HONG KONG STATISTICAL SOCIETY

2016 EXAMINATIONS − SOLUTIONS

GRADUATE DIPLOMA – MODULE 2

The Society is providing these solutions to assist candidates preparing for the examinations in 2017.

The solutions are intended as learning aids and should not be seen as "model answers".

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 that there are half-marks in some questions. Please round up any odd half-mark in a question in the total mark for that question.

1. (i) $$E(X)=\int_\theta^1 x(1-\theta)^{-1}dx=(1-\theta)^{-1}[\tfrac{1}{2}x^2]_\theta^1=\frac{1-\theta^2}{2(1-\theta)}=\frac{1+\theta}{2}$$ [1]

(or by a geometric argument)

Method of moments: set sample mean equal to population mean and solve for θ, [1]

so the estimator $\hat{\theta}$ satisfies $\bar{X}=\frac{1+\hat{\theta}}{2}$

i.e. $\hat{\theta}=2\bar{X}-1$ [1]

(ii) $$E(\bar{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}[nE(X_i)]=\frac{1+\theta}{2}$$, so

$$E(\hat{\theta})=2E(\bar{X})-1=(1+\theta)-1=\theta$$

i.e. $\hat{\theta}$ is unbiased [1]

$$E(X^2)=\int_\theta^1 x^2(1-\theta)^{-1}=(1-\theta)^{-1}[\tfrac{x^3}{3}]_\theta^1=\frac{1-\theta^3}{3(1-\theta)}=\tfrac{1}{3}(1+\theta+\theta^2)$$ [1]

$$\text{var}(X)=\tfrac{1}{3}(1+\theta+\theta^2)-\left(\frac{1+\theta}{2}\right)^2=\tfrac{1}{12}(4+4\theta+4\theta^2-3-6\theta-3\theta^2)$$

$$=\tfrac{1}{12}(\theta^2-2\theta+1)=\frac{(1-\theta)^2}{12}$$ [1]

Hence $\text{var}(\hat{\theta})=4\,\text{var}(\bar{X})=\frac{4(1-\theta)^2}{12n}=\frac{(1-\theta)^2}{3n}$. [1]

(iii) Cramér-Rao lower bound: under certain regularity conditions (may be implicit) [1] the variance of any unbiased estimator of a parameter θ is bounded below by

$$E\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right]$$ [1]

($-E\left[\left(\frac{\partial^2 l}{\partial\theta^2}\right)\right]$ also acceptable), where $l=\log(L(\theta))$ and $L(\theta)$ is the likelihood function.

One of the regularity conditions for the C-R lower bound is that the range of values for x should not depend on θ. This does not hold here, so the C-R is not applicable [1].

(iv) For $\theta < y < 1$, $P(Y > y) = P(X_1, X_2, ..., X_n > y) = \prod_{i=1}^{n} P(X_i > y)$ [1]

$$= \left(1 - \tfrac{(y-\theta)}{(1-\theta)}\right)^n = \left(\tfrac{1-y}{1-\theta}\right)^n \quad [1]$$

$$F(y) = 1 - \left(\tfrac{1-y}{1-\theta}\right)^n \text{ [1] so } f(y) = \frac{n(1-y)^{n-1}}{(1-\theta)^n} \quad [1]$$

(v) Mean square error of $\tilde{\theta}$ is

$E[(\tilde{\theta}-\theta)^2]$ [1] $= E[(1-c(1-Y)-\theta)^2] = E[(1-\theta-c(1-Y))^2]$

$$= (1-\theta)^2 - 2c(1-\theta)E[1-Y] + c^2E[(1-Y)^2] \quad [1]$$

$$= (1-\theta)^2 - \frac{2c(1-\theta)n(1-\theta)}{(n+1)} + \frac{c^2n(1-\theta)^2}{(n+2)} \quad [1]$$

[Candidates may possibly start from MSE = Variance + Bias^2. If they do and succeed in getting the correct final expression they should get 3 marks, with 1 or 2 marks for partially successful attempts.]

