# Mathematical Tripos: Part IB Solutions

Lucy Li

July 13, 2022

## 1 Question 2

In a genetics experiment, a sample of n individuals was found to include a,b,c of the three possible genotypes GG, Gg, gg respectively. The population frequency of a gene of type G is $\theta/(\theta+1)$, where  is unknown, and it is assumed that the individuals are unrelated and that two genes in a single individual are independent. Show what $\theta$ is proportional to and calculate the MLE.

**Solution**

The likelihood function is given as: $\frac{\theta}{\theta+1}^{2a} \cdot \frac{2\theta}{(\theta+1)^2}^{b} \cdot \frac{1}{\theta+1}^{2c}$

The log likelihood function is given as:

$(2a+b)\log\theta - (2a+2b+2c)\log(\theta+1) = 0$

Solving this, we get:

$\theta = \frac{2a+b}{b+2c}$

## 2 Question 3

For some $n \geq 2$, suppose that $X_1, \dots, X_n$ are iid random variables uniformly distributed on $[\theta, 2\theta]$ for some $\theta > 0$. Show that $\tilde{\theta} = \frac{2}{3}X_1$ is an unbiased estimator of $\theta$. Show that $T(\mathbf{X}) = (\min_i X_i, \max_i X_i)$ is a minimal sufficient statistic for $\theta$. Use the Rao-Blackwell theorem to find an unbiased estimator $\hat{\theta}$ of $\theta$ which is a function of $T$ and whose variance is strictly smaller than the variance of $\tilde{\theta}$ for all $\theta > 0$.

**Solution**

The likelihood function is given by:

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\theta} 1[min(X_i) > \theta, max(Xi) < 2\theta]$$

Hence, a sufficient statistic is $T(\mathbf{X}) = (\min_i X_i, \max_i X_i)$

$\tilde{\theta} = \frac{2}{3}X_1$ is unbiased because $E[X_1] = \frac{3\theta}{2}$

An unbiased estimator for $\theta$ is:

$$\begin{aligned}\widehat{\theta}^* &= E\left(\frac{2}{3}X_1 \mid \min_i X_i = a, \max_i X_i = b\right) \\ &= \frac{2a}{3n} + \frac{2b}{3n} + \frac{n-2}{n}\frac{2}{3}\frac{a+b}{2} = \frac{a+b}{3}\end{aligned}$$

## 3 Question 7

(a) Let $X_1, \dots X_n$ be iid with $X_i \sim U[0, \theta]$. Find the maximum likelihood estimator $\hat{\theta}$ of $\theta$. Show that the distribution of $R(\mathbf{X}, \theta) = \hat{\theta}/\theta$ does not depend on $\theta$, and use $R(\mathbf{X}, \theta)$ to find a $100(1-\alpha)\%$ confidence interval for $\theta$ for $0 < \alpha < 1$.

(a) **Solution** Please refer to the section 7. As we know, $\hat{\theta} = max(X_i)$.

Substituting the MLE into the CDF, we get:

$$P(x_n \leq x) = P(x_1 \leq x, x_2 \leq x, \dots, x_n \leq x)$$

$$= (F(x))^n$$

Differentiating the above, we get the pdf, which is expressed as

$$f_{\hat{\theta}} = n(F(x))^{n-1} f(x)$$

We know that the CDF of a uniform distribution is given as $F(x) = \frac{1}{\theta}x$, the pdf is given as $f(x) = \frac{1}{\theta}$, substituting those in, we get:

$$f_{\hat{\theta}/\theta} = f_{\hat{\theta}}(\theta x) \cdot \theta = nx^{n-1}, \quad (0 < x < 1)$$

, this does not depend on $\theta$.

In order to find the confidence interval, we can set (for the upper bound):

$$\int_0^u nx^{n-1} = (1 - \alpha/2)$$

In order to find the confidence interval, we can set (for the lower bound):

$$\int_0^l nx^{n-1} = (\alpha/2)$$

(b) The lengths (in minutes) of calls to a call centre may be modelled as iid exponentially distributed random variables, and n such call lengths are observed. The original sample is lost, but the data manager has noted down n and t where t is the total length of the n calls in minutes. Derive a 95 percent confidence interval for the probability that a call is longer than 2 minutes if n = 50 and t = 105.3.

(b) **Solution** First of all, we can try to write out the cdf of the exponential distribution.

$$f_X(x) = \int_2^\infty \lambda e^{-\lambda x} dx$$

$$= [1 - e^{-\lambda x}]_2^\infty = e^{-2\lambda}$$

We don't actually know what is the distribution of the $\lambda$ in the expression.

However, what we know is that that the the distribution of the calls is modelled by $Gamma(50, \lambda)$, hence, the distribution of $\lambda t$ may be expressed as $Gamma(50, 1)$. Then, we can just read off from the $Gamma(50, 1)$ table for what the confidence interval is.

