1. **Frequency interpretation**
   The probability of an outcome is the relative frequency with which the outcome would be obtained if the experiment were repeated a large number of times under similar conditions. For example, if a coin is tossed 1,000,000 times and a head appears n times then

$$Pr(Head) = \frac{n}{1,000,000}.$$

   We would expect this probability to be about 0.5.

   Classical interpretation
   This is based on the concept of equally likely outcomes. If the outcome of an experiment must be one of n different outcomes and these n outcomes are equally likely then the probability of each outcome is $1/n$. For example, when throwing a die

$$Pr(\text{observe a 3}) = \frac{1}{6}.$$

   Subjective interpretation
   The probability of an outcome represents a person's own judgement of the likelihood that the outcome will be obtained. This judgement is based on the person's own beliefs and information about the experiment. For example,

$$Pr(\text{Newcastle Utd win the Premiership this season}) = 0.8.$$

Drawbacks – any from

Frequency

1. It does not say how many times the experiment should be repeated.
2. "Similar conditions" is a vague concept.
3. It is not appropriate for many probability calculations of one-off events.
4. Standard statistical methods using the frequentist approach are not totally objective since they require subjective judgements about the validity of probability models, choice of hypotheses and interpretation of results.

Classical

1. Only applies to equally likely outcomes.
2. Whether outcomes are equally likely is a subjective assessment, even when symmetry arguments can be used.
3. It is not appropriate for many probability calculations of one-off events.

Subjective

1. It is not objective – but perhaps it is more obvious (honest) about when subjective beliefs are used.
2. It requires people to be *coherent*: they will not make any wagers which they are certain to lose.

2. If $E_1, E_2, \ldots, E_n$ are a *partition* of $\mathcal{S}$ and F is any event with $Pr(F) > 0$ then

$$Pr(E_i|F) = \frac{Pr(F|E_i)Pr(E_i)}{\sum_{j=1}^{n} Pr(F|E_j)Pr(E_j)}, \qquad i = 1, 2, \ldots, n.$$

Proof
The Total Law of Probability states that $Pr(F) = \sum_{j=1}^{n} Pr(F|E_j)Pr(E_j)$. Therefore, using the definition of conditional probability, for $i = 1, 2, \ldots, n$

$$\begin{aligned} Pr(E_i|F) &= \frac{Pr(E_i \cap F)}{Pr(F)} \\ &= \frac{Pr(F|E_i)Pr(E_i)}{Pr(F)} \\ &= \frac{Pr(F|E_i)Pr(E_i)}{\sum_{j=1}^{n} Pr(F|E_j)Pr(E_j)} \end{aligned}$$

using the Law of Total Probability.

3. (a) The **Factorisation Theorem** is
Under certain regularity conditions

$$\boldsymbol{T} \text{ is sufficient for } \theta \iff f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) = h(\boldsymbol{x})\, g(\boldsymbol{T}, \theta) \text{ for some functions } h \text{ and } g.$$

(b) The joint density function is

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} = \theta^{-n} e^{-n\bar{x}/\theta} = h(\boldsymbol{x})\, g(\bar{x}, \theta),$$

where $h(\boldsymbol{x}) = 1$ and $g(t, \theta) = \theta^{-n} e^{-nt/\theta}$. Therefore, by the Factorisation Theorem, $\bar{X}$ is sufficient for θ.
Intuitive explanation: something mentioning $E(\bar{X}|\theta) = \theta$.

4. (a) Bayes Theorem: If $E_1, E_2, \ldots, E_n$ are a partition of the sample space and F is any event with $Pr(F) > 0$ then

$$Pr(E_i|F) = \frac{Pr(F|E_i)Pr(E_i)}{\sum_{j=1}^{n} Pr(F|E_j)Pr(E_j)}, \qquad i = 1, 2, \ldots, n.$$

Proof: The Total Law of Probability states that $Pr(F) = \sum_{j=1}^{n} Pr(F|E_j)Pr(E_j)$. Therefore, using the definition of conditional probability, for $i = 1, 2, \ldots, n$

$$Pr(E_i|F) = \frac{Pr(E_i \cap F)}{Pr(F)} = \frac{Pr(F|E_i)Pr(E_i)}{Pr(F)} = \frac{Pr(F|E_i)Pr(E_i)}{\sum_{j=1}^{n} Pr(F|E_j)Pr(E_j)}$$

using the Law of Total Probability.