Differentiate w.r.t. c and equate to zero:

$$-\frac{2n(1-\theta)^2}{n+1} + \frac{2nc(1-\theta)^2}{n+2} = 0 \text{ [1] so c=} \frac{n+2}{n+1}. \text{ [1]}$$

The second derivative is positive, so this corresponds to a minimum. [1]

[An alternative argument not using calculus would be acceptable.]

2. (i) Suppose that $\hat{\theta}$ is the MLE of a parameter θ and $\phi = g(\theta)$ is a (1-1) function [1] of θ. Then $g(\hat{\theta})$ is the MLE of ϕ. [1]

(ii) The likelihood function is $L(\pi; y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}$ [1]. Taking logs, $l = \log(L) = const + y\,log(\pi) + (n-y)\log(1-\pi)$. [1]

Differentiating and equating to zero, $\frac{y}{\hat{\pi}} = \frac{(n-y)}{(1-\hat{\pi})}$, [1] so $\hat{\pi} = \frac{y}{n}$. [1].

$\frac{\partial^2 l}{\partial \pi^2} < 0$, this is the MLE [1].

(iii) $P(X_i < T) = \frac{1}{\theta}\int_0^T e^{-x/\theta}dx$ [1] $=1-e^{-T/\theta}$ [1] $=\phi$, say, and $P(X_i > T) = 1-\phi$.

We have a binomial [1] with n trials, probability of success ϕ, and *y* successes. [1]

Hence the MLE of ϕ is $\frac{y}{n}$ [1].

From part (i) $\hat{\phi} = 1-e^{-T/\hat{\theta}}$, [1] so $-T/\hat{\theta} = \log(1-\hat{\phi}) = \log(1-\frac{y}{n})$ [1] and

$\hat{\theta} = -T\left[\log\left(1-\frac{y}{n}\right)\right]^{-1}$ as required. [1]

(iv) The approximate distribution of $\hat{\theta}$ is $N(\theta, I_\theta^{-1})$, where $I_\theta = E\left[\left(\frac{\partial l}{\partial \theta}\right)^2\right] = E\left[-\frac{\partial^2 l}{\partial \theta^2}\right]$

(either expression OK) and *l* is the log likelihood. [1] mark is awarded for normality, [1] for the mean, [1] for either expression for the variance, and [1] for saying that *l* is the log likelihood. There is no need to use the notation I_θ. The reciprocal of one of the given expressions could be quoted directly as the variance of $\hat{\theta}$.

An approximate 95% confidence interval has end-points $\hat{\theta} \pm 1.96 I_\theta^{-1/2}$, [1] 3\. (i) A statistic $T(X_1, X_2, ...X_n)$ is a function of $X_1, X_2, ...X_n$ but not of θ. [1]

It is sufficient for θ if the conditional distribution of $X_1, X_2, ...X_n$, given the value of T, does not depend on θ. [1] What this means is that T contains all the information about θ that is available in $X_1, X_2, ...X_n$. [1]

(ii) The likelihood is

$$L(\theta; \underline{x}) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \exp\{A(\theta)B(x_i) + C(x_i) + D(\theta)\} \text{ [1]}$$

$$= \exp\{A(\theta)\sum_i B(x_i) + \sum_i C(x_i) + nD(\theta)\} \text{ [1]}$$

[This final expression is not strictly needed in answering (ii), but is needed in (iii). It should be awarded a mark whether it appears in the answer to (ii) or to (iii).]

