THE ROYAL STATISTICAL SOCIETY

2009 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

MODULAR FORMAT

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, 2009. Question 1

(i) $E\left(\sum_{i=1}^{n} W_i\right) = \sum_{i=1}^{n} E(W_i) = n\mu$.

$E(W_i^2) = \text{Var}(W_i) + (E(W_i))^2 = \sigma^2 + \mu^2$.

$$\therefore E\left(\sum_{i=1}^{n} W_i^2\right) = \sum_{i=1}^{n} E(W_i^2) = n(\sigma^2 + \mu^2).$$

For $i = 1, 2, \ldots, n - 1$, we have

$$\text{Cov}(W_i, W_{i+1}) = \rho\sigma^2 = E(W_i W_{i+1}) - E(W_i)E(W_{i+1}),$$

and therefore $E(W_i W_{i+1}) = \rho\sigma^2 + \mu^2$.

$$\therefore E\left(\sum_{i=1}^{n-1} W_i W_{i+1}\right) = (n-1)(\rho\sigma^2 + \mu^2), \quad \text{as required.}$$

(ii) Method of moments estimators are obtained as follows.

$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} W_i$.

We have $\hat{\sigma}^2 + \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} W_i^2$, so $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} W_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} W_i\right)^2$.

We have $\hat{\rho}\hat{\sigma}^2 + \hat{\mu}^2 = \frac{1}{n-1}\sum_{i=1}^{n-1} W_i W_{i+1}$, so $\hat{\rho} = \dfrac{\frac{1}{n-1}\sum_{i=1}^{n-1} W_i W_{i+1} - \left(\frac{1}{n}\sum_{i=1}^{n} W_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} W_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} W_i\right)^2}$.

Solution continued on next page

(iii) $E\left(\hat{\sigma}^2\right) = E\left(\frac{1}{2}\left(W_1^2 + W_2^2\right)\right) - E\left[\left(\frac{1}{2}\left(W_1 + W_2\right)\right)^2\right]$

$$= \sigma^2 + \mu^2 - \frac{1}{4}\left(E\left(W_1^2\right) + E\left(W_2^2\right) + 2E\left(W_1W_2\right)\right)$$

$$= \sigma^2 + \mu^2 - \frac{1}{4}\left(2\sigma^2 + 2\mu^2 + 2\rho\sigma^2 + 2\mu^2\right) = \sigma^2\left(1 - \frac{1}{2} - \frac{1}{2}\rho\right) = \frac{\sigma^2\left(1-\rho\right)}{2}.$$

$$\therefore \text{Bias} = E\left(\hat{\sigma}^2\right) - \sigma^2 = -\frac{\sigma^2\left(1+\rho\right)}{2}.$$

(iv) $\text{Var}\left(W_1 + W_2\right) = \text{Var}\left(W_1\right) + \text{Var}\left(W_2\right) + 2\text{Cov}\left(W_1, W_2\right) = 2\sigma^2 + 2\rho\sigma^2.$

$$\therefore \text{Var}\left(\hat{\mu}\right) = \text{Var}\left(\frac{1}{2}\left(W_1 + W_2\right)\right) = \frac{\sigma^2\left(1+\rho\right)}{2}.$$

(v) If $W_1 = W_2$, $\hat{\rho}$ is undefined. Assume then that $W_1 \neq W_2$. We then have

$$\hat{\rho} = \frac{W_1W_2 - \frac{1}{4}\left(W_1^2 + W_2^2 + 2W_1W_2\right)}{\frac{1}{2}\left(W_1^2 + W_2^2\right) - \frac{1}{4}\left(W_1^2 + W_2^2 + 2W_1W_2\right)} = \frac{\frac{1}{2}W_1W_2 - \frac{1}{4}\left(W_1^2 + W_2^2\right)}{-\frac{1}{2}W_1W_2 + \frac{1}{4}\left(W_1^2 + W_2^2\right)} = -1.$$

Clearly this is not an estimator to be relied on. It is not possible to obtain a sensible estimate of a correlation based on only two observations.

