

STATISTICAL INFERENCE

- A Quick Guide -

Huan Q. Bui

Colby College

PHYSICS & MATHEMATICS
Statistics

Class of 2021

April 29, 2020

Preface

Greetings,

This guide is based on SC482: Statistical Inference, taught by Professor Liam O'Brien. The guide consists of lecture notes and material from *Introduction to Mathematical Statistics, 8th edition* by Hogg, McKean, and Craig. A majority of the text will be reading notes and solutions to selected problems.

As this is intended only to be a reference source, I might not be as meticulous with my explanations as I have been in some other guides.

Enjoy!

Contents

Preface	2
1 Special Distributions	7
1.1 The Binomial and Related Distributions	8
1.1.1 Negative Binomial & Geometric Distribution	8
1.2 Multinomial Distribution	9
1.3 Hypergeometric Distribution	9
1.4 The Poisson Distribution	9
1.5 The Γ, χ^2, β distributions	10
1.5.1 The Γ and exponential distribution	10
1.5.2 The χ^2 distribution	11
1.5.3 The β distribution	11
1.6 The Normal distribution	12
1.6.1 Contaminated Normal	13
1.7 The Multivariate Normal	13
1.8 The t - and F -distributions	14
1.8.1 The t -distribution	14
1.8.2 The F -distribution	15
1.8.3 The Student's Theorem	16
2 Elementary Statistical Inferences	19
2.1 Sampling & Statistics	20
2.1.1 Point estimators	20
2.1.2 Histogram estimates of pmfs and pdfs	21
2.2 Confidence Intervals	23
2.2.1 CI for difference in means	24
2.2.2 CI for difference in proportions	25
2.3 Order Statistics	25
2.3.1 Quantiles	26
2.3.2 CI for quantiles	26
2.4 Introduction to Hypothesis Testing	27
2.5 Additional comments about statistical test	28
2.5.1 Observed Significance Level, p -value	28
2.6 Chi-Square Tests	28
2.7 The Method of Monte Carlo	31

2.7.1	Inverse Transform	31
2.7.2	Accept-Reject Generation Algorithm	32
2.7.3	Evaluating definite integrals	32
2.8	Bootstrapping	34
2.8.1	Bootstrapping for Hypothesis Testing	34
3	Consistency and Limiting Distributions	37
3.1	Convergence in Probability	37
3.1.1	Sampling and Statistic	38
3.2	Convergence in Distribution	38
3.2.1	Bounded in Probability	39
3.2.2	Δ -method	39
3.2.3	Moment Generating Function Technique	40
3.3	Central Limit Theorem	40
4	Maximum Likelihood Methods	43
4.1	Maximum Likelihood Estimation	44
4.2	Rao-Cramér Lower Bound and Efficiency	47
4.3	Maximum Likelihood Tests	52
4.4	Multiparameter Case: Estimation	55
4.5	Multiparameter Case: Testing	57
5	Sufficiency	59
5.1	Measures of Quality of Estimators	59
5.2	A Sufficient Statistic for a Parameter	60
5.3	Properties of a Sufficient Statistic	61
5.4	Completeness and Uniqueness	62
5.4.1	Completeness	62
5.4.2	Uniqueness	63
5.5	The Exponential Class of Distributions	64
5.6	Functions of a Parameter	65
5.7	The Case of Several Parameters	67
5.8	Minimal Sufficiency and Ancillary Statistics	68
5.8.1	Other types of families of distributions	68
5.9	Sufficiency, Completeness, and Independence	69
6	Optimal Test of Hypotheses	71
6.1	Most Powerful Tests	72
6.2	Uniformly Most Powerful Test	73
7	Problems	75
7.1	Problem Set 1	76
7.2	Problem Set 2	88
7.3	Problem set 3	102
7.4	Problem set 4	109
7.5	Problem set 5	117

CONTENTS 5

7.6 Problem set 6	129
7.7 Problem set 7	143
7.8 Problem set 8	156

Part 1

Special Distributions

1.1 The Binomial and Related Distributions

If we let the random variable X equal the number of observed successes in n independent Bernoulli trials, each with success probability of p , then X follows the binomial distribution.

A binomial pmf is given by

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots \\ 0, & \text{else} \end{cases} \quad (1.1)$$

Using the binomial expansion formula, we can easily check that

$$\sum_x p(x) = 1 \quad (1.2)$$

The mgf of a binomial distribution is obtained by:

$$M_{\text{bin}}(t) = E[e^{tx}] = \sum_x e^{tx} p(x) = [(1-p) + pe^t]^n \quad \forall t \in \mathbb{R} \quad (1.3)$$

With this, we can find the mean and variance for $p(x)$:

$$\mu = M'(0) = n, \quad \sigma^2 = M''(0) = np(1-p) \quad (1.4)$$

Theorem 1.1.1. Let X_1, X_2, \dots, X_m be independent binomial random variables such that $X_i \sim \text{bin}(n_i, p)$, $i = 1, 2, \dots, m$. Then

$$Y = \sum_{i=1}^m X_i \sim \text{bin}\left(\sum_{i=1}^m n_i, p\right) \quad (1.5)$$

Proof: We prove this via the mgf for Y . By independence, we have that

$$M_Y(t) = \prod_{i=1}^m (1 - p + pe^t)^{n_i} = (1 - p + pe^t)^{\sum_{i=1}^m n_i} \quad (1.6)$$

The mgf completely determines the distribution which Y follows, so we're done. \square

1.1.1 Negative Binomial & Geometric Distribution

Consider a sequence of independent Bernoulli trials with constant probability p of success. The random variable Y which denotes the total number of failures in this sequence before the r th success follows the negative binomial distribution.

A negative binomial pmf is given by

$$p_Y(t) = \begin{cases} \binom{(y+r-1)}{r-1} p^r (1-p)^y & y = 0, 1, 2, \dots \\ 0, & \text{else} \end{cases} \quad (1.7)$$

The mgf of this distribution is

$$M(t) = p^r [1 - (1-p)e^t]^{-r} \quad (1.8)$$

When $r = 1$, Y follows the geometric distribution, whose pmf is given by

$$p_Y(y) = p(1-p)^y, \quad y = 0, 1, 2, \dots \quad (1.9)$$

The mgf of this distribution is

$$M(t) = p[1 - (1-p)e^t]^{-1} \quad (1.10)$$

1.2 Multinomial Distribution

We won't worry about this for now.

1.3 Hypergeometric Distribution

We won't worry about this for now.

1.4 The Poisson Distribution

The Poisson distribution gives the probability of observing x occurrences of some rare events characterized by rate $\lambda > 0$. The pmf is given by

$$p(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{else} \end{cases} \quad (1.11)$$

We say a random parameter with the pmf of the form of $p(x)$ follows the Poisson distribution with parameter λ .

The mgf of a Poisson distribution is given by

$$M(t) = e^{-\lambda(e^t - 1)} \quad (1.12)$$

From here, we can find the mean and variance:

$$\mu = M'(0) = \lambda, \quad \sigma^2 = M''(0) = \lambda \quad (1.13)$$

Theorem 1.4.1. If X_1, \dots, X_n are independent random variables, each $X_i \sim \text{Poi}(\lambda_i)$, then

$$\boxed{Y = \sum_{i=1}^n X_i \sim \text{Poi}\left(\sum_{i=1}^n \lambda_i\right)} \quad (1.14)$$

Proof: We once again prove this via the mgf of Y :

$$M_Y(t) = \prod_{i=1}^n e^{\lambda_i(e^t - 1)} = e^{\sum_{i=1}^n \lambda_i(e^t - 1)} \quad (1.15)$$

□

1.5 The Γ, χ^2, β distributions

The gamma function of $\alpha > 0$ is given by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy, \quad (1.16)$$

which gives $\Gamma(1) = 1$ and $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

1.5.1 The Γ and exponential distribution

A continuous random variable $X \sim \Gamma(\alpha, \beta)$ where $\alpha > 0$ and $\beta > 0$ whenever its pdf is

$$\boxed{f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & 0 < x < \infty \\ 0, & \text{else} \end{cases}} \quad (1.17)$$

The mgf for X is obtained via the change of variable $y = x(1 - \beta t)/\beta$, where $t < 1/\beta$:

$$\boxed{M(t) = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x(1-\beta t)/\beta} dx = \frac{1}{(1 - \beta t)^\alpha}} \quad (1.18)$$

From here, we can find the mean and variance:

$$\boxed{\mu = M'(0) = \alpha\beta, \quad \sigma^2 = \alpha\beta^2} \quad (1.19)$$

The $\Gamma(1, \beta)$ distribution is a special case, and it is called the **exponential distribution** with parameter $1/\beta$.

Theorem 1.5.1. Let X_1, \dots, X_n be independent random variables, with $X_i \sim \Gamma(\alpha_i, \beta)$. Then

$$Y = \sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n \alpha_i, \beta\right) \quad (1.20)$$

Proof: Can you guess via which device we prove the statement above? \square

1.5.2 The χ^2 distribution

The χ^2 distribution is a special case of the gamma distribution where $\alpha = r/2, r \in \mathbb{N}^*$ and $\beta = 2$. If a continuous r.v. $X \sim \chi^2(r)$ then its pdf is

$$f(x) = \begin{cases} \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2}, & 0 < x < \infty \\ 0, & \text{else} \end{cases} \quad (1.21)$$

Its mgf is

$$M(t) = (1 - 2t)^{-r/2}, \quad t < \frac{1}{2} \quad (1.22)$$

Theorem 1.5.2. Let $X \sim \chi^2(r)$ and $k > -r/2$ be given. Then $E[X^k]$ exists and is given by

$$E[X^k] = \frac{2^k \Gamma(r/2 + k)}{\Gamma(r/2)} \quad (1.23)$$

Proof: is proof is purey computational and is left to the reader. \square

From here, we note that all moments of the χ^2 distribution exist.

Theorem 1.5.3. Let X_1, \dots, X_n be r.v. with $X_i \sim \chi^2(r_i)$. Then

$$Y = \sum_{i=1}^n X_i \sim \chi^2\left(\sum_{i=1}^n r_i\right) \quad (1.24)$$

Proof: we once again find the mgf for Y . \square

1.5.3 The β distribution

The β distribution differs from the other continuous ones we've discussed so far because its support are bounded intervals.

I will skip most of the details here, except mentioning that we can derive the beta distribution from a pair of independent Γ random variables. Suppose $Y = X_1/(X_1 + X_2)$ where $X_i \sim \Gamma(\alpha, \beta)$ then the pdf of Y is that of the beta distribution:

$$g(y) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1}, & 0 < y < 1 \\ 0, & \text{else} \end{cases} \quad (1.25)$$

The mean and variance of Y are

$$\mu = \frac{\alpha}{\alpha + \beta}, \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2} \quad (1.26)$$

1.6 The Normal distribution

I have dedicated a large chunk in the [QFT](#) notes to evaluating Gaussian integrals, so I won't go into that here.

$X \sim \mathcal{N}(\mu, \sigma^2)$ whenever its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right), \quad -\infty < x < \infty \quad (1.27)$$

where μ and σ^2 are the mean and variance of X , respectively.

The mgf of X can be obtained via the substitution $X = \sigma Z + \mu$:

$$M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \quad (1.28)$$

We note the following correspondence for $X = \sigma Z + \mu$:

$$X \sim \mathcal{N}(\mu, \sigma^2) \iff Z \sim \mathcal{N}(0, 1) \quad (1.29)$$

Theorem 1.6.1. $X \sim \mathcal{N}(\mu, \sigma^2) \implies V = (X - \mu)^2 / \sigma^2 \sim \chi^2(1)$, i.e. a standardized, squared normal follows a chi-square distribution.

Proof: The proof isn't too hard. Let us write V as W^2 and so $W \sim \mathcal{N}(0, 1)$. We consider the cdf $G(v)$ for V , with $v \geq 0$:

$$G(v) = P(W^2 \leq v) = P(-\sqrt{v} \leq W \leq \sqrt{v}) = 2 \int_0^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw \quad (1.30)$$

with $G(v) = 0$ whenever $v < 0$. From here, we can see that the pdf for v , under the change of notation $w \rightarrow \sqrt{y}$, is

$$g(v) = G'(v) = \frac{d}{dv} \left\{ \int_0^v \frac{1}{\sqrt{2\pi}\sqrt{y}} e^{-y/2} dy \right\}, \quad 0 \geq v \quad (1.31)$$

or 0 otherwise. This means

$$g(v) = \begin{cases} \frac{1}{\sqrt{\pi}\sqrt{2}} v^{1/2-1} e^{-v/2}, & 0 < v < \infty \\ 0, & \text{else} \end{cases} \quad (1.32)$$

Using the fact that $\Gamma(1/2) = \sqrt{\pi}$ and by verifying that $g(v)$ integrates to unity we show $V \sim \chi^2(1)$. \square

Theorem 1.6.2. Let X_1, \dots, X_n be independent r.v. with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Then for constants a_1, \dots, a_n

$$Y = \sum_{i=1}^n a_i X_i \sim \mathcal{N} \left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2 \right) \quad (1.33)$$

Proof: We once again prove this kind of theorems via the mgf for Y :

$$\begin{aligned} M(t) &= \prod_{i=1}^n \exp \left(t a_i \mu_i + \frac{1}{2} a_i^2 \sigma_i^2 \right) \\ &= \exp \left\{ t \sum_{i=1}^n a_i \mu_i + \frac{1}{2} t^2 \sum_{i=1}^n a_i^2 \sigma_i^2 \right\} \end{aligned} \quad (1.34)$$

which is the mgf for the normal with the corresponding mean and variance above. \square

Corollary: Let $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Then

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \sim \mathcal{N}(\mu, \sigma^2/n) \quad (1.35)$$

Proof: the proof is left to the reader.

1.6.1 Contaminated Normal

We won't worry about this for now.

1.7 The Multivariate Normal

I'll just jump straight to the n -dimensional generalization. Evaluations of high-dimensional Gaussian integrals and moments can also be found in the [QFT](#) notes.

We say an n -dimensional random vector \mathbf{X} has a multivariate normal distribution if its mgf is

$$M_{\mathbf{X}}(t) = \exp \left(\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t} \right) \quad (1.36)$$

for all $\mathbf{t} \in \mathbb{R}^n$, where $\boldsymbol{\Sigma}$ is a symmetric, positive semi-definite matrix and $\boldsymbol{\mu} \in \mathbb{R}^n$. For short, we say $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Theorem 1.7.1. Suppose $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is positive definite. Then

$$\mathbf{Y} = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(1) \quad (1.37)$$

Theorem 1.7.2. If $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then

$$\mathbf{Y} = \mathbf{AX} + \mathbf{b} \sim \mathcal{N}_n(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}) \quad (1.38)$$

Proof: The proof once again uses the mgf for \mathbf{Y} , but also some linear algebra manipulations. \square

There are many other theorems and results related to this topic, but I won't go into them for now.

1.8 The t - and F -distributions

These two distributions are useful in certain problems in statistical inference.

1.8.1 The t -distribution

Suppose $W \sim \mathcal{N}(0, 1)$ and $V \sim \chi^2(r)$ and that they are independent. Then the joint pdf of W and V , called $h(w, v)$, is the product of the pdf's of W and V :

$$h(w, v) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \frac{1}{\Gamma(r/2)2^{r/2}} v^{r/2-1} e^{-v/2}, & w \in \mathbb{R}, v > 0 \\ 0, & \text{else} \end{cases} \quad (1.39)$$

Now we define a new variable $T = W/\sqrt{V/r}$ and consider the transformation:

$$t = \frac{w}{\sqrt{v/r}} \quad u = v \quad (1.40)$$

which bijectively maps the parameter space $(w, v) = \mathbb{R} \times \mathbb{R}^+$ to $(t, u) = \mathbb{R} \times \mathbb{R}^+$. The absolute value of the Jacobian of the transformation is given by

$$|J| = \left| \det \begin{pmatrix} \partial_t w & \partial_u w \\ \partial_t v & \partial_u v \end{pmatrix} \right| = \frac{\sqrt{u}}{\sqrt{r}}. \quad (1.41)$$

With this, the joint pdf of T and $U \equiv V$ is given by

$$g(t, u) = |J| h\left(\frac{t\sqrt{u}}{\sqrt{r}}, u\right) = \begin{cases} \frac{u^{r/2-1}}{\sqrt{2\pi}\Gamma(r/2)2^{r/2}} \exp\left[-\frac{u}{2}\left(1+\frac{t^2}{r}\right)\right] \frac{\sqrt{u}}{\sqrt{r}}, & t \in \mathbb{R}, u \in \mathbb{R}^+ \\ 0, & \text{else} \end{cases} \quad (1.42)$$

By integrating out u we obtain the marginal pdf for T :

$$\begin{aligned} g_1(t) &= \int_{-\infty}^{\infty} g(t, u) du \\ &= \int_0^{\infty} \frac{u^{(r+1)/2-1}}{\sqrt{2\pi r}\Gamma(r/2)2^{r/2}} \exp\left[-\frac{u}{2}\left(1+\frac{t^2}{r}\right)\right] du. \end{aligned} \quad (1.43)$$

Via the substitution $z = u[1 + (t^2/r)]/2$ we can evaluate the integral to find for $t \in \mathbb{R}$

$$g_1(t) = \boxed{\frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}}} \quad (1.44)$$

A r.v. T with this pdf is said to follow the t -distribution (or the Student's t -distribution) with r degrees of freedom. The t -distribution is symmetric about 0 and has a unique maximum at 0. As $r \rightarrow \infty$, the t -distribution converges to $\mathcal{N}(0, 1)$.

The mean of $T \sim \text{Stu}(r)$ is zero. The variance can be found to be $\text{Var}(T) = E[T^2] = \frac{r}{r-2}$, so long as $r > 2$.

1.8.2 The F-distribution

Let $U \sim \chi^2(r_1), V \sim \chi^2(r_2)$ be given. Then the joint pdf of U and V is once again the product of their pdf's:

$$h(u, v) = \begin{cases} \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} u^{r_1/2-1} v^{r_2/2-1} e^{-(u+v)/2}, & u, v \in \mathbb{R}^+ \\ 0, & \text{else} \end{cases} \quad (1.45)$$

Define the new random variable

$$W = \frac{U/r_1}{V/r_2} \quad (1.46)$$

whose pdf $g_1(w)$ we are interested in finding. Consider the transformation

$$w = \frac{u/r_1}{v/r_2}, \quad z = v \quad (1.47)$$

which bijectively maps $(u, v) = \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow (w, z) = [\mathbb{R}^+ \times \mathbb{R}^+]$. Like last time, the absolute value of the Jacobian can be found to be

$$|J| = \frac{r_1}{r_2} z. \quad (1.48)$$

The joint pdf $g(w, z)$ of the random variables W and $Z = V$ is obtained from by scaling $h(u, v)$ by $|J|$ and applying the variable transformation:

$$g(w, z) = \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2)2^{\frac{r_1+r_2}{2}}} \left(\frac{r_1 zw}{r_2} \right)^{\frac{r_1-2}{2}} z^{\frac{r_2-2}{2}} \exp \left[-\frac{z}{2} \left(\frac{r_1 w}{r_2} + 1 \right) \right] \frac{r_1 z}{r_2} \quad (1.49)$$

so long as $(w, z) \in \mathbb{R}^+ \times \mathbb{R}^+$ and 0 otherwise. We then proceed to find the marginal pdf $g_1(w)$ of W by integrating out z . By considering the change of variables:

$$y = \frac{z}{2} \left(\frac{r_1 w}{r_2} + 1 \right) \quad (1.50)$$

we can evaluate the integral and find the marginal pdf of W to be

$$g_1(w) = \begin{cases} \frac{\Gamma[(r_1+r_2)/2](r_1/r_2)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)} \frac{w^{r_1/2-1}}{(1+r_1 w/r_2)^{(r_1+r_2)/2}}, & w \in \mathbb{R}^+ \\ 0, & \text{else} \end{cases} \quad (1.51)$$

W , which is the ratio of two independent chi-square variables U, V , is said to follow an F -distribution with degrees of freedom r_1 and r_2 . We call the ratio $W = (U/r_1)/(V/r_2)$ the “ F ” ratio.

The mean of W is $E[F] = \frac{r_2}{r_2-2}$. When r_2 is large, $E[F] \rightarrow 1$.

1.8.3 The Student's Theorem

Here we will create the connection between the normal distribution and the t -distribution. This is an important result for the later topics on inference for normal random variables.

Theorem 1.8.1. Let X_1, \dots, X_n be iid r.v. with $X_i \sim \mathcal{N}(\mu, \sigma^2) \forall i$. Define the r.v.’s

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1.52)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (1.53)$$

Then

- (a) $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
- (b) \bar{X} and S^2 are independent.
- (c) $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.
- (d) The variable $\bar{T} = (\bar{X} - \mu)/(S/\sqrt{n})$ follows the Student's t -distribution with $n-1$ degrees of freedom.

$$\bar{T} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim \text{Stu}(n-1). \quad (1.54)$$

Proof: The proof is good, so I will reproduce it here. Because $X_i \sim \mathcal{N}(\mu, \sigma^2) \forall i$, $\mathbf{X} \sim \mathcal{N}_n(\mu\mathbf{1}, \sigma^2\mathbf{I})$, where $\mathbf{1}$ denotes the n -vector whose components are all 1. Now, consider $\mathbf{v}^\top = (1/n)\mathbf{1}^\top$. We see that $\bar{X} = \mathbf{v}^\top \mathbf{X}$. Define the random vector $\mathbf{Y} = (X_1 - \bar{X}, \dots, X_n - \bar{X})^\top$ and consider the (true) equality:

$$\mathbf{W} = \begin{pmatrix} \bar{X} \\ \mathbf{Y} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{v}^\top \\ \mathbb{I} - \mathbf{1}\mathbf{v}^\top \end{pmatrix}}_{\text{the transformation}} \mathbf{X} \quad (1.55)$$

which just restates our definitions nicely. We see that \mathbf{W} is a result of a linear transformation of multivariate normal random vector, and so it follows that $\mathbf{W} \sim \mathcal{N}_{n+1}$ with mean

$$E[\mathbf{W}] = \begin{pmatrix} \mathbf{v}^\top \\ \mathbb{I} - \mathbf{1}\mathbf{v}^\top \end{pmatrix} \mu\mathbf{1} = \begin{pmatrix} \mu \\ \mathbf{0}_n \end{pmatrix} \quad (1.56)$$

and the covariance matrix

$$\Sigma = \begin{pmatrix} \mathbf{v}^\top \\ \mathbb{I} - \mathbf{1}\mathbf{v}^\top \end{pmatrix} \sigma^2 \mathbb{I} \begin{pmatrix} \mathbf{v}^\top \\ \mathbb{I} - \mathbf{1}\mathbf{v}^\top \end{pmatrix}^\top = \sigma^2 \begin{pmatrix} \frac{1}{n} & \mathbf{0}_n^\top \\ \mathbf{0}_n & \mathbb{I} - \mathbf{1}\mathbf{v}^\top \end{pmatrix} \quad (1.57)$$

From here, part (a) is proven. Next, observe that Σ is diagonal, and so all covariances are zero. This means \bar{X} is independent of \mathbf{Y} . But because $S^2 = (n-1)^{-1}\mathbf{Y}^\top \mathbf{Y}$, \bar{X} is independent of S^2 as well. So, (b) is proven.

Now, consider the r.v.

$$V = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \quad (1.58)$$

Each summand of V is a square of an $\mathcal{N}(0, 1)$ r.v., and so each follows a $\chi^2(1)$. Because V is a sum of squares of n such $\chi^2(1)$'s, $V \sim \chi^2(n)$. Next, we can rewrite V as

$$V = \sum_{i=1}^n \left(\frac{(X_i - \bar{X}) + (\bar{X} - \mu)}{\sigma} \right)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2. \quad (1.59)$$

By (b), the summands in the last equation are independent. The second term is a square of a $\mathcal{N}(0, 1)$, so it follows a $\chi^2(1)$. Taking mgfs of both sides, we get

$$(1 - 2t)^{-n/2} = \underbrace{E[\exp\{t(n-1)S^2/\sigma^2\}]}_{M(c)}(1 - 2t)^{-1/2}. \quad (1.60)$$

Solving for the mgf of $(n-1)S^2/\sigma^2$ we get part (c). Finally, writing T as

$$T = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{(n-1)S^2/(\sigma^2(n-1))}} \quad (1.61)$$

and using (a)-(c) gives us (d). *Hint:* consider what distributions the numerator and denominator of T follow. \square

Part 2

Elementary Statistical Inferences

2.1 Sampling & Statistics

In statistical inferences, our ignorance about the pdf/pmf of a random variable X can be classified in two ways:

- The pdf/pmf is unknown.
- The pdf/pmf is assumed/known but its parameter vector θ is not.

We consider the second class of classification for now.

Definition: If the random variables X_1, X_2, \dots, X_n are iid, then these random variables constitute a **random sample** of size n from the common distribution.

Definition: Let X_1, \dots, X_n denote a sample on a random variable X . Let $T = T(X_1, \dots, X_n)$ be a function of the sample. Then T is called a **statistic**.

2.1.1 Point estimators

Definition: (*Unbiasedness*) Let X_1, \dots, X_n denote a sample on a random variable X with pdf $f(x; \theta)$, $\theta \in \Omega$. Let $T = T(X_1, \dots, X_n)$ be a statistic. We say that T is an *unbiased estimator* of θ if $E[T] = \theta$.

We now introduce the concept of the **maximum likelihood estimator (mle)**. The information in the sample and the parameter θ are involved in the joint distribution of the random sample. We write this as

$$L(\theta) = L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta). \quad (2.1)$$

This is called the **likelihood function** of the random sample. A measure of the center of $L(\theta)$ seems to be an appropriate estimate of θ . We often use the value of θ at which $L(\theta)$ is maximized. If this value is unique, then it is called the **maximum likelihood estimator (mle)**, denoted as $\hat{\theta}$:

$$\hat{\theta} = \text{Argmax} L(\theta). \quad (2.2)$$

We often work with the log of the likelihood in practice, which is the function $l(\theta) = \log(L(\theta))$. The logarithm is a strictly increasing function, so its maximum is obtained exactly when the maximum of $L(\theta)$ is obtained. In most models, the pdf and pmf are differentiable functions of θ , in which cases $\hat{\theta}$ solves the equation:

$$\partial_\theta l(\theta) = 0 \quad (2.3)$$

This is equivalent to saying $\hat{\theta}$ maximizes $l(\theta)$. If θ is a vector of parameters, this results in a system of equations to be solved simultaneously. These equations are called the **estimating equations**, (EE).

2.1.2 Histogram estimates of pmfs and pdfs

Let X_1, \dots, X_n be a random sample on a random variable X with cdf $F(x)$. A histogram of the sample is an estimate of the pmf or pdf depending on whether X is discrete or continuous. We make no assumptions on the form of the distribution of X . In particular, we don't assume the parametric form of the distribution, hence the histogram is often called the **nonparametric** estimator.

The distribution of X is discrete

Assume X is a discrete r.v. with pmf $p(x)$. Consider a sample X_1, \dots, X_n . Suppose $X \in \mathcal{D} = \{a_1, \dots, a_n\}$, then intuitively the estimate of $p(a_j)$ is the relative frequency of a_j . More formally, for $j = 1, \dots, m$ we define the statistic

$$I_j(X_i) = \begin{cases} 1 & X_i = a_j \\ 0 & X_i \neq a_j \end{cases} \quad (2.4)$$

Then the estimate of $p(a_j)$ is the average

$$\hat{p}(a_j) = \frac{1}{n} \sum_{i=1}^n I_j(X_i) \quad (2.5)$$

The estimators $\{\hat{p}(a_1), \dots, \hat{p}(a_m)\}$ constitute the nonparametric estimate of the pmf $p(x)$. We note that $I_j(X_i)$ has a Bernoulli distribution with probability $p(a_j)$, and so

$$E[\hat{p}(a_j)] = \frac{1}{n} \sum_{i=1}^n E[I_j(X_i)] = \frac{1}{n} \sum_{i=1}^n p(a_j) = p(a_j), \quad (2.6)$$

which means $\hat{p}(a_j)$ is an *unbiased estimator* of $p(a_j)$.

Now, suppose that the space of X is infinite, i.e., $\mathcal{D} = \{a_1, \dots\}$ then in practice we select a value, say a_m , and make the groupings

$$\{a_1\}, \{a_2\}, \dots, \{a_m\}, \tilde{a}_{m+1} = \{a_{m+1}, \dots\} \quad (2.7)$$

Let $\hat{p}(\tilde{a}_{m+1})$ be the proportion of the sample items that are greater than or equal to a_{m+1} . Then the estimates $\{\hat{p}(a_1), \dots, \hat{p}(a_{m+1})\}$ form our estimate of $p(x)$. To merge groups, the rule of thumb is to select m so that the frequency of the category a_m exceeds twice the combined frequencies of the categories a_{m+1}, a_{m+2}, \dots

A histogram is a *barplot* pf $\hat{p}(a_j)$ versus a_j . When a_j contains no ordinal information (e.g. hair colors, etc) then such histograms consist of nonabutting bars and are called *bar charts*. When the space \mathcal{D} is ordinal, then the histograms is an abutting bar chart plotted in the natural order of the a_j 's.

The distribution of X is continuous

Assume X is a continuous r.v. with pdf $f(x)$. Consider a sample X_1, \dots, X_n . We first sketch an estimate for this pdf at a specified value of x . For a given $h > 0$, we consider the interval $(x-h, x+h)$. By MVT, we have for $\xi, |x - \xi| < h$:

$$P(|X - x| < h) = \int_{x-h}^{x+h} f(t) dt \approx 2hf(x). \quad (2.8)$$

The LHS is the proportion of the sample items that fall in the interval $(x - h, x + h)$. This suggests the use of the estimate of $f(x)$ at a given x :

$$\hat{f}(x) = \frac{\#\{|X_i - x| < h\}}{2hn}. \quad (2.9)$$

More formally, the indicator statistic is, for $i = 1, \dots, n$

$$I_i(x) = \begin{cases} 1 & x - h < X_i < x + h \\ - & \text{else} \end{cases}, \quad (2.10)$$

from which we obtain the nonparametric estimator of $f(x)$:

$$\hat{f}(x) = \frac{1}{2hn} \sum_{i=1}^n I_i(x). \quad (2.11)$$

Since the sample items are iid:

$$E[\hat{f}(x)] = \frac{1}{2hn} nf(\xi) 2h = f(\xi) \rightarrow f(x) \quad \text{as } h \rightarrow 0. \quad (2.12)$$

Therefore $\hat{f}(x)$ is *approximately* (as opposed to *exact* in the discrete case) an unbiased estimator of $f(x)$. I_i is called the **rectangular kernel** with **bandwidth** $2h$.

Provided realized values x_1, \dots, x_n of the random sample of X with pdf $f(x)$, there are many ways to obtain a histogram estimate of $f(x)$. First, select an integer m , an $h > 0$, and a value $a < \min(x_i)$, so that the m intervals cover the range of the sample. These intervals form our classes. Let $A_j = (a + (2j-3)h, a + (2j-1)h]$ for $j = 1, \dots, m$. Let $\hat{f}_h(x)$ denote our histogram estimate. For $a - h < x \leq (2m-1)h$, x is in one and only one A_j . Then for $x \in A_j$, we define

$$\hat{f}_h(x) = \frac{\#\{x_i \in A_j\}}{2hn} \geq 0. \quad (2.13)$$

We see that

$$\begin{aligned}
\int_{-\infty}^{\infty} \hat{f}_h(x) dx &= \int_{a-h}^{a+(2m-1)h} \hat{f}_h(x) dx \\
&= \sum_{j=1}^m \int_{A_j} \frac{\#\{x_i \in A_j\}}{2hn} dx \\
&= \frac{1}{2hn} \sum_{j=1}^m \#\{x_i \in A_j\} [h(2j-1 - 2j+3)] \\
&= \frac{2h}{2hn} n \\
&= 1.
\end{aligned} \tag{2.14}$$

So $\hat{f}_h(x)$ satisfies the properties of a pdf.

2.2 Confidence Intervals

Definition: Let X_1, \dots, X_n be a sample on a r.v. X which has the pdf $f(x; \theta), \theta \in \Omega$. Let $0 < \alpha < 1$ be specified. Let $L = L(X_1, \dots, X_n)$ and $U = U(X_1, \dots, X_n)$ be two statistics. We say that the interval (L, U) is a $(1 - \alpha)100\%$ **confidence interval** for θ if

$$1 - \alpha \equiv P_{\theta}[\theta \in (L, U)]. \tag{2.15}$$

That is, the probability that the interval includes θ is $1 - \alpha$, which is called the **confidence coefficient** or the **confidence level** of the interval.

Under normality, the confidence interval for μ is given by

$$\boxed{(\bar{x} - t_{\alpha/2, n-1}s/\sqrt{n}, \bar{x} + t_{\alpha/2, n-1}s/\sqrt{n})} \tag{2.16}$$

where $t_{\alpha/2, n-1}$ is the upper $\alpha/2$ critical points of a t -distribution with $n - 1$ df. This CI is referred to as the $(1 - \alpha)100\%$ **t -interval** for μ . s is referred to as the **standard error** of \bar{X} .

The Central Limit Theorem: Let X_1, \dots, X_n denote the observations of a random sample from a distribution that has mean μ and finite variance σ^2 . Then the distribution function of the r.v. $W_n = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ converges to Φ , the distribution function of the $\mathcal{N}(0, 1)$ distribution, as $n \rightarrow \infty$.

When the sample is large, the CI for μ can be given by

$$\boxed{(\bar{x} - z_{\alpha/2}s/\sqrt{n}, \bar{x} + z_{\alpha/2}s/\sqrt{n})} \tag{2.17}$$

In general, for the same α , the t -CI is larger (and hence more conservative) than the z -CI. When σ is known, we replace s by σ .

The larger sample CI for p is given by

$$\boxed{(\hat{p} - z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n})} \quad (2.18)$$

where $\sqrt{\hat{p}(1-\hat{p})/n}$ is called the standard error of \hat{p} .

2.2.1 CI for difference in means

By independence of samples,

$$\text{Var}(\hat{\Delta}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \quad (2.19)$$

where $\hat{\Delta} = \bar{X} - \bar{Y}$. We can readily show that $\hat{\Delta}$ is an unbiased estimator of $\Delta = \mu_1 - \mu_2$. Let the sample variances

$$S_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_i - \bar{X})^2 \quad (2.20)$$

be given. Then the random variable follows the $\mathcal{N}(0, 1)$:

$$\boxed{Z = \frac{\hat{\Delta} - \Delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \mathcal{N}(0, 1)} \quad (2.21)$$

The approximate $(1 - \alpha)100\%$ CI for $\Delta = \mu_1 - \mu_2$ is then given by

$$\left((\bar{x} - \bar{y}) - z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, (\bar{x} - \bar{y}) + z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right) \quad (2.22)$$

This is a large sample $(1 - \alpha)100\%$ CI for $\mu_1 - \mu_2$.

