Part IB - Statistics
Definitions

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

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.

2 Estimation, bias and mean squared error

Definition (Statistic)**.** A *statistic* is an estimate of θ. It is a function T of a data. If $\mathbf{X} = \mathbf{x} = (x_1, \cdots, x_n)$, then our estimate is written as $\hat{\theta} = T(\mathbf{x})$. $T(\mathbf{X})$ is an *estimator* of θ.

The distribution of $T = T(\mathbf{X})$ is its sampling distribution.

Note that capital $\mathbf{X}$ denotes a random variable and $\mathbf{x}$ is an observed value. So $T(\mathbf{X})$ is a random variable and $T(\mathbf{x})$ is a particular value.

Definition (Bias)**.** Let $\hat{\theta} = T(\mathbf{X})$ be an estimator of θ. The *bias* of $\hat{\theta}$ is the difference between its expected value and true value.

$$\mathrm{bias}(\hat{\theta}) = \mathbb{E}_\theta(\hat{\theta}) - \theta.$$

Note that the subscript $_\theta$ does not represent the random variable, but the thing we want to estimate. This is inconsistent with the use for, say the pmf.

An estimator is *unbiased* if it has no bias, i.e. $\mathbb{E}_\theta(\hat{\theta}) = \theta$.

2.1 Mean squared error

Definition (Mean squared error)**.** The *mean squared error* of an estimator $\hat{\theta}$ is $\mathbb{E}_\theta[(\hat{\theta} - \theta)^2]$.

Sometimes, we use the *root mean squared error*, that is the square root of the above.