Foundations of Statistical Inference

Julien Berestycki

Department of Statistics
University of Oxford

MT 2016

Lecture 13 : Empirical Bayes.

Empirical Bayes

Bayes estimators have good risk properties (for example, the posterior mean is usually admissible for quadratic loss).

Empirical Bayes

Bayes estimators have good risk properties (for example, the posterior mean is usually admissible for quadratic loss).

However, Bayes estimators may be hard to compute (for example, the posterior mean is an integral, or sum), particularly for hierarchical models.

Empirical Bayes

Bayes estimators have good risk properties (for example, the posterior mean is usually admissible for quadratic loss).

However, Bayes estimators may be hard to compute (for example, the posterior mean is an integral, or sum), particularly for hierarchical models.

In Empirical Bayes, we use Bayesian reasoning to find estimators which can then be used in classical frequentist. EB uses a particular strategy to simplify hierarchical models.

Empirical Bayes

Recall the setup for Bayesian inference for hierarchical models.

$$\begin{aligned} X &\sim f(x;\theta) \\ \theta &\sim \pi(\theta;\psi) \\ \psi &\sim g(\psi) \end{aligned}$$

Empirical Bayes

Recall the setup for Bayesian inference for hierarchical models.

$$\begin{aligned} X &\sim f(x;\theta) \\ \theta &\sim \pi(\theta;\psi) \\ \psi &\sim g(\psi) \end{aligned}$$

Our prior for θ has a parameter ψ which also has a prior. The posterior is

$$\pi(\theta,\psi|x) \propto L(\theta;x)\pi(\theta;\psi)g(\psi)$$

Empirical Bayes

Recall the setup for Bayesian inference for hierarchical models.

$$\begin{aligned} X &\sim f(x;\theta) \\ \theta &\sim \pi(\theta;\psi) \\ \psi &\sim g(\psi) \end{aligned}$$

Our prior for θ has a parameter ψ which also has a prior. The posterior is

$$\pi(\theta,\psi|x) \propto L(\theta;x)\pi(\theta;\psi)g(\psi)$$

If we want minimim risk for quadratic loss (for eg) we should use

$$\hat{\theta} = \int\int \theta\pi(\theta,\psi|x)d\theta d\psi$$

Empirical Bayes

EB

The EB trick is to avoid doing ψ-integrals by replacing ψ with an estimate $\hat{\psi}$, derived from the data, and consider the model

$$\begin{aligned} X &\sim f(x;\theta) \\ \theta &\sim \pi(\theta;\hat{\psi}) \end{aligned}$$

Empirical Bayes

EB

The EB trick is to avoid doing ψ-integrals by replacing ψ with an estimate $\hat{\psi}$, derived from the data, and consider the model

$$\begin{aligned} X &\sim f(x;\theta) \\ \theta &\sim \pi(\theta;\hat{\psi}) \end{aligned}$$

This EB approximation to the full posterior 'chops off' a layer of the hierarchy. The reduced model has posterior

$$\hat{\pi}(\theta|x) \propto L(\theta;x)\pi(\theta;\hat{\psi}),$$

and a Bayes estimator $\hat{\theta}_{EB}$ is calculated using $\hat{\pi}(\theta|x)$. For example, for quadratic loss,

$$\hat{\theta} = \int \theta \hat{\pi}(\theta|x) d\theta.$$

Empirical Bayes

We still need an estimator for ψ. There are several choices.

Empirical Bayes

We still need an estimator for ψ. There are several choices.

We can use the MLE $\hat{\psi} = \arg\max_{\psi} p(x|\psi)$ for ψ in the marginal likelihood

$$p(x|\psi) = \int L(\theta; x)\pi(\theta; \psi)d\theta.$$

Empirical Bayes

We still need an estimator for ψ. There are several choices.

