Week 2
Spring 2009

Lecture 3. The Canonical normal means estimation problem (cont.).

Shrink toward a common mean.

Theorem. Let $X \sim N\left(\theta, \sigma^2 I_n\right)$. Let $0 < C \le 2(n-3)$ (hence $n \ge 4$). Define

$$\delta(X) = \overline{X} + \left(1 - \frac{C\sigma^2}{\left\|X - \overline{X}\right\|^2}\right)\left(X - \overline{X}\right),\ \delta_+(X) = \overline{X} + \left(1 - \frac{C\sigma^2}{\left\|X - \overline{X}\right\|^2}\right)_+\left(X - \overline{X}\right).$$

Then

$$R(\theta, \delta_+) < R(\theta, \delta) \le n\sigma^2.$$

Homework problem: prove the theorem above (cf. Lindley and Smith (1972, JRSSB)).

An example from Efron and Morris.

Efron and Morris (1975,1977) looked at the batting averages of a sample of 18 baseball players for the 1970 season (Batting averages are the proportion of "base hits" for a player out of his total "at bats").

Explanation for the table:

$$\begin{aligned}
Y_i &\sim \frac{1}{45} Bin(45, \gamma_i) : \text{ batting average for the first 45 at bats} \\
P_i &\sim \frac{1}{n_i} Bin(n_i, \gamma_i) : \text{ batting average for the rest of the season} \\
n_i &: \text{ number of at bats for the rest of the season} \\
X_i &: X_i = \arcsin(2Y_i - 1) \approx N\left(\mu_i, \frac{1}{45}\right),\ \mu_i = \arcsin(2\gamma_i - 1) \\
R_i &: R_i = \arcsin(2P_i - 1)
\end{aligned}$$

Variance stabilizing transformations

Let X_i, $i = 1, 2, \ldots, K$, be a sequence of i.i.d.r.v.'s with distribution in the exponential family $\mathcal{P}$ with parameter set Θ,

$$P_\theta(dx) = \exp\{\theta U(x) - V(\theta)\}\, \mu(dx)$$

where U is a measurable map, and $V(\theta)$ is the cumulant generating function associated with the exponential family. Let $Y_i = U(X_i)$, then $\sum_{i=1}^K Y_i$ is a sufficient statistics for the i.i.d. model. Set for brevity, $S_K = \sum_{i=1}^K Y_i$ and $\mu(\theta) = V'(\theta) = EY_1$, $I(\theta) = V'(\theta) = Var_\theta(Y_1)$.

According to the central limit theorem, the sequence $\sqrt{K}\left(S_K/K - \mu(\theta)\right)$ converges weakly to the normal r.v. with zero mean and variance $I(\theta)$. Define

1. 1970 Batting Averages for 18 Major League Players and Transformed Values X_i, θ_i

i	Player	Y_i = batting average for first 45 at bats	p_i = batting average for remainder of season	At bats for remainder of season	X_i	θ_i
		(1)	(2)	(3)	(4)	(5)
1	Clemente (Pitts, NL)	.400	.346	367	−1.35	−2.10
2	F. Robinson (Balt, AL)	.378	.298	426	−1.66	−2.79
3	F. Howard (Wash, AL)	.356	.276	521	−1.97	−3.11
4	Johnstone (Cal, AL)	.333	.222	275	−2.28	−3.96
5	Berry (Chi, AL)	.311	.273	418	−2.60	−3.17
6	Spencer (Cal, AL)	.311	.270	466	−2.60	−3.20
7	Kessinger (Chi, NL)	.289	.263	586	−2.92	−3.32
8	L. Alvarado (Bos, AL)	.267	.210	138	−3.26	−4.15
9	Santo (Chi, NL)	.244	.269	510	−3.60	−3.23
10	Swoboda (NY, NL)	.244	.230	200	−3.60	−3.83
11	Unser (Wash, AL)	.222	.264	277	−3.95	−3.30
12	Williams (Chi, AL)	.222	.256	270	−3.95	−3.43
13	Scott (Bos, AL)	.222	.303	435	−3.95	−2.71
14	Petrocelli (Bos, AL)	.222	.264	538	−3.95	−3.30
15	E. Rodriguez (KC, AL)	.222	.226	186	−3.95	−3.89
16	Campaneris (Oak, AL)	.200	.285	558	−4.32	−2.98
17	Munson (NY, AL)	.178	.316	408	−4.70	−2.53
18	Alvis (Mil, NL)	.156	.200	70	−5.10	−4.32

