THE ROYAL STATISTICAL SOCIETY

2003 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

PAPER II – STATISTICAL THEORY & METHODS

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.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2003. Question 1

(i) The likelihood L of the sample is

$$L = \prod_{i=1}^{n} f(x_i) = \theta^n \prod_{i=1}^{n} (1+x_i)^{-(\theta+1)}$$

i.e. we have $\log L = n\log\theta - (\theta+1)\sum_{i=1}^{n}\log(1+x_i)$.

$$\therefore \frac{d(\log L)}{d\theta} = \frac{n}{\theta} - \sum_{i=1}^{n}\log(1+x_i) \qquad \text{(A)}$$

and setting this equal to zero gives $\hat{\theta} = \dfrac{n}{\sum\log(1+x_i)}$. Further, $\dfrac{d^2(\log L)}{d\theta^2} = -\dfrac{n}{\theta^2}$, confirming that this is a maximum.

Hence, by the invariance property of maximum likelihood estimators,

$$\hat{\gamma} = \frac{1}{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\log(1+x_i).$$

(ii) $$P\{\log(1+X_i) > w\} = P(X_i > e^w - 1) = \theta\int_{e^w-1}^{\infty} \frac{1}{(1+x)^{\theta+1}}\,dx$$

$$= \left[\frac{-1}{(1+x)^{\theta}}\right]_{e^w-1}^{\infty} = 0 + \frac{1}{e^{\theta w}} = e^{-\theta w}.$$

Hence the cdf of this is $1 - e^{-\theta w}$ and the pdf is $\theta e^{-\theta w}$, so the distribution is exponential with mean $1/\theta = \gamma$.

$\therefore E[\hat{\gamma}] = \dfrac{1}{n}.nE[\log(1+X_i)] = \gamma$. Thus $\hat{\gamma}$ is an unbiased estimator of γ.

(iii) $$\frac{d}{d\gamma}(\log L) = \frac{d}{d\theta}(\log L)\frac{d\theta}{d\gamma}$$

$= \left\{n\gamma - \sum\log(1+x_i)\right\}\left\{-\dfrac{1}{\gamma^2}\right\}$ [using result (A) above] $= -\dfrac{n}{\gamma} + \dfrac{1}{\gamma^2}\sum_{i=1}^{n}\log(1+x_i)$.

$\therefore \dfrac{d^2}{d\gamma^2}(\log L) = \dfrac{n}{\gamma^2} - \dfrac{2}{\gamma^3}\sum_{i=1}^{n}\log(1+x_i)$, $\quad \therefore E\left[-\dfrac{d^2}{d\gamma^2}\log L\right] = -\dfrac{n}{\gamma^2} + \dfrac{2}{\gamma^3}n\gamma = \dfrac{n}{\gamma^2}$, and the C-R lower bound is γ^2/n. From (ii), $\text{Var}(\hat{\gamma}) = \gamma^2/n$, so the bound is attained.

(iv) No. Because the bound is attainable for γ, it cannot be attainable for a non-linear function of γ, such as $\theta = 1/\gamma$.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2003. Question 2

(i) Given a random sample of data $\boldsymbol{X}$ from a distribution having parameter θ, a statistic $T(\boldsymbol{X})$ is sufficient for θ if the conditional distribution of $\boldsymbol{X}$ given $T(\boldsymbol{X})$ does not involve θ.

(ii) Let $Y = \min(X_i)$. Defining the indicator function $I_\theta(x_i)$ to be 0 when $x_i < \theta$ and to be 1 when $x_i \geq \theta$, the likelihood function is $L(\theta) = \prod_{i=1}^{n} e^{\theta - x_i} I_\theta(x_i)$. Also, we have $\prod_{i=1}^{n} I_\theta(x_i) = I_\theta(y)$ and so $L(\theta) = e^{n\theta} I_\theta(y) e^{-\Sigma x_i}$. Therefore, by the factorisation theorem, Y is sufficient for θ.