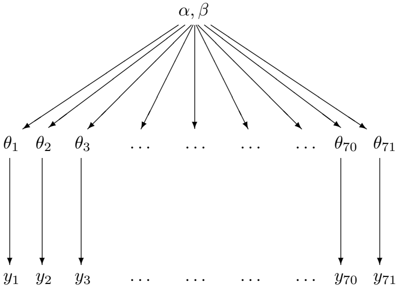
We can use the MLE $\hat{\psi} = \arg\max_{\psi} p(x|\psi)$ for ψ in the marginal likelihood

$$p(x|\psi) = \int L(\theta; x)\pi(\theta; \psi)d\theta.$$

Method of moments estimators are also used.

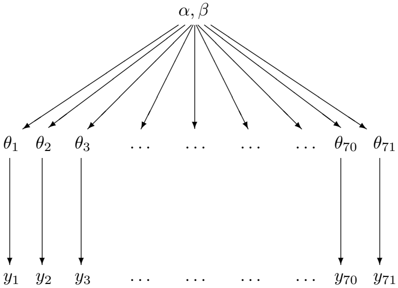
Example: recall from lecture 8

Study of tumors in rodents:



Example: recall from lecture 8

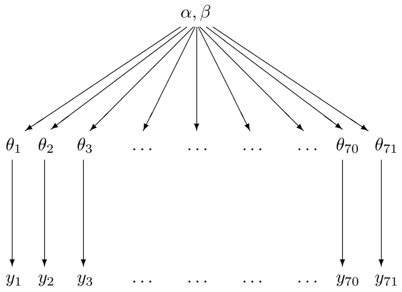
Study of tumors in rodents:



Assume # tumors $\sim$ Bin(n, θ) and a conjugate prior $\theta \sim$ Beta(α, β).

Example: recall from lecture 8

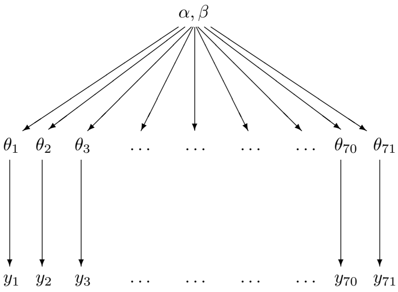
Study of tumors in rodents:



Assume # tumors $\sim$ Bin(n, θ) and a conjugate prior $\theta \sim$ Beta(α, β). New experiment $n = 14$ and $Y = 4$.

Example: recall from lecture 8

Study of tumors in rodents:



Assume # tumors $\sim$ Bin(n, θ) and a conjugate prior $\theta \sim$ Beta(α, β). New experiment $n = 14$ and $Y = 4$.Posterior is

$$p(\theta|y) = Beta(\alpha + 4, \beta + 10).$$

Example: recall from lecture 8

Empirical Bayes

Previous experiments:

0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/47	15/46	9/24

Current experiment:

4/14

Table 5.1 *Tumor incidence in historical control groups and current group of rats, from Tarone (1982). The table displays the values of* $\frac{y_j}{n_j}$*: (number of rats with tumors)/(total number of rats).*

Example: recall from lecture 8

Empirical Bayes

Previous experiments:

0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/47	15/46	9/24

Current experiment:
4/14

Table 5.1 *Tumor incidence in historical control groups and current group of rats, from Tarone (1982). The table displays the values of $\frac{y_j}{n_j}$: (number of rats with tumors)/(total number of rats).*

Moment estimators: Pick $\hat{\alpha}, \hat{\beta}$ to match empirical mean variance of the n_j/θ_j. Get $\hat{\alpha} = 1.4, \hat{\beta} = 8.6, p(\theta|y) \sim Beta(5.4, 18.6)$

Example: recall from lecture 8

Empirical Bayes

Previous experiments:

0/20	0/20	0/20	0/20	0/20	0/20	0/20	0/19	0/19	0/19
0/19	0/18	0/18	0/17	1/20	1/20	1/20	1/20	1/19	1/19
1/18	1/18	2/25	2/24	2/23	2/20	2/20	2/20	2/20	2/20
2/20	1/10	5/49	2/19	5/46	3/27	2/17	7/49	7/47	3/20
3/20	2/13	9/48	10/50	4/20	4/20	4/20	4/20	4/20	4/20
4/20	10/48	4/19	4/19	4/19	5/22	11/46	12/49	5/20	5/20
6/23	5/19	6/22	6/20	6/20	6/20	16/52	15/47	15/46	9/24

Current experiment:
4/14

Table 5.1 *Tumor incidence in historical control groups and current group of rats, from Tarone (1982). The table displays the values of* $\frac{y_j}{n_j}$*: (number of rats with tumors)/(total number of rats).*

Moment estimators: Pick $\hat{\alpha}, \hat{\beta}$ to match empirical mean variance of the n_j/θ_j. Get $\hat{\alpha} = 1.4, \hat{\beta} = 8.6, p(\theta|y) \sim Beta(5.4, 18.6)$
Posterior mean is 0.223, lower than $4/14 = 0.286$. Current experiment has unusually high number of tumors.

Example

The James-Stein estimator is an EB-estimator

Example

The James-Stein estimator is an EB-estimator

Data $x_i \sim N(\theta_i, 1)$, $i = 1, \ldots, p$ (so one observation x_i for each parameter θ_i). The MLE for θ_i is simply $\hat{\theta}_{MLE,i} = x_i$. Construct an EB estimator for quadratic loss.

Example

The James-Stein estimator is an EB-estimator

Data $x_i \sim N(\theta_i, 1)$, $i = 1, \dots, p$ (so one observation x_i for each parameter θ_i). The MLE for θ_i is simply $\hat{\theta}_{MLE,i} = x_i$. Construct an EB estimator for quadratic loss.

Suppose the prior is $\theta_i \sim N(0, \tau^2)$ (we have some freedom here, as we are mainly interested in the risk-related properties of the final estimator).

Example

The James-Stein estimator is an EB-estimator

Data $x_i \sim N(\theta_i, 1)$, $i = 1, \ldots, p$ (so one observation x_i for each parameter θ_i). The MLE for θ_i is simply $\hat{\theta}_{MLE,i} = x_i$. Construct an EB estimator for quadratic loss.

Suppose the prior is $\theta_i \sim N(0, \tau^2)$ (we have some freedom here, as we are mainly interested in the risk-related properties of the final estimator).

If we knew τ we would have (completing the square - see Lecture 7 p16)

$$\theta_i | (x_i, \tau) \sim N\left(\frac{x_i \tau^2}{1+\tau^2}, \frac{\tau^2}{1+\tau^2}\right).$$

Example

To get an estimate for τ we compute the marginal distribution for X_i given τ, which is $X_i \sim N(0, \tau^2 + 1)$. The MLE for τ is then $\hat{\tau}^2 = \frac{1}{p} \sum_i X_i^2 - 1$, and this gives

$$\begin{aligned} \hat{\theta}_{EB,i} &= \frac{X_i \hat{\tau}^2}{1 + \hat{\tau}^2} \\ &= \left(1 - \frac{p}{\sum_i X_i^2}\right) X_i \end{aligned}$$

which is the James-Stein estimator

$$\hat{\theta}_{JS,i} = \left(1 - \frac{a}{\sum_i X_i^2}\right) X_i$$

with $a = p$. This isn't the minimum risk JS estimator for quadratic loss, (that is $a = p - 2$) but it already beats the MLE for all θ. (to get the best JS estimator (with $a = p - 2$) use a method of moments estimator for τ. See Young and Smith Section 3.5)

Example: Poisson

Data $x_i \sim$ Poisson(θ_i), $i = 1, \ldots, n$ (so one observation x_i for each parameter θ_i).

Example: Poisson

Data $x_i \sim$ Poisson(θ_i), $i = 1, \ldots, n$ (so one observation x_i for each parameter θ_i). The MLE for θ_i is simply $\hat{\theta}_{MLE,i} = x_i$.

