PRIFYSGOL ABERYSTWYTH UNIVERSITY

## ADRAN MATHEMATEG / DEPARTMENT OF MATHEMATICS

## ARHOLIADAU SEMESTER 2 / SEMESTER 2 EXAMINATIONS

## MAI / MAY 2020

## MA36010 - Comparative Statistical Inference

The questions on this paper are written in English.

If you have questions about the paper during the exam, contact the module co-ordinator, Dr Jukka Kiukas, on jek20@aber.ac.uk.

You should write out solutions to the paper and upload them to blackboard as a single PDF file.

**Amser a ganiateir - 3 awr**
*Mae'n rhaid cyflwyno eich atebion erbyn 12:30 (amser y DU).*

- Gellir rhoi cynnig ar bob cwestiwn.
- Rhoddir mwy o ystyriaeth i berfformiad yn rhan B wrth bennu marc dosbarth cyntaf.
- Mae tablau ystadegol ar gael ar Blackboard.
- Mae taflen fformiwlâu wedi cael ei hatodi.
- Mae modd i fyfyrwyr gyflwyno atebion i'r papur hwn naill ai yn y Gymraeg neu'r Saesneg.

**Time allowed - 3 hours**
*Submission must be completed by 12:30 (UK time).*

- All questions may be attempted.
- Performance in section B will be given greater consideration in assigning a first class mark.
- Statistical tables are available on Blackboard.
- A formula sheet is attached.
- Students may submit answers to this paper in either Welsh or English.

 

## Section A

**1**. The random variable $K$ follows the Binomial distribution $\mathrm{Bin}(14, p)$. A single observation of $K$ was taken, with outcome $k = 5$.

(a) Write down the likelihood and the score function, and hence find the Maximum Likelihood Estimate (MLE) of $p$. [6 marks]

(b) A Bayesian statistician's prior beliefs for $p$ are that it is equally likely to take any value in $(0, 1)$. Show that their posterior density for $p$ is a member of the Beta family of distributions, and state the relevant parameters. [4 marks]

(c) Find the posterior mean, and confirm that it lies between the prior mean and the MLE. [3 marks]

**2**. The likelihood of an unknown parameter, $\alpha > 0$, has been determined from a random sample of $n$ observations, $X_1, \ldots, X_n$, as

$$L(\alpha) = \alpha^{-n} \prod_{i=1}^{n} (1 + x_i)^{-1-1/\alpha}.$$

(a) Find the score function for $\alpha$. [2 marks]

(b) Show that $\alpha$ has a Minimum Variance Bound Unbiased Estimator (MVBUE) $T$, and find it. [3 marks]

(c) Deduce the expected value of $T$, and hence the Fisher Information for $\alpha$. [4 marks]

(d) Find the Fisher Information for $\phi = \ln \alpha$. Does $\phi$ have an MVBUE? Justify your answer. [4 marks]

(e) Write the likelihood in terms of the MVBUE $T$ found in (b), and explain why $T$ is a sufficient statistic for $\alpha$. [3 marks]

**3**. A proportion $p$ of components manufactured by a production line are defective. Six randomly selected components were tested, and three of them were found defective.

(a) Show that the likelihood is proportional to $p^3(1 - p)^3$. [2 marks]

(b) Suppose that the Bayesian prior for $p$ is taken to be the Beta distribution with mean 2/5 and standard deviation 1/5.

(i) Find the parameters of the prior. [3 marks]

(ii) Identify the posterior distribution as a member of the Beta family. [3 marks]

(iii) Find the posterior mean and variance of $p$. [2 marks]

(c) Assume instead prior ignorance with the Haldane prior $f(p) \propto p^{-1}(1 - p)^{-1}$.

(i) Determine the probability density function (pdf) of the posterior distribution, including the normalisation constant. [3 marks]

(ii) Calculate the posterior predictive probability that the next component from the production line is defective. [4 marks]

**4**. Accidents on a certain road occur randomly at an average rate of $\lambda$ per month, so that the number of accidents in any fixed time period follows a Poisson distribution. The rate had been $\lambda = 0.5$ before the speed limit was lowered. Afterwards the number, $N$, of accidents was recorded in a ten month period, to find out if the rate had decreased.

(a) Consider the test of hypotheses $H_0 : \lambda = 0.5$ versus $H_1 : \lambda = 0.1$, using the critical region $\{N \leq 1\}$. Find (i) the size, and (ii) the power of the test. [5 marks]

(b) Consider now the composite hypothesis $H_1 : \lambda < 0.5$, again with the critical region $\{N \leq 1\}$. Using the pmf of the Poisson distribution, find the formula of the power function $\Omega(\lambda)$. [3 marks]

**5**. The statistic $X$ has a sampling distribution whose cumulative distribution function (cdf) is $F(x) = \frac{2}{\pi} \arctan(\theta x^2)$, $x > 0$, where $\theta > 0$ is an unknown parameter.

