# Solutions to the Exercises

## Chapter 1

1. A coffee shop buys roasted coffee from a supplier. In order to assess the quality of the supplied coffee, the manager of the shop conducts a tasting experiment where she selects a small portion of coffee beans from different batches and tastes the coffee from each portion. For each portion she gives a score in the scale $1, 2, \ldots, 10$ with 10 corresponding to coffee of the best taste and uses the results to assess the quality of the coffee.

   Identify the population, parameter, and statistic.

   **Population:** In this case we want to draw conclusions about all roasted coffee delivered from the supplier.

   **Parameter:** In this tasting experiment we give a score to the taste from each batch. The parameter could be the average score from all possible batches even those which we didn't taste. Because each tasting experiment is a score from 1 to 10, the parameter space is the interval $[1, 10]$.

   **Statistic:** The statistic that helps us estimate the value of the parameter is the sample average. The possible values of the statistic are in the range of 1 to 10. It is not easy to come up with the distribution of the test statistic. If we assume that the central limit theorem applies, then the distribution of the sample average is approximately the normal distribution.

2. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$. Derive the sampling distribution of $\bar{X}$ given in Example 1.3.

   Because each $X_i \sim N$ and $\bar{X}$ is a linear combination of the $X_i$'s, the distribution of $\bar{X}$ will also be normal. To find the mean and variance, we write $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ so

$$
\begin{aligned}
\mathsf{E}\,\bar{X} &= \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) & \mathsf{Var}\,\bar{X} &= \mathsf{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n}\,\mathsf{E}\left(\sum_{i=1}^n X_i\right) & &= \frac{1}{n^2}\,\mathsf{Var}\left(\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^n \mathsf{E}\,X_i & &= \frac{1}{n^2}\sum_{i=1}^n \mathsf{Var}\,X_i, \text{ by independence} \\
&= \frac{1}{n}\sum_{i=1}^n \mu & &= \frac{1}{n^2}\sum_{i=1}^n \sigma^2 \\
&= \frac{1}{n} n\mu = \mu & &= \frac{1}{n^2} n\sigma^2 = \sigma^2/n.
\end{aligned}
$$

3. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$.

   a) Derive the sampling distribution of $\bar{X}$.

   b) Derive the asymptotic distribution of $\bar{X}$ from the central limit theorem.

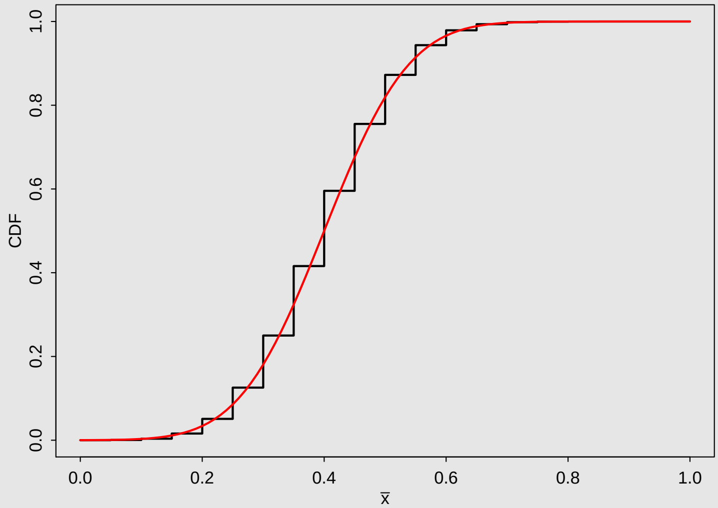
c) Draw a graph of the exact and approximate CDFs when $n = 20$ and $p = 0.4$.

a) Let $W = \sum X_i$. Then $W \sim \text{Bin}(n, p)$ and $\mathbb{P}(W = w) = \binom{n}{w} p^w (1-p)^{n-w}$.

Therefore, $\mathbb{P}(\bar{X} = x) = \mathbb{P}(W = nx) = \binom{n}{nx} p^{nx}(1-p)^{n-nx}$. Thus, $\bar{X}$ is the sample proportion from $n$ Bernoulli trials.

b) The mean and variance of each $X_i$ is $\mathsf{E}\, X_i = p$ and $\mathsf{Var}\, X_i = p(1-p)$. By the central limit theorem $\bar{X} \sim N\left(p, \dfrac{p(1-p)}{n}\right)$ approximately.

c) The two CDFs can be seen in the figure below




with the red line denoting the approximate normal CDF and the black line the exact binomial proportion. The exact CDF is a step function because the random variable is discrete.