Now, suppose $X \sim \mathcal{N}(\mu_1, \sigma^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma^2)$ (i.e. X and Y are normally distributed with the same variance) are independent. We want to show $\Delta \sim t$ -distribution. We know that $\bar{X} \sim \mathcal{N}(\mu_1, \sigma^2/n_1)$ and $\bar{Y} \sim \mathcal{N}(\mu_2, \sigma^2/n_2)$, so it is true that

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{1/n_1 + 1/n_2}} \sim \mathcal{N}(0, 1). \quad (2.23)$$

This quantity will later be the numerator of our T -statistic. Now, let

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (2.24)$$

then S_p^2 , the **pooled estimator** of σ^2 , is also an unbiased estimator of σ^2 . Because $(n_i - 1)S_i^2/\sigma^2 \sim \chi^2(n - 1)$, we have that $(n - 2)S_p^2/\sigma^2 \sim \chi^2(n - 2)$. And so

$$T = \frac{\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{1/n_1 + 1/n_2}}}{\sqrt{(n - 2)S_p^2/(n - 2)\sigma^2}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t_{n-2} \quad (2.25)$$

From here, it is easy to work out the $(1 - \alpha)100\%$ CI for $\mu_1 - \mu_2$:

$$\left((\bar{x} - \bar{y}) - t_{\alpha/2, n-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{x} - \bar{y}) + t_{\alpha/2, n-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) \quad (2.26)$$

There is some difficulty when the unknown variances σ in the distributions of X and Y are not equal.

2.2.2 CI for difference in proportions

Our estimator of the difference in proportions $p_1 - p_2$ is $\bar{X} - \bar{Y} \equiv \hat{p}_1 - \hat{p}_2$ where $X \sim b(1, p_1)$ and $Y \sim b(1, p_2)$. Of course, we know that $\sigma_1^2 = p_1(1 - p_1)$ and $\sigma_2^2 = p_2(1 - p_2)$. From here, the approximate $(1 - \alpha)100\%$ confidence interval for $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \quad (2.27)$$

2.3 Order Statistics

Let X_1, X_2, \dots, X_n denote a random sample from a distribution of the continuous type having a pdf $f(x)$ that has support $S = (a, b)$, where $\infty \leq a < b \leq \infty$. Let $Y_1 < Y_2 < \dots < Y_n$ represent X_1, X_2, \dots, X_n when the latter are arranged in ascending order of magnitude. We call $Y_i, i = 1, 2, \dots, n$, the i th order statistic of the random sample X_1, X_2, \dots, X_n . We have a theorem which gives the joint pdf of Y_1, Y_2, \dots, Y_n .

Theorem 2.3.1. The joint pdf of Y_1, Y_2, \dots, Y_n is given by

$$g(y_1, y_2, \dots, y_n) = \begin{cases} n!f(y_1)\dots f(y_n) & a < y_1 < \dots < y_n < b \\ 0 & \text{else} \end{cases}. \quad (2.28)$$

Proof: The support of X_1, \dots, X_n can be partitioned into $n!$ mutually disjoint sets that map onto the support of the Y_i 's. Obviously the Jacobian for

each transformation is either 1 or -1 .

$$\begin{aligned} g(y_1, y_2, \dots, y_n) &= \sum_{i=1}^{n!} |J_i| f(y_1) \dots f(y_n) \\ &= \begin{cases} n! f(y_1) \dots f(y_n) & a < y_1 < \dots < y_n < b \\ 0 & \text{else} \end{cases} \end{aligned} \quad (2.29)$$

2.3.1 Quantiles

Let X be a random variable with a continuous cdf $F(x)$. For $0 < p < 1$, define the p th quantile of X to be $\xi_p = F^{-1}(p)$. Let X_1, X_2, \dots, X_n be a random sample from the distribution of X and let $Y_1 < Y_2 < \dots < Y_n$ be the corresponding order statistics. Let k be the greatest integer less than or equal to $p(n+1)$. We next define an estimator of ξ_p after making the following observation. The area under the pdf $f(x)$ to the left of Y_k is $F(Y_k)$. The expected value of this area is

$$E[F(Y_k)] = \int_a^b F(y_k) g_k(y_k) dy_k \quad (2.30)$$

where $g_k(y_k)$ is the pdf of Y_k . Consider the transformation $z = F(y_k)$, then the integral becomes

$$E[F(Y_k)] = \int_0^1 \frac{n!}{(k-1)!(n-k)!} z^k (1-z)^{n-k} dz = \dots = \frac{k}{n+1} \quad (2.31)$$

where we recognize the similarity between the integral and the integral of a beta pdf. So, on the average, there is $k/(n+1)$ of the total area to the left of Y_k . Because $p = k/(n+1)$, it seems reasonable to take Y_k as an estimator of the quantile ξ_p . Hence, we call Y_k the **p th sample quantile**. It is also called the **$100p$ th percentile** of the sample.

A **five-number** summary of the data consists of the following five sample quantiles: the minimum (Y_1), the first quartile ($Y_{.25(n+1)}$), the median, the third quartile ($Y_{.75(n+1)}$), and the maximum (Y_n). For this section, we use the notation Q_1 , Q_2 , and Q_3 to denote, respectively, the first quartile, median, and third quartile of the sample.

The five-number summary is the basis for a useful and quick plot of the data. This is called a **boxplot** of the data. In the **box and whisker** plots, we also define a potential outlier. Let $h = 1.5(Q_3 - Q_1)$. The **lower/upper fence** is defined by $L/UF = Q_{1/3} \mp h$. Points lying outside the (LF, UF) interval are called **potential outliers**.

2.3.2 CI for quantiles

Let X be a continuous random variable with cdf $F(x)$. For $0 < p < 1$, define the $100p$ th distribution percentile to be ξ_p , where $F(\xi_p) = p$. For a sample of

size n on X , let $Y_1 < Y_2 < \dots < Y_n$ be the order statistics. Let $k = [(n+1)p]$. Then the 100 p th sample percentile Y_k is a point estimate of ξ_p .

Let $i < [(n+1)p] < j$, and consider the order statistics $Y_i < Y_j$ and the event $Y_i < \xi_p < Y_j$. The event $Y_i < \xi_p < Y_j$ is equivalent to obtaining between i (inclusive) and j (exclusive) successes in n independent trials. So,

$$P(Y_i < \xi_p < Y_j) = \sum_{w=i}^{j-1} \binom{n}{w} p^w (1-p)^{n-w}. \quad (2.32)$$

For the median, we denote $\xi_{1/2}$ the median of $F(x)$, i.e. $\xi_{1/2}$ solves $F(x) = 1/2$. Let Q_2 denote the sample median, which is a point estimator of $\xi_{1/2}$. Take $c_{\alpha/2}$ such that $P[S \leq c_{\alpha/2}] = \alpha/2$ where $S \sim b(n, 1/2)$. Then note also that $P[S \leq c_{\alpha/2}] = \alpha/2$. From here we have

$$P[Y_{c_{\alpha/2}} < \xi_{1/2} < Y_{n-c_{\alpha/2}}] = 1 - \alpha. \quad (2.33)$$

So, if $y_{\alpha/2+1}$ and $y_{n-\alpha/2}$ are the realized values of the order statistics $Y_{c_{\alpha/2}+1}$ and $Y_{n-c_{\alpha/2}}$ then the interval

$$(y_{c_{\alpha/2}}, y_{n-c_{\alpha/2}}) \quad (2.34)$$

is a $(1 - \alpha)100\%$ confidence interval for $\xi_{1/2}$.

2.4 Introduction to Hypothesis Testing

Suppose a r.v. $X \sim f(x; \theta)$, where $\theta \in \Omega$. Suppose that $\theta \in \omega_0$ or $\theta \in \omega_1$ where ω_0 and ω_1 are disjoint subsets of Ω and $\omega_0 \cup \omega_1 = \Omega$. We label these hypotheses as

$$\begin{aligned} H_0 : \theta &\in \omega_0 \\ H_1 : \theta &\in \omega_1. \end{aligned} \quad (2.35)$$

H_0 is called the **null hypothesis**. H_1 is called the **alternative hypothesis**. Type I error occurs when we decide that $\theta \in \omega_1$ when in fact $\theta \in \omega_0$. Type II error occurs when we decide the opposite.

We require the **critical region**, C , to complete the testing structure for the general problem. Consider the r.v. X and the hypotheses given above. C is such that

$$\begin{aligned} \text{Reject } H_0 \text{ if } (X_1, \dots, X_n) &\in C \\ \text{Reject } H_1 \text{ if } (X_1, \dots, X_n) &\in C^c. \end{aligned} \quad (2.36)$$

Type I error occurs if H_0 is rejected when it is true. **Type II** error occurs if H_0 is retained when H_1 is true.

Definition: We say a critical region C is of **size** α if

$$\boxed{\alpha = \max_{\theta \in \omega_0} P_\theta[(X_1, \dots, X_n)] \in C} \quad (2.37)$$

Over all critical regions of size α , we want to consider critical regions that have lower probabilities of Type II error, i.e., for $\theta \in \omega_1$, we want to maximize

$$1 - P_\theta[\text{Type II Error}] = P_\theta[(X_1, \dots, X_n) \in C] \quad (2.38)$$

The probability on the right side of the equation above is called the **power** of the test at θ . It is the probability that the test detects the alternative when $\theta \in \omega_1$ is the true parameter. Minimizing Type II error requires maximizing the test power. The **power function** of C is

$$\gamma_C(\theta) = P_\theta[(X_1, \dots, X_n) \in C]; \quad \theta \in \omega_1. \quad (2.39)$$

Given two critical regions C_1, C_2 both of size α . C_1 is better than C_2 if $\gamma_{C_1}(\theta) \geq \gamma_{C_2}(\theta)$ for all $\theta \in \omega_1$.

A **simple** hypothesis completely specifies the underlying distribution. A **composite** hypothesis can be composed of many simple hypotheses and hence do not completely specify the distribution. α is often referred to as the **significance level** of the test associated with that critical region.

2.5 Additional comments about statistical test

2.5.1 Observed Significance Level, p -value

Suppose $H_0 : \mu = \mu_0$ and $H_1 : \mu > \mu_0$, where μ_0 is maximized. Then we reject H_0 in favor of H_1 if $\bar{X} \geq k$ where \bar{X} is the sample mean. The p -value is the probability that under H_0 , $\bar{X} \geq \bar{x}$:

$$\boxed{p\text{-value} = P_{H_0}(\bar{X} \geq \bar{x})} \quad (2.40)$$

If $\alpha > p$ then we reject H_0 in favor of H_1 . Else, we fail to reject H_1 .

2.6 Chi-Square Tests

Let r.v. $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for $i = 1, \dots, n$. Let X_1, \dots, X_n be mutually independent. The joint pdf is then

$$\frac{1}{\sigma_1 \dots \sigma_n (\sqrt{2\pi})^n} \exp \left[-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2 \right] \quad (2.41)$$

The r.v.

$$\sum_{i=1}^n \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2 \quad (2.42)$$

has a $\chi^2(n)$ distribution.

We will now consider some r.v.s that have approximate χ^2 distribution. Suppose $X_1 \sim b(n, p_1)$. Consider the r.v. defined by

$$Y = \frac{X_1 - p_1}{\sqrt{np_1(1-p_1)}} \quad (2.43)$$

which has, as $n \rightarrow \infty$, an approximate $\mathcal{N}(0, 1)$. We know from earlier discussions that $Y^2 \sim \mathcal{N}(0, 1)$. Now, let $X_2 = n - X_1$ and $p_2 = 1 - p_1$. Let $Q_1 = Y^2$. Then we have

$$Q_1 = \frac{(X_1 - np_1)^2}{np_1(1-p_1)} = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_1 - np_1)^2}{n(1-p_1)} = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2}. \quad (2.44)$$

In general, let X_1, \dots, X_{k-1} have a multinomial distribution with the parameters n and p_1, \dots, p_{k-1} . Let $X_k = n - (X_1 + \dots + X_{k-1})$ and let $p_k = 1 - (p_1 + \dots + p_{k-1})$. Define Q_{k-1} by

$$Q_{k-1} = \sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i} \quad (2.45)$$

As $n \rightarrow \infty$, $Q_{k-1} \sim \chi^2(k-1)$. This makes the r.v. Q_{k-1} a basis of the tests of certain statistical hypotheses. For instance, when the joint pdf of X_1, X_2, \dots, X_{k-1} (and $X_k = n - X_1 - \dots - X_{k-1}$) is a multinomial pmf with parameters n and p_1, \dots, p_{k-1} (and $p_k = 1 - p_1 - \dots - p_{k-1}$), we can consider the simple null hypothesis $H_0 : p_1 = p_{10}, \dots, p_{k-1} = p_{(k-1)0}$ where $p_{10}, \dots, p_{(k-1)0}$ are specified numbers. Under this null, the r.v.

$$Q_{k-1} = \sum_1^k \frac{(X_i - np_{i0})^2}{np_{i0}} \quad (2.46)$$

has an approximate $\chi^2(k-1)$. Intuitively, when H_0 is true, np_{i0} must be the expected value of X_i , which means Q_{k-1} is not too large. Thus, we reject H_0 if $Q_{k-1} \geq c$. The critical value c is specified by the significance level α , $\mathbf{c} = \text{qchisq}(1-\alpha, k-1)$. This is frequently called the **goodness-of-fit** test.

We can have a chi-square test for **homogeneity**. Consider two multinomial distributions with parameters $n_j, p_{1j}, \dots, p_{kj}$ and $j = 1, 2$. Let X_{ij} where $i =$

$1, \dots, k$ and $j = 1, 2$ be frequencies. Suppose n_1, n_2 large and the observations are independent, then the r.v.

$$\sum_{j=1}^2 \sum_{i=1}^n \frac{(X_{ij} - n_j p_{ij})^2}{n_j p_{ij}} \sim \chi^2(2k - 2) \quad (2.47)$$

because it is the sum of two independent r.v.'s each of which $\sim \chi^2(k - 1)$. The null hypothesis we consider is

$$H_0 : p_{11} = p_{12}; \dots; p_{k1} = p_{k2}, \quad (2.48)$$

where each $p_{i1} = p_{i2}$ where $i = 1, \dots, k$ is unspecified. It turns out that the mle of $p_{i1} = p_{i2}$ is given by

$$\theta = \frac{X_{i1} + X_{i2}}{n_1 + n_2} \quad (2.49)$$

which makes intuitive sense. Note that we need only $k - 1$ points estimates, and so the r.v.

$$Q_{k-1} = \sum_{j=1}^2 \sum_{i=1}^k \frac{\{X_{ij} - n_j[(X_{i1} + X_{i2})/(n_1 + n_2)]\}^2}{n_j[(X_{i1} + X_{i2})/(n_1 + n_2)]} \sim \chi^2(2k - 2 - (k - 1) = k - 1). \quad (2.50)$$

With this, we can test if two multinomial distributions are the same.

We can also test for **independence**. Suppose the result of an experiment is classified by only two attributes A (of a possible outcomes) and B (of b possible outcomes). These events are A_1, \dots, A_a for attribute A and B_1, \dots, B_b for attribute B . Then consider $p_{ij} = P(A_i \cap B_j)$. Say the experiment is repeated n independent times and X_{ij} denotes the frequency of the event $A_i \cap B_j$. There are $k = ab$ such events, so the r.v.

$$Q_{ab-1} = \sum_{j=1}^b \sum_{i=1}^a \frac{(X_{ij} - np_{ij})^2}{np_{ij}} \sim \chi^2(ab - 1), \quad (2.51)$$

provided n is large. To test for independence, $H_0 : P(A_i \cap B_j) = P(A_i)P(B_j)$ for all i, j . To test H_0 , we cannot compute Q_{ab-1} , but instead compute

$$\sum_{j=1}^b \sum_{i=1}^a \frac{[X_{ij} - n(X_{i.}/n)(X_{.j}/n)]^2}{n(X_{i.}/n)(X_{.j}/n)} \sim \chi^2(ab - 1 - (a + b - 2) = (a - 1)(b - 1)) \quad (2.52)$$

where

$$\hat{p}_{i.} = \frac{X_{i.}}{n}, \quad X_{i.} = \sum_{j=1}^b X_{ij}, i = 1, \dots, a \quad (2.53)$$

$$\hat{p}_{.j} = \frac{X_{.j}}{n}, \quad X_{.j} = \sum_{i=1}^a X_{ij}, j = 1, \dots, b \quad (2.54)$$

Just a sanity check, the chi-square statistic always has the form of $\sum \text{Expected} - \text{Observed}^2/\text{Expected}$. All tests' statistics have this form. The differences are subtle and are context-based.

2.7 The Method of Monte Carlo

The idea of Monte Carlo methods is to use random numbers to simulate random phenomena and to make numerical approximations. In general, we use Monte Carlo methods for

- Inverse transform sampling: take a random uniform $(0, 1)$ and transform it into a different distribution.
- Accept-Reject Algorithm, which is a method that uses a random uniform generator to produce a set of random numbers that follows some other distribution.
- To approximate the value of definite integrals.

2.7.1 Inverse Transform

For example, we want to simulate coin flips of a coin that is biased and comes up heads with probability p . Here's the algorithm:

- Generate a random uniform $(0, 1) \rightarrow u_1$
- If $u_1 < p \implies$ heads, else tails.
- Repeat

For any multinomial distribution with probabilities p_1, \dots, p_k we can follow this process:

- Generate a random uniform $u - 1 \sim (0, 1)$.
- If $u_1 \leq p_1 \implies$ assign outcome 1.
- Elif $u_1 \leq p_1 + p_2 \implies$ assign outcome 2.
- ...
- Elif $u_1 \leq p_1 + \dots + p_{k-1} \implies$ assign outcome $k - 1$.
- Else, assign outcome k .

Basically, what we're doing here is using the CDF to check against u .

In the continuous case, we generate $u \sim (0, 1)$ but we want a random variable with some other density $X \sim f_X(x)$. Assuming that $X = T(U)$, starting with the CDF of X :

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P(T(u) \leq x) \\ &= P(u \leq T^{-1}(x)) \\ &= F_u(T^{-1}(x)) \\ &= T^{-1}(x) \end{aligned} \tag{2.55}$$

where the last equality follows from the fact that $F_u(u)$ is just the identity function. So we have

$$F(x) = T^{-1}(x) \tag{2.56}$$

which means F and T are inverses of each other.

For example, we can use the inverse transform to generate random $\text{Exp}(\beta)$:

$$f_X(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x \in \mathbb{R}^+ \tag{2.57}$$

We want to find $F_X(x)$ first:

$$F_X(x) = \int_0^x \frac{1}{\beta} e^{-x'/\beta} dx' = 1 - e^{-x/\beta}. \tag{2.58}$$

And so it is easy to see that

$$T^{-1}(x) = 1 - e^{-x/\beta} \implies u = 1 - e^{-x/\beta} \implies x = \beta \log(1 - u). \tag{2.59}$$

From here, we can generate a sample of uniform u 's to get $X \sim \text{Exp}(\beta)$.

Note that a disadvantage to this method is the fact that for this to work we must be able to write down the inverse CDF in some closed form. The advantage, though, is that this method is very efficient if it works. This is because for each u we generate we get an x . This is not the case for the methods we will discuss next.

2.7.2 Accept-Reject Generation Algorithm

2.7.3 Evaluating definite integrals

This is based on the idea that

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx. \tag{2.60}$$

For example,

$$\int_0^1 e^{-x^2/2} dx = E[e^{-x^2/2}] \quad (2.61)$$

where $x \sim U(0, 1)$.

We could estimate $g(\cdot)$ by

$$\bar{g} = \frac{1}{J} \sum_{i=1}^J [g(x^{(j)})] \quad (2.62)$$

where $x^{(j)} \sim f(x)$. If we take enough random variables, $\bar{g} \rightarrow E[g]$. We could also use the CLT to calculate error bounds:

$$\bar{g} \pm z_{\alpha/2} SE(\bar{g}), \quad (2.63)$$

where

$$SE(\bar{g}) = \frac{1}{J^2} \sum_{i=1}^J \left(g(x^{(j)}) - \bar{g} \right)^2. \quad (2.64)$$

For example, we can try to evaluate

$$\int_0^\infty x^4 e^{-x^2/2} dx. \quad (2.65)$$

We identify $g(x) = x^4$, and e^{-x} as the density for $\text{Exp}(1)$. What we can do is generate a sample of random $\text{Exp}(1)$, lu into x^4 then take the average value of x^4 .

2.8 Bootstrapping

The basic idea of bootstrapping is this: We have a sample of data $x_1, \dots, x_n = \vec{X}$. We will replicate sample infinitely many times. This will be a model for the population.

Note that this process doesn't work well in the case that our sample is poorly representative. In practice, we obtain many bootstrap samples X^* 's where each X^* is a resample from our original sample where we randomly select X_i^* from \vec{X} **with replacement**.

The bootstrap sample is the same size as the original sample.

If we're interested in estimating a parameter using an estimator $\hat{\theta} = f(\vec{X})$, then we can calculate an estimate $\hat{\theta}^* = f(\vec{X}^*)$ from each bootstrap sample. The bootstrap distribution of $\hat{\theta}^*$ models the sampling distribution of $\hat{\theta}$.

For example, let x_1, \dots, x_n be a random sample from some population with an unknown mean μ . We can think about taking a simple bootstrap sample, X_i^* , then calculate its mean. Of course, this is an unbiased estimator for μ .

If we take many bootstrap samples X_i^* and calculate the mean of each, we could generate a bootstrap for \bar{X}^* . In an estimation setting, the most common use of bootstrap distribution is to estimate the SE of $\hat{\theta}$.

We can also find CI using percentiles/normal approximation this way.

2.8.1 Bootstrapping for Hypothesis Testing

Say we have $H_0 : \mu = \mu_0$ and $H_a : \mu > \mu_0$. To test these hypotheses with bootstrapping, we shift the original data such that the shifted mean is μ_0 , i.e. we do $X_i - \bar{X} + \mu_0$ to all observations.

From here, we bootstrap and generate a sampling distribution for the mean of the bootstrap samples. Then, we look and count how many bootstraps means are greater than the observed means and find the associated p -value.

We note that the advantage of doing this is we don't make any distributional assumptions. The disadvantage (kind of) is that this process can be a little too computationally expensive. However, with modern computers, this is no longer a major problem.

For example, say we want to compare two means. $H_0 : \Delta = 0$, $H_a : \Delta \neq 0$. Say we have two samples (of different sample sizes) from cdfs $F(x)$ and $F(x-\Delta)$,

respectively. Under the null, these samples come from the same distribution. So, we can follow these steps:

- Combine the samples into a single sample.
- Take one bootstrap of size n_1 and one of size n_2 .
- Calculate the difference in means.
- Repeat many times
- Count how many bootstrap differences are further from 0 than the original observed difference in means.
- Extract the p-value (either by percentile or by SE).

Part 3

Consistency and Limiting Distributions

3.1 Convergence in Probability

Definition: Let $\{X_n\}$ be a sequence of r.v. and let X be a r.v. defined on a sample pace. X_n converges in probability to X if, for all $\epsilon < 0$,

$$\lim_{n \rightarrow \infty} P[|X_n - X| \geq \epsilon] = 0, \quad (3.1)$$

i.e.,

$$\lim_{n \rightarrow \infty} P[|X_n - x| < \epsilon] = 1. \quad (3.2)$$

If so, we write

$$X_n \xrightarrow{P} X. \quad (3.3)$$

Theorem 3.1.1. (handy theorem) If $\hat{\theta}_n$ is an unbiased estimator of θ , then $\hat{\theta}_n \xrightarrow{P} \hat{\theta}$ if $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$. In which case, we call $\hat{\theta}$ a consistent estimator of θ .

Theorem 3.1.2. (Weak Law of Large Numbers). Let $\{X_n\}$ be a sequence of iid r.v. having common mean μ and variance σ^2 , then

$$\bar{X} \equiv \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu. \quad (3.4)$$

Proof: The proof uses Chebychev's inequality. Let $\epsilon > 0$ be given, then

$$P[|X_n - \bar{X}| \geq \epsilon] = P[|\bar{X} - \mu| \geq (\epsilon\sqrt{n}/\sigma)(\sigma/\sqrt{n})] \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow 0, \quad n \rightarrow \infty \quad (3.5)$$

□

Theorem 3.1.3. If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then

$$X_n + Y_n \xrightarrow{P} X + Y. \quad (3.6)$$

Proof: The proof is quite easy. It uses the fact that P is monotone relative to set containment and the triangle inequality. \square

Theorem 3.1.4. If $X_n \xrightarrow{P} X$ then $aX_n \xrightarrow{a} aX$.

Proof: The proof is also very easy, so I won't show it here. \square

Theorem 3.1.5. If $X_n \xrightarrow{P} a$ and the real function g is continuous at a then

$$g(X_n) \xrightarrow{P} g(a). \quad (3.7)$$

Proof: The proof is analysis-like. It's not so hard so I (again) won't show it here.

Theorem 3.1.6. If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then

$$X_n Y_n \xrightarrow{P} XY. \quad (3.8)$$

Proof: This proof uses the result from the previous theorem. The key is to write $X_n Y_n$ as a combination of X_n^2 , Y_n^2 , and $(X_n - Y_n)^2$. Applying the previous to obtain the desired conclusion. \square

3.1.1 Sampling and Statistic

Definition: (Consistency) Let X be a r.v. with cdf $F(x, \theta)$ with $\theta \in \Omega$. Let X_1, \dots, X_n be a sample from the distribution of X and let T_n denote a statistic. T_n is a **consistent** estimator of θ iff

$$T_n \xrightarrow{P} \theta. \quad (3.9)$$

3.2 Convergence in Distribution

Definition: (Convergence in Distribution) Let $\{X_n\}$ be a sequence of r.v.s and let X be a r.v.. Let F_{X_n} and F_X be, respectively, the cdfs of X_n and X . Let $C(F_X)$ denote the set of all points where F_X is continuous. We say that X_n converges in distribution to X if

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x), \quad \forall x \in C(F_X). \quad (3.10)$$

We denote this convergence by $X_n \xrightarrow{D} X$.

Stirling's Formula:

$$\Gamma(k+1) \approx \sqrt{2\pi} k^{k+1/2} e^{-k} \quad (3.11)$$

when k is large .

Theorem 3.2.1. If X_n converges to X in probability, then X_n converges to X in distribution.

Theorem 3.2.2. If X_n converges to the constant b in distribution, then X_n converges to b in probability.

Theorem 3.2.3. Suppose X_n converges to X in distribution and Y_n converges in probability to 0, then $X_n + Y_n$ converges to X in distribution.

Theorem 3.2.4. Suppose X_n converges to X in distribution and g is a continuous function on the support of X . Then $g(X_n)$ converges to $g(X)$ in distribution.

Theorem 3.2.5. (Slutsky's Theorem) Let X_n, X, A_n , and B_n be random variables and let a and b be constants. If $X_n \xrightarrow{D} X$, $A_n \xrightarrow{P} a$, and $B_n \xrightarrow{P} b$ then

$$A_n + B_n X_n \xrightarrow{D} a + bX. \quad (3.12)$$

3.2.1 Bounded in Probability

Definition: We say that the sequence of random variables $\{X_n\}$ is bounded in probability if, for all $\epsilon > 0$, there exists a constant $B_\epsilon > 0$ and an integer N_ϵ such that

$$n \geq N_\epsilon \implies P[|X_n| \leq B_\epsilon] \geq 1 - \epsilon. \quad (3.13)$$

Theorem 3.2.6. Let $\{X_n\}$ be a sequence of r.v. and let X be a r.v.. If $X_n \rightarrow X$ in distribution, then $\{X_n\}$ is bounded in probability.

Theorem 3.2.7. Let $\{X_n\}$ be a sequence of r.v. bounded in probability and let $\{Y_n\}$ be a sequence of r.v. that converges to 0 in probability. Then

$$X_n Y_n \xrightarrow{P} 0. \quad (3.14)$$

3.2.2 Δ -method

Little o notation: $a = o(b)$ if and only if $a/b \rightarrow 0$ as $b \rightarrow 0$.

Theorem 3.2.8. Suppose $\{Y_n\}$ is a sequence of r.v. that is bounded in probability. Suppose $X_n = o_P(Y_n)$, then $X_n \xrightarrow{P} 0$, as $n \rightarrow \infty$.

Theorem 3.2.9. Let $\{X_n\}$ be a sequence of r.v. such that

$$\sqrt{n}(X_n - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2). \quad (3.15)$$

Suppose $g(x)$ is differentiable at θ and $g'(\theta) \neq 0$, then

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} \mathcal{N}(0, \sigma^2(g'(\theta))^2) \quad (3.16)$$

3.2.3 Moment Generating Function Technique

Theorem 3.2.10. Let $\{X_n\}$ be a sequence of r.v. with mgf $M_{X_n}(t)$ that exists for $-h < t < h$ for all n . Let X be a r.v. with mdf $M(t)$, which exists for $|t| \leq h_1 \leq h$. If $\lim_{n \rightarrow \infty} M_{X_n}(t) = M(t)$ for $|t| \leq h_1$, then $X_n \xrightarrow{D} X$.

3.3 Central Limit Theorem

Theorem 3.3.1. (CLT) Let X_1, \dots, X_n denote the observations of a r.v. from a distribution that has mean μ and positive variance σ^2 . Then the r.v.

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} Z \sim \mathcal{N}(0, 1). \quad (3.17)$$

Proof: Assume that the mgf $M(t) = E[e^{tX}]$ exists for $-h < t < h$, then the function

$$m(t) = E[e^{t(X-\mu)}] = e^{-\mu t} M(t) \quad (3.18)$$

also exists for $-h < t < h$. $m(t)$ is the mgf for $X - \mu$, so $m(0) = 1, m'(0) = E[X - \mu] = 0$, and $m''(0) = E[(X - \mu)^2] = \sigma^2$. By Taylor theorem, there exists a number $\xi \in [0, t]$ such that

$$\begin{aligned} m(t) &= m(0) + m'(0)t + \frac{m''(\xi)t^2}{2} \\ &= 1 + \frac{m''(\xi)t^2}{2} \\ &= 1 + \frac{\sigma^2 t^2}{2} + \frac{[m''(\xi) - \sigma^2]t^2}{2}. \end{aligned} \quad (3.19)$$

Now consider $M(t; n)$:

$$\begin{aligned} M(t; n) &= E \left[\exp \left(t \frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \right) \right] \\ &= \dots \\ &= \left\{ E \left[\exp \left(t \frac{X - \mu}{\sigma\sqrt{n}} \right) \right] \right\}^n \\ &= \left[m \left(\frac{t}{\sigma\sqrt{n}} \right) \right]^n, \quad -h < \frac{t}{\sigma\sqrt{n}} < h \\ &= \left\{ 1 + \frac{t^2}{2n} + \frac{[m''(\xi) - \sigma^2]t^2}{2} \right\}^n. \end{aligned} \quad (3.20)$$

Taking $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} [m''(\xi) - \sigma^2] = 0, \quad (3.21)$$

and so

$$\lim_{n \rightarrow \infty} M(t; n) = e^{t^2/2}, \quad t \in \mathbb{R}. \quad (3.22)$$

So $Y_n \sim \mathcal{N}(0, 1)$.

Part 4

Maximum Likelihood Methods

4.1 Maximum Likelihood Estimation

Recall the likelihood function:

$$\mathcal{L}(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta), \quad \theta \in \Omega \quad (4.1)$$

where $f(x_i; \theta)$ is the pdf which the variables X_i follow that depends on the parameter θ and $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top$ is the sample. It's often more convenient to use the log likelihood:

$$l(\theta) = \ln \mathcal{L}(\theta) = \sum_{i=1}^n \log f(x_i; \theta). \quad (4.2)$$

$\hat{\theta}$ is the mle of θ if $\hat{\theta}$ maximizes $l(\theta)$. Let θ_0 denote the true value of θ . We will look at theorem which shows that the maximum of $\mathcal{L}(\theta)$ asymptotically separates the true model at θ_0 from models at $\theta \neq \theta_0$. To prove this theorem, we look at *regularity conditions*:

Regularity Conditions. Regular conditions are

- The cdfs are distinct, i.e., $\theta \neq \theta' \implies F(x_i; \theta) \neq F(x_i; \theta')$.
- The pdfs have common support for all θ .
- The point θ_0 is an interior point in Ω .

Theorem 4.1.1. Assume that θ_0 is the true parameter and that

$$E_{\theta_0}[f(X_i, \theta)/f(X_i, \theta_0)] \quad (4.3)$$

exists. Under the first two regularity conditions

$$\lim_{n \rightarrow \infty} P_{\theta_0}[\mathcal{L}(\theta_0, \mathbf{X}) > \mathcal{L}(\theta, \mathbf{X})] = 1, \quad \forall \theta \neq \theta_0. \quad (4.4)$$

Definition 4.1.1. (Maximum Likelihood Estimator). We say that $\hat{\theta} = \hat{\theta}(\mathbf{X})$ is a maximum likelihood estimator (mle) of θ if

$$\hat{\theta} = \text{Argmax} \mathcal{L}(\theta, \mathbf{X}) \quad (4.5)$$

where the notation means $\mathcal{L}(\theta, \mathbf{X})$ attains maximum at $\hat{\theta}$.

Theorem 4.1.2. (Invariance Property) Let X_1, \dots, X_n be iid with pdf $f(x; \theta), \theta \in \Omega$. For a specified function g , let $\eta = g(\theta)$ be a parameter of interested. Suppose $\hat{\theta}$ is the mle of θ . Then $g(\hat{\theta})$ is the mle of $\eta = g(\theta)$.

Theorem 4.1.3. Assume that X_1, \dots, X_n satisfy the regularity conditions, where θ_0 is the true parameter, and further that $f(x; \theta)$ is differentiable w.r.t. $\theta \in \Omega$. Then the likelihood equation,

$$\partial_\theta \mathcal{L}(\theta) = 0 \iff \partial_\theta l(\theta) = 0 \quad (4.6)$$

has a solution $\hat{\theta}_n$ such that $\hat{\theta}_n \xrightarrow{P} \theta_0$.

Theorem 4.1.4. Assume that X_1, \dots, X_n satisfy the regularity conditions, where θ_0 is the true parameter, and that $f(x; \theta)$ is differentiable w.r.t. $\theta \in \Omega$. Suppose the likelihood equation has the **unique** solution $\hat{\theta}_n$. Then $\hat{\theta}_n$ is a consistent estimator of θ_0 .

So far, we know that two things that make mle good:

- Invariance property
- Consistency of MLEs.