Example: Poisson

Data $x_i \sim$ Poisson(θ_i), $i = 1, \ldots, n$ (so one observation x_i for each parameter θ_i). The MLE for θ_i is simply $\hat{\theta}_{MLE,i} = x_i$. Construct an EB estimator for quadratic loss.

Example: Poisson

Data $x_i \sim$ Poisson(θ_i), $i = 1, \dots, n$ (so one observation x_i for each parameter θ_i). The MLE for θ_i is simply $\hat{\theta}_{MLE,i} = x_i$. Construct an EB estimator for quadratic loss.

Suppose the prior for θ_i's is iid Exponential(λ) i.e. $\pi(\theta_i|\lambda) = \lambda e^{-\lambda\theta_i}$.

$$p(x_i \mid \lambda) = \int_0^\infty \frac{e^{-\theta_i}\theta_i^{x_i}}{x_i!}\lambda e^{-\lambda\theta_i} d\theta_i$$

Example: Poisson

Data $x_i \sim$ Poisson(θ_i), $i = 1, \ldots, n$ (so one observation x_i for each parameter θ_i). The MLE for θ_i is simply $\hat{\theta}_{MLE,i} = x_i$. Construct an EB estimator for quadratic loss.

Suppose the prior for θ_i's is iid Exponential(λ) i.e. $\pi(\theta_i|\lambda) = \lambda e^{-\lambda\theta_i}$.

$$
\begin{aligned}
p(x_i \mid \lambda) &= \int_0^\infty \frac{e^{-\theta_i}\theta_i^{x_i}}{x_i!}\lambda e^{-\lambda\theta_i} d\theta_i \\
&= \left(\frac{1}{1+\lambda}\right)^{x_i} \frac{\lambda}{1+\lambda}
\end{aligned}
$$

Example: Poisson

Data $x_i \sim$ Poisson(θ_i), $i = 1, \ldots, n$ (so one observation x_i for each parameter θ_i). The MLE for θ_i is simply $\hat{\theta}_{MLE,i} = x_i$. Construct an EB estimator for quadratic loss.

Suppose the prior for θ_i's is iid Exponential(λ) i.e. $\pi(\theta_i|\lambda) = \lambda e^{-\lambda\theta_i}$.

$$
\begin{aligned}
p(x_i \mid \lambda) &= \int_0^\infty \frac{e^{-\theta_i}\theta_i^{x_i}}{x_i!}\lambda e^{-\lambda\theta_i} d\theta_i \\
&= \left(\frac{1}{1+\lambda}\right)^{x_i} \frac{\lambda}{1+\lambda}
\end{aligned}
$$

$\Rightarrow$ given λ the $x_i's$ are iid geometric $(\lambda/(1+\lambda))$. Mean is $\frac{1-p}{p} = \lambda^{-1}$

Example: Poisson

Data $x_i \sim$ Poisson(θ_i), $i = 1, \ldots, n$ (so one observation x_i for each parameter θ_i). The MLE for θ_i is simply $\hat{\theta}_{MLE,i} = x_i$. Construct an EB estimator for quadratic loss.

Suppose the prior for θ_i's is iid Exponential(λ) i.e. $\pi(\theta_i|\lambda) = \lambda e^{-\lambda\theta_i}$.

$$\begin{aligned} p(x_i \mid \lambda) &= \int_0^\infty \frac{e^{-\theta_i}\theta_i^{x_i}}{x_i!}\lambda e^{-\lambda\theta_i} d\theta_i \\ &= \left(\frac{1}{1+\lambda}\right)^{x_i} \frac{\lambda}{1+\lambda} \end{aligned}$$

$\Rightarrow$ given λ the $x_i's$ are iid geometric $(\lambda/(1+\lambda))$. Mean is $\frac{1-p}{p} = \lambda^{-1}$

The MLE of λ based on $x_1, \ldots, x_n$ is $\widehat{\lambda} = 1/\bar{x}$, where $\bar{x} = \frac{1}{n}\sum_1^n x_i$.

