Mathematical Statistics II

Introduction to Point Estimation

Jesse Wheeler

Contents

1 Introduction

Overview

- We will formally introduce the idea of point estimation.
- In addition to an introduction, we will introduce the concept of the empirical distribution, as well as methods of moment estimators.
- The material for this section largely comes from Chapter 8 of Rice (2007).

2 Point Estimation: An introduction

Point estimation

- In the previous lecture(s), we provided an example of Bayesian vs Frequentsist point-estimation via first principles.
- That is, using the various interpretations, we could reason an estimate for the probability p in a binomial experiment.
- We are now interested in studying approaches for more general cases.
- Given a dataset and a chosen model, how can we estimate parameters?
- We will first start with some notation, and motivating examples.
- Term *model* in this class will generally refer to a probability model, and can be based on a discrete or continuous probability measure.

  Normal Model

  The Normal (or Gaussian) family of distributions arises often in the real world. Examples include human heights (conditioned on gender), rainfall amounts, and many biological measurements are approximately normal (or log-normal).

Given a set of observations $x_1, x_2, \ldots, x_n$, we may *model* these as iid normal $X_i \sim N(\mu, \sigma^2)$, and our goal being using the data to estimate the values of μ or σ.

Regression

Sometimes the probability model is *implicit*, but present. Consider the regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i.$$

We often think of fitting this regression model by minimizing the average squared-error: $(Y_i - \hat{Y}_i)^2$. However, this approach typically corresponds to an implicit probability model for the error terms ε_i, namely a normal distribution with mean 0. In this case, we might want to estimate β_0, β_1, and σ^2, which is $\mathrm{Var}(\varepsilon_i)$.

Poisson Process

Another common example is a Poisson Process model. Many real-world phenomena are well-approximated by a Poisson process, over space or time. Examples include arrival times at a gas station, number of meteors landing in a geographic area, radioactive decay, etc. Here, there is only one parameter we want to estimate using data, namely the rate λ.

Parameter Estimation

- All of the above examples have the common feature that we pick a *model*, and we want to use the model to describe the data-generating process.
- More accurately, however, we pick a candidate *family* of models; (Gaussian family, Poisson Family, Linear Regression family, etc).
- Generally, the exact model needed within a *family* of models is determined by a few parameters.
  - If the family is Gaussian, the model is determined by μ and σ^2.
  - If the family is Poisson, the model is determined by λ.
  - If the family is linear-Gaussian regression, the model is determined by β_0, β_1, and σ^2.

Example: Gamma-Rainfall

- The Gamma distribution depends on two parameters, α and λ:

$$f_X(x; \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}.$$

- The Gamma distribution is quite flexible, and works as a useful model for various situations.
- One example is modeling rainfall amounts per-storm under two conditions, cloud seeding vs not cloud seeding (simulated data, couldn't find original data).
- A Gamma distribution fits both samples well, but we get different parameters α and λ for the two different samples
- Differences in the respective distributions are reflected in differences in the parameters α and λ.

