Week 5
Spring 2009

Lecture 9. Empirical Bayes, hierarchical Bayes and random effects (cont.)

Random effects model
One-Way ANOVA.

Suppose in a given year there are J_i daughters of each p bulls in a sire-proving experiment with

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},\ i = 1, \ldots p;\ j = 1, \ldots, J_i; \varepsilon_{ij} \sim N\left(0, \sigma_e^2\right) \text{ independent}$$

being the record of daughter j of sire (bull) i. Here the parameters are $\left\{\mu,\ \alpha_i,\ \sigma^2 : i = 1, \ldots, p\right\}$ with $\sum_{i=1}^{p} \alpha_i = 0$. The total number of scalar observations is $n = \Sigma_{i=1}^{p} J_i$. When $J_i = J$, a constant for all i, the model is called "balanced".

We may write the observations in a non-overparametrized fashion as

$$Y_{ij} = \theta_i + \varepsilon_{ij},\ i = 1, \ldots p;\ j = 1, \ldots, J_i; \varepsilon_{ij} \sim N\left(0, \sigma_e^2\right).$$

The sufficient statistics are

$$\begin{aligned} X_i &= \overline{Y}_{i\cdot} = \frac{1}{J_i} \sum_{j=1}^{J_i} Y_{ij} \sim N\left(\theta_i, \frac{\sigma_e^2}{J_i}\right), \\ V &= SSE = \sum_i \sum_{j=1}^{J_i} \left(Y_{ij} - X_i\right)^2 \sim \sigma_e^2 \chi_{n-p}^2. \end{aligned}$$

Random Effects Model.

Eisenhart (1947, Biometrics) and Henderson (1953, Biometrics). The Gaussian random effects model begins from the structure of one-way ANOVA but with the $\{\alpha_i\}$ being modeled as random effects than fixed parameters

$$\begin{aligned} Y_{ij} &= \mu + A_i + \varepsilon_{ij},\ i = 1, \ldots p;\ j = 1, \ldots, J_i; \\ A_i &\sim N\left(0, \sigma_A^2\right),\ i = 1, \ldots p;\ \text{independent} \\ \varepsilon_{ij} &\sim N\left(0, \sigma_e^2\right) \text{ independent, conditional on } \{A_i\} \end{aligned}$$

i.e.,

$$\begin{aligned} A_i &\sim N\left(0, \sigma_A^2\right),\ i = 1, \ldots p;\ \text{independent} \\ Y_{ij}|A_i &\sim N\left(\mu + A_i, \sigma_e^2\right),\ i = 1, \ldots p;\ j = 1, \ldots, J_i; \varepsilon_{ij} \sim N\left(0, \sigma_e^2\right) \text{ independent} \end{aligned}$$

The one-way random effects model is equivalent to an empirical Bayes model,

$$\begin{aligned} X_i|\theta_i &\sim N\left(\theta_i, \frac{\sigma_e^2}{J_i}\right), \theta_i \sim N\left(\mu, \sigma_A^2\right) \\ SSE &= \sum_i \sum_{j=1}^{J_i} \left(Y_{ij} - X_i\right)^2 \sim \sigma_e^2 \chi_{n-p}^2. \end{aligned}$$

with

$$E\left(\theta_i|X_i\right) = \mu + \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma_e^2}{J_i}}\left(X_i - \mu\right) = \frac{\frac{\sigma_e^2}{J_i}}{\sigma_A^2 + \frac{\sigma_e^2}{J_i}}\mu + \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma_e^2}{J_i}}X_i$$

Because θ has traditionally been viewed in the random effects literature as a latent variable, rather than as a parameter, inferential statements about its true value are referred to as *predictions*, rather than *estimates*.

BLUP (Best Linear Unbiased predictor). In the special case μ, σ_A^2, σ_e^2 are known then the natural estimator/predictor of the θ vector has coordinates

$$\left(\delta_{BLUP}\right)_i = \mu + \left(1 - \frac{\sigma_e^2/J_i}{\sigma_A^2 + \sigma_e^2/J_i}\right)\left(X_i - \mu\right).$$

If the parameters are unknown it is then traditional to substitute suitable estimates for these parameters.