(a) Show that $T = \theta X^2$ is a pivotal function for $\theta$. [4 marks]

(b) Calculate the confidence coefficient of the interval $(0.1/X^2, 10/X^2)$. [3 marks]

(c) Find, in terms of $X$, a 90% confidence lower bound for $\theta$. [3 marks]

**6**. You may assume, that given a random sample $\mathbf{X} = (X_1, \ldots, X_n)$ from a $N(\mu, \sigma^2)$ distribution, and a $N(a, b^2)$ prior for $\mu$, the posterior distribution is Normal with

$$\mathbb{E}(\mu|\mathbf{x}) = \frac{n\overline{x}/\sigma^2 + a/b^2}{n/\sigma^2 + 1/b^2}, \qquad \mathrm{Var}(\mu|\mathbf{x}) = \left(n/\sigma^2 + 1/b^2\right)^{-1},$$

where $\overline{x}$ is the sample mean.

The weight of a product from a manufacturing line is Normally distributed with unknown mean $\mu$ (kg) and standard deviation 3 kg. A Bayesian analyst places the Normal distribution with mean 30 kg and standard deviation 6 kg as the prior for $\mu$, and observes the weights of three products to be 31.2, 33.1, and 29.3 kg.

(a) Identify the posterior distribution of $\mu$. [3 marks]

(b) Calculate the posterior probability that $\mu$ exceeds the prior mean. [3 marks]

## Section B

**7**. Let $\theta > 0$ be a parameter, and $Y_1, \ldots, Y_n$ a random sample from the distribution whose pdf is

$$f(y) = \theta(2y+1)\, e^{-y(y+1)\theta}, \quad y > 0.$$

(a) Write down the likelihood for $\theta$, and find a sufficient statistic $S$ for $\theta$. [3 marks]

(b) (i) Find the MLE of $\theta$, in terms of $S$. [4 marks]

(ii) Find the Fisher information, and hence the MVB of $\theta$. [2 marks]

(iii) Is the MLE of $\theta$ an MVBUE? Explain your reasoning. [2 marks]

(c) Consider now the Bayesian approach, using the Jeffreys prior.

(i) Deduce the kernel of the posterior distribution, and identify it as a member of a standard family of distributions. [3 marks]

(ii) Find the posterior mean, in terms of $S$. What do you observe when comparing it with the result in part (b)(i)? [2 marks]

**8**. Consider a sequence of independent Bernoulli trials with success probability $p$. Let $N$ be the number of trials up to (and including) the $r$th success. Then $N \sim \mathrm{Pasc}(r, p)$. Suppose the value $N = n$ was observed.

(a) Write down the likelihood of $p$. [2 marks]

(b) Given a Beta prior $\beta(a, b)$, show that the posterior distribution belongs to the Beta family, and hence derive the update rules for the parameters of the Beta distribution. [3 marks]

(c) A biased coin with probability $p$ for "heads" was tossed repeatedly until getting "heads" three times. Given that this took 8 tosses, and assuming the uniform prior for $p$ (as in Q1b), find the 95% credible posterior upper bound for $p$. [7 marks]

**9**. A random sample, $Y_1, \ldots, Y_{25}$, is drawn from a Normal distribution, $N(\mu, 10)$, with unknown mean $\mu$. Consider the statistics $\overline{Y} = \frac{1}{25}\sum_{i=1}^{25} Y_i$ and $S = \frac{1}{25}\sum_{i=1}^{25}(Y_i - \overline{Y})^2$.

(a) Using the identity $\sum_{i=1}^{n}(Y_i - \mu)^2 = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 + n(\overline{Y} - \mu)^2$, write down the likelihood of $\mu$ in terms of the two statistics $\overline{Y}$ and $S$, and hence show that the value of the likelihood at its maximum is $(20\pi)^{-25/2} e^{-5S/4}$. [3 marks]

(b) Find an 8% relative likelihood interval for $\mu$. [4 marks]

(c) A Bayesian statistician, who has a uniform prior for $\mu$ (i.e. the prior pdf is proportional to a constant) interprets the interval in part (b) as a $100(1-\alpha)\%$ credible highest density posterior interval for $\mu$. Find $\alpha$. [4 marks]

**10**. A random sample, $X_1, \ldots, X_{20}$, is drawn from the Exponential distribution, $\mathrm{Expon}(\theta)$, where $\theta > 0$ is an unknown parameter. Consider the test of two simple hypotheses, $H_0 : \theta = 8$ versus $H_1 : \theta = 15$, using the sample sum $S = \sum_{i=1}^{20} X_i$.