(b) Let M = operator makes mistake, A = served by Operator A,... . Then we have $Pr(A) = 0.5$, $Pr(B) = Pr(C) = 0.25$, $Pr(M|A) = 0.08$, $Pr(M|B) = 0.05$ and

$Pr(M|C) = 0.03$. Therefore, since A, B and C form a partition

$$\begin{aligned}
Pr(A|M) &= \frac{Pr(M|A)Pr(A)}{Pr(M|A)Pr(A) + Pr(M|B)Pr(B) + Pr(M|C)Pr(C)} \\
&= \frac{0.08 \times 0.5}{0.08 \times 0.5 + 0.05 \times 0.25 + 0.03 \times 0.25} \\
&= \frac{0.04}{0.04 + 0.0125 + 0.0075} \\
&= \frac{0.04}{0.06} \\
&= \frac{2}{3}.
\end{aligned}$$

5. Suppose that data $\boldsymbol{x}$ are to be observed with distribution $f(\boldsymbol{x}|\theta)$. A family $\mathfrak{F}$ of prior distributions for θ is said to be *conjugate* to $f(\boldsymbol{x}|\theta)$ if for every prior distribution $\pi(\theta) \in \mathfrak{F}$, the posterior distribution $\pi(\theta|\boldsymbol{x})$ is also in $\mathfrak{F}$.

The likelihood function is

$$\begin{aligned}
f(x|\theta) &= Pr(X = x|\theta) \\
&= \frac{\theta^x e^{-\theta}}{x!} \\
&\propto \theta^x e^{-\theta}
\end{aligned}$$

Suppose the prior distribution is $\theta \sim \Gamma(g, h)$. Then, using Bayes Theorem, the posterior distribution is

$$\begin{aligned}
\pi(\theta|x) &\propto \pi(\theta) f(x|\theta) \\
&\propto \theta^{g-1} e^{-h\theta} \times \theta^x e^{-\theta}, \quad \theta > 0 \\
&\propto \theta^{g+x-1} e^{-(h+1)\theta}, \quad \theta > 0
\end{aligned}$$

and so the posterior distribution is $\theta|x \sim \Gamma(g + x, h + 1)$.

As the prior and posterior distributions both come from the Gamma family, this is the conjugate prior distribution for this data model.

6. One way of determining (or quantifying) a subjective value for the probability $Pr_H(A)$ of an event A given information H is to consider a series of possible bets with outcome

win £c if A occurs and £0 if A^c occurs.

The bet is a "fair" bet if the expected winnings are zero, that is, if a stake of £cp is used when it is believed that $Pr_H(A) = p$.

One problem with this approach is that, in general, p will depend on c: a person who is willing to bet £1 on the spin of a coin to win £2 if it lands heads may refuse to to bet if the stakes are raised to £1000 – most people are risk-averse. Therefore, we shall restrict our attention to the $c = 1$ case: pay £p for the bet

win £1 if A occurs and £0 if A^c occurs.

Bets can be made "honest" by randomising between whether the person "hosts" the bet or "places" the bet. For example, suppose you believe that $Pr_H(A) = 0.5$. An "honest" bet would mean that you would buy the bet for a maximum stake of £0.50. However, if you weren't honest you might try to buy the bet for any amount less than £0.50, say £0.20. If you were hosting the bet, you would take the bet for any amount more than £p, say for £0.80. These conflicting interests can be offset if, when choosing p, it is equally likely that you are hosting the bet or placing the bet. In such circumstances it is in your own interests to give the value of p that you believe to be "correct".

7. (a) T is sufficient for θ if it contains all the information in the data about θ.

(b) See solution for question 3(a).

(c)

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= 2^n \prod_{i=1}^n x_i \times \theta^n e^{-\theta \sum_{i=1}^n x_i^2} \\ &= h(\boldsymbol{x}) g\left(\sum_{i=1}^n x_i^2, \theta\right), \end{aligned}$$

where $h(\boldsymbol{x}) = 2^n \prod x_i$ and $g(t, \theta) = \theta^n e^{-\theta t}$. Therefore, by the Factorisation Theorem, $T = \sum_{i=1}^n X_i^2$ is sufficient for θ.

