Exam 2014, questions and answers

Statistics (University of Melbourne)

MAST20005 Exam Solution, Semester 2, 2014

1. (a)

$$f(x_1,\ldots,x_n;\theta) \propto \frac{\prod_{i=1}^n x_i}{\theta^{2n}} \exp\{-\sum_{i=1}^n x_i/\theta\}$$
$$= \left\{\prod_{i=1}^n x_i\right\} \exp\{-\sum_{i=1}^n x_i/\theta - 2n\log\theta\}$$

and the factorization theorem yields that $Y = \sum_{i=1}^n X_i$ is sufficient for θ.

(b) The log-likelihood and score functions are, respectively:

$$\ell(\theta) = -2n\log(\theta) - \sum_{i=1}^n x_i/(\theta)$$

$$s(\theta) = \frac{\partial \ell(\theta)}{\partial\theta} = -\frac{2n}{\theta} - \frac{\sum_{i=1}^n x_i}{\theta^2}$$

(c) Solving $s(\theta) = 0$ gives the MLE $\hat\theta = \overline{X}/2$.

(d)

$$-\frac{\partial^2 \ell(\theta)}{\partial\theta^2} = -\frac{2n}{\theta^2} + \frac{2\sum_{i=1}^n X_i}{\theta^3},$$

which has expected value

$$E\left(\frac{\partial^2 \ell(\theta)}{\partial\theta^2}\right) = -\frac{2n}{\theta^2} + \frac{2n(2\theta)}{\theta^3} = \frac{2n}{\theta^2}.$$

Hence the Rao-Crámer lower bound is $\theta^2/2n$.

(e) The maximum likelihood estimate is $\hat\theta = 10.5/2 = 5.25$ and an approximate 95% confidence interval is $5.25 \pm 1.96 \times \sqrt{5.25^2/70} = (4.02, 6.48)$.

2. (a) The method of moments estimator is obtained by solving

$$\sum_i x_i/n = E(X) = \alpha\beta$$
$$\sum_i x^2/n = Var(X) + [E(X)]^2 = \alpha\beta^2 + \alpha^2\beta^2.$$

Let $\overline{X} = \sum_i X_i/n$ and $\overline{X^2} = \sum_i X_i^2/n$. Then the solution is

$$\hat\alpha = \overline{X}^2/(\overline{X^2} - \overline{X}), \quad \hat\beta = (\overline{X}^2 - \overline{X})/\overline{X}.$$

(b) Estimates are $\hat\alpha = 1.25$ and $\hat\beta = 3.38$.

3. The posterior distribution is

$$f(\lambda|x_1,\ldots,x_n) \propto \lambda^{\sum_i x_i + \alpha - 1} e^{-(n+\beta)\lambda}.$$

This is the form of a Gamma distribution. So we can conclude that the posterior distribution is a Gamma distribution, namely Gamma$(\alpha + \sum_i x_i, \beta + n)$. The minimiser of the squared loss is the posterior mean, which can be expressed by the weighted average

$$\frac{\alpha + \sum_i x_i}{\beta + n} = \frac{2 + 138}{2 + 10} = 11.7.$$

4. (a) The cdf of X_i is $P(X_i \leq x_i) = 1 - e^{-x_i/\theta}$. Thus,

$$\begin{aligned} P\left(\min\{X_1, \ldots, X_n\} > x\right) &= P\left(X_1 > x, \cdots, X_n > x\right) &(1)\\ &= \prod_{i=1}^{n} P(X_i > x) &(2)\\ &= \prod_{i=1}^{n} \exp(-x/\theta) = \exp\left(-xn/\theta\right). &(3) \end{aligned}$$

   (b) Since $E(W) = \theta/n$, an unbiased estimator is $T = nW$. Then $Var(T) = n^2 Var(W) = n^2\theta^2/n^2 = \theta^2$, while the C-R lower bound is θ^2/n.

   (c) Note that $n\theta \times W/\theta$ follows $Exp(1)$. The inverse cdf for $Exp(1)$ is

$$F^{-1}(p) = -\ln(1-p), \quad 0 \leq p < 1.$$

   Thus, $F^{-1}(0.025) = 0.0253$ and $F^{-1}(0.975) = 3.689$. Since

$$0.95 = P\left(0.0253 \leq \frac{nW}{\theta} \leq 3.689\right),$$

   a 95% CI for θ is $[nw/0.0253, nw/3.689]$.

