THE ROYAL STATISTICAL SOCIETY

2010 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

MODULE 2

STATISTICAL INFERENCE

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

Note. In accordance with the convention used in the Society's examination papers, the notation log denotes logarithm to base *e*. Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Graduate Diploma, Module 2, 2010. Question 1

(i) The joint density is

$$f\left(t_1, ..., t_n\right) = \prod_{i=1}^{n}\left(\mu^{-1}e^{-t_i/\mu}\right) = \underset{\substack{\uparrow \\ g\left(\Sigma t_i;\mu\right)}}{\mu^{-n}e^{-\frac{1}{\mu}\Sigma t_i}} \times \underset{\substack{\uparrow \\ h(t_1,\ \ldots,\ t_n)}}{1}$$

Since the joint density is the product of a factor (simply 1) not involving the parameter μ and a factor only dependent on the observations $t_1, \ldots, t_n$ through Σt_i, it follows by the factorisation theorem that $Y = \Sigma t_i$ is sufficient for μ.

(ii) $$E\left(T_i\right) = \int_0^\infty t.\frac{1}{\mu}e^{-t/\mu}dt = \left[-te^{-t/\mu}\right]_0^\infty + \int_0^\infty e^{-t/\mu}dt = [0] + \left[-\mu e^{-t/\mu}\right]_0^\infty = \mu .$$

$$\therefore E\left(\hat{\mu}_n\right) = \frac{1}{n}\sum_{i=1}^{n}E\left(T_i\right) = \frac{1}{n}\left(n\mu\right) = \mu \qquad \text{(for all } n\text{).}$$

So $\hat{\mu}_n$ is an unbiased estimator of μ.

$$E\left(T_i^2\right) = \int_0^\infty t^2.\frac{1}{\mu}e^{-t/\mu}dt = \left[-t^2e^{-t/\mu}\right]_0^\infty + \int_0^\infty 2te^{-t/\mu}dt$$

$$= [0] + 2\mu\int_0^\infty \frac{t}{\mu}e^{-t/\mu}dt = 2\mu E\left(T_i\right) = 2\mu^2 .$$

$\therefore \text{Var}(T_i) = E(T_i^2) - \{E(T_i)\}^2 = 2\mu^2 - \mu^2 = \mu^2.$

$$\therefore \text{Var}\left(\hat{\mu}_n\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}\left(T_i\right) \qquad \text{(note that the } T_i \text{ are independent)}$$

$$= \frac{1}{n^2}\left(n\mu^2\right) = \frac{\mu^2}{n} .$$

(iii) $\hat{\mu}_n$ is an unbiased estimator of μ (for all n) and its variance tends to zero as $n \to \infty$. So $\hat{\mu}_n$ is a consistent estimator of μ.

Solution continued on next page

Part (iv)

We require

$$E\left(\sqrt{T_i}\right) = \int_0^\infty t^{1/2}.\frac{1}{\mu}e^{-t/\mu}dt .$$

This can be evaluated by noting that $\dfrac{t^{1/2}e^{-t/\mu}}{\mu^{3/2}\Gamma\left(\frac{3}{2}\right)}$ is the pdf of a gamma distribution, or explicitly as follows.

First, use the substitution $u = t^{1/2}$, with which the integral becomes

$$\int_0^\infty \frac{u}{\mu}e^{-u^2/\mu}2u\,du$$ (now use symmetry about 0)

$$= \frac{1}{\mu}\int_{-\infty}^\infty u^2 e^{-u^2/\mu}du$$ (now create the pdf of a Normal distribution)

$$= \frac{\sqrt{2\pi\mu/2}}{\mu}\int_{-\infty}^\infty \frac{u^2}{\sqrt{2\pi\mu/2}}e^{-u^2/\mu}du$$

The integrand is u^2 times the pdf of $N(0, \mu/2)$; thus the integral is the second moment (about zero) of that distribution; and this is the variance plus the square of the mean

$$= \sqrt{\frac{\pi}{\mu}} \times \frac{\mu}{2} = \tfrac{1}{2}\sqrt{\pi\mu}\,,$$ as required.

Therefore, noting that T_1 and T_2 are independent,

$$E(\tilde{\mu}) = \frac{4}{\pi}E\left(\sqrt{T_1}\right)E\left(\sqrt{T_2}\right) = \frac{4}{\pi}\frac{\pi\mu}{4} = \mu .$$ So $\tilde{\mu}$ is an unbiased estimator of μ.

