Section for Statistical Theory

TA: Cong Mu

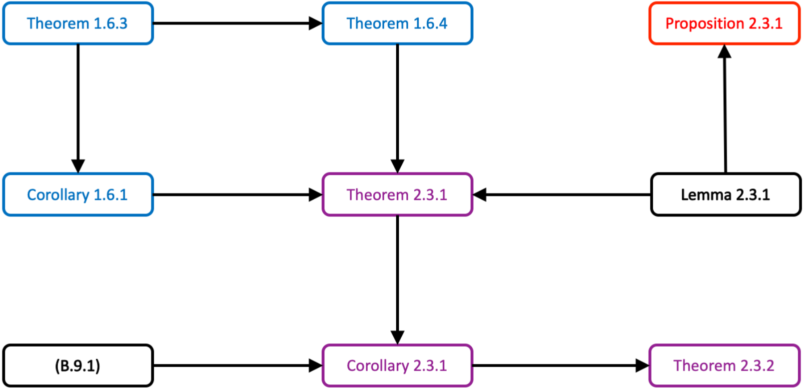
Office Hour: Wednesday 09:30AM - 11:30AM

October 1 & 2, 2020

Overview

1. Review
   - Preliminaries
   - Exponential Families
   - Maximum Likelihood
2. Problems

Overview


Theorem 1.6.3
Theorem 1.6.4
Proposition 2.3.1
Corollary 1.6.1
Theorem 2.3.1
Lemma 2.3.1
(B.9.1)
Corollary 2.3.1
Theorem 2.3.2

Preliminaries

(B.9.1) (page 518 in [B&D, 2015])

The point $\boldsymbol{x}_0$ belongs to the interior S^0 of the convex set S iff for every $\boldsymbol{d} \neq \boldsymbol{0}$

$$\left\{\boldsymbol{x} \,:\, \boldsymbol{d}^\top \boldsymbol{x} > \boldsymbol{d}^\top \boldsymbol{x}_0\right\} \cap S^0 \neq \emptyset,$$
$$\left\{\boldsymbol{x} \,:\, \boldsymbol{d}^\top \boldsymbol{x} < \boldsymbol{d}^\top \boldsymbol{x}_0\right\} \cap S^0 \neq \emptyset.$$

Exponential Families

k-Parameter Exponential Families

The family of distributions of a model $\{P_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k\}$ is said to be a k-parameter exponential family if there exist real-valued functions $\eta_1(\boldsymbol{\theta}), \cdots, \eta_k(\boldsymbol{\theta}), B(\boldsymbol{\theta})$ defined for $\boldsymbol{\theta} \in \Theta$, real-valued functions $T_1(x), \cdots, T_k(x), h(x)$ defined for $x \in \mathcal{X}$ such that the density (frequency) functions $p(x, \boldsymbol{\theta})$ of $P_{\boldsymbol{\theta}}$ can be written as

$$p(x, \boldsymbol{\theta}) = h(x) \exp\left[\boldsymbol{\eta}^\top(\boldsymbol{\theta}) \boldsymbol{T}(x) - B(\boldsymbol{\theta})\right], \tag{1.6.10}$$

where

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = \begin{bmatrix} \eta_1(\boldsymbol{\theta}) & \cdots & \eta_k(\boldsymbol{\theta}) \end{bmatrix}^\top,$$
$$\boldsymbol{T}(x) = \begin{bmatrix} T_1(x) & \cdots & T_k(x) \end{bmatrix}^\top.$$

Canonical k-Parameter Exponential Families

A useful reparametrization of the k-parameter exponential family by letting the model be indexed by $\boldsymbol{\eta}$ rather than $\boldsymbol{\theta}$ has the form as

$$q(x, \boldsymbol{\eta}) = h(x) \exp\left[\boldsymbol{\eta}^\top \boldsymbol{T}(x) - A(\boldsymbol{\eta})\right],$$

where

$$A(\boldsymbol{\eta}) = \begin{cases} \log \int h(x) \exp\left[\boldsymbol{\eta}^\top \boldsymbol{T}(x)\right] dx & \text{continuous case} \\ \log \sum h(x) \exp\left[\boldsymbol{\eta}^\top \boldsymbol{T}(x)\right] & \text{discrete case} \end{cases}.$$

Note that $x \in \mathcal{X}$ and $\boldsymbol{\eta} \in \mathcal{E}$ where $\mathcal{E}$ is the collection of all $\boldsymbol{\eta}$ such that $A(\boldsymbol{\eta})$ is finite.