Graduate Diploma, Module 2, 2009. Question 2

$$P(N=n)=\frac{p^n}{-n\log(1-p)}$$

(i) $$E(N) \;=\; -\frac{1}{\log(1-p)}\sum_{n=1}^{\infty}\frac{np^n}{n} \;=\; -\frac{1}{\log(1-p)}\sum_{n=1}^{\infty}p^n \;=\; -\frac{p}{(1-p)\log(1-p)}.$$

(ii) For independent observations $N_1, N_2, \ldots, N_{40}$, the likelihood is

$$L(p)=\frac{p^{\Sigma N_i}}{\prod N_i.(-\log(1-p))^{40}}.$$

$$\therefore \log L(p)=\Sigma N_i \log p-40\log(-\log(1-p))-\log(\Pi N_i).$$

$$\therefore \frac{d\log L}{dp}=\frac{\Sigma N_i}{p}+\left(\frac{40}{\log(1-p)}\times\frac{1}{1-p}\right).$$

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

$$\frac{\Sigma N_i}{\hat{p}}+\frac{40}{(1-\hat{p})\log(1-\hat{p})}=0.$$

(iii) The Fisher information is $I=-E\left(\frac{d^2\log L}{dp^2}\right)$ (the second derivative is quoted in the question)

$$=-\left(-\frac{\sum_{i=1}^{40}E(N_i)}{p^2}+\frac{40(1+\log(1-p))}{[(1-p)\log(1-p)]^2}\right)$$

$$=\frac{-40p}{p^2(1-p)\log(1-p)}-\frac{40(1+\log(1-p))}{[(1-p)\log(1-p)]^2}$$

$$=\frac{40(-(1-p)\log(1-p)-p-p\log(1-p))}{p[(1-p)\log(1-p)]^2}$$

$$=\frac{40(-\log(1-p)-p)}{p[(1-p)\log(1-p)]^2}.$$

Solution continued on next page

Therefore an approximate 95% confidence interval for p is given by

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\left[(1-\hat{p})\log(1-\hat{p})\right]^2}{40\left(-\log(1-\hat{p})-\hat{p}\right)}}\ .$$

If $\hat{p} = 0.8$, this confidence interval is $0.8 \pm 1.96\sqrt{\dfrac{0.8\left(0.2\log(0.2)\right)^2}{40\left(-\log(0.2)-0.8\right)}}$,

i.e. $0.8 \pm (1.96 \times 0.0506)$, i.e. 0.8 ± 0.099 or (0.701, 8.899).

(iv) We use the Newton-Raphson method, starting from $\hat{p}_0 = 0.75$.

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

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

We have, inserting $\hat{p}_0 = 0.75$ and $\Sigma N_i = 100$,

$$\left.\frac{d\log L}{dp}\right|_{p=\hat{p}_0} = \frac{100}{0.75} + \frac{40}{0.25\log(0.25)} = 17.918\,,$$

and

$$\left.\frac{d^2\log L}{dp^2}\right|_{p=\hat{p}_0} = -\frac{100}{0.75^2} + \frac{40\left(1+\log(0.25)\right)}{\left(0.25\log(0.25)\right)^2} = -306.421\,.$$

$$\therefore\ \hat{p}_1 = 0.75 - \frac{17.918}{-306.421} = 0.75 + 0.0585 = 0.808.$$

<u>Graduate Diploma, Module 2, 2009. Question 3</u>

A random interval (X_1, X_2) is a 95% confidence interval for θ if $P\left(X_1 < \theta < X_2\right) = 0.95$ for all possible values of θ.