Example: Poisson

Now, under the EB simplification, set $\lambda = \hat{\lambda}$, so that

$$\hat{\pi}(\theta|x) \propto L(\theta; x)\pi(\theta|\hat{\lambda}) \; = \; \prod_i e^{-\theta_i} \theta_i^{x_i} \hat{\lambda} e^{-\hat{\lambda}\theta_i}$$

Example: Poisson

Now, under the EB simplification, set $\lambda = \hat{\lambda}$, so that

$$\hat{\pi}(\theta|x) \propto L(\theta; x)\pi(\theta|\hat{\lambda}) \; = \; \prod_i e^{-\theta_i}\theta_i^{x_i}\hat{\lambda}e^{-\hat{\lambda}\theta_i}$$

and we recognise $\theta_i|x \sim \Gamma(x_i + 1, \hat{\lambda} + 1)$ in this EB approximation.

Example: Poisson

Now, under the EB simplification, set $\lambda = \hat{\lambda}$, so that

$$\hat{\pi}(\theta|x) \propto L(\theta; x)\pi(\theta|\hat{\lambda}) \; = \; \prod_i e^{-\theta_i}\theta_i^{x_i}\hat{\lambda}e^{-\hat{\lambda}\theta_i}$$

and we recognise $\theta_i|x \sim \Gamma(x_i + 1, \hat{\lambda} + 1)$ in this EB approximation. This leads to an estimator

$$\hat{\theta}_{EB,i} \;\; = \;\; \int \theta_i \hat{\pi}(\theta_i|x) d\theta_i$$

Example: Poisson

Now, under the EB simplification, set $\lambda = \hat{\lambda}$, so that

$$\hat{\pi}(\theta|x) \propto L(\theta; x)\pi(\theta|\hat{\lambda}) = \prod_i e^{-\theta_i}\theta_i^{x_i}\hat{\lambda}e^{-\hat{\lambda}\theta_i}$$

and we recognise $\theta_i|x \sim \Gamma(x_i + 1, \hat{\lambda} + 1)$ in this EB approximation. This leads to an estimator

$$\begin{aligned} \hat{\theta}_{EB,i} &= \int \theta_i \hat{\pi}(\theta_i|x) d\theta_i \\ &= \frac{x_i + 1}{\hat{\lambda} + 1} = \bar{x}\frac{x_i + 1}{\bar{x} + 1} \end{aligned}$$

Example: Poisson

Now, under the EB simplification, set $\lambda = \hat{\lambda}$, so that

$$\hat{\pi}(\theta|x) \propto L(\theta; x)\pi(\theta|\hat{\lambda}) \; = \; \prod_i e^{-\theta_i}\theta_i^{x_i}\hat{\lambda}e^{-\hat{\lambda}\theta_i}$$

and we recognise $\theta_i|x \sim \Gamma(x_i + 1, \hat{\lambda} + 1)$ in this EB approximation. This leads to an estimator

$$\begin{aligned} \hat{\theta}_{EB,i} &= \int \theta_i \hat{\pi}(\theta_i|x) d\theta_i \\ &= \frac{x_i + 1}{\hat{\lambda} + 1} = \bar{x}\frac{x_i + 1}{\bar{x} + 1} \end{aligned}$$

We can rewrite this

$$\hat{\theta}_{EB,i} = x_i + \frac{\bar{x}}{\bar{x} + 1}(\bar{x} - x_i)$$

showing that this EB estimator shrinks the estimates towards the common mean.

