Statistics

Week 13: Maximum Likelihood (Chapter 15)

ESD, SUTD

Term 5, 2017

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Established in collaboration with MIT

Information

- **Assignment 4** out today, due next Monday.
- Updated slides as a single pdf will be available on *eDimension*.
- Tomorrow: revision.
- Anyone doing Project Option 2?
- Please complete the **course survey**.
- **Exam**: 9–11am, Friday 28 April.

  – Only the second half will be explicitly tested, though you need to know the concepts learned in the first half.

  – An A4 sheet with handwritten notes on both sides, and a non-programmable calculator are allowed.

  – You need to understand how least square regression works.

Outline

Parameter estimation, revisited

We talked about estimators, in particular unbiased estimators, but the ways in which we constructed them have been ad hoc.

For example, in Week 8 we had to estimate the parameter λ in a Poisson distribution. We used the fact that λ is also the expectation, and then estimated it using the sample mean.

What if we didn't know how to relate λ to the expectation? We now describe a general approach to parameter estimation.

The idea is to pick the value of the parameter which *maximizes the probability* of observing our data.

For example, suppose we want to estimate the parameter p of a Bernoulli distribution. A random sample of size 5 drawn from this distribution reads: 1, 1, 1, 1, 1.

Which of these is most likely? $p = 0;\ p = 1/2;\ p = 2/3;\ p = 1.$

Likelihood function

Given a probability distribution, let θ be a parameter to be estimated, and denote the probability density (or mass) function by $f(x|\theta)$.

The joint density for n iid random observations, $x_1, x_2, \ldots, x_n$, is

$$L(\theta) := \prod_{i=1}^{n} f(x_i|\theta).$$

$L(\theta)$ is called the **likelihood function** of θ.

If the maximum of this function occurs at $\hat{\theta}$, then $\hat{\theta}$ is the value of the parameter which maximizes the probability of observing our data, and we will use it as our estimate for θ.

Maximum likelihood estimate

$\hat{\theta}$ is called the **maximum likelihood estimate**. In many cases, we can find $\hat{\theta}$ by solving for

$$\frac{\mathrm{d}}{\mathrm{d}\theta} L(\theta) = 0, \quad \text{or equivalently,} \quad \frac{\mathrm{d}}{\mathrm{d}\theta} \log(L(\theta)) = 0.$$

('log' stands for natural log.)

Example – Poisson

$$L(\theta) = e^{-n\theta}\,\theta^{x_1+x_2+\cdots+x_n}\,\frac{1}{x_1!x_2!\cdots x_n!},$$

solving for $\dfrac{\mathrm{d}}{\mathrm{d}\theta} \log(L(\theta)) = 0$, we find that $\hat{\theta} = \bar{x}$.

We can use the *2nd derivative* to check that it is a maximum.

Maximum likelihood estimate – applications

- Maximum likelihood is used to find the coefficients in *logistic regression*.

- AIC $= 2k - 2\log(\hat{L})$, where $k =$ the number of estimated parameters in the model, and $\hat{L}$ is the maximized value of the likelihood function.

- Maximum likelihood is sometimes used to introduce *Bayesian statistics*, which is an alternative approach to the *frequentist statistics* taught in this course.

Exercises

1. (Exponential) If $f(x|\theta) = \theta e^{-\theta x}$, estimate θ based on x_i.

2. (Normal) Find the maximum likelihood estimates for μ and σ^2 in a normal distribution, whose pdf is given by

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right).$$

Are the estimators unbiased?

3. (Uniform) Suppose $x_1 = 1.2, x_2 = 3.5, x_3 = 2.7$ is a random sample from a *continuous uniform distribution* over $[0,\ \theta]$.

Then $f(x|\theta) = 1/\theta$ if $0 \leq x \leq \theta$, and 0 otherwise. Find $\hat{\theta}$.

Generalize your result. Does this give an unbiased estimator?