

# **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  $\Theta$ .
- 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  $\Theta$  the random parameter vector, and  $x^*$  the observed data. Bayes Theorem states:

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

- 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  $\theta$ .

## 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  $\theta$  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  $\theta$ , 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  $\theta$  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

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

- 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  $\text{Poisson}(\lambda)$ . Find a Bayesian estimate of  $\Lambda = \lambda$  given the observed data  $x^*$ .

## Bayesian point-estimate examples II

### Poisson posterior, uniform prior

Revisit the Poisson( $\lambda$ ) 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
```

## Real-data example: Poisson Distribution II

### Comparing Estimates

Using the data above, compare estimates using the MoM, MLE, and the two Bayesian approaches, as well as the corresponding errors related to these estimates.

## Posteriors and Likelihood

- In the problem above, we saw that we get very similar estimates using MLE or Bayesian approaches, regardless of which prior we picked.
- We can argue why this will often be the case, especially for IID data.
- Previously, we saw:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

- When  $n$  gets large, the likelihood dominates in this equation.  
In the IID case:

$$\text{likelihood} = \prod_{i=1}^n f(x_i^* | \theta).$$

## Posteriors and Likelihood II

- In particular, each new data point scales the likelihood larger and larger, to the point where the prior has little impact on the posterior distribution.
- See the accompanying Lecture 4 R code for a visual demonstration of this using the Poisson distribution.

# **Introduction to Numeric Integration**

---

## Numeric Integration

- As we saw in the previous example, one of the primary challenges of Bayesian estimation is the integration in the denominator of the posterior.
- Bayesian statistics has really exploded since the late 20th century, largely thanks to improved computational tools that help with the numeric integration.
- For this set of lectures, we only briefly introduce this topic. Depending on time and interest, we can explore this topic more later in the semester.

## **Choice of Priors**

---

## Conjugate priors

- The Poisson( $\lambda$ ) example showed two main approaches:
  - The traditional (subjective) Bayesian, who takes seriously the choice of prior, and chose a Gamma density to aid computations.
  - The utilitarian (objective) Bayesian, who picked an uninformative prior.
- The former approach was aided by what is known as a **conjugate prior**.

## Conjugate priors II

### Definition: Conjugate priors

Suppose the prior distribution belongs to a family of distributions,  $G$ , and the data come from a family of distributions  $H$ .

$G$  is said to be conjugate to  $H$  if the posterior is in the family  $G$ .

- Example: If the data-model is  $\text{Poisson}(\lambda)$ , then the family  $H$  is the family of Poisson distributions. The Gamma family ( $G$ ) of distributions is conjugate to the Poisson family, because if Gamma is selected as the prior distribution, then the posterior distribution (under data model  $H$ ) is still Gamma ( $G$ ), with updated parameters.

## Conjugate priors III

- Much of the Bayesian statistics of the 20th century relied on conjugate priors to help with integration, or were confined to models with very few parameters.
- Recent developments in computing, both hardware, software, and theory of Bayesian computing, has enabled fitting much more complex models using arbitrary priors.
- Still, it's worth discussing conjugate priors, and we will provide a few examples.

## Jeffrey's priors

- TODO

# Hierarchical Bayes

---

# Hierarchical Bayes

# **Empirical Bayes**

---

# Empirical Bayes

TODO

# **Uncertainty quantification**

---

## Uncertainty in Bayes estimates

- TODO

## References and Acknowledgements

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

- Compiled on January 29, 2026 using R version 4.5.2.
- Licensed under the [Creative Commons Attribution-NonCommercial license](#). Please share and remix non-commercially, mentioning its origin.
- We acknowledge [students and instructors for previous versions of this course / slides](#).