(iii) Write $K_1(T; \theta) = \exp\{A(\theta)\sum_i B(x_i) + nD(\theta)\}$ [1] and

$K_2(\underline{x}) = \exp\{\sum_i C(x_i)\}$. [1] Then $L(\theta; \underline{x}) = K_1(T; \theta)K_2(\underline{x})$, where

$T = \sum_i B(x_i)$ is a sufficient statistic for θ. [1]

(iv) The likelihood function is

$$L(\theta; \underline{x}) = \prod_i \frac{x_i}{\theta} \exp\left(\frac{-x_i^2}{2\theta}\right)$$

$$= \exp\{A(\theta)\sum_i B(x_i) + \sum_i C(x_i) + nD(\theta)\} \text{ [1]}$$

where $A(\theta) = \frac{-1}{2\theta}$, $B(x_i) = \sum_i x_i^2$, $C(x_i) = \log(x_i)$ and $D(\theta) = -\log(\theta)$. [1]

Hence the distribution is a member of the one-parameter exponential family [1] and $B(x_i) = \sum_i x_i^2$ is a sufficient statistic for θ. [1]

(v) A prior distribution represents knowledge about the probability distribution of an unknown parameter θ before any data are considered. [1]

Combining the prior distribution with the likelihood function gives the posterior distribution. [1] A conjugate prior distribution is such that the resulting posterior distribution belongs to the same family as the prior distribution. [1]

The likelihood function for the Rayleigh distribution can be written as

$$L(\theta;\underline{x}) = \frac{\prod_i x_i}{\theta^n} \exp\left(\frac{-\sum_i x_i^2}{2\theta} \right) \text{ [1]}$$

When multiplied by a prior distribution of the given form, we get a posterior distribution proportional to $\theta^{-\beta_1} \exp(-\beta_2/\theta)$, where $\beta_1 = \alpha_1 + n$ [1] and

$\beta_2 = \alpha_2 + \frac{1}{2}\sum_i x_i^2.$ [1]

This is of the same form as the prior distribution with 'updated' parameter values [1] so the family of prior distributions given is indeed conjugate. [1]

4 (i) Suppose that we are testing a null hypothesis H0: $\underline{\theta} \in \omega$ against the alternative H1: $\underline{\theta} \in \Omega - \omega$ [1]. The test statistic for a generalised likelihood ratio test of H0 vs. H1 is

$$\lambda = \underset{\underline{\theta} \in \omega}{Max}\{L(\underline{\theta};\underline{x})\} \Big/ \underset{\underline{\theta} \in \Omega}{Max}\{L(\underline{\theta};\underline{x})\} \text{ , [1]}$$

where $L(\underline{\theta};\underline{x})$ is the likelihood function. [0.5]

Reject H0 for small values of λ. [0.5]

(ii) For large samples $-2\log\lambda \sim \chi^2_d$ [0.5] where *d* is the difference between the number of freely varying independent parameters in Ω and in ω. [0.5]

For H0: $\theta = \theta_0$ vs. H1: $\theta \neq \theta_0$, *d*=1 [1] and $\lambda = L(\theta_0;\underline{x})/L(\hat{\theta};\underline{x})$, where $\hat{\theta}$ is the maximum likelihood estimator for θ. [1]

$-2\log\lambda = 2[l(\hat{\theta};\underline{x}) - l(\theta_0;\underline{x})]$, where $l(\theta;\underline{x})$ is the log-likelihood [1]. Using the approximation $\Pr[2(l(\hat{\theta};\underline{x}) - l(\theta_0;\underline{x})) \leq \chi^2_{1;\alpha}] = 1-\alpha$, where $\chi^2_{1;\alpha}$ is the upper α critical point of χ^2_1, [1] a confidence interval for θ with confidence coefficient $(1-\alpha)$ is given by those values of θ for which

$l(\theta;\underline{x}) \geq l(\hat{\theta};\underline{x}) - \frac{1}{2}\chi^2_{1;\alpha}$ [1]

(iii) For a random sample of n observations $x_1, x_2, ..., x_n$ from a Poisson distribution with mean μ the likelihood function is $\prod_{i=1}^{n} \frac{\mu^{x_i} e^{-\mu}}{x_i!} = \frac{\mu^{\sum_{i=1}^{n} x_i} e^{-n\mu}}{\prod_{i=1}^{n} x_i}$. [1]

The log likelihood is $l(\mu;\underline{x}) = const + \sum_i x_i \log\mu - n\mu$. [1]

Differentiating, equating the derivative to zero, [1] and checking the second derivative to confirm a *max*imum,[0.5] gives the MLE as $\hat{\mu} = \sum_i x_i \Big/ n = \bar{x}$ [1].