Solution continued on next page

Part (v)

Again noting that T_1 and T_2 are independent,

$$E\left(\tilde{\mu}^2\right) = \frac{16}{\pi^2} E\left(T_1\right) E\left(T_2\right) = \frac{16\mu^2}{\pi^2}.$$

$$\therefore \text{Var}\left(\tilde{\mu}\right) = E\left(\tilde{\mu}^2\right) - \left\{E\left(\tilde{\mu}\right)\right\}^2 = \frac{16\mu^2}{\pi^2} - \mu^2 = \mu^2\left(\frac{16}{\pi^2} - 1\right) = 0.6211\mu^2.$$

So the relative efficiency of $\tilde{\mu}$ compared to $\hat{\mu}_2$ is

$$\frac{\text{Var}\left(\hat{\mu}_2\right)}{\text{Var}\left(\tilde{\mu}\right)} = \frac{\frac{\mu^2}{2}}{0.6211\mu^2} = 0.805 \quad \text{(or 80.5\%).}$$

(i) X_i has the truncated Poisson distribution with zero missing. Letting Y denote the full Poisson distribution, with mean λ, we have (for i = 1, 2, 3, …)

$$P(X_i = k) \;=\; P(Y = k \mid Y > 0) \;=\; \frac{P(Y = k)}{P(Y > 0)} \;=\; \frac{\dfrac{e^{-\lambda}\lambda^k}{k!}}{1 - e^{-\lambda}} \;=\; \frac{\lambda^k e^{-\lambda}}{k!\left(1 - e^{-\lambda}\right)} \;.$$

(ii) The likelihood is

$$L(\lambda) \;=\; \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!\left(1 - e^{-\lambda}\right)} \;=\; \frac{\lambda^{\Sigma x_i} e^{-n\lambda}}{\Pi x_i!\left(1 - e^{-\lambda}\right)^n} \;.$$

$$\therefore \log L(\lambda) = \Sigma x_i \log\lambda - n\lambda - n\log\left(1 - e^{-\lambda}\right) - \log\left(\Pi x_i!\right).$$

$$\therefore \frac{d\log L}{d\lambda} = \frac{\Sigma x_i}{\lambda} - n - \frac{ne^{-\lambda}}{1 - e^{-\lambda}} \;.$$

The maximum likelihood estimator $\hat{\lambda}$ therefore satisfies

$$\frac{\Sigma x_i}{\hat{\lambda}} - n - \frac{ne^{-\hat{\lambda}}}{1 - e^{-\hat{\lambda}}} = 0 \;.$$

(iii) We have

$$\frac{d^2 \log L}{d\lambda^2} = -\frac{\Sigma x_i}{\lambda^2} + \frac{ne^{-\lambda}}{1 - e^{-\lambda}} + \frac{ne^{-2\lambda}}{\left(1 - e^{-\lambda}\right)^2} = -\frac{\Sigma x_i}{\lambda^2} + \frac{ne^{-\lambda}}{\left(1 - e^{-\lambda}\right)^2} \;.$$

$$\therefore E\left(-\frac{d^2 \log L}{\lambda^2}\right) = \frac{\Sigma E(X_i)}{\lambda^2} - \frac{ne^{-\lambda}}{\left(1 - e^{-\lambda}\right)^2} \;.$$

Now, $E(X_i) = \dfrac{1}{\left(1 - e^{-\lambda}\right)} \displaystyle\sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!}$

The summation can be extended to include also k = 0, as this is a zero term; hence the summation is E(Poisson(λ))

$$= \frac{\lambda}{1 - e^{-\lambda}} \;.$$

$$\therefore E\left(-\frac{d^2 \log L}{\lambda^2}\right) = \frac{n\lambda}{\lambda^2\left(1-e^{-\lambda}\right)} - \frac{ne^{-\lambda}}{\left(1-e^{-\lambda}\right)^2} = \frac{n}{\lambda\left(1-e^{-\lambda}\right)^2}\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right).$$

Therefore the asymptotic variance of $\hat{\lambda}$ is

$$\frac{\lambda\left(1-e^{-\lambda}\right)^2}{n\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)}.$$

(iv) We use the Newton-Raphson method.

Iterations continue according to the scheme described below until convergence occurs.

$$\hat{\lambda}_1 = \hat{\lambda}_0 - \frac{\left.\dfrac{d \log L}{d\lambda}\right|_{\lambda=\hat{\lambda}_0}}{\left.\dfrac{d^2 \log L}{d\lambda^2}\right|_{\lambda=\hat{\lambda}_0}}.$$

A starting value is required. A reasonable initial estimate of λ is simply $\bar{x}$, as this would be the estimate if the Poisson distribution had not been truncated.

In the question, an initial estimate of 2.0 is given. Inserting this, n = 30 and $\Sigma X_i = 50$, we have

$$\left.\frac{d \log L}{d\lambda}\right|_{\lambda=\hat{\lambda}_0} = \frac{50}{2} - 30 - \frac{30e^{-2}}{1-e^{-2}} = -9.6956,$$

and

$$\left.\frac{d^2 \log L}{d\lambda^2}\right|_{\lambda=\hat{\lambda}_0} = -\frac{50}{4} + \frac{30e^{-2}}{\left(1-e^{-2}\right)^2} = -7.0696.$$

$$\therefore \hat{\lambda}_1 = 2.0 - \left(\frac{-9.6956}{-7.0696}\right) = 0.628(5).$$

Part (a)

(i) The likelihood is $f\left(\mathbf{x}\,|\,\theta\right)=\prod_{i=1}^{n} f\left(x_i\,|\,\theta\right)$.

