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}.
(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 \[ E(W_i W_{i+1}) = \rho \sigma^2 + \mu^2. \]

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

(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 \( \hat{\rho} = \frac{1}{\frac{1}{n-1} \sum_{i=1}^{n-1} W_i W_{i+1} - \left( \frac{1}{n} \sum_{i=1}^{n} W_i \right)^2}. \)
(iii) \[ E(\hat{\sigma}^2) = E\left(\frac{1}{2}(W_1^2 + W_2^2)\right) - E\left[\frac{1}{2}(W_1 + W_2)^2\right] \]
\[ = \sigma^2 + \mu^2 - \frac{1}{4}\left(E(W_1^2) + E(W_2^2) + 2E(W_1W_2)\right) \]
\[ = \sigma^2 + \mu^2 - \frac{1}{4}\left(2\sigma^2 + 2\mu^2 + 2\rho\sigma^2 + 2\mu^2\right) = \sigma^2\left(1 - \frac{1}{2} - \frac{1}{2}\rho\right) = \frac{\sigma^2(1-\rho)}{2}. \]
\[ \therefore \text{Bias} = E(\hat{\sigma}^2) - \sigma^2 = -\frac{\sigma^2(1+\rho)}{2}. \]

(iv) \[ \text{Var}(W_1 + W_2) = \text{Var}(W_1) + \text{Var}(W_2) + 2\text{Cov}(W_1, W_2) = 2\sigma^2 + 2\rho\sigma^2. \]
\[ \therefore \text{Var}(\hat{\mu}) = \text{Var}\left(\frac{1}{2}(W_1 + W_2)\right) = \frac{\sigma^2(1+\rho)}{2}. \]

(v) If \( W_1 = W_2 \), \( \hat{\rho} \) is undefined. Assume then that \( W_1 \neq W_2 \). We then have
\[ \hat{\rho} = \frac{W_1W_2 - \frac{1}{4}(W_1^2 + W_2^2 + 2W_1W_2)}{\frac{1}{2}(W_1^2 + W_2^2) - \frac{1}{4}(W_1^2 + W_2^2 + 2W_1W_2)} = \frac{1}{2}W_1W_2 - \frac{1}{4}(W_1^2 + W_2^2) = -1. \]

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.
\[ P(N = n) = \frac{p^n}{-n \log(1 - p)} \]

(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

\[ L(p) = \frac{p^{\sum N_i}}{\prod N_i (-\log(1 - p))^{40}}. \]

\[ \therefore \log L(p) = \sum N_i \log p - 40 \log (-\log(1 - p)) - \log (\Pi N_i). \]

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

The maximum likelihood estimator \( \hat{p} \) therefore satisfies

\[ \frac{\sum N_i}{\hat{p}} + \frac{40}{(1 - \hat{p}) \log(1 - \hat{p})} = 0. \]

(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

\[
\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}[(1-\hat{p})\log(1-\hat{p})]}{40(-\log(1-\hat{p})-\hat{p})}}.
\]

If \( \hat{p} = 0.8 \), this confidence interval is

\[
0.8 \pm 1.96\sqrt{\frac{0.8(0.2\log(0.2))^2}{40(-\log(0.2)-0.8)}}
\]

i.e. \( 0.8 \pm (1.96 \times 0.0506) \), i.e. \( 0.8 \pm 0.099 \) 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.

\[
\hat{p}_1 = \hat{p}_0 - \frac{d \log L}{d \hat{p}} \bigg|_{p=\hat{p}_0} - \frac{d^2 \log L}{d \hat{p}^2} \bigg|_{p=\hat{p}_0}.
\]

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

\[
\frac{d \log L}{dp} \bigg|_{p=\hat{p}_0} = \frac{100}{0.75} + \frac{40}{0.25\log(0.25)} = 17.918,
\]

and

\[
\frac{d^2 \log L}{dp^2} \bigg|_{p=\hat{p}_0} = -\frac{100}{0.75^2} + \frac{40(1+\log(0.25))}{(0.25\log(0.25))^2} = -306.421.
\]

\[
\therefore \hat{p}_1 = 0.75 - \frac{17.918}{-306.421} = 0.75 + 0.0585 = 0.808.
\]
A random interval \((X_1, X_2)\) is a 95% confidence interval for \(\theta\) if:

\[
P(X_1 < \theta < X_2) = 0.95
\]

for all possible values of \(\theta\).

