The University of New South Wales

Department of Statistics

School of Mathematics

Statistical Inference, MATH5905

Lecture notes

Associate Professor Spiridon Penev

Foreword

These notes do **not** substitute the lectures in Statistical Inference for masters students. You are strongly recommended to attend each and every lecture because the conceptual bases of the discussed methods, as well as some additional derivations and explanations will then be focused on. This volume is therefore not meant to be a substitute for a textbook, or lecture attendance.

These notes are a compilation from several sources and other notes. Some of the sources are listed in your handout.

I will appreciate if you would let me know about any ways these notes could be further improved.

1 Lecture 1: REVISION: ELEMENTS OF PROBABILITY

Some standard univariate distributions like the binomial, Poisson, normal, Cauchy, logistic, exponential, double exponential (also called Laplace distribution), are assumed to be known. These are summarised in the Table of Common Distributions on pages 621–626 of **CB** and a hard copy is given to everyone.

The revision below involves mainly the following sections of the CB reference:

- **Probability (Sections 1.2, 1.3)** Conditional probability and independence:

  $$P(A|B) = \frac{P(A \bigcap B)}{P(B)} = \frac{P(A)}{P(B)} P(B|A)$$

  is the conditional probability of A given B.

  A and B are *independent* if $P(A \bigcap B) = P(A)P(B)$

  and if A and B are independent then

  $$P(A|B) = P(A).$$

  Bayes rule: for a partition $\{A_i\}$ of the sample space S :

  $$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

- **Random variables and distributions (univariate) (Sections 1.4, 2.1)**

  Transformations.

  Theorem. Let X be a random variable with cdf function $F_X(.)$ and density $f_X(.)$. Let $Y = g(X)$ and $F_Y(.)$ be the cdf of Y. Put

  $$S_X = \{x : f_X(x) > 0\}; S_Y = \{y : y = g(x) \text{ for some } x \in S_X.\}$$

  i) If g is increasing on S_X then $F_Y(y) = F_X(g^{-1}(y))$ for $y \in S_y$.

  ii) If g is decreasing on S_X and X is continuous random variable then $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in S_y$.

  Theorem (Probability integral transform) Let X be a continuous random variable with a cdf $F_X(.)$. The the random variable $Y = F_X(X)$ is uniformly distributed on $[0, 1]$.

- **Expectations, Variances and correlations (Sections 2.2, 2.3, 4.5)**

  $$E(g(X)) = \int g(x)f_X(x)dx, E(g(X)) = \sum g(x_i)P(X = x_i) = \sum g(x_i)f_X(x_i)$$

  for the continuous and for the discrete case, respectively.

  Variance: $Var(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2$.

  Covariance: $Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$.

  Correlation: $\rho = Cor(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$.

- **Multivariate distributions (Sections 4.1, 4.2)**

  Random vector $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_p \end{pmatrix} \in R^p$, $p \geq 2$ has p different components each of which is a random variable with a cumulative distribution function *(cdf)* $F_{X_i}(x_i), i = 1, 2, ..,$p. Each of the functions $F_{X_i}(.)$ is called a *marginal distribution.* The *joint cdf* of the random vector $\mathbf{X}$ is

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p) = F_{\mathbf{X}}(x_1, x_2, .., x_p)$$

  In case of a *discrete* vector of observations $\mathbf{X}$ the *probability mass function* is defined as

$$P_{\mathbf{X}}(\mathbf{x}) = P(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p)$$

  If a *density* $f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, x_2, .., x_p)$ exists such that

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} .. \int_{-\infty}^{x_p} f_{\mathbf{X}}(\mathbf{t}) dt_1 \ldots dt_p$$

  then $\mathbf{X}$ is a *continuous* random vector with a joint density function of p arguments $f_{\mathbf{X}}(\mathbf{x})$. In this case $f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^p F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \partial x_2 .. \partial x_p}$ holds. In case $\mathbf{X}$ has p independent components then

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1}(x_1) F_{X_2}(x_2) \ldots F_{X_p}(x_p)$$

  holds and, equivalently, also

$$P_{\mathbf{X}}(\mathbf{x}) = P_{X_1}(x_1) P_{X_2}(x_2) \ldots P_{X_p}(x_p), f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) f_{X_2}(x_2) f_{X_p}(x_p)$$

  holds.

