Note: Statistical Inference

Oct 2024

Lecturer: *Typed by: Zhuohua Shen*

Contents

Chapter 1

Preliminary

Chapter 2

Statistical inference fundamentals

References: most of the contents are from the undergraduate course STA3020 (by Prof. Jianfeng Mao in 2022-2023 T1, and Prof. Jiasheng Shi in 2023-2024 T2) and postgraduate course STAT5010 (by Kin Wai Keith Chan in 2024-2025 T1), with main textbook Casella and Berger [1]

2.1 Statistical Models

See Chapter 3 of [1]. Suppose $X_i \sim_{\text{iid}} \mathbb{P}_*$, where $\mathbb{P}_*$ refers to the unknown data generating process (DGPg), we find $\widehat{\mathbb{P}} \approx \mathbb{P}_*$. A statistical model is a set of distributions $\mathscr{F} = \{\mathbb{P}_\theta : \theta \in \Theta\}$, where Θ is the parameter space. A parametric model is the model with $\dim(\Theta) < \infty$, while a nonparametric model satisfies $\dim(\Theta) = \infty$.

Definition 2.1.1 (**Exponential family**). A k-dimensional exponential family (EF) $\mathscr{F} = \{f_\theta : \theta \in \Theta\}$ is a model consisting of pdfs of the form

$$f_\theta(x) = c(\theta)h(x)\exp\left\{\sum_{j=1}^{k} \eta_j(\theta)T_j(x)\right\} \tag{2.1}$$

where $c(\theta), h(x) \geq 0$, $\Theta = \{\theta : c(\theta) \geq 0, \eta_j(\theta)$ being well defined for $1 \leq j \leq k\}$. Let $\eta_j = \eta_j(\theta)$, the canonical form is

$$f_\eta(x) = b(\eta)h(x)\exp\left\{\sum_{j=1}^{k} \eta_j T_j(x)\right\}, \tag{2.2}$$

- k-dim natural exponential family (NEF): $\mathscr{F}' = \{f_\eta : \eta \in \Xi\}$;
- natural parameter $\eta = (\eta_1, \ldots, \eta_k)^\top$;
- natural parameter space: $\Xi = \{\eta \in \mathbb{R}^k : 0 < b(\eta) < \infty\}$;
- the NEF $\mathscr{F}'$ is of full rank if Ξ contains an open set in $\mathbb{R}^k$;
- the EF is a curved exponential family if $p = \dim(\Theta) < k$.

Properties of EF:

- Let $X \sim f_\eta$, where $\eta \in \Xi$ such that (i) f_η is of the form (2.2) with $B(\eta) = -\log b(\eta)$, and (ii) Ξ contains an open set in $\mathbb{R}^k$. Then, for $j, j' = 1, \ldots, k$, $\mathbb{E}\{T_j(X)\} = \partial B(\eta)/\partial \eta_j$ and $\mathbb{C}\text{ov}\{T_j(X), T_{j'}(X)\} = \partial^2 B(\eta)/\left(\partial \eta_j \partial \eta_{j'}\right)$.
- Stein's identity:

Definition 2.1.2 (**Location-scale family**). Let f be a density.

- A location-scale family is given by $\mathscr{F} = \{f_{\mu,\sigma} : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^{++}\}$, where $f_{\mu,\sigma}(x) = f\left((x-\mu)/\sigma\right)/\sigma$.
- location parameter: μ; scale parameter: σ; standard density: f;
- A location family is $\mathscr{F} = \{f_{\mu,1} : \mu \in \mathbb{R}\}$.
- A scale family is $\mathscr{F} = \{f_{0,\sigma} : \sigma \in \mathbb{R}^{++}\}$

Representation: $X = \mu + \sigma Z$, $Z \sim f_{0,1}(\cdot)$.

- See some examples in Example 3.9, Keith's note 3, and Table 1 in Shi's note L1.
- Transform between location parameter and scale parameter by taking log.

Definition 2.1.3 (**Identifiable family**). If $\forall \theta_1, \theta_2 \in \Theta$ that

$$\theta_1 \neq \theta_2 \quad \Rightarrow \quad f_{\theta_1}(\cdot) \neq f_{\theta_2}(\cdot),$$

then $\mathscr{F}$ is said to be an identifiable family, or equivalently $\theta \in \Theta$ is identifiable.

A typical feature of non-identifiable EF is that $p = \dim(\Theta) \geqslant k$. Typically,

- $p < k$, curved (must).
- $p = k$, of full rank.
- $p > k$, non-identifiable.