The generalised likelihood ratio test statistic for testing H0: $\mu = 5$ against a two-sided alternative is $-2\log\lambda = 2[l(\hat{\mu};\underline{x}) - l(5;\underline{x})]$. [1]

$\hat{\mu} = \bar{x} = 27/9 = 3,$ [0.5]

$-2\log\lambda = 2[27(\log 3 - \log 5) - 9(3-5)] = 2[-13.79 + 18] = 8.42$. [1]

This is well above the 5% critical value for χ_1^2 (3.84) [0.5], so the null hypothesis is rejected. So there is evidence that the accident rate has changed [0.5] (actually decreased).

(iv) From part (ii), the endpoints of the interval satisfy the equation $l(\mu;\underline{x}) - l(3;\underline{x}) = -{}^{3.84}\!/\!_2$ [1] giving $(27\log\mu - 9\mu) - (27\log 3 - 9\text{x}3) = -1.92$, [1] which simplifies to $3\log\mu - \mu = 0.0825$ as required. [1]

5 (a) Denote the vector of values $x_1, x_2, ..., x_n$ by $\underline{x}$. Find a function $g(\underline{x};\theta)$ of $\underline{x}$ and θ which is monotonic in θ [1] and whose probability distribution is known and does not depend on θ. [1] This is called a pivotal quantity. [1] Make a probability statement regarding $g(\underline{x};\theta)$ i.e $\Pr[g_1 \le g(\underline{x};\theta) \le g_2] = 1-\alpha$ where α is fixed (typically 0.05 or 0.01) [1] and g_1, g_2 do not depend on θ. Now manipulate the inequalities in the probability statement to get θ 'in the middle' [1] i.e $\Pr[\theta_1(\underline{X}) \le \theta \le \theta_2(\underline{X})] = 1-\alpha$. The interval $(\theta_1(\underline{X}), \theta_2(\underline{X}))$ is a $100(1-\alpha)\%$ confidence interval for θ. [1]

[If the monotonicity condition is not mentioned, full marks can still be achieved provided that it is mentioned that the resulting confidence set need not be a single interval.]

(b) In the Bayesian framework, θ is considered to be a random variable [1] and so has a probability distribution. A prior distribution is specified, [1] before taking any data into account. The likelihood function represents the information about θ contained in the data $\underline{x}$. [1] The posterior distribution for θ is obtained by taking the product of the prior distribution and the likelihood function and normalising so that it integrates to1. [1] A $100(1-\alpha)\%$ credible interval for θ is given by (θ_1, θ_2) such that $\Pr[\theta_1 \le \theta \le \theta_2] = 1-\alpha$ according to the posterior distribution. [1]

(c) Suppose that an estimate $\hat{\theta}(\underline{x})$ based on $\underline{x} = (x_1, x_2, ..., x_n)$ is available for θ. [1] Take a sample of size n *with replacement* from $x_1, x_2, ... x_n$ - call it $\underline{x}_1^*$ - and calculate $\hat{\theta}(\underline{x}_1^*)$. [1] Repeat the sampling with replacement (a large number) B times, to give B estimates of θ. [1] Arrange these B estimates in ascending order to give $\hat{\theta}^*_{[1]} \le \hat{\theta}^*_{[2]} \le ... \le \hat{\theta}^*_{[B]}$. [1] Let $B\alpha/2 = m$ (ideally choose B so that this is an integer). Then a $100(1-\alpha)\%$ bootstrap percentile confidence interval for θ is $(\hat{\theta}^*_{[m]}, \theta^*_{[B-m+1]})$. [1]