(i) $$f\left(x_1, ..., x_n\right) = \frac{\left(\Pi x_i\right)^{k-1} e^{-\frac{1}{\alpha}\Sigma x_i}}{\left((k-1)!\right)^n \alpha^{nk}} = \left(\frac{e^{-\frac{1}{\alpha}\Sigma x_i}}{\alpha^{nk}}\right)\left(\frac{\left(\Pi x_i\right)^{k-1}}{\left((k-1)!\right)^n}\right)$$

$$g\left(\Sigma x_i ; \alpha k\right) \quad h(x_1, \ldots, x_n)$$

Since the joint density is the product of a factor not involving the parameter α and a factor only dependent on the observations x_1, …, x_n through Σx_i, it follows by the factorisation theorem that $Y = \Sigma X_i$ is sufficient for α.

(ii) First, the moment generating function (mgf) of $Y = \sum_{i=1}^{n} X_i$ is

$$M_Y\left(t\right) = \prod_{i=1}^{n}\left(\text{mgf of } X_i\right) = \left(1-\alpha t\right)^{-nk} \qquad (\text{for } t < \alpha^{-1}).$$

Now writing $W = \dfrac{2Y}{\alpha}$, we have that the mgf of W is

$$M_W\left(t\right) = E\left(e^{tW}\right) = E\left(e^{\frac{2tY}{\alpha}}\right) = E\left(e^{Y\left(\frac{2t}{\alpha}\right)}\right) = M_Y\left(\frac{2t}{\alpha}\right),$$

i.e. $$M_W\left(t\right) = \left(1-\alpha\frac{2t}{\alpha}\right)^{-nk} = \left(1-2t\right)^{-nk},$$

and this is the mgf of χ^2_{2nk}. Therefore, by the 1:1 correspondence of mgfs and distributions, $W \sim \chi^2_{2nk}$.

(iii) Use the standard χ^2_{2nk} tables to find r_1 satisfying $P\left(\chi^2_{2nk} < r_1\right) = 0.025$ and r_2 satisfying $P\left(\chi^2_{2nk} < r_2\right) = 0.975$. Then we have $P\left(r_1 < \dfrac{2Y}{\alpha} < r_2\right) = 0.95$, for all α. Thus a 95% confidence interval for α is $\left(\dfrac{2Y}{r_2}, \dfrac{2Y}{r_1}\right)$.

Solution continued on next page

(iv) $n = 10$, $k = 3$. So the number of degrees of freedom is $2 \times 10 \times 3 = 60$. From χ^2 tables, we have $r_1 = 40.482$ and $r_2 = 83.298$.

$\therefore$ the 95% confidence interval for α is

$$\frac{2\Sigma X_i}{83.298} \text{ to } \frac{2\Sigma X_i}{40.482}, \quad \text{i.e. } \frac{\Sigma X_i}{41.649} \text{ to } \frac{\Sigma X_i}{20.241}.$$

To find the expected length of this interval, we need $E(\Sigma X_i)$, i.e. $E(Y)$ as defined above. Using the mgf of Y, this is the derivative of $M_Y(t)$ at $t = 0$.

We have

$$\frac{dM_Y(t)}{dt} = nk\alpha(1-\alpha t)^{-nk-1}.$$

which, on inserting $t = 0$ together with $n = 10$ and $k = 3$, gives simply 30α.

$\therefore$ the expected length of the 95% confidence interval is

$$30\alpha\left(\frac{1}{20.241} - \frac{1}{41.649}\right) = 0.7618\alpha .$$

<u>Graduate Diploma, Module 2, 2009. Question 4</u>

Suppose that θ is the parameter of a distribution. We want to test

$$H_0: \theta = \theta_0 \qquad H_1: \theta = \theta_1$$

where θ_0 and $\theta_1 \left(\neq \theta_0\right)$ are given.

Let α be the required significance level. The Neyman-Pearson approach is to choose the test with the largest power at θ_1, subject to its size being $\leq \alpha$. The Neyman-Pearson lemma shows that this property is satisfied by a likelihood ratio test.