2.2 Principles of Data Reduction

Statistics: $T = T(X_{1:n})$, a function of $X_{1:n}$ and free of any unknown parameter.

2.2.1 Sufficiency Principle

Sufficiency principle: If $T = T(X_{1:n})$ is a "sufficient statistics" for θ, then any inference on θ will depend on $X_{1:n}$ only through T.

Definition 2.2.1 (Sufficient, minimal sufficient, ancillary, and complete statistics). Suppose $X_{1:n} \sim_{\text{iid}} \mathbb{P}_\theta$, where $\theta \in \Theta$. Let $T = T(X_{1:n})$ be a statistic. Then T is sufficient (SS) for θ

- $\Leftrightarrow$ (def) $[X_{1:n} \mid T = t]$ is free of θ for each t.
- $\Leftrightarrow$ (technical lemma) $T(x_{1:n}) = T(x'_{1:n})$ implies that $f_\theta(x_{1:n})/f_\theta(x'_{1:n})$ is free of θ.
- $\Leftrightarrow$ (Neyman-Fisher factorization theorem) $\forall \theta \in \Theta$, $x_{1:n} \in \mathscr{X}^n$, $f_\theta(x_{1:n}) = A(t,\theta)B(x_{1:n})$.
- $\Leftrightarrow$ Define $\Lambda(\theta', \theta'' \mid x_{1:n}) := f_{\theta'}(x_{1:n})/f_{\theta''}(x_{1:n})$. $\forall \theta', \theta'' \in \Theta$, $\exists$ function $C_{\theta',\theta''}$ such that $\Lambda(\theta', \theta'' \mid x_{1:n}) = C_{\theta',\theta''}(t)$, for all $x_{1:n} \in \mathscr{X}^n$ where $t = T(x_{1:n})$.

T is minimal sufficient (MSS) for θ

- $\Leftrightarrow$ (def) (1) T is a SS for θ; (2) $T = g(S)$ for any other SS S.
- $\Leftrightarrow$ (1) T is a SS for θ; (2) $S(x_{1:n}) = S(x'_{1:n})$ implies $T(x_{1:n}) = T(x'_{1:n})$ for any SS S.
- $\Leftrightarrow$ (Lehmann-Scheffé theorem) $\forall x_{1:n}, x'_{1:n} \in \mathscr{X}^n$, $f_\theta(x_{1:n})/f_\theta(x'_{1:n})$ is free of $\theta \Leftrightarrow T(x_{1:n}) = T(x'_{1:n})$.

$A = A(X_{1:n})$ is ancillary (ANS) if the distribution of A does not depend on θ.
T is complete (CS) if $\forall \theta \in \Theta$, $\mathbb{E}_\theta g(T) = 0$ implies $\forall \theta \in \Theta$, $\mathbb{P}_\theta\{g(T) = 0\} = 1$.

Properties

- (Transformation) If $T = r(T')$, then (i) T is SS $\Rightarrow$ T' is SS; (ii) T' is CS $\Rightarrow$ T is CS; (iii) r is one-to-one, then if one is SS/MSS/CS, then the another is.
- (Basu's Lemma) $X_i \sim_{\text{iid}} \mathbb{P}_\theta$, A is ANS and T s CSS, then $A \perp\!\!\!\perp T$.
- (Bahadur's theorem) $X_i \sim_{\text{iid}} \mathbb{P}_\theta$, if an MSS exists, then any CSS is also an MSS.
  - Then if a CSS exists, then any MSS is also a CSS $\Rightarrow$ CSS=MSS.
  - All or nothing: start with MSS T, check whether T is CS. (i) Yes, it is both CSS and MSS, then the set of MSS=CSS; (ii) No, there is no CSS at all.
- (Exp-family) If $X_i \sim_{\text{iid}} f_\eta$ in (2.2), then $T = (\sum_{i=1}^n T_1(X_i), \ldots, \sum_{i=1}^n T_k(X_i))$ is a SS, called natural sufficient statistic. If Ξ contains an open set in $\mathbb{R}^k$ (i.e., $\mathscr{F}'$ is of full rank), then T is MSS and CSS.