(iii) $P(Y > y)$ implies $P(X_1 > y,\ X_2 > y,\ \ldots,\ X_n > y)$, i.e. $P(Y > y) = \prod_{i=1}^{n} P(X_i > y)$.
Now, $P(X > y) = \int_y^\infty e^{\theta - x} dx = \left[-e^{\theta - x}\right]_y^\infty = e^{\theta - y}$, so $P(Y > y) = e^{n(\theta - y)}$, for $y > \theta$.
Hence the cdf is $F(y) = 1 - e^{n(\theta - y)}$ and the pdf is $f(y) = dF(y)/dy = ne^{n(\theta - y)}$, for $y > \theta$.

(iv) We have that Y has a shifted exponential distribution. Hence $E(Y) = \theta + \frac{1}{n}$ and $\text{Var}(Y) = \frac{1}{n^2}$, so that $E(Y - c) = \theta - c + \frac{1}{n}$ and $\text{Var}(Y - c) = \frac{1}{n^2}$. From these, $\text{Bias}(Y - c) = \frac{1}{n} - c$ and $MSE = \text{Bias}^2 + \text{Var} = \left(\frac{1}{n} - c\right)^2 + \frac{1}{n^2}$, which is clearly minimised when $c = 1/n$. Thus $Y - (1/n)$ has smallest variance of all estimators of the form $Y - c$.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2003. Question 3

(i) The likelihood for a sample (x_1, x_2, …, x_n) is $L(\theta) = \text{Const.} \times \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}$, and so the likelihood ratio is $\lambda = \dfrac{L\left(\frac{2}{3}\right)}{L\left(\frac{3}{4}\right)} = \dfrac{\left(\frac{2}{3}\right)^{\Sigma x_i}\left(\frac{1}{3}\right)^{n-\Sigma x_i}}{\left(\frac{3}{4}\right)^{\Sigma x_i}\left(\frac{1}{4}\right)^{n-\Sigma x_i}} = \left(\dfrac{8}{9}\right)^{\Sigma x_i}\left(\dfrac{4}{3}\right)^{n-\Sigma x_i}$. Using the Neyman-Pearson lemma, we reject H_0 when $\lambda > c$, where c is chosen to give the required level of test, α. Now, λ is an increasing function of Σx_i, hence of $\hat{\theta}$, and an equivalent rule is therefore to reject H_0 when $\hat{\theta} < k$, where k is chosen to give test level α.

(ii) $n\hat{\theta}$ is binomial with parameters (n, θ). Hence the large-sample distribution of $\hat{\theta}$ is $\mathrm{N}(\theta, \theta(1-\theta)/n)$. When $\theta = 3/4$ this is $\mathrm{N}\left(\frac{3}{4}, \frac{3}{16n}\right)$, and when $\theta = 2/3$ it is $\mathrm{N}\left(\frac{2}{3}, \frac{2}{9n}\right)$.

(iii) For $\alpha = 0.05$, choose k such that $P\left(\hat{\theta} < k \middle| \theta = \frac{3}{4}\right) = 0.05$. That is, we want $\Phi\left(\dfrac{k - \frac{3}{4}}{\sqrt{3/16n}}\right) = 0.05$, or $\dfrac{k - \frac{3}{4}}{\sqrt{3/16n}} = -1.6449$, giving $k = \dfrac{3}{4} - \dfrac{1.6449}{4}\sqrt{\dfrac{3}{n}}$.

(iv) For power 0.95, $P\left(\hat{\theta} < k \middle| \theta = \frac{2}{3}\right) = 0.95$, i.e. $\Phi\left(\dfrac{k - \frac{2}{3}}{\sqrt{2/9n}}\right) = 0.95$ or $\dfrac{k - \frac{2}{3}}{\sqrt{2/9n}} = 1.6449$, giving $k = \dfrac{2}{3} + \dfrac{1.6449}{3}\sqrt{\dfrac{2}{n}}$.

