Matrix Models

Start with the "matrix mean model"

$$Y = M + E \quad , \quad ① \; E[E]=0 \quad V[e_{ij}]=\sigma^2 \quad C[e_{ij}, e_{i'j'}]=0$$

$$② \; \{e_{ij}\} \sim \text{iid} \; N(0,\sigma^2)$$

without any assumed structure on M, just normal means model:

$$\text{vec}(Y) = y = m + e = \text{vec}(M) + \text{vec}(E).$$

Commonly used structures:

Additive effects: $M = \mu 11^T + a1^T + 1b^T$, $m_{ij} = \mu + a_i + b_j$

(Q: what is the rank of M?)

Estimation: OLS/MLEs are

$$\hat{\mu} = \bar{y}_{..}$$
$$\hat{a}_i = \bar{y}_{i.} - \bar{y}_{..} \qquad \hat{a} = Y1/p - \hat{\mu}1$$
$$\hat{b}_j = \bar{y}_{.j} - \bar{y}_{..} \qquad \hat{b} = Y^T1/n - \hat{\mu}1$$

Residuals: let $C_K = I_K - 11^T/K$, so $C_n x = x - 11^Tx/n = x - 1\bar{x}$

Note that

$$C_n Y = Y - 11^TY/n = Y - 1[\hat{b}^T + 1^T\hat{\mu}]$$
$$= Y - 1\hat{b}^T - 11^T\hat{\mu}$$

$$C_nYC_p = Y - \hat{a}1^T - 11^T\hat{\mu} - (1\hat{b}^TC_p - \hat{\mu}\cancel{11^TC_p}^{\,0})$$

$$= Y - \hat{a}1^T - 11^T\hat{\mu} - 1\hat{b}^T$$

$$= Y - [\hat{\mu}11^T + \hat{a}1^T + 1\hat{b}^T]$$

C_K is the K-dim "centering matrix."

It is the proj matrix onto the null space of $\underset{\sim}{1}_K$:

$$C_K 1_K = 1_K - 1_K 1_K^T 1_K / K = 0.$$

<u>Multiplicative effects:</u>

$$M = f g^T \qquad f \in \mathbb{R}^n,\ g \in \mathbb{R}^p$$
$$m_{ij} = f_i g_j \qquad \text{(rank-1 model)}$$

$$M = F G^T \qquad F \in \mathbb{R}^{n \times r},\ G \in \mathbb{R}^{p \times r}$$
$$m_{ij} = f_i^T g_j \qquad \text{(rank } r\text{-model)}$$

<u>Estimation</u>

<u>Thm</u> (Eckart - Young) Let $Y = UDV^T$ be the SVD of Y.

Then $\|Y - M\|^2$ is minimized over rank-r matrices M by

$$\hat{M} = d_1 u_1 v_1^T + \cdots + d_r u_r v_r^T = U_r D_{r,r} V_r^T$$

= the "truncated" SVD of Y.

The error in the approximation is

$$\|Y - \hat{M}\|^2 = \underbrace{\sum_{k=1}^{p} d_k^2}_{\|Y\|^2} - \underbrace{\sum_{k=1}^{r} d_k^2}_{\|\hat{M}\|^2} = \sum_{k=r+1}^{p} d_k^2$$

variation in Y not in $\hat{M}$.

So for the model $Y = FG^T + E$, the OLSE/MLE is

$$\widehat{FG^T} = U_r D_r V_r^T$$

Note that F, G not separately identifiable:

$$\hat{F}\hat{G}^T = \hat{F} R R^T \hat{G}^T \quad \text{for any } R \in \mathcal{O}_r .$$

AMMI models (Gauch 1992, Gabriel 1978)

$$M = \mu 1 1^T + a 1^T + 1 b^T + F G^T$$
$$m_{ij} = \mu + a_i + b_j + f_i^T g_j$$

Estimation: $\hat{\theta} = (\hat{\mu}, \hat{\underset{\sim}{a}}, \hat{\underset{\sim}{b}}, \hat{M}) = \arg\min \| Y - (\mu 1 1^T + a 1^T + 1 b^T + M) \|^2$

OLS estimates: let
① $\hat{\mu} = \bar{y}_{..} = 1^T Y 1 / np$

② $\hat{\underset{\sim}{a}} = \bar{y}_{i.} - \bar{y}_{..} = Y 1/p - 1 1^T Y 1/np$

③ $\hat{\underset{\sim}{b}} = \bar{y}_{.j} - \bar{y}_{..} = Y^T 1/p - 1 1^T Y 1/np$

④ $\hat{M} = U_r D_r V_r^T$, where UDV^T is SVD of $\hat{E} = Y - [1 1^T \hat{\mu} + \hat{a} 1^T + 1 b^T]$

* Note: If you change order, you may not get OLS estimates!

