

# MAS2903

## Problems Class 2

Dr. Lee Fawcett

**Semester 2, 2018—2019**

# Mid-semester test feedback

- It was hard, so WELL DONE!

# Mid-semester test feedback

- It was hard, so WELL DONE!
- Not sure you were expecting *MATCH* output (although I did say you should revise up to page 66)

# Mid-semester test feedback

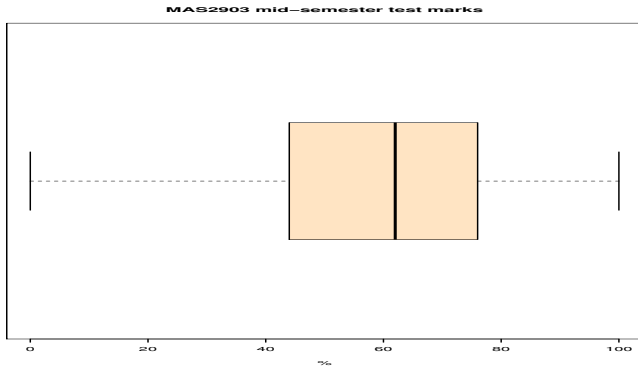
- It was hard, so WELL DONE!
- Not sure you were expecting *MATCH* output (although I did say you should revise up to page 66)
- Expect *MATCH* output in the exam!

# Mid-semester test feedback

- It was hard, so WELL DONE!
- Not sure you were expecting *MATCH* output (although I did say you should revise up to page 66)
- Expect *MATCH* output in the exam!
- Official marks in NESS have been adjusted slightly to account for overly-harsh marking – done fairly, across the board

# Mid-semester test feedback

| Q1 | Median | Mean | Q3 | Max. | St. dev. |
|----|--------|------|----|------|----------|
| 44 | 62     | 58   | 76 | 100  | 26       |



## Specific feedback: Question 1 (3 marks)

- Generally very well done

## Specific feedback: Question 1 (3 marks)

- Generally very well done
- Some people not reading carefully – just assuming that the spinner was classical, and that the jockey was a Bayesian...
- Some students choosing a different word for each scenario

## Specific feedback: Question 2 (8 marks)

- Done very well by almost everyone

## Specific feedback: Question 2 (8 marks)

- Done very well by almost everyone
- Language: "Speed cameras = 16/47" (as opposed to  $\Pr(SC|R) = 16/47$ )

## Specific feedback: Question 3 (18 marks)

- Hardly anyone wrote “roulette method” (and instead wrote “Prior knowledge” or “Gamma distribution”)

## Specific feedback: Question 3 (18 marks)

- Hardly anyone wrote “roulette method” (and instead wrote “Prior knowledge” or “Gamma distribution”)
- Lots of students didn't know where to find parameter values

## Specific feedback: Question 3 (18 marks)

- Hardly anyone wrote “roulette method” (and instead wrote “Prior knowledge” or “Gamma distribution”)
- Lots of students didn't know where to find parameter values
- Hardly anyone could phrase the question correctly, **using the percentiles**

## Specific feedback: Question 3 (18 marks)

- Hardly anyone wrote “roulette method” (and instead wrote “Prior knowledge” or “Gamma distribution”)
- Lots of students didn't know where to find parameter values
- Hardly anyone could phrase the question correctly, **using the percentiles**
- Need to state the domain – i.e.  $\theta > 0$  (-1/2 mark)

## Specific feedback: Question 3 (18 marks)

- Hardly anyone wrote “roulette method” (and instead wrote “Prior knowledge” or “Gamma distribution”)
- Lots of students didn't know where to find parameter values
- Hardly anyone could phrase the question correctly, **using the percentiles**
- Need to state the domain – i.e.  $\theta > 0$  (-1/2 mark) **COME ON!**

## Specific feedback: Question 3 (18 marks)

- Hardly anyone wrote “roulette method” (and instead wrote “Prior knowledge” or “Gamma distribution”)
- Lots of students didn't know where to find parameter values
- Hardly anyone could phrase the question correctly, **using the percentiles**
- Need to state the domain – i.e.  $\theta > 0$  (-1/2 mark) **COME ON!**
- Comments at end – done quite well, by those who attempted it. Should also comment on prior/data discrepancies

## Specific feedback: Question 4 (14 marks)

- Part (a)
  - Not including limits on the sum/product (so WHAT?!)

