MAS2903: Introduction to Bayesian Methods

Solutions to all other questions in Chapter 5

Once you have submitted both assignments, solutions to the assessed questions from Chapter 5 (**questions 4, 9, 10, 11, 12, 14, 30, 40 and 41**) will be available in Blackboard.

By the end of semester 2, we will have worked through **questions 5, 6, 7, 8, 18, 27, 29, 33, 34 and 35** in the fortnightly problems classes; slides with solutions can be be found on the course webpage and Blackboard, along with ReCap recordings.

Students' solutions to **questions 2, 3, 12, 13, 15, 16, 19 and 24** can be found in PowerPoint slides on the course webpage.

In this document, you will find solutions to all remaining questions from Chapter 5.

1. (a) Probability estimates – values to be placed on the $0 \longrightarrow 1$ scale:

(i) Frequency interpretation: take a sample of students from the class (your friends, perhaps?) and estimate the probability as the proportion of students in your sample whose BMI is at least 22.5. Could also use a subjective assessment, just by "looking", or by using information given in the lecture notes about the NU BMI study.

(ii) Subjective assessment: For example, even if you don't know much about football, you can easily check to see where Newcastle are in the league right now – currently 8th top. Without any further research, you could then specify a high probability, given that they are only two places from a top six position. After more research, and perhaps using your own knowledge about the Premier League, you might think a top six position at the end of the season is unlikely, given their up-and-coming games – thus specifying a much lower probability.

(iii) Classical interpretation: because we "randomly select", each number is equally likely. The probability is then 13/90.

(b) Interval estimates:

(i) We could repeat the experiment 100 times to get 100 estimates of the probability. Reading off the 2.5 and 97.5 percentiles could then give an estimate of the probability interval.

(ii) The estimate itself is subjective, so the range will be too. The range should reflect your degree of certainty – the more you know about football, and Newcastle United, the narrower your range!

(iii) Since we have a fixed probability, the range will be $(13/90, 13/90)$.

17. From the sample, we have

$$n = 27; \quad \bar{x} = 2.4685.$$

From the question, we have

$$\begin{aligned} \text{Precision for the data} &= \tau = 1/0.27^2 \\ \text{Prior mean} &= b = 2.7 \\ \text{Prior precision} &= d = 1/0.3^2. \end{aligned}$$

From Example 2.6 in the lecture notes, we know that $\mu|\boldsymbol{x} \sim N(B, 1/D)$, where

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

giving

$$D = 1/0.3^2 + \frac{27}{0.27^2} = 381.4815$$

and

$$B = \frac{2.7/0.3^2 + 27 \times 2.4685/0.27^2}{1/0.3^2 + 27/0.27^2} = 2.4752.$$

Thus we have

$$\mu|\boldsymbol{x} \sim N(2.4752, 1/381.4815).$$

Therefore,

$$\Pr(\mu < 2.5) = \Pr\left(Z < \frac{2.5 - 2.4752}{\sqrt{1/381.4815}}\right) = \Pr(Z < 0.4844) = \Phi(0.4844).$$

From tables, or using R (i.e. `pnorm(0.4844)`), we get $\Pr(\mu < 2.5) = 0.6859$.

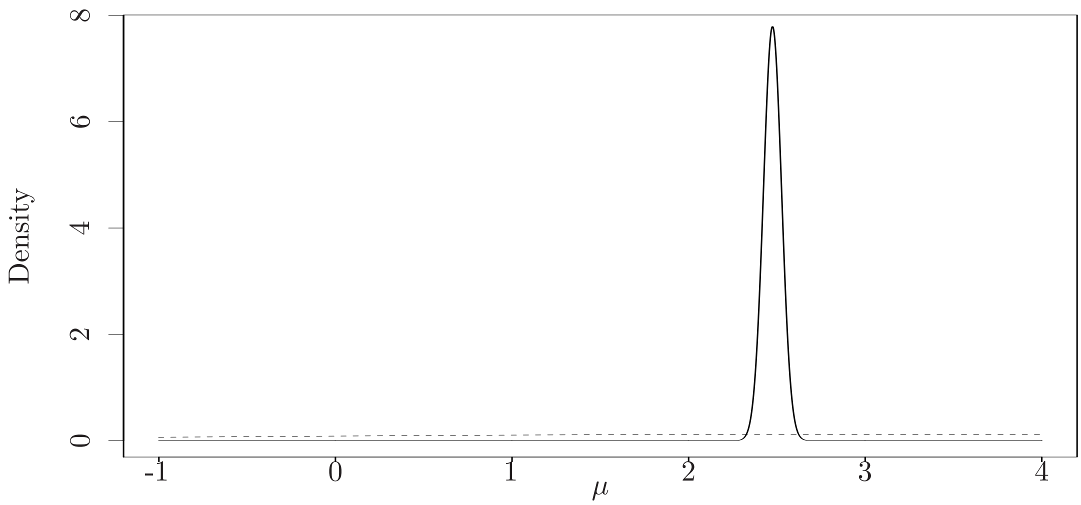
In summary, we have:

