Parametric Statistics

Fisher Information

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

December 11, 2023

Lecture Summary

8.8 Fisher Information

Point Estimation Summary

- Bayesian Approach: Treat θ as an R.V.
  - Find the posterior probability of θ: $f(\theta|x_1, \ldots, x_n)$.
  - Pick the estimator that minimizes some loss function.
  - You can compute $P(a \leq \theta \leq b)$.
- Frequentist Approach: θ is an unknown number.
  - Find the likelihood function of the data for each value of θ: $f(x_1, \ldots, x_n|\theta)$.
  - Pick the estimator that maximizes the likelihood of the data.
  - Use the sampling distribution of the estimator to compute confidence intervals $P(A \leq \theta \leq B)$.

Properties of MLE estimators

Consistency, Asymptotic Normality

Let $\{f(x \mid \theta) : \theta \in \Omega\}$ be a parametric model, where $\theta \in \mathbb{R}$ is a single parameter. Let $X_1, \ldots, X_n \overset{IID}{\sim} f(x \mid \theta_0)$ for $\theta_0 \in \Omega$, and let $\hat{\theta}$ be the MLE based on $X_1, \ldots, X_n$. Under certain regularity conditions, $\hat{\theta}$ is consistent and asymptotically normal, with

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta_0)}\right)$$

$I(\theta)$ is defined by the two equivalent expressions

$$I(\theta) := \mathrm{Var}_\theta[z(X, \theta)] = -\mathbb{E}_\theta\left[z'(X, \theta)\right],$$

where Var_θ and $\mathbb{E}_\theta$ denote variance and expectation with respect to $X \sim f(x \mid \theta)$, and

$$z(x, \theta) = \frac{\partial}{\partial \theta} \log f(x \mid \theta), \quad z'(x, \theta) = \frac{\partial^2}{\partial \theta^2} \log f(x \mid \theta).$$

Conditions

- All PDFs/PMFs $f(x \mid \theta)$ in the model have the same support,
- θ_0 is an interior point (i.e., not on the boundary) of Ω,
- The log-likelihood $l(\theta)$ is differentiable in θ, and
- $\hat{\theta}$ is the unique value of $\theta \in \Omega$ that solves the equation $0 = l'(\theta)$.

Fisher Information

$z(x, \theta)$ is called the score function, and $I(\theta)$ is called the Fisher information.

$$I(\theta) := \text{Var}_\theta[z(X, \theta)] = -\mathbb{E}_\theta\left[z'(X, \theta)\right],$$

Fisher information provides a way to measure the amount of information that a random variable contains about some parameter θ of the random variable's assumed probability distribution.

- Find the Fisher Information for the mean of the Normal Distribution with known mean.
- Find the Fisher Information for the parameter od the Bernoulli distribution.

Properties of Estimators

Unbiased Estimators

An estimator is unbiased if

$$\text{Bias}(\hat{\theta}, \theta) = E(\hat{\theta}) - \theta = 0.$$

Mean Squared Error of an Estimator

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right]$$

Properties of Estimators

Unbiased Estimators

An estimator is unbiased if

$$\text{Bias}(\hat{\theta}, \theta) = E(\hat{\theta}) - \theta = 0.$$

Mean Squared Error of an Estimator

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= \mathbb{E}\left[(\hat{\theta} - \theta)^2\right] \\ &= \text{Var}(\hat{\theta} - \theta) + (\mathbb{E}[\hat{\theta} - \theta])^2 \\ &= \text{Var}(\hat{\theta}) + \text{Bias}^2(\hat{\theta}, \theta) \end{aligned}$$

Cramer-Rao Lower Bound

Consider a parametric model $\{f(x \mid \theta) : \theta \in \Omega\}$ (satisfying certain mild regularity assumptions) where $\theta \in \mathbb{R}$ is a single parameter. Let T be any unbiased estimator of θ based on data $X_1, \ldots, X_n \overset{IID}{\sim} f(x \mid \theta)$. Then

$$\mathrm{Var}_\theta[T] \geq \frac{1}{nI(\theta)}$$

Efficient Estimator

An unbiased estimator T is an efficient estimator of its expectation θ if $\mathrm{Var}_\theta[T] = \frac{1}{nI(\theta)}$ for every value of $\theta \in \Omega$.

Cramer-Rao Lower Bound

Consider a parametric model $\{f(x \mid \theta) : \theta \in \Omega\}$ (satisfying certain mild regularity assumptions) where $\theta \in \mathbb{R}$ is a single parameter. Let T be any unbiased estimator of θ based on data $X_1, \ldots, X_n \overset{IID}{\sim} f(x \mid \theta)$. Then

$$\mathrm{Var}_\theta[T] \geq \frac{1}{nI(\theta)}$$

Efficient Estimator

An unbiased estimator T is an efficient estimator of its expectation θ if $\mathrm{Var}_\theta[T] = \frac{1}{nI(\theta)}$ for every value of $\theta \in \Omega$.

MLE estimators are asymptotically efficient.

Properties of the MLE estimators

Invariance

If $\hat{\theta}$ is the maximum likelihood estimator of θ, then $g(\hat{\theta})$ is the maximum likelihood estimator of $g(\theta)$.

The proof is very easy if g is a one-to-one function, more complicated otherwise.

- Example: Variance of the Bernoulli distribution: $p(1-p)$.
- Example: Odds for the Bernoulli distribution: $\frac{p}{1-p}$.

The delta method

If a function $g : \mathbb{R} \to \mathbb{R}$ is differentiable at θ_0 with $g'(\theta_0) \neq 0$, and if

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, v\left(\theta_0\right)\right)$$

for some variance $v(\theta_0)$, then

$$\sqrt{n}\left(g(\hat{\theta}) - g\left(\theta_0\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \left(g'\left(\theta_0\right)\right)^2 v\left(\theta_0\right)\right)$$