## Specific feedback: Question 4 (14 marks)

### ■ Part (a)

- Not including limits on the sum/product (so WHAT?!)
- Not pulling the 3 outside the product to give  $3^n$

# Specific feedback: Question 4 (14 marks)

## ■ Part (a)

- Not including limits on the sum/product (so WHAT?!)
- Not pulling the 3 outside the product to give  $3^n$
- Other errors: Thinking

$$\prod_{i=1}^n x_i^2 = \bar{x}^2 \quad \text{or} \quad n\bar{x}^2$$

# Specific feedback: Question 4 (14 marks)

## ■ Part (a)

- Not including limits on the sum/product (so WHAT?!)
- Not pulling the 3 outside the product to give  $3^n$
- Other errors: Thinking

$$\prod_{i=1}^n x_i^2 = \bar{x}^2 \quad \text{or} \quad n\bar{x}^2$$

- Also, pulling the 3 out of the product and not raising it to the power  $n$

# Specific feedback: Question 4 (14 marks)

## ■ Part (a)

- Not including limits on the sum/product (so WHAT?!)
- Not pulling the 3 outside the product to give  $3^n$
- Other errors: Thinking

$$\prod_{i=1}^n x_i^2 = \bar{x}^2 \quad \text{or} \quad n\bar{x}^2$$

- Also, pulling the 3 out of the product and not raising it to the power  $n$

## ■ Part (b)

- Writing  $T = \sum_{i=1}^n x_i^3$  – i.e. lower case  $x$

# Specific feedback: Question 4 (14 marks)

## ■ Part (a)

- Not including limits on the sum/product (so WHAT?!)
- Not pulling the 3 outside the product to give  $3^n$
- Other errors: Thinking

$$\prod_{i=1}^n x_i^2 = \bar{x}^2 \quad \text{or} \quad n\bar{x}^2$$

- Also, pulling the 3 out of the product and not raising it to the power  $n$

## ■ Part (b)

- Writing  $T = \sum_{i=1}^n x_i^3$  – i.e. lower case  $x$
- Generally, just errors in notation

## Specific feedback: Question 4 (14 marks)

### ■ Part (c)

- Most students found the correct proportional function but then never went on to say

$$\lambda|\mathbf{x} \sim \text{InvGa}(a + n, b + t).$$

## Specific feedback: Question 4 (14 marks)

### ■ Part (c)

- Most students found the correct proportional function but then never went on to say

$$\lambda|\mathbf{x} \sim \text{InvGa}(a + n, b + t).$$

- Students often thought the posterior was a standard gamma distribution

## Specific feedback: Question 4 (14 marks)

### ■ Part (c)

- Most students found the correct proportional function but then never went on to say

$$\lambda|\mathbf{x} \sim \text{InvGa}(a + n, b + t).$$

- Students often thought the posterior was a standard gamma distribution
- General algebraic mistakes – e.g.

$$e^{-b/\lambda} \cdot e^{(-1/\lambda)t} = e^{-(b+1)t/\lambda}$$

## Specific feedback: Question 4 (14 marks)

### ■ Part (c)

- Most students found the correct proportional function but then never went on to say

$$\lambda|\mathbf{x} \sim \text{InvGa}(a + n, b + t).$$

- Students often thought the posterior was a standard gamma distribution
- General algebraic mistakes – e.g.

$$e^{-b/\lambda} \cdot e^{(-1/\lambda)t} = e^{-(b+1)t/\lambda}$$

## Specific feedback: Question 5 (7 marks)

- Not having square root over  $d + n\tau$

## Specific feedback: Question 5 (7 marks)

- Not having square root over  $d + n\tau$
- Incorrectly identifying  $\tau$  and  $d$

## Specific feedback: Question 5 (7 marks)

- Not having square root over  $d + n\tau$
- Incorrectly identifying  $\tau$  and  $d$
- Mistakes getting inequalities the wrong way round
- General arithmetic mistakes

## ■ **Question 4** (Air France disaster)

- Done well by everyone
- Some issues with “odds” – see car example in Chapter 1!

## ■ **Question 4** (Air France disaster)

- Done well by everyone
- Some issues with “odds” – see car example in Chapter 1!

## ■ **Question 9** (Rayleigh likelihood and sufficient statistic)

- Main problem – students not choosing the correct functions  $h$  and  $g$  and ending up with two sufficient statistics... only one parameter!
- Otherwise done well

## ■ Question 10 (Henry Cavendish on the Earth's density)

- Generally done well
- A lot of

$$\Pr(5.45 < \mu < 5.55) = \Pr(\mu < 5.55) - \Pr(\mu > 5.45)$$

## ■ Question 11 (Posterior $\propto$ Prior $\times$ Likelihood)

- Exponential prior – PDF should not contain  $x$ ! We have  $\theta \sim \text{Exp}(1)$ , so

$$\pi(\theta) = 1 \cdot e^{-1 \cdot \theta} = e^{-\theta}.$$

- Confusion over the binomial range

## ■ **Question 12** (Bicycle tyre punctures)

- Main comment – students bnot commenting on the change from prior to posterior!