Using the notation defined in the question, the loss when θ is estimated by $\hat{\theta}$ (a function of $X_1, X_2, \ldots, X_n$) is $\ell(\hat{\theta};\theta)$. Then the risk is

$$R\left(\theta\right)=E_\theta\left[\ell(\hat{\theta};\theta)\right]=\int_{\aleph}\ell(\hat{\theta};\theta)\,f\left(\mathbf{x}\,|\,\theta\right)d\mathbf{x}\,.$$

Note. $\int_{\aleph}\ldots\ldots d\mathbf{x}$ represents integration over the sample space.

Thus, as $\pi\left(\theta\right)$ is the prior density of θ, the Bayes risk of $\hat{\theta}$ is

$$r_\pi\left(\hat{\theta}\right)=E_\pi\left(R\left(\theta\right)\right)=\int_{\Theta}R\left(\theta\right)\pi\left(\theta\right)d\theta=\int_{\Theta}\int_{\aleph}\ell(\hat{\theta};\theta)\,f\left(\mathbf{x}\,|\,\theta\right)\pi\left(\theta\right)d\mathbf{x}\,d\theta\,.$$

Note. $\int_{\Theta}\ldots\ldots d\theta$ represents integration over the parameter space.

The Bayes estimator of θ is the value $\hat{\theta}$ that minimises $r_\pi\left(\hat{\theta}\right)$.

(ii) We note that $f\left(\mathbf{x}\,|\,\theta\right)\pi\left(\theta\right)=\pi\left(\theta\,|\,\mathbf{x}\right)f\left(\mathbf{x}\right)$

where $\pi\left(\theta\,|\,\mathbf{x}\right)$ is the posterior density of θ given $\mathbf{x}$

and $f\left(\mathbf{x}\right)$ is the marginal density of $\mathbf{x}$, $f\left(\mathbf{x}\right)=\int_{\Theta}f\left(\mathbf{x}\,|\,\theta\right)\pi\left(\theta\right)d\theta$.

Thus, from the definition above, we can write the Bayes risk as

$$r_\pi\left(\hat{\theta}\right)=\int_{\Theta}\int_{\aleph}\ell(\hat{\theta};\theta)\pi\left(\theta\,|\,\mathbf{x}\right)f\left(\mathbf{x}\right)d\mathbf{x}\,d\theta=\int_{\aleph}f\left(\mathbf{x}\right)\int_{\Theta}\ell(\hat{\theta};\theta)\pi\left(\theta\,|\,\mathbf{x}\right)d\theta\,d\mathbf{x}\,.$$

Thus to minimise $r_\pi\left(\hat{\theta}\right)$ at a given $\mathbf{x}$, we minimise $\int_{\Theta}\ell(\hat{\theta};\theta)\pi\left(\theta\,|\,\mathbf{x}\right)d\theta$ which is the posterior expected loss.

(iii) The posterior expected loss, given **x**, is $L(\hat{\theta}) = \int_{-\infty}^{\infty} |\hat{\theta} - \theta| \, \pi(\theta | \mathbf{x}) \, d\theta$. We have to find the $\hat{\theta}$ that minimises this.

First, we note that $\dfrac{d|y|}{dy} = \begin{cases} +1 & \text{if } y > 0 \\ -1 & \text{if } y < 0 \end{cases}$. Using this,

$$\frac{d\,L(\hat{\theta})}{d\hat{\theta}} = \int_{-\infty}^{\infty} \frac{d|\hat{\theta} - \theta|}{d\hat{\theta}} \pi(\theta | \mathbf{x}) \, d\theta = \int_{-\infty}^{\hat{\theta}} \pi(\theta | \mathbf{x}) d\theta - \int_{\hat{\theta}}^{\infty} \pi(\theta | \mathbf{x}) d\theta .$$

Setting this equal to zero gives that $\hat{\theta}$ is the median of $\pi(\theta | \mathbf{x})$.

Part (b)

We have that the posterior distribution of μ is N(100, 1) and that the loss function is

$$\pi(m_1, m_2; \mu) = \begin{cases} m_2 - m_1 & m_1 \le \mu \le m_2 \\ 5 + m_2 - m_1 & \text{otherwise} \end{cases} .$$

So the posterior expected loss is

$$L = (5 + m_2 - m_1) P(\mu < m_1) + (m_2 - m_1) P(m_1 \le \mu \le m_2) + (5 + m_2 - m_1) P(\mu > m_2)$$

$$= m_2 - m_1 + 5\{\Phi(m_1 - 100) + 1 - \Phi(m_2 - 100)\} \qquad \text{where } \Phi \text{ is the cdf of N(0, 1).}$$

This is to be minimised by choice of m_1 and m_2.