Multiplicative Regression Model: (Gabriel 1978)

$$Y = AW^T + XB^T + FG^T + E \qquad \star$$

$$y_{ij} = a_i^T w_j + b_j^T x_i + f_i^T g_j + e_{ij}$$

- $Y \in \mathbb{R}^{n \times p}$, $W \in \mathbb{R}^{p \times q_1}$, $X \in \mathbb{R}^{n \times q_2}$ observed
  - W: col cov, X: row cov
- $A \in \mathbb{R}^{n \times q_1}$, $B \in \mathbb{R}^{p \times q_2}$, $F \in \mathbb{R}^{n \times r}$, $G \in \mathbb{R}^{p \times r}$ to be est.
- E uncorrelated homoscedastic mean zero noise

$\underset{\sim}{a_i}$ = effects of col. vars on ith row, $\underset{\sim}{b_j}$ = effects of row vars on ith col.

Very Useful Result:

Thm (Gabriel 1978)

$$\min_A \min_B \min_{FG^T} \| Y - [AW^T + XB^T + FG^T \|^2$$

$$= \min_{FG^T} \| (I - P_x) Y (I - P_w) - FG^T \|^2$$

where $P_w = w(w^T w)^{-1} w^T$, $P_x = X(X^T X)^{-1} X^T$

Intuition: Recall $P_w W = W$, $(I - P_w) W = 0$
$(I - P_x) X = 0$

so if $Y = AW^T + XB^T + FG^T + E$

then $(I-P_x)Y(I-P_w) = \tilde{Y} = (I-P_x)FG^T(I-P_w) + (I-P_x)E(I-P_w)$

$$\tilde{Y} = \tilde{F}\hat{G}^T + \hat{E}$$

where $\tilde{F}$ is rank r if $n - q_2 \geq r$
$\tilde{G}$ is rank r if $p - q_1 \geq r$

Implication: OLS fit to model (*) can be obtained by

1. obtaining $\hat{A}, \hat{B}$ from OLS fit to linear model $Y = AW^T + XB^T + E$
2. obtaining $\widehat{FG^T}$ from SVD of $(Y - \hat{A}W^T - X\hat{B}^T)$

Exercise: show how to use this result for AMMI model.

Model Reduction

Suppose $Y = \mu 11^T + a1^T + 1b^T + M + E$, $\text{rank}(M) = r < p$

If a, b not of interest, then model can be reduced to a multiplicative model:

$$C_n Y C_p = \tilde{Y} = \tilde{M} + \tilde{E} = C_n M C_p + C_n E C_p$$

$\text{rank}(\tilde{M}) = r$ if $p > r$, $E[\tilde{E}] = 0$

So it seems the model can be reduced to one with no additive effects.

However: (1) $\text{rank}(\tilde{M}) = r$ but the param space is no longer the space of rank 1 matrices.

$\tilde{M}$ also satisfies $\tilde{M}1 = 0 \quad 1^T\tilde{M} = 0$

(2) Elements of $\tilde{E}$ not uncorrelated

$$\tilde{E} = C_n E C_p$$
$$\tilde{e} = (C_p \otimes C_n)$$

$$\text{Cov}(\tilde{e}) = E[\tilde{e}\tilde{e}^T] = (C_p \otimes C_n) E[ee^T](C_p \otimes C_n)$$

$$= (C_p \otimes C_n)(C_p \otimes C_n)$$
$$= (C_p \otimes C_n).$$

Useful Trick: let $C_n = D_n D_n^T$, $D_n \in \mathbb{R}^{n \times n-1}$ $D_n^T D_n = I_{n-1}$
$C_p = D_p D_p^T$ $D_p \in \mathbb{R}^{p \times p-1}$ $D_p^T D_p = I_{p-1}$

If $Y = \mu 1 1^T + a1^T + 1b^T + M + E$, $\text{rank}(M) = r$, $\{e_{ij}\} \sim_{\text{IID}} N(0, \sigma^2)$

$$D_n^T Y D_p = D_n^T M D_p + D_n^T E D_p = \hat{M} + \hat{E} \in \mathbb{R}^{n-1 \times p-1}$$

where (1) param space for $\tilde{M}$ is space of rank r matrices (I think!)

(2) $\{\tilde{e}_{ij}\} \sim_{\text{IID}} N(0, \sigma^2)$ (yes)