NORMAL MODEL FOR MULTIVARIATE DATA ST3457 27/11/17 (1)

$\underline{Y}_1, \dots, \underline{Y}_m$

$\underline{Y}_i = \begin{pmatrix} Y_{1,i} \\ Y_{2,i} \\ \vdots \\ Y_{p,i} \end{pmatrix}$ COLUMN VECTOR, $p \times 1$

SAMPLING MODEL: $\underline{Y}_1, \dots, \underline{Y}_m \mid \underline{\vartheta}, \Sigma \overset{iid}{\sim} N_p(\underline{\vartheta}, \Sigma)$

RECALL: IN THE UNIVARIATE NORMAL CASE WE CONSIDERED 2 PRIOR SPECIFICATIONS.

- $p(\vartheta, \sigma^2) = p(\sigma^2) \cdot p(\vartheta \mid \sigma^2)$, WHERE
  $p(\sigma^2)$ = INVERSE-GAMMA
  $p(\vartheta \mid \sigma^2)$ = NORMAL $(\mu_0, \frac{\sigma^2}{k_0})$

  WE CHECKED THAT
  $p(\sigma^2 \mid y_1, \dots, y_m)$ = INVERSE-GAMMA.
  $p(\vartheta \mid y_1, \dots, y_m, \sigma^2)$ = NORMAL

  $\Rightarrow$ CONJUGATE PRIOR FOR (ϑ, σ^2).

- $p(\vartheta, \sigma^2) = p(\vartheta) \cdot p(\sigma^2)$, WHERE
  $p(\vartheta)$ = NORMAL $(\mu_0, \frac{\sigma_0^2}{k_0})$
  $p(\sigma^2)$ = INVERSE-GAMMA

  IN THIS CASE $\sigma^2 \mid y_1, \dots, y_m$ IS NOT INVERSE-GAMMA (NO CONJUGACY). BUT
  $p(\sigma^2 \mid y_1, \dots, y_m, \vartheta)$ = INVERSE-GAMMA.

MOREOVER, AS BEFORE, (2)

$p(\vartheta \mid y_1, \dots, y_m, \sigma^2)$ = NORMAL.

$\Rightarrow$ BOTH THE FULL CONDITIONAL DISTRIBUTIONS ARE KNOWN IN CLOSED FORM (CONVENIENT IF WE WANT TO USE A GIBBS SAMPLER TO SAMPLE FROM THE JOINT POSTERIOR OF (ϑ, σ^2)).

$\Rightarrow$ SEMI-CONJUGATE PRIOR FOR (ϑ, σ^2)

FOR THE MULTIVARIATE NORMAL MODEL WE'LL FOCUS ON A SEMI-CONJUGATE PRIOR SPECIFICATION

WE CONCLUDED THAT IF $\underline{\vartheta} \sim p(\underline{\vartheta}) \sim N_p(\mu_0, \Lambda_0)$

$\underline{\vartheta} \mid \underline{y}_1, \dots, \underline{y}_m, \Sigma \sim N_p(\underline{\mu}_m, \Lambda_m)$

WHERE $\underline{\mu}_m = (A_0 + A_1)^{-1}(\underline{b}_0 + \underline{b}_1)$

$\Lambda_m = (A_0 + A_1)^{-1}$

WHERE $A_0 = \Lambda_0^{-1}$; $A_1 = m \cdot \Sigma^{-1}$

$b_0 = \Lambda_0^{-1} \cdot \mu_0$; $b_1 = \Sigma^{-1} \cdot m \cdot \bar{\underline{y}}$

THAT IS $\Lambda_m = (\Lambda_0^{-1} + m \Sigma^{-1})^{-1}$.