We have

$$\frac{\partial L}{\partial m_1} = -1 + 5\phi(m_1 - 100) \quad \text{and} \quad \frac{\partial L}{\partial m_2} = 1 - 5\phi(m_2 - 100)$$

where ϕ is the pdf of N(0, 1).

Taking $\phi(1.175)$ as (approximately) 0.2, as is given in the question, it follows that these derivatives are zero at $m_1 = 98.825$ and $m_2 = 101.175$ respectively. To check that these give a minimum, we note that

$$\frac{\partial^2 L}{\partial m_1^2} > 0, \quad \frac{\partial^2 L}{\partial m_2^2} > 0 \quad \text{and} \quad \frac{\partial^2 L}{\partial m_1 \partial m_2} = 0 .$$

L is therefore minimised at $m_1 = 98.825$ and $m_2 = 101.175$, so these give the Bayes rule.

Graduate Diploma, Module 2, 2010. Question 4

Part (a)

First define the *size* of a test as follows: size = P(reject H_0 | H_0 true).

Also define the *power* of a test as follows: power = P(reject H_0 | H_1 true).

Then a test (between two simple hypotheses) is the most powerful test at level α if its size is equal to α and no other test with size $\leq \alpha$ has greater power.

Part (b)

We have the Weibull distribution with probability density

$$f(y) = \theta\phi y^{\phi-1}\exp(-\theta y^{\phi}) \qquad \text{for } y > 0,$$

where θ (> 0) is unknown and ϕ (> 0) is known.

The hypotheses are

H_0: $\theta = 0.5$; H_1: $\theta = 1.0$.

(i) The likelihood is $L(\theta) = \prod_{i=1}^{n} f(y_i) = (\theta\phi)^n (\Pi y_i)^{\phi-1}\exp(-\theta\Sigma y_i^{\phi})$.

The form of the most powerful test is to reject H_0 if $\dfrac{L(0.5)}{L(1.0)} \leq k$ for some k.

So we require

$$\frac{0.5^n \phi^n (\Pi y_i)^{\phi-1}\exp\left(-\tfrac{1}{2}\Sigma y_i^{\phi}\right)}{\phi^n (\Pi y_i)^{\phi-1}\exp\left(-\Sigma y_i^{\phi}\right)} \leq k$$

ie $\exp\left(\tfrac{1}{2}\Sigma y_i^{\phi}\right) \leq k'$ for some k'

ie $\Sigma y_i^{\phi} \leq k''$ for some k''.

(k'' is chosen so that the size of the test is equal to the specified significance level.)

Solution continued on next page

(ii) Let $W_i = Y_i^{\phi}$. For $w > 0$, the cdf of W_i is

$$F(w) = P(W_i \le w) = P(Y_i^{\phi} \le w) = P(Y_i \le w^{1/\phi})$$

$$= \int_0^{w^{1/\phi}} \theta\phi y^{\phi-1} \exp(-\theta y^{\phi}) dy = \left[-\exp\left(-\theta y^{\phi}\right)\right]_0^{w^{1/\phi}} = 1 - e^{-\theta w} .$$

This might be recognised as the cdf of an exponential distribution (with mean $1/\theta$), or this can be seen perhaps more familiarly by finding the pdf of W_i, ie

$$\frac{d}{dw}\left(1 - e^{-\theta w}\right) = \theta e^{-\theta w} .$$

(iii) The test statistic found in part (i) is Σw_i. From the result given in the question, we have $2\theta \Sigma W_i \sim \chi^2_{2n}$. Thus, for $n = 20$, under H_0 we have $\Sigma W_i \sim \chi^2_{40}$.

For significance level 0.05, k'' (see part (i)) must satisfy $P\left(\chi^2_{40} \le k''\right) = 0.05$.

Thus, from χ^2 tables, k'' = 26.509, and the most powerful test is to reject H_0 if Σw_i ($=\Sigma y_i^{\phi}$) $\le$ 26.509.

(iv) Under H_1 we have $2\Sigma W_i \sim \chi^2_{40}$. Thus the power is

$$P(\Sigma W_i \le 26.509 | \theta = 1) = P(2\Sigma W_i \le 53.018 | \theta = 1) = P\left(\chi^2_{40} \le 53.018\right).$$

From the tables, $P\left(\chi^2_{40} \le 51.805\right) = 0.9$ and $P\left(\chi^2_{40} \le 55.758\right) = 0.95$. Thus by linear interpolation we get

$$\text{Power} \approx 0.9 + 0.05\left(\frac{53.018 - 51.805}{55.758 - 51.805}\right) = 0.915 .$$