	Prior	Likelihood	Posterior
Mode	2.7	$\bar{x} = 2.4685$	2.4752
Mean	2.7	—	2.4752
St. dev.	0.3	—	0.0512

Comments:

- We clearly have substantial information provided by the data – our posterior beliefs regarding μ have been shifted down considerably, to be much closer to the mean of the data than our prior mean
- In light of the data, our beliefs regarding μ are *much* more precise, with a six-fold reduction in the standard deviation

This is noticeable when we produce a sketch of prior versus posterior:



18. From example 2.6 in the lecture notes, we know that the posterior precision is given by $d+n\tau$, where n is the number of observations and τ is the precision of our measurements.

(a) Thus, the standard devision of the posterior distribution is $1/sqrtd + n\tau$, and we want

$$\begin{aligned} \frac{1}{\sqrt{d+n\tau}} &= 0.1 \\ \longrightarrow n &= \frac{100-d}{\tau}. \end{aligned}$$

Here, we have $d = 1$ and $\tau = 1/4$, giving $n = 396$.

(b) When $n = 100$ the posterior standard deviation is

$$\frac{1}{\sqrt{d+100\tau}} = \frac{1}{\sqrt{d+25}}.$$

The prior variance is $1/d$, which can never be negative; Thus, d cannot be negative. As $d \to 0$, we have $1/5 = 0.2$ for the posterior standard deviation; as d increases, the posterior standard deviation defined above will always decrease.

20. (a) Benefits of working within the Bayesian framework over a standard frequentist analysis:

- Can inform the analysis with the opinions of experts. This can have huge practical benefits when working with small datasets – including a significant increase in the precision of parameter estimates, leading to substantial reductions in the width of confidence intervals. In the case study, this filtered through to increased precision for *return level estimates* – for example, the rainfall event we might expect to see once per century.
- In a Bayesian analysis, confidence intervals have a more intutitve interpretation. For example, a 95% frequentist confidence interval (ℓ, u) for a parameter μ does not hold the property $\Pr(\ell < \mu < u) = 0.95$. However, a 95% Bayesian confidence interval *does* have this interpretation!
- Although not covered in this case study, a Bayesian analysis has the advantage of being able to handle *prediction* naturally, through the *predictive distribution.*

(b) The main difficulty lay in 'converting' the epxert hydrologist's beliefs into something meaningful about the parameters in the distribution being used. This was especially difficult for the shape parameter in the generalised extreme value distribution, and to some extene the scale parameter; such parameters might not easily be understood by non-statisticians.

(c) In the case study, distribution theory was used to transform elicited prior distributions for quantities the expert might feel more comfortable with (here we used quantiles corresponding to the 10, 50 and 200 year rainfall event for this location), to a distribution for the original parameters in the model. The quantiles used were functions of these parameters.

(d) An *improper* prior distribution is one that does not have total desity under the curve equalling 1; they are not probability densities.

(e) *Markov chain Monte Carlo.*

21. There is a typo in the question: it *should* say "... and in question **42** you have already...".

From question 42, we have that $\theta \sim Ga(11.05, 0.63)$. The likelihood is

$$f(\boldsymbol{x}|\theta) = \frac{e^{-\theta}\theta^{15}}{15!} \times \frac{e^{-\theta}\theta^{6}}{6!} \times \frac{e^{-\theta}\theta^{10}}{10!} \propto e^{-3\theta}\theta^{31}.$$

Thus, the posterior is

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto \theta^{10.05}e^{-0.63\theta} \times e^{-3\theta}\theta^{31} \\ &= \theta^{41.05}e^{-3.63\theta},\end{aligned}$$

that is, $\theta|\boldsymbol{x} \sim Ga(42.05, 3.63)$. A report would be formed in the usual way – that is, comparing prior and posterior beliefs about θ using the following summary table to assist with the discussion:

	Mean	St. Dev.
Prior	17.5	5.28
Posterior	11.6	1.79

The discussion might also be supported by sketches of prior versus posterior.

