Mathematical Statistics II

The Bayesian Approach to Parameter Estimation

Jesse Wheeler

Introduction

Bayesian Estimation

- Much of this work is based on Rice (2007, Section 8.6).
- We have already discussed the philosophy of Bayesian statistics.
- We start with a prior belief about parameter values, and update these beliefs using observed data.
- The resulting distribution is called the *posterior*, and it represents our updated belief after observing data.
- This is very natural idea that is closely related to the idea of likelihood: likelihood quantifies some degree of belief about a parameter value.

Review

Some Review

- Before we begin, we will first do a bit of review.
- In the context of Bayesian inference, we treat unknown parameter vectors as random variables, which I will denote Θ.
- Thus, our probability model can be expressed as $f(x|\Theta = \theta)$, which we often shorten to $f(x|\theta)$.

Some Review II

Bayes' Theorem

Let X be the random vector representing observed data, and Θ the random parameter vector, and x^* the observed data. Bayes Theorem states:

$$\pi_{\Theta|X}(\theta|x^*) = \frac{f_{X|\Theta}(x^*|\theta)\pi_\Theta(\theta)}{f_X(x^*)}$$
$$= \frac{f_{X|\theta}(x^*|\theta)\pi_\Theta(\theta)}{\int f_{X|\Theta}(x^*|\tau)\pi_\Theta(\tau)\, d\tau}$$

- As before, f is taken to be either a pmf or pdf, depending on the problem.

Some Review III

Flipping 10 coins

Our friend hands us a coin from another country, and we want to estimate $\theta = p$, the probability that the coin lands heads.

Suppose we flip a coin 10 times, and see n heads. Find a Bayesian estimate for θ.

Some Review IV

- Even in the simple problem above, we see two of the primary challenges with Bayesian parameter estimation:
  - How do we choose the prior distribution $\pi(\theta)$? A generally safe and accepted approach is a uniform prior. However, this formally only exists if θ is bounded, which is not always the case. Also, it represents a prior belief: given a new coin, do we really think all values of p are equally likely, or maybe values close to $p = 0.5$ are more likely than extreme values $p = 0, 1$? Since the prior represents our beliefs about θ, is a uniform prior actually appropriate? If it isn't appropriate, how exactly should we specify the prior?
  - Even in this very simple model and prior, the denominator $f(x)$ was difficult to compute. What about more complex models and priors? A large amount of Bayesian computation and theory is dedicated to solving this problem.

Some Review V

Proposition: the MAP and MLE

Let θ be a parameter of interest, and x^* the observed data. If our prior distribution is proportional to 1, i.e., $\pi(\theta) \propto 1$ (which is effectively a uniform prior on a bounded interval), then

$$\hat{\theta}_{\mathsf{MAP}} = \hat{\theta}_{\mathsf{MLE}}.$$

- This is true for the Coin-tossing example; look back at the likelihood function and posterior, and use R to plot them both.

Examples

Bayesian point-estimate examples

Poisson model

Suppose we have observations n observations, which we wish to model as IID Poisson(λ). Find a Bayesian estimate of $\Lambda = \lambda$ given the observed data x^*.

Bayesian point-estimate examples II

Poisson posterior, uniform prior

Revisit the Poisson(λ) model, while taking the alternative approach of using a uniform prior.

Real-data example: Poisson Distribution

- Now let's look at a real-data example. These data are the 23 observations from the asbestos-filter problem.

```
x <- c(
  31, 29, 19, 18, 31, 28, 34, 27, 34, 30, 16, 18,
  26, 27, 27, 18, 24, 22, 28, 24, 21, 17, 24
)
x
```

```
 [1] 31 29 19 18 31 28 34 27 34 30 16 18 26 27 27 18 24 22
[19] 28 24 21 17 24
```

- TODO: Compare Bayes 1 to Bayes 2 to MLE

References and Acknowledgements

Rice JA (2007). *Mathematical statistics and data analysis*, volume 371. 3 edition. Thomson/Brooks/Cole Belmont, CA.

- Compiled on January 28, 2026 using R version 4.5.2.
-
- We acknowledge students and instructors for previous versions of this course / slides.