Using a random sample from a  $\text{Bin}(k, \theta)$  (with  $k$  known), determine the posterior distribution for  $\theta$  assuming

- (i) vague prior knowledge;
- (ii) the Jeffreys prior distribution;
- (iii) a very large sample.

The **conjugate prior distribution** is a  $Beta(g, h)$  distribution. Using this prior distribution, the posterior density is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^{g-1} (1-\theta)^{h-1} \times \prod_{i=1}^n \theta^{x_i} (1-\theta)^{k-x_i}, \quad 0 < \theta < 1\end{aligned}$$

The **conjugate prior distribution** is a  $Beta(g, h)$  distribution. Using this prior distribution, the posterior density is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^{g-1}(1-\theta)^{h-1} \times \prod_{i=1}^n \theta^{x_i}(1-\theta)^{k-x_i}, \quad 0 < \theta < 1\end{aligned}$$

The **conjugate prior distribution** is a  $Beta(g, h)$  distribution. Using this prior distribution, the posterior density is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \pi(\theta) f(\mathbf{x}|\theta) \\ &\propto \theta^{g-1}(1-\theta)^{h-1} \times \prod_{i=1}^n \theta^{x_i}(1-\theta)^{k-x_i}, \quad 0 < \theta < 1 \\ &\propto \theta^{g+n\bar{x}-1}(1-\theta)^{h+nk-n\bar{x}-1}, \quad 0 < \theta < 1 \\ \text{i.e. } \theta|\mathbf{x} &\sim Beta(G = g + n\bar{x}, H = h + nk - n\bar{x}).\end{aligned}$$

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

As the beta distribution restricts values to the range  $(0,1)$ , there is a finite upper limit to the variance.

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

As the beta distribution restricts values to the range  $(0,1)$ , there is a finite upper limit to the variance.

Intuitively, the maximum variance is achieved when the probability density is pushed to the extremes of the range, that is, equal mass at  $\theta = 0$  and  $\theta = 1$  – this distribution is obtained in the limit  $g \rightarrow 0$  and  $h \rightarrow 0$ .

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

As the beta distribution restricts values to the range  $(0,1)$ , there is a finite upper limit to the variance.

Intuitively, the maximum variance is achieved when the probability density is pushed to the extremes of the range, that is, equal mass at  $\theta = 0$  and  $\theta = 1$  – this distribution is obtained in the limit  $g \rightarrow 0$  and  $h \rightarrow 0$ .

Thus we will take this limit to represent vague prior information.

We represent **vague prior information** by taking a conjugate prior distribution with large variance.

As the beta distribution restricts values to the range  $(0,1)$ , there is a finite upper limit to the variance.

Intuitively, the maximum variance is achieved when the probability density is pushed to the extremes of the range, that is, equal mass at  $\theta = 0$  and  $\theta = 1$  – this distribution is obtained in the limit  $g \rightarrow 0$  and  $h \rightarrow 0$ .

Thus we will take this limit to represent vague prior information. Hence the posterior distribution under vague prior information is

$$\theta | \mathbf{x} \sim \text{Beta}(n\bar{x}, nk - n\bar{x}).$$

The **Jeffreys prior distribution** is

$$\pi(\theta) \propto \sqrt{I(\theta)}.$$

The **Jeffreys prior distribution** is

$$\pi(\theta) \propto \sqrt{I(\theta)}.$$

Now

$$\begin{aligned} f(\mathbf{x}|\theta) &\propto \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{k - x_i} \\ &\propto \theta^{n\bar{x}} (1 - \theta)^{kn - n\bar{x}}. \end{aligned}$$

Therefore

$$\log f(\mathbf{x}|\theta) = \textit{constant} + n\bar{x} \log \theta + n(k - \bar{x}) \log(1 - \theta)$$

Therefore

$$\begin{aligned}\log f(\mathbf{x}|\theta) &= \textit{constant} + n\bar{x} \log \theta + n(k - \bar{x}) \log(1 - \theta) \\ \Rightarrow \quad \frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) &= \frac{n\bar{x}}{\theta} - \frac{n(k - \bar{x})}{1 - \theta}\end{aligned}$$

Therefore

$$\begin{aligned}\log f(\mathbf{x}|\theta) &= \text{constant} + n\bar{x} \log \theta + n(k - \bar{x}) \log(1 - \theta) \\ \Rightarrow \quad \frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) &= \frac{n\bar{x}}{\theta} - \frac{n(k - \bar{x})}{1 - \theta} \\ \Rightarrow \quad \frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta) &= -\frac{n\bar{x}}{\theta^2} - \frac{n(k - \bar{x})}{(1 - \theta)^2}\end{aligned}$$

