STAT 3690 Lecture 08

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Feb 9nd, 2022

Assumptions

- Model: $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $n > p$
- Parameter space: $\Theta = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}, \boldsymbol{\Sigma} > 0\}$

Method of moments (MM) estimators for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- No requirement on normality
- Steps
  1. Equate raw moments to their sample counterparts:
  $$\begin{cases} \mathrm{E}(\mathbf{X}) = \bar{\mathbf{X}} \\ \mathrm{E}(\mathbf{X}\mathbf{X}^\top) = n^{-1}\sum_i \mathbf{X}_i\mathbf{X}_i^\top \end{cases} \Leftrightarrow \begin{cases} \boldsymbol{\mu} = \bar{\mathbf{X}} \\ \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top = n^{-1}\sum_i \mathbf{X}_i\mathbf{X}_i^\top \end{cases}$$
  2. Solve the above equations w.r.t. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and obtain estimators
  $$\begin{cases} \hat{\boldsymbol{\mu}}_{\mathrm{MM}} = \bar{\mathbf{X}} \\ \widehat{\boldsymbol{\Sigma}}_{\mathrm{MM}} = n^{-1}\sum_i \mathbf{X}_i\mathbf{X}_i^\top - \bar{\mathbf{X}}\bar{\mathbf{X}}^\top = n^{-1}(n-1)\mathbf{S}, \end{cases}$$
  where $\mathbf{S} = (n-1)^{-1}\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top$

Maximum likelihood (ML) estimation for parameters of MVN (J&W Sec 4.3)

- Likelihood function
$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^n \left[\frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\Big\{ -\frac{1}{2}(\mathbf{X}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\mu})\Big\}\right] \\ &= \frac{1}{\sqrt{(2\pi)^{np}\{\det(\boldsymbol{\Sigma})\}^n}} \exp\Big\{ -\frac{1}{2}\sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\mu})\Big\} \end{aligned}$$
- Log likelihood
$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \ln\{L(\boldsymbol{\mu}, \boldsymbol{\Sigma})\} = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\{\det(\boldsymbol{\Sigma})\} - \frac{1}{2}\sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\mu})$$

- ML estimator
$$(\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}}) = \arg\max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\bar{\mathbf{X}}, n^{-1}(n-1)\mathbf{S})$$

- Properties of $(\hat{\boldsymbol{\mu}}_{\text{ML}}, \widehat{\boldsymbol{\Sigma}}_{\text{ML}})$
  - Consistency: $(\hat{\boldsymbol{\mu}}_{\text{ML}}, \widehat{\boldsymbol{\Sigma}}_{\text{ML}}) \xrightarrow{P} (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
  - Efficiency: As $n \to \infty$, the covariance of $(\hat{\boldsymbol{\mu}}_{\text{ML}}, \widehat{\boldsymbol{\Sigma}}_{\text{ML}})$ achieves the Cramer-Rao lower bound.
  - Invariance: For any function g, the ML estimator of $g(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is $g(\hat{\boldsymbol{\mu}}_{\text{ML}}, \widehat{\boldsymbol{\Sigma}}_{\text{ML}})$.

Sampling distributions of $\bar{\mathbf{X}}$ and $\mathbf{S}$ (J&W Sec 4.4)

- Recall the univariate case: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, $n > p$
  - $S^2 \perp\!\!\!\perp \bar{X}$
  - $\sqrt{n}(\bar{X} - \mu)/\sigma \sim N(0, 1)$
  - $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, where $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$
  - $\sqrt{n}(\bar{X} - \mu)/S \sim t(n-1)$
- The multivariate case
  - $\mathbf{S} \perp\!\!\!\perp \bar{\mathbf{X}}$, i.e., $\widehat{\boldsymbol{\Sigma}}_{\text{ML}} \perp\!\!\!\perp \hat{\boldsymbol{\mu}}_{\text{ML}}$
  - $\sqrt{n}\boldsymbol{\Sigma}^{-1/2}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim MVN_p(\mathbf{0}, \mathbf{I})$
  - $(n-1)\mathbf{S} = n\widehat{\boldsymbol{\Sigma}}_{\text{ML}} \sim W_p(n-1, \boldsymbol{\Sigma})$
  - $n(\bar{\mathbf{X}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim$ Hotelling's $T^2(p, n-1)$