Example 4.1.1. Let X_1, \dots, X_n be iid $\text{Exp}(\theta)$ r.v.'s. We want to find $\hat{\theta}_{\text{MLE}}$. We know that

$$f(x_i; \theta) = \frac{1}{\theta} e^{-x_i/\theta}; \quad x_i > \theta, \theta > 0. \quad (4.7)$$

The likelihood function is

$$\mathcal{L}(\theta) = \frac{1}{\theta^n} e^{-\sum_{i=1}^n x_i/\theta}. \quad (4.8)$$

The log likelihood is

$$l(\theta) = -n \ln \theta - \frac{1}{\theta} \sum_{x_i}. \quad (4.9)$$

from which we can easily solve the mle

$$\hat{\theta} = \frac{1}{n} \sum_i x_i = \bar{x}. \quad (4.10)$$

Since all regular conditions are satisfied, this is a *good* mle.

Example 4.1.2. Let X_1, \dots, X_n be iid $U(0, \theta)$ r.v.. We want to find the $\hat{\theta}_{ML}$. Now,

$$f(x_i; \theta) = \frac{1}{\theta}; 0 \leq x_i \leq \theta; \theta > 0. \quad (4.11)$$

We note that the second and third regularity conditions do not hold. Next,

$$\mathcal{L}(\theta) = \frac{1}{\theta^n} \implies l(\theta) = -n \ln \theta. \quad (4.12)$$

We want $\hat{\theta}$ to be as small as possible to maximize $\mathcal{L}(\theta)$, but it also has to be bigger than all of the observations. Thus, $\hat{\theta}_{ML} = \max_i(X_i)$.

Example 4.1.3. Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$ r.v.. We want to find $\hat{\mu}$ and $\hat{\sigma}^2$. We know that

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}; x_i, \mu \in \mathbb{R}, \sigma^2 > 0. \quad (4.13)$$

The log likelihood is easy:

$$l(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2. \quad (4.14)$$

Then

$$\partial_\mu l(\mu, \sigma^2) = -\frac{1}{\sigma^2} \sum (x_i - \mu) = 0 \implies \hat{\mu} = \bar{x}. \quad (4.15)$$

Also,

$$\partial_{\sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \bar{x})^2 = 0 \implies \hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2. \quad (4.16)$$

Because the regularity conditions are satisfied, these are consistent estimators.

Example 4.1.4. Let $X_i \sim \text{Bernoulli}$ with parameter p , i.e., $P(x_i = 1) = p$ and $P(x_i = 0) = 1 - p$. Then

$$P(x_1, \dots, x_n) = p^{\sum x_i} (1-p)^{n-\sum x_i}. \quad (4.17)$$

Then the log likelihood function is just

$$l(p) = \sum x_i \ln p + \left(n - \sum x_i\right) \ln(1-p). \quad (4.18)$$

It follows that

$$\partial_p l(p) = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0 \implies \hat{p} = \frac{\sum x_i}{n}. \quad (4.19)$$

This is a consistent estimator because the regularity conditions are satisfied.

Example 4.1.5. Let $X_i \sim \text{Poi}(\lambda)$. We want to find $\hat{\lambda}$. We have

$$p(x_i | \lambda) = \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}, \quad x_i = 1, 2, \dots, \quad \lambda > 0. \quad (4.20)$$

The log likelihood function is

$$l(\lambda) = -ns + \sum x_i \ln \lambda - \sum \ln(x_i!). \quad (4.21)$$

So,

$$\partial_\lambda l(\lambda) = 0 \implies \hat{\lambda} = \bar{x}. \quad (4.22)$$

From WLLN, $\bar{x} \rightarrow \mu$, so as long as the regularity conditions are satisfied (which they are), then we have a consistent estimator for μ .

4.2 Rao-Cramér Lower Bound and Efficiency

Additional Regularity Conditions.

- The pdf of $f(x; \theta)$ is twice differentiable as a function of θ .
- The integral $\int f(x; \theta) dx$ can be differentiated twice under the integral sign as a function of θ .

All four regularity conditions we have seen so far combined means that the parameter θ does not appear in the endpoints of the interval in which $f(x; \theta) > 0$ and that we can interchange integration and differentiation w.r.t θ . The derivation is below is the the continuous case, the the discrete case can be handled in a similar manner. I'll summarize the derivation in a few steps below:

$$1 = \int_{-\infty}^{\infty} f(x; \theta) dx \xrightarrow{\partial_\theta} 0 = \int_{-\infty}^{\infty} \partial_\theta f(x; \theta) dx. \quad (4.23)$$

Next,

$$\partial_\theta f(x; \theta) = \frac{\partial_\theta f(x; \theta)}{f(x; \theta)} f(x; \theta) \implies 0 = \int_{-\infty}^{\infty} \partial_\theta \ln f(x; \theta) f(x; \theta) dx. \quad (4.24)$$

And so writing this as an expectation:

$$E[\partial_\theta \ln f(X; \theta)] = 0. \quad (4.25)$$

Now, if we take the second derivative of the identity integral we get

$$0 = \int_{-\infty}^{\infty} \partial_\theta^2 \ln f(x; \theta) f(x; \theta) dx + \int_{-\infty}^{\infty} (\partial_\theta \ln f(x; \theta))^2 f(x; \theta) dx. \quad (4.26)$$

The second term on the RHS can be written as an expectation, called the **Fisher information**, denoted $I(\theta)$:

$$I(\theta) = E[(\partial_\theta \ln f(X; \theta))^2] = -E[\partial_\theta^2 \ln f(X; \theta)] \quad (4.27)$$

Now, because $E[\partial_\theta \ln f(x; \theta)] = 0$ we can see that

$$I(\theta) = \text{Var}[\partial_\theta \ln f(X; \theta)] \quad (4.28)$$

The important function

$$\partial_\theta \ln f(x; \theta) \quad (4.29)$$

is called the **score function**. Recall that it determines the estimating equations for the mle, i.e., the mle $\hat{\theta}$ solves

$$\sum_{i=1}^n \partial_\theta \ln f(x_i; \theta) = 0. \quad (4.30)$$

For an n -sample of iid r.v., the Fisher information is

$$\mathcal{I}(\theta) = nI(\theta) = \text{Var}(\partial_\theta \ln \mathcal{L}(\theta, \mathbf{X})). \quad (4.31)$$

Theorem 4.2.1. (Rao-Cramér Lower Bound.) Let X_1, \dots, X_n be iid with pdf $f(x; \theta), \theta \in \Omega$. Assume that all four regularity conditions hold. Let $Y = u(X_1, \dots, X_n)$ be a statistic with mean $E[Y] = k(\theta)$. Then

$$\text{Var}(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)} \quad (4.32)$$

Proof. Here's a sketch of the proof. Define

$$Z = \sum_{i=1}^n \partial_\theta \ln f(X_i; \theta). \quad (4.33)$$

Then $E[Z] = 0$ and $\text{Var}[Z] = nI(\theta)$. Now, verify that

$$k'(\theta) = E[YZ] = E[Y]E[Z] + \rho\sigma_Y\sqrt{nI(\theta)} \quad (4.34)$$

where ρ is the correlation coefficient between Y and Z . Using $E[Z] = 0$ and rearrange, we get the desired result. \square

Theorem 4.2.2. Under the assumptions of the Rao-Cramér lower bound theorem, if $Y = u(X_1, \dots, X_n)$ is an unbiased estimator of θ , so that $k(\theta) = \theta$, then the Rao-Cramér inequality becomes

$$\text{Var}(Y) \geq \frac{1}{nI(\theta)} \quad (4.35)$$

Example 4.2.1. Consider a single observation from a Poisson- λ . We want to find $I(\lambda)$. Well,

$$P(x_i; \lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}, \quad x_i \in \mathbb{N}, \lambda > 0 \in \Omega. \quad (4.36)$$

The regularity conditions are satisfied. Now,

$$I(\lambda) = E \left[(\partial_\lambda \ln P(x_i, \lambda))^2 \right]. \quad (4.37)$$

We can find

$$\partial_\lambda \ln P(x_i, \lambda) = -1 + \frac{x_i}{\lambda}. \quad (4.38)$$

And so

$$\begin{aligned} I(\lambda) &= E \left[x_i^2/\lambda^2 - 2x_i/\lambda + 1 \right] = \frac{E[x_i^2]}{\lambda^2} - \frac{2}{\lambda}E[x_i] + 1 \\ &= \frac{1}{\lambda^2}(\lambda + \lambda^2) - \frac{2}{\lambda}\lambda + 1 = \frac{1}{\lambda}. \end{aligned} \quad (4.39)$$

We can do it using the other expectation too, but let's not worry about that. What about a sample of n iid such r.v.'s? The answer is just n/λ .

Example 4.2.2. Let $X_i \sim \text{Poi}(\lambda)$. Find $\hat{\mu}_{ML}$. Well,

$$l(\lambda) = -n\lambda + \sum_{i=1}^n x_i \ln \lambda - \sum \ln x_i!. \quad (4.40)$$

And so,

$$\partial_\lambda l(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i = 0 \implies \hat{\mu}_{ML} = \bar{X}. \quad (4.41)$$

This is unbiased. We also know that

$$\text{Var}(\bar{X}) = \frac{\lambda}{n}. \quad (4.42)$$

How does this compare with the CRLB? We have an unbiased estimator, and so

$$\text{Var}(\bar{X}) \geq \frac{1}{nI(\lambda)} = \frac{\lambda}{n}. \quad (4.43)$$

In this case, our estimator \bar{X} does achieve the CRLB, which makes it a good mle.

Example 4.2.3. Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$ where μ is unknown but σ^2 is. What is $nI(\mu)$? Well, for a single observation,

$$f(x_i; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \mu)^2/2\sigma^2}; \quad x_i, \mu \in \mathbb{R}. \quad (4.44)$$

Next,

$$l(x_i; \mu) = \frac{-1}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x_i - \mu)^2. \quad (4.45)$$

This the score function. The information is

$$I(\mu) = E[l(x_i; \mu)^2] = -E[\partial_\mu^2 f(x_i; \mu)] = \frac{1}{\sigma^2}. \quad (4.46)$$

And so,

$$nI(\mu) = \frac{n}{\sigma^2}. \quad (4.47)$$

And so, the CRLB for an unbiased estimator is σ^2/n .

Definition 4.2.1. (Efficient Estimator.) Let Y be an unbiased estimator of a parameter θ in the case of point estimator. The statistic Y is called an **efficient estimator** of θ if and only if $\text{Var}(Y)$ attains the Rao-Cramér lower bound.

Definition 4.2.2. (**Efficiency.**) In cases in which we can differentiate w.r.t a parameter under an integral or summation symbol, the ratio of the Rao-Cramér lower bound to the actual variance of any unbiased estimator of a parameter is called the **efficiency** of that estimator.

Additional Regularity Condition. (the total is 5 after this)

- The pdf $f(x; \theta)$ is three times differentiable as a function of θ . Further, for all $\theta \in \Omega$, there exists a constant c and a function $M(x)$ such that

$$|\partial_\theta^3 \ln f(x; \theta)| \leq M(x), \quad (4.48)$$

with $E_{\theta_0}[M(X)] < \infty$ for all $\theta_0 - c < \theta < \theta_0 + c$ and all x in the support of X .

Theorem 4.2.3. Assume X_1, \dots, X_n are iid with pdf $f(x; \theta)$ for $\theta_0 \in \Omega$ such that the 5 regularity conditions are satisfied. Suppose further that the Fisher information satisfies $0 < I(\theta_0) < \infty$. Then any consistent sequence of solutions of the mle equations satisfies

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}\left(0, \frac{1}{I(\theta_0)}\right) \quad (4.49)$$

In practical terms, this tells us that the MLEs for large n have a normal distribution and that they have variance that approaches CRLB.

Definition 4.2.3. Let X_1, \dots, X_n be iid with pdf $f(x; \theta)$. Suppose $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, \dots, X_n)$ is an estimator of θ_0 such that $\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{D} \mathcal{N}(0, \sigma_{\hat{\theta}_{1n}}^2)$. Then

- The **asymptotic efficiency** of $\hat{\theta}_{1n}$ is defined to be

$$e(\hat{\theta}_{1n}) = \frac{1/I(\theta_0)}{\sigma_{\hat{\theta}_{1n}}^2} \quad (4.50)$$

- The estimator $\hat{\theta}_{1n}$ is said to be **asymptotically efficient** if the ratio in the previous item is 1.
- Let $\hat{\theta}_{2n}$ be another estimator such that $\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{D} \mathcal{N}(0, \sigma_{\hat{\theta}_{2n}}^2)$. Then the **asymptotic relative efficiency** (ARE) of $\hat{\theta}_{1n}$ to $\hat{\theta}_{2n}$ is the reciprocal of the ratio of their respective asymptotic variances, i.e.,

$$e(\hat{\theta}_{1n}, \hat{\theta}_{2n}) = \frac{\sigma_{\hat{\theta}_{2n}}^2}{\sigma_{\hat{\theta}_{1n}}^2}. \quad (4.51)$$

Note: MLEs are asymptotically efficient estimators.

Theorem 4.2.4. Under the assumptions of Theorem 4.2.3., suppose $g(x)$ is a continuous function of x that is differentiable at θ_0 such that $g'(\theta_0) \neq 0$. Then

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta_0)) \xrightarrow{D} \mathcal{N}\left(0, \frac{g'(\theta_0)^2}{I(\theta_0)}\right). \quad (4.52)$$

Proof. The proof uses previous theorems and the Δ -method. \square

Theorem 4.2.5. Under the assumptions of Theorem 4.2.3.,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{I(\theta_0)} \frac{1}{\sqrt{n}} \sum_{i=1}^n \partial_\theta \ln f(X_i; \theta_0) + R_n \quad (4.53)$$

where $R_n \xrightarrow{P} 0$.

Example 4.2.4. Let X_1, \dots, X_n be iid $\text{Exp}(\theta)$ r.v., Find the information contained in the sample about θ . Well,

$$f(x_i; \theta) = \frac{1}{\theta} e^{-x_i/\theta}, x_i > 0, \theta > 0. \quad (4.54)$$

The information is given by

$$I(\theta) = -E[\partial_\theta^2 l(\theta)] = \dots = -\frac{1}{\theta^2} + \frac{2}{\theta^3} E[x_i] = \frac{1}{\theta^2}. \quad (4.55)$$

So the amount of information contained in the sample is $nI(\theta) = n/\theta^2$.

Now consider two estimators for θ : \bar{Y} and $n \min X_i$. We know that \bar{Y} is unbiased. Let's check if $n \min X_i$ is unbiased: $E[n \min X_i] = nE[\min X_i]$. Well,

$$\begin{aligned} E[\min X_i] &= \int_0^\infty y g_{\min}(y) dy \\ &= \int_0^\infty y n [1 - F(y)]^{n-1} f(y) dy \\ &= \dots \\ &= \int_0^\infty y \frac{n}{\theta} e^{-ny/\theta} dy \\ &= \frac{\theta}{n}. \end{aligned} \quad (4.56)$$

So $E[n \min X_i] = \theta$, unbiased.

Now, the CRLB is $1/nI(\theta) = \theta^2/n$, which is the smallest possible variance for an unbiased estimator. Now, $\text{Var}(\bar{Y}) = \theta^2/n$. And so this estimator has efficiency of 1 and is an efficient estimator. $\text{Var}(n \min X_i)$ turns out to be $\theta^2 \geq \theta^2/n$. So, the relative efficiency of $\hat{\theta}_1$ and $\hat{\theta}_2$ is $1/n$. The asymptotic efficiency is going to be zero.

4.3 Maximum Likelihood Tests

Consider the two-sided hypothesis

$$H_0 : \theta = \theta_0 \quad H_a : \theta \neq \theta_0. \quad (4.57)$$

The log likelihood is given by

$$l(\theta) = \sum_{i=1}^n \ln f(X_i; \theta). \quad (4.58)$$

If θ_0 is the true value of θ , then asymptotically $\mathcal{L}(\theta_0)$ is the maximum value of $\mathcal{L}(\theta)$. Consider the ratio of two likelihood functions

$$\Lambda = \frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\hat{\theta})} \leq 1. \quad (4.59)$$

If H_0 is true, Λ large and is close to 1. If H_1 is true, Λ is smaller. For a specified significance level α , we get a decision rule: Reject H_0 if $\Lambda \leq c$ where c is such that $\alpha = P_{\theta_0}(\Lambda \leq c)$. This is called the **likelihood ratio test**.

Theorem 4.3.1. Assume the same regularity conditions as for Theorem 4.2.3, under the null hypothesis $H_0 : \theta = \theta_0$,

$$-2 \ln \Lambda \xrightarrow{D} \chi^2(1). \quad (4.60)$$

Proof. The proof uses Taylor expansion about θ_0 of $l(\theta)$ to first order, evaluated at the mle $\hat{\theta}$. \square

Define the test statistic $\chi_L^2 = -2 \ln \Lambda$. The theorem above suggests the decision rule: Reject H_0 if $\chi_L^2 \geq \chi_\alpha^2(1)$.

Example 4.3.1. Let X_1, \dots, X_n be iid $\mathcal{N}(\mu, \sigma^2)$ r.v. where both μ and σ^2 are unknown. $H_0 : \mu = \mu_0$ and $H_a : \mu \neq \mu_0$. The null space is $\{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}$ and the alternative space is $\{(\mu, \sigma^2) : \mu \neq \mu_0, \sigma^2 > 0\}$. The union of these two spaces is $\{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$. Under the null, $\mu = \mu_0$ and σ^2 is unknown.

$$\partial_{\sigma^2} l(\mu_0, \sigma^2) = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu_0)^2 = 0 \implies \hat{\sigma}_0^2 = \frac{1}{n} \sum (x_i - \mu_0)^2. \quad (4.61)$$

And so

$$\mathcal{L}_0(\hat{\theta}) = (2\pi)^{-n/2} (\hat{\sigma}_0)^{-n/2} \exp \left[-\frac{1}{2\hat{\sigma}_0^2} \sum (x_i - \mu_0)^2 \right]. \quad (4.62)$$

This is the likelihood in null space. Under the joint null alternative space where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$,

$$l(\mu, \sigma^2) = -\frac{n}{2} \ln(2\bar{x}) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2. \quad (4.63)$$

We need to estimate both μ and σ^2 . We have

$$\partial_\mu l(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum (x_i - \mu) = 0 \implies \hat{\mu} = \bar{x} \quad (4.64)$$

$$\partial_{\sigma^2} l(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \sum (x_i - \mu)^2 = 0 \implies \hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \mu)^2. \quad (4.65)$$

The likelihood in this case is

$$\mathcal{L}(\hat{\theta}) = (2\pi)^{-n/2} (\hat{\sigma}^2)^{-n/2} \exp \left[-\frac{1}{2\hat{\sigma}^2} \sum (x_i - \bar{x})^2 \right]. \quad (4.66)$$

Now, we reject if

$$\Lambda = \frac{\mathcal{L}_0}{\mathcal{L}} = \dots = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \right)^{n/2} < K \quad (4.67)$$

which is equivalent to when

$$\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} = \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \mu_0)^2} < K^{2/n} \equiv K'. \quad (4.68)$$

Now, note that

$$\sum (x_i - \mu_0)^2 = \sum [(x_i - \bar{x}) + (\bar{x} - \mu_0)]^2 = \sum (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2. \quad (4.69)$$

So, we reject whenever

$$\frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2} < K^{2/n} = \frac{1}{1 + \frac{n(\bar{x} - \mu_0)^2}{\sum (x_i - \bar{x})^2}} < K' \quad (4.70)$$

which is equivalent to

$$\frac{n(\bar{x} - \mu_0)^2}{\sum (x_i - \bar{x})^2} > \frac{1}{K} - 1 \equiv K'' \iff \frac{n(\bar{x} - \mu_0)^2}{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \frac{n(\bar{x} - \mu_0)^2}{S^2} > (n-1)K'', \quad (4.71)$$

where K'' is another constant. So we reject whenever

$$\left| \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \right| > \sqrt{(n-1)K''} \quad (4.72)$$

where the LHS is a t -statistic with $(n-1)$ df.

In practice, there are two other likelihood-related tests. A natural statistic is based on the asymptotic distribution of $\hat{\theta}$. Consider the statistic

$$\chi_W^2 = \left\{ \sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) \right\}^2. \quad (4.73)$$

Because $I(\theta)$ is a continuous function, $I(\hat{\theta}) \xrightarrow{P} I(\theta_0)$ under the null hypothesis. So, under the null hypothesis, χ_W^2 has an asymptotic χ^2 -distribution with one degree of freedom. So the decision rule is to reject H_0 if $\chi_W^2 \geq \chi_\alpha^2(1)$. This is often referred to as a **Wald**-type test.

The third test is called the **scores**-type test. The **scores** are the components of the vector

$$\mathbf{S}(\theta) = (\partial_\theta \ln f(X_1; \theta), \dots, \partial_\theta \ln f(X_n; \theta))'. \quad (4.74)$$

In our notation we have

$$\frac{1}{\sqrt{n}} l'(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \partial_\theta \ln f(X_i; \theta_0). \quad (4.75)$$

Define the statistic

$$\chi_R^2 = \left(\frac{I'(\theta_0)}{\sqrt{nI(\theta_0)}} \right)^2 \quad (4.76)$$

The decision rule is to reject H_0 in favor of H_a if $\chi_R^2 \geq \chi_\alpha^2(1)$.

Example 4.3.2. (MLR test) Suppose we want to test whether or not the mean number of goals scored in all English league soccer games is 3. Suppose $X \sim \text{Poi}(\lambda)$. We set $H_0 : \lambda = 3$ and $H_a : \lambda \neq 3$. The joint pdf is

$$\mathcal{L}(\lambda) = p(x_i | \lambda) = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod x_i!}, \quad x_i = 0, 1, 2, 3, \dots \quad (4.77)$$

Say $n = 380$ and $\bar{x} = 2.534$. We can find the MLE of λ , $\hat{\lambda} = \bar{x}$. The null space is $\{\lambda : \lambda = 3\}$, and the alternative space is $\{\lambda : \lambda \neq 3\}$. The joint space is $\{\lambda : \lambda > 0\}$. Now,

$$\mathcal{L}(\hat{\lambda}_0) = \frac{e^{-380 \cdot 3} 3^{963}}{\dots} \quad (4.78)$$

and

$$\mathcal{L}(\hat{\lambda}) = \frac{e^{-380 \cdot 2.534} 2.534^{963}}{\dots} \quad (4.79)$$

and so

$$\Lambda = \frac{\mathcal{L}_0}{\mathcal{L}} \quad (4.80)$$

Asymptotically, $-2 \ln \Lambda \sim \chi^2(1)$. We reject if $-2 \ln \Lambda > \chi_{1,0.05}^2 = 3.84$. We find that $-2 \ln \Lambda = 29.03 > 3.84$, so we reject.

Example 4.3.3. (Wald test)

$$\chi_W^2 = \left[\frac{\hat{\lambda} - \lambda_0}{1/\sqrt{nI(\hat{\lambda})}} \right]^2 \sim \chi^2(1)$$

under H_0 . The information can be found to be n/λ . Plugging in the numbers we find

$$\chi_W^2 = \left[\frac{\bar{x} - \lambda_0}{1/\sqrt{n/\bar{x}}} \right]^2 = \left[(2.534 - 3)/(1/\sqrt{380 \cdot 1/2.534}) \right]^2 = 32.56 > 3.84,$$

so we still reject.

Example 4.3.4. (Score test)

$$\chi_R^2 = \frac{(I'(\theta_0))^2}{nI(\theta_0)} = \dots = \frac{(-380 + 963/3)^2}{380/3} = 27.48 \sim \chi^2.$$

So we reject as well.

So, χ_R^2 is in general not as conservative as χ_W^2 . The LRT is somewhere in the middle.

4.4 Multiparameter Case: Estimation

In this section we consider the case where $\boldsymbol{\theta}$ is a vector of p parameters. Let X_1, \dots, X_n be iid with common pdf $f(x; \boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^p$. The likelihood function and its log are given by

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta}) \tag{4.81}$$

$$l(\boldsymbol{\theta}) = \ln \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \ln f(x_i; \boldsymbol{\theta}). \tag{4.82}$$

There are more regularity conditions to make the theory work, but we won't worry about those. We will also worry only about the continuous case, because the discrete cases can be treated in similar ways.

Here are some facts: just like before, $\mathcal{L}(\boldsymbol{\theta})$ is maximized at the true value of $\boldsymbol{\theta}$. The quantity that maximizes $\mathcal{L}(\boldsymbol{\theta})$ is called the **maximum likelihood estimator** and we denote it $\hat{\boldsymbol{\theta}}$. It also holds that $\hat{\eta} = g(\hat{\boldsymbol{\theta}})$ if $\eta = g(\boldsymbol{\theta})$.

Theorem 4.4.1. Let X_1, \dots, X_n be iid with pdf $f(x; \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Omega$. Assume that the regularity conditions hold, then

1. The likelihood equation,

$$\partial_{\boldsymbol{\theta}} l(\boldsymbol{\theta}) = \mathbf{0} \quad (4.83)$$

has a solution $\hat{\boldsymbol{\theta}}_n$ such that $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}$.

2. For any sequence that satisfies the first item

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}_p(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta})). \quad (4.84)$$

Theorem 4.4.2. Let X_1, \dots, X_n be iid with pdf $f(x; \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Omega$. Assume the regularity conditions hold. Let $\hat{\boldsymbol{\theta}}_n$ be a sequence of consistent solutions of the likelihood equation. Then $\hat{\boldsymbol{\theta}}_n$ are asymptotically efficient estimates; that is, for $j = 1, \dots, p$,

$$\sqrt{n}(\hat{\theta}_{i,j} - \theta_j) \xrightarrow{D} \mathcal{N}(0, [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{jj}). \quad (4.85)$$

Example 4.4.1. Let X_i be iid $\mathcal{N}(\mu, \sigma^2)$. Find MLE for μ and σ^2 . Now, $\boldsymbol{\theta} = (\mu, \sigma^2)$, and

$$\mathcal{L}(\boldsymbol{\theta}) = \prod f(x_i | \boldsymbol{\theta}) = \prod \left[(2\pi\sigma^2)^{1/2} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \right]. \quad (4.86)$$

Then

$$\partial_{\mu} l(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum (x_i - \mu) = 0 \implies \hat{\mu} = \bar{x} \quad (4.87)$$

and

$$\partial_{\sigma^2} l(\boldsymbol{\theta}) = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu)^2 = 0 \implies \hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \mu)^2. \quad (4.88)$$

What is the information in the multiparameter setting? It's a matrix with elements

$$I_{jk} = -E \left[\frac{\partial^2 \ln f(x_i | \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right] \quad (4.89)$$

Back to the normal situation:

$$I(\boldsymbol{\theta}) = -E \begin{bmatrix} -\frac{1}{\sigma^2} & -\frac{1}{\sigma^4}(x_i - \mu) \\ -\frac{1}{\sigma^4}(x_i - \mu) & \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x_i - \mu)^2 \end{bmatrix} \quad (4.90)$$

$$= \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & -\frac{1}{2\sigma^4} + \frac{1}{\sigma^6}\sigma^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix} \quad (4.91)$$

CRLB for \bar{x} is

$$\frac{1}{n I_{11}(\boldsymbol{\theta})} = \frac{\sigma^2}{n}. \quad (4.92)$$

The CRLB for $\hat{\sigma}^2$ is

$$\frac{1}{n I_{22}(\boldsymbol{\theta})} = \frac{2\sigma^4}{n}. \quad (4.93)$$

4.5 Multiparameter Case: Testing

Here we talk about likelihood ratio tests in the multiparameter setting. Recall that the test statistic is

$$\Lambda = \frac{\mathcal{L}(\hat{\theta}_0)}{\mathcal{L}(\hat{\theta})} \quad (4.94)$$

where the numerator is maximized in null space and the denominator is maximized in the joint (null + alternative) space. We reject if $\Lambda < c$. c is found from the null distribution.

Example 4.5.1. Let $X_i \sim \exp(\theta)$ and $Y_i \sim \exp(\mu)$. $H_0 : \mu = \theta$ and $H_a : \mu \neq \theta$. The null space is $\{\{\theta, \mu\} : \theta = \mu > 0\}$. The alternative space is $\{\{\theta, \mu\} : \theta \neq \mu, \theta, \mu \neq 0\}$. The joint space is just $\{\{\theta, \mu\}, \theta > 0, \mu > 0\}$. It's easy to see that

$$\mathcal{L}(\theta, \mu | X, Y) = \frac{1}{\theta^n} e^{\sum^n x_i / \theta} \frac{1}{\mu^m} e^{\sum^m y_i / \mu}. \quad (4.95)$$

From here one finds that under $H_0 : \theta = \mu$:

$$l(\Theta_0) = -(n+m) \ln \theta - \frac{1}{\theta} \left(\sum x_i + \sum y_i \right). \quad (4.96)$$

And so

$$\partial_\theta l(\theta) = 0 \implies \hat{\theta}_0 = \frac{1}{n+m} \left(\sum x_i + \sum y_i \right). \quad (4.97)$$

With this, we can plug back in to calculate the numerator:

$$\mathcal{L}(\hat{\Theta}_0) = \dots = \left(\frac{1}{n+m} \sum x_i + \sum y_i \right)^{-n-m} \exp(-n-m). \quad (4.98)$$

In the joint space,

$$l(\Theta) = -n \ln \theta - \frac{1}{\theta} \sum x_i - m \ln \mu - \frac{1}{\mu} \sum y_j. \quad (4.99)$$

And so

$$\partial_\theta l(\Theta) = 0 \implies \hat{\theta} = \bar{x} \quad (4.100)$$

$$\partial_\mu l(\mu) = 0 \implies \hat{\mu} = \bar{y}. \quad (4.101)$$

With these,

$$\mathcal{L}(\hat{\Theta}) = \dots \quad (4.102)$$

$$= \frac{1}{(\frac{1}{n} \sum x_i)^n} \exp \left[-\frac{1}{\frac{1}{n} \sum x_i} \sum x_i \right] \frac{1}{(\frac{1}{m} \sum y_i)^m} \exp \left[-\frac{1}{\frac{1}{m} \sum y_i} \sum y_i \right] \quad (4.103)$$

$$= \frac{1}{(\frac{1}{n} \sum x_i)^n} e^{-n} \frac{1}{(\frac{1}{m} \sum y_i)^m} e^{-m}. \quad (4.104)$$

Putting everything together, we find

$$\Lambda = \frac{\mathcal{L}_0}{\mathcal{L}} = \dots = \frac{\frac{1}{(n+m)^{-n-m}} (\sum x_i + \sum y_j)^{-n-m}}{\frac{1}{n^n} \frac{1}{m^m} (\sum x_i)^{-n} (\sum y_i)^{-m}}. \quad (4.105)$$

We reject if $\Lambda < c$, iff

$$\frac{(n+m)^{n+m}}{n^n m^m} \frac{(\sum x_i)^n (\sum y_i)^m}{(\sum x_i + \sum y_j)^{n+m}} < c \quad (4.106)$$

iff (letting c absorb the constant)

$$\frac{(\sum x_i)^n (\sum y_i)^m}{(\sum x_i + \sum y_j)^{n+m}} < c' \quad (4.107)$$

What does the distribution of Λ look like? Notice that we reject if

$$\Lambda = \frac{(n+m)^{n+m}}{n^n m^m} \frac{(\sum x_i)^n (\sum y_i)^m}{(\sum x_i + \sum y_j)^{n+m}} \quad (4.108)$$

$$= \frac{(n+m)^{n+m}}{n^n m^m} \left(\frac{\sum x_i}{\sum x_i + \sum y_j} \right)^n \left(\frac{\sum y_i}{\sum x_i + \sum y_j} \right)^m \quad (4.109)$$

$$= \frac{(n+m)^{n+m}}{n^n m^m} T^n (1-T)^m < c \quad (4.110)$$

which is akin to saying that we reject if $T < a$ or $T > b$ when these constants satisfy

$$a^n (1-a)^m = b^n (1-b)^n. \quad (4.111)$$

What is the distribution of T under H_0 ? Note that under H_0 , $\sum x_i \sim \Gamma(n, \theta)$ and $\sum y_i \sim \Gamma(m, \mu = \theta)$. Use transformation method to show that $T \sim \beta(n, m)$.

Reminder: in large samples, $-2 \ln \Lambda \sim \chi_1^2$

Part 5

Sufficiency

5.1 Measures of Quality of Estimators

Definition 5.1.1. For a given positive integer n , $Y = u(X_1, \dots, X_n)$ is called a **minimum variance unbiased estimator** (MVUE) of the parameter θ if Y is unbiased, i.e., $E[Y] = \theta$, and if $\text{Var}(Y) \leq \text{Var}(Y')$ where Y' is any other unbiased estimator of θ .

From a different standpoint, consider the data X_1, \dots, X_n from a density $f(x_i; \theta)$ and let $Y = u(X_1, \dots, X_n)$ be a statistic from which we will generate the estimator. Then $\delta(Y)$ is the estimator of θ . δ is called the **decision rule**. A value of $\delta(Y)$ is called a **decision**.

The decision can be right or wrong, and how far off we are can be quantified by a **loss function** $L(\theta, \delta(Y))$. The expectation value of $L(\theta, \delta(Y))$ is called the **risk function**.

$$R(\theta, \delta) = E[L(\theta, \delta(Y))] = \int_{-\infty}^{\infty} L(\theta, \delta(Y)) f_Y(y; \theta) dy. \quad (5.1)$$

We would like to minimize this, but usually can't be the loss is dependent on θ . Instead we might choose as our decision function the one that minimizes the maximum of the risk function, i.e. pick the best worst case scenario.

If

$$\boxed{\max_{\theta} [R(\theta, \delta_0(Y))] \leq \max_{\theta} [R(\theta, \delta_0(Y))]} \quad (5.2)$$

then δ_0 is called the **minimax** decision function.

Example 5.1.1. Consider sample X_1, \dots, X_{25} r.v. from $\mathcal{N}(\theta, 1)$. Let $Y = \bar{X}$ and $L[\theta, \delta(Y)] = (\theta - \delta(Y))^2$. Compare two decision functions δ given by

$\delta_1(y) = y$ and $\delta_2(y) = 0$ for $y \in \mathbb{R}$. Then,

$$R(\theta, \delta_1) = E[(\theta - Y)^2] = \frac{\sigma^2}{n} = \frac{1}{25}R(\theta, \delta_2) = E[(\theta - 0)^2] = \theta^2 \quad (5.3)$$

If $\theta = 0$, then δ_2 is a good decision rule, and $R(0, \delta_2) = 0$. If θ differs by more than $1/5$ then $R(\theta, \delta_2) > R(\theta, \delta_1)$. So we choose the function that has the smallest max:

$$\max_{\theta} R(\theta, \delta_1) = \max_{\theta} (1/25) = 1/25. \quad (5.4)$$

If we restrict estimators to those such that $E[\delta(Y)] = \theta$ and $L(\theta, \delta(Y)) = (\theta - \delta(Y))^2$, the decision function that minimizes the risk function yields an unbiased estimator with the minimum variance. We call this the **MVUE**.