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

\[
g(\Sigma x_i; \alpha k) \ h(x_1, \ldots, x_n)
\]

Since the joint density is the product of a factor not involving the parameter \(\alpha\) and a factor only dependent on the observations \(x_1, \ldots, x_n\) through \(\Sigma x_i\), it follows by the factorisation theorem that \(Y = \Sigma X_i\) is sufficient for \(\alpha\).

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

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

Now writing \(W = \frac{2Y}{\alpha}\), we have that the mgf of \(W\) is

\[
M_W(t) = E\left(e^{\frac{2Y}{\alpha}}\right) = E\left(e^{\frac{2Y}{\alpha}}\right) = M_Y\left(\frac{2t}{\alpha}\right),
\]

i.e.

\[
M_W(t) = \left(1 - \alpha \frac{2t}{\alpha}\right)^{-nk} = (1 - 2t)^{-nk},
\]

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

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

**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 \( \chi^2 \) tables, we have \( r_1 = 40.482 \) and \( r_2 = 83.298 \).

\[
\therefore \text{the 95\% confidence interval for } \alpha \text{ is}
\frac{2\Sigma X_i}{83.298} \text{ to } \frac{2\Sigma X_i}{40.482}, \quad \text{i.e. } \frac{\Sigma X_i}{41.649} \text{ to } \frac{\Sigma X_i}{20.241}.
\]

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

which, on inserting \( t = 0 \) together with \( n = 10 \) and \( k = 3 \), gives simply \( 30\alpha \).

\[
\therefore \text{the expected length of the 95\% confidence interval is}
30\alpha \left( \frac{1}{20.241} - \frac{1}{41.649} \right) = 0.7618\alpha.
\]
Suppose that $\theta$ is the parameter of a distribution. We want to test

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

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

Let $\alpha$ be the required significance level. The Neyman-Pearson approach is to choose the test with the largest power at $\theta_1$, subject to its size being $\leq \alpha$. The Neyman-Pearson lemma shows that this property is satisfied by a likelihood ratio test.

(i) 

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

The likelihood is 

$$L(\mu, \theta) = \prod_{i=1}^{n} f(x_i) = \frac{e^{-\mu(n+\theta \sum a_i)} \mu^{\Sigma x_i} \prod (1+a_i \theta)^{x_i}}{\prod (x_i !)}.$$ 

The likelihood ratio is 

$$\frac{L(10, 1)}{L(10, 0)} = \frac{e^{-10n-10\sum a_i} 10^{\sum x_i} \prod (1+a_i)_{x_i} / \prod (x_i !)}{e^{-10n} 10^{\sum x_i} \prod (1)^{x_i} / \prod (x_i !)} = e^{-10\sum a_i} \prod_{i=1}^{n} (1+a_i)^{x_i}.$$ 

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

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

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

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

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

Therefore, by the central limit theorem, 

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

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

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

Solution continued on next page
The likelihood ratio is
\[
\frac{L(\mu, 0)}{L(10, 0)} = \frac{e^{-n \mu} \mu^{\sum x_i} / \Pi(x_i !)}{e^{-10n \mu 10^{\sum x_i}} / \Pi(x_i !)} = e^{-n(\mu-10)} \left( \frac{\mu}{10} \right)^{\sum x_i}.
\]

Therefore the critical region consists of values such that
\[
e^{-n(\mu-10)} \left( \frac{\mu}{10} \right)^{\sum x_i} \geq c
\]
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.

