Lecture 4. Bayes estimation, minimaxity and Admissibility.

Bayes estimator
Observe a normally distributed $n$-dimensional random variable $X,$

$$X \sim N(\theta, \Sigma)$$

where $\theta$ and $\Sigma$ are parameters. We assume that $\theta$ has a proper prior distribution $G$. A Bayes estimator, denoted by $\delta_G$, solves the following minimization problem:

$$\int R(\theta, \delta_G) G(d\theta) = \inf_{\delta} \{ r(G, \delta) \}$$

where

$$r(G, \delta) = \int R(\theta, \delta) G(d\theta)$$

When $G$ has a density w.r.t. Lebesgue measure, the conditional density of $\theta$ given $X = x$ is

$$f(\theta|x) = \frac{f_{\theta}(x) g(\theta)}{\int f_{\theta}(x) g(\theta) d\theta}.$$ 

When the loss is squared error, $L(\theta, \delta) = (\theta - \delta)^T M (\theta - \delta)$ with $M$ positive definite, then the posterior mean is the Bayes estimator,

$$\delta_G = \frac{\int \theta f_{\theta}(x) g(\theta) d\theta}{g^*(x)}$$

with $g^*(x) = \int f_{\theta}(x) g(\theta) d\theta$, since

$$r(G, \delta) = E[E[L(\theta, X)|X]].$$

If $G$ is a general (non-negative) measure, it is typical not true that

$$\inf_{\delta} \left\{ \int R(\theta, \delta) G(d\theta) \right\} < \infty.$$ 

We call $\delta_G$ a Bayes estimator if

$$\inf_{\delta} \left\{ \int [R(\theta, \delta) - R(\theta, \delta_G)] G(d\theta) \right\} \geq 0,$$

and call the posterior mean "formal" or "generalized" Bayes estimator.

Remark: An estimator $\delta_G(x)$ is called a generalized Bayes estimator with respect to $G$, if the posterior expected loss $E[L(\theta, \delta)|X = x]$ is minimized at $\delta = \delta_G$ for all $x$. An estimator $\delta$ is called extended Bayes there exists a sequence of proper priors $G_i$ and Bayes estimators $\delta_{G_i}$ such that $\lim_{i} r(G_i, \delta) = 1$.
An estimator \( \delta_G(x) \) is called a (pointwise) \( \text{limit of Bayes estimators} \) if there exists a sequence of proper priors \( G_i \) and Bayes estimators \( G_i \), such that \( \delta_{G_i}(x) \to \delta_G(x) \) a.s.

**Example:** Let \( X \sim N(\theta, 1) \) and \( g(\theta) = e^{\alpha \theta} \). The posterior mean is \( X + a \), but it is not a Bayes estimator, since

\[
\int [R(\theta, X) - R(\theta, X + a)] G(d\theta) < 0.
\]

Is it an extended Bayes?

**Conjecture (from John Hartigan):** \{admissible estimator\} = \{Bayes estimator\}?

**Alternate form for Bayes estimators (for normal location problem)**

Define

\[
\nabla \psi = (\partial^2 \psi / \partial x_i \partial x_j)_{n \times n}.
\]

**Theorem.** Let \( X \sim N(\theta, \Sigma) \) with \( \Sigma \) known. Let \( G \) be any prior such that \( g^*(x) < \infty \) for all \( x \). Then

\[
E(\theta | X = x) = x + \Sigma \nabla (\log (g^*(x)))
\]

\[
\text{Cov}(\theta | x) = \Sigma + \Sigma \nabla^2 (\log (g^*(x))) .
\]

**Proof of the theorem:** When \( g^*(x) < \infty \) for all \( x \), then \( g^*(x) \) is analytic in each coordinate variable \( x_i \), and partial derivatives of all orders can be computed under the integral sign (why?). Then

\[
\nabla g^*(x) = \int \nabla \varphi_\Sigma (x - \theta) G(d\theta) = \int \Sigma^{-1} (\theta - x) \varphi_\Sigma (x - \theta) G(d\theta),
\]

which implies

\[
E(\theta | X = x) = x + \frac{\int (\theta - x) \varphi_\Sigma (x - \theta) G(d\theta)}{g^*(x)} = x + \Sigma \frac{\nabla g^*(x)}{g^*(x)}
\]

It can be shown

\[
\text{Cov}(\theta | x) = E \left[ (\theta - E(\theta | x)) (\theta - E(\theta | x))^T | x \right]
\]

\[
= E[\langle \theta - x \rangle (\theta - x)^T | x] - (E(\theta | x) - x) (E(\theta | x) - x)^T
\]

\[
= E[\langle \theta - x \rangle (\theta - x)^T | x] - \Sigma \nabla \log (g^*(x)) [\nabla \log (g^*(x))]^T \Sigma
\]

where

\[
E[\langle \theta - x \rangle (\theta - x)^T | x] = \Sigma (\nabla_2 (g^*(x))) \Sigma + g^*(x) \Sigma .
\]