Therefore

$$\log f(\mathbf{x}|\theta) = \text{constant} + n\bar{x} \log \theta + n(k - \bar{x}) \log(1 - \theta)$$

$$\Rightarrow \frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) = \frac{n\bar{x}}{\theta} - \frac{n(k - \bar{x})}{1 - \theta}$$

$$\Rightarrow \frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta) = -\frac{n\bar{x}}{\theta^2} - \frac{n(k - \bar{x})}{(1 - \theta)^2}$$

$$\Rightarrow I(\theta) = E_{\mathbf{x}|\theta} \left[ -\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{X}|\theta) \right]$$

Therefore

$$\begin{aligned}\log f(\mathbf{x}|\theta) &= \text{constant} + n\bar{x} \log \theta + n(k - \bar{x}) \log(1 - \theta) \\ \Rightarrow \frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) &= \frac{n\bar{x}}{\theta} - \frac{n(k - \bar{x})}{1 - \theta} \\ \Rightarrow \frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta) &= -\frac{n\bar{x}}{\theta^2} - \frac{n(k - \bar{x})}{(1 - \theta)^2} \\ \Rightarrow I(\theta) &= E_{\mathbf{x}|\theta} \left[ -\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{X}|\theta) \right] \\ &= \frac{nE_{\mathbf{x}|\theta}(\bar{X})}{\theta^2} + \frac{n[k - E_{\mathbf{x}|\theta}(\bar{X})]}{(1 - \theta)^2}.\end{aligned}$$

Now  $E_{\mathbf{x}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$ . Therefore

Now  $E_{\mathbf{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$ . Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Now  $E_{\mathbf{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$ . Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

Now  $E_{\mathbf{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$ . Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

$$\pi(\theta) \propto \sqrt{I(\theta)}$$

Now  $E_{\mathbf{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$ . Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

$$\begin{aligned}\pi(\theta) &\propto \sqrt{I(\theta)} \\ &\propto \sqrt{\frac{nk}{\theta(1-\theta)}}, \quad 0 < \theta < 1\end{aligned}$$

Now  $E_{\mathbf{X}|\theta}(\bar{X}) = E_{X|\theta}(X) = k\theta$ . Therefore

$$I(\theta) = \frac{nk}{\theta} + \frac{nk}{1-\theta} = \frac{nk}{\theta(1-\theta)}.$$

Hence the Jeffreys prior for this model is

$$\begin{aligned}\pi(\theta) &\propto \sqrt{I(\theta)} \\ &\propto \sqrt{\frac{nk}{\theta(1-\theta)}}, \quad 0 < \theta < 1 \\ &\propto \theta^{-1/2}(1-\theta)^{-1/2}, \quad 0 < \theta < 1.\end{aligned}$$

This is a  $Beta(1/2, 1/2)$  prior distribution and so the resulting posterior distribution is

$$\theta|\mathbf{x} \sim Beta(1/2 + n\bar{x}, 1/2 + nk - n\bar{x}).$$

The **asymptotic posterior distribution** (as  $n \rightarrow \infty$ ) is

$$\theta|\mathbf{x} \sim N\left(\hat{\theta}, J(\hat{\theta})^{-1}\right),$$

where

$$J(\theta) = -\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta) = \frac{n\bar{x}}{\theta^2} + \frac{n(k - \bar{x})}{(1 - \theta)^2}.$$

Now

$$\begin{aligned}\frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) = 0 &\implies \frac{n\bar{x}}{\hat{\theta}} - \frac{n(k - \bar{x})}{1 - \hat{\theta}} = 0 \\ &\implies \hat{\theta} = \frac{\bar{x}}{k}\end{aligned}$$

Now

$$\begin{aligned}\frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) = 0 &\implies \frac{n\bar{x}}{\hat{\theta}} - \frac{n(k - \bar{x})}{1 - \hat{\theta}} = 0 \\ &\implies \hat{\theta} = \frac{\bar{x}}{k} \\ &\implies J(\hat{\theta}) = \frac{nk^3}{\bar{x}(k - \bar{x})}\end{aligned}$$

Now

$$\frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta) = 0 \quad \implies \quad \frac{n\bar{x}}{\hat{\theta}} - \frac{n(k - \bar{x})}{1 - \hat{\theta}} = 0$$

$$\implies \quad \hat{\theta} = \frac{\bar{x}}{k}$$

$$\implies \quad J(\hat{\theta}) = \frac{nk^3}{\bar{x}(k - \bar{x})}$$

$$\implies \quad J(\hat{\theta})^{-1} = \frac{\bar{x}(k - \bar{x})}{nk^3}.$$

Therefore, for large  $n$ , the posterior distribution for  $\theta$  is

$$\theta|\mathbf{x} \sim N\left(\frac{\bar{x}}{k}, \frac{\bar{x}(k - \bar{x})}{nk^3}\right) \quad \text{approximately.}$$