Part IB — Statistics

Theorems

Based on lectures by D. Spiegelhalter

Notes taken by Dexter Chua

Lent 2015

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

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

0 Introduction

1 Estimation

1.1 Estimators

1.2 Mean squared error

1.3 Sufficiency

Theorem (The factorization criterion)**.** T is sufficient for θ if and only if

$$f_{\mathbf{X}}(\mathbf{x} \mid \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x})$$

for some functions g and h.

Theorem. Suppose $T = T(\mathbf{X})$ is a statistic that satisfies

$$\frac{f_{\mathbf{X}}(\mathbf{x}; \theta)}{f_{\mathbf{X}}(\mathbf{y}; \theta)} \text{ does not depend on } \theta \text{ if and only if } T(\mathbf{x}) = T(\mathbf{y}).$$

Then T is minimal sufficient for θ.

Theorem (Rao-Blackwell Theorem)**.** Let T be a sufficient statistic for θ and let $\tilde{\theta}$ be an estimator for θ with $\mathbb{E}(\tilde{\theta}^2) < \infty$ for all θ. Let $\hat{\theta}(\mathbf{x}) = \mathbb{E}[\tilde{\theta}(\mathbf{X}) \mid T(\mathbf{X}) = T(\mathbf{x})]$. Then for all θ,

$$\mathbb{E}[(\hat{\theta} - \theta)^2] \leq \mathbb{E}[(\tilde{\theta} - \theta)^2].$$

The inequality is strict unless $\tilde{\theta}$ is a function of T.

1.4 Likelihood

1.5 Confidence intervals

1.6 Bayesian estimation

2 Hypothesis testing

2.1 Simple hypotheses

Lemma (Neyman-Pearson lemma). Suppose $H_0 : f = f_0$, $H_1 : f = f_1$, where f_0 and f_1 are continuous densities that are nonzero on the same regions. Then among all tests of size less than or equal to α, the test with the largest power is the likelihood ratio test of size α.

2.2 Composite hypotheses

Theorem (Generalized likelihood ratio theorem). Suppose $\Theta_0 \subseteq \Theta_1$ and $|\Theta_1| - |\Theta_0| = p$. Let $\mathbf{X} = (X_1, \cdots, X_n)$ with all X_i iid. Then if H_0 is true, as $n \to \infty$,

$$2 \log \Lambda_{\mathbf{X}}(H_0 : H_1) \sim \chi_p^2.$$

If H_0 is not true, then $2 \log \Lambda$ tends to be larger. We reject H_0 if $2 \log \Lambda > c$, where $c = \chi_p^2(\alpha)$ for a test of approximately size α.

2.3 Tests of goodness-of-fit and independence

2.3.1 Goodness-of-fit of a fully-specified null distribution

2.3.2 Pearson's Chi-squared test

2.3.3 Testing independence in contingency tables

2.4 Tests of homogeneity, and connections to confidence intervals

2.4.1 Tests of homogeneity

2.4.2 Confidence intervals and hypothesis tests

Theorem.

(i) Suppose that for every $\theta_0 \in \Theta$ there is a size α test of $H_0 : \theta = \theta_0$. Denote the acceptance region by $A(\theta_0)$. Then the set $I(\mathbf{X}) = \{\theta : \mathbf{X} \in A(\theta)\}$ is a $100(1 - \alpha)\%$ confidence set for θ.

(ii) Suppose $I(\mathbf{X})$ is a $100(1 - \alpha)\%$ confidence set for θ. Then $A(\theta_0) = \{\mathbf{X} : \theta_0 \in I(\mathbf{X})\}$ is an acceptance region for a size α test of $H_0 : \theta = \theta_0$.

2.5 Multivariate normal theory

2.5.1 Multivariate normal distribution

Proposition.

(i) If $\mathbf{X} \sim N_n(\mu, \Sigma)$, and A is an $m \times n$ matrix, then $A\mathbf{X} \sim N_m(A\boldsymbol{\mu}, A\Sigma A^T)$.

(ii) If $\mathbf{X} \sim N_n(\mathbf{0}, \sigma^2 I)$, then

$$\frac{|\mathbf{X}|^2}{\sigma^2} = \frac{\mathbf{X}^T\mathbf{X}}{\sigma^2} = \sum \frac{X_i^2}{\sigma^2} \sim \chi_n^2.$$

Instead of writing $|\mathbf{X}|^2/\sigma^2 \sim \chi_n^2$, we often just say $|\mathbf{X}|^2 \sim \sigma^2\chi_n^2$.