Graduate Diploma, Module 2, 2010. Question 5

(i) The likelihood is $L(\sigma_1, \sigma_2, \sigma_3) = \prod_{i=1}^{3} \prod_{j=1}^{n_i} \left(2\pi\sigma_i^2\right)^{-\frac{1}{2}} \exp\left(-\frac{x_{ij}^2}{2\sigma_i^2}\right)$

$$= (2\pi)^{-\Sigma n_i/2} \left(\prod_{i=1}^{3} \sigma_i^{-n_i} \right) \exp\left(-\frac{1}{2} \sum_i \left(\sigma_i^{-2} \sum_j x_{ij}^2 \right) \right).$$

$\therefore$ the log likelihood is

$$\log L(\sigma_1, \sigma_2, \sigma_3) = -\frac{1}{2} \sum n_i \log(2\pi) - \sum n_i \log \sigma_i - \frac{1}{2} \sum_i \left(\sigma_i^{-2} \sum_j x_{ij}^2 \right).$$

Thus for each k (= 1, 2, 3),

$$\frac{d \log L}{d\sigma_k} = -\frac{n_k}{\sigma_k} + \sigma_k^{-3} \sum_j x_{kj}^2$$

which on setting equal to zero gives solution $\hat{\sigma}_k^2 = \frac{1}{n_k} \sum_j x_{kj}^2$.

To investigate whether these give a maximum, we need to consider for each k the second derivatives $\frac{d^2 \log L}{d\sigma_k^2} = \frac{n_k}{\sigma_k^2} - 3\sigma_k^{-4} \sum_j x_{kj}^2$. Evaluated at $\hat{\sigma}_k$, this is

$$\frac{n_k}{\hat{\sigma}_k^2} - \frac{3\sum_j x_{kj}^2}{\hat{\sigma}_k^4} = \frac{n_k}{\hat{\sigma}_k^2} - \frac{3n_k\hat{\sigma}_k^2}{\hat{\sigma}_k^4} = -\frac{2n_k}{\hat{\sigma}_k^2} < 0$$

and it follows that $\hat{\sigma}_k^2$ as obtained above indeed gives the maximum likelihood estimator for each k.

Solution continued on next page

(ii) H_0 is $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$.

Let $\hat{\sigma}^2$ denote the restricted maximum likelihood estimator under H_0. The restricted likelihood is

$$(2\pi)^{-\Sigma n_i/2}\,\sigma^{-\Sigma n_i}\exp\left(-\frac{1}{2\sigma^2}\sum_i\sum_j x_{ij}^{\,2}\right)$$

and the restricted log likelihood is

$$-\frac{1}{2}\sum n_i\log(2\pi)-\sum n_i\log\sigma-\frac{1}{2\sigma^2}\sum_i\sum_j x_{ij}^{\,2}\,.$$

Differentiating this with respect to σ gives

$$-\frac{\sum n_i}{\sigma}+\frac{1}{\sigma^3}\sum_i\sum_j x_{ij}^{\,2}$$

and setting this equal to 0 gives

$$\hat{\sigma}^2=\frac{\sum_i\sum_j x_{ij}^{\,2}}{\sum n_i}$$

(we can again check that this is a maximum by considering the second derivative evaluated at this point).

Thus the restricted maximum log likelihood is

$$-\frac{1}{2}\sum n_i\log(2\pi)-\sum n_i\log\left(\frac{\sum_i\sum_j x_{ij}^{\,2}}{\sum n_i}\right)^{\frac{1}{2}}-\frac{1}{2}\sum n_i$$

$$=-\frac{1}{2}\sum n_i\log(2\pi)-\frac{1}{2}\sum n_i\log\left(\frac{\sum_i\sum_j x_{ij}^{\,2}}{\sum n_i}\right)-\frac{1}{2}\sum n_i\,.$$

Solution continued on next page

From part (i), the unrestricted maximum log likelihood is

$$-\frac{1}{2}\sum n_i \log(2\pi) - \sum n_i \log\left(\frac{\sum_j x_{ij}^{\,2}}{n_i}\right)^{\frac{1}{2}} - \frac{1}{2}\sum_i \left(\frac{n_i}{\sum_j x_{ij}^{\,2}} \sum_j x_{ij}^{\,2}\right)$$

$$= -\frac{1}{2}\sum n_i \log(2\pi) - \frac{1}{2}\sum n_i \log\left(\frac{\sum_j x_{ij}^{\,2}}{n_i}\right) - \frac{1}{2}\sum_i n_i \ .$$

