<u>1st tutorial</u>

Solution: 1 (a)

At the analysis of the ECASS 2 study.

$n_0 = 100$

$y_0 = 8$

Here, $a_0 = y_0 + 1 = 8 + 1 = 9$

$b_0 = n_0 - y_0 + 1 = 100 - 8 + 1 = 93$

$\therefore \theta = \frac{y_0}{n_0} = \frac{8}{100} = 0.08$

Prior distribution follows the binomial distribution

$$P(y_0 \mid \theta) = \binom{n_0}{y_0} \theta^{y_0} (1-\theta)^{n_0 - y_0}$$

Distribution of θ follows the beta distribution.

$$P(\theta) = \frac{1}{B(a_0, b_0)} \theta^{a_0 - 1} (1-\theta)^{b_0 - 1}$$

Likelihood function of $P(y|\theta) \cdot P(\theta)$ follows

$$P(y|\theta) \cdot P(\theta) = \binom{n}{y} \theta^{y + a_0 - 1} (1-\theta)^{n - y + b_0 - 1}$$

$\therefore$ The posterior distribution

$$P(\theta \mid y) = \frac{P(y|\theta) P(\theta)}{P(y|\theta) \cdot P(\theta) + P(y|\theta^c) \cdot P(\theta^c)}$$

which follows beta distribution

$$\therefore P(y|\theta) \cdot P(\theta) = \binom{50}{10} \cdot \frac{1}{B(9,93)} (0.08)^{10+9-1} (1-0.08)^{50-10+93-1}$$

$$\therefore P(\theta|y) \approx B(19,133) = \frac{\Gamma 19 \, \Gamma 133}{\Gamma 19+133}$$

(b) Here given,

the posterior distribution of θ is a Beta $(19, 133)$ distribution.

Let, $a = 19$,
$b = 133$

$$\text{Posterior mean} = \frac{a}{a+b} = \frac{19}{19+133} = 0.125$$

$$\text{Posterior mode} = \frac{a-1}{a+b-1} = \frac{19-1}{19+133-1} = 0.12$$

$$\text{Posterior variance} = \frac{ab}{(a+b)^2 (a+b+1)}$$

$$= \frac{19 \times 133}{(19+133)^2 (19+133+1)}$$

$$= 7.1486 \times 10^{-4}$$

$\therefore$ Posterior variance $= 0.000714869$

CamScanner

~~(2)~~ a

Solution: 2 (a)

Deriving the posterior distribution when the data comes from binomial:

Let $X_1, X_2, \ldots, X_n$ from i.i.d random variables.

$$\sum X_i \sim \text{Binomial}$$

The prior distribution is from the beta distribution

$\sim \text{Beta}(\alpha, \beta)$

$$g(\theta) = \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

and the data comes from binomial distribution

$$g(y|\theta) = \binom{n}{y} \cdot \theta^{y} (1-\theta)^{n-y}$$

where, $g(y,\theta) = f(y|\theta) \cdot g(\theta)$

$$= \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$$

$$= \binom{n}{y} \cdot \frac{1}{B(\alpha,\beta)} \theta^{\alpha+y-1} (1-\theta)^{n+\beta-y-1}$$

then,

$$g(y) = \int_0^1 g(y,\theta)\, d\theta$$

$$= \int_0^1 \binom{n}{y} \frac{1}{B(\alpha,\beta)} \theta^{\alpha+y-1} (1-\theta)^{\beta+n-y-1}$$

$$= \frac{1}{B(\alpha,\beta)} \binom{n}{y} \int_0^1 \theta^{\alpha+y-1} (1-\theta)^{\beta+n-y-1}$$

$$g(y) = \frac{1}{B(\alpha,\beta)} \binom{n}{y} B(\alpha+y, \beta+n-y)$$

we know that,

Posterior distribution is $g(\theta|y) = \dfrac{g(y|\theta) \cdot g(\theta)}{g(y)} = \dfrac{g(y,\theta)}{g(y)}$

$$= \frac{\binom{n}{y} \frac{1}{B(\alpha,\beta)} \theta^{\alpha+y-1} (1-\theta)^{\beta+n-y-1}}{\frac{1}{B(\alpha,\beta)} \times \binom{n}{y} \times B(\alpha+y, \beta+n-y)}$$

$$= \frac{\theta^{\alpha+y-1} (1-\theta)^{\beta+n-y-1}}{B(\alpha+y, \beta+n-y)}$$