**Chapter 2**

1. Consider the method of moments estimator of Example 2.6. Identify potential drawbacks of this estimator.

This estimator can sometimes be nonsense. For example, suppose we observe $x_1 = 1$, $x_2 = 1$, $x_3 = 7$. Then $\bar{x} = 3$ and so $\hat{\theta} = 2\bar{x} = 6$. However we know that $\theta$ must be bigger than all $x_i$'s so $\theta > 7$ and $\hat{\theta} < 7$.

2. Verify the formula from the mgf of the gamma distribution in Example 2.7 and use it to derive its mean and variance.

For the Gamma$(\alpha, \beta)$ distribution,

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$

Then

$$M_X(t) = \int_0^\infty e^{tx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\, \mathrm{d}x$$

$$
\begin{aligned}
&= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\beta-t)x}\,\mathrm{d}x \\
&= \frac{\beta^\alpha}{(\beta-t)^\alpha} \int_0^\infty \frac{(\beta-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\beta-t)x}\,\mathrm{d}x \\
&= \frac{\beta^\alpha}{(\beta-t)^\alpha}, \text{ for } t < \beta. \\
&= (1-t/\beta)^{-\alpha}, \text{ for } t < \beta. \\
\Rightarrow M_X^{(1)}(t) &= \frac{d}{dt} M_X(t) = \frac{\alpha}{\beta}(1-t/\beta)^{-\alpha-1} \\
\Rightarrow \mu_1 &= M_X^{(1)}(0) = \frac{\alpha}{\beta} \\
\Rightarrow M_X^{(2)}(t) &= \frac{d^2}{dt^2} M_X(t) = \frac{\alpha(\alpha+1)}{\beta^2}(1-t/\beta)^{-\alpha-2} \\
\Rightarrow \mu_2 &= M_X^{(2)}(0) = \frac{\alpha(\alpha+1)}{\beta^2} \\
\Rightarrow \mathsf{Var}(X) &= \mu_2 - \mu_1^2 = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}.
\end{aligned}
$$

3. Explain why the MLE of Example 2.10 is biased but do not derive its bias.

Because each $X_i < \theta$, then $X_{(n)} = \max\{X_1, \ldots, X_n\} < \theta$, so $\mathsf{E}\, X_{(n)} < \theta$. Therefore $\text{Bias}_\theta(X_{(n)}) = \mathsf{E}\, X_{(n)} - \theta < 0$. So $\text{Bias}_\theta(X_{(n)}) \neq 0$ which means that $X_{(n)}$ is biased for $\theta$.

4. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$, $\theta \in (0,1)$. Derive the MLE for $\theta$.

The pmf for $X$ is

$$f(x|\theta) = \theta^x (1-\theta)^{1-x},\ x \in \{0,1\},\ \theta \in [0,1].$$

Taking logarithms,

$$
\begin{aligned}
\log f(x|\theta) &= x \log\theta + (1-x)\log(1-\theta) \\
\Rightarrow \ell(\theta|\mathbf{x}) &= \sum_{i=1}^n \log f(x_i|\theta) \\
&= \sum_{i=1}^n \{x_i \log\theta + (1-x_i)\log(1-\theta)\} \\
&= \log\theta \sum_{i=1}^n x_i + \log(1-\theta)(n - \sum_{i=1}^n x_i) \\
\Rightarrow \ell'(\theta|\mathbf{x}) &= \frac{1}{\theta}\sum_{i=1}^n x_i - \frac{1}{1-\theta}(n - \sum_{i=1}^n x_i) \\
0 &= \frac{1}{\hat\theta}\sum_{i=1}^n x_i - \frac{1}{1-\hat\theta}(n - \sum_{i=1}^n x_i) \\
0 &= \frac{n}{\hat\theta}\bar{x} - \frac{n}{1-\hat\theta}(1-\bar{x}) \\
0 &= \frac{\bar{x}}{\hat\theta} - \frac{1-\bar{x}}{1-\hat\theta}
\end{aligned}
$$

$$\hat{\theta} = \bar{x}$$

5. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ where both $\mu$ and $\sigma^2$ are unknown parameters. Derive a level $1-\alpha$ confidence interval for $\sigma^2$.