- Wishart distribution
  - Def: $W_p(\boldsymbol{\Sigma}, n)$ is the distribution of $\sum_{i=1}^n \mathbf{Y}_i \mathbf{Y}_i^\top$ with $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \overset{\text{iid}}{\sim} MVN_p(\mathbf{0}, \boldsymbol{\Sigma})$
    - A generalization of χ^2-distribution: $W_p(\boldsymbol{\Sigma}, n) = \chi^2(n)$ if $p = \boldsymbol{\Sigma} = 1$
  - Propoties
    - $\mathbf{A}\mathbf{A}^\top > 0$ and $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n) \Rightarrow \mathbf{A}\mathbf{W}\mathbf{A}^\top \sim W_p(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, n)$
    - $\mathbf{W}_i \overset{\text{iid}}{\sim} W_p(\boldsymbol{\Sigma}, n_i) \Rightarrow \mathbf{W}_1 + \mathbf{W}_2 \sim W_p(\boldsymbol{\Sigma}, n_1 + n_2)$
    - $\mathbf{W}_1 \perp\!\!\!\perp \mathbf{W}_2$, $\mathbf{W}_1 + \mathbf{W}_2 \sim W_p(\boldsymbol{\Sigma}, n)$ and $\mathbf{W}_1 \sim W_p(\boldsymbol{\Sigma}, n_1) \Rightarrow \mathbf{W}_2 \sim W_p(\boldsymbol{\Sigma}, n - n_1)$
    - $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n)$ and $\boldsymbol{a} \in \mathbb{R}^p \Rightarrow$
    $$\frac{\boldsymbol{a}^\top \mathbf{W} \boldsymbol{a}}{\boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a}} \sim \chi^2(n)$$
    - $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n)$, $\boldsymbol{a} \in \mathbb{R}^p$ and $n \geq p \Rightarrow$
    $$\frac{\boldsymbol{a}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}}{\boldsymbol{a}^\top \mathbf{W}^{-1} \boldsymbol{a}} \sim \chi^2(n - p + 1)$$
    - $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n) \Rightarrow$
    $$\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{W}) \sim \chi^2(np)$$

- Hotelling's T^2 distribution
  - A generalization of (Student's) t-distribution
  - If $\mathbf{X} \sim MVN_p(\mathbf{0}, \mathbf{I})$ and $\mathbf{W} \sim W_p(\mathbf{I}, n)$, then
  $$\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X} \sim T^2(p, n)$$
  - $Y \sim T^2(p, n) \Leftrightarrow \frac{n-p+1}{np} Y \sim F(p, n-p+1)$

- Wilk's lambda distribution
  - Wilks's lambda is to Hotelling's T^2 as F distribution is to Student's t in univariate statistics.
  - Given independent $\mathbf{W}_1 \sim W_p(\boldsymbol{\Sigma}, n_1)$ and $\mathbf{W}_2 \sim W_p(\boldsymbol{\Sigma}, n_2)$ with $n_1 \geq p$,
  $$\Lambda = \frac{\det(\mathbf{W}_1)}{\det(\mathbf{W}_1 + \mathbf{W}_2)} = \frac{1}{\det(\mathbf{I} + \mathbf{W}_1^{-1}\mathbf{W}_2)} \sim \Lambda(p, n_1, n_2)$$
  - Resort to approximations for computation: $\{(p - n_2 + 1)/2 - n_1\} \ln \Lambda(p, n_1, n_2) \approx \chi^2(n_2 p)$