Using this expression for k together with the expression in (iii) means that we require

$$\frac{3}{4} - \frac{1.6449}{4}\sqrt{\frac{3}{n}} = \frac{2}{3} + \frac{1.6449}{3}\sqrt{\frac{2}{n}} \quad \text{or} \quad \frac{1}{12} = \frac{1.6449}{\sqrt{n}}\left(\frac{1}{4}\sqrt{3} + \frac{1}{3}\sqrt{2}\right).$$

Thus we get $\sqrt{n} = 12 \times 1.6449 \times 0.9044 = 17.8521$ and $n = 318.7$, so we take $n = 319$.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2003. Question 4

(i) $P(0) = \theta$ $P(1) = \theta(1-\theta)$ $P(\geq 2) = 1 - \theta - \theta(1-\theta) = (1-\theta)^2$.

Thus the likelihood of n_0 zeros, n_1 ones and n_2 with two or more flaws is

$$L = \theta^{n_0}\{\theta(1-\theta)\}^{n_1}\{1-\theta\}^{2(n-n_0-n_1)} = \theta^{n_0+n_1}(1-\theta)^{2n-2n_0-n_1}.$$

(ii) $\log L(\theta) = (n_0+n_1)\log\theta + (2n-2n_0-n_1)\log(1-\theta)$.

$$\therefore \frac{d}{d\theta}(\log L) = \frac{n_0+n_1}{\theta} - \frac{2n-2n_0-n_1}{1-\theta}.$$

Setting this equal to zero gives that $\hat{\theta}$ satisfies $(n+n_0)(1-\hat{\theta}) = (2n-2n_0-n_1)\hat{\theta}$, so that $\hat{\theta} = \dfrac{n_0+n_1}{2n-n_0}$.

Further, $\dfrac{d^2}{d\theta^2}(\log L) = -\dfrac{n_0+n_1}{\theta^2} - \dfrac{2n-2n_0-n_1}{(1-\theta)^2}$, which confirms that $\hat{\theta}$ is a maximum, and the sample information when $\theta = \hat{\theta}$ (given by $-E\left(\frac{d^2\log L}{d\theta^2}\right)$ evaluated at $\theta = \hat{\theta}$) is

$$\frac{(2n-n_0)^2}{n_0+n_1} + \frac{(2n-n_0)^2}{2n-2n_0-n_1} \quad \text{(using } 1-\hat{\theta} = \frac{2n-2n_0-n_1}{2n-n_0}\text{)}.$$

(iii) An approximate 90% confidence interval for θ is $\hat{\theta} \pm \dfrac{1.6449}{\sqrt{(\text{sample information})}}$.

In the case when $n = 100$, $n_0 = 90$ and $n_1 = 7$, we have $2n - n_0 = 110$, $n_0 + n_1 = 97$ and $2n - 2n_0 - n_1 = 13$.

Thus $\hat{\theta} = \dfrac{97}{110} = 0.882$ and the sample information is $\dfrac{110^2}{97} + \dfrac{110^2}{13} = 1055.5115$.

Thus the confidence interval is

$$0.882 \pm \frac{1.6449}{32.489}, \quad \text{i.e. } 0.882 \pm 0.051 \quad \text{or} \quad (0.831, 0.933).$$

Graduate Diploma, Statistical Theory & Methods, Paper II, 2003. Question 5

(i) $\alpha = 0.025$, $\beta = 0.075$.

For observations $x_1, x_2, \ldots, x_n$ the likelihood is $L_n(\theta) = \dfrac{2^n \prod_{i=1}^{n} x_i}{\theta^{2n}} \exp\left(-\dfrac{\Sigma x_i^2}{\theta^2}\right)$, and for the given H_0 and H_1 the likelihood ratio is $\lambda_n = \dfrac{L_n(2)}{L_n(1)} = \dfrac{1}{2^{2n}} \exp\left(\dfrac{3}{4}\sum x_i^2\right)$.

