Statistics 2 - Notes

Dom Hutchinson

October 3, 2019

Contents

1 Estimation **2**
1.1 Introduction 2

2 The Likelihood Function **3**

3 Maximum Likelihood Estimates **4**

0 Appendix **5**
0.1 Notation 5
0.2 R 5
0.3 Probability Distributions 5

1 Estimation

1.1 Introduction

Definition 1.1 - *Probabiltiy Space,* $(\Omega, \mathcal{F}, \mathbb{P})$
A mathematical construct for modelling the real world. A *Probabilty Space* has three elements

i) Ω - Sample space.

ii) $\mathcal{F}$ - Set of events.

iii) $\mathbb{P}$ - Probability measure.

and most fulfil the following conditions

i) $\Omega \in \mathcal{F}$;

ii) $\forall\ A \in \mathcal{F} \implies A^c \in \mathcal{F}$;

iii) $\forall\ A_0, \dots, A_n \in \mathcal{F} \implies \left(\bigcup_i A_i\right) \in \mathcal{F}$;

iv) $\mathbb{P}(\Omega) = 1$; and,

v) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ for disjoint $A_1, A_2, \dots$ (Countable Additivity).

Definition 1.2 - *Random Variable*
A function which maps an event in the sample space to a value *e.g.* $X : \Omega \to \mathbb{R}$.

Remark 1.1 - *Probability Density Function for iid Random Variable Vector*
For $\mathbf{X} \sim f_n(\cdot; \theta)$ where each component of $\mathbf{X}$ is independent and identically distribution the probability density function of $\mathbf{X}$ is

$$f_n(\mathbf{x}; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Definition 1.3 - *Expectation*
The mean value for a random variable. For rv X

$$\mathbb{E}(X) := \sum_{x \in \chi} x f_X(x) \quad \& \quad \mathbb{E}(X) := \int_{\mathbb{R}} x f_X(x) dx$$

Theorem 1.1 - *Expection of a Function*
For a function $g : \mathbb{R} \to \mathbb{R}$ and rv X with pmf f_X

$$\mathbb{E}(g(X)) := \sum_{g(x) \in \chi} x f_X(x) \quad \& \quad \mathbb{E}(g(X)) := \int_{\mathbb{R}} g(x) f_X(x) dx$$

Theorem 1.2 - *Expectation of a Linear Operator*
For rv X with pmf f_X & $a, b \in \mathbb{R}$

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

Definition 1.4 - *Variance*
For rv X

$$\text{var}(X) := \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Theorem 1.3 - *Variance of a Linear Operator*
For rv X and $a, b \in \mathbb{R}$

$$\text{var}(aX + b) = a^2\text{var}(X)$$

Definition 1.5 - *Moment of a Random Variable*
For rv X the n^{th} moment of X is defined as $\mathbb{E}(X^n)$.
N.B. - $\mathbb{E}(X^n) \neq \mathbb{E}(X)^n$.

Definition 1.6 - *Covariance*
For rv X & Y

$$\text{cov}(X, Y) := \mathbb{E}\big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Theorem 1.4 - *Properties of Covaraince*
Let X & Y be independent random varaibles

i) $\text{cov}(X, X) = \text{var}(X)$;

ii) $\text{cov}(X, Y) = 0$

Theorem 1.5 - *Variance of two Random Variables with linear operators*

$$\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\text{cov}(X, Y)$$

Theorem 1.6 - *Independent Random Variables*
Random variables $X_1, \dots, X_n$ are independent iff

$$\mathbb{P}(X_1 \leq a_1, \dots, X_n \leq a_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq a_i) \ \forall \ a_1, \dots, a_n \in \mathbb{R}$$

2 The Likelihood Function

Definition 2.1 - *Likelihood Function*
Define $\mathbf{X} \sim f_n(\cdot; \theta^*)$ for some unknown $\theta^* \in \Theta$ and let $\mathbf{x}$ be an observation of $\mathbf{X}$.
A *Likelihood Function* is any function, $L(\cdot; \mathbf{x}) : \Theta \to [0, \infty)$, which is proportional to the PMF/PDF of the obeserved realisation $\mathbf{x}$.

$$L(\theta; \mathbf{x}) := C f_b(\mathbf{x}; \theta) \ \forall \ C > 0$$

N.B. Sometimes this is called the *Observed* Likelihood Function since it is dependent on observed data.