a function $F : R \to R$: $F^{'}(\lambda) = I\left(\mu^{-1}(\lambda)\right)^{-1/2}$ such that $F^{'}(\mu(\theta)) = I(\theta)^{-1/2}$. The so called delta method gives

$$\sqrt{K}\left\{F(S_K/K) - F(\mu(\theta))\right\} \xrightarrow{d} N(0,1).$$

We then call F a variance stabilization transformation. For two finite constants a and c, it is also true that

$$\sqrt{K}\left\{F\left(\frac{S_K+a}{K+c}\right) - F(\mu(\theta))\right\} \xrightarrow{d} N(0,1).$$

This suggests we have freedom to choose a and c in practice.

Lecture 4. Bayes estimation, minimaxity and Admissibility.

Bayes estimator

Proper prior

Observe a normally distributed n-dimensional random variable X,

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

where θ and Σ are parameters. We assume that θ has a proper prior distribution G. A *Bayes estimator,* denoted by δ_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 θ 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 for all M, i.e.,

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

where $g^*(x) = \int f_\theta(x) g(\theta) d\theta$, since

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

Improper prior

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 δ_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 δ is called *extended Bayes* there exists a sequence of proper priors G_i and Bayes estimators δ_{G_i} such that $\lim_i r(G_i, \delta) =$

$\lim_i r(G_i, \delta_{G_i})$. An estimator $\delta_G(x)$ is called a (pointwise) *limit of Bayes* estimators if there exists a sequence of proper priors G_i and Bayes estimators δ_{G_i} such that $\delta_{G_i}(x) \to \delta_G(x)$ a.s..

Example: Let $X \sim N(\theta, 1)$ and $g(\theta) = e^{a\theta}$ with $a \neq 0$. The posterior mean is $X + a$, but it is not 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

$$\nabla_2 h = (\partial^2 h / \partial x_i \partial x_i)_{n \times n}.$$

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

$$\begin{aligned} E(\theta | X = x) &= x + \Sigma \nabla (\log(g^*(x))) \\ Cov(\theta | x) &= \Sigma + \Sigma \nabla_2 (\log(g^*(x))). \end{aligned}$$

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

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

Then we have

$$\begin{aligned} Cov(\theta | x) &= E\left[(\theta - E(\theta|x))(\theta - E(\theta|x))^T |x\right] \\ &= E[(\theta - x)(\theta - x)^T |x] - (E(\theta|x) - x)(E(\theta|x) - x)^T \\ &= E[(\theta - x)(\theta - x)^T |x] - \Sigma \nabla \log(g^*(x)) [\nabla \log(g^*(x))]^T \Sigma \end{aligned}$$

It can be shown that

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

(check it for the case Σ diagonal by differentiating φ twice). Then the formula for $Cov(\theta|x)$ follows easily.

Example. Consider a normal prior $\theta \sim N(\mu, \Gamma)$. Then the posterior distribution of θ 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 δ is a generalized Bayes, then δ is admissible iff δ is Stein admissible (under very mild regularity condition).

Question: Is the positive part James-Stein estimator Stein admissible?

Minimaxity of $\boldsymbol{\delta_0}$= X

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

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

Then δ' 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 θ under the squared error loss.

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

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

Admissibility of $\boldsymbol{\delta_0}$= X for n $\leq$ 2.We will show that next time.