8. We can write the likelihood as

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \left(\prod_{i=1}^n {}^k\mathrm{C}_{x_i}\right) \times \theta^{n\bar{x}}(1-\theta)^{n(k-\bar{x})} \\ &= h(\boldsymbol{x}) g(\bar{x}, \theta), \end{aligned}$$

where

$$h(\boldsymbol{x}) = \left(\prod_{i=1}^n {}^k\mathrm{C}_{x_i}\right) \qquad \text{and}$$

$$g(t, \theta) = \theta^{nt}(1-\theta)^{n(k-\bar{x})}.$$

Therefore, by the Factorisation Theorem, $T = \bar{X}$ is sufficient for θ.

9. (a) The likelihood function is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^n \binom{r}{x_i} \theta^{x_i}(1-\theta)^{r-x_i} = k\theta^{n\bar{x}}(1-\theta)^{n(r-\bar{x})}$$

(b) Using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) f(\boldsymbol{x}|\theta) \\ &\propto \frac{\theta^{g-1}(1-\theta)^{h-1}}{B(a,b)} \times k\theta^{n\bar{x}}(1-\theta)^{n(r-\bar{x})}, \quad 0 < \theta < 1 \\ &\propto \theta^{g+n\bar{x}}(1-\theta)^{h+n(r-\bar{x})}, \quad 0 < \theta < 1, \end{aligned}$$

that is, $\theta|\boldsymbol{x} \sim Beta(G = g + n\bar{x}, H = h + n(r - \bar{x}))$.

(c) The prior and posterior distributions are both of the Beta family and so the Beta distribution is the conjugate prior distribution to this binomial model.

(d) Consider the 1-1 reparameterisation $m = g/(g+h)$ and $s = g+h$. Then

$$Var(\theta) = \frac{gh}{(g+h)^2(g+h+1)} = \frac{m(1-m)}{s+1}.$$

This ratio is maximised by maximising $m(1-m)$ and minimising $s+1$. Now since $g > 0$ and $h > 0$, we have $0 < m < 1$ and $s > 0$. Therefore, the variance is maximised when $m = 1/2$ and as $s \to 0$, that is, as $g \to 0$ and $h \to 0$.

10. (a) (i) Gamma – since λ is a rate, it is not restricted to lie in the range $(0,1)$ (beta) nor can it be negative (Normal).

(ii) For a $Ga(a,b)$ distribution,

$$\frac{a}{b} = 8 \qquad \text{and so} \qquad a = 8b.$$

Thus

$$\begin{aligned} \frac{a}{b^2} &= \frac{8}{3} \\ \frac{8b}{b^2} &= \frac{8}{3} \\ \frac{8}{b} &= \frac{8}{3} \qquad \text{and so} \\ b &= 3. \end{aligned}$$

Thus, $a = 24$, and so

$$\pi(\lambda) \propto \lambda^{23} e^{-3\lambda},$$

as required.

(b) We have

$$f(\boldsymbol{x}|\lambda) \propto \lambda^{6+5+7} e^{-3\lambda} = \lambda^{18} e^{-3\lambda}.$$

(c) Using Bayes' Theorem, we have

$$\begin{aligned} \pi(\lambda|\boldsymbol{x}) &\propto \lambda^{23+18} e^{-3\lambda-3\lambda} \\ &\propto \lambda^{41} e^{-6\lambda}, \end{aligned}$$

and so $\lambda|\boldsymbol{x} \sim Ga(42,6)$. This gives a posterior mean and variance of 7 and 1.167 (respectively); in light of the data, our beliefs have shifted down slightly, and we have observed reduction in standard deviation.

11. Piano – subjective; call centre – classical; *Bayesian Beauty* – subjective.

12. (a) (i) Using the results in the list of distributions, if $\theta \sim Ga(g=4, h=2)$ then the prior mean and standard deviation are

$$E(\theta) = \frac{g}{h} = \frac{4}{2} = 2 \qquad \text{and} \qquad SD(\theta) = \frac{\sqrt{g}}{h} = \frac{2}{2} = 1.$$

The prior mode is

$$Mode(\theta) = \frac{g-1}{h} = \frac{3}{2} = 1.5.$$

Sketch: must have labelled axes and identify (i) mode at $\theta = 1.5$; (ii) $\pi(0) = 0$; (iii) $\pi(\theta) \to 0$ as $\theta \to \infty$.