If we replace $E[\delta(Y)] = \theta$ with some other condition, then $\delta(Y)$ (if it exists) that minimizes $E[(\theta - \delta(Y))^2]$ uniformly in θ , is called the **minimum-max squared estimator**.

The loss function given by $E[(\theta - \delta(Y))^2] = \text{MSE}(\delta(Y)) = [\text{Bias}(\delta(Y))]^2 + \text{Var}(\delta(Y))$. This makes sense, because if unbiased (bias = 0) then minimizing risk exactly corresponds to minimizing the variance.

Note that the $\text{MSE} - L(\theta, \delta(Y)) = E[(\theta - \delta(Y))^2]$ isn't the only loss function that people use. For example, we could also use $L(\theta, \delta(Y)) = |\delta - \delta(Y)|$, which is called the absolute error loss function.

In a nutshell, the MSE is bias² + variance. So if we limit ourselves to unbiased estimators and minimize the variance then we'll get the MVUE.

We might want to trade some bias for a reduction in variance.

5.2 A Sufficient Statistic for a Parameter

A sufficient statistic is one that captures all of the available information in the sample concerning a parameter

Definition 5.2.1. Let X_1, \dots, X_n denote a random sample of size n from a distribution that has pdf/pmf $f(x; \theta)$, $\theta \in \Omega$. Let $Y_1 = u_1(X_1, \dots, X_n)$ be a statistic whose pdf/pmf is $f_{Y_1}(y_1; \theta)$. Then Y_1 is a **sufficient statistic** for θ iff

$$\frac{\prod_{i=1}^n f(x_i; \theta)}{f_{Y_1}[u_1(X_1, \dots, X_n); \theta]} = H(x_1, \dots, x_n) \quad (5.5)$$

where $H(x_1, \dots, x_n)$ does not depend upon $\theta \in \Omega$.

Another way to put this is that if we condition the joint pdf of X_1, \dots, X_n on Y_1 and that conditional distribution, $f(X_1, \dots, X_n; Y_1)$ does not depend on θ then Y_1 is a sufficient statistic for θ .

Example 5.2.1. Consider $X_i \sim \text{Ber}(p)$ iid. Then $f(x_i) = p^{x_i}(1-p)^{1-x_i}$. Is $Y = \sum x_i$ a sufficient statistic for p ? Well,

$$\frac{\prod^n P(X_i = x_i)}{P(Y_1 = y_1)} = \frac{p^y(1-p)^{n-y}}{\binom{n}{y} p^y(1-p)^{n-y}} = \frac{1}{\binom{n}{y}} \quad (5.6)$$

which does not depend on p , so Y is sufficient.

Theorem 5.2.1 (Neyman – Factorization Theorem). Let X_1, \dots, X_n be a random sample from a distribution that has a pdf or pmf $f(x; \theta), \theta \in \Omega$. The statistic $Y_1 = u_1(X_1, \dots, X_n)$ is a sufficient statistic for θ iff we can find two nonnegative functions k_1, k_2 such that

$$f(x_1; \theta) \dots f(x_n; \theta) = k_1[u_1(x_1, \dots, x_n); \theta]k_2(x_1, \dots, x_n)$$

where k_2 does not depend on θ .

Example 5.2.2. Let $X_1, \dots, X_n \sim \text{Poi}(\lambda)$. We want to find a sufficient statistic for λ . Well,

$$\begin{aligned} p(X_1, x_2, \dots, X_n; \lambda) &= \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \\ &= e^{-n\lambda} \lambda^{\sum x_i} \frac{1}{\prod x_i!} \\ &= \left[e^{-n\lambda} \lambda^{\sum x_i} \right] \left[\frac{1}{\prod x_i!} \right] \end{aligned} \quad (5.7)$$

The first term is k_1 and the second term is k_2 . $\sum x_i$ is sufficient for λ . We can verify that $\sum x_i$ is sufficient using the definition. We need to know the distribution of $\sum x_i$, which is just $\text{Poi}(n\lambda)$. Well,

$$\frac{P(x_1, \dots, X_n; \lambda)}{P_Y(y)} = \frac{\frac{e^{-n\lambda} \lambda^y}{\prod_{i=1}^n x_i!}}{\frac{e^{-n\lambda} (n\lambda)^y}{y!}} = \frac{y!}{\prod x_i!} \frac{1}{ny}. \quad (5.8)$$

which is something that does not depend on λ , which means $\sum x_i$ is sufficient.

5.3 Properties of a Sufficient Statistic

There are many sufficient statistics for a parameter θ . Any “statistic” that contains all the information in the sample useful for estimating θ is efficient. The observations themselves (X_1, \dots, X_n) are sufficient, as well as the order statistics $(X_{(1)}, \dots, X_{(n)})$. Any 1-1 function of a sufficient statistic is sufficient.

Factorization theorem says: $f(x_1, \dots, x_n) = k_1[Y_1; \theta]K_2[x_1, \dots, x_n]$ where Y_1 is a sufficient statistic.

The sufficient statistic that best captures the information about θ in the sample is called the **minimal sufficient statistic**.

Theorem 5.3.1 (Rao-Blackwell). Let X_1, \dots, X_n , n a positive integer, denote a random sample from a distribution that has a pdf/pmf $f(x_i; \theta), \theta \in \Omega$. Let $Y_1 = u_1(X_1, \dots, X_n)$ be a sufficient statistic for θ , and $Y_2 = u_2(X_1, \dots, X_n)$, not a function of Y_1 alone, be an unbiased estimator of θ . Then $E[Y_2|y_1] = \varphi(y_1)$ defines a statistic $\varphi(Y_1)$. This statistic $\varphi(Y_1)$ is a function of the sufficient statistic for θ ; it is an unbiased estimator of θ , and its variance is less than or equal to that of Y_2 .

Rao-Blackwell tells us that if we want to find the minimum variance unbiased estimator (MVUE) we can restrict our search to functions of sufficient statistics.

Theorem 5.3.2. Let X_1, \dots, X_n be random sample from $f(x_i; \theta)$. If a sufficient statistic $Y_1 = u_1(X_1, \dots, X_n)$ for θ exists and if a mle $\hat{\theta}$ of θ also exists uniquely, then $\hat{\theta}$ is a function of $Y_1 = u_1(X_1, \dots, X_n)$.

Example 5.3.1. Let X_1, \dots, X_n be a random sample from a Weibull distribution.

$$f(x_i; \theta) = \frac{2x_i}{\theta} e^{-x_i^2/\theta}; \quad x_i > 0. \quad (5.9)$$

The likelihood function is

$$\mathcal{L}(\theta) = \cdots = \left(\frac{2}{\theta} \right)^n \left(\prod_{i=1}^n x_i \right) e^{-\sum x_i^2/\theta}. \quad (5.10)$$

Find sufficient statistic :

$$\begin{aligned} K_1(Y_1; \theta) &= \left(\frac{2}{\theta} \right)^n e^{-\sum x_i^2/\theta} \\ K_2(X_1, \dots, X_n) &= \prod_{i=1}^n x_i. \end{aligned} \quad (5.11)$$

So, $Y_1 = \sum x_i^2$ is our sufficient statistic. What is the expectation of Y_1 ? It turns out that $E[Y_1] = n\theta \implies \hat{\theta} = \frac{1}{n} \sum x_i^2$, and so $E[\hat{\theta}] = \theta \implies$ MVUE.

5.4 Completeness and Uniqueness

5.4.1 Completeness

Definition 5.4.1. Let the r.v. Z of either the continuous type or the discrete type have a pdf or pmf that is one member of the family $\{h(z; \theta) : \theta \in \Omega\}$. If the condition $E[u(Z)] = 0 \forall \theta \in \Omega$, requires that $u(z)$ be zero except on a set of points that has probability zero for each $h(z; \theta), \theta \in \Omega$, then the family $\{h(z; \theta) : \theta \in \Omega\}$ is called the **complete family** of probability density or mass functions.

Example 5.4.1. If $az^2 + bz + c = 0$ for more than two values of z then $a = b = c = 0$. Use this show that the family of binomial distributions with

$n = 2, p = \theta$ is complete.

Set $E[u(Z)] = 0$. Well,

$$\begin{aligned} E[u(Z)] &= \sum_{z=0}^2 u(z) \binom{2}{z} \theta^z (1-\theta)^{2-z} \\ &= u(0)(1-\theta)^2 + 2u(1)\theta(1-\theta) + u(2)\theta^2 \\ &= [\dots]\theta^2 + [\dots]\theta + [\dots] \\ &= 0 \end{aligned}$$

iff $u(0) = u(1) = u(2) = 0$. So, the binomial family $(2, \theta)$ is complete.

Example 5.4.2. Let

$$f(x; \theta) = \begin{cases} 1/2\theta, & x \in (-\theta, \theta), \theta \in \mathbb{R}^+ \\ 0, \text{else} \end{cases} \quad (5.12)$$

Is this family complete? Well, take $u(X) = X$, then because $E[X] = \dots = 0$ but $u(X) = X \neq 0$, the family is NOT complete.

5.4.2 Uniqueness

Theorem 5.4.1 (Lehmann and Scheffé). Let X_1, \dots, X_n , n a fixed positive integer, denote a random sample from a distribution that has a pdf or pmf $f(x; \theta), \theta \in \Omega$, let $Y_1 = u_1(X_1, \dots, X_n)$ be a sufficient statistic for θ , and let the family $\{f_{Y_1}(y_1; \theta) : \theta \in \Omega\}$ be complete. If there is a function of Y_1 that is an unbiased estimator of θ , then this function of Y_1 is the unique MVUE of θ .

In this case, we say that Y_1 is complete and sufficient.

Example 5.4.3. Let X_1, \dots, X_n be r.v. iid of Uniform $(0, \theta)$ with $0 < \theta < \infty$. We want to find a sufficient statistic for θ . Well,

$$f(x_1, \dots, x_n) = \frac{1}{\theta^n} I(X_{(1)} > 0) I(X_1 < \theta). \quad (5.13)$$

Using factorization theorem, we have that

$$\begin{aligned} K_1[Y_1, \theta] &= \frac{1}{\theta^n} I(Y_{(n)} > 0) \\ K_2[X_1, \dots, X_n] &= I(X_{(1)} > 0). \end{aligned} \quad (5.14)$$

So, $Y_{(1)}$ (the max) is sufficient for θ . This turns out to be minimal too. Now,

$$g_{(n)}(Y) = ny^{n-1}\theta^{-n} \quad (5.15)$$

is the density for $Y_{(1)}$. Suppose $g(t)$ is a function that satisfies $E[g(t)] = 0$, so

$$E[g(t)] = \int_0^\theta g(t) nt^{n-1}\theta^{-n} dt = 0. \quad (5.16)$$

If this holds then taking ∂_θ of both sides we get $\theta^{-1}ng(\theta) = 0 \implies g(\theta) = 0$. So the family of $f_{Y_1}(y_1)$ is complete.

Now, $E[Y_1] = \int_0^\theta xnx^{n-1}/\theta = n\theta/(n+1)$, and so $(n+1)Y_1/n$ is an unbiased estimator that is a function of a sufficient statistic that has a complete distribution (Y_1 is complete sufficient statistic). So $(n+1)/nX_{(n)} \equiv (n+1)\max/n$ is a MVUE.

5.5 The Exponential Class of Distributions

If we can determine that a distribution belongs to this class, a complete sufficient can be readily determined. Consider a family $\{f(x, \theta) : \theta \in \Omega\}$ where $\Omega = \{\theta : \gamma < \theta < \delta\}$ where γ, δ are constants in the extended reals (they may be $\pm\infty$) and

$$f(x; \theta) = \begin{cases} \exp [p(\theta)K(x) + H(x) + q(\theta)], & x \in \mathcal{S} \\ 0, & \text{else} \end{cases} \quad (5.17)$$

where \mathcal{S} is the support of X .

Definition 5.5.1 (Regular Exponential Class). A pdf of the form above is said to be a member of the **regular exponential class** of probability density or mass functions if

1. \mathcal{S} the support of X , does not depend on θ .
2. $p(\theta)$ is a nontrivial continuous function of $\theta \in \Omega$
3.
 - If X is continuous, then each $K'(x) \neq 0$ and $H(x)$ is a continuous function of $x \in \mathcal{S}$.
 - If X is a discrete, then $K(x)$ is a nontrivial function of $x \in \mathcal{S}$.

If the density is the regular exponential class, for a sample, we can write the density as $\exp [p(\theta) \sum K(x_i) + \sum H(x_i) + nq(\theta)]$, and so the factorization theorem tells us that the statistic $Y = \sum K(x_i)$ is sufficient.

Theorem 5.5.1. Let X_1, \dots, X_n denote a random sample from a distribution that represents a regular case of the exponential class, with pdf or pmf given as above. Consider the statistic $Y_1 = \sum^n K(x_i)$. Then

- The pdf or pmf for Y_1 has the form

$$f_{Y_1}(y_1; \theta) = R(y_1) \exp [p(\theta)y_1 + nq(\theta)] \quad (5.18)$$

for $y_1 \in \mathcal{S}_{Y_1}$ and some function $R(y_1)$. Neither \mathcal{S}_{Y_1} nor $R(y_1)$ depends on θ .

- $E[Y_1] = -nq'(\theta)/p'(\theta)$.

- $\text{Var}[Y_1] = n \frac{1}{p'(\theta)^3} \{p''(\theta)q'(\theta) - q''(\theta)p'(\theta)\}$

Theorem 5.5.2. Consider $f(x; \theta)$, $\gamma < \theta < \delta$, be a pdf or pmf of a r.v. X whose distribution is a regular case of the exponential class. Then if X_1, \dots, X_n is a random sample from this distribution then $Y_1 = \sum^n K(X_i)$ is a complete sufficient statistic for θ .

Theorem 5.5.3. If Y_1 is a complete sufficient statistic and $\varphi(Y_1)$ is an unbiased estimator for θ , then $\varphi(Y_1)$ is the MVUE.

Example 5.5.1. Consider $\Gamma(2, \theta)$. Then

$$\begin{aligned} f(x; \theta) &= xe^{-x/\theta}\theta^{-2}, \quad x > 0 \\ &= \exp[-x/\theta + \ln x - 2\ln\theta] \end{aligned} \tag{5.19}$$

where $p(\theta) = -1/\theta$, $K(x) = x$, $H(x) = \ln x$, and $q(\theta) = -2\ln\theta$. For a sample of size n , we just have $\exp[(-1/\theta)\sum x_i + \sum \ln x_i - 2n\ln\theta]$, and so $Y_1 = \sum x_i \sim \Gamma(2n, \theta)$ is a complete sufficient statistic. We know what $\Gamma(2n, \theta)$ looks like, so we can go ahead and put it in the form

$$f_{Y_1}(y_1) = \frac{y_1^{2n-1}}{\Gamma(2n)} \exp\left[\frac{-1}{\theta}y_1 - 2n\ln\theta\right] \equiv R(y_1) \exp[p(\theta)y_1 + nq(\theta)]. \tag{5.20}$$

So, $p'(\theta) = 1/\theta^2$ and $q'(\theta) = -2/\theta$ and $p''(\theta) = -2/\theta^3$ and $q''(\theta) = 2/\theta^2$. Using the theorem we get $E[Y_1] = 2n\theta$ and $\text{Var}[Y_1] = 2n\theta^2$, which is what we would expect for $\Gamma(2n, \theta)$.

Example 5.5.2. Let X_1, \dots, X_n be iid with $\mathcal{N}(\theta, \sigma^2)$. Find the MVUE of θ . Well,

$$\begin{aligned} f(x; \theta) &= (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-1}{2\sigma^2}(x - \theta)^2\right] \\ &= \exp\left[\frac{\theta}{\sigma^2}x - \frac{x^2}{2\sigma^2} - \ln\sqrt{2\pi\sigma^2} - \frac{\theta^2}{2\sigma^2}\right]. \end{aligned} \tag{5.21}$$

Identifying the functions, we find that $Y_1 = \sum X_i$ is a complete sufficient statistic for θ . Also, $E[Y_1] = n\theta$, so \bar{X} is the MVUE for θ .

5.6 Functions of a Parameter

In this section we're interested in finding the MVUE of a parameter, say θ . Unlike MLEs, MVUEs don't have the invariance property. Additionally, MVUEs don't have known asymptotic distributions either. There are two main ways we can try to find the MVUE of a function of θ :

- We can use the conditional expectation of an unbiased estimator given a sufficient statistic.

- We can look at the expected value of a sufficient statistic (plugged into the function) and correct the bias (if we can).

Example 5.6.1. Let's say we have X_1, \dots, X_n , iid with $X_i \sim \text{Poi}(\theta)$. We want to find the MVUE for $P(X \leq 1)$. Well,

$$P(X \leq 1) = \sum_{k=0}^1 \frac{e^{-\theta}\theta^k}{k!}. \quad (5.22)$$

Now, a sufficient statistic for θ is $Y = \sum X_i$. Then

$$\begin{aligned} P(X_1 \leq 1 | Y = y) &= P(X_1 = 0 | Y = y) + P(X_1 = 1 | Y = y) \\ &= \frac{P(X_1 = 0 \cap Y = \sum_{i=2}^n x_i = y)}{P(Y = y)} + \frac{P(X_1 = 1 \cap Y = \sum_{i=2}^n x_i = y - 1)}{P(Y = y)} \\ &= \frac{e^\theta e^{-(n-1)\theta} ((n-1)\theta)^y / y!}{e^{-n\theta} (n\theta)^y / y!} + \frac{\theta e^\theta e^{-(n-1)\theta} ((n-1)\theta)^{y-1} / (y-1)!}{e^{-n\theta} (n\theta)^y / y!} \\ &= \dots \\ &= \left(\frac{n-1}{n} \right)^y \left[1 + \frac{y}{n-1} \right]. \end{aligned} \quad (5.23)$$

This is the MVUE of $P(X \leq 1)$.

Example 5.6.2. Let X_1, \dots, X_n with $X_i \sim \text{Exp}(\theta)$. Find the MVUE for $\text{Var}[Y_1]$. We know that $\text{Var}[Y_1] = \theta^2$. We note that this is a member of the regular exponential class. The likelihood function is

$$\mathcal{L}(\theta) = \theta^{-n} \exp \left[-\frac{1}{\theta} \sum x_i \right]. \quad (5.24)$$

$\sum X_i$ is a sufficient statistic and is complete, so \bar{X} is an MVUE. Now, we try $(\bar{X})^2$, well:

$$E[\bar{X}^2] = \text{Var}[\bar{X}] + E[\bar{X}]^2 = \frac{\theta^2}{n} + \theta^2 = \frac{n+1}{n} \theta^2 \quad (5.25)$$

is biased. So, $n\bar{X}^2/(n+1)$ is an unbiased estimator of θ^2 , and it is a function of the minimal sufficient statistic $\sum X_i$. So it is the MVUE of $\text{Var}[Y_1] = \theta^2$

The distribution of the MVUE is not necessarily known. However, CIs are often generated by using bootstrap samples and calculating the $\hat{\theta}_{MVUE}$ and $SE(\hat{\theta}_{MVUE})$ for each where

$$SE = \sqrt{\frac{1}{\#\text{bootstrap}-1} \sum \left[\sum \hat{\theta}^* - \bar{\hat{\theta}}^* \right]^2} \quad (5.26)$$

5.7 The Case of Several Parameters

Definition 5.7.1. Let X_1, \dots, X_n denote a random sample from a distribution that has a pdf or pmf $f(x; \boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^p$. Let \mathcal{S} denote the support of X . Let \mathbf{Y} be an m -dimensional random vector of statistics $\mathbf{Y} = (Y_1, \dots, Y_m)'$ where $Y_i = u(X_1, \dots, X_n)$ for $i = 1, \dots, m$. Denote the pdf or pmf of \mathbf{Y} by $f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta})$ for $\mathbf{y} \in \mathbb{R}^m$. The random vector of statistics \mathbf{Y} is **jointly sufficient** for $\boldsymbol{\theta}$ iff

$$\frac{\prod_{i=1}^n f(x_i; \boldsymbol{\theta})}{f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta})} = H(x_1, \dots, x_n) \quad (5.27)$$

for all $x_i \in \mathcal{S}$ where $H(\dots)$ does not depend on $\boldsymbol{\theta}$.

This means that the factorization theorem that we used before extends into the multiparameter case.

Example 5.7.1. Suppose X_1, \dots, X_n are iid $\mathcal{N}(\mu, \sigma^2)$ r.v.'s. Find sufficient statistics for both unknowns. The likelihood function is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \right] \\ &= \dots \\ &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp \left[-\frac{1}{2\sigma^2} \left(\sum x_i^2 - 2\mu \sum x_i - n\mu^2 \right) \right] \end{aligned} \quad (5.28)$$

The factorization theorem tells us that $\sum x_i^2$ and $\sum x_i$ are sufficient statistics for μ, σ^2 . Now, \bar{X} is an unbiased estimator for μ and $S^2 = (1/(n-1)) \sum X_i^2$ is an unbiased estimator for σ^2 , they are both MVUEs.

Definition 5.7.2. Let X be a r.v. with pdf/pmf $f(x; \boldsymbol{\theta})$ with $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^m$. If X is continuous, assume that the support of X , $\mathcal{S} = (a, b)$ where a and/or b can be $\pm\infty$. If X is discrete, assume that $\mathcal{S} = \{a_1, \dots, \}\$. Suppose $f(x; \boldsymbol{\theta})$ is of the form

$$f(x; \boldsymbol{\theta}) = \begin{cases} \exp [\sum^m x_i p_j(\boldsymbol{\theta}) K_j(x) + H(x) + q(\theta_1, \dots, \theta_m)], & x \in \mathcal{S} \\ 0, \text{else} \end{cases} \quad (5.29)$$

Then we say that this pdf/pmf is a member of the **exponential class**. We say it is a **regular case** of the exponential family if, in addition,

- \mathcal{S} does not depend on the vector of parameters $\boldsymbol{\theta}$
- The space Ω contains a nonempty, m -dimensional open rectangle
- the $p_j(\boldsymbol{\theta})$ with $j = 1, \dots, m$ are nontrivial, functionally independent, continuous functions of $\boldsymbol{\theta}$.
- and, depending on whether X is continuous or discrete, one of the following holds, respectively:

- if X is continuous, then the derivatives $K'_j(x)$ for $j = 1, \dots, m$ are continuous on (a, b) and no one is a linear homogeneous function of the others, and $H(x)$ is a continuous function on (a, b) .
- if X is discrete, the $K_j(x)$ with $j = 1, \dots, m$ are nontrivial functions of x on \mathcal{S} and no one is a linear homogeneous function of the others
- .

Theorem 5.7.1. If f is a member of the regular exponential class with $\mathbf{Y} = \sum K_j(x)$ being a sufficient statistic then

$$R(\mathbf{Y}) \exp \left[\sum p_j(\boldsymbol{\theta}) Y_j + nq(\boldsymbol{\theta}) \right] \quad (5.30)$$

is the joint density of the sufficient statistic.

5.8 Minimal Sufficiency and Ancillary Statistics

The sufficient statistics that best summarizes the information about $\boldsymbol{\theta}$ is called minimal. In general, if the distribution depends on K parameters then we can find K minimal sufficient statistics.

Example 5.8.1. Let X_1, \dots, X_n be random sample from the uniform $(\theta - 1, \theta + 1)$. $f(x_1, \dots, x_n) = 1/2$; $Y_{(1)} > \theta - 1$ and $Y_{(n)} < \theta + 1$. Both $Y_{(1)}$ and $Y_{(n)}$ are jointly sufficient for θ (and are minimal).

Theorem 5.8.1. Complete and sufficient statistics are all minimal, but not all minimal sufficient statistics are complete.

5.8.1 Other types of families of distributions

Theorem 5.8.2. Let $f(x)$ be a valid density function, then $(1/\sigma)f((x - \mu)/\sigma)$ is also a valid density for $\sigma > 0$.

Definition 5.8.1 (Location family). Let $f(x)$ be any density. Then the family of density functions $f(x - \mu)$, indexed by $-\infty < \mu < \infty$, is called the **location family** with standard density $f(x)$, and μ is called the location parameter for the family.

Example 5.8.2. • Normal distributions with the same variance

- Cauchy
- Double exponential
- Shifted exponential

Definition 5.8.2 (Scale family). Let $f(x)$ be any density. Then the family of densities, assuming that $\sigma > 0$, given by $(1/\sigma)f(x/\sigma)$ is called a scale family with standard density $f(x)$ and scale parameter σ .

Example 5.8.3. • Normals with fixed μ .

- Gamma distributions with fixed α
- Double exponential
-

Definition 5.8.3 (Location-scale family). Let $f(x)$ be any density. For any μ finite and $\sigma > 0$. Then the family $(1/\sigma)f(x/\sigma)$ indexed by parameter μ, σ is called the location-scale family with standard density $f(x)$. μ is the location parameter and σ is called the scale parameter.

Example 5.8.4.

Normal distributions

Cauchy distributions

Definition 5.8.4 (Location-invariant statistics). Location-invariant statistics are statistics that do not change when used within a location framework.

Example 5.8.5. $S^2 = (1/(n-1)) \sum (x_i - \bar{x})^2$ for $x_i \in \mathcal{N}(\mu, \sigma^2)$. What is S^2 for $X_i + \mu$? The answer is NO. So S^2 is an location-invariant statistic.

Example 5.8.6. The range is also a location-invariant statistic.

Definition 5.8.5 (Scale-invariant statistics). Scale-invariant statistics are statistics that do not change when used within a scale framework.

Example 5.8.7. Suppose $Z_i = \sigma X_i$. Then $X_1/(X_1 + X_2)$ is an scale-invariant statistic. Another example is the ratio between the min and the max.

Definition 5.8.6 (Location-scale-invariant statistics). Location-scale-invariant statistics are statistics that do not change when used within a location-scale framework.

Example 5.8.8. Consider $Z_i = \mu + \sigma X_i$ then $(X_i - \mu)/S_x$ is a location-scale invariant statistic.

5.9 Sufficiency, Completeness, and Independence

Ancillary statistics are statistics with distributions that do not depend on the parameter(s) of interest and contain no information about the parameter(s) of interest. Ancillary statistics are kind of the like the “opposite” of sufficient statistics.

Definition 5.9.1 (Ancillary statistic). An ancillary statistic is a statistic whose distribution does not depend on θ . To be more precise, the statistic $S(\vec{x})$ is ancillary for θ if it is the same for all values of $\theta \in \Omega$.

Example 5.9.1. S^2 is an ancillary statistic. From a normal, S^2 dependence on θ gets canceled out. Explicitly, let X_1, \dots, X_n be iid $\mathcal{N}(\mu, \sigma^2)$ and let $S^2 = (1/(n-1)) \sum (x_i - \bar{x})^2$. We know that $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$. So $S^2 \sim (\sigma^2/(n-1))\chi_{n-1}^2$. Now, $\chi_{n-1}^2 = \Gamma((n-1)/2, 2)$. So, S^2 has nothing to do with μ . So we say S^2 is ancillary for the mean: $\{\{\mu, \sigma^2\} : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\}$, but not ancillary for the variance: $\{\{\mu, \sigma^2\} : \mu \in \mathbb{R}, \sigma^2 > 0\}$. It's also worth noting that S^2 can be ancillary for any location family.

Theorem 5.9.1 (Basu's Theorem). Let X_1, \dots, X_n be iid from a density $f(x; \theta)$ with $\theta \in \Omega$. Suppose that the statistic Y_1 is complete sufficient for θ and $Z = u(X_1, \dots, X_n)$ be any other statistic that is NOT a function of Y_1 alone. If the distribution of Z does not depend on θ (that is, Z is ancillary for θ) then Z is INDEPENDENT of the statistic Y_1 .

Consider a random sample from $\mathcal{N}(\mu, \sigma^2)$, then $\sum X_i$ is complete sufficient and S^2 is ancillary for μ . Basu's Theorem tells us that $\sum X_i$ (or \bar{X}) and S^2 are independent.

Part 6

Optimal Test of Hypotheses

6.1 Most Powerful Tests

Recall that **power** is the probability that H_0 is true provided that H_a is true.

Definition 6.1.1. Let C denote the subset of the sample space. We say that C is the **best critical region** of size α for testing the simple hypothesis $H_0 : \theta = \theta'$ against the alternative simple hypothesis $H_1 : \theta = \theta''$ if

- $P_{\theta'}[\mathbf{X} \in C] = \alpha$
- And for every subset A of the sample space

$$P_{\theta'} = \alpha \implies P_{\theta''}[\mathbf{X} \in C] \geq P_{\theta''}[\mathbf{X} \in A]. \quad (6.1)$$

Theorem 6.1.1 (Neyman-Pearson Theorem). Let X_1, \dots, X_n denote a random sample from a distribution that has pdf or pmf $f(x; \theta)$. Then the likelihood of X_1, \dots, X_n is

$$\mathcal{L}(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta). \quad (6.2)$$

Let θ' and θ'' be distinct fixed values of θ so that $\Omega = \{\theta : \theta = \theta', \theta''\}$, and let k be a positive number. Let C be a subset of the sample space such that

1. For each point $\mathbf{x} \in C$

$$\frac{\mathcal{L}(\theta'; \mathbf{x})}{\mathcal{L}(\theta''; \mathbf{x})} \leq k \quad (6.3)$$

2. For each point $\mathbf{x} \in C^c$

$$\frac{\mathcal{L}(\theta'; \mathbf{x})}{\mathcal{L}(\theta''; \mathbf{x})} \geq k \quad (6.4)$$

3. $\alpha = P_{H_0}[\mathbf{X} \in C]$.

Then C is a best critical region of size α for testing the simple hypothesis $H_0 : \theta = \theta'$ against the alternative hypothesis $H_1 : \theta = \theta''$.

The test that maximizes the power of θ_A for a critical region is

$$\frac{\mathcal{L}(\theta')}{\mathcal{L}(\theta'')} < K. \quad (6.5)$$

The value of K is chosen so that the test has the desired α level. And we call such a test the **most powerful** test for H_0 versus H_a .

Example 6.1.1. Suppose we have $f(x) = (1/\theta^2)xe^{-x/\theta}$ with $x > 0; \theta > 0$. Find the most powerful test for $H_0 : \theta = 1$ and $H_A : \theta = 2$ (actually we're interested in the composite alternative $\theta > 1$). Well, the likelihood functions are

$$\mathcal{L}(\theta_0) = xe^{-x}, \quad \mathcal{L}(\theta_A) = (1/4)e^{-x/2}. \quad (6.6)$$

The test is based on

$$\frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\theta_A)} = 4e^{-x/2} < K \implies x > -4 \ln(K/4) \equiv K'. \quad (6.7)$$

We want to find K' for $\alpha = 0.05$. Well,

$$P(x > K' | \theta = \theta_0 = 1) = \alpha = 0.05 = \int_{K'}^{\infty} xe^{-x} dx = 0.05 \quad (6.8)$$

so

$$K' = 4.473. \quad (6.9)$$

So, we reject if $x < 4.473$. This is the most powerful (MP) test.

6.2 Uniformly Most Powerful Test

Definition 6.2.1. A test is called **uniformly most powerful (UMP)** test if the form of the rejection region does not depend on the specific value of the alternative parameter θ_A (that is, if $X > C$ for any constant C).

In our example above, the test is *UMP* for $H_A : \theta > 1$.

Example 6.2.1. Suppose X is a single observation from $f(x; \theta) = \theta x^{\theta-1}$ with $0 \leq x \leq 1$, with $H_0 : \theta = 2, H_a : \theta = 1$. We can easily find that

$$\frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\theta_A)} = \frac{2x}{1} = 2x < K \implies x < K'. \quad (6.10)$$

Set $\alpha = 0.05$, then we integrate under the null

$$\int_0^{K'} 2x dx = 0.05 \implies K' = \sqrt{0.05}. \quad (6.11)$$

So we reject if $x < \sqrt{0.05}$. The form of this does not depend on θ_A , so this test is the UMP for $H_a : \theta < 2$. What is the power of this test? Well,

$$P(x < \sqrt{0.05} | \theta = 1) = \int_0^{0.05} 1 dx = 0.231, \quad (6.12)$$

pretty lower power, but most powerful nevertheless. However, this test is NOT UMP for $H_a : \theta > 2$ because we end up with $x > K'$.

Part 7

Problems

7.1 Problem Set 1

3.6.4

- (a) X has a standard normal distribution:

```
x=seq(-6,6,.01); plot(dnorm(x)^x)
```

- (b) X has a t -distribution with 1 degree of freedom.

```
lines(dt(x,1)^x,lty=2)
```

- (c) X has a t -distribution with 3 degrees of freedom.

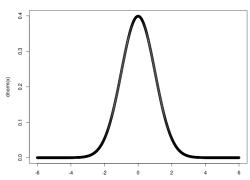
```
lines(dt(x,3)^x,lty=2)
```

- (d) X has a t -distribution with 10 degrees of freedom.

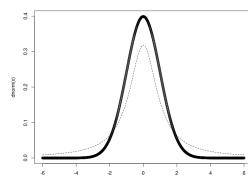
```
lines(dt(x,10)^x,lty=2)
```

- (e) X has a t -distribution with 30 degrees of freedom.

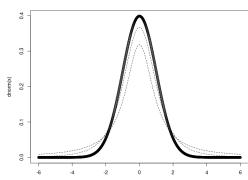
```
lines(dt(x,30)^x,lty=2)
```



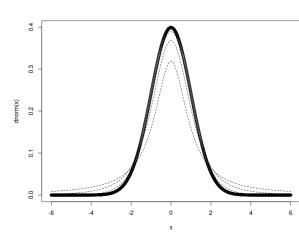
(a)



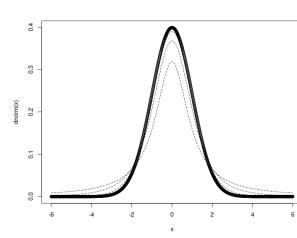
(b)



(c)



(d)



(e)

3.6.5

(a) $P(|X| \geq 2) = 2 \times [1 - P(X \leq 2)] = \mathbf{0.046}.$

```
> 2*(1 - pnorm(2))  
[1] 0.04550026
```

(b) $P(|X| \geq 2) = 2 \times [1 - P(X \leq 2)] = \mathbf{0.295}.$

```
> 2*(1 - pt(2,1))  
[1] 0.2951672
```

(c) $P(|X| \geq 2) = 2 \times [1 - P(X \leq 2)] = \mathbf{0.139}.$

```
> 2*(1 - pt(2,3))  
[1] 0.139326
```

(d) $P(|X| \geq 2) = 2 \times [1 - P(X \leq 2)] = \mathbf{0.073}.$

```
> 2*(1 - pt(2,10))  
[1] 0.07338803
```

(e) $P(|X| \geq 2) = 2 \times [1 - P(X \leq 2)] = \mathbf{0.055}.$

```
> 2*(1 - pt(2,30))  
[1] 0.05462504
```

3.6.11: Let $T = W/\sqrt{V/r}$, where the independent variables $W \sim \mathcal{N}(0, 1)$ and $V \sim \chi^2(r)$. Show that $T^2 \sim F(r_1 = 1, r_2 = r)$. *Hint:* What is the distribution of the numerator of T^2 ?