Theorem 1.6.3 (page 59 in [B&D, 2015])

Let $\mathcal{P}$ be a canonical k-parameter exponential family generated by $(\boldsymbol{T}, h)$ with corresponding natural parameter space $\mathcal{E}$ and function $A(\boldsymbol{\eta})$. Then

(a) $\mathcal{E}$ is convex.

(b) $A : \mathcal{E} \to \mathbb{R}$ is convex.

(c) If $\mathcal{E}$ has nonempty interior $\mathcal{E}^0$ in $\mathbb{R}^k$ and $\boldsymbol{\eta}_0 \in \mathcal{E}^0$, then $\boldsymbol{T}(X)$ has under $\boldsymbol{\eta}_0$ a moment-generating function M given by

$$M(\boldsymbol{s}) = \exp\left[A(\boldsymbol{\eta}_0 + \boldsymbol{s}) - A(\boldsymbol{\eta}_0)\right]$$

valid for all $\boldsymbol{s}$ such that $\boldsymbol{\eta}_0 + \boldsymbol{s} \in \mathcal{E}$. Since $\boldsymbol{\eta}_0$ is an interior point this set of $\boldsymbol{s}$ includes a ball about $\mathbf{0}$.

Corollary 1.6.1 (page 59 in [B&D, 2015])

Under the conditions of **Theorem 1.6.3**

$$\begin{aligned} E_{\boldsymbol{\eta}_0}\left[\boldsymbol{T}(X)\right] &= \dot{A}\left(\boldsymbol{\eta}_0\right), \\ \mathrm{Var}_{\boldsymbol{\eta}_0}\left[\boldsymbol{T}(X)\right] &= \ddot{A}\left(\boldsymbol{\eta}_0\right), \end{aligned}$$

where

$$\dot{A}\left(\boldsymbol{\eta}_0\right) = \begin{bmatrix} \frac{\partial A}{\partial \eta_1}\left(\boldsymbol{\eta}_0\right) \\ \vdots \\ \frac{\partial A}{\partial \eta_k}\left(\boldsymbol{\eta}_0\right) \end{bmatrix}, \qquad \ddot{A}\left(\boldsymbol{\eta}_0\right) = \begin{bmatrix} \frac{\partial^2 A}{\partial \eta_1 \partial \eta_1}\left(\boldsymbol{\eta}_0\right) & \cdots & \frac{\partial^2 A}{\partial \eta_1 \partial \eta_k}\left(\boldsymbol{\eta}_0\right) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 A}{\partial \eta_k \partial \eta_1}\left(\boldsymbol{\eta}_0\right) & \cdots & \frac{\partial^2 A}{\partial \eta_k \partial \eta_k}\left(\boldsymbol{\eta}_0\right) \end{bmatrix}.$$

Theorem 1.6.4 (page 60-61 in [B&D, 2015])

Suppose $\mathcal{P} = \{q(x, \boldsymbol{\eta}) \mid \boldsymbol{\eta} \in \mathcal{E}\}$ is a canonical exponential family generated by $(\boldsymbol{T}_{k \times 1}, h)$ with natural parameter space $\mathcal{E}$ such that $\mathcal{E}$ is open. Then the following are equivalent.

(i) $\mathcal{P}$ is of rank k.

(ii) $\boldsymbol{\eta}$ is a parameter (identifiable).

(iii) $\mathrm{Var}_{\boldsymbol{\eta}}[\boldsymbol{T}(X)]$ is positive definite.

(iv) $\boldsymbol{\eta} \to \dot{A}(\boldsymbol{\eta})$ is 1-1 on $\mathcal{E}$.

(v) A is strictly convex on $\mathcal{E}$.

Lemma 2.3.1 (page 121 in [B&D, 2015])

Suppose we are given a function $\ell : \Theta \to \mathbb{R}$ where $\Theta \subset \mathbb{R}^p$ is open and ℓ is continuous. Suppose also that

$$\lim \{\ell(\boldsymbol{\theta}) \; : \; \boldsymbol{\theta} \to \partial\Theta\} = -\infty. \tag{2.3.1}$$

Then there exists $\widehat{\boldsymbol{\theta}} \in \Theta$ such that

$$\ell\left(\widehat{\boldsymbol{\theta}}\right) = \max\{\ell(\boldsymbol{\theta}) \; : \; \boldsymbol{\theta} \in \Theta\}$$

Maximum Likelihood

Proposition 2.3.1 (page 122 in [B&D, 2015])

Suppose $X \sim \{P_{\boldsymbol{\theta}} \ : \ \boldsymbol{\theta} \in \Theta\}$, $\Theta \subset \mathbb{R}^p$ open, with corresponding densities $p\left(x, \boldsymbol{\theta}\right)$. If further $\ell_x\left(\boldsymbol{\theta}\right) \equiv \log p\left(x, \boldsymbol{\theta}\right)$ is strictly concave and $\ell_x\left(\boldsymbol{\theta}\right) \to -\infty$ as $\boldsymbol{\theta} \to \partial\Theta$, then the MLE $\widehat{\boldsymbol{\theta}}\left(x\right)$ exists and is unique.

