SOLUTIONS TO TAKE HOME EXAM FOR BAYESIAN INFERENCE IN THEORY AND PRACTISE 2007-03-27

MATTIAS VILLANI

Question 1a. *"Assume that you want to investigate the proportion (θ) of defective items manufactured at a production line. Your collegue takes a random sample of 30 items. Three were defective in the sample. Assume a uniform prior for θ. Compute the posterior of θ."*

Solution 1a. Data: $x_1, ..., x_{30} | \theta \overset{iid}{\sim} Bern(\theta)$. Likelihood $p(x_1, ..., x_n | \theta) = \theta^s (1 - \theta)^f$, where $s = 3$ and $f = 27$.

Uniform prior: $\theta \sim Beta(\alpha, \beta)$, with $\alpha = \beta = 1$. Prior density:

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

Posterior density:

$$\begin{aligned} p(\theta | x_1, ..., x_n) &\propto p(x_1, ..., x_n | \theta) p(\theta) \\ &\propto \theta^s (1 - \theta)^f \theta^{\alpha - 1} (1 - \theta)^{\beta - 1} \\ &= \theta^{\alpha + s - 1} (1 - \theta)^{\beta + f - 1}, \end{aligned}$$

which we recongize as being proportional to the Beta density with parameters $\alpha + s$ and $\beta + f$. That is: the posterior is given by the $Beta(\alpha + s, \beta + f)$ density. In this specific example (where $\alpha = \beta = 1$, $s = 3$, $f = 27$), the posterior is therefore the $Beta(4, 28)$ density. Figure 1 and table below were produced with my Excel spreadsheet *bernoullimodellen.xls*.

Prior (Beta)
a   1
b   1

Posterior (Beta)
a+y   4
b+n-y   28

Data
Number of trials (n)   30
Number of successes (y)   3

Prior and posterior summaries
Prior mean   0,5
Prior standard dev.   0,288675135
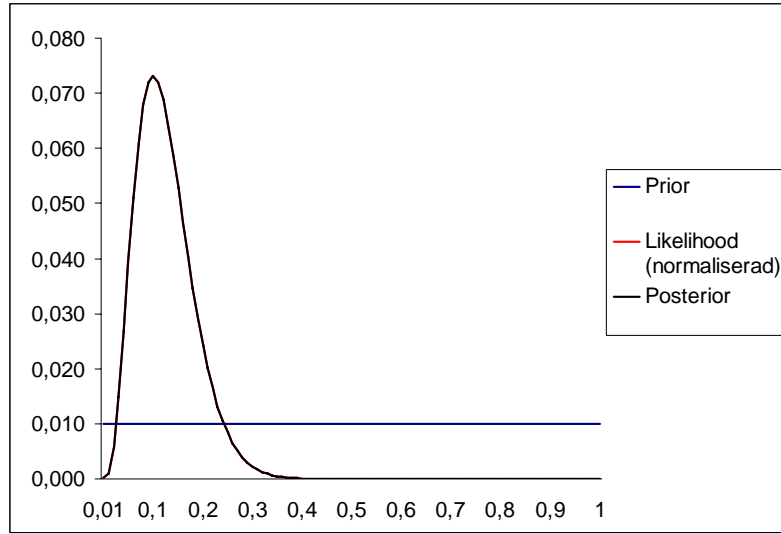ML estimate   0,1
Posterior mean   0,125

Figure 1. Figure 1. Prior and posterior for exercise 1a. Prior, likelihood and posterior for the Beta(1,1) prior.

Posterior standard dev.   0,057570773

Question 1b. *"Your collegue now tells you that he did not decide on the sample size before the sampling was performed. His sampling plan was to keep on sampling items until he had found three defective ones. It just happend that the 30th item was the third one to be defective. Redo the posterior calculculation, this time under the new sampling scheme. Discuss the results."*

Solution 1b. The likelihood function is now given by the negative binomial distribution:

$$p(n|\theta) = \binom{n-1}{s-1}\theta^s(1-\theta)^f,$$

where I note that it is now n that is stochastic (in 1a it was s that was stochastic). But when we do a Bayesian analysis we are looking at the posterior $p(\theta|Data)$, where $Data$ is fixed. It does not matter if s or n was stochastic before we collected the data, what matters is the form of the likelihood as a function of θ. Evidently, the likelihood function, as a function of θ, is the same in 1a and 1b. Since the prior is assumed to be the same, so are the posteriors in 1a and 1b. The posterior of θ is therefore again the $Beta(4,28)$ density. Bayesian inference therefore respects the *likelihood principle*: it is only the shape of the likelihood that should matter for inference.