22. If we use a vague prior, we use a conjugate prior with as large a varance as possible. The conjugate prior for the Poisson model is the gamma distribution (as this gives a gamma posterior); in example 3.10 of the lecture notes, we saw that if $\theta \sim Ga(a, b)$, then $\text{Var}(\theta) \to \infty$ as $a, b \to 0$.

In general, the likelihood is:

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

where $\bar{x}$ is the sample mean. The posterior, assuming the prior $\theta \sim Ga(a, b)$, is

$$\pi(\theta|\boldsymbol{x}) \propto \theta^{a-1}e^{-b\theta} \times e^{-n\theta}\theta^{n\bar{x}} = \theta^{a+n\bar{x}-1}e^{-(n+b)\theta},$$

that is, $\theta|\boldsymbol{x} \sim Ga(a + n\bar{x}, b + n)$. For vague prior knowledge, we let $a, b \to 0$, giving

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

In question 21, this would have resulted in a $Ga(31, 3)$ posterior (as opposed to the $Ga(42.05, 3.63)$ we obtained in question 21). This would have resulted in:

	Mean	St. Dev.
Posterior	10.33	1.05

23. If $\theta \sim Be(22.6, 3.4)$, then

$$Mode(\theta) = \frac{22.6 - 1}{22.6 + 3.4 - 2} = 0.9.$$

giving the required "most likely value". Using R, we have:

```
> pbeta(0.6,22.6, 3.4)
[1] 0.0008882217
```

This implies that $\Pr(\theta < 0.6) = 0.0008882217 \approx 0.001 = 10^{-3}$, as required.

25. (a) Using the trial roulette method in MATCH, we would specify a range for θ of $(0, 200)$. Placing ten chips in the bin for $(80, 100)$, we can then specify the number of chips for all other bins relative to this (i.e. 5 chips for the $(40, 60)$ bin, etc.). There is a typo in the question: just before the start of part (a), the question should read "... very few occasions have there been less than 40 casualties".

Building up the prior in this way results in $\theta \sim Ga(10, 0.11)$.

(b) The prior is conjugate – combining a Poisson likelihood with a gamma prior gives a gamma posterior, and so both prior and posterior are from the same family.

(c) We have:

$$\begin{aligned} \text{Prior mode} &= 82 \text{ casualties} \\ \text{Prior st. dev.} &= 28.75 \text{ casualties} \\ 1\%\text{-ile} \longrightarrow 99\%\text{-ile} &: 38 \longrightarrow 173 \text{ casualties} \end{aligned}$$

26. (a) Using the bisection method in *MATCH*, with lower and upper limits of 0 and 20 respectively, gives $\lambda \sim Ga(1.23, 0.22)$.

(b) Assuming a Poisson observation of 6, we can find the posterior as:

$$\pi(\lambda|x=6) \propto e^{-\lambda}\lambda^6 \times \lambda^{0.23}e^{-0.22\lambda} = e^{-1.22\lambda}\lambda^{6.23},$$

giving $\lambda|\boldsymbol{x} \sim Ga(7.23, 1.22)$.

28. Non-examinable for this course.

31. There is a typo in this question: it should say "... in question **44**". Interesting for comparison to the Bayesian confidence intervals in question 44, but non-examinable for this course... anyway, here is the solution:

Both θ_C and θ_I in question 44 are binomial proportions (the proportion of rates exceeding 0.12), and from MAS2901 you know that for a binomial proportion the 95% confidence interval is given by:

$$\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}.$$

For the Coastal location, we have $n = 10$ and $\hat{\theta}_C = 0.7$, giving $(0.416, 0.984)$. For the Inland location, we have $\hat{\theta}_I = 0$, giving $(0, 0)$.

