

Note: Statistical Inference

Oct 2024

Lecturer:

Typed by: Zhuohua Shen

Contents

1	Preliminary	5
2	Statistical inference fundamentals	7
2.1	Statistical Models	7
2.2	Principles of Data Reduction	8
2.2.1	Sufficiency Principle	8
2.2.2	Likelihood principle	8
3	Multivariate Inference Fundamentals	9
3.1	Random vectors and distributions	9
3.1.1	Multivariate normal distribution	9
3.1.2	Basic multivariate distributions	10
3.2	Asymptotic properties	10
3.2.1	Asymptotic distributions of sample means and covariance matrices	10

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** (DGP), we find $\hat{\mathbb{P}} \approx \mathbb{P}_*$. A **statistical model** is a set of distributions $\mathcal{F} = \{\mathbb{P}_\theta : \theta \in \Theta\}$, 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) $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ is a model consisting of pdfs of the form

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

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

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

- k -dim **natural exponential family** (NEF): $\mathcal{F}' = \{f_\eta : \eta \in \Xi\}$;
- **natural parameter** $\eta = (\eta_1, \dots, \eta_k)^\top$;
- **natural parameter space**: $\Xi = \{\eta \in \mathbb{R}^k : 0 < b(\eta) < \infty\}$;
- the NEF \mathcal{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 $j, j' = 1, \dots, k$, $\mathbb{E}\{T_j(X)\} = \partial B(\eta) / \partial \eta_j$ and $\text{Cov}\{T_j(X), T_{j'}(X)\} = \partial^2 B(\eta) / (\partial \eta_j \partial \eta_{j'})$.
- **Stein's identity**:

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

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

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

- $p < k$, 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) $\forall \theta \in \Theta, x_{1:n} \in \mathcal{X}^n, f_\theta(x_{1:n}) = A(t, \theta)B(x_{1:n})$.

\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 $x_{1:n} \in \mathcal{X}^n$ 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) $\forall x_{1:n}, x'_{1:n} \in \mathcal{X}^n, f_\theta(x_{1:n})/f_\theta(x'_{1:n})$ 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 is 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 $\Rightarrow \text{CSS} = \text{MSS}$.
 - **All or nothing:** start with MSS T , check whether T is CS. (i) Yes, it is both CSS and MSS, then the set of $\text{MSS} = \text{CSS}$; (ii) No, there is no CSS at all.
- (Exp-family) If $X_i \sim \text{iid } f_\eta$ in (2.2), then $T = (\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i))$ is a SS, called **natural sufficient statistic**. If Ξ contains an open set in \mathbb{R}^k (i.e., \mathcal{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 $d\mathbb{E}_\theta g(T)/d\theta = 0$.
- 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 estimators for θ as a function of T . E.g., $X_1, X_2 \sim \text{iid } N(\theta, \theta^2)$, $T = (X_1, X_2)$, $g(T) = X_1 - X_2 \sim N(0, 2\theta^2)$.

Remark 2.2.2. • ANS A is useless on its own, but useful together with other information.

- $\mathbb{P}(A(\mathbf{X}) \mid \theta)$ is free of θ , but for non-SS T , $\mathbb{P}(A(\mathbf{X}) \mid T(\mathbf{X}))$ 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 $\mathbf{x} = (x_1, \dots, x_p)^\top \in \mathbb{R}^p$ be a random vector,