Interpretation of frequentist interval: the parameter θ is fixed but unknown – the interval is random. [1] In many repetitions of finding $100(1-\alpha)\%$ confidence intervals, the intervals with contain the true value of the parameter $100(1-\alpha)\%$ of the time in the long run, but there no information on whether any individual interval will do so. [1]

Interpretation of Bayesian interval: θ is considered to be a random variable so the interval is an interval between two quantiles of its posterior distribution. [1] So in this case the end-points of the interval are fixed (not random – unlike the frequentist interval). [1]

6. (i) Let $F_0(x)$ be the c.d.f. of the null distribution. [1]

If the random sample is ordered as $X_{[1]} \le X_{[2]} \le ... \le X_{[n]}$ define

$$F_n(x) = \begin{cases} 0 & \text{if } 0 < x < X_{[1]} \\ \dfrac{k}{n} & \text{if } X_{[k]} \le x < X_{[k+1]},\ k = 1, 2, ..., n-1 \\ 1 & \text{if } x \ge X_{[n]} \end{cases} \qquad [1]$$

Then the test statistic for the one-sample Kolmogorov-Smirnov test is

$\underset{x}{Sup} \; | F_0(x) - F_n(x) |$ [1]

[The null hypothesis will be rejected for large values of the test statistic. If this is stated here by candidates, but not explicitly mentioned in their solution to (ii), 1 additional mark should be awarded here.]

(ii) Recognise that $1 - e^{-\frac{1}{2}x}$ is the c.d.f. for an exponential distribution with mean 2 [1] and that $| F_0(x) - F_n(x) |$ must reach its maximum value at or immediately below one of the observed values of x. [1] It rises to 0.167 just below $x_{[1]}$, then drops to 0.033; it rises to 0.238 just below $x_{[2]}$, then drops to 0.038; rises to 0.359 just below $x_{[3]}$, then drops to 0.159; rises to 0.293 just below $x_{[4]}$, then drops to 0.093; rises to 0.099 just below $x_{[5]}$ before jumping to 0.101, and finally decreasing to zero. [2 marks if all these calculations are correct, 1 mark if method is clearly known but calculations are wrong.]

So the test statistic has value 0.359. [1] Large values of the test statistic lead to rejection of the null hypothesis [1] and 0.359 is well short of the given critical values [1]. Thus there is insufficient evidence to reject H_0 [1]

(iii) The t-test is a test for the mean [1] whereas the sign test is a test for the median [1]

The distribution of the test statistic for the t-test assumes approximate normality for the data which is clearly not the case here. [1] The sign test is non-parametric and does not depend on the distribution of the data and so can be used here. [1]

(iv) If the distribution is exponential with mean 2, then its median m is found by solving $0.5 = 1 - e^{-\frac{1}{2}m}$ [1] so $\text{m} = 2\log 2 = 1.386$ [1]. Hence test $H_0 : m = 1.386$ against $H_1 : m \ne 1.386$ [1]

The test statistic S is the number of observations greater than 1.386, which is 3. [1]   Since this is right in the centre of the distribution of S [Bin(5, 0.5)], there is clearly no evidence against the null hypothesis. [1]

[Although I'm not expecting it, I would award full marks if a candidate calculated $Pr(X>2) = 0.368$, and used a test statistic equal to the number of observations exceeding 2, with null distribution Bin(5, 0.368)].

7. (a) If p is the probability of H0 then $\frac{p}{(1-p)}$ is the odds of the H0. [1]

Given two simple hypotheses H_0, (null) H_1 (alternative) the Bayes factor is the likelihood under H_0 divided by the likelihood under H_1 [1] [The reciprocal of this would also be acceptable, although the wording of the question should steer candidates towards this definition].

