Lecture 14. Empirical Bayes, hierarchical Bayes and random effects

Empirical Bayes and Hierarchical Bayes.

Review: Let $X \sim N(\theta, \Sigma)$. Consider a normal prior $\theta \sim N(\mu, \Gamma)$. Then the posterior distribution of $\theta$ is

$$
\theta | X \sim N(\mu + \Gamma (\Sigma + \Gamma)^{-1} (X - \mu), \Gamma (\Sigma + \Gamma)^{-1} \Sigma)
$$

Empirical Bayes

Let $\Sigma = I$, $\mu = 0$ and $\Gamma = \gamma I$. Then

$$
\delta_\gamma (x) = \frac{\gamma^2}{1 + \gamma^2} x = \left(1 - \frac{1}{1 + \gamma^2}\right)x.
$$

Since

$$
E \left( \frac{n - 2}{\|x\|^2} \right) = \frac{1}{1 + \gamma^2},
$$

we may estimate $\frac{1}{1 + \gamma^2}$ by $\frac{n - 2}{\|x\|^2}$ to yield the James-Stein estimator.

Hierarchical Bayes

Let $\Sigma = I$, $\Gamma = \gamma I$ and $\gamma \sim H$.

Recall that

$$
g(\theta) = 1/ \|\theta\|^{n - 2} \propto \int_0^\infty \gamma \varphi_{\gamma^2 I}(\theta) \, d\gamma
$$

then the harmonic prior corresponds $h(\gamma) \sim \gamma$.

Theorem. Let

$$
h(\gamma) \propto \frac{\gamma}{(1 + \gamma^2)^{2-a}}
$$

The the generalized Bayes estimators (i) exist if $a < 1 + n/2$; (ii) are admissible if $n \geq 3$ and $3 - n/2 \leq a \leq 2$; (iii) are minimax if they exist and $n \geq 3$ and $3 - n/2 \leq a$.

Proof of the theorem: for part (ii)

$$
g(\theta) \propto \int_0^\infty \frac{1}{(1 + \gamma^2)^{2-a} \gamma^{n-1}} \exp \left( - \frac{\|\theta\|^2}{2\gamma^2} \right) \, d\gamma
$$

then

$$
\|\theta\|^{n+2-2a} g(\theta) \to c \text{ as } \|\theta\| \to \infty.
$$

It can be shown that the growth condition

$$
\int_{\|\theta\| > 2} \frac{g(\theta)}{\|\theta\|^2 (\log \|\theta\|)^2} \, d\theta < \infty
$$
the flatness condition hold when \(3 - n/2 \leq a < 2\) (Homework problem). for part (iii)

\[
\delta(x) = \left(1 - \frac{r\left(||x||^2\right)2(n-2)}{||x||^2}\right)x
\]

where

\[
0 \leq r\left(||x||^2\right) = \frac{1}{2n-4}\left(n + 2 - 2a - \frac{2e^{-||x||^2/2}}{\int_0^{n/2-a}e^{-u||x||^2}du}\right).
\]

Hierarchical Bayes vs. Empirical Bayes.
These two forms of analysis are closely related. The hierarchical formulation

\[
X \sim N(\theta, I), \ \theta \sim N(\theta, \gamma^2 I)
\]

is common to both of them.

The empirical Bayes method uses the data to produce some heuristic estimator of \(\gamma\). Hierarchical Bayes methods treat the hierarchical parameter, \(\gamma\), in a Bayesian fashion.

There is an additional heuristic connection between the two methodologies. Note that the hierarchical Bayes estimator can be written as

\[
E(\theta|x) = E\left(E(\theta|x, \gamma^2) \mid x\right).
\]

The inner expectation on the right hand side of the equation can be considered to be an estimator \(\delta_{\gamma^2}\) such as the one that appears in the empirical Bayes derivation. Hence the hierarchical Bayes estimator, \(\delta_{\text{hier}}\), say is the mean of these \(\delta_{\gamma^2}\) with respect to the Bayesian conditional distribution of \(\gamma^2\) given \(x\). Write

\[
\delta_{\text{hier}} = \delta_{\gamma^2}.
\]

In this way the hierarchical Bayes estimator can also be viewed as an empirical Bayes estimator.
Lecture 15. Empirical Bayes, hierarchical Bayes and random effects


Selected writings of Robbins:

Example.
Observe \( X_i \sim \text{Poisson} (\theta_i) \)

\[
P (X_i = x_i | \theta_i) = \frac{\exp(-\theta_i) \theta_i^{x_i}}{x_i!}.
\]

We want to estimate the unknown parameter \( \theta_i \). Assume that \( \theta_1, \theta_2, \ldots, \theta_n \) are i.i.d. with distribution \( G \). The generalized Bayes w.r.t. squared error loss is

\[
\delta_i (x_i) = \frac{\int \theta \exp(-\theta) \theta_i^{x_i} G (d\theta)}{\int \exp(-\theta) \theta_i^{x_i} G (d\theta)}
\]

\[
= (x_i + 1) \frac{\int \exp(-\theta) \theta_i^{x_i+1} G (d\theta)}{\int \exp(-\theta) \theta_i^{x_i} G (d\theta)}
\]

\[
= (x_i + 1) \frac{G^* (x_i + 1)}{G^* (x_i)}
\]

where \( G^* \) is the marginal distribution of \( X_i \). We know for every fixed \( x_i \),

number of terms \( X_1, X_2, \ldots, X_n \) which are equal to \( x + 1 \)

number of terms \( X_1, X_2, \ldots, X_n \) which are equal to \( x \)

\[
\rightarrow \frac{G^* (x_i + 1)}{G^* (x_i)}
\]

Example (read Efron, 2003, Ann. Stat.).

Extension to exponential family.
Let \( X_i \sim f (x|\theta_i) = \exp (x\theta_i - \psi (\theta_i)) h (x) \) and \( \theta_i \sim G \), then

\[
\delta (x) = \frac{\int \theta \exp (x\theta - \psi (\theta)) h (x) G (d\theta)}{\int \exp (x\theta - \psi (\theta)) h (x) G (d\theta)}
\]

\[
= h (x) \frac{d}{dx} \left( \frac{G^* (x)}{h (x)} \right).
\]
Then the question now is how to estimate $G^*(x)$.

**Connection to compound decision theory**


Let $X_i \sim f(x|\theta_i)$ and write

$$R(\theta, \delta) = \frac{1}{n} \sum_{i=1}^{n} EL(\theta_i, \delta_i(X))$$

For separable decision rules of the form $\delta_i(X) = t(X_i)$, the compound risk is equal to the average risk

$$R(\theta, \delta) = \int \int L(\theta, t(X)) f(x|\theta) dx G(d\theta).$$

where $G(A) = n^{-1} \sum_{\theta \in A} f(x|\theta) dx$. Robbin’s proposal is to seek asymptotically minimax procedure satisfying

$$R(\theta, \delta) = R^*(\theta, \delta_G) + o(1) \text{ as } n \to \infty.$$