We know that $\dfrac{(n-1)S^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$ so $Y = \dfrac{(n-1)S^2}{\sigma^2}$ is a pivot function because its distribution does not depend on the unknown parameter $\sigma^2$. Let $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ denote the $\alpha/2$ and $1-\alpha/2$ left quantiles of the $\mathcal{X}^2_{n-1}$ distribution respectively. Then we have the following inequality which holds with probability $1-\alpha$

$$\begin{aligned}
&\chi^2_{\alpha/2} < Y < \chi^2_{1-\alpha/2} \\
&\Rightarrow \chi^2_{\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2} \\
&\Rightarrow \frac{1}{\chi^2_{1-\alpha/2}} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi^2_{\alpha/2}} \\
&\Rightarrow \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}},
\end{aligned}$$

so a level $1-\alpha$ confidence interval is $\left[\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}}, \dfrac{(n-1)S^2}{\chi^2_{\alpha/2}}\right]$

6. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{U}(0, \theta)$, $\theta > 0$. By considering an appropriate pivot construct a level $1-\alpha$ confidence interval for $\theta$.

Let $W_i = X_i/\theta$. Then $W_i \sim U(0,1)$ so if we let $W_{(n)} = \max\{W_1, \ldots, W_n\}$, then the distribution of $W_{(n)}$ does not depend on $\theta$. Note that by letting $X_{(n)} = \max\{X_1, \ldots, X_n\}$, $W_{(n)} = X_{(n)}/\theta$.

The CDF of $W_{(n)}$ is

$$\begin{aligned}
F_W(w) &= \mathbb{P}(W_{(n)} \leq w) \\
&= \mathbb{P}(\max\{W_1, \ldots, W_n\} \leq w) \\
&= \mathbb{P}(W_1 \leq w, \ldots, W_n \leq w) \\
&= \mathbb{P}(W_1 \leq w) \ldots \mathbb{P}(W_n \leq w), \text{ by independence} \\
&= w \times \ldots \times w \\
&= w^n
\end{aligned}$$

So if we let $c_1 = (\alpha/2)^{\frac{1}{n}}$ and $c_2 = (1-\alpha/2)^{\frac{1}{n}}$, with probability $1-\alpha$,

$$\begin{aligned}
&c_1 < W_{(n)} < c_2 \\
&c_1 < X_{(n)}/\theta < c_2 \\
&\frac{1}{c_2} < \theta/X_{(n)} < \frac{1}{c_1} \\
&\frac{X_{(n)}}{c_2} < \theta < \frac{X_{(n)}}{c_1},
\end{aligned}$$

so a level $1-\alpha$ confidence interval is $\left[\frac{X_{(n)}}{c_2}, \frac{X_{(n)}}{c_1}\right]$

## Chapter 3

1. Consider the setup of Example 3.3 but with $\mu_1 < \mu_0$. Derive the hypothesis test in this case for $\sigma^2$ known and for $\sigma^2$ unknown.

If $\mu_1 < \mu_0$, then the likelihood ratio, $\frac{L(\mu_1|\mathbf{x})}{L(\mu_0|\mathbf{x})}$ is a decreasing function of $\bar{x}$. Therefore, for every rule of the form

$$\text{Reject H}_0 \text{ if } \frac{L(\mu_1|\mathbf{x})}{L(\mu_0|\mathbf{x})} > c,$$

for a critical value $c$, there is an equivalent rule of the form

$$\text{Reject H}_0 \text{ if } \bar{x} < c^*,$$

for a different critical value $c^*$. We proceed the same way taking care to use $<$ where we had $>$. This suggests a rule

$$\text{Reject H}_0 \text{ if } z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < z_\alpha.$$

when $\sigma^2$ is known and

$$\text{Reject H}_0 \text{ if } t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < t_{n-1;\alpha}.$$

when $\sigma^2$ is unknown.

2. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$, $\theta \in (0,1)$. Derive a hypothesis test for testing $\text{H}_0\colon \theta = \theta_0$ v.s. $\text{H}_1\colon \theta > \theta_0$.

We consider the equivalent hypothesis test

$$\text{H}_0\colon \theta = \theta_0 \text{ v.s. } \text{H}_1\colon \theta = \theta_1 \text{ for some } \theta_1 > \theta_0.$$

The likelihood function is given by

$$L(\theta|\mathbf{x}) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}.$$

so the likelihood ratio is

$$\begin{aligned}
\frac{L(\theta_1|\mathbf{x})}{L(\theta_0|\mathbf{x})} &= \left(\frac{\theta_1}{\theta_0}\right)^{\sum x_i} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-\sum x_i} \\
&= \left(\frac{\theta_1}{\theta_0}\right)^{\sum x_i} \left(\frac{1-\theta_0}{1-\theta_1}\right)^{\sum x_i} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n} \\
&= \left(\frac{\theta_1/(1-\theta_1)}{\theta_0/(1-\theta_0)}\right)^{\sum x_i} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n}.
\end{aligned}$$