In Bayesian inference for a parameter θ, the posterior distribution of θ is given by $q(\theta \mid \underline{x}) = \frac{L(\theta;\underline{x})\,p(\theta)}{h(\underline{x})}$, where $p(\theta)$ is the prior distribution, $L(\theta;\underline{x})$ is the likelihood function, and $h(\underline{x})$ is the marginal distribution of the data. [1]

Suppose the simple hypotheses are $H_0:\theta=\theta_0; H_1:\theta=\theta_1$. Then

$$\text{Posterior odds} = \frac{q(\theta_0 \mid \underline{x})}{q(\theta_1 \mid \underline{x})} = \frac{L(\theta_0;\underline{x})\,p(\theta_0)}{h(\underline{x})} \Bigg/ \frac{L(\theta_1;\underline{x})\,p(\theta_1)}{h(\underline{x})} \quad [1]$$

$$= \frac{p(\theta_0)}{p(\theta_1)}\frac{L(\theta_0;\underline{x})}{L(\theta_1;\underline{x})} = \text{prior odds x Bayes factor, as required.}[1]$$

(b) (i) $\Pr(X_i = x_i) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$ [1]

so $L(\lambda;\underline{x}) = \prod_{i=1}^{n}\frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_i x_i}}{\prod_i x_i!}$ [1]

The Bayes factor is this likelihood evaluated at $\lambda=5$ divided by the likelihood evaluated at $\lambda=10$. The factorial terms cancel leaving

$e^{-5n}5^{\sum_i x_i} \Big/ e^{-10n}10^{\sum_i x_i}$ [1] $= e^{5n}(0.5)^{\sum_i x_i}$ [1]

(ii) Posterior odds will be greater than prior odds if the Bayes factor exceeds 1. [1]

This occurs if $e^{5n}(0.5)^{\sum_i x_i} > 1$ or $(0.5)^{\sum_i x_i} > e^{-5n}$

$\sum_i x_i \log(0.5) > -5n$; [1] $\sum_i x_i < \frac{-5n}{\log(0.5)}$; [1] $\bar{x} < \frac{5}{\log(2)}$, as required [1]

(iii) Prior distribution has p.d.f. $0.2e^{-0.2\lambda}$, $\lambda > 0$ [0.5] and the likelihood function is $\frac{e^{-n\lambda}\lambda^{\sum_i x_i}}{\prod_i x_i!}$ [0.5].

The likelihood for a composite hypothesis such as H_1 is obtained by integrating the product of this likelihood function and the prior distribution over values of λ contained in the hypothesis, [1] so

$$L(H_1;\underline{x}) = \int_0^\infty \frac{e^{-n\lambda}\lambda^{\sum_i x_i} 0.2e^{-0.2\lambda}}{\prod_i x_i!} d\lambda = \frac{0.2}{\prod_i x_i!}\int_0^\infty \lambda^{\sum_i x_i} e^{-\lambda(n+0.2)} d\lambda \text{ [1]}$$

$$= \frac{0.2}{\prod_i x_i!(n+0.2)^{(\sum_i x_i)+1}} \int_0^\infty (n+0.2)^{(\sum_i x_i)+1} \lambda^{(\sum_i x_i)+1-1} e^{-\lambda(n+0.2)} d\lambda \text{ [1]}$$

The integral is in the form of the gamma function given in the hint with

$\nu = n + 0.2$ and $\alpha = (\sum_i x_i) + 1$ [1] so $L(H_1;\underline{x}) = \dfrac{0.2\Gamma(\sum_i x_i + 1)}{\prod_i x_i!(n+0.2)^{(\sum_i x_i)+1}}$ [1] and

the Bayes factor is

$$\frac{L(H_0;\underline{x})}{L(H_1;\underline{x})} = \frac{\dfrac{e^{-5n}5^{\sum_i x_i}}{\prod_i x_i}}{\dfrac{0.2\Gamma(\sum_i x_i + 1)}{\prod_i x_i!(n+0.2)^{(\sum_i x_i)+1}}} = \frac{e^{-5n}(5n+1)^{\sum_i x_i + 1}}{\Gamma(\sum_i x_i + 1)} \text{ as required. [1]}$$

8. (i) A strategy d_i is inadmissible if there is another strategy d_j for which $U(d_i,\theta_k) \le U(d_j,\theta_k)$ for all k, where $U(d_i,\theta_k)$ is the utility of strategy when state of nature holds, [1] with at least one inequality strict [1].

