Exercise (1)
The posterior modes are just the MLEs \( t_i = y_i/n_i \) in each case, and these values are 0.25, 0.30, 0.73 and 0.76, respectively. Also, cases \( i = 2 \) and \( i = 4 \) have more precise posteriors since they are based on larger \( n_i \) values. Simply matching with the graphs implies the matches

\[
i = 1: D, \quad i = 2: A, \quad i = 3: C, \quad i = 4: B
\]

Exercise (2)

(a) \( \text{Gamma}(6 + 11, 2 + 1) = \text{Gamma}(17, 3) \). The variance is \( 17/3^2 = 1.89 \).

(b) The prior \( \text{Gamma}(17, 3) \) updates to the new posterior \( \text{Gamma}(17 + 7, 3 + 1) = \text{Gamma}(24, 4) \), with variance \( 24/4^2 = 1.5 \).

(c) The posterior \( \text{Gamma}(a_t, b_t) \) after month \( t \) gets updated to \( \text{Gamma}(a_t + x_{t+1}, b_t + 1) \) when month \( t + 1 \) data value \( x_{t+1} \) is recorded. As time progresses more data is received and the spread of the posterior continues to decrease, reflecting increased information about \( \theta \).

(d) \( \text{Gamma}(6 + 198, 2 + 12) = \text{Gamma}(204, 14) \).

Exercise (3)

(a) \( (\theta_0 | y_0) \sim \text{Beta}(176, 320) \) and \( (\theta_1 | y_1) \sim \text{Beta}(40, 124) \).

(b) The means are the MLEs \( t_0 = 0.35 \) and \( t_1 = 0.24 \), and the variances are \( \tau_0^2 = t_0(1 - t_0)/n_0 \) so that \( \tau_0^2 = 0.000463 \) and \( \tau_1^2 = 0.001128 \). The SDs are \( \tau_0 = 0.0215 \) and \( \tau_1 = 0.0336 \) so the 95% intervals are (0.31,0.40) and (0.17,0.31).

(c) We know that \( \delta \sim N(t_1 - t_0, \tau_1^2 + \tau_0^2) = N(-0.1135, 0.04^2) \). So the 95% and 99% intervals are\((0.19, 0.04)\) and \((0.22, 0.01)\).

(d) \( \theta_1 = \theta_0 \) implies \( \delta = 0 \). This has little support under the above posterior – the value \( 0 \) is lies well outside the 99% interval for \( \delta \), so it quite improbable.

(e) It is highly likely that \( \delta < 0 \), so that \( \theta_1 < \theta_0 \). Members of the Cambridge Inn voting population are clearly more likely to have voted for Caberwal.

Exercise (4)

(a) The statement attempts to assign a value of 0.99 to \( p(\theta = 1|x = 1) \), or equivalently the value of 0.01 to \( p(\theta = 0|x = 1) \). The reasoning implicit in the statement is that the official matches \( p(\theta = 0|x = 1) \) with \( p(x = 1|\theta = 0) \) to get his answer. Though the objective is appropriate, this analysis is simply wrong: \( p(\theta = 0|x = 1) \) is NOT equal to \( p(x = 1|\theta = 0) \). It must be computed by Bayes’ theorem, which involves additional information, namely both \( p(x = 1|\theta = 1) \) and \( p(\theta = 1) \). So, in addition to being just wrong, there is not enough information given to compute this (posterior) probability.

\( nb: \) this is called the “fallacy of the transposed conditional” – incorrectly matching the required posterior probability \( p(\theta = 0|x = 1) \) with the test error rate \( p(x = 1|\theta = 0) \) – and is a famous error that arises all the time among medics, lawyers, etc.
(b) It is certainly true that “a positive result is 100 times more likely for a drug user than for non-users” - this is direct as the likelihood ratio \( r = p(x = 1|\theta = 1)/p(x = 1|\theta = 0) = 100 \). However, the conclusion that “She is most probably a drug user” can only be based on actually computing \( p(\theta = 1|x = 1) \), and there is still a missing piece: the prior probability that she is a drug user, \( p(\theta = 1) \). This could be very, very small, in which case the posterior probability may be small too, even though the the likelihood ratio is 100.

**Exercise (5)**

(a) Given the sample size \( k \), the likelihood function depends on the data only through the value of \( \bar{x} \); therefore \( \bar{x} \) is a sufficient statistic for \( \theta \). This means that the investigator needs only to record \( \bar{x} \); not every single observation.

(b) Differentiate the log likelihood function and set to zero to solve for \( \hat{\theta} \) as stated. (An extra point if the second derivative is checked and found to be negative at the mode).

(c) Directly by Bayes’ theorem, the posterior density is proportional to \( \theta^{a+k-1} \exp(-\theta(b + k\bar{x})) \), implying that \( (\theta|X) \sim Gamma(a + k, b + k\bar{x}) \). The posterior mean is \( (a + k)/(b + k\bar{x}) \) which reduces to \( \hat{\theta} = 1/\bar{x} \) at \( a = b = 0 \).

(d) \( 1/\bar{x} = 0.267 \).

(e) Now \( (\theta|X) \sim Gamma(100, 375) \). Then \( \theta = \phi/375 \) where \( \phi \sim Gamma(100, 1) \), the standard gamma with shape parameter 100. Quantiles of \( \theta \) are therefore given by dividing the corresponding quantiles of \( \phi \) by 375. Hence the limits of the central 95\% posterior interval for \( \theta \) are 81.4/375 and 120.5/375, or 0.217 and 0.321.

(f) The classical value 0.212 lies below the lower 2.5\% point of the posterior for \( \theta \), hence this value is way down in the lower tail of the posterior, among the 5\% least likely values. It looks implausible, and \( \theta \) is very likely bigger: indeed, since \( Pr(\theta > 0.217|X) = 0.975 \) we know that \( Pr(\theta > 0.212|X) > 0.975 \).

(g) Approximate the posterior for \( \phi \) implied by the two, independent posteriors for \( \theta \) and \( \theta^* \) by generating a large sample of, say, \( k = 5000 \) values from the two posteriors, and computing the corresponding sample of values of \( \phi = 100(1 - \theta^*/\theta) \). As a point estimate, take the mean or median of this sample. For \( Pr(\phi > 0.03|Data) \) simply find the proportion of \( \phi \) values in the posterior sample that exceed 0.03.