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

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

- $\mathbb{E}\mathbf{a}^\top \mathbf{x} = \mathbf{a}^\top \mathbb{E}\mathbf{x}$, $\text{Var}(\mathbf{a}^\top \mathbf{x}) = \mathbf{a}^\top \Sigma \mathbf{a}$, and $\text{Cov}(\mathbf{a}^\top \mathbf{x}, \mathbf{b}^\top \mathbf{x}) = \mathbf{a}^\top \Sigma \mathbf{b}$.
- $\mathbb{E}A\mathbf{x} = A\mathbb{E}\mathbf{x}$, $\text{Var}(A\mathbf{x}) = A\Sigma A^\top$, and $\text{Cov}(A\mathbf{x}, B\mathbf{x}) = A\Sigma B^\top$.
- If $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, where $A \in \mathbb{R}^{q \times p}$, $\mathbf{b} \in \mathbb{R}^q$, then $\boldsymbol{\mu}_{\mathbf{y}} = A\boldsymbol{\mu}_{\mathbf{x}} + \mathbf{b}$ and $\Sigma_{\mathbf{y}} = A\Sigma_{\mathbf{x}}A^\top$.

Let $\mathbf{Z} \in \mathbb{R}^{p \times q}$ 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

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

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

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

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 $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$, where the i th row $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip})$ is the i th observation vector, $i = 1, \dots, n$.

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

- $\bar{\mathbf{x}} = n^{-1}X^\top \mathbf{1}_n$, $A = (X - \mathbf{1}_n \bar{\mathbf{x}}^\top)^\top (X - \mathbf{1}_n \bar{\mathbf{x}}^\top) \succeq \mathbf{0}$.
- $\mathbb{E}\bar{\mathbf{x}} = \boldsymbol{\mu}$, $\text{Var}(\bar{\mathbf{x}}) = n^{-1}\Sigma$, $\mathbb{E}A = (n-1)\Sigma$, and $\mathbb{E}S = \Sigma$.

3.1.1 Multivariate normal distribution

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

Theorem 3.1.6 (Fundamental properties). Let $\mathbf{x} \sim N_p$, we have

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

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

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

Theorem 3.1.7 (Properties of multivariate normal). If $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then we have

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

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

Then $\mathbf{x}_1 \sim N_q(\boldsymbol{\mu}_1, \Sigma_{11})$, $\mathbf{x}_1 \perp\!\!\!\perp \mathbf{x}_2$ 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) $\mathbf{x}_1 - \Sigma_{12}\Sigma_{22}^-\mathbf{x}_2 \sim N_q(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^-\boldsymbol{\mu}_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{21})$, and $\perp\!\!\!\perp \mathbf{x}_2$.
 - (b) $[\mathbf{x}_1 | \mathbf{x}_2] \sim N_q(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^-(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{21})$.
6. (**Cramér**) If $p \times 1$ random vectors $\mathbf{x} \perp\!\!\!\perp \mathbf{y}$ and $\mathbf{x} + \mathbf{y} \sim N_p$, then both $\mathbf{x}, \mathbf{y} \sim N_p$.

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 $U_1, U_2, U_3 \sim \text{iid} N(0, 1)$, $Z \perp\!\!\!\perp U_{1:3}$. 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 $[X_1 | Z] \sim N(0, 1)$, free of Z , so $X_1 \sim N(0, 1)$, and $X_2 \sim N(0, 1)$. But (X_1, X_2) not normal. The converse is true if the components of \mathbf{x} are all independent and normal, or if \mathbf{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 $\text{im}(\Sigma_{21}) \subset \text{im}(\Sigma_{22})$. So $\exists B \in \mathbb{R}^{q \times (p-q)}$ satisfying $\Sigma_{12} = B\Sigma_{22}$.

3.1.2 Basic multivariate distributions

- (**Sample mean**) If $\mathbf{x}_{1:n} \sim \text{iid} N_p(\boldsymbol{\mu}, \Sigma)$, then $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, n^{-1}\Sigma)$, and $n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \Sigma^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim \chi_p^2$. The squared generalized distance (Mahalanobis distance) $d_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})^T S^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) \xrightarrow{d} \chi_p^2$.
- **MLE** of $(\boldsymbol{\mu}, \Sigma)$ is $(\bar{\mathbf{x}}, A/n)$.

Definition 3.1.8 (Wishart distribution).

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 $\mathbf{x}_{1:n} \sim \text{iid} [\boldsymbol{\mu}, \Sigma]$, then

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

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](#)