**Example.** Consider a normal prior \( \theta \sim N(\mu, \Gamma) \). Then the posterior distribution of \( \theta \) is

\[
\theta | X \sim N(\mu + \Gamma (\Sigma + \Gamma)^{-1} (X - \mu), \Gamma (\Sigma + \Gamma)^{-1} \Sigma).
\]
Homework problem: Show that $g^*(x)$ is analytic in each coordinate variable $x_i$ when $g^*(x) < \infty$ for all $x$. Can the positive part James-Stein estimator for the canonical normal means estimation be generalized Bayes for squared error loss?

Conjecture (from Larry Brown): If $\delta$ is a generalized Bayes, then $\delta$ is admissible iff $\delta$ is Stein admissible (under very mild regularity condition).

Minimaxity of $\delta_0 = X$

Lemma. For a given procedure $\delta'$ suppose there is a sequence of prior distributions $\{G_i\}$ such that

$$\lim_{i \to \infty \int R(\theta, \delta_{G_i}) G_i(d\theta) = \sup_{\theta} R(\theta, \delta').$$

Then $\delta'$ is minimax.

The squared error loss: $L(\theta, \delta) = (\theta - \delta)^T M (\theta - \delta)$

Theorem. For the normal location problem, $\delta_0 = X$ is a minimax estimator of $\theta$ under the squared error loss.

Proof of the theorem: Let $G_i = N(0, i^2 I)$. Then

$$\lim_{i \to \infty \int R(\theta, \delta_{G_i}) G_i(d\theta) = \sup_{\theta} R(\theta, \delta') = Tr(\Sigma M).$$

Admissibility of $\delta_0 = X$ for $n \leq 2$. We will show that next time.
Lecture 5. Bayes estimation, minimaxity and Admissibility (cont.).

Admissibility

Conditions on priors and admissibility: conditions on the prior measure which guarantees that the corresponding generalized Bayes procedure is admissible.

Let $S_2 = \{ x \in \mathbb{R}^n : ||x|| \leq 2 \}$. Define

$$J_x (h) = \int h(\theta) \varphi (x - \theta) \, d\theta$$

Assumption

Growth Condition:

$$\int_{S_2^c} \frac{g(\theta)}{||\theta||^2 \log^2 (||\theta||)} < \infty$$

Asymptotic flatness condition:

$$\int_{S_2^c} J_x \left\{ \frac{\nabla g}{g} - \frac{J_x(\nabla g)}{J_x(g)} \right\} \, dx < \infty$$

It can be shown that $\int_{S_2^c} ||\nabla g||^2 / g \, dx < \infty$ implies the flatness condition.

**Theorem.** Let $G$ be a prior satisfying two conditions above. Then $\delta_G$ is admissible.

For the normal mean estimation problem with squared error loss,

**Blyth’s Method.** Let $\delta$ be an estimator. Let $\{ G_j \}$ be a sequence of finite prior measures such that: (i) $r(G_j, \delta) - r(G_j, \delta_G) \to 0$ as $n \to \infty$; (ii) $\inf_j \{ G_j (S) \} > 0$. Then $\delta$ is an admissible estimator.

**hint:** If $R(\theta, \delta') \leq R(\theta, \delta)$ for all $\theta$ and with strict inequality for some $\theta$, then $R(\theta, \delta'') < R(\theta, \delta)$ for all $\theta$. For all $j$,

$$r(G_j, \delta_G) \leq r(G_j, \delta'') \leq \int_S R(\theta, \delta'') G_j (d\theta) + \int_{S^c} R(\theta, \delta) G_j (d\theta) = r(G_j, \delta) + \int_S \left[ R(\theta, \delta'') - R(\theta, \delta) \right] G_j (d\theta),$$

contradiction!