(i) $$f\left(x_i\right) = \frac{e^{-\mu(1+a_i\theta)}\left(\mu\left(1+a_i\theta\right)\right)^{x_i}}{x_i!}.$$

The likelihood is $L\left(\mu,\theta\right) = \prod_{i=1}^{n} f\left(x_i\right) = \dfrac{e^{-\mu\left(n+\theta\Sigma a_i\right)}\mu^{\Sigma x_i}\Pi\left(1+a_i\theta\right)^{x_i}}{\Pi\left(x_i!\right)}$.

The likelihood ratio is

$$\frac{L\left(10,1\right)}{L\left(10,0\right)} = \frac{e^{-10n-10\Sigma a_i}10^{\Sigma x_i}\Pi\left(1+a_i\right)^{x_i}/\Pi\left(x_i!\right)}{e^{-10n}10^{\Sigma x_i}\Pi\left(1\right)^{x_i}/\Pi\left(x_i!\right)} = e^{-10\Sigma a_i}\prod_{i=1}^{n}\left(1+a_i\right)^{x_i}.$$

Therefore the critical region consists of values of the x_i such that

$e^{-10\Sigma a_i}\prod_{i=1}^{n}\left(1+a_i\right)^{x_i} \geq k$, where k is a constant,

i.e. such that $\Sigma x_i \log\left(1+a_i\right) \geq c$, where c is a constant.

(ii) Under H_0 ($\mu = 10$ and $\theta = 0$), we have $E(X_i) = \text{Var}(X_i) = 10$ and thus

$E(X_i \log(1 + a_i)) = 10\log(1 + a_i)$ and $\text{Var}(X_i \log(1 + a_i)) = 10(\log(1 + a_i))^2$.

Therefore, by the central limit theorem,

$\Sigma X_i \log\left(1+a_i\right) \sim \text{N}\left(10\Sigma\log\left(1+a_i\right)\right), 10\Sigma\left(\log\left(1+a_i\right)\right)^2$ under H_0.

We require $P\left(\Sigma X_i \log\left(1+a_i\right) \geq c \mid H_0\right) = 0.05$. Thus we have

$$\therefore c = 10\Sigma\log\left(1+a_i\right) + 1.645\sqrt{10\Sigma\left(\log\left(1+a_i\right)\right)^2}.$$

Solution continued on next page

(iii) (a) The likelihood ratio is

$$\frac{L(\mu,0)}{L(10,0)}=\frac{e^{-n\mu}\mu^{\Sigma x_i}/\Pi(x_i!)}{e^{-10n}10^{\Sigma x_i}/\Pi(x_i!)}=e^{-n(\mu-10)}\left(\frac{\mu}{10}\right)^{\Sigma x_i}.$$

Therefore the critical region consists of values such that

$$e^{-n(\mu-10)}\left(\frac{\mu}{10}\right)^{\Sigma x_i}\geq c$$

i.e. such that $\Sigma x_i \geq c'$ (since $\mu > 10$). Thus the same form of test is obtained for all $\mu > 10$, so this test is uniformly most powerful.

(b) The likelihood ratio is

$$\frac{L(10,\theta)}{L(10,0)}=\frac{e^{-10(n+\theta\Sigma a_i)}10^{\Sigma x_i}\,\Pi(1+a_i\theta)^{x_i}/\Pi(x_i!)}{e^{-10n}10^{\Sigma x_i}/\Pi(x_i!)}$$

$$=e^{-10\theta\Sigma a_i}\Pi(1+a_i\theta)^{x_i}.$$

Therefore the critical region consists of values such that

$$\Sigma x_i \log(1+a_i\theta)\geq c.$$

Different tests are obtained for different values of θ so there is no uniformly most powerful test.

Graduate Diploma, Module 2, 2009. Question 5

(i) Let $\hat{\theta}_{\backslash i}$ be the corresponding estimator based on the $n - 1$ observations with X_i missing, i.e. $\hat{\theta}_{\backslash i} = \hat{\theta}_{n-1}\left(X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_n\right)$, for $i = 1, 2, \ldots, n$.

Now define $\tilde{\theta}_i = n\hat{\theta}_n - (n-1)\hat{\theta}_{\backslash i}$, for $i = 1, 2, \ldots, n$.