Solution: Let the independent random variables U, V be given, with $W \sim \mathcal{N}(0, 1)$ and $U \sim \chi^2(r)$. The random variable T^2 , where $T = W/\sqrt{V/r}$ is given by

$$T^2 = \left(\frac{W}{\sqrt{V/r}} \right)^2 = \frac{W^2}{V/r}. \quad (7.1)$$

Because $W \sim \mathcal{N}(0, 1)$, we have that $W^2 \sim \chi^2(1)$ (by theorem). Now, T^2 has the form

$$T^2 = \frac{W^2}{V/r} = \frac{W^2/1}{V/r} \quad (7.2)$$

where 1 is the df of $\chi^2(1)$ which W follows, and r is the df of $\chi^2(r)$ which U follows. Thus, $T^2 \sim F(1, r)$, by the definition of the F -distribution. \square

3.6.15: Let X_1, X_2 be iid with common distribution having the pdf

$$f(x) = \begin{cases} e^{-x}, & 0 < x < \infty \\ 0, & \text{else} \end{cases} \quad (7.3)$$

Show that $Z = X_1/X_2$ has an F -distribution.

Solution: It suffices to show that Z can be written as a ratio of two χ^2 -distributed independent random variables. To this end, we can consider the mgf $M_X(t)$ of X_1 , which is also identically that of X_2 since X_1, X_2 are iid:

$$M_X(t) = E[e^{tx}] = \int_0^\infty e^{tx} e^{-x} dx = (1-t)^{-1}. \quad (7.4)$$

However, this does not quite match the mgf for a $\chi^2(2)$. To circumvent this problem, we rewrite

$$Z = \frac{X_1}{X_2} = \frac{2X_1/2}{2X_2/2} = \frac{(X_1 + X_1)/2}{(X_2 + X_2)/2}, \quad (7.5)$$

as we expect $r = 2$. Let $Y_1 = X_1 + X_1$. Then we have trivially $Y_1 = 2X_1$, and so $|J| = 1/2$. With this, Y_1 has the pdf

$$\tilde{f}_Y(y) = |J|f(x) = \frac{1}{2}f(x) = \begin{cases} \frac{1}{2}e^{-y/2}, & 0 < y < \infty \\ 0, & \text{else} \end{cases}. \quad (7.6)$$

From here, we find the mgf of Y_1 to be

$$M_{Y_1}(t) = E[e^{ty}] = \frac{1}{2} \int_0^\infty e^{ty} e^{-y/2} dy = (1-2t)^{-1} = (1-2t)^{-2/2}, t < \frac{1}{2}. \quad (7.7)$$

By symmetry, $M_{Y_2}(t)$ is identically $M_{Y_1}(t)$, and both are the mgf for $\chi^2(r = 2)$. Because each mgf uniquely determines a pdf, $Y_1, Y_2 \sim \chi^2(r = 2)$ identically and independently (for each depends exclusively on X_1, X_2 , respectively). Therefore,

$$Z = \frac{(X_1 + X_1)/2}{(X_2 + X_2)/2} = \frac{Y_1/2}{Y_2/2} \quad (7.8)$$

follows the F -distribution with degrees of freedom $r_1 = r_2 = 2$, by definition. \square

3.6.16: Let X_1, X_2, X_3 be independent r.v. with $X_i \sim \chi^2(r_i)$.

- (a) Show that $Y_1 = X_1/X_2$ and $Y_2 = X_1 + X_2$ are independent and that $Y_2 \sim \chi^2(r_1 + r_2)$.
- (b) Deduce that

$$\frac{X_1/r_1}{X_2/r_2} \text{ and } \frac{X_3/r_3}{(X_1 + X_2)/(r_1 + r_2)} \quad (7.9)$$

are independent F -variables.

Solution:

- (a) We consider the transformation

$$y_1 = u(x_1, x_2) = \frac{x_1}{x_2} \quad (7.10)$$

$$y_2 = v(x_1, x_2) = x_1 + x_2. \quad (7.11)$$

whose inverse is

$$\begin{aligned} x_1 &= \bar{u}(y_1, y_2) = \frac{y_1 y_2}{1 + y_1} \\ x_2 &= \bar{v}(y_1, y_2) = \frac{y_2}{1 + y_1}. \end{aligned} \quad (7.12)$$

The absolute value of the Jacobian is

$$|J| = \left| \det \begin{pmatrix} \partial_{y_1} \bar{u} & \partial_{y_2} \bar{u} \\ \partial_{y_1} \bar{v} & \partial_{y_2} \bar{v} \end{pmatrix} \right| = \frac{y_2}{(1 + y_1)^2}, \quad (7.13)$$

which maps one-to-one from the space of $X_1, X_2 \in \mathbb{R}^+ \times \mathbb{R}^+$ onto the space of $Y_1, Y_2 \in \mathbb{R}^+ \times \mathbb{R}^+$. Since X_1, X_2 are independent, we consider the joint pdf of X_1, X_2 :

$$h(x_1, x_2) = \begin{cases} \frac{x_1^{r_1/2-1} x_2^{r_2/2-1}}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} e^{-(x_1+x_2)/2}, & 0 < x_1, x_2 < \infty \\ 0, & \text{else} \end{cases} \quad (7.14)$$

from which we can deduce the joint pdf for Y_1, Y_2 :

$$\begin{aligned} \tilde{h}(y_1, y_2) &= |J|h\left(\frac{y_1 y_2}{1 + y_1}, \frac{y_2}{1 + y_1}\right) \\ &= \begin{cases} \frac{y_2(y_1 y_2)^{r_1/2-1} y_2^{r_2/2-1} (1+y_1)^{-r_1/2-r_2/2+\frac{1}{2}}}{(1+y_1)^2 \Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} e^{-y_2/2}, & 0 < y_1, y_2 < \infty \\ 0, & \text{else} \end{cases} \\ &= \begin{cases} \frac{y_2^{r_1/2+r_2/2-1} y_1^{r_1/2-1} (1+y_1)^{-r_1/2-r_2/2}}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} e^{-y_2/2}, & 0 < y_1, y_2 < \infty \\ 0, & \text{else} \end{cases} \end{aligned} \quad (7.15)$$

Without further computation we see that $\tilde{h}(y_1, y_2)$ can be written as a product of two nonnegative functions of y_1 and y_2 . In view of Theorem 2.4.1, Y_1 and Y_2 are independent. \square

Next, we wish to show $Y_2 \sim \chi^2(X_1, X_2)$, to which end we find the marginal pdf $g_2(y_2)$ of Y_2 :

$$\begin{aligned} g_2(y_2) &= \int_0^\infty \tilde{h}(y_1, y_2) dy_1 \\ &= \mathfrak{C} \int_0^\infty y_1^{r_1/2-1} (1+y_1)^{-r_1/2-r_2/2} dy_1 \\ &= \mathfrak{C} \frac{\Gamma(r_1/2)\Gamma(r_2/2)}{\Gamma[(r_1+r_2)/2]} \end{aligned} \quad (7.16)$$

where \mathfrak{C} contains all the y_1 -independent elements. From here, via simple back-substitution we obtain the marginal pdf for Y_2 :

$$g_2(y_2) = \begin{cases} \frac{y_2^{(r_1+r_2)/2-1}}{\Gamma[(r_1+r_2)/2]2^{(r_1+r_2)/2}} e^{-y_2/2}, & 0 < y_2 < \infty \\ 0, & \text{else} \end{cases}, \quad (7.17)$$

i.e., $Y_2 \sim \chi^2(r_1 + r_2)$. \square

Mathematica code:

```
In[20]:= Integrate[
x^(r1/2 - 1) (1 + x)^(-r1/2 - r2/2), {x, 0, Infinity}]

Out[20]= ConditionalExpression[(Gamma[r1/2] Gamma[r2/2])/
Gamma[(r1 + r2)/2], Re[r2] > 0 && Re[r1] > 0]
```

- (b) By definition, because X_1, X_2 are independent random variables with $X_i \sim \chi^2(r_i)$,

$$\Omega = \frac{X_1/r_1}{X_2/r_2} \sim F(r_1, r_2). \quad (7.18)$$

Also, because $X_3 \sim \chi^2(r_3)$ and $(X_1 + X_2) \sim \chi^2(r_1 + r_2)$ (from (a)), we have

$$\Lambda = \frac{X_3/r_3}{(X_1 + X_2)/(r_1 + r_2)} \sim F(r_3, r_1 + r_2) \quad (7.19)$$

as well. Furthermore, because

$$\Omega = \frac{X_1/r_1}{X_2/r_2} = \frac{r_2}{r_1} Y_1 \quad (7.20)$$

$$\Lambda = \frac{r_1 + r_2}{r_3} \frac{X_3}{Y_2} \quad (7.21)$$

and because X_1, X_2, X_3 are independent, we have that Y_1, Y_2, X_3 are independent. Therefore, it is necessary that $\Omega \sim F(r_1, r_2)$ and $\Lambda \sim F(r_3, r_1 + r_2)$ are independent as well. \square

4.1.1 Twenty motors were put on test under a high-temperature setting. The lifetimes in hours of the motors under these conditions are given below. Also, the data are in the file **lifetimemotor.rda** at the site listed in the Preface. Suppose we assume that the lifetime of a motor under these conditions, X , has a $\Gamma(1, \theta)$ distribution.

1	4	5	21	22	28	40	42	51	53
58	67	95	124	124	160	202	260	303	363

- (a) Obtain a histogram of the data and overlay it with a density estimate, using the code `hist(x,pr=T); lines(density(x))` where the R vector x contains the data. Based on this plot, do you think that the $\Gamma(1, \theta)$ model is credible?
- (b) Assuming a $\Gamma(1, \theta)$ model, obtain the maximum likelihood estimate $\hat{\theta}$ of θ and locate it on your histogram. Next overlay the pdf of a $\Gamma(1, \hat{\theta})$ distribution on the histogram. Use the R function `dgamma(x,shape=1,scale=hat_theta)` to evaluate the pdf.
- (c) Obtain the sample median of the data, which is an estimate of the median lifetime of a motor. What parameter is it estimating (i.e., determine the median of X)?
- (d) Based on the mle, what is another estimate of the median of X ?

Solution:

- (a) For some reason R does not recognize the dataset as of numeric type. Because the dataset is small enough, I recoded and fed it by hand to the data vector y :

```
> lines(density(y))
> y <- c(1,4,5,21,22,28,40,42,51,53,58,67,
95,124,124,160,202,260,303,363)
> hist(y,pr=T)
> lines(density(y))
```

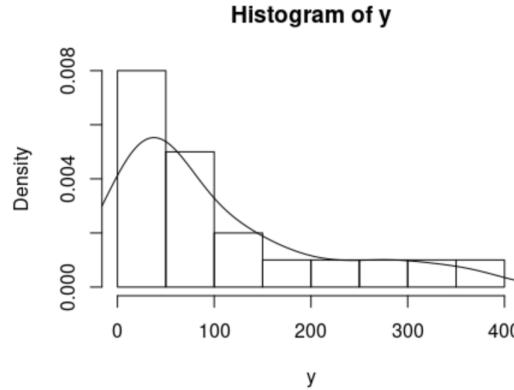
The $\Gamma(1, \theta)$, or $\text{Exp}(\theta)$, model seems to be **credible** as far as the histogram is concerned. However, the overlaying density does not look like a $\Gamma(1, \theta)$. \square

- (b) Assuming the $\Gamma(1, \theta)$ model, then the pdf on the support \mathbb{R}^+ is given by

$$f(y) = \frac{1}{\theta} e^{-y/\theta}, \quad (7.22)$$

from which we obtain the logarithm of the likelihood function:

$$l(\theta) = \log \left(\prod_{i=1}^n \frac{1}{\theta} e^{-y_i/\theta} \right) = -n \log \theta - \frac{1}{\theta} \sum_{i=1}^n y_i. \quad (7.23)$$



The first partial derivative wrt θ is then

$$\partial_\theta l(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n y_i. \quad (7.24)$$

Setting $\partial_\theta l(\theta) = 0$, we get (by inspection) that $l(\theta)$ is extremized iff $\theta = (1/n) \sum_{i=1}^n y_i = \bar{y}$. We also have that $\partial_{\theta\theta} < 0 \forall \theta \in \mathbb{R}^+$, which means $l(\theta)$ is maximized globally at \bar{y} . From here, the statistic

$$\hat{\theta} = \bar{Y} = \mathbf{101.15} \quad (7.25)$$

is the mle of θ . (Also note that because $E[Y] = \theta \implies E[\bar{Y}] = \theta$, $\hat{\theta}$ is an unbiased estimator of θ .)

```
> mean(y)
[1] 101.15
> abline(v = mean(y), lwd=3, lty=2)
> z=dgamma(y, shape=1, scale=mean(y))
> lines(z~y, lty=2)
```

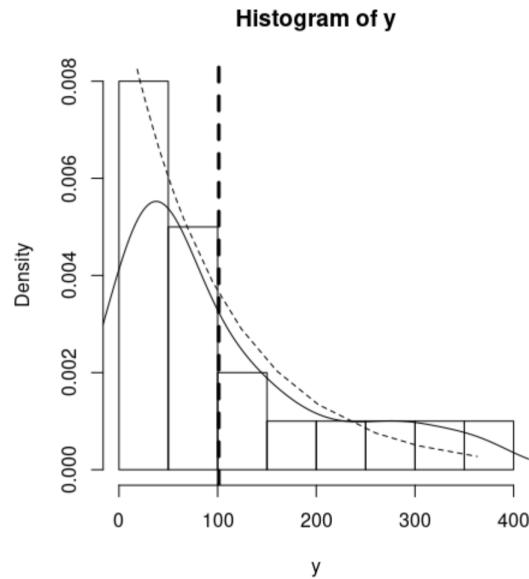
(c) The sample median of the data is **55.5**

```
> median(y)
[1] 55.5
```

The median of $Y \sim \Gamma(1, \theta) \equiv \text{Exp}(\theta)$ is the value of y' at which

$$0.5 = \int_0^{y'} \frac{1}{\theta} e^{-y'/\theta} dy = 1 - e^{-y'/\theta} \implies y' = \theta \ln 2, \quad (7.26)$$

which means that the median of $Y \sim \Gamma(1, \theta) \equiv \text{Exp}(1, \theta)$ is the half-life, $\theta \ln 2$. Since the sample median is just θ multiplied by $\ln 2$, the sample median also estimates the parameter θ .



- (d) From part (a), we know that $\hat{\theta} = \bar{Y}$, the sample mean, is the mle of θ , the population mean. From part (c), we have shown that the median of $Y \sim \Gamma(1, \theta)$ is simply $\theta \ln 2$. By simple inspection we see that $\hat{\theta} \ln 2 = \bar{Y} \ln 2$ is the (*unbiased*) mle of $\theta \ln 2$, the median of Y . \square

4.1.3 Suppose the number of customers X that enter a store between the hours 9:00 a.m. and 10:00 a.m. follows a Poisson distribution with parameter θ . Suppose a random sample of the number of customers that enter the store between 9:00 a.m. and 10:00 a.m. for 10 days results in the values

9 7 9 15 10 13 11 7 2 12

1. Determine the maximum likelihood estimate of θ . Show that it is an unbiased estimator.
2. Based on these data, obtain the realization of your estimator in part (a). Explain the meaning of this estimate in terms of the number of customers.

Solution:

1. Let $X \sim \text{Poi}(\theta)$ be given, then the pmf of X is given by

$$p(x) = \begin{cases} \frac{\theta^x e^{-\theta}}{x!}, & x \in \mathbb{N} \\ 0, & \text{else} \end{cases}. \quad (7.27)$$

Assuming the X_i 's $\sim \text{Poi}(\theta)$ are iid, where $i = 1, \dots, n$, then the logarithm of the likelihood function is

$$\begin{aligned} l(\theta) &= \log \left(\prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} \right) \\ &= \log \left(e^{-n\theta} \theta^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!} \right) \\ &= -n\theta + \left(\sum_{i=1}^n x_i \right) \log \theta - \sum_{i=1}^n \log x_i!. \end{aligned} \quad (7.28)$$

Setting $\partial_\theta l(\theta) = 0$, we solve for θ :

$$\partial_\theta l(\theta) = -n + \frac{1}{\theta} \sum_{i=1}^n x_i = 0 \iff \theta = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad (7.29)$$

By inspection, $\partial_{\theta\theta} l(\theta) < 0 \forall \theta \in \mathbb{R}^+$, and so the statistic

$$\hat{\theta} = \bar{Y} \quad (7.30)$$

is the mle of θ . Further, it is an unbiased estimator of θ simply because

$$E[Y] = \theta \implies E[\bar{Y}] = \theta. \quad (7.31)$$

□

2. Part (a) says the sample means is the mle of θ . The means of the given sample is **9.5**.

```
> mean(c(9 ,7, 9, 15, 10, 13, 11, 7, 2, 12))  
[1] 9.5
```

This says that on average, 9.5, or about 9-10 customers enter the store between the hours 9:00 a.m. and 10:00 a.m.. □

4.1.8 Recall that for the parameter $\eta = g(\theta)$, the mle of η is $g(\hat{\theta})$, where $\hat{\theta}$ is the mle of θ . Assuming that the data in Example 4.1.6 were drawn from a Poisson distribution with mean λ , obtain the mle of λ and then use it to obtain the mle of the pmf. Compare the mle of the pmf to the nonparametric estimate. Note: For the domain value 6, obtain the mle of $P(X \geq 6)$.

Solution: Based on the previous problem, the mle of λ is the sample means, which has the value **2.13**.

```
> mean(c(2,1,1,1,1,5,1,1,3,0,2,1,1,3,4,2,1,2,2,6,5,2,3,2,4,1,3,1,3,0))
[1] 2.133333
```

Because the sample means \bar{x} is the mle of λ , and the pmf is given by

$$p(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x \in \mathbb{N} \\ 0, & \text{else} \end{cases}, \quad (7.32)$$

the mle of the pmf is given by

$$\tilde{p}(x) = \begin{cases} \frac{\bar{x}^x e^{-\bar{x}}}{x!}, & x \in \mathbb{N} \\ 0, & \text{else} \end{cases}. \quad (7.33)$$

Next, we compare the mle of the pmf to the nonparametric estimate:

j	0	1	2	3	4	5	≥ 6
$\hat{p}(j)$	0.067	0.367	0.233	0.167	0.067	0.067	0.033
$\tilde{p}(j)$	0.118	0.253	0.270	0.192	0.102	0.044	0.022

Mathematica code for $P(j \geq 6)$ for $\tilde{p}(j)$:

```
P[x_] := (2.1333333)^x*E^(-2.1333333)/x!
N[Sum[P[y], {y, 6, Infinity}]]
0.0218705
```

7.2 Problem Set 2

4.2.2. Consider the data on the lifetimes of motors given in Exercise 4.1.1. Obtain a large sample 95% confidence interval for the mean lifetime of a motor.

Solution: Large sample 95% CI's have the form

$$(\bar{x} - z_{\alpha/2} \frac{S}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{S}{\sqrt{n}}) \quad (7.34)$$

Here, $\bar{x} = 101.15$, $n = 20$, $s = 105.4091$, $z_{\alpha/2} = 1.96$. So, the desired CI is

$$(101.15 - 1.96 \frac{105.4091}{\sqrt{20}}, 101.15 + 1.96 \frac{105.4091}{\sqrt{20}}) = \boxed{(54.95, 147.35)} \quad (7.35)$$

□

4.2.6. \bar{X} is the sample mean of a sample of size n from $\mathcal{N}(\mu, 9)$. Find n such that

$$P(\bar{X} - 1 \leq \mu \leq \bar{X} + 1) = 0.90 \quad (7.36)$$

Solution: $\sigma^2 = 9 \implies \sigma = 3$. We have

$$\begin{aligned} 0.90 &= P(\bar{X} - 1 \leq \mu \leq \bar{X} + 1) \\ &= P(\mu - 1 \leq \bar{X} \leq \mu + 1) \\ &= P(-1 \leq \bar{X} - \mu \leq 1) \\ &= P\left(\frac{-1}{3/\sqrt{n}} \leq \frac{\bar{X} - \mu}{3/\sqrt{n}} \leq \frac{1}{3/\sqrt{n}}\right). \end{aligned} \quad (7.37)$$

In other words,

$$z_{0.05} = \frac{1}{3/\sqrt{n}} = \frac{\sqrt{n}}{3} = 1.644854 \implies n = 24.35 \approx \boxed{25}. \quad (7.38)$$

□

4.2.18. X_i 's $\sim \mathcal{N}(\mu, \sigma^2)$, with μ, σ^2 unknown. A confidence interval for σ^2 can be found as follows. We know that $(n - 1)S^2/\sigma^2$ is a random variable with a $\chi^2(n-1)$ distribution. Thus we can find constants a and b so that $P((n-1)S^2/\sigma^2 < b) = 0.975$ and $P(a < (n-1)S^2/\sigma^2 < b) = 0.95$. In R, $b = qchisq(0.975, n-1)$, while $a = qchisq(0.025, n-1)$.

- (a) Show that this second probability statement can be written as

$$P((n-1)S^2/b < \sigma^2 < (n-1)S^2/a) = 0.95. \quad (7.39)$$

- (b) If $n = 9$ and $S^2 = 7.93$, find a 95% confidence interval for σ^2 .

- (c) If μ is known, how would you modify the preceding procedure for finding a confidence interval for σ^2 ?

Solution:

- (a) We simply re-arrange things in the probability statement:

$$\begin{aligned} 0.95 &= P(a < (n-1)S^2/\sigma^2 < b) \\ &= P(\sigma^2 < (n-1)S^2/a \wedge \sigma^2 > (n-1)S^2/b) \\ &= P((n-1)S^2/b < \sigma^2 < (n-1)S^2/a). \end{aligned} \quad (7.40)$$

- (b) When $n = 9, S^2 = 7.93$, we have $a = 2.179731$ and $b = 17.53455$. Then the 95% CI for σ^2 is

$$\left(\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a} \right) = \left(\frac{8 \times 7.93}{17.53455}, \frac{8 \times 7.93}{2.179731} \right) = \boxed{(3.618, 29.10451)} \quad (7.41)$$

- (c) If μ is known, the unbiased estimator for the population standard deviation becomes proportional to $1/\sqrt{n}$, not $1/\sqrt{n-1}$. Because of this, we modify some numerics in our procedure from $n-1$ to n . From here, we make the following changes

$$\begin{aligned} (n-1)S^2/\sigma^2 &\sim \chi^2(n-1) \rightarrow nS^2/\sigma^2 \sim \chi^2(n) \\ P(nS^2/\sigma^2 < b) &= 0.975 \\ P(a < nS^2/\sigma^2 < b) &= 0.95. \end{aligned} \quad (7.42)$$

The new CI will look like $(nS^2/b < \sigma^2 < nS^2/a)$.

□

4.2.21. Let two independent random samples, each of size 10, from two normal distributions $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ yield $\bar{x} = 4.8, s_1^2 = 8.64, \bar{y} = 5.6, s_2^2 = 7.88$. Find a 95% confidence interval for $\mu_1 - \mu_2$.

Solution: The 95% CI for difference in means in this case looks like

$$\left((\bar{x} - \bar{y}) - t_{\alpha/2, n_1 + n_2 - 2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{x} - \bar{y}) + t_{\alpha/2, n_1 + n_2 - 2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right). \quad (7.43)$$

The pooled variance is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{9 \times 8.64 + 9 \times 7.88}{18} = 8.26. \quad (7.44)$$

Plugging in numbers we find the CI, with $t_{0.025, 18} = 2.100922$:

$$\begin{aligned} & \left((4.8 - 5.6) - 2.100922 \sqrt{8.26} \sqrt{\frac{1}{10} + \frac{1}{10}}, (4.8 - 5.6) + 2.100922 \sqrt{8.26} \sqrt{\frac{1}{10} + \frac{1}{10}} \right) \\ &= \boxed{(-3.500, 1.900)} \end{aligned} \quad (7.45)$$

□

4.2.22. Let two independent random variables, Y_1 and Y_2 , with binomial distributions that have parameters $n_1 = n_2 = 100$, p_1 , and p_2 , respectively, be observed to be equal to $y_1 = 50$ and $y_2 = 40$. Determine an approximate 90% confidence interval for $p_1 - p_2$.

Solution: The 90% CI for the difference in proportions looks like

$$\begin{aligned} & \left((\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}, \right. \\ & \quad \left. (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \right) \end{aligned} \quad (7.46)$$

where $\alpha_{0.05} = 1.644854$. Plugging in numbers, we find

$$\begin{aligned} & \left(0.5 - 0.4 - 1.644854 \sqrt{\frac{(0.5)(0.5)}{100} + \frac{(0.4)(0.6)}{100}}, \right. \\ & \quad \left. 0.5 - 0.4 + 1.644854 \sqrt{\frac{(0.5)(0.5)}{100} + \frac{(0.4)(0.6)}{100}} \right) \\ & = \boxed{(-0.01513978, 0.2151398)} \end{aligned} \quad (7.47)$$

□

4.2.27. Let X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m be two independent random samples from the respective normal distributions $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, where the four parameters are unknown. To construct a confidence interval for the ratio, σ_1^2/σ_2^2 , of the variances, form the quotient of the two independent 2 variables, each divided by its degrees of freedom, namely,

$$F = \frac{\frac{(m-1)S_2^2}{\sigma_2^2}/(m-1)}{\frac{(n-1)S_1^2}{\sigma_1^2}/(n-1)} = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \quad (7.48)$$

where S_1^2, S_2^2 are respectively sample variances.

- (a) What kind of distribution does F have?
- (b) Rewrite the second probability statement as

$$P \left[a \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < b \frac{S_1^2}{S_2^2} \right] = 0.95. \quad (7.49)$$

The observed values, s_1^2 and s_2^2 , can be inserted in these inequalities to provide a 95% confidence interval for σ_1^2/σ_2^2 .

Solution:

- (a) $F \sim F(m-1, n-1)$, by definition.
- (b) We just rearrange the quantities in the probability statement:

$$\begin{aligned} 0.95 &= P(a < F < b) \\ &= P \left(a < \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} < b \right) \\ &= P \left(\frac{\sigma_1^2}{\sigma_2^2} < b \frac{S_1^2}{S_2^2} \wedge \frac{\sigma_1^2}{\sigma_2^2} > a \frac{S_1^2}{S_2^2} \right) \\ &= P \left(a \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < b \frac{S_1^2}{S_2^2} \right). \end{aligned} \quad (7.50)$$

□

4.5.1. Show that the approximate power function given in expression (4.5.12) of Example 4.5.3 is a strictly increasing function of μ . Show then that the test discussed in this example has approximate size α for testing

$$H_0 : \mu \leq \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0. \quad (7.51)$$

Solution: The approximate power function is given by

$$\gamma(\mu) \approx \Phi\left(-z_\alpha - \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}\right) \quad (7.52)$$

$\partial_\mu \gamma(\mu)$ is necessarily positive $\forall \mu \in \mathbb{R}$ for $\gamma(\mu)$ to be strictly increasing. So we check:

$$\begin{aligned} \partial_\mu \gamma(\mu) &= \partial_\mu \left[\Phi\left(-z_\alpha - \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}\right) \right] \\ &= \Phi'\left(-z_\alpha - \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma}\right) \frac{\sqrt{n}}{\sigma}. \end{aligned} \quad (7.53)$$

We note that $\sqrt{n}/\sigma > 0$ and $\Phi'(\dots) > 0$ necessarily because Φ is a cdf (for the $\mathcal{N}(\mu, \sigma^2)$). With this, we have shown that $\gamma(\mu)$ is strictly increasing in μ .

Under the hypotheses and the fact that $\gamma(\mu)$ is strictly increasing in μ , $\alpha = \max_{\mu \leq \mu_0} \gamma(\mu)$ is maximized whenever $\mu \leq \mu_0$ is maximized, i.e. $\mu = \mu_0$:

$$\max_{\mu \leq \mu_0} \gamma(\mu) = \gamma(\mu_0) = \Phi(-z_\alpha) = \alpha. \quad (7.54)$$

So the test has approximate size α . □

4.5.2. For the Darwin data tabled in Example 4.5.5, verify that the Student t-test statistic is 2.15.

Solution: $\alpha = 0.05$. The sample mean and standard deviation for the differences are

$$\bar{x} = 2.62 \quad (7.55)$$

$$s_x = 4.71826. \quad (7.56)$$

The t-statistic is then

$$t_{df=14} = \frac{\bar{x} - 0}{s_x} = \frac{2.62}{4.71826/\sqrt{15}} \approx 2.150627 \quad (7.57)$$

R code:

```
> mean(darwin$cross)-mean(darwin$self)
[1] 2.62

> sd(darwin$cross - darwin$self)
[1] 4.71826
```

□

4.5.5. Let X_1, X_2 be a random sample of size $n = 2$ from the distribution having pdf $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $\theta < x < \infty$, zero elsewhere. We reject $H_0 : \theta = 2$ and accept $H_1 : \theta = 1$ if the observed values of X_1, X_2 , say x_1, x_2 , are such that

$$\frac{f(x_1; 2)f(x_2; 2)}{f(x_1; 1)f(x_2; 1)} \leq \frac{1}{2} \quad (7.58)$$

Here $\Omega = \{\theta : \theta = 1, 2\}$. Find the significance level of the test and the power of the test when H_0 is false.

Solution: We reject H_0 whenever

$$\begin{aligned} \frac{1}{2} &\geq \frac{f(x_1; 2)f(x_2; 2)}{f(x_1; 1)f(x_2; 1)} \\ &= \frac{(1/2)e^{-x_1/2}(1/2)e^{-x_2/2}}{e^{-x_1}e^{-x_2}} \\ &= \frac{1}{4}e^{x_1/2}e^{x_2/2} \\ &= \frac{1}{4}e^{(x_1+x_2)/2} \implies x_1 + x_2 \leq 2\ln(2). \end{aligned} \quad (7.59)$$

The significance level of the test α is the probability of rejecting H_0 when it is true, i.e.

$$\alpha = P(x_1 + x_2 \leq 2\ln(2) | \theta = 2). \quad (7.60)$$

Recall that the non-zero part of the pdf for $\Gamma(k, \theta)$ is given by

$$\begin{aligned} f(x) &= \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \quad x \in \mathbb{R}^+ \\ &= \frac{1}{\theta^1} e^{-x/\theta}, \quad k = 1, \end{aligned} \quad (7.61)$$

we have that $X_1, X_2 \sim \Gamma(k = 1, \theta = 2)$, iid, implies $X_1 + X_2 \sim \Gamma(k = 2, \theta = 2)$. From here, it is “easy” to calculate α :

$$\begin{aligned} \alpha &= P(x_1 + x_2 \leq 2\ln(2) | \theta = 2) = \int_0^{2\ln(2)} \frac{1}{\Gamma(2)\theta^2} \xi e^{-\xi/\theta} d\xi \\ &= \int_0^{2\ln(2)} \frac{1}{4} \xi e^{-\xi/2} d\xi \\ &= \frac{1}{2}(1 - \ln(2)) \approx \boxed{0.1534} \end{aligned} \quad (7.62)$$

The power of the test is the probability of rejecting H_0 when H_0 is false. In this case, we make a similar calculation, only setting $\theta = 1$ (since H_0 false):

$$\begin{aligned} P(x_1 + x_2 \leq 2\ln(2) | \theta = 1) &= \int_0^{2\ln(2)} \xi e^{-\xi} d\xi \\ &= \frac{3}{4} - \frac{\ln(2)}{2} \approx \boxed{0.403426} \end{aligned} \quad (7.63)$$

□

4.5.12. Let X_1, X_2, \dots, X_8 be a random sample of size $n = 8$ from a Poisson distribution with mean μ . Reject the simple null hypothesis $H_0 : \mu = 0.5$ and accept $H_1 : \mu > 0.5$ if the observed $\sum_{i=1}^8 x_i \geq 8$.

- (a) Show that the significance level is $1 - \text{ppois}(7, 8 * .5)$.
- (b) Use R to determine $\gamma(0.75)$, $\gamma(1)$, and $\gamma(1.25)$.
- (c) Modify the code in Exercise 4.5.9 to obtain a plot of the power function.

Solution:

- (a) The significance level α is the probability of rejecting H_0 when H_0 is true. Under the null, $\mu = 0.5$, and the r.v.

$$X_1 + \cdots + X_8 \sim \text{Poi}(8\mu) \equiv \text{Poi}(8 \times 0.5). \quad (7.64)$$

α is given by

$$\begin{aligned} \alpha = \gamma(\mu) &= P\left(\sum_{i=1}^8 x_i \geq 8 | \mu = 0.5\right) \\ &= 1 - P\left(\sum_{i=1}^8 x_i < 7\right) \\ &= \boxed{1 - \text{ppois}(7, 8 * .5)} \\ &= 0.05113362 \end{aligned} \quad (7.65)$$

(b)

$$\begin{aligned} \gamma(0.75) &= \boxed{0.2560202} \\ \gamma(1) &= \boxed{0.5470392} \\ \gamma(1.25) &= \boxed{0.7797794} \end{aligned} \quad (7.66)$$

R code:

```
> 1 - ppois(7, 8 * 0.5)
[1] 0.05113362

> 1 - ppois(7, 8 * 0.75)
[1] 0.2560202

> 1 - ppois(7, 8 * 1)
[1] 0.5470392

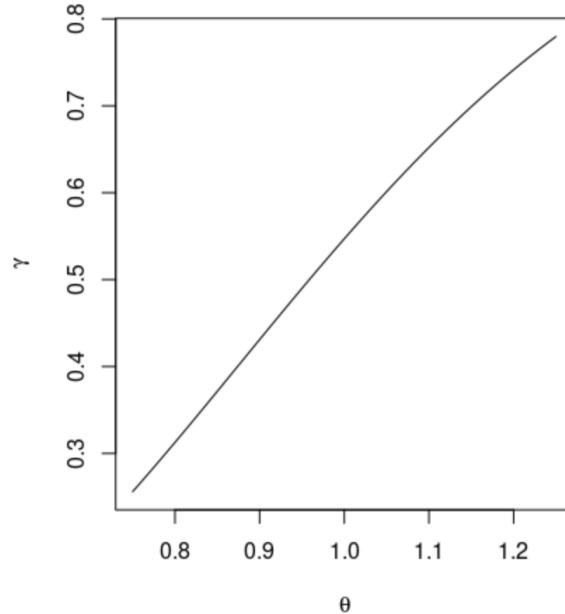
> 1 - ppois(7, 8 * 1.25)
[1] 0.7797794
```

(c) **R code:**

```
> theta=seq(.75,1.25,.25); gam=1-ppois(7,theta*8)
> plot(gam~theta,pch=" ",xlab=expression(theta),ylab=expression(gamma))
> lines(gam~theta)
```

We're interested in the range of $\theta \in [0.75, 1.25]$. I'm making the step size small to make the power function look smooth.