(ii) The likelihood function is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} \propto \theta^{n\bar{x}} e^{-n\theta}.$$

(b) Using Bayes Theorem, the posterior density is

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) f(\boldsymbol{x}|\theta) \\ &\propto \theta^3 e^{-2\theta} \times \theta^{24} e^{-12\theta}, \quad \theta > 0 \\ &\propto \theta^{27} e^{-14\theta}, \quad \theta > 0,\end{aligned}$$

that is, $\theta|\boldsymbol{x} \sim Ga(G = 28, H = 14)$.

(i) The posterior mean, mode and standard deviation are

$$E(\theta|\boldsymbol{x}) = \frac{G}{H} = \frac{28}{14} = 2, \qquad Mode(\theta|\boldsymbol{x}) = \frac{G-1}{H} = \frac{27}{14} = 1.929$$

and

$$SD(\theta|\boldsymbol{x}) = \frac{\sqrt{G}}{H} = \frac{28}{14} = 0.378.$$

Means: no change

Modes: posterior mode much closer to posterior mean than in the prior

SDs: posterior SD is much smaller - by a factor of 2.6.

(ii) In large samples, the posterior distribution is approximately $N(\hat{\theta}, J(\hat{\theta})^{-1})$, where $\hat{\theta}$ is the MLE and

$$J(\theta) = -\frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{x}|\theta).$$

Now

$$\log f(\boldsymbol{x}|\theta) = n\bar{x} \log\theta - n\theta$$

and so

$$\frac{\partial}{\partial\theta} \log f(\boldsymbol{x}|\theta) = \frac{n\bar{x}}{\theta} - n \quad \text{and} \quad \frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{x}|\theta) = -\frac{n\bar{x}}{\theta^2}.$$

Therefore, $\hat{\theta}$ satisfies

$$\frac{n\bar{x}}{\hat{\theta}} - n = 0 \qquad \Rightarrow \qquad \hat{\theta} = \bar{x}$$

and this is the MLE since the second derivative is always negative. Also

$$J(\hat{\theta}) = \frac{n\bar{x}}{\hat{\theta}^2} = \frac{n}{\bar{x}}.$$

Therefore, in large samples, the posterior distribution is (approximately)

$$\theta|\boldsymbol{x} \sim N\left(\bar{x}, \frac{\bar{x}}{n}\right).$$

13.
- $Pr(\theta \in C_B) = Pr(\theta \in C_H) = 0.95$ whereas $Pr(\theta \in C_F) = 0$ or 1;
- However, the interval C_F covers the true value θ on 95% of occasions — in repeated applications of the formula;
- C_H contains the most likely values of θ (in terms of the posterior distribution).

14. (a) Factorisation Theorem: under certain regularity conditions

$$\boldsymbol{T} \text{ is sufficient for } \theta \iff f(\boldsymbol{x}|\theta) = h(\boldsymbol{x})\, g(\boldsymbol{T}, \theta) \text{ for some functions } h \text{ and } g.$$

(b) The joint probability density function is

$$\begin{aligned}
f(\boldsymbol{x}|\mu) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x_i-\mu)^2}{2}\right\} \\
&= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum (x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\left[\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right]\right\} \\
&= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum x_i^2\right\} \times \exp\left\{n\mu\bar{x} - \frac{n\mu^2}{2}\right\} \\
&= \qquad\qquad h(\boldsymbol{x}) \times g\left(\bar{x}, \theta\right).
\end{aligned}$$

Therefore, by the Factorisation Theorem, $T = \bar{X}$ is sufficient for μ.

(c) Using Bayes Theorem, the posterior density function is

$$\begin{aligned}
\pi(\mu|\boldsymbol{x}) &\propto \pi(\mu) f(\boldsymbol{x}|\mu) \\
&\propto \frac{d}{\sqrt{2\pi}} e^{-d^2\mu^2/2} \times (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum x_i^2\right\} \exp\left\{n\mu\bar{x} - \frac{n\mu^2}{2}\right\} \\
&\propto \exp\left\{-\frac{(n+d^2)\mu^2}{2} + n\mu\bar{x}\right\} \\
&\propto \exp\left\{-\frac{(n+d^2)}{2}\left(\mu - \frac{n\bar{x}}{n+d^2}\right)^2\right\}.
\end{aligned}$$

Therefore

$$\mu|\boldsymbol{x} \sim N\left(\frac{n\bar{x}}{n+d^2}, \frac{1}{n+d^2}\right).$$

15. (a) Weak prior knowledge:

(i) A conjugate distribution for the parameter θ in the data model $f(\boldsymbol{x}|\theta)$ is one for which the prior and posterior distributions are from the same parametric family. Vague prior knowledge is represented by taking a diffuse conjugate prior distribution – usually taking the limit $Var(\theta) \to \infty$.