Definition 2.2 - *Log-Likelihood Function*
Let $\mathbf{X} \sim f_n(\cdot; \theta^*)$ for some unknown $\theta^* \in \Theta$ and $\mathbf{x}$ be an observation of $\mathbf{X}$.
The *Log-Likelihood Function* is the natural log of a *Likelihood Function*

$$\ell(\theta; \mathbf{x}) := \ln f_n(\mathbf{x}; \theta) + C, \ C \in \mathbb{R}$$

Theorem 2.1 - *Multidiensional Transforms*
Let $\mathbf{X}$ be a continuous random vector in $\mathbb{R}^n$ with PDF $f_\mathbf{X}$; $g : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous diferentiable bijection; and, $h := g^{-1}$.
Then $\mathbf{Y} = g(\mathbf{X})$ is a continuous random vector and its PDF is

$$f_\mathbf{Y}(\mathbf{y}) = f_\mathbf{X}(h(\mathbf{y})H_h(\mathbf{Y})$$

where

$$J_h := \left|\det\left(\frac{\partial h}{\partial \mathbf{y}}\right)\right|$$

Proposition 2.1 - *Invaraince of Likelihood Function by bijective transformation of the observations independent of θ*
Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be a bijetive transformation which is independent of θ; and $\mathbf{Y} := g(\mathbf{X})$.
Then $\mathbf{Y}$ is a random variable with PDF/PMG

$$f_{\mathbf{Y}}(\mathbf{y};\theta) \propto f_{\mathbf{X}}(g^{-1}(\mathbf{y});\theta)$$

Hence, if $\mathbf{y} = g(\mathbf{x})$ then $L_{\mathbf{Y}}(\theta;\mathbf{y}) \propto L_{\mathbf{X}}(\theta;\mathbf{x})$
Proof 2.2 - *Proposition 2.1*
Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be a bijective transformation which is independent of θ; $h := g^{-1}$; $\mathbf{X}, \mathbf{Y}$ be a rvs st $\mathbf{Y} := g(\mathbf{X})$.

i) *Discrete Case* - Consider the case when $\mathbf{X}$ is a discrete rv. Then

$$\begin{array}{rcl} f_{\mathbf{Y}}(y;\theta) & = & \mathbb{P}(\mathbf{Y} = \mathbf{y};\theta) \\ & = & \mathbb{P}(g^{-1}(\mathbf{Y}) = g^{-1}(\mathbf{y});\theta) \\ & = & \mathbb{P}(h(\mathbf{Y}) = h(\mathbf{y});\theta) \\ & = & \mathbb{P}(\mathbf{X} = h(\mathbf{y});\theta) \\ & = & f_{\mathbf{X}}(g^{-1}(\mathbf{y});\theta) \end{array}$$

ii) *Continuous Case* - Consider the case when $\mathbf{X}$ is a continuous rv.
Then, by **Theorem 2.1**

$$f_{\mathbf{Y}}(\mathbf{y};\theta) = f_{\mathbf{X}}(g^{-1}(\mathbf{y});\theta)J_{g^{-1}}(\mathbf{y})$$

Since $J_{g^{-1}}$ does not depend on θ this case is solved.

Thus in botoh cases $L_{\mathbf{Y}}(\theta;y) = f_{\mathbf{Y}}(y;\theta) \propto f_{\mathbf{X}}(g^{-1}(\mathbf{y});\theta) = L_{\mathbf{X}}(\theta;\mathbf{x})$.