The jack-knife estimator is then given by $\tilde{\theta}_J = \frac{1}{n}\sum_{i=1}^{n}\tilde{\theta}_i$.

We have $E\left(\hat{\theta}_n\right) = \theta + \frac{k}{n}$, and so $E\left(\hat{\theta}_{\backslash i}\right) = \theta + \frac{k}{n-1}$.

$$\therefore E\left(\tilde{\theta}_i\right) = nE\left(\hat{\theta}_n\right) - (n-1)E\left(\hat{\theta}_{\backslash i}\right) = n\theta + k - (n-1)\theta - k = \theta .$$

$\therefore E\left(\tilde{\theta}_J\right) = \frac{1}{n}\sum E\left(\tilde{\theta}_i\right) = \frac{1}{n}\sum_{i=1}^{n}\theta = \theta$, i.e. $\tilde{\theta}_J$ is an unbiased estimator of θ.

(ii) (a) We use $\hat{c}_n = \frac{S}{\bar{X}} = \dfrac{\sqrt{\dfrac{U - (T^2/n)}{n-1}}}{\dfrac{T}{n}} = \frac{n}{\sqrt{n-1}}\sqrt{\frac{U}{T^2} - \frac{1}{n}}$.

For the sample with X_i missing, i.e. $X_1, ..., X_{i-1}, X_{i+1}, ..., X_n$, the sum of the observations is $T - X_i$ and the sum of the squares of the observations is $U - X_i^2$.

$$\therefore \hat{c}_{\backslash i} = \frac{n-1}{\sqrt{n-2}}\sqrt{\frac{U - X_i^2}{(T - X_i)^2} - \frac{1}{n-1}} .$$

$$\therefore \tilde{c}_i = \frac{n^2}{\sqrt{n-1}}\sqrt{\frac{U}{T^2} - \frac{1}{n}} - \frac{(n-1)^2}{\sqrt{n-2}}\sqrt{\frac{U - X_i^2}{(T - X_i)^2} - \frac{1}{n-1}} .$$

Therefore the jack-knife estimator $\tilde{c}$ is

$$\therefore \tilde{c} = \frac{1}{n}\sum_{i=1}^{n}\tilde{c}_i = \frac{n^2}{\sqrt{n-1}}\sqrt{\frac{U}{T^2} - \frac{1}{n}} - \frac{(n-1)^2}{n\sqrt{n-2}}\sum_{i=1}^{n}\sqrt{\frac{U - X_i^2}{(T - X_i)^2} - \frac{1}{n-1}} .$$

Solution continued on next page

(b) Let $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n}\left(\tilde{c}_i - \tilde{c}\right)^2}{n-1}$.

Then an approximate 95% confidence interval for the coefficient of variation is $\tilde{c} \pm t\sqrt{\dfrac{\hat{\sigma}^2}{n}}$ where t is the upper 2.5% point of t_{n-1}.

(c) Take say 1000 bootstrap samples. In the ith sample, sample n values at random with replacement from $X_1, X_2, ..., X_n$ and use these values to find $c_i^* = \dfrac{\text{sample s.d.}}{\text{sample mean}}$.

Order the 1000 estimates: $c_{(1)}^* < c_{(2)}^* < ... < c_{(1000)}^*$.

Then an approximate 95% confidence interval is $c_{(25)}^*$ to $c_{(975)}^*$.

Graduate Diploma, Module 2, 2009. Question 6

$$f(y_i) = \frac{1}{2}\alpha e^{-\alpha|y_i|} \qquad (i = 1, 2, ..., m)$$

(i) The likelihood is $L(\alpha) = \left(\frac{1}{2}\right)^m \alpha^m e^{-\alpha\Sigma|y_i|}$.

$\therefore$ the log likelihood is $\log L(\alpha) = m\log\left(\frac{1}{2}\right) + m\log\alpha - \alpha\,\Sigma|y_i|$.