(ii) A Jeffreys' prior distribution is

$$\pi(\theta) \propto I(\theta)^{1/2}$$

where

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

(b) The likelihood function is

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \prod_{i=1}^{n} \frac{\theta^k x_i^{k-1} e^{-\theta x_i}}{\Gamma(k)} \\ &\propto \theta^{nk} e^{-n\bar{x}\theta}. \end{aligned}$$

(i) If $\theta \sim Ga(g, h)$ then, using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) f(\boldsymbol{x}|\theta) \\ &\propto \theta^{g-1} e^{-h\theta} \times \theta^{nk} e^{-n\bar{x}\theta}, \quad \theta > 0 \\ &\propto \theta^{g+nk-1} e^{-(h+n\bar{x})\theta}, \quad \theta > 0, \end{aligned}$$

that is, $\theta|\boldsymbol{x} \sim Ga(g+nk, h+n\bar{x})$. Therefore, the *Gamma* distribution is the conjugate prior distribution (for this data model).
Reparameterising the prior in terms of its mean m and variance v gives

$$g = \frac{m^2}{v} \quad \text{and} \quad h = \frac{m}{v}.$$

Therefore, as $v \to \infty$, we have $g \to 0$ and $h \to 0$. Under this limit, the posterior distribution is $\theta|\boldsymbol{x} \sim Ga(nk, n\bar{x})$.

(ii) The log-likelihood function is

$$\begin{aligned} \log f(\boldsymbol{x}|\theta) &= nk \log\theta + (k-1)\sum \log x_i - n\bar{x}\theta - n\log\Gamma(k) \\ &\Rightarrow \quad \frac{\partial}{\partial\theta} \log f(\boldsymbol{x}\theta) = \frac{nk}{\theta} - n\bar{x} \\ &\Rightarrow \quad \frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{x}|\theta) = -\frac{nk}{\theta^2} \\ &\Rightarrow \quad I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{x}|\theta)\right] = \frac{nk}{\theta^2}. \end{aligned}$$

Hence, a Jeffreys prior for this model is

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

Taking this prior distribution and using Bayes Theorem, the posterior density is

$$\begin{aligned} \pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) f(\boldsymbol{x}|\theta) \\ &\propto \frac{1}{\theta} \times \theta^{nk} e^{-n\bar{x}\theta}, \quad \theta > 0 \\ &\propto \theta^{nk-1} e^{-n\bar{x}\theta}, \quad \theta > 0, \end{aligned}$$

that is, $\theta|\boldsymbol{x} \sim Ga(nk, n\bar{x})$.

(iii) The posterior distribution assuming vague prior knowledge and a Jeffreys' prior distribution are the same.

16. The asymptotic form of the posterior distribution is

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

where

$$J(\theta) = -\frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{x}|\theta).$$

17. (a) The asymptotic posterior distribution is given by

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

The likelihood function is

$$f(\boldsymbol{x}|\theta) \quad = \quad (2\theta)^n \prod_{i=1}^{n} x_i e^{-\theta \sum_{i=1}^{n} x_i^2},$$

giving

$$\log f(\boldsymbol{x}|\theta) \quad = \quad n\log 2 + n\log\theta + \sum_{i=1}^{n} \log x_i - \theta \sum_{i=1}^{n} x_i^2.$$

Thus

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\theta) \quad = \quad \frac{n}{\theta} - \sum_{i=1}^{n} x_i^2 \qquad \text{and} \tag{1}$$

$$\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{x}|\theta) \quad = \quad -\frac{n}{\theta^2}. \tag{2}$$

Equating (1) to zero, replacing θ with $\hat{\theta}$ and solving for $\hat{\theta}$ gives