   (d) Since $x_{(1)} = 0.16$ a 95% confidence interval for θ is $[0.217, 31.6]$.

5. (a) $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$.

$$s_p^2 = \frac{(29 \times .71^2 + 29 \times .89^2)}{60 - 2} = 5.08^2$$

   Then

$$t = \frac{8.63 - 10.97}{5.08\sqrt{1/30 + 1/30}} = -1.784$$

   Now, $t_{0.025}(58) = -2.002$ so that at the 5% level of significance we cannot reject H_0.

   (b) Since $0.025 < P(T_{58} \leq -1.784) < 0.05$, we have $0.05 <$ p-value < 0.1

   (c) Consider the test statistic

$$f = \frac{s_1^2}{s_2^2} = \frac{0.071^2}{0.089^2} = 0.6364.$$

   Under the null hypothesis this is drawn from a $F(29, 29)$ distribution so we reject $H_0 : \sigma_1^2 = \sigma_2^2$ in favour of $H_1 : \sigma_1^2 \neq \sigma_2^2$ at the 5% level if $f < 0.476$ or $f > 2.101$. Thus we cannot reject H_0 at the 5% level of significance.

6. (a) $H_0 : X \overset{d}{=} \text{Pois}(2)$, the goodness-of-fit statistic is computed by $q = \sum(o - e)^2/e \approx 3.21$ which is smaller than the critical value $c_{0.95}(\mathcal{X}_5^2) = 11.07$. Therefore, there is not enough evidence to reject the theoretical model proposed by the researcher.

   (b) Since we estimate an additional unknown parameter (λ), the goodness-of-fit statistics follows a chisquare distribution with 4 degrees of freedom. Since $c_{0.95}(\mathcal{X}_4^2) = 9.49$, we cannot reject the theoretical model proposed by the researcher.

7. (a) There are 3 negative signs, so the p-value is $P(X \leq 3; n = 8, p = 0.5) = 0.36$ and at the 5% level we cannot reject H_0.

   (b) Ranks are

```
9   14  13  15  12  10  11   2
16   5   3   7   6   1   4   8
```

   Sum of the ranks of Carb-replacement group is thus 50. Under the null hypothesis, this has mean

$$\mu = \frac{8(8+8+1)}{2} = 68,$$

   and variance

$$\sigma_W^2 = \frac{8 \times 8 \times (17)}{12} = 9.52^2.$$

   Thus

$$z = \frac{50 - 68}{9.52} = -1.8907,$$

   and $z_{0.975} = -1.96$, so we cannot reject H_0 at the 5% level of significance.

8. (a) $F = 0.2141$ which has an $F(4, 9)$ distribution and the p-value is 0.9240.

   (b) $F = 38.8982$ which has an $F(2, 13)$ distribution. The p-value is approximately 0. $F = 4.8239$ which has an $F(2, 13)$ distribution. The p-value is 0.0271.

   (c) $\hat{\sigma}^2 = 0.01289$

   (d) At the 5% level there was no significant interaction between type of coal and NaOH concentration. There was a significant difference in the mean acidity for both coal type and ethanolic NaOH concentration.

   (e) Reject H_0 when $|T| > t_{0.975}(15) = 2.13145$, where the statistic T is defined by

$$T = \frac{\overline{X}_{1\cdot} - (\overline{X}_{2\cdot} + \overline{X}_{3\cdot})/2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2} + \dfrac{1}{n_3}}},$$

   where

$$S_p = \frac{1}{n-k}\sum_{i=1}^{k}(n_i - 1)S_i^2$$

   and S_i^2's denotes the sample variance in group i. The above T statistic follows the t-student distribution with $n - k = 15$ degrees of freedom under H_0. This can shown by noting that under then null hypothesis,

$$Z = \frac{\overline{X}_{1\cdot} - (\overline{X}_{2\cdot} + \overline{X}_{3\cdot})/2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}}} \sim N(0,1)$$

and $U = (n-k)S_p^2/\sigma^2 \sim \mathcal{X}_{n-k}$ indepentdently. Thus, $Z/\sqrt{U/(n-k)} \sim t_{n-k}$.