STA732
Statistical Inference

Lecture 09: Bayesian estimation

Yuansi Chen

Spring 2023

Duke University

https://www2.stat.duke.edu/courses/Spring23/sta732.01/

Recap from Lecture 08

1. Construct minimum risk equivariant (MRE) estimator via conditioning on maximal invariant statistics
2. Pitman estimator of location
3. MRE for location is unbiased under squared error loss
4. MRE usually admissible

Where we are

- We have finished the first approach of arguing for “the best” estimator in point estimation: by restricting to a small set of estimatiors
    - Unbiased estimators
    - Equivariant estimators
- We begin the second approach: global measure of optimality
    - average risk
    - minimax risk

Goal of Lecture 09

1. Bayes risk, Bayes estimator
2. Examples
3. Bayes estimators are usually biased
4. Bayes estimators are usually admissible

Chap. 7 in Keener or Chap. 4 in Lehmann and Casella

Bayes risk, Bayes estimator

Recall the components of a decision problem

- Data X
- Model family $\mathscr{P} = \{P_\theta : \theta \in \Omega\}$, a collection of probability distributions on the sample space
- Loss function L, $L(\theta, d)$ measures the loss incurred by the decision d when compared with the parameter obtained from θ
- Risk function R, $R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta)]$

The frequentist motivation of the Bayesian setup

Motivation

It is in general hard to find uniformly minimum risk estimator. Oftentimes, we have risks that cross. This difficulty will not arise if the performance is measured via a single number.

Def. Bayes risk

The Bayes risk is the average-case risk, integrated w.r.t. some measure Λ, called prior.

The frequentist motivation of the Bayesian setup

Motivation

It is in general hard to find uniformly minimum risk estimator. Oftentimes, we have risks that cross. This difficulty will not arise if the performance is measured via a single number.

Def. Bayes risk

The Bayes risk is the average-case risk, integrated w.r.t. some measure Λ, called prior.

Remark

For now, assume $\Lambda(\Omega) = 1$ (Λ is a prob measure). Later we might deal with improper prior.

Bayes risk

$$R_{\text{Bayes}}(\Lambda, \delta) = \int_{\Omega} R(\theta, \delta) d\Lambda(\theta)$$
$$= \mathbb{E}R(\Theta, \delta)$$

where Θ is the randoma variable with distribution Λ.

$$\mathbb{E}R(\Theta, \delta) = \mathbb{E}[\mathbb{E}[L(\Theta, \delta(X)) \mid X]]$$

Both X and Θ are considered random.

The frequentist understanding: average risk makes sense without believing the parameter is random

Bayes estimator

An estimator δ which minimizes the average risk $R_{\text{Bayes}}(\Lambda, \cdot)$ is a Bayes estimator.

Construct Bayes estimator

Thm 7.1 in Keener

Suppose $\Theta \sim \Lambda$, $X \mid \Theta = \theta \sim P_\theta$, and $L(\theta, d) \geq 0$ for all $\theta \in \Omega$ and all d. If

- $\mathbb{E}[L(\Theta, \delta_0)] < \infty$ for some δ_0
- for a.e. x, there exists a $\delta_\Lambda(x)$ minimizing

$$\mathbb{E}[L(\Theta, d) \mid X = x]$$

  with respect to d

Then δ_Λ is a Bayes estimator.

In words: the Bayes estimator can be found by minimizing the conditional distribution $\mathbb{E}[L(\theta, d) \mid X = x]$, one x at a time

proof of Thm 7.1

Posterior

Def. Posterior

The conditional distribution of Θ given X, written as $\mathcal{L}(\Theta \mid X)$ is called the posterior distribution

Remark

- Λ is usually interpreted as prior belief about Θ before seeing the data
- $\mathcal{L}(\Theta \mid X)$ is the belief after seeing the data

Posterior calcultation with density

Suppose prior density $\lambda(\theta)$, likelihood $p_\theta(x)$, then the posterior density is

$$\lambda(\theta \mid x) = \frac{\lambda(\theta) p_\theta(x)}{q(x)}$$

where $q(x) = \int_\Omega \lambda(\theta) p_\theta(x) d\theta$ is the marginal density of X.

Then the Bayes estimator has the form

$$\delta_\Lambda(x) = \arg\min_d \int_\Omega L(\theta, d) \lambda(\theta \mid x) d\theta$$

Posterior mean is Bayes estimator for squared error loss

Suppose $L(\theta, d) = (g(\theta) - d)^2$ then the Bayes estimator is the posterior mean

proof:

Examples

Binomial model with Beta prior

Suppose $X \mid \Theta = \theta \sim \text{Binomial}(n, \theta)$ with density $\theta^x (1-\theta)^{n-x} \binom{n}{x}$, $\Theta \sim \text{Beta}(\alpha, \beta)$ with density $\theta^{\alpha-1} (1-\theta)^{\beta-1} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Find the Bayes estimator under squared error loss.

Weighted squared error loss

Suppose $L(\theta, d) = w(\theta)\left(g(\theta) - d\right)^2$. Find a Bayes estimator.

Normal mean estimation

$X \mid \Theta = \theta \sim \mathcal{N}(\theta, \sigma^2),$
$\Theta \sim \mathcal{N}(\mu, \tau^2).$

Find the Bayes estimator of mean under squared error loss

What if we have n i.i.d. data points?

Binary classification

Suppose the parameter space $\Omega = \{0, 1\}$.

$\mathbb{P}(X = x \mid \Theta = 0) = f_0(x)$ and $\mathbb{P}(X = x \mid \Theta = 1) = f_1(x)$. The prior is $\pi(1) = p, \pi(0) = 1 - p$.

Determine a Bayes estimator under 0-1 loss $L(\theta, d) = \begin{cases} 0 & d = \theta \\ 1 & d \neq \theta \end{cases}$

Bayes estimators are usually biased

Unbiased estimator under squared error loss is not Bayes

Thm Lehmann Casella 4.2.3

If δ is unbiased for $g(\theta)$ with $R_{\text{Bayes}}(\Lambda, \delta) < \infty$ then δ is not Bayes under squared error loss unless its average risk is zero

$$\mathbb{E}\left[(\delta(X) - g(\Theta))^2\right] = 0$$

proof:

Bayes estimators are usually admissible

Uniqueness of Bayes estimator under strictly convex loss

Thm. Lehmann Casella 4.1.4

Let Q be the marginal distribution of X, i.e., $Q(E) = \int \mathbb{P}_\theta(E) d\Lambda(\theta)$. Suppose L is strictly convex. If

1. $R_{\text{Bayes}}(\Lambda, \delta_\Lambda) < \infty$,
2. $Q(E) = 0$ implies $P_\theta(E) = 0, \forall\theta$,

then the Bayes estimator δ_Λ is unique (a.e. with respect to P_θ for all θ).

proof: Use the following lemma

Lem. Lehmann Casella exercise 1.7.26

Let ϕ be a strictly convex function over an interval I. If there exists a value $a_0 \in I$ minimizing ϕ, then a_0 is unique.

A unique Bayes estimator is admissible

Thm. Lehmann Casella 5.2.4

A unique Bayes estimator (a.s. for all P_θ) is admissible.

proof:

Summary

- Bayes estimator is defined as the minimizer of the average risk over a prior on θ.
- Bayes estimator can be constructed by conditioning the risk on each x
- Bayes estimators are biased under squared error loss
- Bayes estimators are admissible under strictly convex loss

What is next?

- Where do priors come from?
- Pros and cons of Bayes

Thank you