Proof techniques

- Prove T is not sufficient for θ: show if $\exists x_{1_n}, x'_{1:n} \in \mathcal{X}^n$ and $\theta', \theta'' \in \Theta$, such that $T(x_{1:n}) = T(x'_{1:n})$ and $\Lambda(\theta', \theta'' \mid x_{1:n}) \neq \Lambda(\theta', \theta'' \mid x'_{1:n})$.
- Prove A is an ANS: consider location-scale representation.
- Prove T is a CS: use definition or take $\mathrm{d}\mathbb{E}_\theta g(T)/\mathrm{d}\theta = 0$.
- Disprove T is CS:
  - Construct an ANS $S(T)$ based on T, then $\mathbb{E}S(T)$ is free of θ, then $g(T) = S(T) - \mathbb{E}S(T)$ is free of θ but $g(T) \neq 0$ w.p.1.
  - (Cancel the 1st moment) Find two unbiased estiamtors for θ as a function of T. E.g., $X_1, X_2 \sim_{\text{iid}} \mathrm{N}(\theta, \theta^2)$, $T = (X_1, X_2)$, $g(T) = X_1 - X_2 \sim \mathrm{N}(0, 2\theta^2)$.

Remark 2.2.2.

- ANS A is useless on its own, but useful together with other information.
- $\mathbb{P}(A(\boldsymbol{X}) \mid \theta)$ is free of θ, but for non-SS T, $\mathbb{P}(A(\boldsymbol{X}) \mid T(\boldsymbol{X}))$ is not necessarily free of θ.

2.2.2 Likelihood principle

Chapter 3

Multivariate Inference Fundamentals

Reference:

- Robb J. Muirhead - Aspects of multivariate statistical theory [2].
- CUHK STAT4002 - Applied Multivariate Analysis (2023 Spring), by Zhixiang Lin.

3.1 Random vectors and distributions

Definition 3.1.1. Let $\boldsymbol{x} = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$ be a random vector,

- Mean $\mathbb{E}\boldsymbol{x} = \boldsymbol{\mu} = (\mathbb{E}x_1, \ldots, \mathbb{E}x_p)^\top = (\mu_j)$.
- Covariance matrix $\mathbb{V}\mathrm{ar}(\boldsymbol{x}) = \mathbb{C}\mathrm{ov}(\boldsymbol{x}) = \Sigma = \mathbb{E}[(\boldsymbol{x} - \mathbb{E}\boldsymbol{x})(\boldsymbol{x} - \mathbb{E}\boldsymbol{x})^\top] = \mathbb{E}\boldsymbol{x}\boldsymbol{x}^\top - \mathbb{E}\boldsymbol{x}\mathbb{E}\boldsymbol{x}^\top = (\sigma_{ij})$, $\Sigma \succeq \mathbf{0}$.
- Correlation matrix $R = D^{-1/2}\Sigma D^{-1/2}$, where $D = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$. We have $R_{ij} = \rho_{ij} = \sigma_{ij}/(\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}})$.
- If $\boldsymbol{y} \in \mathbb{R}^q$ random vector, then $\mathbb{C}\mathrm{ov}(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}[(\boldsymbol{x} - \mathbb{E}\boldsymbol{x})(\boldsymbol{y} - \mathbb{E}\boldsymbol{y})^\top] = \mathbb{E}\boldsymbol{x}\boldsymbol{y}^\top - \mathbb{E}\boldsymbol{x}\mathbb{E}\boldsymbol{y}^\top \in \mathbb{R}^{p\times q}$.

If $\boldsymbol{Z} = (z_{ij}) \in \mathbb{R}^{p\times q}$ is a random matrix,

- $\mathbb{E}\boldsymbol{Z} = (\mathbb{E}z_{ij})$.

Proposition 3.1.2. *Let $\boldsymbol{x} \in \mathbb{R}^p$ be a random vector, $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^p$ be vectors, $A \in \mathbb{R}^{r_1\times p}, B \in \mathbb{R}^{r_2\times p}$ be matrices,*

- $\mathbb{E}\boldsymbol{a}^\top\boldsymbol{x} = \boldsymbol{a}^\top\mathbb{E}\boldsymbol{x}$, $\mathbb{V}\mathrm{ar}(\boldsymbol{a}^\top\boldsymbol{x}) = \boldsymbol{a}^\top\Sigma\boldsymbol{a}$, *and* $\mathbb{C}\mathrm{ov}(\boldsymbol{a}^\top\boldsymbol{x}, \boldsymbol{b}^\top\boldsymbol{x}) = \boldsymbol{a}^\top\Sigma\boldsymbol{b}$.
- $\mathbb{E}A\boldsymbol{x} = A\mathbb{E}\boldsymbol{x}$, $\mathbb{V}\mathrm{ar}(A\boldsymbol{x}) = A\Sigma A^\top$, *and* $\mathbb{C}\mathrm{ov}(A\boldsymbol{x}, B\boldsymbol{x}) = A\Sigma B^\top$.
- *If* $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b}$, *where* $A \in \mathbb{R}^{q\times p}$, $\boldsymbol{b} \in \mathbb{R}^q$, *then* $\boldsymbol{\mu}_{\boldsymbol{y}} = A\boldsymbol{\mu}_{\boldsymbol{x}} + \boldsymbol{b}$ *and* $\Sigma_{\boldsymbol{y}} = A\Sigma_{\boldsymbol{x}}A^\top$.