Clearly d_1 is inadmissible (compare with any other strategy), as is d_2 (compare with d_3, d_4). [1].

A maximin strategy is one for which the minimum utility is maximised. [1] The minimum utility is 1.0, 0.5, -4.5 for d_3, d_4, d_5 respectively (inadmissible strategies don't need to be considered, but candidates will not be penalised if they do so). So d_3 is maximin. [1]

(ii) The Bayes strategy is the one which maximises expected utility, [1] where expectation is taken with respect to the prior distribution of the states of nature. [1]

Prior probabilities of $\theta_1,\theta_2,\theta_3,\theta_4$ are 0.4, 0.4, 0.1, 0.1 respectively [1], and the expected utilities for the three admissible strategies are:

d_3: 0.4x1 + 0.4x1 + 0.1x2 + 0.1x2 = 1.2;

d_4: 0.4x0.5 + 0.4x0.5 + 0.1x2.5 + 0.1x 2.5 = 0.9;

d_5: -0.4x4.5 + 0.4x1 - 0.1x1.5 + 0.1x4 = -1.15. [1]

So d_3 is the Bayes strategy. [1]

(iii) Expected utilities are

d_3: $\pi_1 + 2(1-\pi_1) = 2-\pi_1$

d_4: $0.5\pi_1 + 2.5(1-\pi_1) = 2.5-2\pi_1$

d_5: $-1.75\pi_1 + 1.25(1-\pi_1) = 1.25-3\pi_1$ [1]

It is fairly clear that the expected utility for d_5 is smaller than that for d_3 for any value of π_1 between 0 and 1, so d_5 is never Bayes. [1]

d_3 is better than d_4 if $2-\pi_1 > 2.5-2\pi_1$ i.e. $\pi_1 > 0.5$. So d_3 is Bayes when π_1 >0.5 and d_4 is Bayes for $\pi_1 < 0.5$ [1] (Both are equally good at $\pi_1 = 0.5$).

(iv) Posterior probability for θ_i given advice of small demand, is the product of the prior probability and the probability of small demand, given such advice, divided by the probability of such advice. [1]

For θ_1 this is $\frac{0.4x0.9}{1-\phi} = \frac{0.36}{1-\phi}$. For $\theta_2, \theta_3, \theta_4$, the corresponding values are $\frac{0.04}{1-\phi}, \frac{0.09}{1-\phi}, \frac{0.01}{1-\phi}$. [1]

For d_3 the expected utility is now $\frac{(0.36+0.04)+2(0.09+0.01)}{1-\phi} = \frac{0.6}{1-\phi}$ [1]

Candidates could similarly calculate expected utilities for d_4, d_5, but it would be equally acceptable to say that looking at the table of utilities and the dominance of the prior probability for θ_1, that it is clear that the expected utilities for d_4, d_5 will be less than that for d_3 [1] so d_3 is Bayes when the advice is for small demand. [1]

The expected utility if the advice is used is (Expected utility when advice is large)x(Probability advice is 'large') + (Expected utility when advice is 'small')x(Probability advice is 'small') $= \frac{0.6625\phi}{\phi} + \frac{0.6(1-\phi)}{(1-\phi)} = 1.2625$. [1]

In the absence of advice, part (ii) showed that the optimal strategy was d_3 with expected utility 1.2. So the advice increases expected utility by 0.0625 i.e £62500 compared to a fee of £50000, so the gain from using the consultancy firm's advice is £12500. [1]