So ~~we can say that $g(\theta,y) \sim Beta(\alpha+y, \beta+n-y)$~~

$$g(\theta|y) = \frac{1}{B(\bar\alpha,\bar\beta)} \theta^{\bar\alpha-1} (1-\theta)^{\beta-1}$$

where, $\bar\alpha = \alpha+y$ and $\bar\beta = \beta+n-y$

So, we can say that $g(\theta,y) \sim Beta(\bar\alpha, \bar\beta)$

(b) From (a), we find the posterior distribution is,

$$g(\theta \mid y) = \frac{1}{B(\bar{\alpha},\bar{\beta})} \theta^{\bar{\alpha}-1} (1-\theta)^{\bar{\beta}-1}$$

$$\text{Mean} = \frac{1}{B(\bar{\alpha},\bar{\beta})} \int_0^1 \theta \cdot \theta^{\bar{\alpha}-1} (1-\theta)^{\bar{\beta}-1} \, d\theta$$

$$= \frac{1}{B(\bar{\alpha},\bar{\beta})} \int_0^1 \theta^{\bar{\alpha}+1-1} (1-\theta)^{\bar{\beta}-1} \, d\theta$$

$$= \frac{B(\bar{\alpha}+1,\bar{\beta})}{B(\bar{\alpha},\bar{\beta})}$$

$$= \frac{\Gamma(\bar{\alpha}+1)\,\Gamma\bar{\beta}}{\Gamma(\bar{\alpha}+\bar{\beta}+1)} \times \frac{\Gamma(\bar{\alpha}+\bar{\beta})}{\Gamma\bar{\alpha}\,\Gamma\bar{\beta}}$$

$$= \frac{\bar{\alpha}\,\Gamma\bar{\alpha}\,\Gamma\bar{\beta}}{(\bar{\alpha}+\bar{\beta})\,\Gamma(\bar{\alpha}+\bar{\beta})} \times \frac{\Gamma(\bar{\alpha}+\bar{\beta})}{\Gamma\bar{\alpha}\,\Gamma\bar{\beta}}$$

$$= \frac{\bar{\alpha}}{\bar{\alpha}+\bar{\beta}}$$

$$= \frac{1}{1+\cancel{\bar{\beta}/\beta}}$$

$$\therefore \text{Mean} = \frac{1}{1+\frac{\bar{\beta}}{\bar{\alpha}}}$$

Mode: By using marginal parameter.

$$f'(\theta) = \frac{df(\theta)}{d\theta} = \frac{1}{B(\bar{\alpha},\bar{\beta})} \frac{d}{d\theta} \theta^{\bar{\alpha}-1} (1-\theta)^{\bar{\beta}-1}$$

$$= \frac{1}{B(\bar{\alpha},\bar{\beta})} \left[(\bar{\alpha}-1)\theta^{(\bar{\alpha}-2)} (1-\theta)^{\bar{\beta}-1} - \theta^{\bar{\alpha}-1} (\bar{\beta}-1)(1-\theta)^{\bar{\beta}-2} \right]$$

$$= \frac{1}{B(\bar{\alpha},\bar{\beta})} \left[\theta^{\bar{\alpha}-2} (1-\theta)^{\bar{\beta}-2} \left\{ (\bar{\beta}-1)(1-\theta) - (\bar{\beta}-1)(\theta) \right\} \right]$$

$$f'(\theta) = \frac{1}{B(\bar{\alpha},\bar{\beta})} \theta^{\bar{\alpha}-2} (1-\theta)^{\bar{\beta}-2} \left[(1-\theta)(\bar{\alpha}-1) - (\bar{\beta}-1)\theta \right]$$

We know,

$$f'(\theta) = 0$$

$$\Rightarrow \frac{1}{B(\bar{\alpha},\bar{\beta})} \theta^{\bar{\alpha}-2} (1-\theta)^{\bar{\beta}-2} \left[(1-\theta)(\bar{\alpha}-1) - (\bar{\beta}-1)\theta \right] = 0$$

$$\Rightarrow (1-\theta)(\bar{\alpha}-1) - (\bar{\beta}-1)\theta = 0$$

$$\Rightarrow \bar{\alpha} - 1 - \theta\bar{\alpha} + \theta - \bar{\beta}\theta + \theta = 0$$

$$\Rightarrow \bar{\alpha} - 1 - \theta[\bar{\alpha} - 1 + \bar{\beta} - 1] = 0$$

$$\Rightarrow \theta = \frac{\bar{\alpha}-1}{\bar{\alpha}+\bar{\beta}-2}$$

So the mode of beta distribution is $\frac{\bar{\alpha}-1}{\bar{\alpha}+\bar{\beta}-2}$

Variance from posterior distribution:

$$E(\theta^2) = \int_0^1 \theta^2 f(\theta; \bar{\alpha}, \bar{\beta})\, d\theta$$

$$= \frac{1}{B(\bar{\alpha}, \bar{\beta})} \int_0^1 \theta^2 \cdot \theta^{\bar{\alpha}-1} (1-\theta)^{\bar{\beta}-1} d\theta$$

$$= \frac{1}{B(\bar{\alpha}, \bar{\beta})} \int_0^1 \theta^{\bar{\alpha}+2-1} (1-\theta)^{\bar{\beta}-1} d\theta$$

$$= \frac{B(\bar{\alpha}+2, \bar{\beta})}{B(\bar{\alpha}, \bar{\beta})}$$

$$= \frac{\Gamma\bar{\alpha}+2\ \Gamma\cancel{\bar{\beta}}}{\Gamma\bar{\alpha}+\bar{\beta}+2} \times \frac{\Gamma\bar{\alpha}+\bar{\beta}}{\Gamma\bar{\alpha} * \Gamma\cancel{\bar{\beta}}}$$

$$= \frac{(\bar{\alpha}+1)\Gamma\bar{\alpha}+1\ \Gamma\cancel{\bar{\beta}}}{(\bar{\alpha}+\bar{\beta}+1)(\bar{\alpha}+\bar{\beta})\Gamma\cancel{(\bar{\alpha}+\bar{\beta})}} \quad \frac{\cancel{\Gamma\bar{\alpha}\ \Gamma\bar{\beta}}\ \Gamma(\bar{\alpha}+\bar{\beta})}{\Gamma\bar{\alpha}\ \Gamma\cancel{\bar{\beta}}}$$

$$= \frac{(\bar{\alpha}+1)\bar{\alpha}\,\Gamma\bar{\alpha}}{(\bar{\alpha}+\bar{\beta}+1)(\bar{\alpha}+\bar{\beta})\,\Gamma\bar{\alpha}}$$

$$\therefore E(\theta^2) = \frac{\bar{\alpha}(\bar{\alpha}+1)}{(\bar{\alpha}+\bar{\beta}+1)(\bar{\alpha}+\bar{\beta})}$$

~~From~~ we get, $E(\theta) = \mu = \dfrac{\bar{\alpha}}{\bar{\alpha}+\bar{\beta}}$

$$\therefore Var(\theta) = E(\theta^2) - \{E(\theta)\}^2$$

$$= \frac{\bar{\alpha}(\bar{\alpha}+1)}{(\bar{\alpha}+\bar{\beta}+1)(\bar{\alpha}+\bar{\beta})} - \frac{\bar{\alpha}^2}{(\bar{\alpha}+\bar{\beta})^2}$$

$$= \frac{\bar{\alpha}(\bar{\alpha}+1)(\bar{\alpha}+\bar{\beta}) - \bar{\alpha}^2(\bar{\alpha}+\bar{\beta}+1)}{(\bar{\alpha}+\bar{\beta}+1)(\bar{\alpha}+\bar{\beta})^2}$$

$$= \frac{\cancel{\bar{\alpha}^3} + \cancel{\bar{\alpha}^2} + \cancel{\bar{\alpha}^2\bar{\beta}} + \bar{\alpha}\bar{\beta} - \cancel{\bar{\alpha}^3} - \cancel{\bar{\alpha}^2\bar{\beta}} - \cancel{\bar{\alpha}^2}}{(\bar{\alpha}+\bar{\beta}+1)(\bar{\alpha}+\bar{\beta})^2}$$

$$\therefore Var[\theta] = \frac{\bar{\alpha}\bar{\beta}}{(\bar{\alpha}+\bar{\beta}+1)(\bar{\alpha}+\bar{\beta})^2}$$

This is the variance of posterior distribution.

Solution: 3

In the posterior distribution, there are two types of credible interval

a) Equal ~~two~~ tail credible interval.

b) Highest posterior density interval.

(a) Equal tail credible interval: $100(1-\alpha)\%$ of equal tail credible interval as $100(1-\alpha)\%$ credible interval in which:

$$P[a \leq \theta \leq b] = 1-\alpha$$

$$\Rightarrow F(a) - F(b)$$

The function of equal tail credible interval are same value of θ. Many contain lower value of $P(\theta|y)$ that the value of θ outside the interval. So, we need to construct the highest posterior density interval.

(b) <u>Highest Posterior density function:</u>

A $100(1-\alpha)\%$ HPD is a $100(1-\alpha)\%$ credible interval for which,

for all $\theta_1 \in [a,b]$ and for all $\theta_2 \notin [a,b]$,

$$P(\theta_1|y) \geq P(\theta_2|y)$$

So, HPD interval is a range that contains the value of θ. That is a most plausible posterior of HPD. HPD interval is a interval of evidence that based on posterior distribution but not the likelihood. HPD interval needs the density function.

A 95% credible interval is an interval that contains 95% most plausible. The 95% confidence interval is n contains or doesn't contains. The 95% credible interval in the given data. The interpretation are

A 95% confidence interval is an interval that contains does or doesn't value of the posterior in the given data set.

This is the example of confidence interval and credible interval.

Comparison of credible and confidence interval

Confidence Interval (CI)	Credible Interval (CrI)
1) Confidence interval based on frequentist inference.	1) Credible interval based on bayesian inference
2) In confidence interval parameter θ is fixed but unknown	2) In credible interval parameter θ is treated as random.
3) Used only sample data	3) Uses data and prior belief.
4) 95% CI refers to long-run frequency of the method	4) 95% CrI gives direct probability about θ
5) Confidence interval derived from sampling distribution	5) Credible interval derived from posterior distribution
6) Does not use prior information	6) Uses prior information

Hypothetical example:

A study surveys $n = 100$ students to estimate the proportion θ satisfied with a new teaching method.

Observed satisfied students: $y = 45$

sample proportion : $\hat{p} = 0.45$

95% confidence Interval

$$CI = \hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 0.45 \pm 1.96 \sqrt{\frac{0.45 \times 0.55}{100}}$$

$$= 0.45 \pm 1.96\,(0.0498)$$

$$= 0.45 \pm 0.0976$$

$$CI = [0.352,\ 0.547]$$

If we repeat this survey infinitely many times and compute 95% CI each time using the same method, approximately 95% of those intervals will contain the true parameter θ. ~~That~~ This does not imply that the probability of θ lying in the interval $[0.352, 0.547]$ is 0.95.

95% Credible Interval

Assume prior: $\theta \sim \text{Beta}(1,1)$

Posterior becomes: $\theta | y \sim \text{Beta}(1+45, 1+55) = \text{Beta}(46, 56)$

Posterior mean: $E(\theta | y) = \dfrac{46}{46+56} = 0.451$

Approximately 95% equal-tailed credible interval from posterior quantiles

$$CrI = [0.352, 0.547]$$

Given the observed data and prior belief, there is a 95% probability that the true proportion θ lies in the interval $[0.352, 0.547]$

Solution (4):

Use of credible interval in Bayesian hypothesis testing:

1) In Bayesian inference, a $100(1-\alpha)\%$ credible interval satisfies:

$$P(a < \theta < b \mid data) = 1-\alpha$$

2) In hypothesis testing credible interval is used to assess whether the null hypothesis value is ~~plau~~ plausible under the posterior distribution.

- If the hypothesized value θ_0 is not contained in the CrI, we infer that H_0 is unlikely given the observed data

3) Unlike confidence intervals, it does not rely on repeated sampling. it measures posterior certainty.

Use of contour probability in Bayesian hypothesis testing:

1) A contour probability region represents the denset area of the posterior distribution.

2) The highest posterior density region is constructed so that:

$$P(\theta \in HPD \mid data) = 1 - \alpha$$

and every point inside has higher posterior density than points outside.

3) Used in Bayesian hypothesian testing ~~testing~~ to determine where the parameter most plausibly lies, helping to assess the strength of evidence for or against a hypothesis

Comparison between Bayesian and frequentist hypothesis testing

Bayesian	Frequentist
1) Hypothesis and parameter are treated as random	1) Parameter is fixed but data is random

Bayesian	Frequentist
2) Uses posterior probability, bayes factor, credible interval	2) Uses test statistic, p-value.
3) Can accept H_0 if posterior support is high	3) Cannot accept H_0; only reject or fail to reject
4) Includes prior knowledge	4) No prior ~~knowledge~~ knowledge is used.
5) Interval means: $P(a < \theta < b \mid \text{data})$	5) CI means: long-run frequency, no probability on θ.

Example:

Suppose a coin is tossed 20 times and 14 heads are observed. let

$H_0: \theta = 0.5$

Frequentist Approach

perform a two-sided binomial test

· Suppose the p-values is $P = 0.041$

Since $p < 0.05$, we reject H_0.

If the coin is fair, observing data as extreme as this would be rare in repeated sample.

Bayesian Approach:

Assume a prior $\theta \sim Beta(1,1)$

· Posterior become $Beta(15, 7)$

· Compute a 95% HPD interval is $[0.53, 0.88]$ which does not contain 0.5

there is posterior evidence against H_0.

Given the data and prior, it is unlikely that the coin is fair