Let $\boldsymbol{Z} \in \mathbb{R}^{p\times q}$ be a random matrix, $B \in \mathbb{R}^{m\times p}$, $C \in \mathbb{R}^{q\times n}$, and $D \in \mathbb{R}^{m\times n}$ constants, then

- $\mathbb{E}(B\boldsymbol{Z}C + D) = B\mathbb{E}(\boldsymbol{Z})C + D$.

- The $\Sigma \in \mathbb{R}^{p\times p}$ is a covariance matrix (i.e., $\Sigma = \mathbb{C}\mathrm{ov}(\boldsymbol{x})$ for some random vector $\boldsymbol{x} \in \mathbb{R}^p$) iff $\Sigma \succeq \mathbf{0}$.
  - ($\Leftarrow$): suppose $\mathrm{r}(\Sigma) = r \le p$, write full rank decomposition $\Sigma = CC^\top$, $C \in \mathbb{R}^{p\times r}$. Let $\boldsymbol{y} \sim [\mathbf{0}_r, I_r]$, then $\mathbb{C}\mathrm{ov}(C\boldsymbol{y}) = \Sigma$.
- If Σ is not PD, then $\exists \boldsymbol{a} \ne \mathbf{0}_p$ s.t. $\mathbb{V}\mathrm{ar}(\boldsymbol{a}^\top\boldsymbol{x}) = 0$ so w.p.1., $\boldsymbol{a}^\top\boldsymbol{x} = k$, i.e., $\boldsymbol{x}$ lies in a hyperplane.

Theorem 3.1.3. *If $\boldsymbol{x} \in \mathbb{R}^p$ random, then its distribution is uniquely determined by the distributions of $\boldsymbol{a}^\top\boldsymbol{x}$, $\forall \boldsymbol{a} \in \mathbb{R}^p$.*

The proof uses the fact that a distribution in $\mathbb{R}^p$ is uniquely determined by its ch.f., see Theorem 1.2.2. [2].

Definition 3.1.4. Dataset contains p variables and n observations are represented by $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$, where the ith row $\boldsymbol{x}_i^\top = (x_{i1}, \ldots, x_{ip})$ is the ith observation vector, $i = 1, \ldots, n$.

- (Sample mean vector) $\bar{\boldsymbol{x}} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i = (\bar{x}_1, \ldots, \bar{x}_p)^\top$, where $\bar{x}_j = n^{-1}\sum_{i=1}^n x_{ij}$.
- (Sum of squares and cross product (SSCP) matrix) $\boldsymbol{A} = \sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top$.
- (Sample covariance matrix) $\boldsymbol{S} = (n-1)^{-1}\boldsymbol{A}$.
- (Sample correlation matrix) $\boldsymbol{R} = D^{-1/2}\boldsymbol{S}D^{-1/2}$, where $D^{-1/2} = \mathrm{diag}(1/\sqrt{s_{11}}, \ldots, 1/\sqrt{s_{pp}})$.

- $\bar{\boldsymbol{x}} = n^{-1}\boldsymbol{X}^\top\mathbf{1}_n$, and

$$\begin{aligned}\boldsymbol{A} &= \sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top - n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^\top \\ &= (\boldsymbol{X} - \mathbf{1}_n\bar{\boldsymbol{x}}^\top)^\top(\boldsymbol{X} - \mathbf{1}_n\bar{\boldsymbol{x}}^\top) \succeq \mathbf{0}.\end{aligned}$$

- $\mathbb{E}\bar{\boldsymbol{x}} = \boldsymbol{\mu}$, $\mathbb{V}\mathrm{ar}(\bar{\boldsymbol{x}}) = n^{-1}\Sigma$, $\mathbb{E}\boldsymbol{A} = (n-1)\Sigma$, and $\mathbb{E}\boldsymbol{S} = \Sigma$.