Plot of $\gamma(\mu)$:



□

4.6.4. (Note that it is fine to make a heuristic argument here. Just make sure that it is clear with supporting graphs/figures (hand drawn is fine).) Consider the one-sided t-test for $H_0 : \mu = \mu_0$ versus $H_{A1} : \mu > \mu_0$ constructed in Example 4.5.4 and the two-sided t-test for t-test for $H_0 : \mu = \mu_0$ versus $H_1 : \mu = \mu_0$ given in (4.6.9). Assume that both tests are of size α . Show that for $\mu > \mu_0$, the power function of the one-sided test is larger than the power function of the two-sided test.

Solution: We want to show that for $\mu > \mu_0$, the power function of the one-sided test is larger than the power function of the two-sided test. To this end, let $\gamma_1(\mu)$ denote the power function of the one-sided test, and $\gamma_2(\mu)$ the two-sided test. This gives

$$\gamma_2(\mu) = P(|\text{test-statistic}| \geq t_{\alpha/2,n-1}) = P(\text{test-statistic} \geq t_{\alpha/2,n-1}) \quad (7.67)$$

(test statistic positive because $\mu > \mu_0$), while

$$\gamma_1(\mu) = P(\text{test-statistic} \geq t_{\alpha,n-1}). \quad (7.68)$$

Since $t_{\alpha/2,n-1} > t_{\alpha,n-1}$, we have that

$$\gamma_2(\mu) = P(|\cdot| \geq t_{\alpha/2}) \leq P(\cdot \geq t_{\alpha}) = \gamma_1(\mu). \quad (7.69)$$

And so, for a given $\mu > \mu_0$, the power function of the one-sided test is larger than the power function of the two-sided test. \square

4.6.7. Among the data collected for the World Health Organization air quality monitoring project is a measure of suspended particles in $\mu\text{g}/\text{m}^3$. Let X and Y equal the concentration of suspended particles in $\mu\text{g}/\text{m}^3$ in the city center (commercial district) for Melbourne and Houston, respectively. Using $n = 13$ observations of X and $m = 16$ observations of Y , we test $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$.

- (a) Define the test statistic and critical region, assuming that the unknown variances are equal. Let $\alpha = 0.05$.
- (b) If $\bar{x} = 72.9$, $s_x = 25.6$, $\bar{y} = 81.7$, and $s_y = 28.3$, calculate the value of the test statistic and state your conclusion.

Solution:

- (a) Assuming the unknown variances are equal, we have

$$\tau = \frac{(\bar{y} - \bar{x})}{s_p \sqrt{\frac{1}{13} + \frac{1}{16}}} \quad (7.70)$$

The critical region is given by

$$C := \{(X_1, \dots, X_{13}, Y_1, \dots, Y_{16}) | \tau \geq t_{0.05, 13+16-2} = [1.703288]\}. \quad (7.71)$$

- (b) With the given numbers, we calculate the pooled variance is

$$S_p^2 = \frac{(13-1)(25.6)^2 + (16-1)(28.3)^2}{13+16-2} = 736.21 \quad (7.72)$$

With this,

$$\tau = \frac{(81.7 - 72.9) - 0}{\sqrt{736.21} \sqrt{\frac{1}{13} + \frac{1}{16}}} = [0.8685893] \quad (7.73)$$

Since $0.8685893 < 1.703288$, there is not enough evidence to reject H_0 in favor of H_a .

□

4.6.8. Let p equal the proportion of drivers who use a seat belt in a country that does not have a mandatory seat belt law. It was claimed that $p = 0.14$. An advertising campaign was conducted to increase this proportion. Two months after the campaign, $y = 104$ out of a random sample of $n = 590$ drivers were wearing their seat belts. Was the campaign successful?

1. Define the null and alternative hypotheses.
2. Define a critical region with an $\alpha = 0.01$ significance level.
3. Determine the approximate p -value and state your conclusion.

Solution:

(a) $H_0 : p = 0.14 \quad H_a : p > 0.14$.

(b) The critical region is given by

$$C := \left\{ y \left| \frac{y/n - 0.14}{\sqrt{\frac{0.14(1-0.14)}{590}}} \geq z_{\alpha=0.01} = 2.326348 \right. \right\}. \quad (7.74)$$

(c) The value of the test statistic is

$$z^* = \frac{104/590 - 0.14}{\sqrt{\frac{0.14(1-0.14)}{590}}} = \boxed{2.539069 > 2.326348} \quad (7.75)$$

Since $z^* > z$, there is enough evidence to reject H_0 in favor of H_a (p -value: $\boxed{0.006 < 0.01 = \alpha}$), i.e., there is enough evidence to suggest that the campaign was successful.

R code:

```
> 1-pnorm(2.539069)
[1] 0.005557395
```

□

7.3 Problem set 3

4.7.4

A die was cast $n = 120$ independent times and the following data resulted:

Spots Up	1	2	3	4	5	6
Frequency	b	20	20	20	20	$40 - b$

If we use a chi-square test, for what values of b would the hypothesis that the die is unbiased be rejected at the 0.025 significance level?

Solution: The test statistic under the null hypothesis $H_0 : \partial_i = 1/6 \forall i$ is

$$\chi = \sum_{i=1}^6 \frac{(\text{Freq}_i - 20)^2}{20} = \frac{(b - 20)^2}{20} + 4 \cdot 0 + \frac{(40 - b - 20)^2}{20} = \frac{(b - 20)^2}{10}. \quad (7.76)$$

Under the null hypothesis, $\chi \sim \chi^2(df = 5)$. At $\alpha = 0.025$, we reject the null hypothesis whenever $\chi \geq \text{qchisq}(1-0.025, 5) = 12.8325$, i.e.,

$$\frac{(b - 20)^2}{10} \geq 12.8325 \implies [b \leq 8 \vee b \geq 32] \quad (7.77)$$

□

4.7.5

Consider the problem from genetics of crossing two types of peas. The Mendelian theory states that the probabilities of the classifications (a) round and yellow, (b) wrinkled and yellow, (c) round and green, and (d) wrinkled and green are $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, and $\frac{1}{16}$, respectively. If, from 160 independent observations, the observed frequencies of these respective classifications are 86, 35, 26, and 13, are these data consistent with the Mendelian theory? That is, test, with $\alpha = 0.01$, the hypothesis that the respective probabilities are $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, and $\frac{1}{16}$.

Solution: The test statistic under $H_0 : p_a = 9/16, p_b = 3/16, p_c = 3/16, p_d = 1/16$, the test statistic is

$$\chi^2 = \frac{(86 - 90)^2}{90} + \frac{(35 - 30)^2}{30} + \frac{(26 - 30)^2}{30} + \frac{(13 - 10)^2}{10} = \frac{22}{9}. \quad (7.78)$$

Under H_0 , $\chi^2 \sim \chi^2(df = 4-1 = 3)$. The p-value for this χ^2 is $1 - \text{pchisq}(22/9, 3) = 0.4854149 > 0.01$. So, we don't have enough evidence to reject H_0 , i.e. the data is consistent with the Mendelian theory. \square

4.7.6

Two different teaching procedures were used on two different groups of students. Each group contained 100 students of about the same ability. At the end of the term, an evaluating team assigned a letter grade to each student. The results were tabulated as follows.

Group	Grade					Total
	A	B	C	D	F	
I	15	25	32	17	11	100
II	9	18	29	28	16	100

If we consider these data to be independent observations from two respective multinomial distributions with $k = 5$, test at the 5% significance level the hypothesis that the two distributions are the same (and hence the two teaching procedures are equally effective). For computation in R, use

```
r1=c(15,25,32,17,11);r2=c(9,18,29,28,16);mat=rbind(r1,r2)
chisq.test(mat)
```

Solution: Using the code provided by the problem we get

```
> r1=c(15,25,32,17,11);r2=c(9,18,29,28,16);mat=rbind(r1,r2)
> chisq.test(mat)

Pearson's Chi-squared test

data: mat
X-squared = 6.4019, df = 4, p-value = 0.1711
```

Since $p = 0.1711 > 0.05$, we don't have enough evidence to reject H_0 , i.e. the two teaching procedures are (statistically) equally effective. \square

4.8.9 (use acceptance sampling)

Determine a method to generate random observations for the Cauchy distribution with pdf

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty. \quad (4.8.16)$$

Write an R function that returns a random sample of observations from this Cauchy distribution.

Solution: R code:

```
# acceptance sampling
# approximate the Cauchy distribution

x <- runif(100000,-7,7)
y <- runif(100000,0,1)

z <- cbind(x,y)
accept <- NULL
reject <- NULL

dens <- function(x){
  d <- 1/((pi^2)*(x^2+1))
  return(d)}

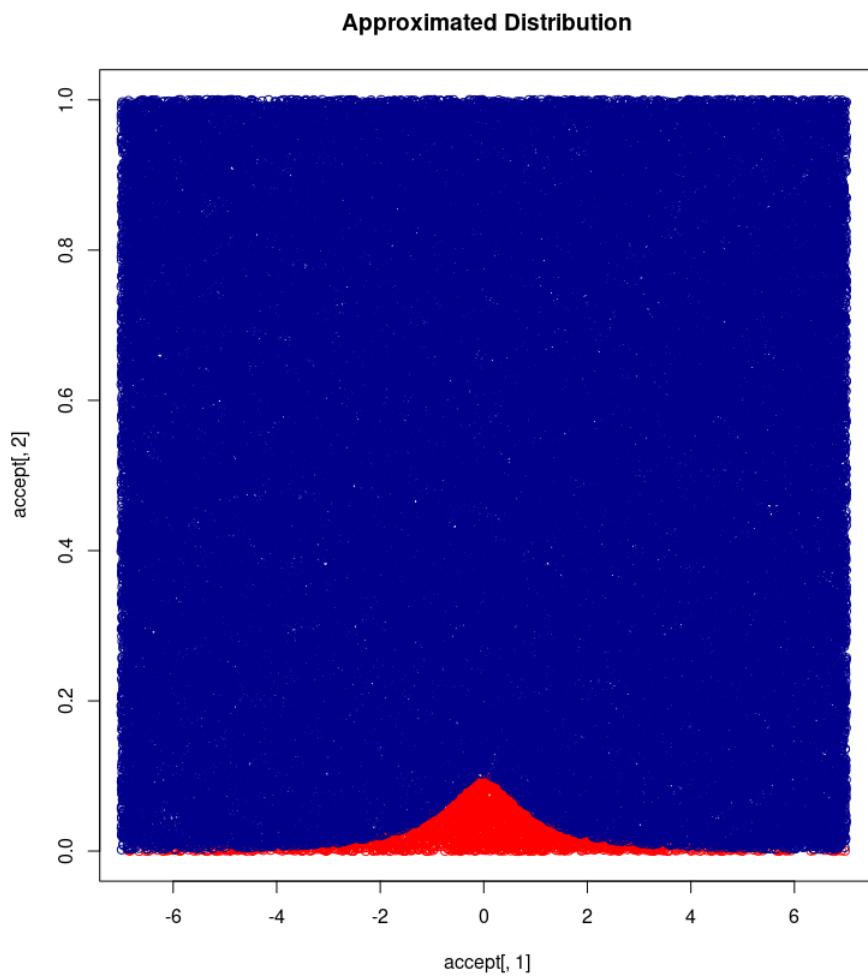
for (i in 1:length(x)){
  d <- dens(x[i])
  if (y[i] < d) {
    accept <- rbind(accept,z[i,])}
  else{
    reject <- rbind(reject,z[i,])}
}

# plot accepted values in red
plot(accept[,1],accept[,2],
main = "Approximated Distribution", ylim=c(0,1), col="red")

# plot rejected values in blue
points(reject[,1], reject[,2], col = "dark blue")
```

Here's a small sample:

```
> accept
x           y
[1,] -0.2257702402 8.676955e-02
[2,] -0.3536258731 6.871887e-02
[3,]  4.1498403600 4.068073e-03
[4,] -0.3986396971 3.033239e-02
[5,]  0.7077048477 1.492551e-02
[6,] -0.0422784868 1.612575e-02
[7,] -0.6391186416 1.029603e-02
[8,]  0.4845524384 6.119987e-02
[9,]  0.3572320105 2.479371e-02
.
.
```



□

4.8.10 (use inverse transformation sampling)

Suppose we are interested in a particular Weibull distribution with pdf

$$f(x) = \begin{cases} \frac{1}{\theta^3} 3x^2 e^{-x^3/\theta^3} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Determine a method to generate random observations from this Weibull distribution. Write an R function that returns such a sample.

Hint: Find $F^{-1}(u)$.

Solution: We first want to find the cdf F , given $f(x)$. Well,

$$F(x) = \int_{-\infty}^x f(x') dx' = \int_0^x \frac{1}{\theta^3} 3x'^2 e^{-x'^3/\theta^3} dx' = \dots = 1 - e^{-x^3/\theta^3}. \quad (7.79)$$

With this, let $u \sim U(0, 1)$ then

$$F^{-1}(u) = \sqrt[3]{-\theta^3 \ln(1-u)} \sim \text{Wei}(k=3, \theta), \quad 0 \leq u \leq 1. \quad (7.80)$$

Suppose $\theta = 2$ then in R, we do the following:

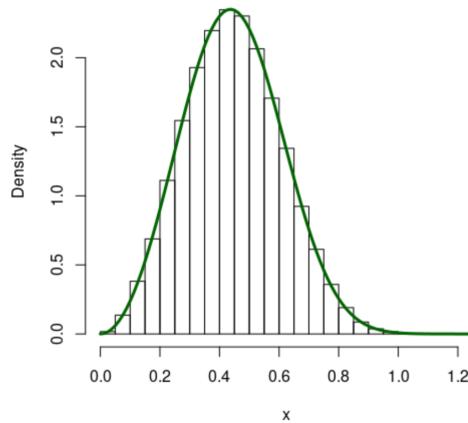
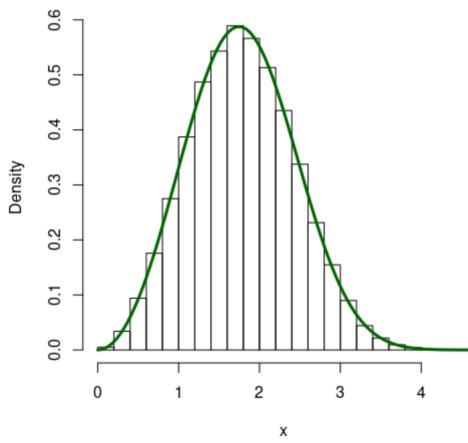
```
# inverse transformation sampling
# generate vector of random uniform(0,1)
u <- runif(100000)

# set beta and transform to random weibull(shape=3, scale=theta)
beta = 2
x <- (-beta^3)*log(1 - u)^(1/3)
hist(x, main="Histogram of Tranformed Variable", pr = TRUE)

# overlay weibull
curve(dweibull(x, shape = 3, scale=2),
col="darkgreen", lwd=3, add=TRUE)
```

Here's a sample:

```
> x <- (-beta^3)*log(1 - u)^(1/3)
> x
1.57511912 1.81610458 2.51359811 1.12564398 1.21991006
1.10365173 0.71966428 2.06591082 3.08423326 1.67620145
1.02533405 1.24429387 0.94192313 2.19097666 1.98069305
0.43007645 1.18581709 1.11303036 1.63002049 2.37205273
2.45962833 1.67162606 1.61624974 1.46149955 1.92907758
1.12735030 1.02997348 2.25870529 1.58137215 2.71199084
1.05306668 1.43121600 1.17977939 1.24985669 2.22292333
1.86914252 3.01206737 1.11948037 1.64757666 1.35904188
1.39351092 2.60002179 1.12125772 2.15576011 2.41963279
```

Histogram of Tranformed Variable, beta=0.5**Histogram of Tranformed Variable, beta=2**

□

7.4 Problem set 4

4.9.4 (for part B, write your own R function to generate the bootstrap distribution of the median - using the code posted on Moodle is an okay starting point)

- Consider the situation discussed in Example 4.9.1. Suppose we want to estimate the median of X_i using the sample median.

- (a) Determine the median for a $\Gamma(1, \beta)$ distribution.
- (b) The algorithm for the bootstrap percentile confidence intervals is general and hence can be used for the median. Rewrite the R code in the function `percentciboot.s` so that the median is the estimator. Using the sample given in the example, obtain a 90% bootstrap percentile confidence interval for the median. Did it trap the true median in this case?

Solution:

- (a) The median $x_{1/2}$ for a $\Gamma(1, \beta)$ is such that

$$\frac{1}{2} = \int_0^{x_{1/2}} \frac{1}{\beta} e^{-x/\beta} = 1 - e^{-x_{1/2}/\beta} \implies \boxed{x_{1/2} = \beta \ln(2)} \quad (7.81)$$

- (b) Here's the R code (I'm not using Prof. O'Brien's code here)

```
> x <- c(131.7, 182.7, 73.3, 10.7, 150.4, 42.3, 22.2, 17.9, 264.0,
154.4, 4.3, 265.6, 61.9, 10.8, 48.8, 22.5, 8.8, 150.6, 103.0, 85.9)
> percentciboot <- function(x,b,alpha){
+   theta=median(x); thetastar=rep(0,b); n=length(x)
+   for(i in 1:b){xstar=sample(x,n,replace=T)
+     thetastar[i]=median(xstar)}
+   thetastar=sort(thetastar); pick=round((alpha/2)*(b+1))
+   lower=thetastar[pick]; upper=thetastar[b-pick+1]
+   list(theta=theta,lower=lower,upper=upper)}
> percentciboot(x,3000,.10)

$theta
[1] 67.6

$lower
[1] 30.1

$upper
[1] 131.7

> median(x)
[1] 67.6

> 100*log(2)
[1] 69.31472
```

The 90% bootstrap percentile CI for the median is given by $(30.1, 131.7)$. The true median is given by $\beta \ln(2) = 100 \ln(2) \approx 69.31472$. So, yes, the 90% CI traps the true median in this case.

□

4.9.11

Let z^* be drawn at random from the discrete distribution that has mass n^{-1} at each point $z_i = x_i - \bar{x} + \mu_0$, where (x_1, x_2, \dots, x_n) is the realization of a random sample. Determine $E(z^*)$ and $V(z^*)$.

Solution:

(a) $E(z^*)$:

$$E(z^*) = \sum_{i=1}^n \frac{x_i - \bar{x} + \mu_0}{n} = \bar{x} - \bar{x} + \mu_0 = \boxed{\mu_0} \quad (7.82)$$

(b) $V(z^*)$:

$$V(z^*) = \sum_{i=1}^n (z_i - E(z^*))^2 = \sum_{i=1}^n (x_i - \bar{x} + \mu_0 - \mu_0)^2 = \boxed{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.83)$$

□

4.9.13

For the situation described in Example 4.9.3, obtain the bootstrap test based on medians. Use the same hypotheses; i.e.,

$$H_0 : \mu = 90 \text{ versus } H_1 : \mu > 90.$$

Solution: Here's the R code:

```
> X <- c(119.7, 104.1, 92.8, 85.4, 108.6, 93.4, 67.1, 88.4, 101.0, 97.2,
+      95.4, 77.2, 100.0, 114.2, 150.3, 102.3, 105.8, 107.5, 0.9, 94.1)
> boottestonemed <-
+   function(x,theta0,b){
+     #
+     # x = sample
+     # theta0 is the null value of the mean
+     # b is the number of bootstrap resamples
+     #
+     # origtest contains the value of the test statistics
+     # for the original sample
+     # pvalue is the bootstrap p-value
+     # teststatall contains the b bootstrap tests
+     #
+     n<-length(x)
+     v<-median(x)
+     z<-x-median(x)+theta0
+     counter<-0
+     teststatall<-rep(0,b)
+     for(i in 1:b){xstar<-sample(z,n,replace=T)
+     vstar<-median(xstar)
+     if(vstar >= v){counter<-counter+1}
+     teststatall[i]<-vstar}
+     pvalue<-counter/b
+     list(origtest=v,pvalue=pvalue,teststatall=teststatall)
+     #list(origtest=v,pvalue=pvalue)
+   }
> boottestonemed(X,90,3000)
$origtest
[1] 98.6

$pvalue
[1] 0.006
```

At such a low p-value ($p = 0.006$), we reject the null hypothesis H_0 . So even though we don't reject H_0 with the test based on sampling mean, we *do* reject H_0 with the test based on medians. \square

5.1.7

5.1.7. Let X_1, \dots, X_n be iid random variables with common pdf

$$f(x) = \begin{cases} e^{-(x-\theta)} & x > \theta, -\infty < \theta < \infty \\ 0 & \text{elsewhere.} \end{cases} \quad (5.1.3)$$

This pdf is called the **shifted exponential**. Let $Y_n = \min\{X_1, \dots, X_n\}$. Prove that $Y_n \rightarrow \theta$ in probability by first obtaining the cdf of Y_n .

Solution: By definition,

$$\begin{aligned} Y_n \xrightarrow{P} \theta &\iff \forall \epsilon > 0, \lim_{n \rightarrow \infty} P[|Y_n - \theta| \geq \epsilon] = 0 \\ &\iff \forall \epsilon > 0, \lim_{n \rightarrow \infty} P[|\min\{X_1, \dots, X_n\} - \theta| \geq \epsilon] = 0 \\ &\iff \forall \epsilon > 0, \lim_{n \rightarrow \infty} P[\min\{X_1, \dots, X_n\} - \theta \geq \epsilon] = 0 \\ &\iff \forall \epsilon > 0, \lim_{n \rightarrow \infty} P[Y_n \geq \epsilon + \theta] = 0 \end{aligned} \quad (7.84)$$

where the second to last equivalence statement comes from the fact that $x_i > \theta \forall i = 1, 2, \dots, n$. To show $Y_n \xrightarrow{P} \theta$, we find the cdf for Y_n :

$$\begin{aligned} F_{Y_n}(y) &= P(Y_n < y) \\ &= 1 - P(Y_n \geq y_m) \\ &= 1 - \prod_{i=1}^n P(X_i \geq y) \\ &= 1 - \prod_{i=1}^n \int_y^\infty e^{-(x-\theta)} dx \\ &= 1 - \prod_{i=1}^n e^{-(y-\theta)} \\ &= 1 - e^{-n(y-\theta)} \\ P(Y_n \geq y) &= e^{-n(y-\theta)}. \end{aligned} \quad (7.85)$$

Let $\epsilon > 0$ be given. Then

$$P(Y_n \geq \epsilon + \theta) = e^{-n(\epsilon+\theta-\theta)} = e^{-n\epsilon}, \quad (7.86)$$

and so

$$\lim_{n \rightarrow \infty} P[|Y_n - \theta| \geq \epsilon] = \lim_{n \rightarrow \infty} P[Y_n \geq \epsilon + \theta] = \lim_{n \rightarrow \infty} e^{-n\epsilon} = 0 \implies Y_n \xrightarrow{P} \theta. \quad (7.87)$$

□

5.2.2 (Investigate = find)

5.2.2. Let Y_1 denote the minimum of a random sample of size n from a distribution that has pdf $f(x) = e^{-(x-\theta)}$, $\theta < x < \infty$, zero elsewhere. Let $Z_n = n(Y_1 - \theta)$. Investigate the limiting distribution of Z_n .

Solution: Let F_{Z_n} and F_Z be, respectively, the cdfs of Z_n and Z . Then by definition

$$Z_n \xrightarrow{D} Z \iff \lim_{n \rightarrow \infty} F_{Z_n}(z) = F_Z(z) \quad (7.88)$$

for all z at which $F_Z(z)$ is continuous. Now, we don't know what F_Z is in this case, but we can find what F_{Z_n} converges to when $n \rightarrow \infty$. To show this, we find F_{Z_n} :

$$\begin{aligned} F_{Z_n}(z) &= P(Z_n \leq z) \\ &= P(n(Y_1 - \theta) \leq z) \\ &= P(Y_1 \leq z/n + \theta) \\ &= 1 - e^{-n(z/n + \theta - \theta)}, \quad \text{calculated in Problem 5.1.7} \\ &= 1 - e^{-z}. \end{aligned} \quad (7.89)$$

And so (obviously)

$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = 1 - e^{-z} \equiv \text{cdf}(\text{Exp}(1)) \quad (7.90)$$

Therefore, $Z_n \xrightarrow{D} Z \sim \text{Exp}(1)$. □

5.2.7

5.2.7. Let X_n have a gamma distribution with parameter $\alpha = n$ and β , where β is not a function of n . Let $Y_n = X_n/n$. Find the limiting distribution of Y_n .

Solution: Let $X_n \sim \Gamma(\alpha = n, \beta)$ be given, where β is not a function of n . Let $Y_n = X_n/n$. To find the limiting distribution of Y_n , we find $\lim_{n \rightarrow \infty} M_{Y_n}(t)$. The reason we don't want to find the cdf F_{Y_n} is that integrals involving the Gamma distribution are often ugly.

$$\begin{aligned} M_{Y_n}(t) &= E[e^{tY_n}] = E[e^{tX_n/n}] \equiv E[e^{t_n X_n}] \\ &= (1 - \beta t_n)^{-n} \\ &= \left(1 - \frac{\beta t}{n}\right)^{-n}. \end{aligned} \quad (7.91)$$

And so using the identity

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^{-n} = e, \quad (7.92)$$

by change of variables we obtain

$$\lim_{n \rightarrow \infty} M_{Y_n}(t) = \lim_{n \rightarrow \infty} \left(1 - \frac{\beta t}{n}\right)^{-n} = e^{\beta t} \quad (7.93)$$

So what is the limiting distribution of Y_n ? By definition,

$$M_Y(t) = E[e^{tY}] = \int_{-\infty}^{\infty} f_Y(y) e^{yt} dy = e^{\beta t}. \quad (7.94)$$

Upon inspection, this equality holds if and only if $f_Y(y) \equiv \delta(y - \beta)$, the delta function centered at β .

$$\int_{-\infty}^{\infty} \delta(y - \beta) e^{yt} dy = e^{\beta t}. \quad (7.95)$$

And so, the limiting distribution of Y_n is the degenerate distribution with parameter β :

$$Y_n \xrightarrow{D} Y \sim \delta(y - \beta) \equiv \begin{cases} 1, & y = \beta \\ 0, & \text{else} \end{cases}. \quad (7.96)$$

□

5.3.9

5.3.9. Let $f(x) = 1/x^2$, $1 < x < \infty$, zero elsewhere, be the pdf of a random variable X . Consider a random sample of size 72 from the distribution having this pdf. Compute approximately the probability that more than 50 of the observations of the random sample are less than 3.

Solution: We have X_1, \dots, X_{72} , with

$$X_i \sim f(x) = \begin{cases} \frac{1}{x^2}, & 1 < x < \infty \\ 0, & \text{else} \end{cases}. \quad (7.97)$$

We first find the probability that any given observation is less than 3:

$$P(X < 3) = \int_1^3 \frac{1}{x^2} dx = \frac{2}{3}. \quad (7.98)$$

So we have a “binomial situation” where the probability of success is $p = 2/3$. Given $n = 72$ trials, we have $\mu = np = 72(2/3) = 48$ and $\sigma = \sqrt{np(1-p)} = 4$. We wish to find the probability of having more than 50 successes. To this end, we use the normal approximation (CLT), which says

$$\frac{Y_{72} - \mu}{\sigma} \sim \mathcal{N}(0, 1). \quad (7.99)$$

And so

$$\begin{aligned} P(Y_{72} > 50) &\approx P\left(Z \geq \frac{51 - 48}{4}\right) \\ &= 1 - \text{pnorm}(3/4) \\ &= \boxed{0.2266274} \end{aligned} \quad (7.100)$$

Or we can also use the continuity correction to get

$$\begin{aligned} P(Y_{72} > 50) &\approx P\left(Z \geq \frac{50.5 - 48}{4}\right) \\ &= 1 - \text{pnorm}((50.5 - 48)/4) \\ &= \boxed{0.2659855} \end{aligned} \quad (7.101)$$

□

5.3.11

5.3.11. We know that \bar{X} is approximately $N(\mu, \sigma^2/n)$ for large n . Find the approximate distribution of $u(\bar{X}) = \bar{X}^3$, provided that $\mu \neq 0$.

Solution: We want to use the Δ -method for this problem. Since $\bar{X} \sim N(\mu, \sigma^2/n)$, we have

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2), \quad (7.102)$$

by a simple change of variables. The function $u(\bar{X}) = \bar{X}^3$ is differential at for all \bar{X} and $u'(\bar{X}) \neq 0$ in general, so by the Δ -method,

$$\begin{aligned} & \sqrt{n}(u(\bar{X}) - u(\mu)) \xrightarrow{D} N(0, \sigma^2(u'(\mu))^2) \\ \iff & \sqrt{n}(u(\bar{X}) - u(\mu)) \xrightarrow{D} N(0, \sigma^2(3\mu^2)^2) \\ \iff & \sqrt{n}(u(\bar{X}) - \mu^3) \xrightarrow{D} N(0, 9\sigma^2\mu^4). \end{aligned} \quad (7.103)$$

But of course, the convergence in distribution above is equivalent to

$$u(\bar{X}) = \bar{X}^3 \xrightarrow{D} N\left(\mu^3, \frac{9\sigma^2\mu^4}{n}\right), \quad (7.104)$$

again by change of variables. \square

7.5 Problem set 5

6.1.1

6.1.1. Let X_1, X_2, \dots, X_n be a random sample on X that has a $\Gamma(\alpha = 4, \beta = \theta)$ distribution, $0 < \theta < \infty$.

- (a) Determine the mle of θ .
- (b) Suppose the following data is a realization (rounded) of a random sample on X . Obtain a histogram with the argument `pr=T` (data are in `ex6111.rda`).

```
9 39 38 23 8 47 21 22 18 10 17 22 14
9 5 26 11 31 15 25 9 29 28 19 8
```

- (c) For this sample, obtain $\hat{\theta}$ the realized value of the mle and locate $4\hat{\theta}$ on the histogram. Overlay the $\Gamma(\alpha = 4, \beta = \hat{\theta})$ pdf on the histogram. Does the data agree with this pdf? Code for overlay:

```
xs=sort(x);y=dgamma(xs,4,1/betahat);hist(x,pr=T);lines(y~xs).
```

Solution:

- (a) The likelihood function is

$$\mathcal{L}(\theta) = \prod_{i=1}^n \frac{1}{\Gamma(4)\theta^4} x_i^{4-1} e^{-\theta x_i} = \frac{1}{\Gamma(4)^n} \frac{1}{\theta^{4n}} e^{-\sum_{i=1}^n x_i/\theta} \prod_{i=1}^n x_i^3. \quad (7.105)$$

The log likelihood function is then

$$l(\theta) = -n \ln(\Gamma(4)) - 4n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n x_i + \ln \left(\prod_{i=1}^n x_i^3 \right). \quad (7.106)$$

Next, solve for $\partial_\theta l(\theta) = 0$:

$$\partial_\theta l(\theta) = -\frac{4n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i = 0 \iff \boxed{\hat{\theta}_{ML} = \frac{1}{4n} \sum_{i=1}^n x_i = \bar{x}} \quad (7.107)$$

- (b) **R code:**

```
> dat = c(9, 39, 38, 23, 8, 47, 21, 22, 18, 10, 17, 22, 14,
+       9, 5, 26, 11, 31, 15, 25, 9, 29, 28, 19, 8)
> hist(dat, pr=T)
```

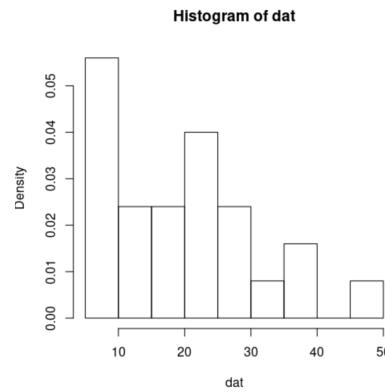


Figure 7.1: (b)

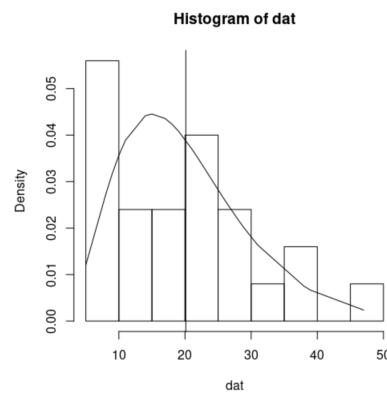
(c) $\hat{\theta} = \bar{x}/4 = 5.03$ **R code:**

```
> mean(dat)/4
[1] 5.03

> xs=sort(dat)
> y=dgamma(xs,4,4/mean(dat))
> hist(dat,pr=T)
> lines(y~xs)
```

Locate $4\hat{\theta}$ on the histogram and overlay with $\Gamma(\alpha = 4, \beta = \hat{\theta} = 5.03)$:

```
> xs=sort(dat)
> y=dgamma(xs,4,4/mean(dat))
> hist(dat,pr=T)
> lines(y~xs)
> abline(v=mean(dat))
```



The data somewhat agrees with this pdf.

□

6.1.2

6.1.2. Let X_1, X_2, \dots, X_n represent a random sample from each of the distributions having the following pdfs:

- (a) $f(x; \theta) = \theta x^{\theta-1}$, $0 < x < 1$, $0 < \theta < \infty$, zero elsewhere.
- (b) $f(x; \theta) = e^{-(x-\theta)}$, $\theta \leq x < \infty$, $-\infty < \theta < \infty$, zero elsewhere. Note that this is a nonregular case.

In each case find the mle $\hat{\theta}$ of θ .

