesians to accept
- Bayes requires MCMC computing, which can be tricky and may tempt us to fit models larger than the data can support – or we can understand...
- **Example:** Choice of “vague” hyperprior for a variance component $\tau^2$. Most widespread current choice:

$$
\tau^2 \sim \text{Gamma}(\epsilon, \epsilon); \quad E(\tau^2) = 1, \quad \text{Var}(\tau^2) = 1/\epsilon.
$$

- But recent work shows that such hyperpriors can actually have significant impact on the resulting posteriors (at least for the variances)
- lead to MCMC convergence failure – or apparent convergence when joint posterior is improper!
So dump EB and return to full Bayes?

- EB still easier for non-Bayesians to accept
- Bayes requires **MCMC** computing, which can be tricky and may tempt us to fit models larger than the data can support – or we can understand...

**Example:** Choice of “vague” hyperprior for a variance component \( \tau^2 \). Most widespread current choice:

\[
\tau^2 \sim \text{Gamma}(\epsilon, \epsilon); \quad E(\tau^2) = 1, \quad \text{Var}(\tau^2) = 1/\epsilon.
\]

- But recent work shows that such hyperpriors can actually have significant impact on the resulting posteriors (at least for the variances)
- lead to MCMC convergence failure – or apparent convergence when joint posterior is improper!
- EB approach (replacing \( \tau^2 \) by \( \hat{\tau}^2 \)) may produce estimates that are still improved, yet safer to use.