Denoting these maximum log likelihoods by $\log L(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$ (unrestricted) and $\log L\left(\hat{\hat{\sigma}}\right)$ (restricted), the generalised likelihood ratio test is to reject H_0 if

$$\log L(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3) - \log L\left(\hat{\hat{\sigma}}\right) \geq k \quad \text{for some } k$$

ie if

$$-\frac{1}{2}\sum n_i \log(2\pi) - \frac{1}{2}\sum n_i \log\left(\frac{\sum_j x_{ij}^{\,2}}{n_i}\right) - \frac{1}{2}\sum_i n_i$$

$$-\left[-\frac{1}{2}\sum n_i \log(2\pi) - \frac{1}{2}\sum n_i \log\left(\frac{\sum_i\sum_j x_{ij}^{\,2}}{\sum n_i}\right) - \frac{1}{2}\sum n_i\right] \geq k$$

ie if

$$-\frac{1}{2}\sum n_i \log\left(\frac{\sum_j x_{ij}^{\,2}}{n_i}\right) + \frac{1}{2}\sum n_i \log\left(\frac{\sum_i\sum_j x_{ij}^{\,2}}{\sum n_i}\right) \geq k$$

ie if

$$\frac{1}{2}\sum n_i \left\{\log\left(\frac{\sum_i\sum_j x_{ij}^{\,2}}{\sum n_i}\right) - \log\left(\frac{\sum_j x_{ij}^{\,2}}{n_i}\right)\right\} \geq k\,, \quad \text{for some } k.$$

Solution continued on next page

(iii) We use the asymptotic result that $2\left(\log L(\hat{\sigma}_1,\hat{\sigma}_2,\hat{\sigma}_3)-\log L\left(\hat{\hat{\sigma}}\right)\right)\sim\chi_2^2$ under H_0 (note 2 degrees of freedom because there are two constraints under H_0).

Inserting the values from the question,

$$2\left(\log L(\hat{\sigma}_1,\hat{\sigma}_2,\hat{\sigma}_3)-\log L\left(\hat{\hat{\sigma}}\right)\right)$$

$$= 50\left(\log\frac{14.3}{130}-\log\frac{5.5}{50}\right) + 40\left(\log\frac{14.3}{130}-\log\frac{3.6}{40}\right) + 50\left(\log\frac{14.3}{130}-\log\frac{5.2}{40}\right)$$

$= (50 \times 0) + (40 \times 0.20067) + (40 \times (-0.16705))$

$= 1.34.$

The 95% point of χ_2^2 is 5.991. So the result is not significant at the 5% level – there is no real evidence against H_0.

Graduate Diploma, Module 2, 2010. Question 6

Part (a)

Nonparametric inference refers to situations where no specific distribution $f(\underline{x} \mid \theta)$ is assumed for the population underlying the data. This being the case, inference about its parameter(s) θ is not possible. However, there are some properties of a distribution, such as its median, which do not require explicit knowledge of $f(\underline{x} \mid \theta)$. Inference in respect of these properties is possible. Although there is no explicit assumption about the form of $f(\underline{x} \mid \theta)$, some weaker assumptions may be needed.

Advantages compared to parametric inference:

- can be used with (non-numeric) ordinal data
- requires few assumptions about the distribution underlying the data
- sometimes nearly as powerful as a corresponding parametric test.

Disadvantages:

- some loss of power
- corresponding parametric tests are often quite robust to some failure in distributional assumptions
- often harder to interpret than if parametric assumptions had been made.

Part (b)

(i) We have H_0: $m = 20$, H_1: $m \neq 20$, where m is the population median.

Let Y be the number of observations that are ≤ 20, so that $Y \sim B(12, \frac{1}{2})$ under H_0. This distribution is symmetrical about $Y = 6$, so the critical region will be of the form $Y \leq y$ or $Y \geq 12 - y$, for some $y = 0, 1, 2, \ldots, 6$.

Under H_0 we have $P(Y = 0) = (\frac{1}{2})^{12} = 0.00024$, $P(Y = 1) = 12(\frac{1}{2})^{12} = 0.00293$, $P(Y = 2) = 0.01611$, $P(Y = 3) = 0.05371$.

Thus, under H_0, $P(Y \leq 2) = 0.0193$ and $P(Y \leq 3) = 0.0730$.

Thus with $y = 2$ the size (significance level) of the test is $2 \times 0.0193 \approx 0.039$, and with $y = 3$ it is greater than 0.05.

So the required critical region is $Y \leq 2$ or $Y \geq 10$.

Solution continued on next page

(ii) An approximate 95% confidence interval for m consists of those values m' such that, if the hypotheses were H_0: $m = m'$ and H_1: $m \neq m'$, then H_0 could not be rejected at the 5% level.

Let $X_{(1)}, X_{(2)}, \ldots, X_{(12)}$ be the order statistics for the 12 observations.

Now let Y be the number of observations that are $\leq m'$. Part (b)(i) shows that H_0 is not rejected if $3 \leq Y \leq 9$ (and is rejected otherwise). However, $Y \geq 3$ means that there are ≥ 3 observations $\leq m'$ and so $X_{(3)} \leq m'$. Similarly, $Y \leq 9$ means that there are ≤ 9 observations $\leq m'$ and so $X_{(10)} > m'$. Thus the approximate 95% confidence interval for m is from $X_{(3)}$ to $X_{(10)}$.