(LET'S CALL PRECISION MATRIX THE INVERSE OF A COVARIANCE MATRIX).

$$\Rightarrow \underbrace{\Lambda_m^{-1}}_{\text{POSTERIOR PRECISION MATRIX}} = \underbrace{\Lambda_0^{-1}}_{\text{PRIOR PRECISION MATRIX}} + \underbrace{m \cdot \Sigma^{-1}}_{\text{DATA PRECISION MATRIX}}$$

$$\mu_m = (\Lambda_0^{-1} + m\Sigma^{-1})^{-1} \cdot (\Lambda_0^{-1}\mu_0 + \Sigma^{-1} m \bar{y})$$ ③

$$= (\Lambda_0^{-1} + m\Sigma^{-1})^{-1} \cdot \Lambda_0^{-1} \underbrace{\mu_0}_{\text{PRIOR GUESS}} +$$

$$+ (\Lambda_0^{-1} + m\Sigma^{-1})^{-1} \Sigma^{-1} m \underbrace{\bar{y}}_{\text{SAMPLE AVERAGE}}$$

$\Rightarrow$ POSTERIOR MEAN IS A WEIGHTED AVERAGE OF PRIOR GUESS AND SAMPLE AVERAGE

VERY SIMILAR INTERPRETATION TO THE ONE WE SAW IN THE UNIVARIATE CASE. (NOTATION HEAVIER!)

NEXT: WHAT PRIOR CAN WE ASSIGN TO Σ?

RECALL: IF WE WANT $N_p(\underline{\theta}, \Sigma)$ TO BE WELL DEFINED, Σ MUST SATISFY SOME CONDITIONS.

$$\underset{p\times p}{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & & \\ \sigma_{12} & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_p^2 \end{pmatrix}$$

SPECIFICALLY:

1. Σ MUST BE POSITIVE DEFINITE, THAT IS

$$\underline{x}^T \Sigma \underline{x} > 0 \quad \text{FOR ANY VECTOR } x \text{ (} p\times 1 \text{ DIMENSIONAL)}$$

(THIS CONDITION GUARANTEES THAT $\sigma_i^2 > 0 \;\; \forall i = 1, \dots, p$ AND THAT THE CORRELATIONS

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2}\sqrt{\sigma_j^2}} \in [-1, 1]$$

2. Σ ~~MATRIX~~ MUST BE SYMMETRIC. ④

INDEED $\sigma_{ij} = \sigma_{ji} \quad \forall i, j$

WE MUST FIND A PRIOR WHICH ASSIGNS PROBABILITY MASS ONLY TO MATRICES SATISFYING 1 AND 2.

IDEA: WE COULD GENERATE VECTORS $\underline{z}_1, \dots, \underline{z}_m$ FROM A MULTIVARIATE NORMAL AND COMPUTE THE SAMPLE COVARIANCE MATRIX STARTING FROM $\underline{z}_1, \dots, \underline{z}_m$. WHAT WE GET IS A REALIZATION OF A RANDOM MATRIX, WHICH SATISFIES CONDITIONS 1 AND 2.

CONSIDER $\underline{z}_1, \dots, \underline{z}_m$, p-DIMENSIONAL VECTORS

$$\underline{z}_i = \begin{pmatrix} z_{1,i} \\ z_{2,i} \\ \vdots \\ z_{p,i} \end{pmatrix}$$

DEFINE

$$\underline{z}_i \cdot \underline{z}_i^T = \underbrace{\begin{pmatrix} z_{1,i}^2 & z_{1,i} z_{2,i} & & \\ z_{2,i} \cdot z_{i,1} & z_{2,i}^2 & & \\ & & \ddots & \\ & & & z_{p,i}^2 \end{pmatrix}}_{p\times p \text{ MATRIX}}$$

$$\sum_{i=1}^{m} \underline{z}_i \cdot \underline{z}_i^T = \Big[\quad\Big] + \Big[\quad\Big] + \dots + \Big[\quad\Big] =$$

$$= \begin{pmatrix} \sum_{i=1}^m z_{1,i}^2 & \sum_{i=1}^m z_{1,i} \cdot z_{2,i} & & \\ \sum_{i=1}^m z_{2,i} z_{1,i} & \sum_{i=1}^m z_{2,i}^2 & & \\ & & \ddots & \\ & & & \sum_{i=1}^m z_{p,i}^2 \end{pmatrix}$$

WE CALL IT SUM OF SQUARES MATRIX (5)

WE CAN CHECK THAT

$$\sum_{i=1}^{m} \underline{z}_i \cdot \underline{z}_i^T = Z^T Z$$

WHERE $Z = \begin{pmatrix} \underline{z}_1 \\ \underline{z}_2 \\ \vdots \\ \underline{z}_m \end{pmatrix}$ $m \times p$ MATRIX

IF $\underline{z}_1, \dots, \underline{z}_m$ ARE LINEARLY INDEPENDENT

$\Rightarrow Z^T Z$ IS POSITIVE DEFINITE AND SYMMETRIC.