Solution:

- (a) The likelihood function is

$$\mathcal{L}(\theta) = \theta^n \prod_{i=1}^n x_i^{\theta-1}, \quad x_i \in (0, 1), 0 < \theta < \infty. \quad (7.108)$$

The log likelihood is

$$l(\theta) = \ln \mathcal{L}(\theta) = n \ln \theta + (\theta - 1) \sum_{i=1}^n \ln x_i. \quad (7.109)$$

Then we solve for θ in $\partial_\theta l(\theta) = 0$:

$$\partial_\theta l(\theta) = \frac{n}{\theta} + \sum_{i=1}^n \ln(x_i) = 0 \implies \boxed{\hat{\theta} = \frac{-n}{\sum_{i=1}^n \ln x_i}} \quad (7.110)$$

- (b) The likelihood function is

$$\mathcal{L}(\theta) = e^{-\sum_{i=1}^n (x_i - \theta)}, \quad \theta < x_i < \infty, -\infty < \theta < \infty. \quad (7.111)$$

Then the log likelihood is

$$l(\theta) = \ln \mathcal{L}(\theta) = -\sum_{i=1}^n x_i + n\theta. \quad (7.112)$$

Next,

$$\partial_\theta l(\theta) = n > 0. \quad (7.113)$$

So, $\hat{\theta}$ must be as large as possible to maximize $\mathcal{L}(\theta)$. But at the same time, $\theta \leq x_i$ for all i , so

$$\boxed{\hat{\theta} = \min_i(X_i)} \quad (7.114)$$

□

6.1.4

6.1.4. Suppose X_1, \dots, X_n are iid with pdf $f(x; \theta) = 2x/\theta^2$, $0 < x \leq \theta$, zero elsewhere. Note this is a nonregular case. Find:

- (a) The mle $\hat{\theta}$ for θ .
- (b) The constant c so that $E(c\hat{\theta}) = \theta$.
- (c) The mle for the median of the distribution. Show that it is a consistent estimator.

Solution:

- (a) The likelihood function is

$$\mathcal{L}(\theta) = \frac{2^n}{\theta^{2n}} \prod_{i=1}^n x_i, \quad 0 < x_i \leq \theta. \quad (7.115)$$

The log likelihood is then

$$l(\theta) = n \ln 2 - 2n \ln \theta + \ln \left(\prod_{i=1}^n x_i \right). \quad (7.116)$$

Next,

$$\partial_\theta l(\theta) = -\frac{2n}{\theta}. \quad (7.117)$$

We cannot set this to zero. However, by inspection, $\mathcal{L}(\theta)$ is maximized whenever θ is minimized while $x_i \leq \theta$ for all i , so

$$\boxed{\hat{\theta} = \max(X_i) = Y_n} \quad (7.118)$$

- (b) To find c we first find $E(\hat{\theta})$. To get this, we must first find its cdf.

$$F_{Y_n}(x) = P(Y_n \leq x) = \prod_{i=1}^n P(x_i \leq x) = \prod_{i=1}^n F_X(x) = \frac{x^{2n}}{\theta^{2n}}. \quad (7.119)$$

Differentiating this w.r.t. x we get the pdf of Y_n :

$$f_{Y_n}(x) = \partial_x F_{Y_n}(x) = 2n \frac{x^{2n-1}}{\theta^{2n}}. \quad (7.120)$$

From here, calculating the expectation is easy:

$$E(Y_n) = \int_0^\theta 2n \frac{x^{2n-1} \cdot x}{\theta^{2n}} dx = \frac{2n}{2n+1} \theta. \quad (7.121)$$

Because we want $E(c\hat{\theta}) = cE(\hat{\theta}) = \theta$, we can just make

$$\boxed{c = \frac{2n+1}{2n}} \quad (7.122)$$

(c) The median $x_{1/2}$ is such that:

$$\frac{1}{2} = \int_0^{x_{1/2}} \frac{2x}{\theta^2} dx = \frac{x_{1/2}^2}{\theta^2}. \quad (7.123)$$

So, $x_{1/2} = \theta/\sqrt{2}$. By the invariance property, we have

$$\hat{x}_{1/2} = \frac{\hat{\theta}}{\sqrt{2}}$$

(7.124)

Since $E(\hat{\theta}) = (2n+1)/2n \cdot \theta$, we see that $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$. Next, look at the variance of $\hat{\theta}$:

$$\begin{aligned} \text{Var}(\hat{\theta}) &= \text{Var}(Y_n) = E[\hat{\theta}^2] - E[\hat{\theta}]^2 \\ &= \int_0^\theta 2n \frac{x^{2n-1} \cdot x^2}{\theta^{2n}} dx - \left(\frac{2n}{2n+1} \right)^2 \theta^2 \\ &= \frac{2n}{2+2n} \theta^2 - \left(\frac{2n}{2n+1} \right)^2 \theta^2. \end{aligned} \quad (7.125)$$

Obviously, $\lim_{n \rightarrow \infty} 2n/(2+2n) - (2n/(2n+1))^2 = 0$, so $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}) = 0$. With this, we conclude $\hat{\theta}$ is a consistent estimator of θ .

□
□

6.1.10

6.1.10. Let X_1, X_2, \dots, X_n be a random sample from a Bernoulli distribution with parameter p . If p is restricted so that we know that $\frac{1}{2} \leq p \leq 1$, find the mle of this parameter.

Solution: The likelihood function is

$$\mathcal{L}(x; p) = \prod_{i=1}^n f(x_i; p) = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}. \quad (7.126)$$

The log likelihood is then

$$l(p) = \sum_{i=1}^n x_i \ln p + \left(n - \sum_{i=1}^n x_i \right) \ln(1-p). \quad (7.127)$$

Taking ∂_p of $l(p)$ gives

$$\partial_p l(p) = \frac{\sum_{i=1}^n x_i}{p} + \frac{n - \sum_{i=1}^n x_i}{1-p}. \quad (7.128)$$

Letting $\partial_p l(p) = 0$, we get

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X} \quad (7.129)$$

as expected. Now, since $1/2 \leq p \leq 1$, we must consider the case where $\bar{X} < 1/2$. If $\bar{X} < 1/2$, then because p cannot take the value of \bar{X} , it must take the boundary value of $1/2$ at which $l(p)$ is maximized. So,

$$\hat{p} = \max\left(\frac{1}{2}, \bar{X}\right)$$

(7.130)

□

6.2.2

6.2.2. Given $f(x; \theta) = 1/\theta$, $0 < x < \theta$, zero elsewhere, with $\theta > 0$, formally compute the reciprocal of

$$nE\left\{\left[\frac{\partial \log f(X : \theta)}{\partial \theta}\right]^2\right\}.$$

Compare this with the variance of $(n+1)Y_n/n$, where Y_n is the largest observation of a random sample of size n from this distribution. Comment.

Solution:

$$nE[(\partial_\theta \ln f(x; \theta))^2] = nE[(\partial_\theta(-\ln \theta))^2] = nE\left[\frac{1}{\theta^2}\right] = \frac{n}{\theta^2}. \quad (7.131)$$

The reciprocal of this is therefore

$$\frac{\theta^2}{n}. \quad (7.132)$$

On the other hand,

$$\text{Var}\left[\frac{(n+1)}{n} Y_n\right] = \left(\frac{n+1}{n}\right)^2 (E[Y_n^2] - E[Y_n]^2). \quad (7.133)$$

The cdf of x is given by $F_X(x) = x/\theta$. The cdf of Y_n is therefore

$$F_Y(x) = \prod_{i=1}^n F_x(x) = \frac{x^n}{\theta^n} \implies f_Y(x) = n \frac{x^{n-1}}{\theta^n}. \quad (7.134)$$

And so

$$E[Y_n] = \int_0^\theta n \frac{x^{n-1} \cdot x}{\theta^n} dx = \frac{n\theta}{n+1}. \quad (7.135)$$

$$E[Y_n^2] = \int_0^\theta n \frac{x^{n-1} \cdot x^2}{\theta^n} dx = \frac{n\theta^2}{n+2}. \quad (7.136)$$

So,

$$\text{Var}\left[\frac{(n+1)}{n} Y_n\right] = \left(\frac{n+1}{n}\right)^2 \left[\frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 \right] = \frac{\theta^2}{n(n+2)}. \quad (7.137)$$

So, obviously,

$$\text{Var}[Y_n] = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{n}. \quad (7.138)$$

However, since the support of $f(x; \theta)$ depends on θ , CRLB does not apply. \square

6.2.7

6.2.7. Recall Exercise 6.1.1 where X_1, X_2, \dots, X_n is a random sample on X that has a $\Gamma(\alpha = 4, \beta = \theta)$ distribution, $0 < \theta < \infty$.

- (a) Find the Fisher information $I(\theta)$.
- (b) Show that the mle of θ , which was derived in Exercise 6.1.1, is an efficient estimator of θ .
- (c) Using Theorem 6.2.2, obtain the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$.
- (d) For the data of Example 6.1.1, find the asymptotic 95% confidence interval for θ .

Solution:

- (a) From 6.1.1. we know that

$$\partial_\theta l(\theta) = -\frac{4n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i \implies \partial_\theta^2 l(\theta) = \frac{4n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^n x_i \quad (7.139)$$

And so

$$\begin{aligned} nI(\theta) &= -E \left[\frac{4n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^n x_i \right] \\ &= -\frac{4n}{\theta^2} + \frac{2}{\theta^3} \sum_{i=1}^n E[x_i] \\ &= -\frac{4n}{\theta^2} + \frac{2}{\theta^3} n4\theta \\ &= -\frac{4n}{\theta^2} + \frac{8n}{\theta^2} \\ &= \boxed{\frac{4n}{\theta^2}} \implies \boxed{I(\theta) = \frac{4}{\theta^2}} \end{aligned} \quad (7.140)$$

- (b) From 6.1.1., $\hat{\theta} = \bar{X}/4$, which is unbiased because

$$E[\hat{\theta}] = \frac{1}{4} E[\bar{X}] = \frac{1}{4} \frac{4n\theta}{n} = \theta. \quad (7.141)$$

This is an efficient estimator of θ because

$$\begin{aligned}\text{Var}[\hat{\theta}] &= \frac{1}{16} \text{Var}[\bar{X}] = \frac{1}{16n^2} \text{Var} \left[\sum_{i=1}^n x_i \right] \\ &= \frac{1}{16n^2} \sum_{i=1}^n \text{Var}[x_i] \\ &= \frac{1}{16n^2} n 4\theta^2 \\ &= \frac{\theta^2}{4n}.\end{aligned}\tag{7.142}$$

So, we see that

$$\text{Var}[\hat{\theta}] = \frac{1}{nI(\theta)},\tag{7.143}$$

which means $\text{Var}(Y)$ attains the Rao-Cramér lower bound.

- (c) The observations are iid with $\Gamma(\alpha = 4, \beta = \theta)$. The regularity conditions are also satisfied. Since the Fisher information $I(\theta) = 4n/\theta^2$ is positive and finite, Theorem 6.2.2. says that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N} \left(0, \frac{1}{I(\theta)} \right) = \boxed{\mathcal{N} \left(0, \frac{\theta^2}{4} \right)}\tag{7.144}$$

- (d) The 95% CI for θ , with $\alpha = 0.025, n = 25, \sigma = \hat{\theta}/2$, is given by

$$\begin{aligned}\left(\hat{\theta} - z_{\alpha/2} \frac{\hat{\theta}/2}{\sqrt{n}}, \hat{\theta} + z_{\alpha/2} \frac{\hat{\theta}/2}{\sqrt{n}} \right) &= \left(5.03 - 1.96 \frac{5.03/2}{5}, 5.03 + 1.96 \frac{5.03/2}{5} \right) \\ &= \boxed{(4.04, 6.02)}\end{aligned}\tag{7.145}$$

□

6.2.8

6.2.8. Let X be $N(0, \theta)$, $0 < \theta < \infty$.

- (a) Find the Fisher information $I(\theta)$.
- (b) If X_1, X_2, \dots, X_n is a random sample from this distribution, show that the mle of θ is an efficient estimator of θ .
- (c) What is the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$?

Solution:

- (a) Well, the likelihood is just the pdf:

$$f(x; \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2\theta}, \theta \in \mathbb{R}^+ \setminus \{\infty\}. \quad (7.146)$$

Then, the log likelihood is

$$l(\theta) = -\frac{1}{2} \ln(2\pi\theta) - \frac{x^2}{2\theta} \implies \partial_\theta l(\theta) = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2}. \quad (7.147)$$

The Fisher information is then

$$\begin{aligned} I(\theta) &= -E[\partial_\theta^2 l(\theta)] \\ &= -E\left[\frac{1}{2\theta^2} - \frac{X^2}{\theta^3}\right] \\ &= -\frac{1}{2\theta^2} + \frac{1}{\theta^3} E[X^2] \\ &= -\frac{1}{2\theta^2} + \frac{1}{\theta^3} (\text{Var}[X] + \cancel{E[X]^2}) \\ &= -\frac{1}{2\theta^2} + \frac{1}{\theta^3} \theta \\ &= \boxed{\frac{1}{2\theta^2}} \end{aligned} \quad (7.148)$$

- (b) We know that the mle of θ is given by

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n (x_i - 0)^2, \quad (7.149)$$

which is just the sample variance with denominator n . Now, $\hat{\theta}$ is an unbiased estimator of θ because

$$E[\hat{\theta}] = \frac{1}{n} \sum_{i=1}^n E[X_i^2 + 2X_i \cdot 0 + 0] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] = \frac{n}{n} \text{Var}[X] = \theta. \quad (7.150)$$

Next we want to find $\text{Var}[\hat{\theta}]$. Well,

$$\begin{aligned}
 \text{Var}[\hat{\theta}] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i^2\right] \\
 &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i^2] \\
 &= \frac{n}{n^2} (E[X_i^4] - E[X_i^2]^2) \\
 &= \frac{n}{n^2} (E[X_i^4] - \text{Var}[X_i]^2) \\
 &= \frac{n}{n^2} (E[X_i^4] - \theta^2). \tag{7.151}
 \end{aligned}$$

To find $E[X_i^4]$ we can either brute force:

$$E[X^4] = \int_{\mathbb{R}} x^4 \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2\theta} dx = 3\theta^2. \tag{7.152}$$

Or use the MGF:

$$\begin{aligned}
 M(t) &= e^{\mu t + \theta t^2/2} \\
 \implies E[X^4] &= M^{(4)}(0) = (3\theta^2 + 6t^2\theta^3 + t^4\theta^4)e^{\theta t^2/2} \Big|_{t=0} = 3\theta^2. \tag{7.153}
 \end{aligned}$$

And so,

$$\text{Var}[\hat{\theta}] = \frac{1}{n} (3\theta^2 - \theta^2) = \boxed{\frac{2\theta^2}{n}} \tag{7.154}$$

Because $\text{Var}[\hat{\theta}] = 2\theta^2/n = 1/nI(\theta)$, the CRLB is attained. Further, because $\hat{\theta}$ is an unbiased estimator of θ , $\hat{\theta}$ is an efficient estimator of θ .

- (c) Since all regularity conditions are met, we can use Theorem 6.2.2., which says that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}(0, 1/I(\theta)) = \boxed{\mathcal{N}(0, 2\theta^2)} \tag{7.155}$$

□

6.2.9

6.2.9. If X_1, X_2, \dots, X_n is a random sample from a distribution with pdf

$$f(x; \theta) = \begin{cases} \frac{3\theta^3}{(x+\theta)^4} & 0 < x < \infty, 0 < \theta < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

show that $Y = 2\bar{X}$ is an unbiased estimator of θ and determine its efficiency.

Solution: Well, we first show that Y is an unbiased estimator of θ :

$$E[Y] = 2E[\bar{X}] = \frac{2}{n}nE[X_i] = 2E[X_i] = 2 \int_0^\infty \frac{3x\theta^3}{(x+\theta)^4} dx = 2 \frac{\theta}{2} = \theta. \quad (7.156)$$

So, yes, Y is an unbiased estimator of θ .

The efficiency of Y is given by

$$\epsilon = \frac{CRLB}{\text{Var}[Y]} = \frac{1}{\text{Var}[Y] \cdot nI(\theta)} \quad (7.157)$$

To find this we next find the Fisher information:

$$\begin{aligned} I(\theta) &= -E[\partial_\theta^2 l(\theta)] \\ &= -E\left[\partial_\theta^2 \ln \frac{3\theta^3}{(X+\theta)^4}\right] \\ &= -E\left[\partial_\theta^2 \ln \frac{3\theta^3}{(X+\theta)^4}\right] \\ &= -E\left[\frac{\theta^2 - 6\theta X - 3X^2}{\theta^2(\theta+X)^2}\right] \\ &= -\int_0^\infty \frac{\theta^2 - 6\theta x - 3x^2}{\theta^2(\theta+x)^2} \cdot \frac{3\theta^3}{(x+\theta)^4} dx = \frac{3}{5\theta^2}. \end{aligned} \quad (7.158)$$

Next, we find the variance of Y :

$$\begin{aligned} \text{Var}[Y] &= \text{Var}[2\bar{X}] = \frac{4}{n^2} \text{Var}\left[\sum_{i=1}^n X_i\right] = \frac{4n}{n^2} \text{Var}[X_i] = \frac{4}{n} (E[X_i^2] - E[X_i]^2) \\ &= \frac{4}{n} \left\{ \int_0^\infty \frac{3x^2\theta^3}{(x+\theta)^4} dx - \left[\int_0^\infty \frac{3x\theta^2}{(x+\theta)^4} dx \right]^2 \right\} \\ &= \frac{4}{n} \left(\theta^2 - \frac{\theta^2}{4} \right) = \frac{3\theta^2}{n}. \end{aligned} \quad (7.159)$$

So,

$$\epsilon = \frac{5\theta^2}{3\theta^2 \cdot 3} = \boxed{\frac{5}{9}} \quad (7.160)$$

7.6 Problem set 6

6.3.1

6.3.1. The following data were generated from an exponential distribution with pdf $f(x; \theta) = (1/\theta)e^{-x/\theta}$, for $x > 0$, where $\theta = 40$.

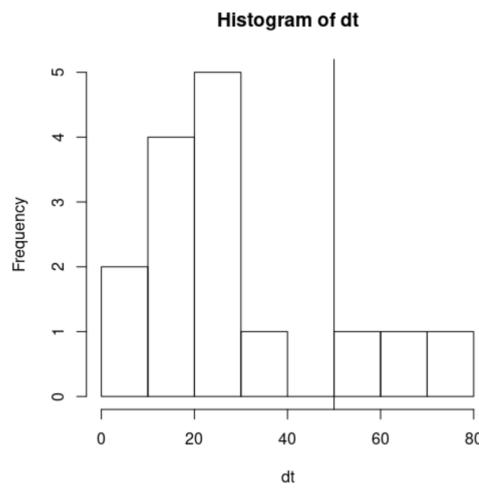
- (a) Histogram the data and locate $\theta_0 = 50$ on the plot.
- (b) Use the test described in Example 6.3.1 to test $H_0 : \theta = 50$ versus $H_1 : \theta \neq 50$. Determine the decision at level $\alpha = 0.10$.

19 15 76 23 24 66 27 12 25 7 6 16 51 26 39

Solution:

(a) **R code:**

```
> dt = c(19, 15, 76, 23, 24, 66, 27, 12, 25, 7, 6, 16, 51, 26, 39)
> hist(dt)
> abline(v=50)
```



- (b) $H_0 : \theta = 50$, $H_a : \theta \neq 50$. Example 6.3.1 says that we reject H_0 if

$$\frac{2}{\theta_0} \sum_{i=1}^n X_i \geq \chi^2_{1-\alpha/2}(2n) \text{ or } \frac{2}{\theta_0} \sum_{i=1}^n X_i \leq \chi^2_{\alpha/2}(2n) \quad (7.161)$$

In our case,

$$\tau = \frac{2}{\theta_0} \sum_{i=1}^n X_i = 17.28, \quad (7.162)$$

while

$$\begin{aligned}\chi^2_{0.10/2}(30) &= 18.49266 \\ \chi^2_{1-0.10/2}(30) &= 43.77297.\end{aligned}\tag{7.163}$$

Since $\tau < \chi^2_{0.10/2}(30)$, we **reject** H_0 .

R code:

```
> sum(dt)
[1] 432
> (2/50)*sum(dt)
[1] 17.28
> length(dt)
[1] 15
> qchisq(0.10/2,30)
[1] 18.49266
> qchisq(1-0.10/2,30)
[1] 43.77297
```

□

6.3.6

6.3.6. Let X_1, X_2, \dots, X_n be a random sample from a $N(\mu_0, \sigma^2 = \theta)$ distribution, where $0 < \theta < \infty$ and μ_0 is known. Show that the likelihood ratio test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ can be based upon the statistic $W = \sum_{i=1}^n (X_i - \mu_0)^2 / \theta_0$. Determine the null distribution of W and give, explicitly, the rejection rule for a level α test.

Solution: The mle for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \mu_0)^2. \quad (7.164)$$

With this,

$$\Lambda = \frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\hat{\theta})} = \left(\frac{\hat{\sigma}^2}{\theta_0} \right)^{n/2} \exp \left[\frac{-1}{2\theta_0} \sum_{i=1}^n (x_i - \mu_0)^2 + \frac{n}{2} \frac{\sum (x_i - \mu_0)^2}{\sum (x_i - \mu_0)^2} \right]. \quad (7.165)$$

This has the form $\Lambda = n^{-n/2} W^{n/2} e^{-W/2} e^{n/2}$, which is unimodal. So, $\Lambda \leq C \implies W \in (c_1, c_2)$ for some c_1, c_2 . This says we can use W as a testing statistic. Now, under $H_0 : \theta = \theta_0$, we have that $X_i \sim N(\mu_0, \theta = \theta_0)$, and so $(X_i - \mu_0) / \sqrt{\theta_0} \sim N(0, 1)$. This means $W = \sum^n (X_i - \mu_0)^2 / \theta_0 \sim \chi^2(n)$.

We reject whenever $W \geq \chi^2_{1-\alpha/2}(n)$ or $W \leq \chi^2_{\alpha/2}(n)$. □

6.3.11

6.3.11. Let X_1, X_2, \dots, X_n be a random sample from a $\Gamma(\alpha = 4, \beta = \theta)$ distribution, where $0 < \theta < \infty$.

- (a) Show that the likelihood ratio test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is based upon the statistic $W = \sum_{i=1}^n X_i$. Obtain the null distribution of $2W/\theta_0$.
- (b) For $\theta_0 = 3$ and $n = 5$, find c_1 and c_2 so that the test that rejects H_0 when $W \leq c_1$ or $W \geq c_2$ has significance level 0.05.

Solution:

- (a) We know that

$$\hat{\beta} = \frac{1}{n\alpha} \sum_{i=1}^n X_i = \frac{\bar{X}}{\alpha} = \frac{\bar{X}}{4}. \quad (7.166)$$

$$\begin{aligned} \Lambda &= \frac{\mathcal{L}(\hat{\theta})}{\mathcal{L}(\theta_0)} = \frac{1/\hat{\theta}^{4n}}{1/\theta_0^{4n}} \exp \left[-\left(\frac{1}{\hat{\theta}} - \frac{1}{\theta_0} \right) \sum x_i \right] \\ &\sim (4n)^{-4n} \left(\sum x_i / \theta_0 \right)^{-4n} \exp \left[-\sum X_i / \theta_0 \right] \exp [4n] \\ &= (4n)^{-4n} W^{-4n} \theta_0^{4n} \exp [-W/\theta_0] \exp [4n]. \end{aligned} \quad (7.167)$$

So, Λ depends on W as desired. Now, since $X_i \sim \Gamma(4, \theta_0)$ under H_0 , we have

$$\frac{2W}{\theta_0} = \frac{2}{\theta_0} \sum X_i \sim \Gamma \left(n\alpha, \frac{2\theta_0}{\theta_0} \right) = \Gamma(4n, 2) = \boxed{\chi^2(4n)} \quad (7.168)$$

(b)

$$\begin{aligned} c_1 &= \frac{\theta_0}{2} \chi_{\alpha/2}^2(4n) = \frac{3}{2} \times 9.590777 = \boxed{14.38617} \\ c_2 &= \frac{\theta_0}{2} \chi_{1-\alpha/2}^2(4n) = \frac{3}{2} \times 34.16961 = \boxed{51.25441} \end{aligned} \quad (7.169)$$

R code:

```
> qchisq(1-0.05/2, 4*5)
[1] 34.16961
> qchisq(0.05/2, 4*5)
[1] 9.590777
> 34.16961*1.5
[1] 51.25441
> 9.590777*1.5
[1] 14.38617
```

□

6.3.18

6.3.18. Let X_1, X_2, \dots, X_n be a random sample from a $\Gamma(\alpha, \beta)$ distribution where α is known and $\beta > 0$. Determine the likelihood ratio test for $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$.

Solution: The setup of this problem is exactly like that of 6.3.11. The likelihood ratio test of $H_0 : \beta = \beta_0$ versus $H_a : \beta \neq \beta_0$ is based upon the statistic $W = \sum X_i$. The null distribution of $2W/\beta_0$ is $\chi^2(n\alpha)$, under H_0 . We reject whenever $W \leq c_1$ or $W \geq c_2$, where

$$\begin{aligned} c_1 &= \frac{\beta_0}{2} \chi_{\alpha/2}^2(\alpha n) \\ c_2 &= \frac{\beta_0}{2} \chi_{1-\alpha/2}^2(\alpha n). \end{aligned} \tag{7.170}$$

□

6.4.1

6.4.1. A survey is taken of the citizens in a city as to whether or not they support the zoning plan that the city council is considering. The responses are: Yes, No, Indifferent, and Otherwise. Let p_1, p_2, p_3 , and p_4 denote the respective true probabilities of these responses. The results of the survey are:

Yes	No	Indifferent	Otherwise
60	45	70	25

- (a) Obtain the mles of p_i , $i = 1, \dots, 4$.
- (b) Obtain 95% confidence intervals, (4.2.7), for p_i , $i = 1, \dots, 4$.

Solution:

- (a) The mle for each p_i is

$$\hat{p}_i = \frac{\# \text{ Observations of } i}{n}. \quad (7.171)$$

- (b)

$$\begin{aligned} CI_1 &= \left(\hat{p}_1 - z_{\alpha/2} \sqrt{\hat{p}_1(1 - \hat{p}_1)/n}, \hat{p}_1 + z_{\alpha/2} \sqrt{\hat{p}_1(1 - \hat{p}_1)/n} \right) \\ &= \left(\frac{60}{200} - 1.96 \times 0.0324037, \frac{60}{200} + 1.96 \times 0.0324037 \right) \\ &= \boxed{(0.2364887, 0.3635113)} \end{aligned} \quad (7.172)$$

$$\begin{aligned} CI_2 &= \left(\frac{45}{200} - 1.96 \times 0.02952753, \frac{45}{200} + 1.96 \times 0.02952753 \right) \\ &= \boxed{(0.167126, 0.282874)} \end{aligned} \quad (7.173)$$

$$\begin{aligned} CI_3 &= \left(\frac{70}{200} - 1.96 \times 0.03372684, \frac{70}{200} + 1.96 \times 0.03372684 \right) \\ &= \boxed{(0.2838954, 0.4161046)} \end{aligned} \quad (7.174)$$

$$\begin{aligned} CI_4 &= \left(\frac{25}{200} - 1.96 \times 0.02338536, \frac{25}{200} + 1.96 \times 0.02338536 \right) \\ &= \boxed{(0.0791647, 0.7732091)} \end{aligned} \quad (7.175)$$

□

6.4.3

6.4.3. Let X_1, X_2, \dots, X_n be iid, each with the distribution having pdf $f(x; \theta_1, \theta_2) = (1/\theta_2)e^{-(x-\theta_1)/\theta_2}$, $\theta_1 \leq x < \infty$, $-\infty < \theta_2 < \infty$, zero elsewhere. Find the maximum likelihood estimators of θ_1 and θ_2 .

Solution: The likelihood function is

$$\mathcal{L}(\theta_1, \theta_2) = \frac{1}{\theta_2^n} \exp \left[-\frac{1}{\theta_2} \sum_{i=1}^n (x_i - \theta_1) \right]. \quad (7.176)$$

with the support given in the problem. We find

$$\ln \mathcal{L} = -n \ln \theta_2 - \frac{1}{\theta_2} \sum_{i=1}^n (x_i - \theta_1). \quad (7.177)$$

And so,

$$\begin{aligned} \partial_{\theta_1} \ln \mathcal{L}(\theta_1, \theta_2) &= \frac{n}{\theta_2} > 0 \implies \boxed{\hat{\theta}_1 = \text{Argmax } \ln \mathcal{L}(\theta_1, \theta_2) = \min_i X_i} \\ \partial_{\theta_2} \ln \mathcal{L}(\hat{\theta}_1, \theta_2) &= -\frac{n}{\theta_2} + \frac{1}{\theta_2^2} \sum_{i=1}^n (x_i - \hat{\theta}_1) = 0 \implies \boxed{\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \min_i X_i)} \end{aligned} \quad (7.178)$$

where we have used $\hat{\theta}_1 = \min_i X_i$ in writing the mle for θ_2 . \square

6.4.5

6.4.5. Let $Y_1 < Y_2 < \dots < Y_n$ be the order statistics of a random sample of size n from the uniform distribution of the continuous type over the closed interval $[\theta - \rho, \theta + \rho]$. Find the maximum likelihood estimators for θ and ρ . Are these two unbiased estimators?

Solution: It's not difficult to see that

$$\begin{aligned}\hat{\theta} - \hat{\rho} &= \min Y_i = Y_1 \\ \hat{\theta} + \hat{\rho} &= \max Y_i = Y_n.\end{aligned}\tag{7.179}$$

Solving this for $\hat{\theta}, \hat{\rho}$ gives

$$\begin{aligned}\hat{\theta} &= \frac{Y_n + Y_1}{2} \\ \hat{\rho} &= \frac{Y_n - Y_1}{2}.\end{aligned}\tag{7.180}$$

Next,

$$\begin{aligned}F_{Y_1}(x) &= P(Y_1 \leq x) = P(\min_i Y_i \leq x) \\ &= 1 - P(\min_i Y_i > x) \\ &= 1 - \prod_{i=1}^n P(Y_i > x) \\ &= 1 - (1 - F_{Y_i}(x))^n = 1 - \left(1 - \int_{\theta-\rho}^x \frac{1}{2\rho} dx\right)^n \\ &= 1 - \left(1 - \frac{x - \theta + \rho}{2\rho}\right)^n \\ \implies f_{Y_1}(x) &= \frac{n(\theta + \rho - x)^{n-1}}{(2\rho)^n}\end{aligned}\tag{7.181}$$

And so,

$$E[Y_1] = \int_{\theta-\rho}^{\theta+\rho} x \frac{n(\theta + \rho - x)^{n-1}}{(2\rho)^n} dx = \frac{1-n}{1+n} \rho + \theta\tag{7.182}$$

Next,

$$\begin{aligned}
 F_{Y_n}(x) &= P(Y_n \leq x) = P(\max_i Y_i \leq x) \\
 &= \prod_{i=1}^n P(Y_i \leq x) = \prod_{i=1}^n F_{Y_i} \\
 &= \frac{(x - \theta + \rho)^n}{(2\rho)^n} \\
 \implies f_{Y_n}(x) &= \frac{n(x - \theta + \rho)^{n-1}}{(2\rho)^n}. \tag{7.183}
 \end{aligned}$$

And so,

$$E[Y_n] = \int_{\theta-\rho}^{\theta+\rho} x \frac{n(x - \theta + \rho)^{n-1}}{(2\rho)^n} dx = \frac{n-1}{n+1} \rho + \theta. \tag{7.184}$$

With these,

$$\begin{aligned}
 E[\hat{\theta}] &= E\left[\frac{Y_n + Y_1}{2}\right] = \frac{1}{2}E[2\theta] = \theta \\
 E[\hat{\rho}] &= E\left[\frac{Y_n - Y_1}{2}\right] = \frac{1}{2}E\left[\frac{n-1-1+n}{n+1}\rho\right] = \rho. \tag{7.185}
 \end{aligned}$$

So, $\hat{\theta}$ and $\hat{\rho}$ are unbiased. \square

6.5.3

6.5.3. Consider the survey of citizens discussed in Exercise 6.4.1. Suppose that the hypotheses of interest are $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$. Note that computation can be carried out using the R function `p2pair.R`, which can be downloaded at the site mentioned in the Preface.

- (a) Test these hypotheses at level $\alpha = 0.05$ using the test (6.5.20). Conclude in terms of the problem.
- (b) Obtain the 95% confidence interval, (6.5.21), for $p_1 - p_2$. What does the confidence interval mean in terms of the problem?

Solution:

- (a) $H_0 : p_1 = p_2$ and $H_a : p_1 \neq p_2$.

$$\begin{aligned}\chi_W^2 &= \frac{(\hat{p}_1 - \hat{p}_2)^2}{(\hat{p}_1 + \hat{p}_2 - (\hat{p}_1 - \hat{p}_2)^2)/n} \\ &= \frac{(60/200 - 45/200)^2}{(60/200 + 45/200 - (60/200 - 45/200)^2)/200} \\ &= \boxed{2.167}\end{aligned}\tag{7.186}$$

Clearly, $2.167 < 3.81 = \chi_{0.05}^2(1)$. So, we fail to reject H_0 – there is not enough evidence to reject the null hypothesis that the proportions of people who vote Yes and No are the same.

- (b) The 95% CI for $p_1 - p_2$ is

$$\left(\frac{60}{200} - \frac{45}{200} \right) \pm z_{0.05/2} \left(\frac{\frac{60}{200} + \frac{45}{200} - \left(\frac{60}{200} - \frac{45}{200} \right)^2}{200} \right)^{1/2} \tag{7.187}$$

$$= (-0.025, 0.175). \tag{7.188}$$

Since the CI contains 0, we fail to reject H_0 .

□

6.5.6

6.5.6. Let X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m be independent random samples from the two normal distributions $N(0, \theta_1)$ and $N(0, \theta_2)$.

- (a) Find the likelihood ratio Λ for testing the composite hypothesis $H_0 : \theta_1 = \theta_2$ against the composite alternative $H_1 : \theta_1 \neq \theta_2$.
- (b) This Λ is a function of what F -statistic that would actually be used in this test?