Because $\theta_1 > \theta_0 \Rightarrow \theta_1/(1-\theta_1) > \theta_0/(1-\theta_0)$. This shows that the likelihood ratio is an increasing function of $\sum x_i$. Therefore, for every rule of the form

$$\text{Reject } \mathrm{H}_0 \text{ if } \frac{L(\theta_1|\mathbf{x})}{L(\theta_0|\mathbf{x})} > c,$$

for a critical value $c$, there is an equivalent rule of the form

$$\text{Reject } \mathrm{H}_0 \text{ if } \sum x_i > c^*,$$

for a different critical value $c^*$. If we let $W = \sum X_i$, the distribution of $W$ is binomial $W \sim \text{Bin}(n, \theta)$. So we choose the critical value $c^*$ such that $\mathbb{P}(W \leq c^*) = 1 - \alpha$ where $W \sim \text{Bin}(n, \theta)$.

The $p$-value of the test is, for an observed count $w = \sum x_i$,

$$p\text{-value} = \mathbb{P}(W > w),$$

where $W \sim \text{Bin}(n, \theta)$.

3. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$, $\theta \in (0,1)$. Derive a hypothesis test for testing $\mathrm{H}_0$: $\theta = \theta_0$ v.s. $\mathrm{H}_1$: $\theta \neq \theta_0$.

Continuing from the previous exercise, we consider first the hypothesis tests

$$\mathrm{H}_0\colon\ \theta = \theta_0 \text{ v.s. } \mathrm{H}_2\colon\ \theta = \theta_1 \text{ for some } \theta_1 > \theta_0.$$

and

$$\mathrm{H}_0\colon\ \theta = \theta_0 \text{ v.s. } \mathrm{H}_3\colon\ \theta = \theta_1 \text{ for some } \theta_1 < \theta_0.$$

The rule for the first test at level $\alpha_2$ is

$$\text{Reject } \mathrm{H}_0 \text{ if } \sum x_i > c^*,$$

and for the second test at level $\alpha_3$ is

$$\text{Reject } \mathrm{H}_0 \text{ if } \sum x_i < c^{**},$$

with $c^*$ and $c^{**}$ such that

$$\mathbb{P}(W \leq c^*) = 1 - \alpha_2 \text{ and } \mathbb{P}(W < c^{**}) = \alpha_3,$$

where $W \sim \text{Bin}(n, \theta)$. We choose $\alpha_2 = \alpha_3 = \alpha/2$. Therefore, the rule for the original test is

$$\text{Reject } \mathrm{H}_0 \text{ if } \sum x_i < c^{**} \text{ or if } \sum x_i > c^*.$$

Suppose $w = \sum x_i$ is the observed count. Then $p$-value $= 2\min\{p_2, p_3\}$, where $p_2 = \mathbb{P}(W > w)$ and $p_3 = \mathbb{P}(W < w)$, where $W \sim \text{Bin}(n, \theta)$.

4. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$. Derive a hypothesis test for testing $\mathrm{H}_0$: $\mu = \mu_0$ v.s. $\mathrm{H}_1$: $\mu \neq \mu_0$ for the cases where $\sigma^2$ is known and when it is not. What is the $p$-value of the test?

We consider the two separate hypothesis tests

$$\mathrm{H}_0\colon\ \mu = \mu_0 \text{ v.s. } \mathrm{H}_2\colon\ \mu < \mu_0 \text{ at level } \alpha_2$$

and

$$\mathrm{H}_0\colon\ \mu = \mu_0 \text{ v.s. } \mathrm{H}_3\colon\ \mu > \mu_0 \text{ at level } \alpha_3$$

with $\alpha_2 + \alpha_3 = \alpha$.

In the case where $\sigma^2$ is known, it is easy to see from Example 3.3 that $\mathrm{H}_2$ is accepted if $z < z_{\alpha_2}$ and $\mathrm{H}_3$ is accepted if $z > z_{1-\alpha_3}$, where

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

We can take $\alpha_2 = \alpha_3 = \alpha/2$ so the hypothesis test for $\mathrm{H}_1$ is

$$\text{Reject } \mathrm{H}_0 \text{ if } z < z_{\alpha/2} \text{ or if } z > z_{1-\alpha/2}.$$

