Mathematical Statistics II

Introduction to Point Estimation

Jesse Wheeler

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).

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?

Point estimation II

- 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.

Point estimation III

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.

Point estimation IV

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 $\mathrm{Var}(\varepsilon_i)$.

Point estimation V

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.

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

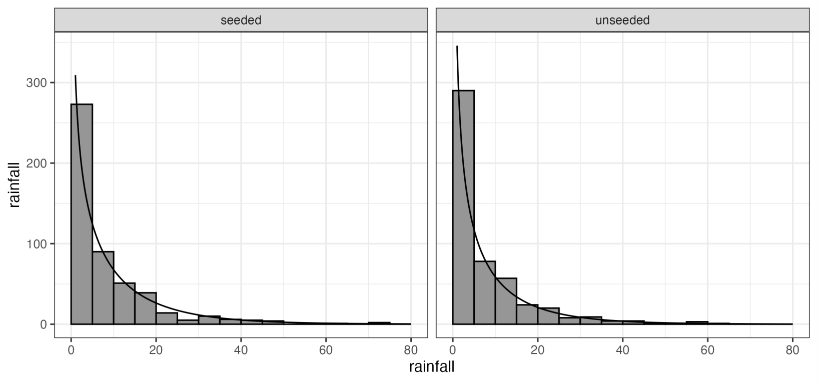

Figure 1: Data and model fit to two different Gamma distributions.

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.

Notation and generalizations II

- 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.

Notation and generalizations III

Normal model example

Suppose we observe the following data: 3.49, 2, 3.38, 1.62, 2.18, 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 $x_{1:5}^*$.

Notation and Generalization (continued)

- 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, and optimality of estimates.

Method of Moments

Motivation

- The Method of Moments (MoM) estimation technique is a simple idea.
- Pick a family of models $f(x;\theta)$, and observed data x^*.
- The family of models will have theoretical moments, i.e., $\mathsf{E}[X^k]$.
- Generally, these moments can be expressed in terms of the model parameters, θ.
- Thus, we will estimate $\hat{\theta}$ so that the data moments match the theoretical moments.

The empirical distribution

- One justification of this approach considers the empirical distribution of observed data.
- Let $X_1, X_2, \dots, X_N$ be random variables, representing a possible data sample.
- We will assume that X_i are iid, from some distribution F_θ (F_θ is the cdf here).
- We will define the empirical distribution function as:

$$F_n(t) = \frac{1}{N} \sum_{i=1}^{N} I[X_i \leq t].$$

- When we observe a specific dataset x^*, we can plug in these numbers to get a specific distribution that is not random.

The empirical distribution II

- A few things to note is that $F_n(t)$ is a proper CDF.
- By the law of large numbers, $F_n(t) \xrightarrow{a.s.} F_\theta(t)$ for every point t.
- The Glivenko–Cantelli theorem also strengthens this statement by saying that the convergence is uniform, in the sense that $\sup_t |F_n(t) - F_\theta(t)|$ converges to zero.
- It can be shown that the kth moment of the empirical distribution is

$$\hat{\mu}_k = \frac{1}{N} \sum_{i=1}^{N} X_i^k.$$

- Method of Moments idea:

The empirical distribution III

- For many commonly used parametric families (e.g., Gaussian, Poisson), the distribution is completely specified by a small set of parameters.
- These parameters are typically explicit functions of the moments of the distribution (e.g., mean and variance for the Gaussian).
- Although the moment generating function (MGF) uniquely determines the entire distribution, in many model families, the relevant parameters are uniquely determined by just the first few moments.
- Therefore, as the empirical moments computed from data converge to the true moments (by the Law of Large Numbers), it is natural to estimate model parameters by equating empirical and theoretical moments—leading to the method of moments estimators.

Method of Moments: generalized version

- To summarize mathematically, let $\mu_k = E[X^k]$ be the theoretical kth moment.
- Let $\hat{\mu}_k = \frac{1}{N}\sum_{i=1}^{N} X_i^k$ be the kth sample moment.
- $\hat{\mu}_k$ is an estimate of μ_k; however, we don't want an estimate of μ, we want an estimate of θ!
- For models with finite parameters, $\theta = (\theta_1, \ldots, \theta_k)$, we can often express θ_i as a function of $(\mu_1, \ldots, \mu_k)$:

$$\theta_i = g_i(\mu_1, \ldots, \mu_k)$$

.

Method of Moments: generalized version II

- Thus, our estimate of θ_i would be found by plugging in the empirical moments:

$$\hat{\theta}_i = g_i(\hat{\mu}_1, \ldots, \hat{\mu}_k).$$

Poisson Distribution

Suppose we observe data $x^*_{1:N}$, and want to fit a Poisson model. Since the Poisson distribution only has one parameter (λ), our goal is to use x^* to estimate λ.

Real-data example

- Before we start looking at real-data examples, let's introduce some basic R coding principles that will help us calculate moments from the data.
- R is a programming language, but for the sake of this class, we'll just treat it as a statistics calculator.
- For now, we will only focus on the most simple data types and operations: creating objects, vectors, and computing summary statistics.

Real-data example II

- First, saving objects in R. We can use = (like most languages), or the assignment operator: <-

```
x <- 2
x + 2
```

```
[1] 4
```

Real-data example III

- A vector in R is a collection of objects of the same data type. In this class, we will only need to use numeric data types

Real-data example IV

```r
x <- c(1, 2, 3, 4, 5)
class(x)
```

```
[1] "numeric"
```

```r
mean(x)
```

```
[1] 3
```

```r
sum(x)
```

```
[1] 15
```

Real-data example V

- Some fast ways of building vectors include:

```r
1:5  # this gives 1, 2, 3, 4, 5
```

```
[1] 1 2 3 4 5
```

```r
seq(1, 10, by = 2)  # Gives 1, 3, 5, 7, 9
```

```
[1] 1 3 5 7 9
```

Real-data example VI

- For generating random numbers, we can use the syntax: `rdist`.

Real-data example VII

```
rnorm(n = 10, mean = 2, sd = 1)
```

```
 [1] 1.7531041 0.7844391 3.5614051 2.4273102 0.7989765 3.0524585
 [7] 0.6949364 1.3073924 2.6026489 1.8022469
```

```
rpois(n = 7, lambda = 5)
```

```
[1] 2 6 1 5 5 9 6
```

```
rbeta(n = 3, shape1 = 0.8, shape2 = 1.3)
```

```
[1] 0.51652672 0.10386537 0.05986089
```

Real-data example VIII

- Lastly (and maybe most important), function documentation and help is readily available by appending a question mark: `?rnorm`

```
?mean
?rnorm
?sd
```

Real-data example IX

Poisson distribution with real data

The National Institute of Science and Technology collected data about asbestos fibers on filters. Asbestos dissolved in water was spread on a filter, and the number of fibers in each of 23 grid squares were counted:

```
 [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: Sampling Distribution and Bootstrap.

Real-data example X

Normal Distribution

TODO: Finish.

References and Acknowledgements

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

- Compiled on January 14, 2026 using R version 4.5.1.
-
- We acknowledge students and instructors for previous versions of this course / slides.