$\frac{d\log L}{d\alpha} = \frac{m}{\alpha} - \Sigma|y_i|$ which on setting equal to zero gives solution $\hat{\alpha} = \frac{m}{\Sigma|y_i|}$.

To investigate whether this is a maximum, consider $\frac{d^2\log L}{d\alpha^2} = -\frac{m}{\alpha^2} < 0$.

$\therefore \hat{\alpha} = m/\Sigma|y_i|$ maximises $\log L(\alpha)$; thus $m/\Sigma|y_i|$ is the maximum likelihood estimator of α.

For $H_0: \alpha = 2$, $H_1: \alpha \neq 2$, the generalised likelihood ratio test has critical region given by

$$-2\left(\log L(2) - \log L(\hat{\alpha})\right) \geq k \qquad \text{(for some constant } k\text{)},$$

i.e. $$-2\left(m\log 2 - 2\Sigma|y_i| - m\log\left(\frac{m}{\Sigma|y_i|}\right) + m\right) \geq k$$

i.e. $$2\Sigma|y_i| - m\log\left(\Sigma|y_i|\right) \geq k' \qquad \text{(for some constant } k'\text{)}.$$

(ii) When m is large, $-2\left(\log L(2) - \log L(\hat{\alpha})\right) \sim \chi_1^2$, approximately, under H_0.

The 95% point of χ_1^2 is 3.841. So choose k in the above equal to 3.841.

Solution continued on next page

(iii) $H_0 : \alpha = \beta, \quad H_1 : \alpha \neq \beta$.

As above, $\hat{\beta} = \dfrac{n}{\Sigma|w_i|}$ and, under H_0, $\hat{\hat{\alpha}} = \hat{\hat{\beta}} = \dfrac{m+n}{\Sigma|y_i| + \Sigma|w_i|}$.

For this generalised likelihood ratio test, the critical region is given by

$$-2\left\{\log L\left(\hat{\hat{\alpha}}, \hat{\hat{\beta}}\right) - \log L\left(\hat{\alpha}, \hat{\beta}\right)\right\} \geq k$$

i.e. $-2\Big\{m\log\left(\hat{\hat{\alpha}}\right) + n\log\left(\hat{\hat{\alpha}}\right) - \hat{\hat{\alpha}}\,\Sigma|y_i| - \hat{\hat{\alpha}}\,\Sigma|w_i|$

$$-m\log\left(\hat{\alpha}\right) + \hat{\alpha}\,\Sigma|y_i| - n\log\left(\hat{\beta}\right) + \hat{\beta}\,\Sigma|w_i|\Big\} \geq k$$

i.e. $-2\left\{(m+n)\log\left(\dfrac{m+n}{\Sigma|y_i| + \Sigma|w_i|}\right) - (m+n)\right.$

$$\left. - m\log\left(\frac{m}{\Sigma|y_i|}\right) + m - n\log\left(\frac{n}{\Sigma|w_i|}\right) + n\right\} \geq k$$

i.e. $-2\left\{(m+n)\log\left(\dfrac{m+n}{\Sigma|y_i| + \Sigma|w_i|}\right) - m\log\left(\dfrac{m}{\Sigma|y_i|}\right) - n\log\left(\dfrac{n}{\Sigma|w_i|}\right)\right. \geq k$.

(iv) There is one constraint under H_0. $\therefore$ $k = 3.841$ (as above).

Inserting the given values in the left-hand side of the above inequality gives

$$-2\left(300\log\left(\frac{300}{140}\right) - 100\log\left(\frac{100}{40}\right) - 200\log\left(\frac{200}{100}\right)\right)$$

which equals 3.233. Since 3.233 < 3.841, there is not significant evidence against H_0 at the 5% level.

Graduate Diploma, Module 2, 2009. Question 7

(a) Prior $\pi(p_1) = 6p_1(1-p_1) \quad (0 < p_1 < 1)$.

(i) Let X_1 = number in sample supporting Candidate 1.