  The *marginal cdf of the first* $k < p$ components of the vector $\mathbf{X}$ is defined in a natural way as follows:

$$P(X_1 \leq x_1, X_2 \leq x_2, .., X_k \leq x_k) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k, X_{k+1} \leq \infty, ..., X_p \leq \infty)$$

$$= F_{\mathbf{X}}(x_1, x_2, .., x_k, \infty, \infty, .., \infty)$$

  The *marginal density* of the first k components can be obtained by partial differentiation:

$$\int_{-\infty}^{\infty} .. \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, .., x_p) dx_{k+1} \ldots dx_p$$

The *conditional density* $\mathbf{X}$ **when** $X_{r+1} = x_{r+1}, \ldots, X_p = x_p$ is defined by

$$f_{(X_1,\ldots,X_r|X_{r+1},\ldots,X_p)}(x_1,\ldots,x_r|x_{r+1},\ldots,x_p) = \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{X_{r+1},\ldots,X_p}(x_{r+1},\ldots,x_p)}$$

The above conditional density is interpreted as the joint density of $X_1, \ldots, X_r$ when $X_{r+1} = x_{r+1}, \ldots, X_p = x_p$ and is only defined when $f_{X_{r+1},\ldots,X_p}(x_{r+1},\ldots,x_p) \neq 0$.

We note that in case of mutual independence the p components, all conditional distributions do **not** depend on the conditions and it holds

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{p} F_{X_i}(x_i), f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{p} f_{X_i}(x_i).$$

Moments

Given the density $f_{\mathbf{X}}(\mathbf{x})$ of the random vector $\mathbf{X}$ the joint moments of order $s_1, s_2 \ldots, s_p$ are defined, in analogy to the univariate case, as

$$E(X_1^{s_1} \ldots X_p^{s_p}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{s_1} \ldots x_p^{s_p} f_{\mathbf{X}}(x_1,\ldots,x_p) dx_1 \ldots dx_p$$

Density transformation formula

Assume , the p existing random variables $X_1, X_2 \ldots, X_p$ with given density $f_{\mathbf{X}}(\mathbf{x})$ have been transformed by a smooth (i.e. differentiable) one-to-one transformation into p new random variables $Y_1, Y_2 \ldots, Y_p$, i.e. a new random vector $\mathbf{Y} \in \mathbf{R^p}$ has been created by calculating

$$Y_i = y_i(X_1, X_2 \ldots, X_p), i = 1, 2, \ldots, p$$

The question is how to calculate the density $g_{\mathbf{Y}}(\mathbf{y})$ of $\mathbf{Y}$ by knowing the transformation functions $y_i(X_1, X_2 \ldots, X_p), i = 1, 2, \ldots, p$ and the density $f_{\mathbf{X}}(\mathbf{x})$ of the original random vector. Since the transformation of the X's into Y's is assumed to be one-to-one, its inverse transformation $X_i = x_i(Y_1, Y_2 \ldots, Y_p), i = 1, 2, \ldots, p$ also exists and then the following density transformation formula applies:

$$g_{\mathbf{Y}}(y_1,\ldots,y_p) = f_{\mathbf{X}}[x_1(y_1,\ldots,y_p),\ldots,x_p(y_1,\ldots,y_p)]|J(y_1,\ldots,y_p)|$$

where $J(y_1, \ldots, y_p)$ is the *Jacobian* of the transformation:

$$J(y_1,\ldots,y_p) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \ldots & \frac{\partial x_1}{\partial y_p} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \ldots & \frac{\partial x_2}{\partial y_p} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial x_p}{\partial y_1} & \frac{\partial x_p}{\partial y_2} & \ldots & \frac{\partial x_p}{\partial y_p} \end{vmatrix}$$

Multivariate Normal Distribution

For the purpose of this course, we will only need the density for the case of non-degenerated multivariate normal. The formula generalizes the formula from the univariate case. Looking at the term $(\frac{x-\mu}{\sigma})^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$ in the exponent of the well known

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2}, -\infty < x < \infty$$

for the univariate density, a natural way to generalize this is to *replace* it by $(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu)$. Here $\mu = E\mathbf{X} \in \mathbf{R^p}$ is the expected value of the random vector $\mathbf{X} \in \mathbf{R^p}$ and the matrix

$$\mathbf{\Sigma} = E(\mathbf{X}-\mu)(\mathbf{X}-\mu)' = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} \in \mathbf{M_{p,p}}$$

(assumed to be positive definite) is the *covariance matrix*. On the diagonals of $\mathbf{\Sigma}$ we get the *variances* of each of the p random variables whereas $\sigma_{ij}, i \neq j$ are the *covariances* between the ith and jth random variable. Sometimes, we also denote σ_{ii} by σ_i^2.

The final result is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{\frac{1}{2}}} e^{-(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu)/\mathbf{2}}, -\infty < x_i < \infty, i = 1, 2\dots, p$$