(a) Quote the distribution of $S$ (from the back page of Statistical Tables), and hence show that $2\theta S \sim \chi^2_{(40)}$. [2 marks]

(b) Write the likelihood in terms of $S$, and hence show that the likelihood ratio is proportional to $e^{7S}$. [3 marks]

(c) Show that the critical region of the most powerful test of size 0.05 is $\{S < k\}$, where $k$ is a constant to be determined using the chi-squared table. [6 marks]

## Standard Distributions and their Recognition

| Name | Notation | Range | Kernel | Constant | Mean | Mode | Variance | Comments |
|---|---|---|---|---|---|---|---|---|
| Beta | $\beta(a,b)$ | $0<z<1$ | $z^{a-1}(1\text{-}z)^{b-1}$ | $\frac{1}{B(a,b)}=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ | $\frac{a}{(a+b)}$ | $\frac{a-1}{a+b-2}$ | $\frac{ab}{(a+b)^2(a+b+1)}$ | $1-Z\sim\beta(b,a)$ <br> $\frac{bZ}{a(1-Z)}\sim F(2a,2b)$ |
| Gamma | $\mathrm{Ga}(a,b)$ | $z>0$ | $z^{a-1}e^{-bz}$ | $\frac{b^a}{\Gamma(a)}$ | $\frac{a}{b}$ | $\frac{(a-1)}{b}$ | $\frac{a}{b^2}$ | $cZ\sim\mathrm{Ga}(a,b/c)$ <br> $2bZ\sim\chi^2(2a)$ |
| Chi-square | $\chi^2(k)$ | $z>0$ | $z^{(k-2)/2}e^{-z/2}$ | $\frac{1}{2^{k/2}\Gamma(k/2)}$ | $k$ | $(k-2)$ | $2k$ | $\chi^2(k)\equiv\mathrm{Ga}(k/2,1/2)$ |
| Inverse chi-square | $\chi^{-2}(k)$ | $z>0$ | $z^{-(k-2)/2}e^{-1/(2z)}$ | $\frac{1}{2^{k/2}\Gamma(k/2)}$ | $\frac{1}{(k-2)}$ | $\frac{1}{(k+2)}$ | $\frac{2}{(k-2)^2(k-4)}$ | $Z^{-1}\sim\chi^2(k)$ |
| Inverse chi | $\chi^{-1}(k)$ | $z>0$ | $z^{-k-1}\exp\{-1/(2z^2)\}$ | $\frac{1}{2^{\frac{1}{2}k-1}\Gamma(k/2)}$ | $\frac{\Gamma\left(\frac{k-1}{2}\right)}{2^{\frac{1}{2}}\Gamma\left(\frac{k}{2}\right)}$ | $\frac{1}{(k+1)^{1/2}}$ | $\frac{1}{(k-2)}-(\text{mean})^2$ | $Z^{-2}\sim\chi^2(k)$ |
| Normal | $N(a,b^2)$ | all $z$ | $\exp[-½\{(z-a)/b\}^2]$ | $\{2\pi b^2\}^{-1/2}$ | $a$ | $a$ | $b^2$ | $\frac{(Z-a)}{b}\sim N(0,1)$ |
| Standard Normal | $N(0,1)$ | all $z$ | $\exp\{-z^2/2\}$ | $\{2\pi\}^{-1/2}$ | 0 | 0 | 1 | |
| (Student's) t | $t(k)$ | all $z$ | $1\Big/\left\{1+\frac{z^2}{k}\right\}^{(k+1)/2}$ | $\frac{\Gamma\left(\frac{k+1}{2}\right)}{(k\pi)^{\frac{1}{2}}\Gamma\left(\frac{k}{2}\right)}$ | 0 | 0 | $\frac{k}{(k-2)}$ | |
| (Fisher's) F | $F(a,b)$ | $z>0$ | $\frac{z^{\frac{1}{2}a-1}}{\left(z+\frac{b}{a}\right)^{\frac{1}{2}(a+b)}}$ | $\left(\frac{b}{a}\right)^{b/2}\frac{\Gamma\left(\frac{a+b}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{b}{2}\right)}$ | $\frac{b}{(b-2)}$ | $\frac{b(a-2)}{a(b+2)}$ | $\frac{2b^2(a+b-2)}{a(b-2)^2(b-4)}$ | $Z^{-1}\sim F(b,a)$ <br> Mean>1>Mode with Median between Mode and Mean <br> Median = 1 whenever $a=b$ |

**Useful fact**: If $Z$ has density $f(z)$ then $W=cZ+d$ has density $g(w)=f(z)/|c|=\frac{f((w-d)/c)}{|c|}$

**Useful identities**: $\sum a_i(z-b_i)^2=\sum a_i(b_i-\bar{b})^2+(z-\bar{b})^2\sum a_i$ where $\bar{b}=\left(\sum a_ib_i\right)/\left(\sum a_i\right)$, $a(z-c)^2+b(z-d)^2=(a+b)\left(z-\frac{ac+bd}{a+b}\right)^2+\frac{ab}{a+b}(c-d)^2$