Estimation

Wenda Zhou

June 2, 2017

Estimation

Probability	:	parameters fixed and known, can generate more data
Statistics	:	parameters unknown, data given

Model

- A model describes how the data we observe is “generated”, depending on some underlying quantities of interest.
- Mathematically, we will consider a family of distributions which are parametrised by the quantities of interest.
- The goal of estimation is to deduce the parameter of interest from data.

Model

Example: coin toss

We have collected data from a series of 10 tosses of a given coin.

Model

We can model the number of heads using a binomial distribution

Parameter

The parameter of interest is p, the probability that the coin lands on heads

Likelihood

In probability, we think of the p.m.f. / p.d.f. as a function of the outcome $f_X(x) = f_X(x, p)$.
In statistics, the outcome x is usually given (it is the data we observe). Instead, we are interested in how the probability of obtaining the data changes according to the parameter.
We call this the likelihood

$$L(p) = L(p \mid x) = f_X(x, p) \tag{1}$$

Likelihood

Example: coin toss

Recall that the p.m.f. for a binomial with parameter p is given by:

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{2}$$

Suppose that we observe k heads in an experiment, then the likelihood is given by:

$$L(p) = \binom{n}{k} p^k (1-p)^{n-k} \tag{3}$$

Estimators

We will study estimators, which are functions of the data which attempt to "guess" the value of the parameter.
Usually, if the parameter is p, then we will write $\hat{p}$ for the estimator.
As estimators depend on the data, they are random quantities.

Performance of estimators

Wish to quantify how “good” an estimator is.
First idea: is our estimator “right” on average?
We say $\hat{\theta}$ is unbiased for θ if:

$$\mathbb{E}_\theta \, \hat{\theta} = \theta \tag{4}$$

Unbiased estimator

Binomial example

Let $X \sim \mathrm{Binom}(n, p)$.
Let $\hat{p} = n^{-1}X$.

$$\mathbb{E}_p\, \hat{p} = \mathbb{E}_p\, \frac{1}{n}X = \frac{1}{n}np = p \tag{5}$$

Hence $\hat{p}$ is unbiased for p.

Unbiased estimator of variance

Let $X_1, \ldots, X_n$ be independent with mean μ and variance σ^2. Suppose we wish to estimate σ^2 by:

$$\hat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{6}$$

This is unbiased:

$$\mathbb{E}\,\hat{\sigma^2} = \sigma^2 \tag{7}$$

Variance of unbiased estimator

Right on average. Now would like not too far on average.
However, if $\hat{\theta}$ is unbiased, we have:

$$\mathbb{E}_\theta(\hat{\theta}-\theta)^2 = \mathbb{E}_\theta(\hat{\theta}-\mathbb{E}\,\hat{\theta}) = \mathrm{Var}\,\hat{p} \tag{8}$$

Hence for unbiased estimator, want low variance.
We define the standard error to be the standard deviation of the estimator.

Mean squared error

Often, unbiasedness is not necessarily desirable. Instead, simply be close on average:
define the mean squared error (or m.s.e.):

$$\mathrm{mse}(\theta) = \mathbb{E}_\theta(\hat{\theta} - \theta)^2 \tag{9}$$

Mean squared error

Often, unbiasedness is not necessarily desirable. Instead, simply be close on average:
define the mean squared error (or m.s.e.):

$$\mathrm{mse}(\theta) = \mathbb{E}_\theta(\hat{\theta} - \theta)^2 \tag{10}$$

Mean squared error

It is possible to decompose the mean-squared error into two parts:

$$\begin{aligned}
\mathrm{mse}(\theta) &= \mathbb{E}_\theta(\hat{\theta} - \theta)^2 \\
&= \mathbb{E}_\theta(\hat{\theta} - \mathbb{E}(\hat{\theta}) + \mathbb{E}(\hat{\theta}) - \theta)^2 \\
&= \mathbb{E}_\theta(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2 + (\mathbb{E}\,\hat{\theta} - \theta)^2 \\
&= \mathsf{Var}\,\hat{\theta} + \mathrm{bias}^2
\end{aligned}$$

Hence have bias-variance tradeoff.

Maximum likelihood estimation

General idea to obtain a good estimator: maximise the (log-)likelihood.

m.l.e. for binomial

Suppose we observe $X = k$, and wish to estimate p.
The likelihood is given by:

$$L(p) = \binom{n}{k} p^k (1-p)^{n-k} \tag{11}$$

Compute log-likelihood:

$$\ell(p) = k \log p + (n-k) \log(1-p) + \log \binom{n}{k} \tag{12}$$

Maximise to obtain:

$$\hat{p} = k/n \tag{13}$$

m.l.e. for exponential

Suppose we have $X_1, \ldots, X_n \sim \text{Exp}(\lambda)$.
The likelihood is given by:

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \\ &= \lambda^n e^{-\lambda \sum_i x_i} \end{aligned}$$

Hence log-likelihood is given by:

$$\ell(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i \tag{14}$$

Maximise to obtain

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} \tag{15}$$

Method of moments

Alternative to m.l.e. – usually inferior, although simpler.
Idea: match sample and population moments.

Population and sample moments

Suppose we have a sample $X_1, \ldots, X_n$.
The first population moment is the population mean:

$$\mu_1 = \mathbb{E}\, X \tag{16}$$

The first sample moment is the sample mean:

$$M_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{17}$$

Population and sample moment

Definition: for $j \geq 2$ the jth (centered) population moment is

$$\mu_j = \mathbb{E}(X - \mathbb{E}\, X)^j \tag{18}$$

and the jth (centered) sample moment is

$$M_j = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^j \tag{19}$$

Population and sample moment

Definition: for $j \geq 2$ the jth (centered) population moment is

$$\mu_j = \mathbb{E}(X - \mathbb{E}\, X)^j \tag{18}$$

and the jth (centered) sample moment is

$$M_j = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^j \tag{19}$$

2^{nd} moment variance

3^{rd} moment skewness

4^{th} moment kurtosis

Method of moments

Example: exponential distribution

Suppose that $X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda)$. The population mean is given by:

$$\mu_1 = \lambda^{-1} \tag{20}$$

Hence we write:

$$M_1 = \hat{\mu}_1 \tag{21}$$

to obtain

$$\hat{\lambda} = \frac{n}{\sum_i X_i} \tag{22}$$

Method of moments

Example: gamma distribution

Suppose that $X_1, \dots, X_n$ follow a gamma distribution with shape k and scale θ.
The p.d.f. is given by:

$$f_X(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta} \tag{23}$$

The mean and variance are given by:

$$\mathbb{E}\, X = k\theta \text{ and } \sigma^2 = k\theta^2 \tag{24}$$

Method of moments

Example: gamma distribution

With two parameters, match first two moments.
Hence obtain

$$M_1 = k\theta$$
$$M_2 = k\theta^2$$

and thus solve to obtain:

$$\hat{k} = \frac{M_1^2}{M_2}$$
$$\hat{\theta} = \frac{M_2}{M_1}$$

Theoretical properties of the m.l.e.

- Not necessarily unbiased!
- Unbiased as $n \to \infty$. consistent
- Asymptotically normal (c.f. central limit theorem)