32. Comparing frequentist with Bayesian:

	Coastal	Inland
Frequentist	$(0.4160, 0.9840)$	$(0, 0)$
Bayesian	$(0.6005, 0.9308)$	$(0, 0.1173)$

Numerical comparisons:

- For the coastal location, the Bayesian confidence interval is considerably narrower than the frequentist interval.
- It was not possible to find a frequentist interval for the inland location since $\hat{\theta}_I = 0$.

Interpretation:

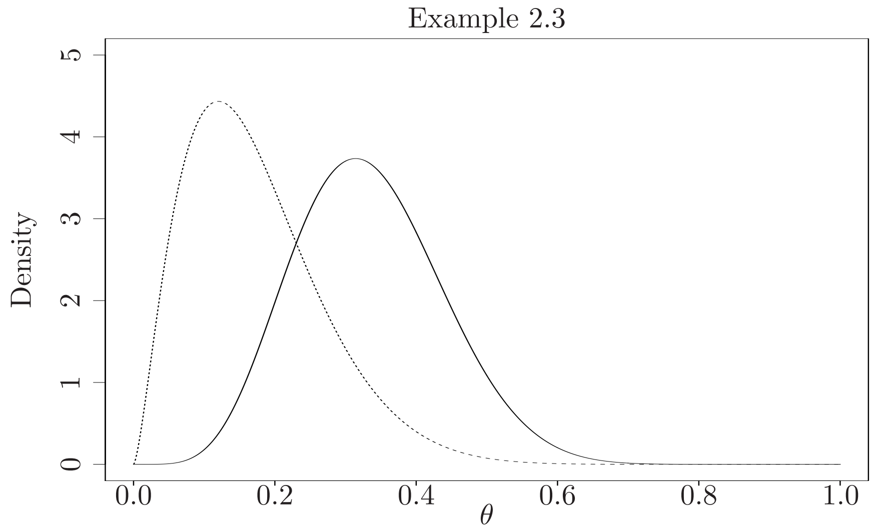
- For the Bayesian confidence intervals, there is a probability of 0.95 that θ_C or θ_I lies between the lower and upper bounds.
- For the frequentist intervals, the probability that θ_C or θ_I lies between the lower and upper bounds is either zero or one, since – in the frequentist paradigm – these parameters are fixed (but unknown) constants.

36. Question is completely instructional.

37. In R:

```
> plot(x,posterior,type="l",xlab="theta",ylab="Density",main="Example 2.3",ylim=c(0,5))
> lines(x,prior,type="l",lty=2)
```

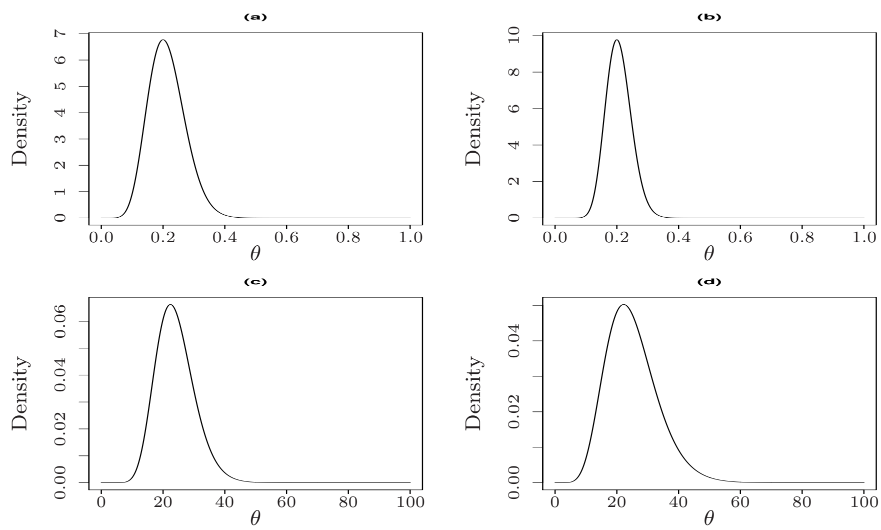
Giving the following plot (which is the same as that in Figure 2.7 in the lecture notes):



