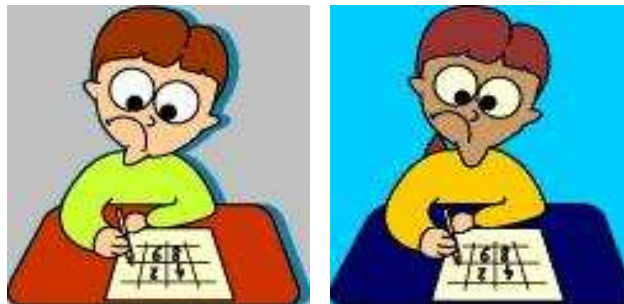




MAS2317/3317

Introduction to Bayesian Statistics



More revision material

Dr. Lee Fawcett, 2014–2015

Section A style questions

1. Describe briefly the frequency, classical and Bayesian interpretations of probability, giving an example in each case. Describe briefly two drawbacks with each interpretation.
2. Suppose $E_1, E_2, \dots, E_n$ form a partition of the sample space. State and prove Bayes Theorem. You may assume (and need not prove) the Law of Total Probability.
3. Suppose $X_1, X_2, \dots, X_n$ are a random sample from an $Exp(1/\theta)$ distribution.

(a) State the Factorisation Theorem.

(b) Show that the sample mean $\bar{X}$ is sufficient for θ . Give an intuitive explanation for this result.

4. (a) Suppose $E_1, E_2, \dots, E_n$ form a partition of the sample space. State and prove Bayes Theorem. You may assume (and need not prove) the Law of Total Probability.
(b) A small store has three checkout operators. Operator A works twice as fast (and so serves twice as many customers) as each of Operators B and C, who work at the same rate. Operator A makes a mistake when giving change 8% of the time. Operators B and C make mistakes 5% and 3% of the time respectively. If a customer does receive the wrong change, what is the probability that they were served by Operator A?

5. Explain what is meant by a conjugate prior distribution. Suppose that the data consist of a *single* observation x on the Poisson random variable X , where $X|\theta \sim Po(\theta)$. Show that the likelihood function for θ is

$$f(x|\theta) \propto \theta^x e^{-\theta}.$$

Hence show that the *Gamma* distribution is the conjugate prior distribution for θ when the data consist of a single observation from a Poisson $Po(\theta)$ distribution.

6. Describe, in detail, how subjective probabilities can be evaluated using bets. Your answer should address how such bets can be made *honest* and how they can cater for *risk averse* people.
7. Suppose $X_1, X_2, \dots, X_n$ are a random sample from a population with probability density function $f(x|\theta)$.

(a) What does it mean to say “ T is sufficient for θ ”?

(b) State the Factorisation Theorem.

(c) Suppose the population follows a Rayleigh distribution with density

$$f(x|\theta) = \begin{cases} 2x\theta e^{-\theta x^2} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\sum_{i=1}^n X_i^2$ is sufficient for θ .

8. Suppose $x_1, x_2, \dots, x_n$ are a random sample from a binomial $B(k, \theta)$ distribution (with k known). Show that the sample mean $\bar{X}$ is sufficient for θ . State any results you use.

9. Suppose that $x_1, x_2, \dots, x_n$ are a random sample from a binomial $\text{Bin}(r, \theta)$ distribution, where r is known.

(a) Show that the likelihood function for θ given the random sample is

$$f(\mathbf{x}|\theta) = k\theta^{n\bar{x}}(1 - \theta)^{n(r-\bar{x})}$$

where $\bar{x}$ is the sample mean and k is a positive constant (with respect to θ).

(b) Suppose your prior beliefs about θ were described by a $\text{Beta}(g, h)$ distribution.

(i) Determine your posterior distribution for θ given the data $\mathbf{x}$.

(ii) Is the Beta distribution a conjugate prior distribution for this model? Explain your answer.

(iii) Show that vague prior knowledge can be represented as $g \rightarrow 0$ and $h \rightarrow 0$. Hint: re-parameterise the distribution using its mean $m = g/(g + h)$ and $s = g + h$.

10. (a) Let λ be the arrival rate of trains, per hour, at Central Station. From the station manager we elicit that $E(\lambda) = 8$ and $\text{Var}(\lambda) = \frac{8}{3}$.

(i) Of the Normal, gamma and beta distributions, which distribution is most appropriate for λ ?

(ii) Given the information we have elicited from the station manager, and your answer to part (a), show that

$$\pi(\lambda) \propto \lambda^{23}e^{-3\lambda}, \quad \lambda > 0.$$

(b) This morning, in the three 1-hour periods before midday, 6, 5 and 7 trains arrived at Central Station. Assuming a Poisson distribution for X , the number of trains arriving at the station each hour, show that

$$f(\mathbf{x}|\lambda) \propto \lambda^{18}e^{-3\lambda}.$$

(c) Obtain the posterior distribution for λ , and briefly explain how our beliefs about the rate of train arrivals has changed in light of the data.

11. Consider the scenarios below. State whether a classical, frequentist or subjective interpretation of probability is being used to estimate θ_1 , θ_2 and θ_3 .

Scenario	Classical? Frequentist? Subjective?
Your friend plays the piano and has her Grade 8 exam in the morning. You and some friends try to evaluate $\theta_1 = \text{Pr}(\text{she passes her exam})$.	
You work in the outbound sales team of a call centre. From a list of all potential customers, the computer selects one completely at random. $\theta_2 = \text{Pr}(\text{customer is female})$.	
You have an interest in horse racing. You visit three bookmakers in an attempt to determine $\theta_3 = \text{Pr}(\text{Bayesian Beauty wins at the Grand National})$.	

- 12.** (a) A trucking company owns a large fleet of well-maintained trucks. Suppose that breakdowns occur at random times. The owner of the company is interested in learning about the daily rate θ at which breakdowns occur. It is known that the number of breakdowns X on a typical day has a Poisson distribution with mean θ .

The owner has some knowledge about the rate parameter θ based on the observed number of breakdowns in previous years and expresses these prior beliefs using a $Ga(4, 2)$ distribution.

- (i) Find the mean, standard deviation and mode of the owner's prior distribution. Draw a rough sketch of these beliefs.
- (ii) Suppose that the daily number of truck breakdowns are obtained for n consecutive days. Assuming these data $\mathbf{x}$ are a random sample, show that the likelihood function for θ is

$$f(\mathbf{x}|\theta) \propto \theta^{n\bar{x}} e^{-n\theta},$$

where $\bar{x}$ is the sample mean.

- (b) The owner obtains data for $n = 12$ days and finds that it has mean $\bar{x} = 2$. Determine and identify the posterior distribution for θ given this information.
 - (i) Find the mean, standard deviation and mode of the owner's posterior distribution. Describe in what ways (if any) these beliefs have changed.
 - (ii) Determine the posterior distribution for θ when the sample size n is large.
- 13.** Consider a statistical model for data which depend on an unknown parameter θ . Describe the similarities and differences in interpretation of
- (i) C_F , a 95% frequentist confidence interval for θ
 - (ii) C_B , a 95% Bayesian confidence interval for θ
 - (iii) C_H , a 95% H.D.I. for θ .

Section B style questions

14. Suppose $X_1, X_2, \dots, X_n$ are a random sample from a $N(\mu, 1)$ distribution.
- (a) State the Factorisation Theorem.
 - (b) Show that the sample mean $\bar{X}$ is sufficient for μ .
 - (c) Derive the posterior density, and hence the posterior distribution, for μ assuming a normal $N(0, d^{-2})$ prior distribution for μ .
15. (a) Suppose that data $\mathbf{x}$ are to be observed with distribution $f(\mathbf{x}|\theta)$.
- (i) What is meant by a conjugate distribution for θ ? Explain how they are used to represent vague prior knowledge for θ .
 - (ii) Describe Sir Harold Jeffreys' method for representing prior ignorance about θ .
- (b) Suppose the population follows a $Gamma(k, \theta)$ distribution, where k is known.
- (i) Verify that the conjugate distribution for θ (for this data model) is the *Gamma* distribution.
 - (ii) Describe how to represent vague prior knowledge using this prior distribution.
 - (iii) Hence, derive the posterior distribution for θ assuming vague prior knowledge.
 - (iv) Derive Jeffreys' ignorance prior distribution for θ and the consequent posterior distribution.
 - (v) By comparing the posterior distributions above, describe the effect (for this data model) of using these different methods to represent very little prior information for θ .
16. Suppose that $f(\mathbf{x}|\theta)$ is the likelihood function for a parameter θ given data $\mathbf{x}$. State the asymptotic form (as $n \rightarrow \infty$) of the posterior distribution.
17. The average annual wind speed above my office is thought to follow a Rayleigh distribution with density
- $$f(x|\theta) = \begin{cases} 2x\theta e^{-\theta x^2} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$
- Suppose that the last n years data $\mathbf{x}$ is available and that these annual means can be assumed to be independent.
- (a) Determine the asymptotic (as $n \rightarrow \infty$) posterior distribution for θ .
 - (b) Verify that the conjugate distribution for θ (for this data model) is the *Gamma* distribution.
18. Explain the *trial roulette* method of prior elicitation, and explain the role of feedback percentiles in prior elicitation.
19. Explain the *bisection* method of prior elicitation, and outline the difficulties that can be encountered when attempting to elicit a prior distribution using this approach.
20. (a) Give one advantage and one disadvantage of using a conjugate family of prior distributions in a Bayesian analysis.
- (b) Define the Jeffreys prior distribution for a model with a single parameter θ .
 - (c) A manufacturer has been developing a new type of light bulb. In the course of the development of these light bulbs it has been determined that the lifetimes of the bulbs follow an $Exp(\theta)$ distribution. Suppose a random sample of bulbs have lifetimes $x_1, x_2, \dots, x_n$.

- (i) Show that the likelihood function for θ is

$$f(\mathbf{x}|\theta) = \theta^n e^{-n\bar{x}\theta}.$$

- (ii) For this model, the conjugate family of prior distributions is the Gamma family. Express the parameters of the conjugate prior distribution in terms of its mean and variance. Hence, determine the posterior distribution for θ assuming vague prior knowledge.
- (iii) Determine the Jeffreys prior distribution for θ . Verify that this is a Gamma distribution and identify its parameters.
- (iv) Hence, determine and identify the posterior distribution for θ when using the Jeffreys prior distribution.
- 21.** A random sample of size n is taken from a normal $N(\theta, 1/\tau)$ distribution (where τ is known), giving sample mean $\bar{x}$. Prior beliefs about θ follow a normal $N(b, 1/d)$ distribution.

- (a) Show that the posterior distribution is $\theta|\mathbf{x} \sim N(B, 1/D)$, where

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

- (b) Determine C_B , a 95% Bayesian confidence interval for θ .
- (c) Determine C_H , the 95% H.D.I. for θ .
- (d) Determine C_F , a 95% frequentist confidence interval for θ .
- (e) Explain any differences in interpretation between these confidence intervals.
- 22.** (a) Suppose that the outcomes of an experiment form a random sample with probability density function $f(x|\theta)$. Our current beliefs about θ are described by the posterior distribution $\pi(\theta|\mathbf{x})$, obtained by combining experimental results $\mathbf{x}$ with a prior distribution for θ . Suppose we are interested in the outcome Y of next experiment.

By considering the distribution of (Y, θ) given the data $\mathbf{x}$, show that the predictive density function for Y given the data $\mathbf{x}$ can be expressed as

$$f(y|\mathbf{x}) = \frac{f(y|\theta)\pi(\theta|\mathbf{x})}{\pi(\theta|\mathbf{x}, y)},$$

where $\pi(\theta|\mathbf{x}, y)$ is the posterior distribution of θ given both $\mathbf{x}$ and y .

- (b) The Head of Department is interested in the student uptake of a new 4th year module. He believes that the proportion θ of 4th year students taking the module each year follows a $Beta(g, h)$ distribution. He is also prepared to assume that students choose to take the new module independently of one another. The first time the module runs, it attracts x out of the n students.
- (i) Determine the posterior distribution for θ .
- (ii) Show that the predictive distribution for the uptake y of the m possible students who will take the module the following year is

$$f(y|x) = \binom{m}{y} \frac{B(g+x+y, h+n+m-x-y)}{B(g+x, h+n-x)}, \quad y = 0, 1, \dots, m.$$