Solution:

- (a) $H_0 : \theta_1 = \theta_2$ and $H_a : \theta_1 \neq \theta_2$. $X_i \sim N(0, \theta_1)$ and $Y_i \sim N(0, \theta_2)$. So,

$$\mathcal{L}(\theta_1, \theta_2) = (2\pi\theta_1)^{-n/2}(2\pi\theta_2)^{-m/2} \exp \left[-\frac{1}{2\theta_1} \sum_{i=1}^n x_i^2 - \frac{1}{2\theta_2} \sum_{i=1}^m y_i^2 \right]. \quad (7.189)$$

Under $H_0 : \theta_1 = \theta_2$:

$$\begin{aligned} \partial_\theta \ln \mathcal{L}(\theta_1 = \theta_2 = \theta) &= 0 \implies -\frac{n+m}{2} - \frac{1}{2\theta} \left[\sum_{i=1}^n x_i^2 + \sum_{i=1}^m y_i^2 \right] = 0 \\ &\implies \hat{\theta} = \frac{1}{n+m} \left[\sum_{i=1}^n x_i^2 + \sum_{i=1}^m y_i^2 \right]. \end{aligned} \quad (7.190)$$

With this,

$$\begin{aligned} \mathcal{L}(\hat{\Theta}_0) &= \left(\frac{2\pi}{n+m} \left[\sum_{i=1}^n x_i^2 + \sum_{i=1}^m y_i^2 \right] \right)^{-\frac{n+m}{2}} \exp \left[-\frac{\sum_{i=1}^n x_i^2 + \sum_{i=1}^m y_i^2}{\frac{2}{n+m} (\sum_{i=1}^n x_i^2 + \sum_{i=1}^m y_i^2)} \right] \\ &= \left(\frac{2\pi}{n+m} \left[\sum_{i=1}^n x_i^2 + \sum_{i=1}^m y_i^2 \right] \right)^{-\frac{n+m}{2}} \exp \left[\frac{-n-m}{2} \right]. \end{aligned} \quad (7.191)$$

In the joint space,

$$\begin{aligned} \hat{\theta}_1 &= \frac{1}{n} \sum_{i=1}^n x_i^2 \\ \hat{\theta}_2 &= \frac{1}{m} \sum_{i=1}^m y_i^2. \end{aligned} \quad (7.192)$$

With these,

$$\mathcal{L}(\hat{\Theta}) = \left(\frac{2\pi}{n} \sum_{i=1}^n x_i^2 \right)^{-n/2} \left(\frac{2\pi}{m} \sum_{i=1}^m y_i^2 \right)^{-m/2} \exp \left[\frac{-n-m}{2} \right]. \quad (7.193)$$

And so we have

$$\begin{aligned}
 \Lambda &= \frac{\mathcal{L}_0}{\mathcal{L}} = \frac{\left(\frac{1}{n+m} [\sum^n x_i^2 + \sum^m y_i^2]\right)^{-\frac{n+m}{2}}}{\left(\frac{1}{n} \sum^n x_i^2\right)^{-n/2} \left(\frac{1}{m} \sum^m y_i^2\right)^{-m/2}} \\
 &= \frac{(n+m)^{\frac{n+m}{2}}}{n^{n/2} m^{m/2}} \frac{(\sum^n x_i^2)^{n/2} (\sum^m y_i^2)^{m/2}}{[\sum^n x_i^2 + \sum^m y_i^2]^{\frac{n+m}{2}}} \\
 &= \frac{(n+m)^{\frac{n+m}{2}}}{n^{n/2} m^{m/2}} \left(\underbrace{\sum^n x_i^2}_{T} \right)^{n/2} \left(\underbrace{\sum^m y_i^2}_{1-T} \right)^{m/2}.
 \end{aligned} \tag{7.194}$$

(b) $\Lambda \leq k$ is equivalent to $W \leq c_1$ or $W \geq c_2$ with

$$W = \frac{\frac{1}{n} \sum^n x_i^2}{\frac{1}{m} \sum^m y_i^2} \sim F(n, m)$$

(7.195)

such that $1/T = 1 + 1/W$. This is because $\sum^n x_i^2/\theta \sim \chi^2(n)$, $\sum^m y_i^2/\theta \sim \chi^2(m)$ under H_0 .

□

6.5.7

6.5.7. Let X and Y be two independent random variables with respective pdfs

$$f(x; \theta_i) = \begin{cases} \left(\frac{1}{\theta_i}\right) e^{-x/\theta_i} & 0 < x < \infty, 0 < \theta_i < \infty \\ 0 & \text{elsewhere,} \end{cases}$$

for $i = 1, 2$. To test $H_0 : \theta_1 = \theta_2$ against $H_1 : \theta_1 \neq \theta_2$, two independent samples of sizes n_1 and n_2 , respectively, were taken from these distributions. Find the likelihood ratio Λ and show that Λ can be written as a function of a statistic having an F -distribution, under H_0 .

Solution: We did this in “class,” so I will skip a few steps. Let $X_i \sim \exp(\theta_1)$ and $Y_i \sim \exp(\theta_2)$. $H_0 : \theta_1 = \theta_2$ and $H_a : \theta_1 \neq \theta_2$. The null space is $\{\{\theta_1, \theta_2\} : \theta_1 = \theta_2 > 0\}$. The alternative space is $\{\{\theta_1, \theta_2\} : \theta_1 \neq \theta_2, \theta_1, \theta_2 \neq 0\}$. The joint space is just $\{\{\theta_1, \theta_2\}, \theta_1 > 0, \theta_2 > 0\}$. It’s easy to see that

$$\mathcal{L}(\theta_1, \theta_2 | X, Y) = \frac{1}{\theta_1^n} e^{\sum x_i / \theta_1} \frac{1}{\theta_2^m} e^{\sum y_i / \theta_2}. \quad (7.196)$$

From here one finds that under $H_0 : \theta_1 = \theta_2 = \theta$:

$$l(\Theta_0) = -(n+m) \ln \theta - \frac{1}{\theta} \left(\sum x_i + \sum y_i \right). \quad (7.197)$$

And so

$$\partial_\theta l(\theta) = 0 \implies \hat{\theta}_0 = \frac{1}{n+m} \left(\sum x_i + \sum y_i \right). \quad (7.198)$$

With this, we can plug back in to calculate the numerator:

$$\mathcal{L}(\hat{\Theta}_0) = \dots = \left(\frac{1}{n+m} \left[\sum x_i + \sum y_j \right] \right)^{-n-m} \exp(-n-m). \quad (7.199)$$

In the joint space,

$$l(\Theta) = -n \ln \theta - \frac{1}{\theta} \sum x_i - m \ln \mu - \frac{1}{\mu} \sum y_j. \quad (7.200)$$

And so

$$\partial_{\theta_1} l(\Theta) = 0 \implies \hat{\theta}_1 = \bar{x} \quad (7.201)$$

$$\partial_{\theta_2} l(\Theta) = 0 \implies \hat{\theta}_2 = \bar{y}. \quad (7.202)$$

With these,

$$\mathcal{L}(\hat{\Theta}) = \dots \quad (7.203)$$

$$= \frac{1}{\left(\frac{1}{n} \sum x_i\right)^n} \exp \left[-\frac{1}{\frac{1}{n} \sum x_i} \sum x_i \right] \frac{1}{\left(\frac{1}{m} \sum y_i\right)^m} \exp \left[-\frac{1}{\frac{1}{m} \sum y_i} \sum y_i \right] \quad (7.204)$$

$$= \frac{1}{\left(\frac{1}{n} \sum x_i\right)^n} e^{-n} \frac{1}{\left(\frac{1}{m} \sum y_i\right)^m} e^{-m}. \quad (7.205)$$

Putting everything together, we find

$$\Lambda = \frac{\mathcal{L}_0}{\mathcal{L}} = \dots = \frac{\frac{1}{(n+m)^{-n-m}} (\sum x_i + \sum y_j)^{-n-m}}{\frac{1}{n^n} \frac{1}{m^m} (\sum x_i)^{-n} (\sum y_j)^{-m}}. \quad (7.206)$$

We reject if $\Lambda < c$, iff

$$\frac{(n+m)^{n+m}}{n^n m^m} \frac{(\sum x_i)^n (\sum y_j)^m}{(\sum x_i + \sum y_j)^{n+m}} < c \quad (7.207)$$

iff (letting c absorb the constant)

$$\frac{(\sum x_i)^n (\sum y_j)^m}{(\sum x_i + \sum y_j)^{n+m}} < c' \quad (7.208)$$

What does the distribution of Λ look like? Notice that we reject if

$$\Lambda = \frac{(n+m)^{n+m}}{n^n m^m} \frac{(\sum x_i)^n (\sum y_j)^m}{(\sum x_i + \sum y_j)^{n+m}} \quad (7.209)$$

$$= \frac{(n+m)^{n+m}}{n^n m^m} \left(\frac{\sum x_i}{\sum x_i + \sum y_j} \right)^n \left(\frac{\sum y_j}{\sum x_i + \sum y_j} \right)^m \quad (7.210)$$

$$= \frac{(n+m)^{n+m}}{n^n m^m} T^n (1-T)^m < c \quad (7.211)$$

which is akin to saying that we reject if $T < a$ or $T > b$. What is the distribution of T under H_0 ? Note that under H_0 , $\sum x_i \sim \Gamma(n, \theta)$ and $\sum y_i \sim \Gamma(m, \theta)$. By transformation method, we can show that $\boxed{T \sim \beta(n, m)}$.

Notice that

$$\frac{1}{T} = 1 + \frac{\sum y_j}{\sum x_i} = 1 + \frac{m \sum y_j / m}{n \sum x_i / n} \quad (7.212)$$

Since $\sum X_i \sim \Gamma(n, \theta)$, $\sum Y_i \sim \Gamma(m, \theta)$ under H_0 we have that $2\theta X_i \sim \chi^2(2n)$ and $2\theta Y_i \sim \chi^2(2m)$. This means we can call

$$\boxed{W = \frac{\sum y_j / m}{\sum x_i / n} = \frac{2\theta \sum y_j / m}{2\theta \sum x_i / n} \sim F(2m, 2n)}. \quad (7.213)$$

□

7.7 Problem set 7

7.2.1

7.2.1. Let X_1, X_2, \dots, X_n be iid $N(0, \theta)$, $0 < \theta < \infty$. Show that $\sum_1^n X_i^2$ is a sufficient statistic for θ .

Solution: On the one hand

$$\prod_{i=1}^n f(x_i; \theta) = (2\pi\theta)^{-n/2} e^{-\sum x_i^2/2\theta}. \quad (7.214)$$

On the other hand, $x_i/\sqrt{\theta} \sim N(0, 1)$, which means $Y = \sum X_i^2/\theta \sim \chi^2(n)$, whose pdf is

$$f_Y \left(\sum X_i^2 / \theta \right) = \frac{1}{2^{n/2} \Gamma(n/2)} \left(\sum X_i^2 / \theta \right)^{n/2-1} e^{-(\sum X_i^2 / \theta)/2}. \quad (7.215)$$

And so,

$$\frac{(2\pi\theta)^{-n/2} e^{-\sum x_i^2/2\theta}}{\theta^{-1} \frac{1}{2^{n/2} \Gamma(n/2)} (\sum X_i^2 / \theta)^{n/2-1} e^{-(\sum X_i^2 / \theta)/2}} \text{ is independent of } \theta, \quad (7.216)$$

where the factor θ^{-1} in the denominator comes from the scaling Jacobian. Since the ratio is independent of θ , we say $\sum X_i^2$ is sufficient.

Alternative, we can also use the factorization theorem here, with $k_2 = 1$ and $k_1 = \prod f(x_i; \theta)$. \square

7.2.4

7.2.4. Let X_1, X_2, \dots, X_n be a random sample of size n from a geometric distribution that has pmf $f(x; \theta) = (1 - \theta)^x \theta$, $x = 0, 1, 2, \dots$, $0 < \theta < 1$, zero elsewhere. Show that $\sum_{i=1}^n X_i$ is a sufficient statistic for θ .

Solution: On the one hand,

$$A = \prod_{i=1}^n f(x_i; \theta) = (1 - \theta)^{\sum x_i} \theta^n. \quad (7.217)$$

On the other hand, since each $X_i \sim \text{Geom}(\theta)$, the sum $Y = \sum_{i=1}^n X_i \sim \text{NegBin}(n, \theta)$, which means

$$B = f_Y \left(y = \sum x_i \right) = \binom{\sum x_i + r - 1}{\sum x_i} \theta^n (1 - \theta)^{\sum x_i}$$

Clearly, the ratio is independent of θ , so we say $\sum X_i$ is sufficient.

Alternative, we can also use the factorization theorem, with $k_2 = 1$ and $k_1 = A$. \square

7.2.7

7.2.7. Show that the product of the sample observations is a sufficient statistic for $\theta > 0$ if the random sample is taken from a gamma distribution with parameters $\alpha = \theta$ and $\beta = 6$.

Solution: Let X_1, \dots, X_n be iid r.v. from $\Gamma(\theta, \beta)$. Consider $Y = \prod X_i$. Then we have

$$\prod_{i=1}^n f(x_i; \theta) = \left[\frac{1}{\Gamma(\theta)^n \beta^{n\theta}} \prod_{i=1}^n x_i^{\theta-1} \right] [e^{\sum x_i / \beta}]$$

where $\beta = 6$ is fixed. We can identify the first term as k_1 and the second term as k_2 . k_1, k_2 are both nonnegative functions. k_1 is a function of Y and θ , and k_2 is a function of only the data. So the factorization theorem tells us Y is sufficient. \square

7.2.8

7.2.8. What is the sufficient statistic for θ if the sample arises from a beta distribution in which $\alpha = \beta = \theta > 0$?

Solution: Let X_1, \dots, X_n be iid r.v. from $B(\theta, \theta)$, then

$$f(x_i; \theta) = \frac{\Gamma(2\theta)x_i^{\theta-1}(1-x_i)^{\theta-1}}{\Gamma^2(\theta)} \quad (7.218)$$

So,

$$\prod_{i=1}^n f(x_i; \theta) = \frac{\Gamma^n(2\theta) \prod_{i=1}^n [x_i^{\theta-1}(1-x_i)^{\theta-1}]}{\Gamma^{2n}(\theta)} \quad (7.219)$$

We identify

$$\prod_{i=1}^n [x_i(1-x_i)]^{\theta-1}$$

as a sufficient statistic by the factorization theorem, where k_1 is the entire $\prod_{i=1}^n f(x_i; \theta)$ and $k_2 = 1$. \square

7.3.1 (Do only for exercises 7.2.1 and 7.2.4)

7.3.1. In each of Exercises 7.2.1–7.2.4, show that the mle of θ is a function of the sufficient statistic for θ .

Solution: Going back to 7.2.1 and 7.2.4 we see that this is the case. In both exercises, we can just set $k_2 = 1$ and $k_1 = \prod_{i=1}^n f(x_i; \theta) = \mathcal{L}(x_i; \theta)$ and we can easily see that the mle of θ is a function of the sufficient statistic for θ .

In 7.2.1, the mle of θ is $(1/n) \sum_{i=1}^n X_i^2$, and the sufficient statistic for θ is $\sum_{i=1}^n X_i^2$. So, $\hat{\theta} = (1/n)Y_\theta$ where Y_θ is the sufficient statistic for θ .

For 7.2.4, we have to find the mle of θ . Well, the log likelihood is

$$\partial_\theta l(\theta) = \sum x_i \frac{-1}{1-\theta} + \frac{n}{\theta} = 0 \implies \hat{\theta} = \frac{1}{\bar{X}+1}.$$

Of course, $\hat{\theta}$ is a function of the sum of the observations, $\sum X_i$, which is the sufficient statistic for θ . \square

7.3.3

7.3.3. If X_1, X_2 is a random sample of size 2 from a distribution having pdf $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere, find the joint pdf of the sufficient statistic $Y_1 = X_1 + X_2$ for θ and $Y_2 = X_2$. Show that Y_2 is an unbiased estimator of θ with variance θ^2 . Find $E(Y_2|y_1) = \varphi(y_1)$ and the variance of $\varphi(Y_1)$.

Solution: We have that $X_1 = Y_1 - Y_2$ and $X_2 = Y_2$ are iid r.v.'s. The joint pdf of Y_1, Y_2 is

$$f_{X_1, X_2} = f_{Y_1, Y_2}(y_1, y_2; \theta) = (1/\theta)^2 e^{-y_2/\theta}. \quad (7.220)$$

We have

$$E[Y_2] = \int_0^\infty y(1/\theta)^2 e^{-y/\theta} dy = \theta, \quad (7.221)$$

so Y_2 is unbiased. Also,

$$\text{Var}[Y_2] = E[Y_2^2] - \theta^2 = \int_0^\infty y(1/\theta)^2 e^{-y/\theta} dy - \theta^2 = 2\theta^2 - \theta^2 = \theta^2.$$

Next,

$$f_{Y_1}(y_1) = \int_0^{y_1} (1/\theta)^2 e^{-y_1/\theta} dy_2 = (1/\theta)^2 y_1 e^{-y_1/\theta}. \quad (7.222)$$

And so,

$$f_{Y_2|Y_1=y_1} = \frac{(1/\theta)^2 e^{-y_2/\theta}}{(1/\theta)^2 y_1 e^{-y_1/\theta}} = \frac{1}{y_1}, \quad y_2 \in (0, y_1) \quad (7.223)$$

So,

$$E[Y_2|y_1] = \frac{y_1}{2}.$$

Finally,

$$\text{Var}[Y_1/2] = \frac{1}{4} \text{Var}[Y_1] = \frac{1}{4} \text{Var}[X_1 + X_2] = \frac{2\theta^2}{4} = \frac{\theta^2}{2}.$$

□

7.3.4

7.3.4. Let $f(x, y) = (2/\theta^2)e^{-(x+y)/\theta}$, $0 < x < y < \infty$, zero elsewhere, be the joint pdf of the random variables X and Y .

(a) Show that the mean and the variance of Y are, respectively, $3\theta/2$ and $5\theta^2/4$.

(b) Show that $E(Y|x) = x + \theta$. In accordance with the theory, the expected value of $X + \theta$ is that of Y , namely, $3\theta/2$, and the variance of $X + \theta$ is less than that of Y . Show that the variance of $X + \theta$ is in fact $\theta^2/4$.

Solution:

(a)

$$E[Y_2] = \int_0^\infty \int_x^\infty y(2/\theta^2)e^{-(x+y)/\theta} dy dx = \frac{3\theta}{2}.$$

$$\text{Var}[Y_2] = \int_0^\infty \int_x^\infty y^2(2/\theta^2)e^{-(x+y)/\theta} dy dx = \frac{3\theta}{2} - \frac{9\theta^2}{4} = \frac{5\theta^2}{4}.$$

(b)

$$f_X(x) = \int_{y=x}^\infty (2/\theta^2)e^{-(x+y)/\theta} dy = (2/\theta)e^{-2x/\theta}, \quad x \in (0, \infty)$$

And so

$$f_Y(y) = f_{X,Y}/f_X = (1/\theta)e^{-(y-x)/\theta}.$$

So,

$$E[Y|x] = \int_x^\infty (1/\theta)e^{-(y-x)/\theta} dy = x + \theta.$$

$$\begin{aligned} \text{Var}[X + \theta] &= \text{Var}[X] = E[X^2] - E^2[X] \\ &= \int_0^\infty \frac{2x^2}{\theta}e^{-2x/\theta} dx - \left(\int_0^\infty \frac{2x}{\theta}e^{-2x/\theta} dx \right)^2 \\ &= \frac{\theta^2}{2} - \frac{\theta^2}{4} = \frac{\theta^2}{4}. \end{aligned}$$

□

7.4.3

7.4.3. Let X_1, X_2, \dots, X_n represent a random sample from the discrete distribution having the pmf

$$f(x; \theta) = \begin{cases} \theta^x (1 - \theta)^{1-x} & x = 0, 1, \quad 0 < \theta < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Show that $Y_1 = \sum_1^n X_i$ is a complete sufficient statistic for θ . Find the unique function of Y_1 that is the MVUE of θ .

Hint: Display $E[u(Y_1)] = 0$, show that the constant term $u(0)$ is equal to zero, divide both members of the equation by $\theta \neq 0$, and repeat the argument.

Solution:

$$\prod_{i=1}^n f(x_i; \theta) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$$

the factorization theorem tells us $\sum x_i$ is sufficient. Now, set $E[u(Y_1)] = 0$ then because

$$f_Y(y) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}, \quad (7.224)$$

we have

$$0 = u(0) \binom{n}{0} \theta^0 (1 - \theta)^n + u(1) \binom{n}{1} \theta^1 (1 - \theta)^{n-1} + \dots \quad (7.225)$$

We can expand this out into an n -degree polynomial in θ that is equal to zero for all θ . The leading coefficient is $u(0)$, so the polynomial is identically zero if $u(0) = 0$. Continue with this argument for the next term, we see that $u(1) = 0$ too. So, $u(x) = 0$ for all natural x . This means $u(Y_1) = 0$, so Y_1 is also complete. So, Y_1 is a complete statistic for θ .

To find a unique function of Y_1 that is the MVUE of θ , it suffices to find $\varphi(Y_1)$ such that $\varphi(Y_1)$ is an unbiased estimator for θ (Lehmann-Scheffé theorem). Well,

$$\varphi(Y_1) = \frac{1}{n} Y_1 = \frac{1}{n} \sum X_i$$

is an unbiased estimator for θ since

$$E[\sum X_i/n] = E[X_i] = \theta.$$

Because Y is complete sufficient, this function is the unique MVUE for θ . \square

7.4.5

7.4.5. Show that the first order statistic Y_1 of a random sample of size n from the distribution having pdf $f(x; \theta) = e^{-(x-\theta)}$, $\theta < x < \infty$, $-\infty < \theta < \infty$, zero elsewhere, is a complete sufficient statistic for θ . Find the unique function of this statistic which is the MVUE of θ .

Solution:

$$\begin{aligned} F_{Y_1}(y) &= P(Y_1 \leq y) = 1 - P(Y_1 \geq y) \\ &= 1 - \prod_{i=1}^n P(X_i \geq y) \\ &= 1 - \prod_{i=1}^n e^{-(y-\theta)} \\ &= 1 - e^{-n(y-\theta)} \end{aligned} \tag{7.226}$$

So,

$$f_{Y_1}(y) = ne^{-n(y-\theta)}. \tag{7.227}$$

So,

$$\frac{\prod_{i=1}^n e^{-(x_i-\theta)}}{ne^{-n(y-\theta)}} = \frac{e^{-\sum x_i}}{ne^{-ny_1}} \tag{7.228}$$

which is independent of θ . So Y_1 is sufficient. Now, suppose $E[u(Y_1)] = 0$, then

$$\int_{\theta}^{\infty} u(y) ne^{-n(y-\theta)} dy = 0 \implies nu(\theta) = 0 \implies u(\theta) = 0 \tag{7.229}$$

so Y_1 is complete. Now, the unique function of this statistic which is the MVUE of θ is a function of Y that is also an unbiased estimator for θ . Further,

$$E[Y_1] = \int_{\theta}^{\infty} ne^{-n(y-\theta)} dy = \frac{1}{n} + \theta.$$

And so the function $Y_1 - \frac{1}{n}$ is an unbiased estimator for θ . Since this is a function of Y_1 which is complete sufficient, $\min X_i - \frac{1}{n}$ is an MVUE of θ . \square

7.4.9

7.4.9. Let X_1, \dots, X_n be iid with pdf $f(x; \theta) = 1/(3\theta)$, $-\theta < x < 2\theta$, zero elsewhere, where $\theta > 0$.

- (a) Find the mle $\hat{\theta}$ of θ .
- (b) Is $\hat{\theta}$ a sufficient statistic for θ ? Why?
- (c) Is $(n+1)\hat{\theta}/n$ the unique MVUE of θ ? Why?

Solution:

1.

$$\mathcal{L}(\theta) = \frac{1}{3^n \theta^n} \prod_{i=1}^n I(-\theta < x_i < 2\theta)$$

this is a decreasing function in θ . Thus it is maximized whenever $\hat{\theta}$ is minimized, which means $\hat{\theta}$ has no choice but to be $\max\{-x_{(1)}, x_{(n)}/2\}$.

2. The factorization theorem tells us that $\max\{-x_{(1)}, x_{(n)}/2\}$ is sufficient because \mathcal{L} can be factored into K_1 which is essentially the entire likelihood function and K_2 as the constantly 1 function (multiplied by a characteristic equation which we won't worry about).
3. The pdf for Y_1 can be found via

$$\begin{aligned} F_{Y_1}(y) &= P(Y_1 \leq y) = 1 - P(Y_1 > y) \\ &= 1 - \prod_{i=1}^n P(X_i > y) \\ &= 1 - \left(1 - \frac{y + \theta}{3\theta}\right)^n. \end{aligned} \tag{7.230}$$

So,

$$f_{Y_1}(y) = \frac{n}{(3\theta)^n} [2\theta - y]^{n-1}. \tag{7.231}$$

Similarly,

$$f_{Y_n}(y) = \frac{n}{(3\theta)^n} [y + \theta]^{n-1}. \tag{7.232}$$

So, if

$$0 = E[u(Y_1)] = \int_{-\theta}^{2\theta} u(y) \frac{n}{(3\theta)^n} [2\theta - y]^{n-1} dy \stackrel{\partial_u}{\Rightarrow} u(-\theta) \frac{n}{(3\theta)^n} (3\theta)^{n-1} = 0 \tag{7.233}$$

$$0 = E[u(Y_n)] = \int_{-\theta}^{2\theta} u(y) \frac{n}{(3\theta)^n} [y + \theta]^{n-1} \xrightarrow{\partial_\theta} u(2\theta) \frac{n}{(3\theta)^n} (3\theta)^{n-1} = 0 \quad (7.234)$$

So by rescaling, $u(\theta) = 0$ identically. So, $\hat{\theta}$ is complete sufficient. So, if $(n+1)\hat{\theta}/n$ is an unbiased estimator for θ , it will be an MVUE for θ . Well, we need to find $E[\hat{\theta}]$, which means we need to find the pdf of $\hat{\theta}$. Well, $\hat{\theta} = \max\{-X_1, X_n/2\}$:

$$\begin{aligned} F_{\hat{\theta}}(\tau) &= P(\hat{\theta} \leq \tau) \\ &= \prod_{i=1}^n P(x_i \in [-\tau, 2\tau]) = \frac{3^n \tau^n}{3^n \theta^n} = \frac{\tau^n}{\theta^n}. \end{aligned} \quad (7.235)$$

So, $f_{\hat{\theta}}(\tau) = (n/\theta)(\tau/\theta)^{n-1}$, with $0 \leq \hat{\theta} \leq \theta$, so

$$E[\hat{\theta}] = \int_0^\theta \tau (n/\theta)(\tau/\theta)^{n-1} d\tau = \frac{n}{1+n}\theta. \quad (7.236)$$

So, $(n+1)\hat{\theta}/n$ is the unique MVUE for θ .

□

7.5.1

7.5.1. Write the pdf

$$f(x; \theta) = \frac{1}{6\theta^4} x^3 e^{-x/\theta}, \quad 0 < x < \infty, \quad 0 < \theta < \infty,$$

zero elsewhere, in the exponential form. If X_1, X_2, \dots, X_n is a random sample from this distribution, find a complete sufficient statistic Y_1 for θ and the unique function $\varphi(Y_1)$ of this statistic that is the MVUE of θ . Is $\varphi(Y_1)$ itself a complete sufficient statistic?

Solution:

$$\frac{1}{6\theta^4} x^3 e^{-x/\theta} = e^{-x/\theta + \ln x^3 - \ln(6\theta^4)} \quad (7.237)$$

We notice that $X_i \sim \Gamma(4, \theta)$. The joint pdf of the X_i 's is

$$\begin{aligned} & \exp \left[\frac{-1}{\theta} \sum x_i + \ln \prod x_i^3 - n \ln(6\theta^4) \right] \\ &= \exp \left[\frac{-1}{\theta} \sum x_i - n \ln(6\theta^4) \right] \exp \left[\ln \prod x_i^3 \right] \end{aligned} \quad (7.238)$$

The factorization theorem tells us that $Y_1 = \sum_{i=1}^n x_i$ is a sufficient statistic. Since $K(x) = x$, theorem 7.5.2. says Y_1 is also complete. So, Y_1 is complete sufficient. Now,

$$E[Y_1] = E \left[\sum X_i \right] = 4n\theta,$$

so $\varphi(Y_1) = (1/4n)Y_1$ is a function of the complete sufficient Y_1 and is an unbiased estimator of θ . This means $(1/4n) \sum X_i$ is a MVUE of θ . Since $\varphi(Y_1)$ is just a rescaled version of Y_1 (hence is one-to-one), $\varphi(Y_1)$ is also a complete sufficient statistic. \square

7.5.3

7.5.3. Let X_1, X_2, \dots, X_n denote a random sample of size n from a distribution with pdf $f(x; \theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, and $\theta > 0$.

- (a) Show that the *geometric mean* $(X_1 X_2 \cdots X_n)^{1/n}$ of the sample is a complete sufficient statistic for θ .
- (b) Find the maximum likelihood estimator of θ , and observe that it is a function of this geometric mean.

Solution:

1. We rewrite the pdf as

$$f(x_i; \theta) = \exp [\ln \theta + (\theta - 1) \ln x_i]. \quad (7.239)$$

We check that $f(x_i; \theta)$ belongs to the regular exponential family. So, the with the joint pdf given by

$$\prod_{i=1}^n f(x_i; \theta) = \exp \left[n \ln \theta + (\theta - 1) \sum_{i=1}^n \ln x_i \right] \quad (7.240)$$

Let $K(x) = \ln x$, then Theorem 7.5.2 says that $Y_1 = \sum_{i=1}^n \ln x_i$ is complete sufficient. Now, consider the one-to-one transformation $\varphi(Y_1) = \exp [Y_1/n]$ which gives the complete sufficient statistic $[\prod x_i]^{1/n}$. So, the geometric mean is a complete sufficient statistic of θ .

- 2.

$$\begin{aligned} \partial_\theta \left[n \ln \theta + (\theta - 1) \sum_{i=1}^n \ln x_i \right] &= 0 \\ \implies \hat{\theta} &= -\frac{n}{\sum_{i=1}^n \ln X_i} = -\frac{1}{\ln [\prod_{i=1}^n X_i]^{1/n}} \end{aligned} \quad (7.241)$$

where of course the denominator is natural log of the geometric mean.

□

7.8 Problem set 8

7.6.1

7.6.1. Let X_1, X_2, \dots, X_n denote a random sample from a distribution that is $N(\theta, 1)$, $-\infty < \theta < \infty$. Find the MVUE of θ^2 .

Hint: First determine $E(\bar{X}^2)$.

Solution:

7.6.9

7.6.1. Let X_1, X_2, \dots, X_n denote a random sample from a distribution that is $N(\theta, 1)$, $-\infty < \theta < \infty$. Find the MVUE of θ^2 .

Hint: First determine $E(\bar{X}^2)$.

Solution:

7.8.1

7.8.1. Let X_1, X_2, \dots, X_n be a random sample from each of the following distributions involving the parameter θ . In each case find the mle of θ and show that it is a sufficient statistic for θ and hence a minimal sufficient statistic.

- (a) $b(1, \theta)$, where $0 \leq \theta \leq 1$.
- (b) Poisson with mean $\theta > 0$.
- (c) Gamma with $\alpha = 3$ and $\beta = \theta > 0$.
- (d) $N(\theta, 1)$, where $-\infty < \theta < \infty$.
- (e) $N(0, \theta)$, where $0 < \theta < \infty$.

Solution:

7.8.3

7.8.3. Let $Y_1 < Y_2 < \dots < Y_n$ be the order statistics of a random sample of size n from a distribution with pdf

$$f(x; \theta_1, \theta_2) = \left(\frac{1}{\theta_2} \right) e^{-(x-\theta_1)/\theta_2} I_{(\theta_1, \infty)}(x),$$

where $-\infty < \theta_1 < \infty$ and $0 < \theta_2 < \infty$. Find the joint minimal sufficient statistics for θ_1 and θ_2 .

Solution:

7.8.5

7.8.5. In Example 7.8.5, a scale model was presented and scale invariance was defined. Using the notation of this example, show that the following statistics are scale-invariant:

- (a) $X_1^2 / \sum_1^n X_i^2$.
- (b) $\min\{X_i\} / \max\{X_i\}$.

Solution:

7.9.4

7.9.4. Let X and Y be random variables such that $E(X^k)$ and $E(Y^k) \neq 0$ exist for $k = 1, 2, 3, \dots$. If the ratio X/Y and its denominator Y are independent, prove that $E[(X/Y)^k] = E(X^k)/E(Y^k)$, $k = 1, 2, 3, \dots$.

Hint: Write $E(X^k) = E[Y^k(X/Y)^k]$.

Solution:

7.9.13

7.9.13. Suppose X_1, X_2, \dots, X_n is a random sample from a distribution with pdf $f(x; \theta) = (1/2)\theta^3 x^2 e^{-\theta x}$, $0 < x < \infty$, zero elsewhere, where $0 < \theta < \infty$:

- (a) Find the mle, $\hat{\theta}$, of θ . Is $\hat{\theta}$ unbiased?

Hint: Find the pdf of $Y = \sum_1^n X_i$ and then compute $E(\hat{\theta})$.

- (b) Argue that Y is a complete sufficient statistic for θ .

- (c) Find the MVUE of θ .

- (d) Show that X_1/Y and Y are independent.

- (e) What is the distribution of X_1/Y ?

Solution:

8.1.2

8.1.2. Let the random variable X have the pdf $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, zero elsewhere. Consider the simple hypothesis $H_0 : \theta = \theta' = 2$ and the alternative hypothesis $H_1 : \theta = \theta'' = 4$. Let X_1, X_2 denote a random sample of size 2 from this distribution. Show that the best test of H_0 against H_1 may be carried out by use of the statistic $X_1 + X_2$.

Solution:

8.1.5

8.1.5. If X_1, X_2, \dots, X_n is a random sample from a distribution having pdf of the form $f(x; \theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, show that a best critical region for testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$ is $C = \{(x_1, x_2, \dots, x_n) : c \leq \prod_{i=1}^n x_i\}$.

Solution:

8.1.8

8.1.8. If X_1, X_2, \dots, X_n is a random sample from a beta distribution with parameters $\alpha = \beta = \theta > 0$, find a best critical region for testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$.

Solution:

8.2.2

8.2.2. Let X have a pdf of the form $f(x; \theta) = 1/\theta$, $0 < x < \theta$, zero elsewhere. Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample of size 4 from this distribution. Let the observed value of Y_4 be y_4 . We reject $H_0 : \theta = 1$ and accept $H_1 : \theta \neq 1$ if either $y_4 \leq \frac{1}{2}$ or $y_4 > 1$. Find the power function $\gamma(\theta)$, $0 < \theta$, of the test.

Solution:

8.2.7

8.2.7. Let X_1, X_2, \dots, X_{25} denote a random sample of size 25 from a normal distribution $N(\theta, 100)$. Find a uniformly most powerful critical region of size $\alpha = 0.10$ for testing $H_0 : \theta = 75$ against $H_1 : \theta > 75$.

Solution:

8.2.11

8.2.11. Let X_1, X_2, \dots, X_n be a random sample from a distribution with pdf $f(x; \theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, where $\theta > 0$. Show the likelihood has mnr in the statistic $\prod_{i=1}^n X_i$. Use this to determine the UMP test for $H_0 : \theta = \theta'$ against $H_1 : \theta < \theta'$, for fixed $\theta' > 0$.

Solution: