

**Week 4**  
Spring 2009

**Lecture 7. 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

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

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

**Hierarchical Bayes**

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

It is easy to see that

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

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  $a < 1 + p/2$ ; (ii) are admissible if  $p \geq 3$  and  $3 - p/2 \leq a \leq 2$ ; (iii) are minimax if they exist and  $p \geq 3$  and  $3 - p/2 \leq a$ .

*Proof of the theorem:*

(i) It is easy to see  $\int_1^\infty \gamma^{2(a-2)-(p-1)} d\gamma$  is finite if and only if  $2(a-2) < p-2$ , i.e.,  $a < 1 + p/2$ .

(ii) We have

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

then

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

It can be shown that the growth condition holds

$$\int_{\|\theta\|>2} \frac{g(\theta)}{\|\theta\|^2 (\log \|\theta\|)^2} d\theta < \infty$$

and the flatness condition holds when  $3 - p/2 \leq a \leq 2$  (leave it to you).

(iii) It can be shown that

$$\delta(x) = \left( 1 - \frac{r(\|x\|^2) 2(p-2)}{\|x\|^2} \right) x$$

where

$$0 \leq r(\|x\|^2) = \frac{1}{2p-4} \left( p+2 - 2a - \frac{2e^{-\|x\|^2/2}}{\int_0^1 v^{p/2-a} e^{-v\|x\|^2} dv} \right).$$

From Baranchik (1970) the estimator is minimax if

$$0 \leq p+2 - 2a - \frac{2e^{-\|x\|^2/2}}{\int_0^1 v^{p/2-a} e^{-v\|x\|^2} dv} \leq 2p-4.$$

Let's just look at the case  $a = 2$ . The general case follows similarly. Write

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

then

$$g^*(x) \propto \int_0^\infty \gamma \varphi_{(1+\gamma^2)I}(x) d\gamma \propto \int_0^1 v^{(p-4)/2} \exp\left(-\frac{\|x\|^2 v}{2}\right) dv,$$

thus

$$\begin{aligned} \frac{\nabla g^*}{g^*} &= \frac{\int_0^1 v^{p/2-1} \exp\left(-\frac{\|x\|^2 v}{2}\right) dv}{\int_0^1 v^{p/2-2} \exp\left(-\frac{\|x\|^2 v}{2}\right) dv} x = \frac{-2 \int_0^1 v^{p/2-1} d \exp\left(-\frac{\|x\|^2 v}{2}\right)}{\int_0^1 v^{p/2-2} \exp\left(-\frac{\|x\|^2 v}{2}\right) dv} \frac{x}{\|x\|^2} \\ &= \left( p-2 - \frac{2e^{-\|x\|^2/2}}{\int_0^1 v^{p/2-2} e^{-v\|x\|^2} dv} \right) \frac{x}{\|x\|}. \end{aligned}$$

**Remark 1** When  $p \geq 5$ , there exist proper Bayes minimax estimators, since  $h$  is the density of a finite measure for  $a < 1$ . For  $p = 3, 4$  there are no proper Bayes minimax estimators from Brown (1971).

### Hierarchical Bayes vs. Empirical Bayes.

These two forms of analysis are closely related. The hierarchical formulation

$$X \sim N(\theta, I), \quad \theta \sim N(\theta, \gamma^2 I)$$

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

$$E(\theta|x) = E(E(\theta|x, \gamma^2) | x).$$

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,  $\delta^{hier}$ , say is the mean of these  $\delta_{\gamma^2}$  with respect to the Bayesian conditional distribution of  $\gamma^2$  given  $x$ . Write

$$\delta^{hier} = \delta_{\hat{\gamma}^2}.$$

In this way the hierarchical Bayes estimator can also be viewed as an empirical Bayes estimator.

**Lecture 8. Empirical Bayes, hierarchical Bayes and random effects (Cont.)**

Robbins (1956): An empirical Bayes approach to statistics, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Jerzy Neyman, ed., vol. 1, Berkeley, California: University of California Press, 1956, pp. 157–163.

Selected writings of Robbins:

1. *What is Mathematics?: An elementary approach to ideas and methods*, with Richard Courant, London: Oxford University Press, 1941.

2. A stochastic approximation method, with Sutton Monro, *Annals of Mathematical Statistics* **22**, 3, 1951, pp. 400–407.

3. Robbins (1956) above.

4. Asymptotically subminimax solutions of compound decision problems, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1951, pp. 131-148.

**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  $\theta_1, \theta_2, \dots, \theta_n$  are i.i.d. with distribution  $G$ . The generalized Bayes w.r.t. squared error loss is

$$\begin{aligned} \delta_i(x_i) &= \frac{\int \theta \frac{\exp(-\theta)\theta^{x_i}}{x_i!} G(d\theta)}{\int \frac{\exp(-\theta)\theta^{x_i}}{x_i!} G(d\theta)} \\ &= (x_i + 1) \frac{\int \frac{\exp(-\theta)\theta^{x_i+1}}{(x_i+1)!} G(d\theta)}{\int \frac{\exp(-\theta)\theta^{x_i}}{x_i!} G(d\theta)} \\ &= (x_i + 1) \frac{G^*(x_i + 1)}{G^*(x_i)} \end{aligned}$$

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

$$\frac{\text{number of terms } X_1, X_2, \dots, X_n \text{ which are equal to } x_i + 1}{\text{number of terms } X_1, X_2, \dots, X_n \text{ which are equal to } x_i} \rightarrow \frac{G^*(x_i + 1)}{G^*(x_i)}$$

**Example** (read Efron, 2003, Ann. Stat.). Applications to missing species problem. Estimate Shakespeare’s vocabulary.

**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

$$\begin{aligned} \delta(x) &= \frac{\int \theta \exp(x\theta - \psi(\theta)) h(x) G(d\theta)}{\int \exp(x\theta - \psi(\theta)) h(x) G(d\theta)} \\ &= \frac{\frac{d}{dx} (G^*(x) / h(x))}{G^*(x) / h(x)}. \end{aligned}$$

Then the question now is how to estimate  $G^*(x)$ .

**Connection to compound decision theory**

Robbins (1951): Asymptotically subminimax solutions of compound decision problems, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1951, pp. 131-148.

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  $G(A) = n^{-1} \sum_{i=1}^n \{ \theta \in A \}$ . Robbin's proposal is to seek asymptotically minimax procedure satisfying

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