# 4 Question 8

Suppose that $X_1 \sim N(\theta_1, 1)$ and $X_2 \sim N(\theta_2, 1)$ independently, where $\theta_1$ and $\theta_2$ are unknown. Show that $(\theta_1 - X_1)^2 + (\theta_2 - X_2)^2$ has a $\chi_2^2$ distribution and that this is the same as Exponential $\left(\frac{1}{2}\right)$, i.e., the exponential distribution with mean 2 . Show that both the square $S$ and circle $C$ in $R^2$, given by

$$S = \{(\theta_1, \theta_2) : |\theta_1 - X_1| \leq 2.236; |\theta_2 - X_2| \leq 2.236\}$$
$$C = \left\{(\theta_1, \theta_2) : (\theta_1 - X_1)^2 + (\theta_2 - X_2)^2 \leq 5.991\right\}$$

are 95% confidence regions for $(\theta_1, \theta_2)$. Hint: $\Phi(2.236) = (1 + \sqrt{.95})/2$, where $\Phi$ is the distribution function of $N(0, 1)$. What might be a sensible criterion for choosing between $S$ and $C$ ?

We know that, if $Z_1, \ldots, Z_k$ are independent, standard normal random variables, then the sum of their squares,

$$Q = \sum_{i=1}^k Z_i^2$$

is distributed according to the chi-squared distribution with $k$ degrees of freedom. This is usually denoted as $Q \sim \chi^2(k)$ or $Q \sim \chi_k^2$

# 5 Question 9

Suppose that the number of defects on a roll of magnetic recording tape is modelled with a Poisson distribution for which the mean $\lambda$ is known to be either 1 or 1.5. Suppose the prior mass function for $\lambda$ is

$$\pi_\lambda(1) = 0.4, \quad \pi_\lambda(1.5) = 0.6.$$

A random sample of five rolls of tape has $\mathbf{x} = (3, 1, 4, 6, 2)$ defects respectively. Show that the posterior distribution for $\lambda$ given $\mathbf{x}$ is

$$\pi_{\lambda|\mathbf{X}}(1 \mid \mathbf{x}) = 0.012, \quad \pi_{\lambda|\mathbf{X}}(1.5 \mid \mathbf{x}) = 0.988$$

# 6 Question 10

Suppose $X_1, \ldots, X_n$ are iid with (conditional) probability density function $f(x \mid \theta) = \theta x^{\theta-1}$ for $0 < x < 1$ (and is zero otherwise), for some $\theta > 0$. Suppose that the prior for $\theta$ is Gamma$(\alpha, \beta), \alpha > 0, \beta > 0$. Find the posterior distribution of $\theta$ given $\mathbf{X} = (X_1, \ldots, X_n)$ and the Bayesian estimator of $\theta$ under quadratic loss. +11 For some $n \geq 3$, let $\epsilon_1, \ldots, \epsilon_n$ be iid with $\epsilon_i \sim N(0, 1)$. Set $X_1 = \epsilon_1$ and $X_i = \theta X_{i-1} + \left(1 - \theta^2\right)^{1/2} \epsilon_i$ for $i = 2, \ldots, n$ and some $\theta \in (-1, 1)$. Find a sufficient statistic for $\theta$ that takes values in a subset of $R^3$.

# 7 MLE for different distributions

MLE for different distributions

**Exponential distribution**

$$f_X(x) = \lambda e^{-\lambda x}$$

likelihood function is given as :

$$L(x_i) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

log-likelihood function is given as :

$$n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i \Rightarrow \frac{d \log(L(x_i))}{d\lambda} = \frac{n}{\lambda} - \sum x_i$$

$$\therefore \boxed{\lambda = \frac{\sum x_i}{n}}$$

**Normal distribution**

$$f_X(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

likelihood function is given as :

$$\mathcal{L}(x_i) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

log-likelihood function is given as :

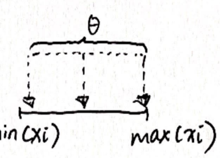
$$n\,\frac{1}{2}\log(2\pi) + n\,\frac{1}{2}\log(\sigma^2) + \left[+\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

taking partial derivative wrt. $\sigma^2$:

$$\frac{\partial \log(\mathcal{L}(x_i))}{\partial(\sigma^2)} = \frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 \longrightarrow \boxed{\hat{\sigma}^2 = \frac{\sum(x_i-\mu)^2}{n}}$$

$$\frac{\partial \log(\mathcal{L}(x_i))}{\partial(\mu)} = \frac{n\mu}{\sigma^2} - \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i \rightarrow \boxed{\hat{\mu} = \bar{x}}$$

**Uniform Distribution**




likelihood $\Rightarrow \prod_{i=1}^{n} \frac{1}{\theta}(1(0<x_i<\theta)) = \frac{1}{\theta^n}(1_{(0<x_i<\theta)})$

log-likelihood function $\Rightarrow -n\log(\theta)(1(0<x_i<\theta))$

$$= -n\log(\theta)(1_{[\min(x_i)>0\,;\ \max(x_i)<\theta]}) \Rightarrow \boxed{\text{setting } \hat{\theta} = \max(x_i) \text{ maximizes MLE}}$$