

Selected exam questions for revision

1. (A1, June 2014)

A sports scientist measures the rate of oxygen consumption of some randomly chosen athletes immediately after exercise. She finds the following 95% frequentist confidence interval C_F for the population mean rate of oxygen consumption μ :

$$(1.997, 2.503) \text{ litres per minute.}$$

In her report, she makes the following statement:

“My study revealed that the population mean rate of oxygen consumption for athletes lies in the range (1.997, 2.503) with probability 0.95”

- (a) Briefly explain why the sports scientist’s interpretation of C_F is incorrect.
- (b) After formulating an appropriate prior distribution for μ , a Bayesian statistician updates his beliefs about μ using the data in the sports scientist’s study, giving

$$\mu|\mathbf{x} \sim N(2.39, 0.01).$$

- (i) Find C_B , a 95% Bayesian confidence interval for μ .
- (ii) Briefly explain how C_B differs from C_F in interpretation.

2. (B6, June 2014)

A random sample of size n is taken from a Normal $N(\mu, 1/\tau)$ distribution (where the precision τ is known) giving a sample mean $\bar{x}$. It can be shown that the likelihood function for μ is

$$f(\mathbf{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2} [s^2 + (\bar{x} - \mu)^2]\right\},$$

where

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

- (a) Suppose our prior beliefs about μ can be summarised by a Normal $N(b, 1/d)$ distribution. Show that the posterior distribution is $\mu|\mathbf{x} \sim N(B, 1/D)$, where

$$B = \frac{db + n\tau\bar{x}}{D} \quad \text{and} \quad D = d + n\tau.$$

- (b) A sports scientist measures the rate of oxygen consumption of 10 randomly chosen athletes immediately after exercise. The sample mean is $\bar{x} = 2.25$ litres per minute. She assumes a Normal distribution for these measurements with mean μ and standard deviation $1/\sqrt{6}$ litres per minute.

A prior study suggests that $\mu \sim N(2.6, 0.025)$.

- (i) Using this information, confirm the posterior distribution used in question **A1**(b), that is

$$\mu|\mathbf{x} \sim N(2.39, 0.01).$$

(ii) Find the posterior probability that the mean rate of oxygen consumption is greater than 2.518 litres per minute.

(c) Obtain the posterior distribution for μ under the assumption of *vague prior knowledge*, and explain why the resulting 95% Bayesian confidence interval would be the same (numerically) as the sports scientist's frequentist confidence interval in question **A1**.

3. (A4, June 2013)

Consider a statistical model for data which depends on an unknown parameter θ . Describe the differences in interpretation of C_F , a 95% frequentist confidence interval for θ and C_B , a 95% Bayesian confidence intervals for θ .

4. (B7, Specimen)

A random sample of size n is taken from a Normal $N(\mu, 1/\tau)$ distribution (with τ known), giving sample mean $\bar{x}$. Prior beliefs about μ follow a normal $N(b, 1/d)$ distribution. It can be shown that the likelihood function for μ is

$$f(\mathbf{x}|\mu) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2} [s^2 + (\bar{x} - \mu)^2]\right\},$$

where

$$s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n.$$

(a) Obtain the posterior distribution $\pi(\mu|\mathbf{x})$.

(b) Determine

(i) C_B , a 95% Bayesian confidence interval for μ ;

(ii) C_H , the 95% highest density interval for μ ;

(iii) C_F , a 95% frequentist confidence interval for μ .

(c) Explain any differences in interpretation between these confidence intervals.

5. (A3, Specimen)

Suppose $X_1, X_2, \dots, X_n$ are a random sample from a Poisson $Po(\lambda)$ distribution.

(a) Write down the likelihood function for λ .

(b) Find the log-likelihood function for λ .

(c) The prior distribution proposed by Sir Harold Jeffreys can be written as $\pi(\lambda) \propto I(\lambda)^{1/2}$, where

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

Show that, for this model, Jeffreys' prior is

$$\pi(\lambda) \propto \frac{1}{\sqrt{\lambda}}, \quad \lambda > 0.$$

6. (A2, June 2014)

Sapelo Island is a barrier island off the southeast coast of the USA, and is on the migratory path of the rare black pelican. The random variable X represents the number of birds that are seen, per day, over the course of a week at an observation spot on the island.

(a) In the first week of April 2014, the following observations were made on X :

- (i) Why might the Poisson distribution be a suitable model for X ?
 - (ii) Assuming that $X \sim Po(\theta)$, obtain the likelihood function $f(\mathbf{x}|\theta)$ assuming your observations are independent.
- (b) After a consultation with an ornithologist (a zoologist with an expertise in birds), we are able to elicit a prior distribution for θ . Combining this prior distribution with the likelihood in part (a) results in the following posterior:

$$\theta|\mathbf{x} \sim Ga(18, 9).$$

Obtain the ornithologist's prior distribution for θ . How do the ornithologist's beliefs compare to our observations?