Example: Poisson

Suppose $X \sim \text{Poisson}(\theta)$ and $\theta \sim \Gamma(\alpha, \beta)$. Using

$$\begin{aligned} E[X] &= E_\psi[E_\theta(X|\theta)] \\ V(X) &= E_\psi[V_\theta(X|\theta)] + V_\psi[E_\theta(X|\theta)] \end{aligned}$$

We get

Example: Poisson

Suppose $X \sim \text{Poisson}(\theta)$ and $\theta \sim \Gamma(\alpha, \beta)$. Using

$$\begin{aligned} E[X] &= E_\psi[E_\theta(X|\theta)] \\ V(X) &= E_\psi[V_\theta(X|\theta)] + V_\psi[E_\theta(X|\theta)] \end{aligned}$$

We get

$$\begin{aligned} E[X] &= E_\psi[\theta] = \alpha/\beta \\ V(X) &= E_\psi[\theta] + V_\psi[\theta] = \alpha/\beta + \alpha/\beta^2. \end{aligned}$$

Hence, if $\bar{x}$ and s^2 are the sample mean and variance, the method of moments yields

Example: Poisson

Suppose $X \sim \text{Poisson}(\theta)$ and $\theta \sim \Gamma(\alpha, \beta)$. Using

$$
\begin{aligned}
E[X] &= E_\psi[E_\theta(X|\theta)] \\
V(X) &= E_\psi[V_\theta(X|\theta)] + V_\psi[E_\theta(X|\theta)]
\end{aligned}
$$

We get

$$
\begin{aligned}
E[X] &= E_\psi[\theta] = \alpha/\beta \\
V(X) &= E_\psi[\theta] + V_\psi[\theta] = \alpha/\beta + \alpha/\beta^2.
\end{aligned}
$$

Hence, if $\bar{x}$ and s^2 are the sample mean and variance, the method of moments yields

$$
\begin{aligned}
\hat{\alpha} &= \bar{x}^2/(s^2 - \bar{x}) \\
\hat{\beta} &= \bar{x}/(s^2 - \bar{x}).
\end{aligned}
$$

Hypothesis testing

$X_i \sim f(x, \theta_i)$ where the θ_i have a particular distribution $\pi(\theta)$.

Hypothesis testing

$X_i \sim f(x, \theta_i)$ where the θ_i have a particular distribution $\pi(\theta)$. We want to test $H_0 : \theta_k = \theta_0^*$ against $H_1 : \theta_k = \theta_1^*$. Since θ_i are iid let

$$p = P[\theta_i = \theta_0^*], \quad 1 - p = P[\theta_i = \theta_1^*].$$

Hypothesis testing

$X_i \sim f(x, \theta_i)$ where the θ_i have a particular distribution $\pi(\theta)$. We want to test $H_0 : \theta_k = \theta_0^*$ against $H_1 : \theta_k = \theta_1^*$. Since θ_i are iid let

$$p = P[\theta_i = \theta_0^*], \quad 1 - p = P[\theta_i = \theta_1^*].$$

Estimate p from data $x_1, \dots, x_k$ by maximising

$$L(p; x) = \prod_{i=1}^{k} \{pf(x_i, \theta_0^*) + (1-p)f(x_i, \theta_1^*)\}$$

so $\bar{p}$ solves

$$\sum_{i=1}^{k} \frac{f(x_i, \theta_0^*) - f(x_i, \theta_1^*)}{pf(x_i, \theta_0^*) + (1-p)f(x_i, \theta_1^*)} = 0.$$

Hypothesis testing

$X_i \sim f(x, \theta_i)$ where the θ_i have a particular distribution $\pi(\theta)$. We want to test $H_0 : \theta_k = \theta_0^*$ against $H_1 : \theta_k = \theta_1^*$. Since θ_i are iid let

$$p = P[\theta_i = \theta_0^*], \quad 1 - p = P[\theta_i = \theta_1^*].$$

Estimate p from data $x_1, \ldots, x_k$ by maximising

$$L(p; x) = \prod_{i=1}^{k} \{p f(x_i, \theta_0^*) + (1 - p) f(x_i, \theta_1^*)\}$$

so $\bar{p}$ solves

$$\sum_{i=1}^{k} \frac{f(x_i, \theta_0^*) - f(x_i, \theta_1^*)}{p f(x_i, \theta_0^*) + (1 - p) f(x_i, \theta_1^*)} = 0.$$

Exercise: If $E[X] = \theta$ then the method of moments yields that $\bar{p}$ solves

$$\bar{x} = p\theta_0^* + (1 - p)\theta_1^*.$$

Hypothesis testing

For a, b loss function reject if

$$\frac{L(\theta_1^*, x_k)}{L(\theta_0^*, x_k)} \geq \frac{\hat{p}a}{(1-\hat{p})b}.$$

Example: For $i = 1, 2, \dots$ suppose $X_i \sim N(\theta_i, \sigma^2)$ with σ^2 known. $H_0 : \theta_k = \mu_0$ against $H_1 : \theta_k = \mu_1$ with $\mu_1 > \mu_0$. We reject H_0 if

$$x_k > \frac{1}{2}(\mu_0 + \mu_1) + \frac{\sigma^2}{\mu_1 - \mu_2} / \log\left(\frac{\hat{p}a}{(1-\hat{p})b}\right)$$

provided $\hat{p} \in (0, 1)$. If $\hat{p} = 0$ then always reject (no matter what x_k is). If $\hat{p} = 1$ then always accept.

Hypothesis testing

For a, b loss function reject if

$$\frac{L(\theta_1^*, x_k)}{L(\theta_0^*, x_k)} \geq \frac{\hat{p}a}{(1-\hat{p})b}.$$

Example: For $i = 1, 2, \ldots$ suppose $X_i \sim N(\theta_i, \sigma^2)$ with σ^2 known. $H_0 : \theta_k = \mu_0$ against $H_1 : \theta_k = \mu_1$ with $\mu_1 > \mu_0$. We reject H_0 if

$$x_k > \frac{1}{2}(\mu_0 + \mu_1) + \frac{\sigma^2}{\mu_1 - \mu_2} / \log\left(\frac{\hat{p}a}{(1-\hat{p})b}\right)$$

provided $\hat{p} \in (0, 1)$. If $\hat{p} = 0$ then always reject (no matter what x_k is). If $\hat{p} = 1$ then always accept.

Non-parametric EB

Assume only that the θ_i are iid from some distribution π.

Non-parametric EB

Assume only that the θ_i are iid from some distribution π. Use the data to estimate the prior or the marginal distribution directly.
(pioneered/championed by Robbins 1950âs; actually older than PEB)

Non-parametric EB

Assume only that the θ_i are iid from some distribution π. Use the data to estimate the prior or the marginal distribution directly.
(pioneered/championed by Robbins 1950âs; actually older than PEB)
Model: $y_i|\theta_i \sim f(y_i|\theta_i) = \text{Poi.}(\theta_i)$ and $\theta_i \overset{iid}{\sim} \pi(\cdot)$

Non-parametric EB

Assume only that the θ_i are iid from some distribution π. Use the data to estimate the prior or the marginal distribution directly.
(pioneered/championed by Robbins 1950âs; actually older than PEB)
Model: $y_i|\theta_i \sim f(y_i|\theta_i) = \text{Poi.}(\theta_i)$ and $\theta_i \overset{iid}{\sim} \pi(\cdot)$
Square error loss $\Rightarrow$ Bayes estimator = post. mean

$$\hat{\theta}_i(y) \quad =$$

Non-parametric EB

Assume only that the θ_i are iid from some distribution π. Use the data to estimate the prior or the marginal distribution directly.
(pioneered/championed by Robbins 1950âs; actually older than PEB)
Model: $y_i|\theta_i \sim f(y_i|\theta_i) = \text{Poi.}(\theta_i)$ and $\theta_i \overset{iid}{\sim} \pi(\cdot)$
Square error loss $\Rightarrow$ Bayes estimator = post. mean

$$\hat{\theta}_i(y) \quad = \quad E[\theta_i|\underline{y}]$$

Non-parametric EB

Assume only that the θ_i are iid from some distribution π. Use the data to estimate the prior or the marginal distribution directly.