(i) We have $\bar{X} \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right)$ and $\bar{Y} \sim \mathrm{N}\left(\alpha\mu,\ \frac{\sigma^2}{n}\right)$.

$$\therefore \frac{\bar{Y}}{\alpha} \sim \mathrm{N}\left(\mu,\ \frac{\sigma^2}{\alpha^2 n}\right) \quad \text{and so} \quad \bar{X} - \frac{\bar{Y}}{\alpha} \sim \mathrm{N}\left(0,\ \frac{\sigma^2}{n}\left(1+\frac{1}{\alpha^2}\right)\right).$$

$$\therefore \frac{\bar{X} - \frac{\bar{Y}}{\alpha}}{\sigma\sqrt{\frac{1}{n}\left(1+\frac{1}{\alpha^2}\right)}} = \frac{\alpha\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n}\left(1+\alpha^2\right)}} \sim \mathrm{N}(0,\ 1).$$

Thus $\dfrac{\alpha\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n}\left(1+\alpha^2\right)}}$ is a pivotal quantity for α, because

(i) it is a function of α but not of (the other unknown parameter) μ

and

(ii) its distribution, N(0, 1), does not involve any unknown parameters.

(ii) The likelihood is

$$L(\alpha,\mu) = \prod_{i=1}^{n}\left(2\pi\sigma^2\right)^{-1/2} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \prod_{i=1}^{n}\left(2\pi\sigma^2\right)^{-1/2} e^{-\frac{(y_i-\alpha\mu)^2}{2\sigma^2}}$$

$$= \left(2\pi\sigma^2\right)^{-n} \exp\left(-\frac{1}{2\sigma^2}\left(\sum(x_i-\mu)^2 + \sum(y_i-\alpha\mu)^2\right)\right).$$

$$\therefore \log L(\alpha,\mu) = -n\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(\sum(x_i-\mu)^2 + \sum(y_i-\alpha\mu)^2\right).$$

Solution continued on next page

$$\therefore \frac{\partial \log L}{\partial \alpha} = \frac{\mu}{\sigma^2}\sum\left(y_i - \alpha\mu\right)$$

and $$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum\left(x_i - \mu\right) + \frac{\alpha}{\sigma^2}\sum\left(y_i - \alpha\mu\right).$$

Setting first $\frac{\partial \log L}{\partial \alpha} = 0$, we get $\sum\left(y_i - \hat{\alpha}\hat{\mu}\right) = 0$.

Using this in $\frac{\partial \log L}{\partial \mu} = 0$ gives $\frac{1}{\sigma^2}\sum\left(x_i - \hat{\mu}\right) = 0$, so that $\hat{\mu} = \bar{x}$.

Inserting this in the previous expression gives $\sum y_i - n\bar{x}\hat{\alpha} = 0$, i.e. $\hat{\alpha} = \bar{y}/\bar{x}$.

Thus the maximum likelihood estimator of α is $\bar{Y}/\bar{X}$, as required [NB – the question indicates that only the first partial derivatives of the log likelihood need be considered].

(iii) We have $\dfrac{\alpha\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n}\left(1+\alpha^2\right)}} \sim$ N(0, 1) and so

$$P\left(-1.96 < \frac{\alpha\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n}\left(1+\alpha^2\right)}} < 1.96\right) = 0.95.$$

$$\therefore P\left(\bar{Y} - 1.96\sigma\sqrt{\frac{1}{n}\left(1+\alpha^2\right)} < \alpha\bar{X} < \bar{Y} + 1.96\sigma\sqrt{\frac{1}{n}\left(1+\alpha^2\right)}\right) = 0.95.$$

Therefore an approximate 95% confidence interval for α is

$$\frac{\bar{Y}}{\bar{X}} \pm 1.96\frac{\sigma}{\bar{X}}\sqrt{\frac{1}{n}\left(1+\hat{\alpha}^2\right)}, \qquad \text{i.e. } \frac{\bar{Y}}{\bar{X}} \pm 1.96\frac{\sigma}{\bar{X}}\sqrt{\frac{1}{n}\left(1+\left(\frac{\bar{Y}}{\bar{X}}\right)^2\right)}.$$

Solution continued on next page

(iv) Step 1: draw a random sample of size n, with replacement, from the set $\{X_1, X_2, \ldots, X_n\}$. Call the values $X_1^*, X_2^*, \ldots, X_n^*$.

Step 2: draw an independent random sample of size n, with replacement, from the set $\{Y_1, Y_2, \ldots, Y_n\}$. Call the values $Y_1^*, Y_2^*, \ldots, Y_n^*$.

Step 3: let $\alpha^* = \dfrac{\Sigma Y_i^*}{\Sigma X_i^*}$.

