Section for Statistical Theory

TA: Cong Mu

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

September 10 & 11, 2020

Overview

1. Review
   - Preliminaries
   - Sufficiency
   - Completeness
   - Exponential Families
2. Problems

Preliminaries

Regular Models

Models $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ will be called regular parametric models if either one of the followings hold

1. All of the P_θ are continuous with densities $p(\boldsymbol{x}, \theta)$.
2. All of the P_θ are discrete with frequency functions $p(\boldsymbol{x}, \theta)$, and the set $\{\boldsymbol{x} \mid p(\boldsymbol{x}, \theta) > 0\}$ is the same for all $\theta \in \Theta$.

Preliminaries

Bias–variance tradeoff

1. Low bias with high variance.
2. High bias with low variance.

MSE

Let $\widehat{\theta}$ be an estimator of a parameter θ, then

$$\mathrm{MSE}\left(\widehat{\theta}\right) = \left(\mathrm{Bias}\left[\widehat{\theta}\right]\right)^2 + \mathrm{Var}\left[\widehat{\theta}\right],$$

where

$$\mathrm{Bias}\left[\widehat{\theta}\right] = \mathrm{E}\left[\widehat{\theta}\right] - \theta.$$

Sufficiency

Sufficient Statistic

A statistic $T(\boldsymbol{X})$ is called sufficient for $P_\theta \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ if the condition distribution of $\boldsymbol{X}$ given $T(\boldsymbol{X}) = t$ does not involve θ.

Factorization Theorem for Sufficient Statistics

In a regular model, a statistic $T(\boldsymbol{X})$ with range $\mathcal{T}$ is sufficient for θ **if and only if** there exists a function $g(t, \theta)$ defined for $t \in \mathcal{T}$ and $\theta \in \Theta$ and a function $h(\boldsymbol{x})$ defined for $\boldsymbol{x} \in \mathcal{X}$ such that for all $\boldsymbol{x} \in \mathcal{X}$ and $\theta \in \Theta$

$$p(\boldsymbol{x}, \theta) = g(T(\boldsymbol{x}), \theta) h(\boldsymbol{x}).$$

Minimal Sufficient Statistic

A statistic $\mathsf{T}(\boldsymbol{X})$ is called minimally sufficient if it is sufficient and provides a greater reduction of the data than any other sufficient statistic $\mathsf{S}(\boldsymbol{X})$, i.e., we can find a transformation r such that $\mathsf{T}(\boldsymbol{X}) = \mathrm{r}(\mathsf{S}(\boldsymbol{X}))$.

Characterization for Minimal Sufficient Statistics

A statistic $\mathsf{T}(\boldsymbol{X})$ is minimally sufficient **if and only if**

$$\frac{\mathrm{p}(\boldsymbol{x}, \theta)}{\mathrm{p}(\boldsymbol{y}, \theta)} \text{ does not involve } \theta \qquad \Longleftrightarrow \qquad \mathsf{T}(\boldsymbol{x}) = \mathsf{T}(\boldsymbol{y}).$$

Completeness

Complete Statistic

A statistic $T(\boldsymbol{X})$ is called complete for $P_\theta \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ if for every measurable function g,

$$E_\theta[g(T)] = 0,\ \forall \theta \qquad \Longrightarrow \qquad P_\theta[g(T) = 0] = 1,\ \forall \theta.$$

Sufficiency and Completeness

Rao-Blackwell Theorem

If $g(X)$ is an estimator of parameter θ, then the Rao–Blackwell estimator $g_*(X) = \mathrm{E}[g(X) | T(X)]$ where $T(X)$ is a sufficient statistic is typically a better estimator of θ. With respect to MSE, that is

$$\mathrm{E}\left[(g_*(X) - \theta)^2\right] \leqslant \mathrm{E}\left[(g(X) - \theta)^2\right].$$

Lehmann–Scheffé Theorem

If an unbiased estimator depends on the data only through a complete and sufficient statistic for some parameter θ, then it is the minimum-variance unbiased estimator (MVUE) of θ.

Exponential Families

One-Parameter Exponential Families

The family of distributions of a model $\{P_\theta \mid \theta \in \Theta\}$ is said to be a one-parameter exponential family if there exist real-valued functions $\eta(\theta), B(\theta)$ defined for $\theta \in \Theta$, real-valued functions $T(x), h(x)$ defined for $x \in \mathcal{X}$ such that the density (frequency) functions $p(x, \theta)$ of P_θ can be written as

$$p(x, \theta) = h(x) \exp\left[\eta(\theta) T(x) - B(\theta)\right].$$

Note that the functions $\eta(\theta), B(\theta)$ and $T(x)$ are not unique. By **Factorization Theorem for Sufficient Statistics**, $T(X)$ is sufficient for θ.

Canonical One-Parameter Exponential Families

A useful reparametrization of the one-parameter exponential family by letting the model be indexed by η rather than θ has the form as

$$q(x, \eta) = h(x) \exp\left[\eta T(x) - A(\eta)\right],$$

where

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

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

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

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

where

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

Exponential Families

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

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

where

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

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

Normal Distribution

- **Multivariate Normal:** $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right].$$

- **Bivariate Normal:** $(X, Y) \sim \mathcal{N}\left(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho\right)$

$$f_{X,Y}(x,y) = \frac{\exp(z)}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}},$$

where

$$z = -\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right].$$

Bivariate Normal

Marginal and Conditional Distribution

The marginal and conditional distributions of bivariate normal are normal. That is, if $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Then $X \sim \mathcal{N}\left(\mu_X, \sigma_X^2\right), Y \sim \mathcal{N}\left(\mu_Y, \sigma_Y^2\right)$ and $X \mid Y = y \sim \mathcal{N}\left(\mu_{X|Y}, \sigma_{X|Y}^2\right)$ where

$$\mu_{X|Y} = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \qquad \sigma_{X|Y}^2 = \left(1 - \rho^2\right)\sigma_X^2$$

Problems

Problem 1.1.1(d), page 67 in [B&D, 2015]

The number of eggs laid by an insect follows a Poisson distribution with unknown mean λ. Once laid, each egg has an unknown chance p of hatching and the hatching of one egg is independent of the hatching of the others. An entomologist studies a set of n such insects observing both the number of eggs laid and the number of eggs hatching for each nest.

Problems

Problem 1.3.11(a), page 78 in [B&D, 2015]

A decision rule δ is said to be unbiased if for all $\theta, \theta' \in \Theta$, we have

$$E_\theta \left[\ell \left(\theta, \delta \left(\boldsymbol{X} \right) \right) \right] \leqslant E_\theta \left[\ell \left(\theta', \delta \left(\boldsymbol{X} \right) \right) \right].$$

Show that if θ is real and $\ell \left(\theta, a \right) = \left(\theta - a \right)^2$, then this definition coincides with the definition of an unbiased estimate of θ.

Problems

Problem 1.5.1, page 84 in [B&D, 2015]

Let $X_1, \cdots, X_n \sim \text{Poisson}(\theta)$ iid where $\theta > 0$.

1. Show directly that $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ is sufficient for θ.
2. Establish the same result using the factorization theorem.

Problems

Problem 1.5.12, page 86 in [B&D, 2015]

Let $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ where P_θ is discrete concentrated on $\mathcal{X} = \{x_1, x_2, \cdots\}$. Let

$$p(x, \theta) = P_\theta[X = x] = L_x(\theta) > 0 \ \text{ on } \ \mathcal{X}.$$

Show that $\frac{L_X(\cdot)}{L_X(\theta_0)}$ is minimal sufficient.

References

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

Thanks for listening!