Statistics

Adam Kelly (ak2316@cam.ac.uk)

January 29, 2022

Statistics the science of making informed decisions. It is an area of mathematics that is widely applicable, and covers the design of experiments, the analysis of data, statistical inference, and the communication of uncertainty and risk. In this course we will focus on formal statistical inference.

This article constitutes my notes for the 'Statistics' course, held in Lent 2022 at Cambridge. These notes are *not a transcription of the lectures*, and differ significantly in quite a few areas. Still, all lectured material should be covered. Note that familiarity with classical probability theory is assumed, and a recap is not included.

Contents

§1 Introduction

§1.1 Parametric Inference

Let $X_1, \dots, X_n$ be i.i.d random variables. We will assume that the distribution of X_i belongs to some family with parameter $\theta \in \Theta$. For example, we could have $X_1 \sim \text{Poisson}(\mu)$ with $\theta = \mu \in \Theta = (0, \infty)$. We can also have a multi-dimensional parameter such as $X_1 \sim N(\mu, \sigma^2)$ with $\theta(\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$. We will use the observed $X = (X_1, \dots, X_n)$ to make inferences about the parameter θ. Such inferences include:

1. Point estimate of $\hat{\theta}(X)$ of θ;
2. An interval estimate $(\hat{\theta}_1(X), \hat{\theta}_2(X))$ of θ;
3. Testing hypothesis about θ, that is, checking whether there is evidence in the data X against a given hypothesis.

In this course we will usually assume that the distribution family of $X_1, \dots, X_n$ is known, and that the parameter is unknown. From time to time we will be able to make slightly more general statements.

§1.2 Estimation

Suppose $X_1, \dots, X_n$ are i.i.d observations with a pdf (or pmf) $f_X(x \mid \theta)$ where θ is an unknown parameter in a parameter space Θ.

Definition 1.1 (Estimator)
An **estimator** is a statistic or function of the data $T(X) = \hat{\theta}$ which does *not* depend on θ, and is used to approximate the true parameter θ.

The distribution of $T(x)$ is its **sampling distribution**.

We can then go on to define the *bias* of an estimator.

Definition 1.2 (Bias)
The **bias** of $\hat{\theta} = T(x)$ is

$$\operatorname{bias}(\hat{\theta}) = \mathbb{E}_\theta(\hat{\theta}) - \theta,$$

where $\mathbb{E}_\theta$ gives expectation in the model $X_i \sim f_X(\cdot \mid \theta)$.

Remark. In general the bias is a function of the true parameter θ, even though it is not explicit in the notation $\operatorname{bias}(\hat{\theta})$.

Definition 1.3 (Unbiased Estimator)
We say that $\hat{\theta}$ is **unbiased** if $\operatorname{bias}(\hat{\theta}) = 0$ for *all* values of the true parameter θ.

Example 1.4 (Defining an Estimator)
Consider the i.i.d random variables $X_1, \dots, X_n \sim N(\mu, 1)$. Then we can define the estimator

$$\hat{\mu} = T(x) = \frac{1}{n} \sum_i X_i = \overline{X_n}.$$

The sampling distribution of $\hat{\mu}$ is $T(x) \sum N(\mu, 1/n)$.

We can see that $\hat{\mu}$ is an unbiased estimator since

$$\mathbb{E}_\theta(\hat{\mu}) = \mathbb{E}_\mu(\overline{X_n}) = \mu,$$

for all $\mu \in \mathbb{R}$.

One way of measuring 'how far' $\hat{\theta}$ is from θ 'on average' is given by the *mean squared error* (mse).

Definition 1.5 (Mean Squared Error)
The **mean squared error** of θ is defined to be

$$\operatorname{mse}(\hat{\theta}) = \mathbb{E}_\theta[(\hat{\theta} - \theta)^2].$$

Again, note that $\operatorname{mse}(\hat{\theta})$ may depend on θ, even though this dependence is omitted from the notation.

Method 1.6 (Bias-Variance Decomposition)

We can decompose the mean squared error of an estimator as follows.

$$\begin{aligned} \operatorname{mse}(\hat{\theta}) &= \mathbb{E}_\theta \left[(\hat{\theta} - \theta)^2 \right] \\ &= \mathbb{E}_\theta \left[(\hat{e\theta}^n - \mathbb{E}_\theta \hat{\theta} + \underbrace{\mathbb{E}_\theta \hat{\theta} - \theta \Big)^2}\, \right] \\ &= \underbrace{\operatorname{Var}_\theta(\hat{\theta})}_{\geqslant 0} + \underbrace{\operatorname{bias}^2_\theta(\hat{\theta})}_{\geqslant 0} \end{aligned}$$

This shows that there is a tradeoff between bias and variance in an estimator.

Example 1.7

Consider the random variable $X \sim \operatorname{Binomial}(n, \theta)$, and suppose n is known, with $\theta \in [0, 1]$ being an unknown parameter.

The standard estimator for this example would be

$$T_u = \frac{X}{n},$$

which can be thought of as the 'proportion of successes observed'. This estimator is clearly unbiased since

$$\mathbb{E}_\theta(T_u) = \frac{\mathbb{E}_\theta X}{n} = \frac{n\theta}{n} = \theta.$$

We can also compute that

$$\operatorname{mse}(T_u) = \operatorname{Var}_\theta(T_u) = \frac{\theta(1-\theta)}{n}.$$

Now consider another estimator

$$T_B = \frac{X+1}{n+2}$$

§2 LECTURE 3 TBC

(TODO: Type this up)

§3 Lecture 4

Remark. If S and T are minimal sufficient then they are in bijection, that is, $T(x) = T(y)$ if and only if $S(x) = S(y)$. So we can think of minimal sufficient statistics as being unique up to bijection.

Theorem 3.1 (Criterion for Minimal Sufficiency)

Suppose that $f_x(x|\theta)/f_x(y \mid \theta)$ is constant in θ if and only if $T(x) = T(y)$. Then T is minimal sufficient.