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i^2}.$$

Also, from (2), we see that

$$J(\hat{\theta})^{-1} = \frac{n}{\left(\sum_{i=1}^{n} x_i^2\right)^2}.$$

(b) Using Bayes' Theorem, with a $Ga(g, h)$ prior, gives

$$\theta|\boldsymbol{x} \sim Ga\left(g + n, h + \sum_{i=1}^{n} x_i^2\right).$$

18. See lecture notes – Example 3.3.

19. See lecture notes – Example 3.4.

20. (a) Advantage: the Bayes update is straightforward as both prior and posterior have the same distributional form and the posterior's parameters have a simple functional form
Disadvantage: a conjugate prior might not fully capture prior beliefs

(b) Jeffreys prior distribution is $\pi(\theta) \propto \sqrt{I(\theta)}$ where

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

is Fisher's information. It is preferred over the standard (sometimes improper) constant prior used to represent prior ignorance because the Jeffreys prior is invariant under 1–1 transformations when these others are not.

(c) (i) The likelihood function is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-n\bar{x}\theta}.$$

(ii) If $\theta \sim Ga(g, h)$ then it has mean and variance

$$m = \frac{g}{h} \qquad \text{and} \qquad v = \frac{g}{h^2}$$

and so

$$g = \frac{m^2}{v} \qquad \text{and} \qquad h = \frac{m}{v}.$$

Taking $\theta \sim Ga(g, h)$ and using Bayes Theorem, the posterior density is

$$\begin{aligned}
\pi(\theta|\boldsymbol{x}) &\propto \pi(\theta) f(\boldsymbol{x}|\theta) \\
&\propto \theta^{g-1} e^{-h\theta} \times \theta^n e^{-n\bar{x}\theta}, \quad \theta > 0 \\
&\propto \theta^{g+n-1} e^{-(h+n\bar{x})\theta}, \quad \theta > 0
\end{aligned}$$

that is, the posterior distribution is a $Ga(G = g + n, h + n\bar{x})$ distribution.
A vague prior is one with as large a variance as possible. Here $g \to 0$ and $h \to 0$ as $v \to \infty$ (for fixed m), and so vague prior knowledge is represented by taking $g \to 0$ and $h \to 0$. In this case, $G = n$ and $H = n\bar{x}$. Thus the posterior distribution for θ assuming vague prior knowledge is a $Ga(n, n\bar{x})$ distribution.

(iii) We have

$$\begin{aligned}
\log f(\boldsymbol{x}|\theta) &= n \log \theta - n\bar{x}\theta \\
\Rightarrow \quad & \frac{\partial}{\partial\theta} \log f(\boldsymbol{x}|\theta) = \frac{n}{\theta} - n\bar{x} \\
\Rightarrow \quad & \frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{x}|\theta) = -\frac{n}{\theta^2} \\
\Rightarrow \quad & I(\theta) = E_{\boldsymbol{X}|\theta}\left[-\frac{\partial^2}{\partial\theta^2} \log f(\theta|\boldsymbol{X})\right] = \frac{n}{\theta^2}.
\end{aligned}$$

Hence Jeffreys prior for this model is

$$\begin{aligned}
\pi(\theta) &\propto \sqrt{I(\theta)} \\
&\propto \sqrt{\frac{n}{\theta^2}}, \qquad \theta > 0 \\
&\propto \theta^{-1}, \qquad \theta > 0.
\end{aligned}$$

This is a $Ga(g, h)$ prior distribution, with $g \to 0$ and $h \to 0$.

(iv) From (ii), taking a $Ga(g, h)$ prior distribution gives a $Ga(G = g + n, h + n\bar{x})$ posterior distribution. Therefore, taking a Jeffreys prior distribution gives a $Ga(n, n\bar{x})$ posterior distribution.

21. (a) (i) The likelihood function for μ is

$$\begin{aligned}
f(\boldsymbol{x}|\mu) = &= \prod_{i=1}^{n} \left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau}{2}(x_i-\mu)^2\right\} \\
&= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \\
&= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2\right\} \\
&= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2\right]\right\} \\
&= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2+(\bar{x}-\mu)^2\right]\right\}.
\end{aligned}$$