3.1.1 Multivariate normal distribution

Definition 3.1.5 (Original definition of multivariate normal)**.** The random vector $\boldsymbol{x} \in \mathbb{R}^p$ is said to have an p-variate normal distribution ($\boldsymbol{x} \sim \mathrm{N}_p$) if $\forall \boldsymbol{a} \in \mathbb{R}^p$, the distribution of $\boldsymbol{a}^\top \boldsymbol{x}$ is univariate normal.

Theorem 3.1.6 (Fundamental properties)**.** *Let $\boldsymbol{x} \sim \mathrm{N}_p$, we have*

1. *Both $\boldsymbol{\mu} = \mathbb{E}\boldsymbol{x}$ and $\Sigma = \mathbb{C}\mathrm{ov}(\boldsymbol{x})$ exist and the distribution of $\boldsymbol{x}$ is determined by $\boldsymbol{\mu}$ and Σ. Write $\boldsymbol{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$.*
2. *(Representation) Let $\Sigma \succeq \mathbf{0}_{p\times p}$, $\mathrm{r}(\Sigma) = r \le p$, and $u_{1:r} \sim_{\mathrm{iid}} \mathrm{N}(0,1)$, i.e., $\boldsymbol{u} \sim \mathrm{N}_r(\mathbf{0}_r, I_r)$, then if C is the full rank decomposition of Σ and $\boldsymbol{\mu} \in \mathbb{R}^p$, then $\boldsymbol{x} = C\boldsymbol{u} + \boldsymbol{\mu} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$.*
   - *Let $\Sigma = HDH^\top$ be the spectral decomposition, then $\boldsymbol{x} = HD^{1/2}\boldsymbol{z} + \boldsymbol{\mu}$, where $\boldsymbol{z} \sim \mathrm{N}_p(\mathbf{0}_p, I_p)$.*
3. *If $\boldsymbol{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, then its ch.f. $\phi_{\boldsymbol{x}}(\boldsymbol{t}) = \exp(i\boldsymbol{\mu}^\top \boldsymbol{t} - \boldsymbol{t}^\top \Sigma \boldsymbol{t}/2)$.*
4. *(Density) $\boldsymbol{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ with $\Sigma \succ \mathbf{0}$, then $\boldsymbol{x}$ has pdf*

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}. \tag{3.1}$$

Note that we guarantee the existence of $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ by means of the representation in point 2.

Theorem 3.1.7 (Properties of multivariate normal)**.** *If $\boldsymbol{x} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, then we have*

1. *(Linearity) Let $B \in \mathbb{R}^{q\times p}, \boldsymbol{b} \in \mathbb{R}^q$ nonrandom, and $B\Sigma B^\top \succ \mathbf{0}$, then $B\boldsymbol{x} + \boldsymbol{b} \sim \mathrm{N}_q(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B^\top)$.*
2. *(Linear combinations) If $\boldsymbol{x}_k \sim \mathrm{N}_p(\boldsymbol{\mu}_k, \Sigma_k) \perp\!\!\!\perp$ for $k = 1, \dots, N$, then for any fixed constants $\alpha_1, \dots, \alpha_N$, $\sum_{k=1}^N \alpha_k \boldsymbol{x}_k \sim \mathrm{N}_p(\sum_{k=1}^N \alpha_k \boldsymbol{\mu}_k, \sum_{k=1}^N \alpha_k^2 \Sigma_k)$.*
   - *The sample mean $\bar{\boldsymbol{x}} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma/N)$.*
3. *(Subset) The marginal distribution of any subset of $k(< p)$ components of $\boldsymbol{x}$ is k-variate normal.*
4. *(Marginal distribution) Partition*

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \boldsymbol{x}_1 \in \mathbb{R}^q, \boldsymbol{x}_2 \in \mathbb{R}^{p-q}, \Sigma_{12} \in \mathbb{R}^{q\times(p-q)}.$$

   Then $\boldsymbol{x}_1 \sim \mathrm{N}_q(\boldsymbol{\mu}_1, \Sigma_{11})$, $\boldsymbol{x}_1 \perp\!\!\!\perp \boldsymbol{x}_2$ iff $\Sigma_{12} = \mathbf{0}$.
5. *(Conditional distribution) Let Σ_{22}^- be a generalized inverse of Σ_{22} (i.e., $\Sigma_{22}\Sigma_{22}^-\Sigma_{22} = \Sigma_{22}$), then*
   (a) $\boldsymbol{x}_1 - \Sigma_{12}\Sigma_{22}^-\boldsymbol{x}_2 \sim \mathrm{N}_q(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^-\boldsymbol{\mu}_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{21})$, and $\perp\!\!\!\perp \boldsymbol{x}_2$.
   (b) $[\boldsymbol{x}_1 \mid \boldsymbol{x}_2] \sim \mathrm{N}_q(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^-(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{21})$.
6. *(Cramér) If $p \times 1$ random vectors $\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{y}$ and $\boldsymbol{x} + \boldsymbol{y} \sim \mathrm{N}_p$, then both $\boldsymbol{x}, \boldsymbol{y} \sim \mathrm{N}_p$.*
7. *(MLE) of (μ, Σ) is $(\bar{\boldsymbol{x}}, A/n)$.*
8. *(Quadratic form) If $\boldsymbol{x}_{1:n} \sim_{\mathrm{iid}} \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, then $n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^\top \Sigma^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \sim \chi_p^2$. The squared generalized distance (Mahalanobis distance) $d_i^2 = (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top \boldsymbol{S}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \xrightarrow{\mathrm{d}} \chi_p^2$.*

For point 3, each component of a random vector is (marginally) normal does not imply that the vector has a multivariate normal distribution. Counterexample: let $U_1, U_2, U_3 \sim_{\mathrm{iid}} \mathrm{N}(0,1)$, $Z \perp\!\!\!\perp U_{1:3}$. Define

$$X_1 = \frac{U_1 + ZU_3}{\sqrt{1+Z^2}}, \quad X_2 = \frac{U_2 + ZU_3}{\sqrt{1+Z^2}}.$$

Then $[X_1|Z] \sim \mathrm{N}(0,1)$, free of Z, so $X_1 \sim \mathrm{N}(0,1)$, and $X_2 \sim \mathrm{N}(0,1)$. But (X_1, X_2) not normal. The converse is true if the components of $\boldsymbol{x}$ are all independent and normal, or if $\boldsymbol{x}$ consists of independent subvectors, each of which is normally distributed.

For the proof of point 5, we use the lemma: if $\Sigma \succeq \mathbf{0}$, then $\ker(\Sigma_{22}) \subset \ker(\Sigma_{12})$, and $\mathrm{range}(\Sigma_{21}) \subset \mathrm{range}(\Sigma_{22})$. So $\exists B \in \mathbb{R}^{q\times(p-q)}$ satisfying $\Sigma_{12} = B\Sigma_{22}$.

3.1.2 The noncentral χ^2 and F distributions

3.2 Asymptotic properties

3.2.1 Asymptotic distributions of sample means and covariance matrices

Refer to section 1.2.2, [2].

Theorem 3.2.1 (CLT for sample means)**.** *Let $\boldsymbol{x}_{1:n} \sim_{\mathrm{iid}} [\boldsymbol{\mu}, \Sigma]$, then*

$$\sqrt{n}(\bar{\boldsymbol{x}}_n - \boldsymbol{\mu}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu}) \xrightarrow{\mathrm{d}} \mathrm{N}_p(\mathbf{0}_p, \Sigma).$$

Theorem 3.2.2 (CLT for sample covariance matrices)**.** *Let* $\boldsymbol{x}_{1:n} \sim_{\mathrm{iid}} [\boldsymbol{\mu}, \Sigma]$ *with finite fourth moments, SSCP matrix* $\boldsymbol{A} = \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top$, *and* $\boldsymbol{S} = (n-1)^{-1}\boldsymbol{A}$. *Let* $V = \mathbb{C}\mathrm{ov}[\mathrm{vec}((\boldsymbol{x}_1 - \boldsymbol{\mu})(\boldsymbol{x}_1 - \boldsymbol{\mu})^\top)]$, *then*

$$\frac{1}{\sqrt{n}}(\mathrm{vec}(\boldsymbol{A}) - n \cdot \mathrm{vec}(\Sigma)) \xrightarrow{\mathrm{d}} \mathrm{N}_{p^2}(\boldsymbol{0}, V),$$

$$\sqrt{n-1}(\mathrm{vec}(\boldsymbol{S}) - \mathrm{vec}(\Sigma)) \xrightarrow{\mathrm{d}} \mathrm{N}_{p^2}(\boldsymbol{0}, V).$$

Note that $V \in \mathbb{R}^{p^2 \times p^2}$ is singular as the LHS vectors above have repeated elements.

Bibliography

[1] G. Casella and R. L. Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002. 2, 2.1

[2] R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, 1982. 3, 3.1, 3.2.1