We have $X_1 \sim \mathrm{B}(n, p_1)$, so $f(x_1|p_1) = \binom{n}{x_1} p_1^{x_1}(1-p_1)^{n-x_1}$.

The posterior density is

$$f(p_1|x_1) \propto \binom{n}{x_1} p_1^{x_1}(1-p_1)^{n-x_1} \times 6p_1(1-p_1) \propto p_1^{x_1+1}(1-p_1)^{n-x_1+1}.$$

We note from the information in the question that this is a beta distribution with $\alpha_1 = x_1 + 2$ and $\alpha_2 = n + 2 - x_1$.

(ii) For a large sample, the posterior distribution has approximately a Normal distribution. From the information in the question,

$$\text{mean} = \frac{x_1+2}{x_1+2+n+2-x_1} = \frac{x_1+2}{n+4},$$

$$\text{variance} = \frac{(x_1+2)(n+2-x_1)}{(n+5)(n+4)^2}.$$

So an approximate Bayesian 95% interval for p_1 is given by

$$\frac{x_1+2}{n+4} \pm 1.96\sqrt{\frac{(x_1+2)(n+2-x_1)}{(n+5)(n+4)^2}}.$$

Solution continued on next page

(b) Prior $\pi(p_1, p_2, p_3) \propto p_1$.

(i) $P(X_1 = x_1, X_2 = x_2, X_3 = x_3) \propto p_1^{x_1} p_2^{x_2} p_3^{x_3}$.

The posterior joint distribution is simply proportional to $p_1^{x_1+1} p_2^{x_2} p_3^{x_3}$, i.e. from the information in the question it is a Dirichlet distribution with $\alpha_1 = x_1 + 2$, $\alpha_2 = x_2 + 1$ and $\alpha_3 = x_3 + 1$.

(ii) With respect to the posterior, using the information in the question,

$$E(p_1 - p_2) = E(p_1) - E(p_2) = \frac{x_1 + 2}{n+4} - \frac{x_2 + 1}{n+4} = \frac{x_1 - x_2 + 1}{n+4},$$

$$\begin{aligned}\text{Var}(p_1 - p_2) &= \text{Var}(p_1) + \text{Var}(p_2) - 2\text{Cov}(p_1, p_2) \\ &= \frac{(x_1+2)(x_2+x_3+2) + (x_2+1)(x_1+x_3+3) + 2(x_1+2)(x_2+1)}{(n+4)^2(n+5)}.\end{aligned}$$

So an approximate Bayesian 95% interval for $p_1 - p_2$ is given by

$$\frac{x_1 - x_2 + 1}{n+4} \pm 1.96\sqrt{\text{Var}(p_1 + p_2)}.$$

Graduate Diploma, Module 2, 2009. Question 8

Decision making: the actions "accept H_0" or "accept H_1" must be taken after analysing the data. Usually no wider issues are involved; the data are relevant only to the immediate situation (e.g. quality control – either want to stop the production line or let it continue).

Strength of evidence: it is not necessarily expected that the current experiment will lead to immediate actions, rather that it will add to previously gained information. Wider issues are involved and it is often felt important that significant evidence is found in several independent studies (e.g. at independent centres). In principle, p-values can be combined (meta-analysis). One application is clinical trials.

The contrast should not be taken too far. In the former case, a value near the critical value ("just accept" H_0 or "just accept" H_1) may lead to a suspension of action until further evidence is obtained. On the other hand, a "very significant" result in the second case may lead to immediate action.

Significance level: in decision making, this will deliberately be chosen to reflect the "cost" of wrongly rejecting H_0 (e.g. stopping the production line when nothing is wrong). In the strength of evidence approach, it is customary to use one of the traditional values (e.g. 0.05), or to quote the exact p-value.

Sample size: in decision making, this will be deliberately chosen to reflect the cost of making wrong decisions (e.g. continuing operating the production line when in fact there is a fault). In the strength of evidence approach, it is common practice to ensure that the sample size is sufficiently large that the power of detecting an effect of practical importance is sufficiently high.