Approaches to Estimate Parameters
The ANOVA Approach
Let

$$SSM = \sum_i \sum_j \left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2 - SSE = \sum_i J_i \left(X_i - \overline{Y}_{\cdot\cdot}\right)^2.$$

Consider the *balanced* case,

$$E\left(SSM\right) = J\left(p-1\right)\left(\sigma_A^2 + \frac{\sigma_e^2}{J}\right) = \left(n - J\right)\left(\sigma_A^2 + \frac{\sigma_e^2}{J}\right).$$

Hence

$$\begin{aligned} \widetilde{\sigma}_e^2 &= \frac{SSE}{n-p} \\ \widetilde{\sigma}_A^2 &= \frac{SSM}{n-J} - \frac{\widetilde{\sigma}_e^2}{J}. \end{aligned}$$

For the balanced model these estimates are minimum variance unbiased estimates.

ML Estimation

For the balanced case, the joint log-likelihood of X and SSE is

$$l(\mu, \sigma_A, \sigma_e) = C - \frac{1}{2}\left\{p \log\left(\sigma_A^2 + \frac{\sigma_e^2}{J}\right) + (n-p)\log\sigma_e^2 + \frac{SSE}{\sigma_e^2} + \sum_{i=1}^{p} \frac{(X_i - \mu)^2}{\sigma_A^2 + \frac{\sigma_e^2}{J}}\right\}$$

We estimate μ by $\overline{X}$.

If $\frac{1}{p}\sum\left(X_i - \overline{X}\right)^2 \geq \frac{SSE}{J(n-p)}$, we estimate $\sigma_A^2 + \frac{\sigma_e^2}{J}$ by $\frac{1}{p}\sum\left(X_i - \overline{X}\right)^2$ and then

$$\widetilde{\sigma}_e^2 = SSE/(n-p),\ \widetilde{\sigma}_A^2 = \frac{1}{p}\sum\left(X_i - \overline{X}\right)^2 - \frac{1}{J(n-p)}SSE;$$

If $\frac{1}{p}\sum\left(X_i - \overline{X}\right)^2 < \frac{SSE}{J(n-p)}$,

$$\widetilde{\sigma}_e^2 = \frac{SSE + J\sum\left(X_i - \overline{X}\right)^2}{n} = \frac{SSM}{n},\ \widetilde{\sigma}_A^2 = 0.$$

REML Estimation

W. A. Thompson (1962, Ann. Math. Statist.).

The general idea is to transform the data to a space of lower dimension by using a matrix K whose columns are orthogonal (in the ordinary sense) to the space spanned by the fixed factor effects. Subject to this structural condition, the REML estimates do not depend on the choice of K.

For the balanced case, the joint log-likelihood of $Z = KX$ and SSE is

$$l(\mu, \sigma_A, \sigma_e) = C - \frac{1}{2}\left\{(p-1)\log\left(\sigma_A^2 + \frac{\sigma_e^2}{J}\right) + (n-p)\log\sigma_e^2 + \frac{SSE}{\sigma_e^2} + \frac{\sum\left(X_i - \overline{X}\right)^2}{\left(\sigma_A^2 + \frac{\sigma_e^2}{J}\right)}\right\}$$

since

$$\begin{aligned} Z &= KX \sim N\left(0, \left(\sigma_A^2 + \frac{1}{J}\sigma_e^2\right) I_{p-1}\right) \\ \|Z\|^2 &= \sum\left(X_i - \overline{X}\right)^2. \end{aligned}$$

If $\frac{1}{p-1}\sum\left(X_i - \overline{X}\right)^2 \geq \frac{SSE}{J(n-p)}$,

$$\widetilde{\sigma}_e^2 = SSE/(n-p),\ \widetilde{\sigma}_A^2 = \frac{1}{p-1}\sum\left(X_i - \overline{X}\right)^2 - \frac{1}{J(n-p)}SSE;$$

If $\frac{1}{p-1}\sum\left(X_i - \overline{X}\right)^2 < \frac{SSE}{J(n-p)}$,

$$\widetilde{\sigma}_e^2 = \frac{SSE + \sum\left(X_i - \overline{X}\right)^2}{n-1} = \frac{SSM}{n-1},\ \widetilde{\sigma}_A^2 = 0.$$

Shrinkage Estimation

The REML estimator is

$$\begin{aligned}\delta_{BLUP} &= \widehat{\mu} + \left(\frac{\widehat{\sigma_A^2}}{\widehat{\sigma_A^2} + \widehat{\sigma_e^2}/J}\right)(X - \widehat{\mu}) \\ &= \overline{X} + \left(1 - \frac{(p-1)\,\frac{V}{J}}{\sum\left(X_i - \overline{X}\right)^2}\right)_+ \left(X - \overline{X}\right).\end{aligned}$$

where $V = \frac{SSE}{(n-p)}$. Lindley and Smith (1972, JRSSB).