Because $z_{\alpha/2} = -z_{1-\alpha/2}$ this test can also be written as

$$\text{Reject } \mathrm{H}_0 \text{ if } |z| > z_{1-\alpha/2}.$$

In the case where $\sigma^2$ is unknown, then

$$t = \frac{z}{\sqrt{\dfrac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

Proceeding the same way as in the first case, our hypothesis test is

$$\text{Reject } \mathrm{H}_0 \text{ if } |t| > t_{n-1;1-\alpha/2}.$$

## Chapter 4

1. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$, $\theta \in [0, 1]$ with prior distribution $\theta \sim \text{Beta}(\alpha, \beta)$ for given values of $\alpha$ and $\beta$.
   a) Derive the posterior distribution for $\theta$ in this case.
   b) Suppose we choose $\alpha = 2$, $\beta = 4$ and, after $n = 10$ trials, we observe 7 successes. Derive the posterior estimate for $\theta$ and use Python to get a 95% confidence interval.

a) The prior pdf for $\theta$ is

$$\pi(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},\ \theta \in (0,1).$$

The pdf for $x$ is

$$f(x|\theta) = \theta^x(1-\theta)^{1-x},$$

so

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}.$$

Multiplying the two, we have

$$f(\mathbf{x}|\theta)\pi(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\sum x_i+\alpha-1}(1-\theta)^{n-\sum x_i+\beta-1}.$$

By integrating,

$$\int_0^1 f(\mathbf{x}|\theta)\pi(\theta)\,\mathrm{d}\theta = \frac{B(\sum x_i+\alpha, n-\sum x_i+\beta)}{B(\alpha,\beta)},$$

so, the posterior pdf is

$$\pi(\theta|\mathbf{x}) = \frac{1}{B(\sum x_i+\alpha, n-\sum x_i+\beta)}\theta^{\sum x_i+\alpha-1}(1-\theta)^{n-\sum x_i+\beta-1},$$

which can be recognised as the $\mathrm{Beta}(\sum x_i+\alpha, n-\sum x_i+\beta)$ distribution.

b) Let $\alpha' = \sum x_i + \alpha = 7+2 = 9$, $\beta' = n - \sum x_i + \beta = 10-7+4 = 7$. Then, we find that the posterior mean is $\tilde{\theta} = \frac{9}{9+7} = 0.5625$.
Using Python, we have

```
>>> from scipy.stats import beta
>>> beta.ppf([0.025,0.975],9,7)
array([ 0.32286977,  0.78733327])
```

so the confidence interval is $[0.323, 0.787]$.

2. Let $X_1,\ldots,X_n \overset{\text{iid}}{\sim} \mathrm{Poisson}(\theta)$, $\theta > 0$, with prior distribution $\theta \sim \mathrm{Gamma}(\alpha,\beta)$ for given values of $\alpha$ (shape parameter) and $\beta$ (rate parameter).

a) Derive the posterior distribution for $\theta$ in this case.

b) Suppose we choose $\alpha = 1$, $\beta = 1$ and, with $n = 5$, we observe $\sum x_i = 3.4$. Derive the posterior estimate for $\theta$ and use Python to get a 95% confidence interval.

a) The prior pdf for $\theta$ is

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta},\ \theta \in (0,\infty).$$

The pdf for $x$ is

$$f(x|\theta) = \frac{\theta^x}{x!}e^{-\theta}$$

so

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \frac{\theta^{\sum x_i}}{\prod x_i!}e^{-n\theta}.$$

Multiplying the two, we have

$$f(\mathbf{x}|\theta)\pi(\theta) = \frac{\beta^\alpha}{\prod x_i!\Gamma(\alpha)}\theta^{\sum x_i+\alpha-1}e^{-(\beta+n)\theta}.$$

This product has the form

$$f(\mathbf{x}|\theta)\pi(\theta) = C\theta^{A-1}e^{-B\theta},$$

with $A = \sum x_i + \alpha$ and $B = \beta + n$ while $C$ does not depend on $\theta$. This can be recognised as a form of the gamma distribution with shape $A$ and rate $B$. Therefore the posterior distribution of $\theta$ is

$$\theta|\mathbf{x} \sim \text{Gamma}(\sum x_i + \alpha, \beta + n).$$

b) Let $A = \sum x_i + \alpha = 3.4 + 1 = 4.4$, $\beta' = \beta + n = 1 + 5 = 6$. Then, we find that the posterior mean is $\tilde{\theta} = \frac{A}{B} = \frac{4.4}{6} = 0.733$.
Using Python, we have

```
>>> from scipy.stats import gamma
>>> gamma.ppf([0.025,0.975],a=4.4,scale=1.0/6.0)
array([ 0.21617058,  1.5606256 ])
```

so the confidence interval is $[0.216, 1.561]$.