STRATEGY: GIVEN A POSITIVE INTEGER ν_0 AND A COVARIANCE MATRIX Φ_0,

- GENERATE $\underline{z}_1, \dots, \underline{z}_m \overset{IID}{\sim} N_p(\underline{0}, \Phi_0)$
- COMPUTE $Z^T Z$ (SUM OF SQUARES MATRIX)

$\Rightarrow Z^T Z$ IS SUCH THAT:

- IF $\nu_0 > p \Rightarrow Z^T Z$ IS POSITIVE DEFINITE WITH PROBABILITY 1
- $Z^T Z$ IS SYMMETRIC WITH PROB. 1.
- $\mathbb{E}[Z^T Z] = \nu_0 \cdot \Phi_0$.

OBS: SINCE THE MEAN VECTOR OF THE NORMAL IS $\underline{0}$,

$$Z^T Z = m \cdot (\text{SAMPLE COVARIANCE MATRIX})$$

(COHERENT WITH THE ORIGINAL IDEA).

$Z^T Z$ IS A RANDOM MATRIX (DEFINED AS A FUNCTION OF $\underline{z}_1, \dots, \underline{z}_m$ WHICH ARE RANDOM) & ITS DISTRIBUTION IS CALLED WISHART (ν_0, Φ_0) (6)

HENCEFORTH, SET $\nu_0 > p$ AND $S_0 = \Phi_0^{-1}$

STRATEGY:

- SAMPLE $\underline{z}_1, \dots, \underline{z}_{\nu_0} \overset{IID}{\sim} N_p(\underline{0}, S_0^{-1})$
- COMPUTE $Z^T Z = \sum_{i=1} \underline{z}_i \underline{z}_i^T$
- SET $\Sigma = (Z^T Z)^{-1}$

$\Rightarrow \Sigma^{-1}$ HAS WISHART (ν_0, S_0^{-1})

$\Rightarrow \Sigma$ HAS INVERSE-WISHART (ν_0, S_0^{-1})

OBS: $\mathbb{E}[\Sigma^{-1}] = \nu_0 \cdot S_0^{-1}$

$\rightarrow \mathbb{E}[\Sigma] = \dfrac{1}{\nu_0 - p - 1} (S_0^{-1})^{-1} = \dfrac{1}{\nu_0 - p - 1} S_0$

PROPERTY OF THE INVERSE-WISHART DISTR.

IF WE HAVE A PRIOR GUESS, SAY Σ_0, ABOUT Σ, WE CAN SET $\boxed{S_0 = (\nu_0 - p - 1) \cdot \Sigma_0}$

$$\Rightarrow \mathbb{E}[\Sigma] = \frac{\nu_0 - p - 1}{\nu_0 - p - 1} \Sigma_0 = \Sigma_0$$

MOREOVER, ν_0 CAN BE INTERPRETED AS THE "PRIOR SAMPLE SIZE", THE LARGER ν_0 IS, THE MORE STRENGTH WE ASSIGN TO THE PRIOR GUESS.

(7)

IF $\Sigma \sim$ INVERSE-WISHART (ν_0, S_0^{-1})

$$p(\Sigma) \propto |\Sigma|^{-\frac{\nu_0+p+1}{2}} \cdot \exp\{-\mathrm{tr}(S_0 \Sigma^{-1})/2\}$$

where $\mathrm{tr}(A) = \sum_{i=1}^{p} a_{i,i}$ FOR ANY $p \times p$ MATRIX A

↳ THE DEFINITION IS GIVEN HERE UP TO A CONSTANT. THE NORMALIZING CONSTANT IS UGLY, BUT LUCKILY WE WON'T NEED TO USE IT.