The likelihood ratio is
\[
\frac{L(10, \theta)}{L(10, 0)} = \frac{e^{-10(n+\theta \Sigma x_i)} 10^{\sum x_i} \Pi(1+a \theta)^{x_i} / \Pi(x_i !)}{e^{-10n \mu 10^{\sum x_i}} / \Pi(x_i !)} = e^{-10\theta \Sigma x_i} \Pi(1+a \theta)^{x_i}.
\]

Therefore the critical region consists of values such that
\[
\Sigma x_i \log(1+a \theta) \geq c.
\]

Different tests are obtained for different values of \( \theta \) so there is no uniformly most powerful test.
(i) Let $\hat{\theta}_i$ be the corresponding estimator based on the $n-1$ observations with $X_i$ missing, i.e. $\hat{\theta}_i = \hat{\theta}_{n-1}(X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_n)$, for $i = 1, 2, ..., n$.

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

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

We have $E(\hat{\theta}_i) = \theta + \frac{k}{n}$, and so $E(\hat{\theta}_i) = \theta + \frac{k}{n-1}$.

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

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

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

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 $U - X_i^2$.

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

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

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

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

Solution continued on next page
(b) Let \( \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (\bar{c}_i - \bar{c})^2}{n-1} \).

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

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

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

Then an approximate 95% confidence interval is \( c_{(25)}^* \) to \( c_{(975)}^* \).
Graduate Diploma, Module 2, 2009. Question 6

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

(i) The likelihood is

\[ L(\alpha) = \left(\frac{1}{2}\right)^m \alpha^m e^{-\alpha \sum |y_i|} \]

\[ \therefore \text{the log likelihood is} \quad \log L(\alpha) = m \log \left(\frac{1}{2}\right) + m \log \alpha - \alpha \sum |y_i| \]

\[ \frac{d \log L}{d \alpha} = \frac{m}{\alpha} - \sum |y_i| \] which on setting equal to zero gives solution \[ \hat{\alpha} = \frac{m}{\sum |y_i|} \]

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

\[ \therefore \hat{\alpha} = \frac{m}{\sum |y_i|} \] maximises \( \log L(\alpha) \); thus \( \frac{m}{\sum |y_i|} \) is the maximum likelihood estimator of \( \alpha \).

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

\[ -2 \left( \log L(2) - \log L(\hat{\alpha}) \right) \geq k \quad \text{(for some constant } k) \]

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

i.e. \[ 2 \sum |y_i| - m \log \left( \sum |y_i| \right) \geq k' \quad \text{(for some constant } k') \].

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

The 95% point of \( \chi^2 \) 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, \( \hat{\beta} = \frac{n}{\Sigma |w_i|} \) and, under \( H_0, \) \( \hat{\alpha} = \hat{\beta} = \frac{m+n}{\Sigma |v_i| + \Sigma |w_i|}. \)

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

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

i.e.

\[
-2 \left\{ m \log \left( \hat{\alpha} \right) + n \log \left( \hat{\beta} \right) - \hat{\alpha} \Sigma |v_i| - \hat{\alpha} \Sigma |w_i| \right\}
- m \log \left( \hat{\alpha} \Sigma |v_i| \right) - n \log \left( \hat{\beta} \Sigma |w_i| \right) \geq k
\]

i.e. \(-2\left\{ (m+n) \log \left( \frac{m+n}{\Sigma |v_i| + \Sigma |w_i|} \right) - (m+n) \right\}
- m \log \left( \frac{m}{\Sigma |v_i|} \right) + m - n \log \left( \frac{n}{\Sigma |w_i|} \right) + n \right\} \geq k
\]

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

(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 $\pi(p_1) = 6p_1(1-p_1)$ \hspace{0.5cm} (0 < p_1 < 1).

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

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

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,

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

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

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 $p_1^{x_1+1} p_2^{x_2+1} p_3^{x_3+1}$, 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,

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

$$\text{Var}(p_1 - p_2) = \text{Var}(p_1) + \text{Var}(p_2) - 2\text{Cov}(p_1, p_2)$$

$$= \frac{(x_1 + 2)(x_2 + 3) + (x_2 + 1)(x_1 + 3) + 2(x_1 + 2)(x_2 + 1)}{(n+4)^2(n+5)}.$$

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

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