The sequential probability ratio test rule is to continue sampling while $A < \lambda_n < B$, accept H_0 if $\lambda_n \geq B$ and reject H_0 (i.e. accept H_1) if $\lambda_n \leq A$. A and B are given by $A = \dfrac{\alpha}{1-\beta} = \dfrac{0.025}{0.925} = \dfrac{1}{37} = 0.027$, $B = \dfrac{1-\alpha}{\beta} = \dfrac{0.975}{0.075} = 13$.

(ii) $E(X^2) = \int_0^\infty \dfrac{2x^3}{\theta^2} e^{-x^2/\theta^2} dx$ put $y = x^2/\theta^2$, so that $dy/dx = 2x/\theta^2$

$$= \int_0^\infty \theta^2 y e^{-y} dy = \theta^2 \Gamma(2) = \theta^2 .$$

The ith item in the sequence making up $\{\log \lambda_n\}$ is $Z_i = -2\log 2 + \frac{3}{4} X_i^2$.

$$E(Z_i | \theta = 2) = -2\log 2 + \tfrac{3}{4}.4 = 1.6137 .$$

$$E(Z_i | \theta = 1) = -2\log 2 + \tfrac{3}{4}.1 = -0.6363 .$$

$$E(N|\theta = 2) \approx \frac{\alpha \log A + (1-\alpha)\log B}{E(Z_i|\theta = 2)} = 1.494 .$$

$$E(N|\theta = 1) \approx \frac{(1-\beta)\log A + \beta \log B}{E(Z_i|\theta = 1)} = 4.948 .$$

(iii) $x_1 = 2.2$. $\lambda_1 = \frac{1}{4}\exp\left(\frac{3}{4} \times 4.84\right) = 9.428$, continue sampling.

$x_2 = 2.5$. $\lambda_2 = \frac{1}{16}\exp\left(\frac{3}{4} \times (2.2^2 + 2.5^2)\right) = 255.93$, accept H_0.

No need to consider x_3.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2003. Question 6

(i) A prior distribution is conjugate for a particular model (e.g. Normal, beta) if the prior and posterior distributions are from the same family.

(ii) Likelihood $L(\mathbf{x}|\theta) = \text{constant} \times \theta^{n/2} \exp\left\{-\frac{1}{2}\theta\sum_{i=1}^{n}\left(x_i - 2 + x_i^{-1}\right)\right\}$.

The posterior distribution is proportional to $g(\theta)L(\mathbf{x}|\theta)$, i.e. it is

$$\text{constant} \times \theta^{\alpha-1+(n/2)} \exp\left\{-\theta\left[\beta + \frac{1}{2}\sum_{i=1}^{n}\left(x_i - 2 + x_i^{-1}\right)\right]\right\},$$

which is gamma with parameters $\alpha + (n/2)$ and $\beta + \frac{1}{2}\sum\left(x_i - 2 + x_i^{-1}\right)$. Hence the gamma prior is conjugate.

(iii) The mean, 20, is α/β. The variance, also 20, is α/β^2. So β must be 1, and α must be 20, and these must be the values used in the prior distribution.

(iv) $\theta|\mathbf{x}$ is gamma$\left[\left(20 + \frac{80}{2}\right), \left(1 + \frac{5.0}{2}\right)\right]$, i.e. gamma(60, 3.5).

The mean of this is 60/3.5 and the variance is $60/(3.5)^2$. These are used in a Normal approximation, which is satisfactory for n = 80. Hence an approximate 90% highest posterior density interval for θ is given by

$$\frac{60}{3.5} \pm 1.6449\frac{\sqrt{60}}{3.5},$$

i.e. 17.143 ± 3.640 or (13.50, 20.78).