Two-sample Rainfall

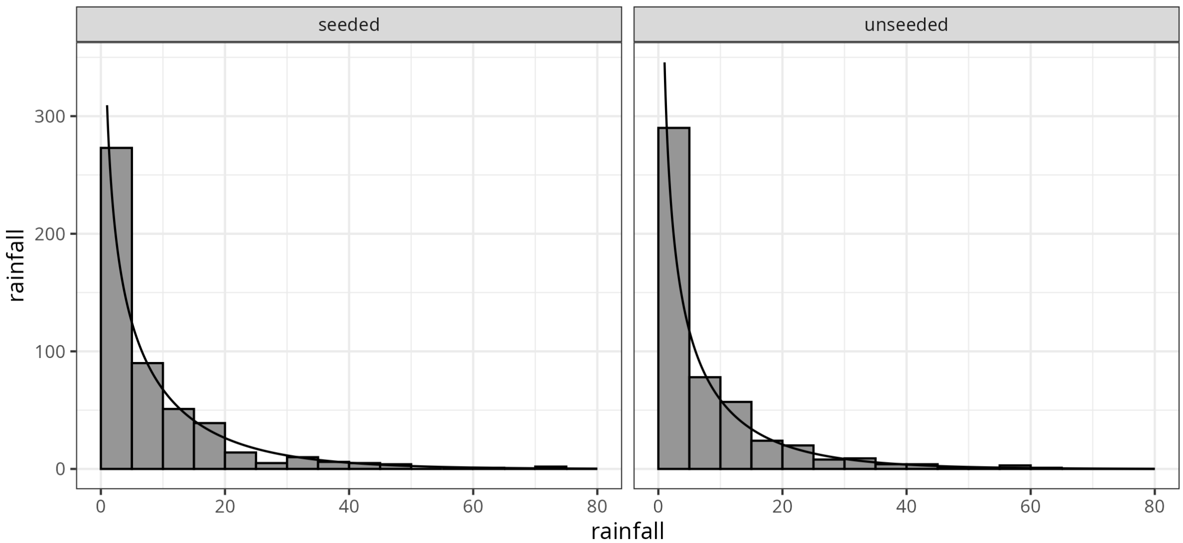

Figure 1: Data and model fit to two different Gamma distributions.

Notation and generalizations

- We will generalize by using the following ideas and notations.
- We will denote the *observed data* as $x_1^*, x_2^*, \ldots, x_N^*$, and use the shorthands $x_{1:N}^*$ if we emphasize the entire collection, and x^* if the emphasis is not needed.
- We assume that the data are realizations of random variables $X_1, X_2, \ldots, X_N$, again using the notation $X_{1:N}$ for the collection of N random variables, or X if this is not needed.
- In general, the data x_i^* and random variables X_i can be multivariate, but focus primarily on the univariate case.
- We will be interested in fitting a probabilistic model $f_{X_{1:N}}(x_{1:N}; \theta)$ using the data. The model may correspond to a discrete probability, or a continuous probability. In these cases, f is usually a pmf of pdf, respectively.
- Subscripts will be dropped occasionally if it is not necessary. For instance, $f(x; \theta)$ is taken to mean the model of all data $x = x_{1:N}$, and would formally be expressed as $f_{X_{1:N}}(x_{1:N}; \theta)$.
- This approach is sometimes called "function overload"; it's not my favorite approach, but it is convenient. The meaning of the function is primarily understood by the arguments and context.
- The function $f(x; \theta)$ belongs to a particular *family* of models, indexed by θ, which is generally multivariate.

Normal model example
Suppose we observe the following data: , and we would like to fit a normal model to the data, assuming the data are iid. Then $x_1^* = 3.49$, $x_2^* = 2$, and so forth, and the model family depends on $\theta = (\mu, \sigma^2)$,

and the model can be expressed as:

$$
\begin{aligned}
f(x;\theta) &= f_{X_{1:5}}(x_{1:5};\mu,\sigma^2) \\
&= \prod_{i=1}^{5} f_{X_i}(x_i;\mu,\sigma^2) \\
&= \prod_{i=1}^{5} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2}
\end{aligned}
$$

Our goal is to estimate μ, σ^2 using the observed data $x^*_{1:5}$.

- Our goal now is to develop general procedures for estimating θ, using observed data x^*, and a proposed family of models $f(x;\theta)$.
- We will develop three main approaches: (1) Method of Moments (2) Maximum Likelihood Estimation, and (3) Bayesian estimation.
- In this section, we will focus only on method of moments estimators.
- Once point estimation techniques are developed, we will provide theory about these estimates and their uncertainty; discussing bias, variance, an optimality of estimates.

3 Empirical distribution

The empirical distribution funciton

Acknowledgments

- Compiled on January 13, 2026 using R version 4.5.2.

- We acknowledge students and instructors for previous versions of this course / slides.

References

Rice JA (2007). *Mathematical statistics and data analysis*, volume 371. 3 edition. Thomson/Brooks/Cole Belmont, CA. 1