Part IB - Statistics

Lectured by D. Spiegelhalter

Lent 2015

Estimation
Review of distribution and density functions, parametric families. Examples: binomial, Poisson, gamma. Sufficiency, minimal sufficiency, the Rao-Blackwell theorem. Maximum likelihood estimation. Confidence intervals. Use of prior distributions and Bayesian inference. [5]

Hypothesis testing
Simple examples of hypothesis testing, null and alternative hypothesis, critical region, size, power, type I and type II errors, Neyman-Pearson lemma. Significance level of outcome. Uniformly most powerful tests. Likelihood ratio, and use of generalised likelihood ratio to construct test statistics for composite hypotheses. Examples, including t-tests and F-tests. Relationship with confidence intervals. Goodness-of-fit tests and contingency tables. [4]

Linear models
Derivation and joint distribution of maximum likelihood estimators, least squares, Gauss-Markov theorem. Testing hypotheses, geometric interpretation. Examples, including simple linear regression and one-way analysis of variance. Use of software. [7]

Contents

1 Introduction and probability review

Lecturer apologizes for not turning up the previous lecture

Definition (Statistics)**.** *Statistics* is a set of principle a nd procedures for gaining and processing quantitative evidence in order to help us make judgements and decisions.

Note that we did not mention data. We don't necessarily need data for statistics (even though most often we do).

In this course, we focus on formal *statistical inference*. In the process, we assume

- we have data generated from some unknown probability model
- we aim to use the data to learn about certain properties of the underlying probability model

In particular, we perform parametric inference:

We assume a random variable X takes values in $\mathcal{X}$. We assume its distribution belongs to a family of distribution (e.g. Poisson) indexed by a scalar or vector parameter θ, taking values in some parameter space Θ. We call this a *parametric family*.

For example, we can have $X \sim \text{Poisson}(\mu)$ and $\theta = \mu \in \Theta = (0, \infty)$.

We assume that we already which family it belongs to, and then try to find out θ.

Suppose $X_1, X_2, \cdots, X_n$ are iid with the same distribution as X. Then $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ is a simple random sample (i.e. our data).

We use the observed $\mathbf{X} = \mathbf{x}$ to make inferences about θ, such as

- giving an estimate $\hat{\theta}(\mathbf{x})$ of the true value of θ.
- Giving an interval estimate $(\hat{\theta}_1(\mathbf{x}), \hat{\theta}_2(\mathbf{x}))$ for θ
- testing a hypothesis about θ, e.g. whether $\theta = 0$.