Prediction

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

April 3, 2012

Optimal prediction

Consider two random variables X and Y. Let $\tilde{X} = f(Y)$ be a predictor of X based on Y. The mean square prediction error is

$$\mathbb{E}(X - \tilde{X})^2$$

The predictor that minimizes the prediction error is the conditional expectation $\mathbb{E}[X|Y]$:

$$\mathbb{E}(X - \tilde{X})^2 =$$
$$\mathbb{E}(X-\mathbb{E}[X|Y])^2+\mathbb{E}(\mathbb{E}[X|Y]-\tilde{X})^2+2\mathbb{C}\mathrm{ov}(X-\mathbb{E}[X|Y],\mathbb{E}[X|Y]-\tilde{X})$$

and

$$\mathbb{C}\mathrm{ov}(X - \mathbb{E}[X|Y], \mathbb{E}[X|Y] - \tilde{X}) =$$
$$\mathbb{C}\mathrm{ov}(\mathbb{E}[X - \mathbb{E}[X|Y]|Y], \mathbb{E}[\mathbb{E}[X|Y] - \tilde{X}|Y])+$$
$$\mathbb{E}\mathbb{C}\mathrm{ov}(X - \mathbb{E}[X|Y], \mathbb{E}[X|Y] - \tilde{X}|Y) = 0$$

("Pythagoras for random variables", $E[X|Y]$ "projection" of X on Y)

Decomposition of X:

$$X = \mathbb{E}[X|Y] + (X - \mathbb{E}[X|Y])$$

where predictor $\mathbb{E}[X|Y]$ and prediction error $X - \mathbb{E}[X|Y]$ uncorrelated. Hence

$$\mathbb{V}\mathrm{ar}X = \mathbb{V}\mathrm{ar}\mathbb{E}[X|Y] + \mathbb{V}\mathrm{ar}(X - \mathbb{E}[X|Y])$$

Note:

$$\mathbb{E}\mathbb{V}\mathrm{ar}(X|Y) = \mathbb{V}\mathrm{ar}(X - \mathbb{E}[X|Y])$$

Pythagoras and conditional expectation

Space of real random variables with finite variance may be viewed as a vector space with inner product and (L_2) norm

$$< X, Y >= \mathbb{C}\mathrm{ov}[X, Y] \quad \|X\| = \sqrt{\mathbb{V}\mathrm{ar}X}$$

Two random variables are orthogonal if they are uncorrelated.
Pythagoras:

$$\|X\|^2 = \|\mathbb{E}[X|Y]\|^2 + \|X - \mathbb{E}[X|Y]\|^2$$

$\mathbb{E}[X|Y]$ may be viewed as projection of X on Y.

Conditional expectation in multivariate normal distribution

Consider jointly normal random vectors X and Y with mean vector

$$\mu = (\mu_X, \mu_Y)$$

and covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix}$$

Then (provided Σ_Y invertible)

$$X|Y = y \sim N(\mu_X + \Sigma_{XY}\Sigma_Y^{-1}(y - \mu_Y), \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX})$$

Proof:

$$X = Z + R$$

where $Z = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - \mu_Y)$ and $R = X - Z$ are independent and R is independent of Y.

Conditional simulation using prediction

Suppose we wish to simulate $X|Y = y$. By previous decomposition

$$X = Z + R$$

we can simulate (X^*, Y^*) and compute Z^* and $R^* = X^* - Z^*$.

Then our conditional simulation is

$$\mathbb{E}[X|Y = y] + R^*$$

Advantageous if it is easier to simulate (X^*, Y^*) and compute Z^* than simulate directly from conditional distribution of $X|Y = y$

(e.g. if Σ_X and Σ_Y easier than $\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}$)

Prediction in linear mixed model (Jiang page 75)

Let $\alpha \sim N(0, G)$ and $Y|\alpha \sim N(X\beta + Z\alpha, R)$.

Then $\mathbb{C}\text{ov}[\alpha, Y] = GZ^\mathsf{T}$ and $\mathbb{V}\text{ar}\, Y = ZGZ^\mathsf{T} + R$.

$$\mathbb{E}[\alpha|Y] = GZ^\mathsf{T}V^{-1}(Y - X\beta)$$

NB:

$$GZ^\mathsf{T}(ZGZ^\mathsf{T} + R)^{-1} = (G^{-1} + Z^\mathsf{T}R^{-1}Z)^{-1}Z^\mathsf{T}R^{-1}$$

e.g. useful if G^{-1} is sparse (like AR-model).

IQ example (Jiang)

$$Y = \mu + \alpha + \epsilon$$

where standard deviation of α and ϵ are 15 and 5 and $\mu = 100$.

Given $Y = 130$, $\mathbb{E}[\mu + \alpha|Y = 130] = 127$.

Example of shrinkage to the mean.

BLUP

Consider random vectors X and Y with with mean vector

$$\mu = (\mu_X, \mu_Y)$$

and covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix}$$

Then best linear unbiased predictor (BLUP) of X given Y is

$$\hat{X} = \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y - \mu_Y)$$

($\mathbb{E}\hat{X} = \mathbb{E}X$, linear in Y and smallest prediction error among linear unbiased predictors)

Covariance matrix of BLUP:

$$\mathbb{C}\mathrm{ov}\hat{X} = \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}$$

Prediction of mixed effect - Henderson's equations

Given β, BLUP of

$$b^\mathsf{T}\beta + a^\mathsf{T}\alpha$$

is

$$b^\mathsf{T}\beta + a^\mathsf{T}GZ^\mathsf{T}V^{-1}(Y - X\beta)$$

Typically β is unknown. Then BLUP is

$$b^\mathsf{T}\hat{\beta} + a^\mathsf{T}GZ^\mathsf{T}V^{-1}(Y - X\hat{\beta})$$

where $\hat{\beta}$ is BLUE (Harville, 1991)

NB: $(\hat{\alpha}, \hat{\beta})$ solution of Henderson's mixed model equations.

EBLUP and EBLUE

Typically covariance matrix depends on unknown parameters.

EBLUPS are obtained by replacing unknown variance parameters by their estimates (similar for EBLUE).

Model assessment

Make histograms, qq-plots etc. for EBLUPs of ϵ and α.

May be advantageous to consider standardized EBLUPS. Standardized BLUP is

$$[\mathbb{C}\mathrm{ov}\hat{\alpha}]^{-1/2}\hat{\alpha}$$

Example: prediction of random intercepts and slopes in orthodont data

```
ort7=lmer(distance~age+factor(Sex)+(1|Subject),data=Orthod
#check of model ort7
#residuals
res=residuals(ort7)
qqnorm(res)
qqline(res)
#outliers occur for subjects M09 and M13
#plot residuals against subjects
boxplot(resort~Orthodont$Subject)
#plot residuals against fitted values
fitted=fitted(ort7)
plot(rank(fitted),resort)
```

```
#extract predictions of random intercepts
raneffects=ranef(ort7)
#qqplot of random intercepts
qqnorm(ranint[[1]])
qqline(ranint[[1]])
#plot for subject M09
M09=Orthodont$Subject=="M09"
plot(Orthodont$age[M09],fitted[M09],type="l",ylim=c(20,32)
points(Orthodont$age[M09],Orthodont$distance[M09])
```