3 Maximum Likelihood Estimates

Definition 3.1 - *Maximum Likelihood Estimate*
Let $\mathbf{X} \sim f_n(\cdot;\theta)$; and $\mathbf{x}$ be a realisation of $\mathbf{X}$.
The *Maximum Likelihood Estimate* is the value $\hat{\theta} \in \Theta$ st

$$\forall\ \theta \in \Theta\ f_n(\mathbf{x};\hat{\theta}) \geq f_n(\mathbf{x},\theta)$$

Equivalently

$$\forall\ \theta \in \Theta\ L(\hat{\theta};\mathbf{x}) \geq L(\theta;\mathbf{x}) \quad \text{or} \quad \ell(\hat{\theta};\mathbf{x}) \geq \ell(\theta;\mathbf{x})$$

i.e. $\hat{\theta}(\mathbf{x}) := \text{argmax}_\theta(L(\theta;\mathbf{x})$.

Remark 3.1 - *The Maximum Likelihood Estimate may* <u>*not*</u> *be unique*

0 Appendix

0.1 Notation

Notation	Denotes
$\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$	Scalar or vector parameter characterising a probability distribution
$\hat{\theta}$	Estimation for the value of the parameter θ
θ^*	True value of the paramter θ
$\mathbb{P}$	Probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$
Ω	Sample space
X	Scalar random variable
$\mathcal{F}$	Sigma field (Set of events)
χ	Support of rv XX. A set set X is definitely in it *i.e.* $\mathbb{P}(X \in \chi; \theta) = 1$
$\mathbf{X}$	Vector consiting of scalar random variables

0.2 R

Command	Result
hist(a)	Plots a histogram of the values in array a
mean(a)	Returns the mean value of array a
rbinom(s,n,p)	Samples n of $Bi(n, p)$ random variables
rep(v,n)	Produces an array of size n where each entry has value v
$x \leftarrow v$	Maps value v to variable x

0.3 Probability Distributions

Definition 0.1 - *Binomial Distribution*
Let X be a discrete random variable modelled by a *Binomial Distribution* with n events and rate of success p.

$$\begin{array}{rcl} p_X(k) & = & \binom{n}{k} p^k (1-p)^{n-k} \\ \mathbb{E}(X)np = & \& & Var(X) = np(1-p) \end{array}$$

Definition 0.2 - *Gamma Distribution*
Let T be a continuous randmo variable modelled by a *Gamma Distribution* with shape parameter α & scale parameter λ. Then

$$\begin{array}{rcl} f_T(x) & = & \dfrac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \text{for } x > 0 \\ \mathbb{E}(T) = \dfrac{\alpha}{\lambda} & \& & Var(T) = \dfrac{\alpha}{\lambda^2} \end{array}$$

N.B. $\alpha, \lambda > 0$.

Definition 0.3 - *Exponential Distribution*
Let T be a continuous random variable modelled by a *Exponential Distribution* with parameter λ. Then

$$\begin{array}{rcl} f_T(t) & = & \mathbb{1}\{t \geq 0\}.\lambda e^{-\lambda t} \\ F_T(t) & = & \mathbb{1}\{t \geq 0\}.\left(1 - e^{-\lambda t}\right) \\ \mathbb{E}(X) = \frac{1}{\lambda} & \& & Var(X) = \frac{1}{\lambda^2} \end{array}$$

N.B. Exponential Distribution is used to model the wait time between decays of a radioactive source.

Definition 0.4 - *Normal Distribution*
Let X be a continuous random variable modelled by a *Normal Distribution* with mean μ &

variance σ^2.
Then

$$\begin{array}{rcl} f_X(x) & = & \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ F_X(x) & = & \frac{1}{\sqrt{2\pi\sigma^2}}\int\limits_{-\infty}^{x} e^{-\frac{(y-\mu)^2}{2\sigma^2}}dy \\ M_X(\theta) & = & e^{\mu\theta+\sigma^2\theta^2(1/2)} \\ \mathbb{E}(X)=\mu & \& & Var(X)=\sigma^2 \end{array}$$

Definition 0.5 - *Poisson Distribution*
Let X be a discrete random variable modelled by a *Poisson Distribution* with parameter λ. Then

$$\begin{array}{rcl} p_X(k) & = & \frac{e^{-\lambda}\lambda^k}{k!} \qquad \text{For } k \in \mathbb{N}_0 \\ \mathbb{E}(X)=\lambda & \& & Var(X)=\lambda \end{array}$$

N.B. Poisson Distribution is used to model the number of radioactive decays in a time period.