It is REML estimator, or the positive part of ANOVA estimator, or Efron–Morris estimator (with slightly different thresholding constants).

An example from a telephone call service (see Brown's notes).

Lecture 10. Sharp Asymptotic Minimaxity for Nonparametric Regression

We consider a general setting

$$y_i = \theta_i + \epsilon z_i,\ z_i \overset{i.i.d.}{\sim} N(0,1),\ \theta \in \Theta$$

where Θ is an ellipsoid in $l_2(\mathbb{N})$:

$$\Theta = \left\{ \theta : \sum_i a_i^2 \theta_i^2 \le M \right\}.$$

Pinsker's Theorem: Let $\Theta = \left\{\theta : \sum a_i^2\theta_i^2 \le M\right\}$ and $a_i \to \infty$, then

$$R_L(\Theta,\epsilon) = \inf_\lambda \left[M\lambda^{-2} + \sum_{i=1}^\infty (1 - a_i/\lambda)_+^2 \right]$$

and

$$R_N(\Theta,\epsilon) = (1 + o(1))\, R_L(\Theta,\epsilon) \text{ as } \epsilon \to 0.$$

Linear Minimaxity

We will show

$$R_L(\Theta,\epsilon) = \inf_\lambda \left[M\lambda^{-2} + \sum_{i=1}^\infty (1 - a_i/\lambda)_+^2 \right].$$

The set Θ is orthosymmetric and convex. Without loss of generality, assume that $\epsilon = 1$.

(**John Hartigan's calculation of avoiding minimax theorem**) Recall that

$$R_L(\Theta,\epsilon) = \inf_{(c_i)} \sup_{\sum_i a_i^2\theta_i^2 \le M} \sum_{i=1}^\infty \left[c_i^2 + \theta_i^2 (1 - c_i)^2 \right]$$

which is equal to

$$\begin{aligned}
& \inf_{(c_i)} \left\{ \sup_{\sum_i y_i \le 1} \left[\sum_{i=1}^\infty y_i (1-c_i)^2 \frac{M}{a_i^2} + \sum_{i=1}^\infty c_i^2 \right] \right\} \\
= \;& \inf_{(c_i)} \left\{ \sup_i \left[(1-c_i)^2 \frac{M}{a_i^2} + \sum_{i=1}^\infty c_i^2 \right] \right\} \\
= \;& \inf_\lambda \inf_{\left\{ \sup_i (1-c_i)^2 \frac{1}{a_i^2} = \lambda^{-2} \right\}} \left\{ \sup_i \left[(1-c_i)^2 \frac{M}{a_i^2} + \sum_{i=1}^\infty c_i^2 \right] \right\}
\end{aligned}$$

When $\sup_i (1-c_i)^2 \frac{1}{a_i^2} = \lambda^{-2}$, then $c_i \geq (1 - a_i/\lambda)_+$ for all i and equality holds for at least one i, which implies

$$\inf_{\left\{\sup_i (1-c_i)^2 \frac{1}{a_i^2} = \lambda\right\}} \left[\sup_i (1-c_i)^2 \frac{M}{a_i^2} + \sum_{i=1}^{\infty} c_i^2\right] = M\lambda^{-2} + \sum_{i=1}^{\infty} (1 - a_i/\lambda)_+^2$$

Thus

$$R_L(\Theta, \epsilon) = \inf_{\lambda} \left[M\lambda^{-2} + \sum_{i=1}^{\infty} (1 - a_i/\lambda)_+^2\right]$$

then solve the quadratic minimization for λ. The solution is

$$\sum_{i=1}^{\infty} a_i (\lambda_* - a_i)_+ = M. \tag{1}$$

For general ϵ, we have

$$\epsilon^2 \sum a_i (\lambda_* - a_i)_+ = M$$

and

$$c_i = \epsilon (1 - a_i/\lambda_*)_+ .$$

The corresponding linear minimax risk is

$$R_L(\Theta, \epsilon) = \epsilon^2 \sum (1 - a_i/\lambda_*)_+ .$$

Example. If a_i is a monotone sequence, let $I = \max\{i, 1 \geq a_i/\lambda_*\}$, then

$$R_L(\Theta, \epsilon) = \frac{M}{\epsilon^2}\lambda_*^{-2} + \sum_{i=1}^{I} (1 - a_i/\lambda_*)^2 .$$