Graduate Diploma, Statistical Theory & Methods, Paper II, 2003. Question 7

(i) The likelihood $L(\mathbf{x}|\theta)$ is $k.\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}$, and the posterior density is

$$g(\theta|\mathbf{x}) \propto g(\theta)L(\mathbf{x}|\theta)$$

i.e. we have

$$g(\theta|\mathbf{x}) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}$$
$$= \theta^{\alpha+\Sigma x_i-1}(1-\theta)^{\beta+n-1-\Sigma x_i}.$$

So $\theta|\mathbf{x}$ is beta($\alpha + \Sigma x_i$, $\beta + n - \Sigma x_i$), and with a squared error loss the Bayes estimator of θ is the mean of this distribution, i.e. $\dfrac{\alpha+\Sigma x_i}{\alpha+\beta+n}$.

(ii) When $\alpha = \beta = \frac{1}{2}\sqrt{n}$, we have $\hat{\theta}_{\mathrm{B}} = \dfrac{\frac{1}{2}\sqrt{n}+\Sigma x_i}{n+\sqrt{n}}$. The expectation of this is $\dfrac{\frac{1}{2}\sqrt{n}+n\theta}{n+\sqrt{n}}$, so its bias is given by

$$\frac{\frac{1}{2}\sqrt{n}+n\theta}{n+\sqrt{n}} - \theta = \frac{\sqrt{n}\left(\frac{1}{2}-\theta\right)}{n+\sqrt{n}} = \frac{\frac{1}{2}-\theta}{1+\sqrt{n}}.$$

Also,

$$\mathrm{Var}\left(\hat{\theta}_{\mathrm{B}}\right) = \mathrm{Var}\left(\frac{\Sigma x_i}{n+\sqrt{n}}\right) = \frac{1}{\left(n+\sqrt{n}\right)^2}n\theta(1-\theta) = \frac{\theta(1-\theta)}{\left(1+\sqrt{n}\right)^2}.$$

The risk is

$$MSE\left(\hat{\theta}_{\mathrm{B}}\right) = \mathrm{Bias}^2 + \mathrm{Variance} = \frac{\left(\frac{1}{2}-\theta\right)^2}{\left(1+\sqrt{n}\right)^2} + \frac{\theta(1-\theta)}{\left(1+\sqrt{n}\right)^2} = \frac{1}{4\left(1+\sqrt{n}\right)^2}.$$

(iii) A Bayes estimator with constant risk for all θ is minimax.

Graduate Diploma, Statistical Theory & Methods, Paper II, 2003. Question 8

Topics to be included in a comprehensive answer include the following, and suitable examples should be given.

Parametric tests are based on assumptions about the values of the parameters in mass or density functions for a family of distributions, for example N(μ, σ^2) or B(n, p), and confidence interval methods use the same theory.

Parametric methods often use a likelihood function based on an assumed model, for example in a likelihood ratio test to compare hypotheses about a parameter in (say) a gamma family.

Moments of a distribution, especially mean and variance, are often used in parametric methods, whereas order statistics (median etc) are more useful for non-parametric inference.

It is less easy to construct confidence-limit arguments in non-parametric inference.

Non-parametric methods need fewer assumptions, for example not requiring a specific distribution as a model.

Prior information for parametric methods includes a model and some values for its parameters, whereas merely the value of an order statistic is often sufficient in a non-parametric test.

Exact probability theory based on samples from Normal distributions can be used for parametric methods, whereas approximate methods are more common for non-parametric methods.

Computing of critical value tables for non-parametric tests is often very complex compared with that required for parametric tests, although some good Normal approximations exist for moderate-sized samples in some standard non-parametric tests.

If both types of test are possible for a set of data (for example a two-sample test), the parametric one is more powerful (provided the underlying modelling assumptions are satisfied) but the non-parametric one may be more robust (in case the assumptions are not).

Ranked (non-numerical) data need the non-parametric approach.