(ii) Applying Bayes Theorem, the posterior density function is

$$\begin{aligned}
\pi(\mu|\boldsymbol{x}) &\propto \pi(\mu)f(\boldsymbol{x}|\mu) \\
&\propto \left(\frac{d}{2\pi}\right)^{1/2} \exp\left\{-\frac{d}{2}(\mu-b)^2\right\} \\
&\qquad\qquad \times \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{n\tau}{2}\left[s^2+(\bar{x}-\mu)^2\right]\right\} \\
&= k_1 \exp\left\{-\frac{1}{2}\left[d(\mu-b)^2+n\tau(\bar{x}-\mu)^2\right]\right\}
\end{aligned}$$

where k_1 is a constant that does not depend on μ. Now the exponent can be simplified by expanding terms in μ and then completing the square, as follows. We have

$$\begin{aligned}
d(\mu-b)^2+n\tau(\bar{x}-\mu)^2 &= d(\mu^2-2b\mu+b^2)+n\tau(\bar{x}^2-2\bar{x}\mu+\mu^2) \\
&= (d+n\tau)\mu^2-2(db+n\tau\bar{x})\mu+db^2+n\tau\bar{x}^2 \\
&= (d+n\tau)\left\{\mu-\left(\frac{db+n\tau\bar{x}}{d+n\tau}\right)\right\}^2+c
\end{aligned}$$

where c does not depend on μ. Then

$$\begin{aligned}
\pi(\mu|\boldsymbol{x}) &= k_1 \exp\left\{-\frac{D}{2}(\mu-B)^2-\frac{c}{2}\right\} \\
&\propto \exp\left\{-\frac{D}{2}(\mu-B)^2\right\} \\
\implies \quad \mu|\boldsymbol{x} &\sim N(B, 1/D).
\end{aligned}$$

(b) Any interval C_B with $Pr(\mu \in C_B|\boldsymbol{x}) = 0.95$ will do.

(c) The H.D.I. for μ contains the most likely values of μ and, as the posterior distribution is symmetric about B, the H.D.I. is an equi-tailed interval:
$C_H = (B - 1.96/\sqrt{D}, B + 1.96/\sqrt{D})$.

(d) The frequentist equi-tailed interval is $C_F = (\bar{x} - 1.96/\sqrt{n\tau}, \bar{x} + 1.96/\sqrt{n\tau})$.

(e) $Pr(\mu \in C_B|\boldsymbol{x}) = Pr(\mu \in C_H|\boldsymbol{x}) = 0.95$ but $Pr(\mu \in C_F|\boldsymbol{x}) = 0$ or 1.
Numerically, C_H and C_F are the same assuming vague prior knowledge for μ.

22. (a) The conditional probability (density) function of (Y, θ) given the data $\boldsymbol{x}$ can be written as

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

or as

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

where $\pi(\theta|\boldsymbol{x}, y)$ is, formally, the posterior probability (density) function of θ given both the data $\boldsymbol{x}$ and the future observation y. Equating these two expressions, we obtain

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

(b) We have $X|\theta \sim B(n, \theta)$ and $\theta \sim Beta(g, h)$.

(i) Using Bayes Theorem, the posterior density function is

$$\begin{aligned}\pi(\theta|x) &\propto \pi(\theta)\, f(x|\theta) \\ &\propto \theta^{g-1}(1-\theta)^{h-1} \times \theta^{x}(1-\theta)^{n-x} \\ &\propto \theta^{g+x-1}(1-\theta)^{h+n-x-1}\end{aligned}$$

and so $\theta|x \sim Beta(g + x, h + n - x)$.

(ii) $Y|\theta \sim B(m, \theta)$, independent of $X|\theta$. The conjugate update clearly adds the number of succcess (failures) to the Beta first (second) parameter. Thus, $\theta|x, y \sim Beta(g + x + y, h + n + m - x - y)$. Therefore, for $y = 0, 1, \ldots, m$

$$\begin{aligned}f(y|x) &= \frac{f(y|\theta)\pi(\theta|x)}{\pi(\theta|x, y)} \\ &= \frac{\binom{m}{y}\theta^{y}(1-\theta)^{m-y} \times \frac{\theta^{g+x-1}(1-\theta)^{h+n-x-1}}{B(g+x,h+n-x)}}{\frac{\theta^{g+x+y-1}(1-\theta)^{h+n+m-x-y-1}}{B(g+x+y,h+n+m-x-y)}} \\ &= \binom{m}{y}\frac{B(g + x + y, h + n + m - x - y)}{B(g + x, h + n - x)}\end{aligned}$$