Math Tools

Note: Statistical Inference

Oct 2024

Lecturer: *Typed by: Zhuohua Shen*

Contents

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 [1]

1 Statistical Models

See Chapter 3 of [1]. Suppose $X_i \sim_{\text{iid}} \mathrm{P}_*$, where P_* refers to the unknown data generating process (DGPg), we find $\widehat{\mathrm{P}} \approx \mathrm{P}_*$. A statistical model is a set of distributions $\mathscr{F} = \{\mathrm{P}_\theta : \theta \in \Theta\}$, 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 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

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

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

$$f_\eta(x) = b(\eta)h(x)\exp\left\{\sum_{j=1}^{k} \eta_j T_j(x)\right\}, \tag{2}$$

- k-dim natural exponential family (NEF): $\mathscr{F}' = \{f_\eta : \eta \in \Xi\}$;
- natural parameter $\eta = (\eta_1, \ldots, \eta_k)^T$;
- 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) 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 $\mathbf{Cov}\{T_j(X), T_{j'}(X)\} = \partial^2 B(\eta)/(\partial\eta_j\partial\eta_{j'})$.
- Stein's identity:

Definition 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 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 $p = \dim(\Theta) > k$. Typically,

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

2 Principles of Data Reduction

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

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 2.1 (Sufficient, minimal sufficient, ancillary, and complete statistics). Suppose $X_{1:n} \sim_{\mathrm{iid}} \mathrm{P}_\theta$, 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 $\forall \theta \in \Theta$, $\mathbf{E}_\theta\, g(T) = 0$ implies $\forall \theta \in \Theta$, $\mathrm{P}_\theta\{g(T) = 0\} = 1$.

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) $X_i \sim_{\mathrm{iid}} \mathrm{P}_\theta$, A is ANS and T s CSS, then $A \perp\!\!\!\perp T$.
- (Bahadur's theorem) $X_i \sim_{\mathrm{iid}} \mathrm{P}_\theta$, 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 $X_i \sim_{\mathrm{iid}} f_\eta$ in (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., $\mathscr{F}'$ 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 $\mathrm{d}\,\mathbf{E}_\theta\, g(T)/\mathrm{d}\theta = 0$.
- Disprove T is CS:
  - Construct an ANS $S(T)$ based on T, then $\mathbf{E}\, S(T)$ is free of θ, then $g(T) = S(T) - \mathbf{E}\, S(T)$ 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., $X_1, X_2 \sim_{\mathrm{iid}} \mathrm{N}(\theta, \theta^2)$, $T = (X_1, X_2)$, $g(T) = X_1 - X_2 \sim \mathrm{N}(0, 2\theta^2)$.

Remark 2.2.
- ANS A is useless on its own, but useful together with other information.
- $\mathrm{P}(A(\boldsymbol{X}) \mid \theta)$ is free of θ, but for non-SS T, $\mathrm{P}(A(\boldsymbol{X}) \mid T(\boldsymbol{X}))$ is not necessarily free of θ.

2.2 Likelihood principle

References

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