38. As for questions 36 and 37.

39. In R:

```
> x1=seq(0,1,0.0001)
> x2=seq(0,100,0.01)
> par(mfrow=c(2,2))
> plot(x1,dbeta(x1,10,37),type="l",xlab="t",ylab="d",main="(a)")
> plot(x1,dbeta(x1,20,77),type="l",xlab="t",ylab="d",main="(b)")
> plot(x2,dgamma(x2,15,0.625),type="l",xlab="t",ylab="d",main="(c)")
> plot(x2,dgamma(x2,9,0.36),type="l",xlab="t",ylab="d",main="(d)")
```



For scenario 1, the most suitable distribution must be one of (a) or (b), since θ is a probability and so is defined on $(0, 1)$. From R:

```
> 1-pbeta(0.5,10,37)<1/100000
[1] FALSE
> 1-pbeta(0.5,20,77)<1/100000
[1] TRUE
```

This suggests that $\theta \sim Beta(20, 77)$ must the most appropriate prior – in other words, distribution (b).

For scenario 2, the most suitable distribution must be one of (c) or (d), since θ is a rate and so is defined over the positive real line. From R:

```
> pgamma(10,15,0.625)
[1] 0.00205863
> pgamma(10,9,0.36)
[1] 0.01167141
```

This suggests that $\theta \sim Ga(9, 0.36)$ must be the most appropriate prior as the distribution function evaluated at 10 is closest to the specified 1/100 probability – in other words, distribution (d).

42. (c) **Feedback: first fit**

$$\begin{aligned}
\theta &\sim Ga(10, 0.54) \\
\pi(\theta) &= \frac{0.54^{10}\theta^{9}e^{-0.54\theta}}{\Gamma(10)}, \theta > 0 \\
E(\theta) &= 18.518 \\
SD(\theta) &= 5.856
\end{aligned}$$

(d) $P(\theta < 7.61) = P(\theta > 34.6) = 0.01$.

(e) "Only once in a hundred years would we expect to see more than around 34 or 35 tropical depressions. Does this seem reasonable?"

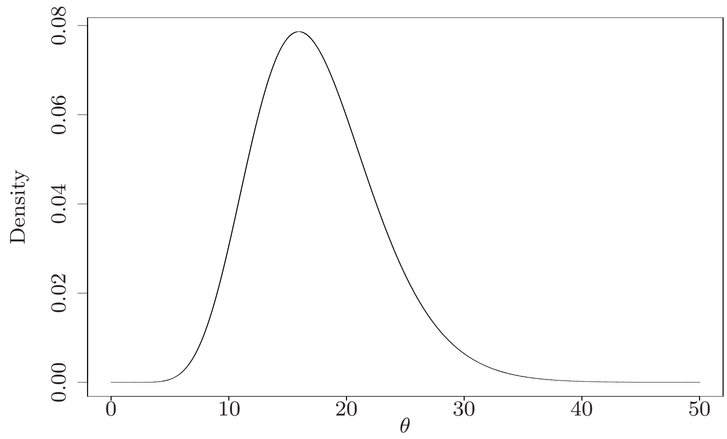
(f) **Feedback after refinement: second fit**

$$\begin{aligned}
\theta &\sim Ga(11.05, 0.63) \\
\pi(\theta) &= \frac{0.63^{11.05}\theta^{10.05}e^{-0.63\theta}}{\Gamma(11.05)}, \theta > 0 \\
E(\theta) &= 17.539 \\
SD(\theta) &= 5.276 \\
P(\theta < 7.61) &= 0.01 \\
P(\theta > 32) &= 0.01
\end{aligned}$$

(g) (i) In R:

```
> x=seq(0,50,0.001)
> plot(x, dgamma(x,11.05,0.63), type="l", xlab="t", ylab="d")
```

Giving the following plot:



(ii) In R:

```
> pgamma(7.61,11.05,0.63)
[1] 0.009866593
> 1-pgamma(32,11.05,0.63)
[1] 0.01028627
```

43. (a) Use $\lambda \sim Be(a, b)$, as λ is a probability.

(b) We have

$$Mode(\lambda) = \frac{a-1}{a+b-2} = 0.7$$

Therefore, we can find that

$$a = \frac{7b-4}{3}.$$

(c) The question *should* say "... like that at the top of page 56...". In R:

```
> f=function(b){
+ answer = pbeta(0.3, ((7*b-4)/3),b)-0.001
+ return(answer)
+ }
```

(d) The question *should* say "... on page 56...". In R:

```
> uniroot(f, lower=1, upper=100)
$root
[1] 5.170688

$f.root
[1] 8.960203e-12

$iter
[1] 12

$estim.prec
[1] 9.430276e-05
```

Therefore, we have $b = 5.17$.