Repeat steps 1 to 3 many times, say 10,000 times, to obtain estimates $\alpha_1^*, \alpha_2^*, \ldots, \alpha_{10,000}^*$.

Order these estimates (smallest to largest) $\alpha_{(1)}^*, \alpha_{(2)}^*, \ldots, \alpha_{(10,000)}^*$.

Then a bootstrap 95% confidence interval for α is $\alpha_{(250)}^*$ to $\alpha_{(9750)}^*$.

<u>Graduate Diploma, Module 2, 2010. Question 8</u>

<u>Part (a)</u>

Let $\pi(\eta_0)$ be the prior probability of the null hypothesis and $\pi(\eta_1) = 1-\pi(\eta_0)$ be the prior probability of the alternative.

Then the prior odds of the null hypothesis = $\dfrac{\pi(\eta_0)}{\pi(\eta_1)}$.

Now denote the data by **d**, then the posterior odds of the null hypothesis = $\dfrac{\pi(\eta_0|\mathbf{d})}{\pi(\eta_1|\mathbf{d})}$, where the π functions here represent the posterior probabilities given the data.

$$\text{The Bayes factor} = \frac{\text{likelihood under null hypothesis}}{\text{likelihood under alternative hypothesis}}$$

$$= \frac{f(\mathbf{d}|\eta_0)}{f(\mathbf{d}|\eta_1)} \quad \text{in obvious notation.}$$

In the usual manner of Bayesian inference, we have $\pi(\eta|\mathbf{d}) = \dfrac{f(\mathbf{d}|\eta)\pi(\eta)}{h(\mathbf{d})}$.

$$\text{Thus posterior odds} = \frac{\pi(\eta_0|\mathbf{d})}{\pi(\eta_1|\mathbf{d})} = \frac{f(\mathbf{d}|\eta_0)\pi(\eta_0)}{f(\mathbf{d}|\eta_1)\pi(\eta_1)} = \text{Bayes factor} \times \text{prior odds.}$$

Solution continued on next page

Part (b)

(i) $P(X_i = x_i) = \theta(1-\theta)^{x_i}$ (for x_i = 0, 1, 2, …).

$\therefore$ the likelihood is $f(\mathbf{x}|\theta) = \prod_{i=1}^{20}\left(\theta(1-\theta)^{x_i}\right) = \theta^{20}(1-\theta)^{\Sigma x_i} = \theta^{20}(1-\theta)^{40}$.

H_0 is $\theta = 0.5$, H_1 is $\theta = 0.25$, with the prior probability of H_0 being 0.75.

$$\therefore \text{ Bayes factor } = \frac{(0.5)^{20}(1-0.5)^{40}}{(0.25)^{20}(1-0.25)^{40}} = 2^{20}\left(\frac{2}{3}\right)^{40} = 0.0948.$$

Also, the prior odds of $H_0 = \dfrac{0.75}{1-0.75} = 3$.

$\therefore$ the posterior odds of $H_0 = 3 \times 0.0948 = 0.2845$.

(ii) Now H_0 is $\theta = 0.5$ and H_1 is $\theta \neq 0.5$. Also, under H_1 the prior distribution of θ is $12\theta^2(1-\theta)$ (for $0 < \theta < 1$).

$$\therefore f(\mathbf{x}|H_1) = \int_0^1 \theta^{20}(1-\theta)^{40} \times 12\theta^2(1-\theta)\,d\theta$$

$$= 12\int_0^1 \theta^{22}(1-\theta)^{41}\,d\theta$$

We manipulate this to create the pdf of the beta distribution with parameters 23 and 42 – see the hint in the question

$$= 12 \times \frac{22!\,41!}{64!}\int_0^1 \frac{\Gamma(65)}{\Gamma(23)\Gamma(42)}\theta^{22}(1-\theta)^{41}\,d\theta$$

$$= 12 \times \frac{22!\,41!}{64!}.$$

$$\therefore \text{ Bayes factor } = \frac{(0.5)^{60}}{12 \times \dfrac{22!\,41!}{64!}} = \frac{(0.5)^{60} \times 64!}{12 \times 22! \times 41!},$$ as required.

Solution continued on next page

(iii) Let $W \sim B(64, 0.5)$. Then we have

$$P(W=22) = \frac{64!}{22!\,42!}(0.5)^{64} = \text{Bayes factor in part (ii)} \times (0.5)^4 \times \frac{12}{42}.$$

The Normal approximation to the distribution of W is N(32, 16). Using this,

$$P(W=22) \approx \Phi\left(\frac{22.5-32}{4}\right) - \Phi\left(\frac{21.5-32}{4}\right)$$

$$= \Phi(-2.375) - \Phi(-2.625)$$

$$= 0.9957 - 0.9912$$

$$= 0.0045.$$

$$\therefore \text{ Bayes factor } \approx 0.0045 \times 2^4 \times \frac{42}{12} = 0.252.$$