**Proposition.**

$$r(G, \delta) - r(G, \delta_G) = \int ||\delta_G - \delta||^2 g^* (x) \, dx$$

**Proof:**

$$r(G, \delta) - r(G, \delta_G) = E_x E \left( ||\theta - \delta_G + \delta \|^2 - ||\theta - \delta_G\|^2 | X \right)$$

$$= E_x E \left( ||\delta_G - \delta||^2 | X \right) = \int ||\delta_G - \delta||^2 g^* (x) \, dx$$
Proof of the theorem: Define \( g_j = h_j^2 g \) where

\[
\begin{align*}
  h_j &= \begin{cases}
    1 - \frac{1}{\log(\|\theta\|)} & \|\theta\| \leq 1 \\
    \frac{\log(\|\theta\|)}{\log(j)} & 1 \leq \|\theta\| \leq j \\
    \|\theta\| & \|\theta\| > j
  \end{cases}
\end{align*}
\]

It is easy to see

\[
\delta_{G_j} = \frac{\int \theta h_j^2 g(\theta) \varphi(x - \theta) d\theta}{\int h_j^2 g(\theta) \varphi(x - \theta) d\theta} \to \delta_G
\]

and

\[
g_j^*(x) = \int h_j^2 g(\theta) \varphi(x - \theta) d\theta \leq g^*(x).
\]

Write

\[
\begin{align*}
  \delta_G(x) &= x + \frac{\nabla g^*}{g^*} = x + \frac{J_x(\nabla g)}{J_x(g)}, \\
  \delta_{G_j}(x) &= x + \frac{\nabla g^*}{g^*} = x + \frac{J_x(h_j^2 \nabla g + g \nabla h_j^2)}{J_x(h_j^2 g)}.
\end{align*}
\]

Hence

\[
r(G_j, \delta_G) - r(G_j, \delta_{G_j})
= \int_{S_2} \|\delta_G - \delta_{G_j}\|^2 g_j^*(x) \, dx + \int_{S_2} \|\delta_G - \delta_{G_j}\|^2 g_j^*(x) \, dx
\leq 2 \int_{S_2} \left\| \frac{J_x(g \nabla h_j^2)}{J_x(g_j)} \right\|^2 g_j^*(x) \, dx + 2 \int_{S_2} \left\| \frac{J_x(\nabla g)}{J_x(g)} - \frac{J_x(h_j^2 \nabla g)}{J_x(h_j^2 g)} \right\|^2 g_j^*(x) \, dx + o(1)
= 2A_j + 2B_j + o(1)
\]

Show \( A_j \to 0 \):

\[
A_j = \int_{S_2} \left\| \frac{J_x(g \nabla h_j^2)}{J_x(g_j)} \right\|^2 g_j^*(x) \, dx
= 4 \int_{S_2} \left\| \frac{J_x(g h_j \nabla h_j)}{J_x(g_j)} \right\|^2 g_j^*(x) \, dx
\leq 4 \int_{S_2} J_x \left( g \|\nabla h_j\|^2 \right) \, dx \quad \text{(Cauchy-Schwarz inequality)}
= 4 \int_{S_2} \|\nabla h_j(\theta)\|^2 g(\theta) \, d\theta
\]

and

\[
\|\nabla h_j(\theta)\|^2 \leq \frac{1}{\|\theta\|^2 \log^2(j)} I_{[1,j]}(\|\theta\|) \leq \frac{1}{\|\theta\|^2 \log^2(\|\theta\| \vee 2)} I_{[1,j]}(\|\theta\|).
\]
Show $B_j \to 0$:

$$\left\| \frac{J_x (\nabla g)}{J_x (g)} - \frac{J_x (h^2 \nabla g)}{J_x (g_j)} \right\|^2 g^*_j (x) = \left\| \frac{J_x \left( g_j \frac{J_x (\nabla g)}{J_x (g)} - h^2 \nabla g \right)}{J_x (g_j)} \right\|^2$$

$$\leq \frac{J_x \left( g_j \left[ \left\| \frac{J_x (\nabla g)}{J_x (g)} - \frac{\nabla g}{g} \right\|^2 \right) \right]}{J_x (g_j)}$$

Admissibility of $\delta_0 = X$ for $n = 1, 2$

Let $g (\theta) = 1$, then

$$\int_{S_2} \frac{g (\theta)}{\| \theta \|^2 \log^2 (\| \theta \|)} = 2 \pi \int_2^{\infty} \frac{1}{r \log^2 r} dr < \infty.$$

**Homework problem** (you pick one part to work on). Let $X_i \sim \text{Poisson} (\lambda_i)$ be independent, $i = 1, 2, \ldots, n$. Denote $X = (X_1, \ldots, X_n)$ and $\lambda = (\lambda_1, \ldots, \lambda_n)$. Under the loss $L (\lambda, \delta) = \sum_{i=1}^{n} (\delta_i - \lambda_i)^2 / \lambda_i$, show that (i) for $n = 1$, $X$ is an admissible estimator of $\lambda$ using Blyth’s method; (2) for $n \geq 2$, $X$ is not an admissible estimator of $\lambda$.

Reference: (i) Clevenson and Zidek (1975), *Simultaneous Estimation of the Means of Independent Poisson Laws*, JASA.