(e) We have $b = 5.17$. Therefore

$$a = \frac{7 \times 5.17 - 4}{3} = 10.73,$$

and so we have $\lambda \sim Be(10.73, 5.17)$.

44. (a) The mean and standard deviation, as found in R, are:

```
> mean(rate.coastal)
[1] 0.1330711
> mean(rate.inland)
[1] 0.05500738
> sd(rate.coastal)
[1] 0.04766444
> sd(rate.inland)
[1] 0.02595326
```

(b) We can estimate θ_C and θ_I as the proportion of gravestones with an annual rate of degradation of at least 0.12mm at the coastal and inland locations, respectively. Looking at the data in R:

```
> rate.coastal
 [1] 0.22025231 0.18373611 0.15552727 0.13451462 0.12195733 0.13273000
 [7] 0.14519588 0.09990640 0.06345654 0.07343473
> rate.inland
 [1] 0.05206567 0.11985773 0.08846387 0.07917771 0.07800750 0.06193125
 [7] 0.04544318 0.04938312 0.04901284 0.03702157 0.03068259 0.03173200
[13] 0.03168775 0.03482604 0.03581794
```

we see that

$$\hat{\theta}_C = \frac{7}{10} = 0.7 \qquad \text{and} \qquad \hat{\theta}_I = \frac{0}{15} = 0.$$

The likelihood functions are:

$$\begin{aligned} f_{\theta_C}(x = 7|\theta_C) &= {}^{10}C_7 \theta_C^7 (1 - \theta_C)^3 \qquad \text{and} \\ f_{\theta_I}(x = 0|\theta_I) &= {}^{15}C_0 \theta_I^0 (1 - \theta_I)^{15} = (1 - \theta_I)^{15}. \end{aligned}$$

(c) (i) The *trial roulette* method might be suitable, since the expert has specified probabilities for different ranges for both θ_C and θ_I.

(ii) Using *MATCH*, we find that

$$\begin{aligned} \theta_C &\sim Beta(10, 2) \qquad \text{and} \\ \theta_I &\sim Beta(1, 9). \end{aligned}$$

(iii) The expert has told us that we can expect a higher proportion of gravestones to have a rate of degradation of at least 0.12mm at Savannah than at Macon, due to the effects of salt in the air at the coastal location. This is reflected in the fitted distributions for θ_C and θ_I; *MATCH* shows $\pi(\theta_C)$ having a mode at around 0.9, whereas $\pi(\theta_I)$ has a mode around 0.02.

(d) For the coastal location, we have

$$\begin{aligned}\pi(\theta_{\rm C}|x=7) &\propto \theta_{\rm C}^{9}(1-\theta_{\rm C}) \times \theta_{\rm C}^{7}(1-\theta_{\rm C})^{3} \\ &\propto \theta_{\rm C}^{16}(1-\theta_{\rm C})^{4},\end{aligned}$$

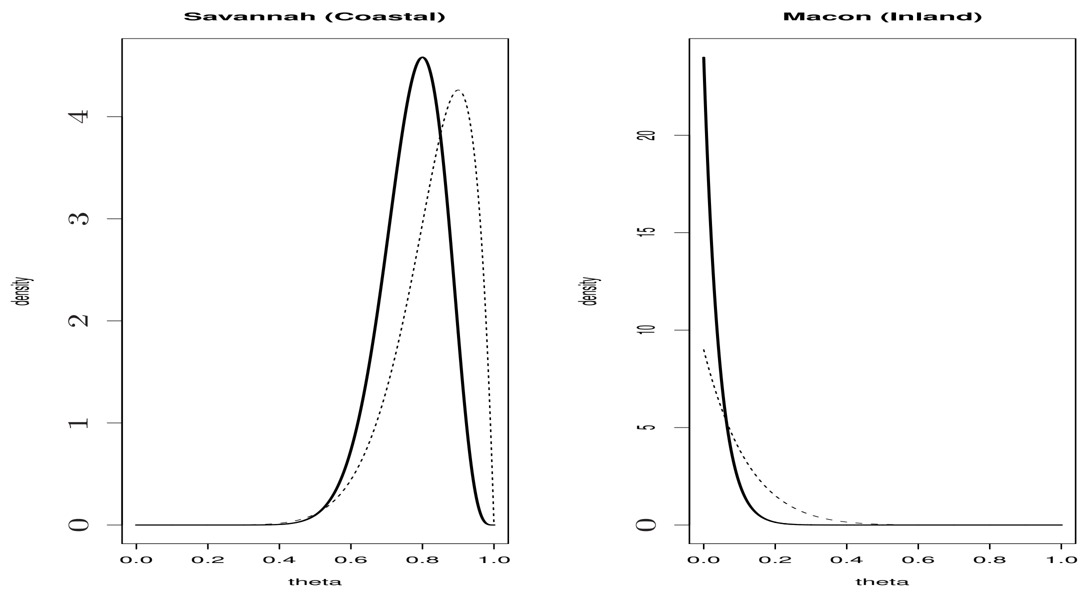
and so $\theta_{\rm C}|x=7 \sim Beta(17,5)$.