Maximum Likelihood

Theorem 2.3.1 (page 122 in [B&D, 2015])

Suppose $\mathcal{P}$ is the canonical exponential family generated by $(\boldsymbol{T}, h)$ and

- (i) The natural parameter space, $\mathcal{E}$, is open.
- (ii) The family is of rank k.

Let x be the observed data vector and set $\boldsymbol{t}_0 = \boldsymbol{T}(x)$.

- (a) If $\boldsymbol{t}_0 \in \mathbb{R}^k$ satisfies

$$P\left[\boldsymbol{c}^\top \boldsymbol{T}(X) > \boldsymbol{c}^\top \boldsymbol{t}_0\right] > 0 \qquad \forall \boldsymbol{c} \neq \boldsymbol{0}, \tag{2.3.2}$$

then the MLE $\widehat{\boldsymbol{\eta}}$ exists, is unique, and is a solution to

$$\dot{A}(\boldsymbol{\eta}) = E_{\boldsymbol{\eta}}[\boldsymbol{T}(X)] = \boldsymbol{t}_0. \tag{2.3.3}$$

- (b) Conversely, if $\boldsymbol{t}_0$ doesn't satisfy (2.3.2), then the MLE doesn't exist and (2.3.3) has no solution.

Maximum Likelihood

Corollary 2.3.1 (page 122 in [B&D, 2015])

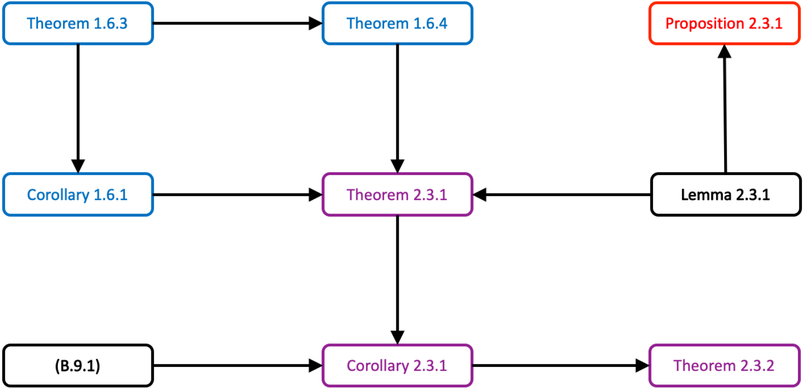
Suppose the conditions of **Theorem 2.3.1** hold. If $\mathrm{C}_{\boldsymbol{T}}$ is the convex support of the distribution of $\boldsymbol{T}(\mathrm{X})$, then $\widehat{\boldsymbol{\eta}}$ exists and is unique iff $\boldsymbol{t}_0 \in \mathrm{C}_{\boldsymbol{T}}^0$ where $\mathrm{C}_{\boldsymbol{T}}^0$ is the interior of $\mathrm{C}_{\boldsymbol{T}}$.

Maximum Likelihood

Theorem 2.3.2 (page 123 in [B&D, 2015])

Suppose the conditions of **Theorem 2.3.1** hold and $\boldsymbol{T}_{k \times 1}$ has a continuous case density on $\mathbb{R}^k$. Then the MLE $\widehat{\boldsymbol{\eta}}$ exists with probability 1 and necessarily satisfies (2.3.3).

Summary



Theorem 1.6.3
Theorem 1.6.4
Proposition 2.3.1
Corollary 1.6.1
Theorem 2.3.1
Lemma 2.3.1
(B.9.1)
Corollary 2.3.1
Theorem 2.3.2

Problems - HW4

- PMF and PDF

$$\sum_{x \in \mathcal{X}} p(x) = 1 \qquad \text{and} \qquad \int_{x \in \mathcal{X}} f(x)dx = 1$$

For example, consier $\mathcal{N}\left(\mu, \sigma^2\right)$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1,$$
$$\int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \sqrt{2\pi}\sigma.$$

Another example, k-parameter exponential families in canonical form.

Problems - HW4

- Combination

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & k \leqslant n \\ 0 & k > n \end{cases},$$

$$k\binom{n}{k} = n\binom{n-1}{k-1},$$

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

By convention, $0!$ is defined to be 1.

References

Bickel, Peter J., and Kjell A. Doksum. (2015) Mathematical statistics: basic ideas and selected topics, volume I. CRC Press.

Thanks for listening!