ORIGINAL GOAL:

WE KNOW $p(\underline{\theta} | \underline{y}_1, \dots, \underline{y}_m, \Sigma)$

WE NEED $p(\Sigma | \underline{y}_1, \dots, \underline{y}_m, \underline{\theta})$

$$p(\Sigma | \underline{y}_1, \dots, \underline{y}_m, \underline{\theta}) \propto p(\Sigma, \underline{y}_1, \dots, \underline{y}_m, \underline{\theta})$$

$$= p(\underline{y}_1, \dots, \underline{y}_m | \underline{\theta}, \Sigma) \cdot p(\underline{\theta}, \Sigma)$$

$$= p(\underline{y}_1, \dots, \underline{y}_m | \underline{\theta}, \Sigma) \cdot \underbrace{p(\underline{\theta})}_{\text{CONSTANT}} \cdot p(\Sigma)$$

SEMI-CONJUGATE PRIOR SPEC.

$$\propto p(\underline{y}_1, \dots, \underline{y}_m | \underline{\theta}, \Sigma) \cdot p(\Sigma)$$

$$\propto |\Sigma|^{-m/2} \cdot \exp\{-\frac{1}{2} \sum_{i=1}^{m} (\underline{y}_i - \underline{\theta})^T \cdot \Sigma^{-1} \cdot (\underline{y}_i - \underline{\theta})\}$$

$$\cdot |\Sigma|^{-\frac{\nu_0+p+1}{2}} \exp\{-\mathrm{tr}(S_0 \cdot \Sigma^{-1}) \cdot \frac{1}{2}\}$$

FROM MATRIX ALGEBRA

(8)

$$\sum_{i=1}^{m} (\underline{y}_i - \underline{\theta})^T \Sigma^{-1} (\underline{y}_i - \underline{\theta}) = \mathrm{tr}(S_\theta \cdot \Sigma^{-1})$$

WHERE $\underbrace{S_\theta}_{p \times p} = \sum_{i=1}^{m} \underbrace{(\underline{y}_i - \underline{\theta})}_{p \times 1} \cdot \underbrace{(\underline{y}_i - \underline{\theta})^T}_{1 \times p}$

$$= |\Sigma|^{-\frac{\nu_0+m+p+1}{2}} \cdot \exp\{-\mathrm{tr}(S_\theta \cdot \Sigma^{-1})/2 - \mathrm{tr}(S_0 \Sigma^{-1})/2\}$$

$$= |\Sigma|^{-\frac{\nu_0+m+p+1}{2}} \exp\{-\mathrm{tr}((S_0 + S_\theta) \cdot \Sigma^{-1})/2\}$$

AND WE RECOGNIZE THE SHAPE OF AN INVERSE-WISHART (WITH UPDATED PARAMETERS)

$$\Rightarrow \Sigma | \underline{y}_1, \dots, \underline{y}_m, \underline{\theta} \sim \text{INVERSE-WISHART}(\nu_0 + m, (S_0 + S_\theta)^{-1})$$

$$\sim \text{INV-WISHART}(\nu_m, S_m^{-1})$$

WHERE $\nu_m := \nu_0 + m$

$S_m := S_0 + S_\theta$

WHAT IS THE POSTERIOR MEAN? (CONDITIONALLY ON $\underline{\theta}$)

$$\mathbb{E}[\Sigma | \underline{y}_1, \dots, \underline{y}_m, \underline{\theta}] = \frac{1}{\nu_m - p - 1} \cdot (S_m^{-1})^{-1}$$

$$= \frac{1}{\nu_m - p - 1} \cdot S_m$$

$$= \frac{1}{\nu_0+m-p-1}\left(S_0+S_\theta\right) =$$

(9)

$$= \frac{\nu_0-p-1}{\nu_0+m-p-1}\left(\frac{1}{\nu_0-p-1}\cdot S_0\right) + \frac{m}{\nu_0+m-p-1}\left(\frac{1}{m} S_\theta\right)$$

$$\underset{\mathbb{E}[\Sigma]}{\parallel}$$

THIS IS AN UNBIASED ESTIMATOR OF POPULATION COVARIANCE MATRIX

TOMORROW AT 3PM IN LB04 OR LB08?