The inf is achieved at

$$\lambda_* = \frac{\sum_{i=1}^{I} a_i^2 + \frac{M}{\epsilon^2}}{\sum_{i=1}^{I} a_i},$$

and

$$\begin{aligned} R_L(\Theta, \epsilon) &= \epsilon^2 \sum_{i=1}^{I} (1 - a_i/\lambda_*) \\ &= \epsilon^2 \left(I - \frac{\sum_{i=1}^{I} a_i}{\lambda_*}\right) \end{aligned}$$

Example. Now consider the application of this result to Sobolev ellipsoid. Define

$$\mathcal{F} = \left\{f : \sum_{k=1}^{\infty} (2\pi k)^{2m} \left(\theta_{2k}^2 + \theta_{2k+1}^2\right) \leq M\right\}.$$

Recall the Lagrangian multiplier parameter $\lambda = \lambda_{\epsilon,M}$ is uniquely determined by the equation

$$\epsilon^2 \sum_{i=1}^{\infty} a_i \left(\lambda_* - a_i\right)_+ = M.$$

Note that $a_{2k} = a_{2k+1} = (2\pi k)^m$, i.e.,

$$a_i \sim (\pi i)^m$$

then

$$(\pi I)^m \sim \lambda_*, \text{ i.e., } I \sim \lambda_*^{1/m}/\pi,$$

since $a_i \le \lambda_* \le a_{i+1}$. We approximate the sum by integrals,

$$\begin{aligned} \sum_{i=1}^{I} a_i &\sim \pi^m \frac{I^{m+1}}{m+1} \sim \pi^{-1} \frac{\lambda_*^{(m+1)/m}}{m+1} \\ \sum_{i=1}^{I} a_i^2 &\sim \pi^{2m} \frac{I^{2m+1}}{2m+1} \sim \pi^{-1} \frac{\lambda_*^{(2m+1)/m}}{2m+1} \end{aligned}$$

and

$$\lambda_* \cdot \frac{1}{m+1} \lambda_*^{(m+1)/m} - \frac{1}{2m+1} \lambda_*^{(2m+1)/m} = \pi M/\epsilon^2$$

i.e.,

$$\lambda_*^{(2m+1)/m} \sim \frac{(m+1)(2m+1)}{m} \pi M/\epsilon^2$$

then

$$\begin{aligned} R_L(\Theta, \epsilon) &= \epsilon^2 \left(I - \frac{\sum_{i=1}^{I} a_i}{\lambda_*} \right) \sim \epsilon^2 \pi^{-1} \frac{m}{m+1} \lambda_*^{1/m} \\ &\sim \epsilon^2 \pi^{-1} \frac{m}{m+1} \left[\frac{(m+1)(2m+1)}{m} \frac{\pi M}{\epsilon^2} \right]^{1/(2m+1)} = P_r M^{(1-2r)} \pi^{-2r} \epsilon^{4r}, \end{aligned}$$

where $r = m/(2m+1)$ and *Pinsker constant* is

$$P_r = \left(\frac{m}{m+1} \right)^{2m/(2m+1)} (2m+1)^{1/(2m+1)}.$$

Nonlinear Minimaxity
We may show

$$R_N(\Theta, \epsilon) = (1 + o(1)) R_L(\Theta, \epsilon) \text{ as } \epsilon \to 0.$$

Strategy: Find a sequence of priors G_ϵ supported in Θ such that

$$R_L(\Theta, \epsilon) = (1 + o(1)) r\left(G_\epsilon, \delta_{G_\epsilon}\right).$$

Then

$$R_N(\Theta, \epsilon) \sim R_L(\Theta, \epsilon)$$

since

$$r(G_\epsilon, \delta_{G_\epsilon}) \leq R_N(\Theta, \epsilon) \leq R_L(\Theta, \epsilon).$$

It is natural to define

$$G_\epsilon = \prod_i N\left(0, \epsilon^2 (\lambda_*/a_i - 1)_+\right)$$

with λ_* determined by equation (1), since $\epsilon^2 (\lambda_*/a_i - 1)_+ / \left((\lambda_*/a_i - 1)_+ + 1\right) = \epsilon^2 (1 - a_i/\lambda_*)_+$. But

$$\mathbb{E} \sum a_i^2 \theta_i^2 = \sum \epsilon^2 a_i^2 (\lambda_*/a_i - 1)_+ = M.$$

then G_ϵ is not supported in Θ! What about γG_ϵ with $\gamma \nearrow 1$? That would work.