(pioneered/championed by Robbins 1950âs; actually older than PEB)

Model: $y_i|\theta_i \sim f(y_i|\theta_i) = \text{Poi.}(\theta_i)$ and $\theta_i \overset{iid}{\sim} \pi(\cdot)$

Square error loss $\Rightarrow$ Bayes estimator = post. mean

$$\begin{aligned}
\hat{\theta}_i(y) &= E[\theta_i|\underline{y}] = E[\theta_i|y_i] \\
&= \frac{\int u^{y_i+1} e^{-u} \pi(du)}{\int u^{y_i} e^{-u} \pi(du)} \\
&= \frac{(y_i+1)p(y_i+1)}{p(y_i)}
\end{aligned}$$

Non-parametric EB

Assume only that the θ_i are iid from some distribution π. Use the data to estimate the prior or the marginal distribution directly.
(pioneered/championed by Robbins 1950âs; actually older than PEB)

Model: $y_i|\theta_i \sim f(y_i|\theta_i) = \text{Poi.}(\theta_i)$ and $\theta_i \overset{iid}{\sim} \pi(\cdot)$
Square error loss $\Rightarrow$ Bayes estimator = post. mean

$$\begin{aligned} \hat{\theta}_i(y) &= E[\theta_i|\underline{y}] = E[\theta_i|y_i] \\ &= \frac{\int u^{y_i+1} e^{-u} \pi(du)}{\int u^{y_i} e^{-u} \pi(du)} \\ &= \frac{(y_i+1)p(y_i+1)}{p(y_i)} \end{aligned}$$

The Robins miracle: $\hat{\theta}_i$ is directly estimable as

$$\hat{\theta}_i = \frac{(y_i+1)\hat{p}(y_i+1)}{\hat{p}(y_i)} = \frac{(y_i+1)[\#y's = (y_i+1)]}{[\#y's = y_i])}$$

Non-parametric EB

Intermediate approach: let θ have a pmf

$$\pi(\theta = \phi_j) = p_j$$

for $j = 1, \ldots, m$ with $\sum p_j = 1$.

Non-parametric EB

Intermediate approach: let θ have a pmf

$$\pi(\theta = \phi_j) = p_j$$

for $j = 1, \ldots, m$ with $\sum p_j = 1$. Either the p_j or the ϕ_j are fixed.

Non-parametric EB

Intermediate approach: let θ have a pmf

$$\pi(\theta = \phi_j) = p_j$$

for $j = 1, \dots, m$ with $\sum p_j = 1$. Either the p_j or the ϕ_j are fixed. As $m \to \infty$ the cdf converges to the thrue cdf of θ.

Non-parametric EB

Intermediate approach: let θ have a pmf

$$\pi(\theta = \phi_j) = p_j$$

for $j = 1, \ldots, m$ with $\sum p_j = 1$. Either the p_j or the ϕ_j are fixed.
As $m \to \infty$ the cdf converges to the thrue cdf of θ.
When the p_j are fixed, can use the EM algorithm to determine $\phi_1 \leq \phi_2 \leq \ldots \leq \phi_m$.

$$L(\phi, x) = \prod_{i=1}^{k} \left\{ \sum_{j=1}^{m} f(x_i; \phi_j) p_j \right\}$$

Summary

1. Parametric EB: suppose θ_i iid $\pi(\theta|\psi)$ and evaluate ψ by $\hat{\psi}$ fom data.

Summary

1. Parametric EB: suppose θ_i iid $\pi(\theta|\psi)$ and evaluate ψ by $\hat{\psi}$ fom data.
2. Non-parametric EB: suppose θ_i iid $\pi(\cdot)$ and estimate $\hat{\pi}$ fom data.