For the inland location, we have

$$\begin{aligned}\pi(\theta_{\rm I}|x=0) &\propto \theta_{\rm I}^{0}(1-\theta_{\rm I})^{8} \times \theta_{\rm I}^{0}(1-\theta_{\rm I})^{15} \\ &\propto \theta_{\rm I}^{0}(1-\theta_{\rm I})^{23},\end{aligned}$$

and so $\theta_{\rm I}|x=0 \sim Beta(1,24)$.

(e) In R, we can obtain the following plots of prior (dashed) and posterior (solid) densities:



(i) Comparing prior and posterior beliefs about $\theta_{\rm C}$ it is obvious that we had substantial prior knowledge since there is very little difference between $\pi(\theta_{\rm C})$ and $\pi(\theta_{\rm C}|\boldsymbol{x})$. The posterior distribution is shifted down slightly, relative the the prior, and the standard deviation for $\theta_{\rm C}$ seems to have reduced slightly after observing the data.

Comparing prior and posterior beliefs about $\theta_{\rm I}$ again, we see that we had fairly substantial prior knowledge. Having observed the data, our beliefs about $\theta_{\rm I}$ have been shifted down even closer to zero, and it seems the standard deviation for $\theta_{\rm I}$ has reduced substantially.

(ii) Having observed the data at both locations, the differences in beliefs about the rate of granite degradation at the coast and inland are still quite pronounced; beliefs about $\theta_{\rm C}$ have been modified somewhat, with a slight downwards shift in the mode, but beliefs about $\theta_{\rm I}$ also shifted down slightly, preserving the expert's original ideas about the difference in rate of degrdation between Savannah and Macon.

(f) It is obvious from the plot in part (e) that the 95% HDI for $\theta_{\rm I}$ must include zero, i.e. we have $(0,b)$. In fact, example 4.1 in the lecture notes shows that, for a 95% HDI for a $Beta(1, 24)$ distribution,

$$b = 1 - 0.05^{1/24} = 0.1173.$$

Thus, for θ_{I}, the 95% HDI is $(0, 0.1173)$.

(g) The 95% HDI for θ_{C} is more awkward to find because the density of a $Beta(17, 5)$ does not increase or decrease monotonically as $\theta_{\text{C}} \to 0$ or $\theta_{\text{C}} \to 1$.

In R, we can find the 95% HDI for θ_{C} using the following code:

```
> g=function(x)
+ {
+ a=x[1]
+ b=x[2]
+ (pbeta(b,17,5)-pbeta(a,17,5)-0.95)^2+0.0001*(dbeta(b,17,5)-dbeta(a,17,5))^2
+ }
> initiala=qbeta(0.025,17,5)
> initialb=qbeta(0.975,17,5)
> res=optim(c(initiala,initialb),g,method="L-BFGS-B",lower=0,upper=1)
> a=res$par[1]
> b=res$par[2]
> a
[1] 0.6005299
> b
[1] 0.9308117
```

giving the 95% HDI for θ_{C} as $(0.6005, 0.9308)$.

(h) In light of the data, it appears there *is* a significant difference in the rate of degradation of granite at the two locations – the HDI for θ_{I} and θ_{C} are completely separate and do not overlap.