- (c) In a new analysis of the same data, a statistician attempts to use $\theta \sim N(3, 1.5)$. Briefly explain the similarities between this prior and the prior you obtained in part (b). What is the main problem with this new prior?

7. (*Question 3, mid-semester test 2015*)

The number of fatal shootings in the state of North Carolina, USA, is assumed Poisson with rate μ per month.

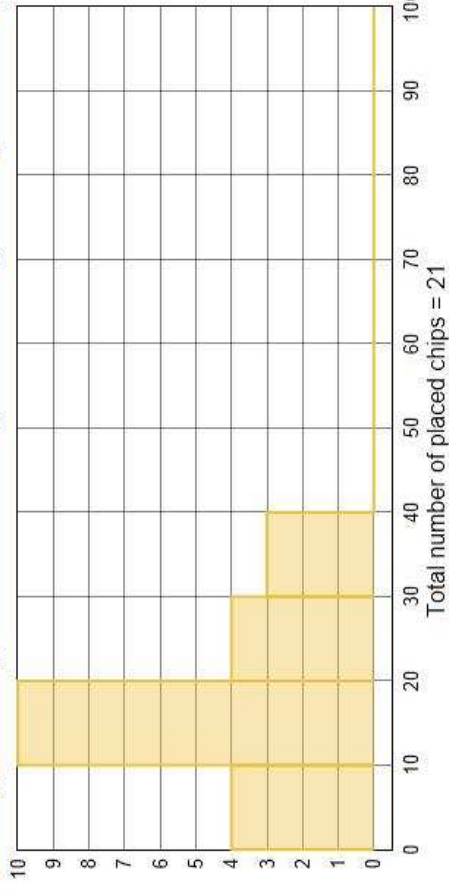
- (a) Assuming a random sample $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, show that

$$f(\mathbf{x}|\mu) \propto \mu^{n\bar{x}} e^{-n\mu}, \quad \mu > 0.$$

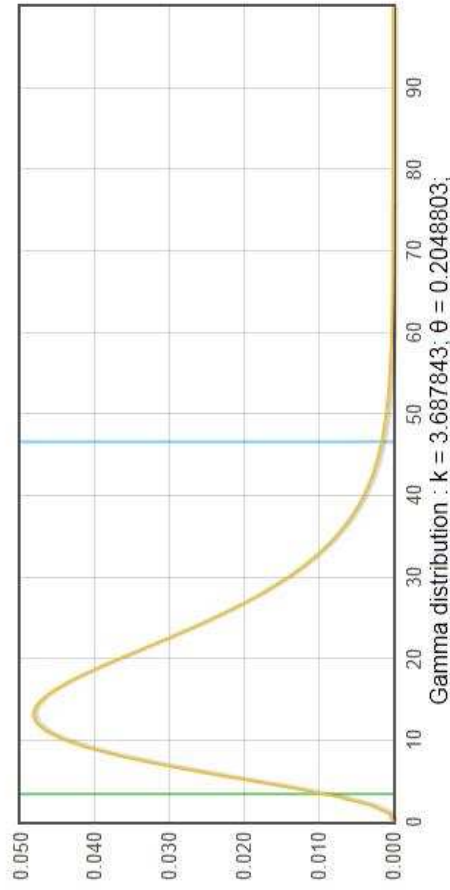
- (b) Discussions with a criminologist regarding μ led to the screenshot shown overleaf, taken from the *MATCH* online elicitation tool.
 - (i) Name the elicitation procedure that has been used.
 - (ii) Write down the fitted distribution for μ , and give the associated density, prior mean and prior standard deviation.
 - (iii) Find $\Pr(\mu > 46.6)$ and interpret this in plain English as part of the feedback and refinement process with the criminologist.
- (c) The number of fatal shootings in North Carolina was recorded for each month in 2014. Combining the associated likelihood in part (a), with the prior identified in part (b), results in the following posterior distribution for μ :

$$\mu|\mathbf{x} \sim Ga(57.7, 12.2).$$

- (i) Find $\bar{x}$, the mean number of fatal shootings observed in North Carolina in 2014.
- (ii) How do the criminologist's beliefs compare to the observed data in our sample?
- (iii) Does the posterior give more weight to the prior or the likelihood? Explain your answer.



Total number of placed chips = 21



Fitting & Feedback

☐ Normal
☐ Student-t
☐ Scaled Beta
☒ Gamma
☐ Log Normal
☐ Log Student-t
☒ Auto-select best fit

Feedback Percentiles

st percentile = 3.39
 th percentile = 46.6

8. (A3, June 2014)

Suppose we have a random sample from an exponential distribution, that is, $X_i|\theta \sim \text{Exp}(\theta)$, $i = 1, \dots, n$.

- (a) Use the Factorisation Theorem to show that $T = \sum_{i=1}^n X_i$ is sufficient for θ .
- (b) Suppose we determine that a unit exponential distribution adequately represents our prior beliefs about the parameter θ , that is, we have

$$\pi(\theta) = e^{-\theta}, \quad \theta > 0.$$

It can be shown that, after observing data $\mathbf{x}$, the posterior distribution for θ is

$$\theta|\mathbf{x} \sim \text{Ga}\left(n+1, \sum_{i=1}^n x_i + 1\right).$$

Use the fact that $T|\theta \sim \text{Ga}(n, \theta)$ to show that the same posterior is obtained knowing just the sufficient statistic T .

9. (B6, June 2013)

In hydrology, the *Pareto* distribution, with probability density function

$$f(x|\kappa, \theta) = \theta \kappa^\theta x^{-(\theta+1)}, \quad \theta, \kappa > 0 \text{ and } x > \kappa,$$

is often applied to extreme events, such as annual maximum one day rainfall accumulations. For Eskdale, a location in the Lake District, suppose we are told that $\kappa = 30$ mm is a plausible lower bound for annual maximum daily rainfall totals.

- (a) Obtain the likelihood function for θ , given that we observe a sample of annual maximum rainfall accumulations $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ at this location (and we assume these observations are independent).
- (b) Suppose our prior beliefs about θ can be described by an exponential $\text{Exp}(g)$ distribution. Show that the posterior distribution for θ is

$$\theta|\mathbf{x} \sim \text{Ga}\left(n+1, g - n \log 30 + \sum_{i=1}^n \log x_i\right).$$

- (c) Describe the term *vague prior knowledge* and explain why, in this example, this is represented as $g \rightarrow 0$.
- (d) In 2010, 2011 and 2012, annual maximum daily rainfall accumulations at Eskdale were 61mm, 55mm and 60mm (respectively). Assuming vague prior knowledge,
 - (i) find the posterior distribution for θ , rounding all parameters in this posterior to the nearest integer for convenience;
 - (ii) show that, for a 95% Bayesian confidence interval (a, b) for θ ,

$$\int_a^b \theta^3 e^{-2\theta} d\theta = \frac{57}{160};$$

- (iii) write down an additional relationship between a and b that must hold for the Bayesian confidence interval to be one of *highest density*.

10. (B7, June 2014)

Suppose that X is the number of tropical storms occurring over the Atlantic Ocean in July 2014, and let θ be the daily rate of occurrence of such storms. A meteorologist assumes that $X|\theta \sim \text{Po}(31\theta)$, with the following prior distribution for θ :

Level of activity	θ	$\pi(\theta)$
Busy	0.25	0.2
Normal	0.1	0.8

- Suppose that six tropical storms are observed over the Atlantic Ocean in July 2014. Determine the posterior distribution for θ .
- Suppose now that we want to predict the number of tropical storms Y over the Atlantic Ocean during August 2014. Determine the predictive probability function for Y , $f(y|x = 6)$, clearly stating any assumptions that you make.
- Complete the table below showing the posterior predictive distribution for Y , and use this table to find an approximate 90% prediction interval for Y .

y	0	1	2	3	4	5	6
$f(y x = 6)$	0.0286	0.0894			0.1333	0.1049	0.0828
y	7	8	9	10	11	12	13
$f(y x = 6)$	0.0684	0.0573	0.0462		0.0243	0.0156	0.0093

- Let ψ be the probability that a tropical storm develops into a hurricane. After a discussion with the meteorologist, we are able to elicit the following prior distribution for ψ :

$$\psi \sim \text{Be}(2, 23), \quad 0 < \psi < 1.$$

One of the last five tropical storms over the Atlantic Ocean developed into a hurricane. For the next five tropical storms observed in the future, show that the predictive probability distribution for the number that will develop into a hurricane is Beta-Binomial $\text{BetaBin}(5, 3, 27)$.

- Suppose we have a random sample $X_1, X_2, \dots, X_n$. For each of the following models for our sample, write down the *conjugate* prior for θ : $\text{Bin}(n, \theta)$; $\text{Po}(\theta)$; $\text{Ga}(n, \theta)$; $\text{Exp}(\theta)$; $N(\theta, \sigma^2)$.