Statistics

Adam Kelly (`ak2316@cam.ac.uk`)

January 20, 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.

Contents

§1   Introduction

§1.1   Parametric Inference

Let $X_1, \ldots, X_n$ 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 $X_1 \sim$ Poisson(μ) 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 $X = (X_1, \ldots, X_n)$ 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 $X_1, \ldots, X_n$ is known, and that the parameter is unknown. From time to time we will be able to make slightly more general statements.

§1.2   A Recap on Probability

Statistics makes heavy use of probability theory, and for that reason we will review some of the relevant material here.

Let Ω be the **sample space** of outcomes in an experiment. A measurable subset of Ω is called an **event**, and the set of events is denoted $\mathcal{F}$.

A **probability measure** $\mathbb{P} : \mathcal{F} \to [0,1]$ is a function which satisfies

1. $\mathbb{P}(\emptyset) = 0$

2. $\mathbb{P}(\Omega) = 1$

3. $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$ if A_i is a sequence of disjoint events.

A **random variable** is a (measurable) function $X : \Omega \to \mathbb{R}$. Note that the random variable is defined without reference to the probability measure.

> **Example 1.1** (Tossing a Coin)
>
> Consider the experiment of tossing a coin. Here we have $\Omega = \{HH, HT, TH, TT\}$, and $\mathcal{F}$ is the power set of Ω. We can have a random variable X to be the number of heads, with
>
> $$X(HH) = 2, \quad X(HT) = 1, \quad X(TH) = 1, \quad X(TT) = 0$$

The way that a random variable is linked to the probability measure is by considering it's **distribution**, that is $F_X : \mathbb{R} \to [0,1]$ with

$$F_X(x) = \mathbb{P}(X \le x).$$

A discrete random variable is one which takes values in a countable set. It's **probability mass function** is given by

$$p_X(x) = \mathbb{P}(X = x).$$

We say that X has a continuous distribution if it has a **probability density function** $f_X(x)$ which satisfies

$$\mathbb{P}(X \in A) = \int_A f_X(x) \, \mathrm{d}x$$

when A is 'sufficiently nice'.

The **expectation** of a random variable X is defined to be

$$\mathbb{E}[X] = \sum_{x \in X} x \cdot p_X(x)$$

when X is discrete, and

$$\mathbb{E}[X] = \int_{\infty}^{\infty} x f_X(x) \, \mathrm{d}x$$

when X is continuous.

The **variance** of X is given by

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

We say that $X_1, \ldots, X_n$ are **independent** if for all $x_1, \ldots, x_n$ we have

$$\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1) \cdots \mathbb{P}(X_n \le x_n).$$

If the variables have probability density functions (or probability mass functions) $f_{X_1}, \ldots, f_{X_n}$ then the joint PDF (or PMF) is

$$f_X(x) = \prod_i f_{X_i}(x_i).$$

> **Example 1.2** (Maximum of Random Variables)
>
> If $Y = \max\{X_1, \ldots, X_n\}$ for independent random variables $X_1, \ldots, X_n$, then
>
> $$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(X_1 \le y, \ldots, X_n \le y) = \prod_i F_{X_i}(y),$$
>
> and then by differentiating we can obtain the PDF of Y (if it exists).

We will now consider some linear transformations of random variables. Let $a \in \mathbb{R}^n$ be a constant. We have

$$\mathbb{E}[a_1 X_1 + \cdots + a_n X_n] = \mathbb{E}[a^T X] = a^T \mathbb{E}[X]$$
$$\mathrm{Var}(a^T X) = \sum_{i,j} a_i a_j \, \mathrm{Cov}(X_i, X_j) = a^T \, \mathrm{Var}(X) a$$

where $\mathrm{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$, and

$$\mathrm{Var}(X)_{i,j} = \mathrm{Cov}(X_i, X_j)$$

is the **variance-covariance matrix**.

Let $X_1, \ldots, X_n$ be i.i.d random variables with $\mathbb{E}[X_1] = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$. Then if we define the **sample mean**

$$S_n = \sum_i X_i, \quad \bar{x}_n = S_n / n.$$

By linearity, $\mathbb{E}[\bar{x}_n] = \mu$ and $\mathrm{Var}(\bar{x}_n) = \sigma^2 / n$.

We define the **standard statistic**

$$Z_n = \frac{S_n - n\mu}{\sigma \sqrt{n}} = \frac{\sqrt{n}(\hat{x}_n - \mu)}{\sigma}$$

which has mean 0 and variance 1.

The **moment generating function** of a random variable X is the Laplace transform of the PDF, that is,

$$M_X(t) = \mathbb{E}[e^{tX}],$$

provided that it exists for t in some neighborhood of 0. This has a nice relation to the moments of X, in that

$$\mathbb{E}[X^n] = \frac{d^n}{dt^n} M_X(t) \Big|_{t=0}.$$