Proposition. Let $\mathbf{X} \sim N_n(\mu, \Sigma)$. We split $\mathbf{X}$ up into two parts: $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$, where $\mathbf{X}_i$ is a $n_i \times 1$ column vector and $n_1 + n_2 = n$.

Similarly write

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where Σ_{ij} is an $n_i \times n_j$ matrix.

Then

(i) $\mathbf{X}_i \sim N_{n_i}(\boldsymbol{\mu}_i, \Sigma_{ii})$

(ii) $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent iff $\Sigma_{12} = 0$.

Proposition. When Σ is a positive definite, then $\mathbf{X}$ has pdf

$$f_{\mathbf{X}}(\mathbf{x}; \mu, \Sigma) = \frac{1}{|\Sigma|^2} \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right].$$

2.5.2 Normal random samples

Theorem (Joint distribution of $\bar{X}$ and S_{XX})**.** Suppose $X_1, \cdots, X_n$ are iid $N(\mu, \sigma^2)$ and $\bar{X} = \frac{1}{n}\sum X_i$, and $S_{XX} = \sum (X_i - \bar{X})^2$. Then

(i) $\bar{X} \sim N(\mu, \sigma^2/n)$

(ii) $S_{XX}/\sigma^2 \sim \chi^2_{n-1}$.

(iii) $\bar{X}$ and S_{XX} are independent.

2.6 Student's t-distribution

Proposition. If $k > 1$, then $\mathbb{E}_k(T) = 0$.

If $k > 2$, then $\operatorname{var}_k(T) = \frac{k}{k-2}$.

If $k = 2$, then $\operatorname{var}_k(T) = \infty$.

In all other cases, the values are undefined. In particular, the $k = 1$ case, this is known as the Cauchy distribution, and has undefined mean and variance.

3 Linear models

3.1 Linear models

Proposition. The least squares estimator satisfies

$$X^T X\hat{\boldsymbol{\beta}} = X^T \mathbf{Y}. \tag{3}$$

3.2 Simple linear regression

Theorem (Gauss Markov theorem)**.** In a full rank linear model, let $\hat{\boldsymbol{\beta}}$ be the least squares estimator of $\boldsymbol{\beta}$ and let $\boldsymbol{\beta}^*$ be any other unbiased estimator for $\boldsymbol{\beta}$ which is linear in the Y_i's. Then

$$\mathrm{var}(\mathbf{t}^T\hat{\boldsymbol{\beta}}) \leq \mathrm{var}(\mathbf{t}^T\boldsymbol{\beta}^*).$$

for all $\mathbf{t} \in \mathbb{R}^p$. We say that $\hat{\boldsymbol{\beta}}$ is the *best linear unbiased estimator* of $\boldsymbol{\beta}$ (BLUE).

3.3 Linear models with normal assumptions

Proposition. Under normal assumptions the maximum likelihood estimator for a linear model is

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y},$$

which is the same as the least squares estimator.

Lemma.

(i) If $\mathbf{Z} \sim N_n(\mathbf{0}, \sigma^2 I)$ and A is $n \times n$, symmetric, idempotent with rank r, then $\mathbf{Z}^T A\mathbf{Z} \sim \sigma^2\chi^2_r$.

(ii) For a symmetric idempotent matrix A, $\mathrm{rank}(A) = \mathrm{tr}(A)$.

Theorem. For the normal linear model $\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, \sigma^2 I)$,

(i) $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(X^T X)^{-1})$

(ii) $\mathrm{RSS} \sim \sigma^2\chi^2_{n-p}$, and so $\hat{\sigma}^2 \sim \frac{\sigma^2}{n}\chi^2_{n-p}$.

(iii) $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent.

3.4 The F distribution

Proposition. If $X \sim F_{m,n}$, then $1/X \sim F_{n,m}$.

3.5 Inference for $\boldsymbol{\beta}$

3.6 Simple linear regression

3.7 Expected response at $\mathbf{x}^*$

3.8 Hypothesis testing

3.8.1 Hypothesis testing

Lemma. Suppose $\mathbf{Z} \sim N_n(\mathbf{0}, \sigma^2 I_n)$, and A_1 and A_2 are symmetric, idempotent $n \times n$ matrices with $A_1 A_2 = 0$ (i.e. they are orthogonal). Then $\mathbf{Z}^T A_1\mathbf{Z}$ and $\mathbf{Z}^T A_2\mathbf{Z}$ are independent.

3.8.2 Simple linear regression

3.8.3 One way analysis of variance with equal numbers in each group