IB Statistics

Martin von Hodenberg (`mjv43@cam.ac.uk`)

January 20, 2022

This article constitutes my notes for the 'IB 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

§0 Introduction

Statistics can be defined as the science of *making informed decisions*. It can include:

1. Formal statistical inference
2. Design of experiments and studies
3. Visualisation of data
4. Communication of uncertainty and risk
5. Formal decision theory

In this course we will only focus on formal statistical inference.

Definition 0.1 (Parametric inference)

Let $X_1, \dots, X_n$ be iid. random variables. We will assume the distribution of X_1 belongs to some family with parameter $\theta \in \Theta$.

Example 0.2

We will give some examples of such families:

1. $X_1 \sim \text{Po}(\mu), \theta = \mu \in \Theta = (0, \infty)$.
2. $X_1 \sim N(\mu, \sigma^2) \quad N(\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$.

We will use the observed $X = (X_1, \dots X_n)$ to make inferences about θ such as:

1. Point estimate $\theta(X)$ of θ.
2. Interval estimate of θ: $(\theta_1(x), \theta_2(x))$
3. Testing hypotheses about θ: for example checking if there is evidence in X against the hypothesis $H_0 : \ \theta = 1$.

Remark. In general, we'll assume the distribution of the family $X_1, \dots, X_n$ is known but the parameter is unknown. Some results (on mean square error, bias, Gauss-Markov theorem) will make weaker assumptions.

§1 Probability

First we will briefly recap IA Probability.

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

Definition 1.1 (Probability measure)

A probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfies:

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

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

Definition 1.2 (Random variable)

A random variable is a (measurable) function $X : \Omega \to \mathbb{R}$.

Example 1.3

Tossing two coins has $\Omega = \{HH, HT, TH, TT\}$. Since Ω is countable, $\mathcal{F}$ is the power set of Ω. We can define X to be the random variable that counts the number of heads. Then

$$X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 0.$$

Definition 1.4 (Distribution function)

The distribution function of X is $F_X(x) = \mathbb{P}(X \leq x)$.

A discrete random variable takes values in a countable set $S \subset \mathbb{R}$. Its probability mass function is

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

A random variable X has a continuous distribution if it has a probability density function $f_X(x)$ which satisfies

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

for measurable sets A.

The expectation of X is

$$\mathbb{E}(X) = \begin{cases} \sum_{x \in X} x p_X(x) & X \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x & X \text{ is continuous} \end{cases}$$

If $g : \mathbb{R} \to \mathbb{R}$, then for a continuous r.v

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\mathrm{d}x.$$

The variance of X is

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

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

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1) \dots \mathbb{P}(X_n \leq x_n).$$

If $X_1, \dots, X_n$ have pdfs or pmfs $f_{X_1}, \dots, f_{X_n}$ then their joint pdf or pmf is

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

If $Y = \max(X_1, \dots, X_n)$ independent, then

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X_1 \leq y, \dots, X_n \leq y) = \prod_i F_{X_i}(y).$$

The pdf of Y (if it exists) is obtained by differentiating F_Y.

§1.1 Linear transformations

Let $(a_1, \dots a_n)^T = a \in \mathbb{R}^n$ be a constant.

$$\mathbb{E}(a_1 X_1 + \ldots + a_n X_n) = \mathbb{E}(a^T X) = a^T \mathbb{E}(X).$$

This gives linearity of expectation (does not require independence).

$$\operatorname{Var}(a^T X) = \sum_{i,j} a_i a_j \underbrace{\operatorname{Cov}(X_i, X_j)}_{=\mathbb{E}((X_i - \mathbb{E}(X_i)(X_j - \mathbb{E}(X_j))))} = a^T \operatorname{Var}(X) a.$$

where the matrix $[\operatorname{Var}(X)]_{ij} = \operatorname{Cov}(X_i, X_j)$. This gives the "bilinearity of variance".

§1.2 Standardised statistics

Let $X_1, \dots, X_n$ be iid. with $\mathbb{E}(X_1) = \mu$, $\operatorname{Var}(X_1) = \sigma^2$. We define $S_n = \sum_i X_i$ and $\overline{X_n} \frac{S_n}{n}$ (the sample mean). By linearity

$$\mathbb{E}(\overline{X_n}) = \mu, \quad \operatorname{Var}(\overline{X_n}) = \frac{\sigma^2}{n}.$$

Define $Z_n = \frac{S_n - n\mu}{n}$. Then $\mathbb{E}(Z_n) = 0$ and $\operatorname{Var}(Z_n) = 1$.

§1.3 Moment generating functions