L10: One sample estimation

1. Maximum likelihood estimators

   (1) Likelihood function and maximum likelihood estimators.
   Let $X = (X_1, ..., X_n) \in R^{p\times n}$ be a random sample from $N(\mu, \Sigma)$. The joint pdf of X, treated as a function of μ and Σ, is the likelihood function (LF).

$$\begin{aligned} L(\mu, \Sigma) &= \prod_{i=1}^n \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(X_i-\mu)'\Sigma^{-1}(X_i-\mu)\right] \\ &= \frac{1}{(2\pi)^{(np)/2}|\Sigma|^{n/2}} \exp\left[-\frac{1}{2}\sum_i (X_i-\overline{X}+\overline{X}-\mu)'\Sigma^{-1}(X_i-\overline{X}+\overline{X}-\mu)\right] \\ &= \frac{1}{(2\pi)^{(np)/2}|\Sigma|^{n/2}} \exp\left[-\frac{1}{2}\sum_i (X_i-\overline{X})'\Sigma^{-1}(X_i-\overline{X})\right] \exp\left[-\frac{n}{2}(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu)\right] \\ &= \frac{1}{(2\pi)^{(np)/2}|\Sigma|^{n/2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1/2}\ \mathrm{CSSCP}\ \Sigma^{-1/2}\right)\right] \exp\left[-\frac{n}{2}(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu)\right]. \end{aligned}$$

   $\widehat{\mu}$ and $\widehat{\Sigma}$ are maximum likelihood estimators (MLEs) for μ and Σ if

$$L(\mu, \Sigma) \le L(\widehat{\mu}, \widehat{\Sigma}) \text{ for all } \mu \text{ and } \Sigma.$$

   (2) MLE of μ
   Note that Maximizing $L(\mu, \Sigma)$ over all μ for all $\Sigma \Longleftrightarrow$ Minimizing $(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu)$.
   But $(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu)$ is minimized when $\mu = \overline{X}$.
   Thus $\widehat{\mu} = \overline{X}$ is the MLE for μ.

   (3) MLE of Σ
   Let $\Sigma^{-1/2}\mathrm{CSSCP}\Sigma^{-1/2} = P\Lambda P'$ be the EVD.

$$\begin{aligned} L(\overline{X}, \Sigma) &= \frac{|\Sigma^{-1/2}\ \mathrm{CSSCP}\ \Sigma^{-1/2}|^{n/2}}{(2\pi)^{np/2}|\mathrm{CSSCP}|^{n/2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1/2}\ \mathrm{CSSCP}\ \Sigma^{-1/2}\right)\right] \\ &= \frac{|P\Lambda P'|^{n/2}}{(2\pi)^{np/2}|\mathrm{CSSCP}|^{n/2}} \exp\left[-\frac{1}{2}\mathrm{tr}(P\Lambda P')\right] \\ &= \frac{1}{(2\pi)^{np/2}|\mathrm{CSSCP}|^{n/2}} (\lambda_1\cdots\lambda_p) \exp\left(-\frac{\lambda_1+\cdots+\lambda_p}{2}\right) \\ &= \frac{1}{(2\pi)^{np/2}|\mathrm{CSSCP}|^{n/2}} \prod_i \lambda_i^{n/2} e^{-\frac{\lambda_i}{2}} = \frac{1}{(2\pi)^{np/2}|\mathrm{CSSCP}|^{n/2}} \prod_i g(\lambda_i) \end{aligned}$$

   where $g(\lambda_i) = \lambda_i^{n/2}e^{-\frac{\lambda_i}{2}}$ is maximized when $\lambda_i = n$. So with $\lambda_i = n$ for all $i = 1, ..., p$, $\Sigma^{-1/2}\ \mathrm{CSSCP}\ \Sigma^{-1/2} = P\Lambda P' = nI$, i.e., $\Sigma = \frac{\mathrm{CSSCP}}{n}$. Hence $L(\overline{X}, \Sigma)$ is maximized when $\Sigma = \frac{\mathrm{CSSCP}}{n}$. Thus $\frac{\mathrm{CSSCP}}{n}$ is a MLE for Σ.

Comments: (i) Recall: $S = \frac{\mathrm{CSSCP}}{n-1}$ is UE for Σ.
(ii) Direct computation shows that $L\left(\overline{X}, \frac{\mathrm{CSSCP}}{n}\right) = \left(\frac{n}{2\pi e}\right)^{np/2} |\mathrm{CSSCP}|^{n/2}$.
(iii) SAS for computing $\overline{X}$, CSSCP and S by SAS:

```
data a; infile "C:\ex.txt";
        input x1 x2 x3;
proc corr nocorr CSSCP COV;
       var x1 x2 x3;
       run;
```

2. Confidence regions

   (1) Pivotal quantity and confidence region
   For parameter vector $\theta \in R^p$ if $C(X) \subset R^p$ where X is the data matrix such that
   $$P(\theta \in C(X)) \ge 1-\alpha,$$
   then the random region $C(X)$ is called a confidence region for θ with confidence coefficient $1-\alpha$.
   $T(X, \theta)$ is called a pivotal quantity if its distribution is free of parameter θ.

Let $\mathcal{A} \subset R^p$ be a high probability density region for pivotal quantity $T(X,\theta)$ such that $P(T(X,\theta) \in \mathcal{A}) \geq 1-\alpha$. With $T(X,\theta) \in \mathcal{A} \Longleftrightarrow \theta \in C(X)$,

$$P(\theta \in C(X)) = P(T(X,\theta) \in \mathcal{A}) \geq 1-\alpha.$$

So $C(X)$ is a $1-\alpha$ confidence region for θ.

(2) Confidence region for β

Based on sample $X \sim N_{p\times n}(\mu 1_n', \Sigma, I_n)$, $T^2 = (\overline{X}-\mu)'\left(\frac{S}{n}\right)^{-1}(\overline{X}-\mu) \sim T^2(p, n-1)$ is a pivotal quantity. By $T^2(p, n-1) = \frac{p(n-1)}{n-p}F(p, n-p)$,

$$P\left(T^2 \leq \frac{p(n-1)}{n-p}F_\alpha(p, n-p)\right) = 1-\alpha.$$

But

$$T^2 \leq \frac{p(n-1)}{n-p}F_\alpha(p, n-p) \Longleftrightarrow \mu \in \left\{(\mu-\overline{X})'\left(\frac{S}{n}\right)^{-1}(\mu-\overline{X}) \leq \frac{p(n-1)}{n-p}F_\alpha(p, n-p)\right\}.$$

Thus $\left\{\mu \in R^p : (\mu-\overline{X})'\left(\frac{S}{n}\right)^{-1}(\mu-\overline{X}) \leq \frac{p(n-1)}{n-p}F_\alpha(p, n-p)\right\}$
is a $1-\alpha$ confidence region for μ.

(3) An ellipsoid

Let $c = \frac{p(n-1)}{n-p}F_\alpha(p, n-p)$ and $\frac{cS}{n} = P\Lambda P'$ be the EVD. By the orthogonal transformation $y = P'(\mu-\overline{X})$,

$$\begin{aligned} & (\mu-\overline{X})'\left(\tfrac{S}{n}\right)^{-1}(\mu-\overline{X}) \leq \tfrac{p(n-1)}{n-p}F_\alpha(p, n-p) \\ \Longleftrightarrow\ & (\mu-\overline{X})'(P\Lambda P')^{-1}(\mu-\overline{X}) \leq 1 \\ \Longleftrightarrow\ & y'\Lambda^{-1}y \leq 1 \Longleftrightarrow \tfrac{y_1^2}{(\sqrt{\lambda_1})^2} + \cdots + \tfrac{y_p^2}{(\sqrt{\lambda_p})^2} \leq 1 \end{aligned}$$

Thus the $1-\alpha$ confidence region for μ is an ellipsoid with center $\overline{X}$ and half-axes $\sqrt{\lambda_1}, \ldots., \sqrt{\lambda_p}$.

Ex1: Find a formula for 95% confidence region for $\mu \in R^3$ based on a sample of size 20.
$(\mu-\overline{X})'\left(\frac{S}{20}\right)^{-1}(\mu-\overline{X}) \leq \frac{3\times 19}{17}F_{0.05}(3, 17) = \frac{57}{17}\times 3.21 = 10.73$
So $\{\mu \in R^3 : (\mu-\overline{X})'(0.5365S)^{-1}(\mu-\overline{X}) \leq 1\}$ is a 95% confidence region for μ.

3. Confidence region for $\theta = L\mu \in R^q$

(1) Transformations

Transformations of populations, parameters, samples and statistics.
$X \in R^{p\times n}$ is a random sample from $N(\mu, \Sigma)$. Then $LX \in R^{q\times n}$ is a random sample from $N(L\theta, L\Sigma L')$.
If $\overline{X}$ and S are the sample mean and sample covariance matrix from X, $L\overline{X}$ and LSL' are sample mean and sample covariance matrix from LX.

(2) Confidence region for $\theta = L\mu$

$\left\{\theta \in R^q : (\theta - L\overline{X})'\left(\frac{LSL'}{n}\right)^{-1}(\theta - L\overline{X}) \leq \frac{q(n-1)}{n-q}F_\alpha(q, n-q)\right\}$
is a $1-\alpha$ confidence region for $\theta = L\mu$.

Ex2: When $q = 1$, write $\theta = L\mu = l'\mu$, $L\overline{X} = l'\overline{X}$ and $LSL' = l'Sl$. Then

$$\begin{aligned} & (l'\mu - l'\overline{X})'\left(\tfrac{l'Sl}{n}\right)^{-1}(l'\mu - l'\overline{X}) \leq \tfrac{1(n-1)}{n-1}F_\alpha(1, n-1) \\ \Longleftrightarrow\ & (l'\mu - l'\overline{X})^2 \leq \tfrac{l'Sl}{n}F_\alpha(1, n-1) \Longleftrightarrow l'\mu \in l'\overline{X} \pm \sqrt{F_\alpha(1, n-1)}\sqrt{l'Sl/n}. \end{aligned}$$

So $l'\overline{X} \pm \sqrt{F_\alpha(1, n-1)}\sqrt{l'Sl/n}$ is the $1-\alpha$ confidence interval for $l'\mu$.

L11: Simultaneous CIs

1. Confidence regions and confidence intervals

   (1) Formulas
   $1-\alpha$ confidence region for $\mu \in R^p$ is the collection of μ satisfying

   $$(\mu - \overline{X})' \left(\frac{S}{n}\right)^{-1} (\mu - \overline{X}) \le T^2_\alpha(p,\, n-1).$$

   $1-\alpha$ confidence region for $\theta = L\mu \in R^q$ is the collection of θ satisfying

   $$(\theta - L\overline{X})' \left(\frac{LSL'}{n}\right)^{-1} (\theta - L\overline{X}) \le T^2_\alpha(q,\, n-1).$$

   $1-\alpha$ confidence interval for $\theta = l'\mu$ is

   $$l'\overline{X} \pm \sqrt{T^2_\alpha(1,\, n-1)} \sqrt{\frac{l'Sl}{n}}.$$

   (2) Table values
   $T^2_\alpha(p,\, n-1) = \frac{p(n-1)}{n-p} F_\alpha(p,\, n-p)$ and $T^2(q,\, n-1) = \frac{q(n-1)}{n-q} F_\alpha(1,\, n-q)$ can be calculated with F-table or APP. While $T^2(1,\, n-1) = F(1,\, n-1)$, because $F(1,\, n-1) = [t(n-1)]^2$, $\alpha = P(F(1,\, n-1) > F_\alpha(1,\, n-1)) = P([t(n-1)]^2 > F_\alpha(1,\, n-1)) = 2P(t(n-1) > \sqrt{F_\alpha(1,\, n-1)}\ )$. So $\alpha/2 = P(t(n-1) > \sqrt{F_\alpha(1,\, n-1)})$. Thus $\sqrt{F_\alpha(1,\, n-1)} = \sqrt{T^2(1,\, n-1)} = t_{\alpha/2}(n-1)$. This value can be looked up from t-table or calculated by APP.

   (3) Statistics
   In the last formula, $\overline{X} \sim N\left(\mu,\, \frac{\Sigma}{n}\right) \Longrightarrow l'\overline{X} \sim N\left(l'\mu,\, \frac{l'\Sigma l}{n}\right)$ where $\sigma^2_{l'\overline{X}} = \frac{l'\Sigma l}{n}$ is the variance of $l'\overline{X}$. When Σ is estimated by S, $\sigma^2_{l'\overline{X}}$ is estimated by $S^2_{l'\overline{X}} = \frac{l'Sl}{n}$.

   (4) Confidence intervals
   $1-\alpha$ CI for $l'\mu$ is $l'\overline{X} \pm t_{\alpha/2}(n-1) S_{l'\overline{X}}$.

   Ex1: A sample of size $n = 61$ from $N(\mu,\, \Sigma)$ produced $\overline{X} = \begin{pmatrix} 17.98 \\ 31.13 \end{pmatrix}$ and $S = \begin{pmatrix} 9.95 & 13.88 \\ 13.88 & 21.26 \end{pmatrix}$.
   Find a 95% confidence region for μ.
   $(\mu - \overline{X})' \left(\frac{S}{n}\right)^{-1} (\mu - \overline{X}) \le T^2_\alpha(p,\, n-1) = T^2_{0.05}(2,\, 60) = \frac{2\times 60}{59} F_{0.05}(2,\, 59) = \frac{120}{59} \times 3.15312 = 6.4131.$
   So $$\left\{ \mu \in R^2 :\ \begin{pmatrix} \mu_1 - 17.98 \\ \mu_2 - 31.13 \end{pmatrix}' \begin{pmatrix} 1.0461 & 1.4592 \\ 1.4592 & 2.2351 \end{pmatrix}^{-1} \begin{pmatrix} \mu_1 - 17.98 \\ \mu_2 - 31.13 \end{pmatrix} \le 1 \right\}$$
   is a 95% confidence region for μ.

   Ex2: In Ex1 find 95% confidence interval for $\mu_1 + \mu_2$.
   $$\begin{aligned} & \overline{X}_1 + \overline{X}_2 \pm t_{\alpha/2}(n-1) S_{\overline{X}_1 + \overline{X}_2} \\ = \ & 17.98 + 31.13 \pm t_{0.025}(60) \sqrt{\tfrac{9.95+13.88+13.88+21.26}{61}} = 49.11 \pm 1.96 \times 0.9832 \\ = \ & 49.11 \pm 1.93 = (47.18,\, 51.04) \end{aligned}$$
   is a 95% confidence interval for $\mu_1 + \mu_2$.

2. Bonferroni simultaneous confidence intervals

   (1) Definition
   Let I_i be a CI for θ_i, $i = 1, ..., k$. If $P(\theta_1 \in I_1, ..., \theta_k \in I_k) \ge 1-\alpha$, then $I_1, ..., I_k$ are simultaneous confidence intervals for $\theta_1, ..., \theta_k$ with overall confidence coefficient $1-\alpha$.

   (2) Bonferroni method
   To construct simultaneous confidence intervals for $\theta_1, ..., \theta_k$ with overall confidence coefficient $1-\alpha$, one can construct them one by one, each has confidence coefficient $1 - \frac{\alpha}{k}$.

Proof. Let $A_i = [\theta_i \in I_i]$ and suppose $P(A_i) = P(\theta_i \in I_i) \geq 1 - \frac{\alpha}{k}$, $i = 1, ..., k$.

We show $P(A_1 \cap \cdots \cap A_k) \geq 1 - \alpha$.

$$\begin{aligned} P(A_1 \cap \cdots \cap A_k) &= 1 - P((A_1 \cap \cdots \cap A_k)^c) = 1 - P(A_1^c \cup \cdots \cup A_k^c) \\ &\geq 1 - [P(A_1^c) + \cdots + P(A_k^c)] = 1 - [1 - P(A_1) + \cdots + 1 - P(A_k)] \\ &= 1 - k + [P(A_1) + \cdots + P(A_k)] \\ &\geq 1 - k + (1 - \tfrac{\alpha}{k} + \cdots + 1 - \tfrac{\alpha}{k}) = 1 - k + (k - \alpha) = 1 - \alpha \end{aligned}$$

(3) Simultaneous CIs for $\theta_i = l_i'\mu \in R$, $i = 1, ..., k$.

$$\theta_i \in l_i'\overline{X} \pm t_{\alpha/(2k)}(n-1) S_{l_i'\overline{X}}, \; i = 1, ..., k$$

are k simultaneous CIs with overall confidence coefficient $1 - \alpha$.

Ex3: In Ex1 find simultaneous CIs for μ_1 and μ_2 with overall confidence coefficient 90%.

With $k = 2$, $\alpha = 0.1$ and $n = 61$, $t_{\alpha/(2k)}(n-1) = t_{0.025}(60) = 1.96$.

$\mu_1 \in \overline{X}_1 \pm 1.96 \times \sqrt{\frac{9.95}{61}} = 17.98 \pm 1.96 \times 0.404 = 17.98 \pm 0.79 = (17.19,\ 18.77)$ and

$\mu_2 \in \overline{X}_2 \pm 1.96 \times \sqrt{\frac{21.26}{61}} = 31.13 \pm 1.96 \times 0.59 = 31.138 \pm 1.16 = (29.97,\ 32.29)$

are simultaneous CIs with overall confidence coefficient 90%.

3. Scheffe's intervals

(1) Extended Cauchy-Schwartz inequality

For $x,\ y \in R^p$, $(x'y)^2 \leq (x'x)(y'y)$ is the well-known Cauchy-Schwartz inequality.

With $A > 0$, replacing x by $A^{-1/2}x$, and y be $A^{1/2}y$ leads to the extended Cauchy-Schwartz inequality

$$(x'y)^2 \leq (x'A^{-1}x)(y'Ay).$$

(2) A lemma

If $x'A^{-1}x \leq c$, then $x'y \in \pm\sqrt{c}\,\sqrt{y'Ay}$, i.e., $-\sqrt{c}\sqrt{y'Ay} \leq x'y \leq \sqrt{c}\sqrt{y'Ay}$.

Proof. If $x'A^{-1}x \leq c$, by the extended Cauchy-Schwartz inequality $(x'y)^2 \leq (x'A^{-1}x)\,(y'Ay)$, $(x'y)^2 \leq c\,(y'Ay)$. Hence $x'y \in \pm\sqrt{c}\,\sqrt{y'Ay}$.

(3) Scheffe's intervals

For $\theta_i = l_i'\mu$, $i = 1, 2, \cdots$,

$$\theta_i \in l_i'\overline{X} \pm \sqrt{T_\alpha^2(p,\ n-1)}\, S_{l_i'\overline{X}},\ i = 1, 2, \cdots$$

are simultaneous CIs with overall confidence coefficient $1 - \alpha$.

Proof. Let $D_i = \left[l_i'\mu \in l_i'\overline{X} \pm \sqrt{T_\alpha^2(p,\ n-1)}\, S_{l_i'\overline{X}}\right]$. We show $P(D_1 \cap D_2 \cap \cdots) \geq 1 - \alpha$.

Let $D = \left[(\mu - \overline{X})' \left(\frac{S}{n}\right)^{-1} (\mu - \overline{X}) \leq T_\alpha^2(p,\ n-1)\right]$. Then $P(D) = 1 - \alpha$.

With $x = \mu - \overline{X}$, $y = l_i$, $A = \frac{S}{n}$, and $c = T_\alpha^2(p,\ n-1)$, by (2)

$$\begin{aligned} D &= \left[(\mu - \overline{X})' \left(\tfrac{S}{n}\right)^{-1} (\mu - \overline{X}) \leq T_\alpha^2(p,\ n-1)\right] \\ &\subset \left[(\mu - \overline{X})'l_i \in \pm\sqrt{T_\alpha^2(p,\ n-1)}\, \sqrt{\tfrac{l_i'Sl_i'}{n}}\right] = \left[l_i'(\mu - \overline{X}) \in \pm\sqrt{T_\alpha^2(p,\ n-1)}\, S_{l_i'\overline{X}}\right] \\ &= \left[l_i'\mu \in l_i'\overline{X} \pm \sqrt{T_\alpha^2(p,\ n-1)}\, S_{l_i'\overline{X}}\right] = D_i \text{ for all } i = 1, 2, \cdots \end{aligned}$$

Thus $D \subset D_1 \cap D_2 \cap \cdots$ Hence $1 - \alpha = P(D) \leq P(D_1 \cap D_2 \cap \cdots)$.

Ex4: In Ex1 find Scheffe's simultaneous confidence intervals for μ_1 and μ_2 with overall confidence coefficient 90%.

$\sqrt{T_{0.1}^2(2,\ 60)} = \sqrt{\frac{2\times 60}{59} F_{0.1}(2,\ 59)} = \sqrt{\frac{2\times 60}{59} \times 2.395} = 2.207.$

$\mu_1 \in \overline{X}_1 \pm 2.207 \times \sqrt{\frac{9.95}{61}} = 17.98 \pm 2.207 \times 0.404 = 17.98 \pm 0.89 = (17.09,\ 18.87)$ and

$\mu_2 \in \overline{X}_2 \pm 2.207 \times \sqrt{\frac{21.26}{61}} = 31.13 \pm 2.207 \times 0.59 = 31.138 \pm 1.30 = (29.84,\ 32.44)$

are Scheffe's intervals for μ_1 and μ_2 with overall confidence coefficient 90%.