Question: 1c. *"Repeat 1a and 1b, but this time using Jeffreys' prior. Discuss the results."*

Solution 1c. Let us first repeat 1a. The derivation of Jeffreys prior for the bernoulli setup with a fixed number of trials (n) is given in my notes for Lecture 2. The result is

$$p(\theta) = |J(\theta)|^{1/2} \propto \theta^{-1/2}(1-\theta)^{-1/2} \propto Beta(\theta|1/2, 1/2).$$

This is proper density. The posterior using this prior is therefore the $Beta(3.5, 27.5)$ density. It is depicted in Figure 2.
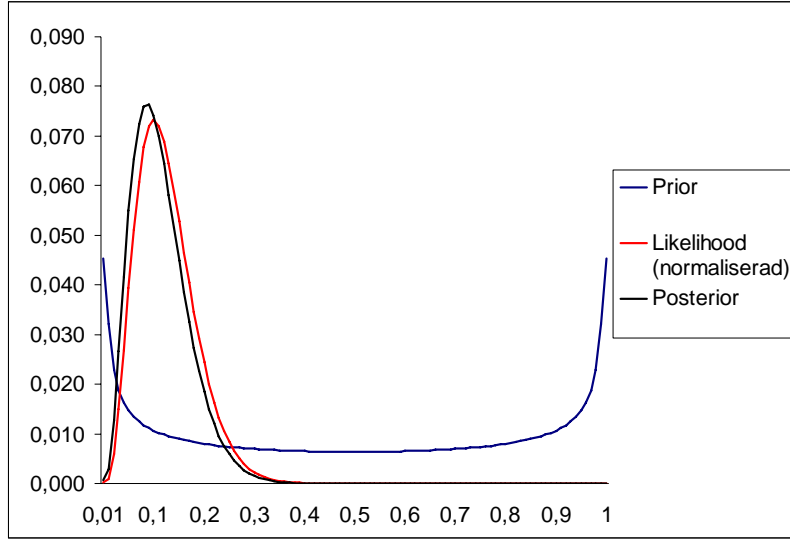
FIGURE 2. Prior, likelihood and posterior for the Jeffreys prior with Binomial data.

For 1b, where the number of trials are stochastic we have (remember the likelihood is now the negative binomial where n is stochastic)

$$\ln p(n|\theta) = s \ln \theta + (n-s) \ln(1-\theta)$$

$$\frac{d \ln p(n|\theta)}{d\theta} = \frac{s}{\theta} - \frac{(n-s)}{(1-\theta)}$$

$$\frac{d^2 \ln p(n|\theta)}{d\theta^2} = -\frac{s}{\theta^2} - \frac{(n-s)}{(1-\theta)^2}$$

$$J(\theta) = - - E_{n|\theta}\left[\frac{d^2 \ln p(n|\theta)}{d\theta^2}\right] = \frac{E_{n|\theta}(s)}{\theta^2} + \frac{E_{n|\theta}(n-s)}{(1-\theta)^2}.$$

We can use that $E_{n|\theta}(s) = s$ (s is a constant, it is n that is stochastic here), and that $E_{n|\theta}(n-s) = E_{n|\theta}(n) - s = \frac{s}{\theta} - s$ from the negative binomial distribution (see for example Hogg and Tanis) to get

$$
\begin{aligned}
J(\theta) &= \frac{s}{\theta^2} + \frac{\frac{s}{\theta} - s}{(1-\theta)^2} = s\left(\frac{1}{\theta^2} + \frac{\frac{1-\theta}{\theta}}{(1-\theta)^2}\right) \\
&= s\left(\frac{1}{\theta^2} + \frac{1}{(1-\theta)\theta}\right) = s\left(\frac{1}{\theta^2(1-\theta)}\right)
\end{aligned}
$$

Thus, the Jeffreys' prior is

$$p(\theta) = |J(\theta)|^{1/2} \propto \theta^{-1}(1-\theta)^{-1/2}.$$

This density is the limit of a $Beta(\epsilon, 1/2)$ density as $\epsilon \to 0$. This is an improper prior. Multiplying this prior with likelihood gives the posterior

$$\binom{n-1}{s-1}\theta^s(1-\theta)^f\theta^{-1}(1-\theta)^{-1/2} \propto \theta^{s-1}(1-\theta)^{f-1/2},$$

which we recognize as proportional to the $Beta(s, f+1/2)$ density. This is a proper posterior density when $s \geq 1$. In our specific example we have the posterior $Beta(3, 27.5)$, which is
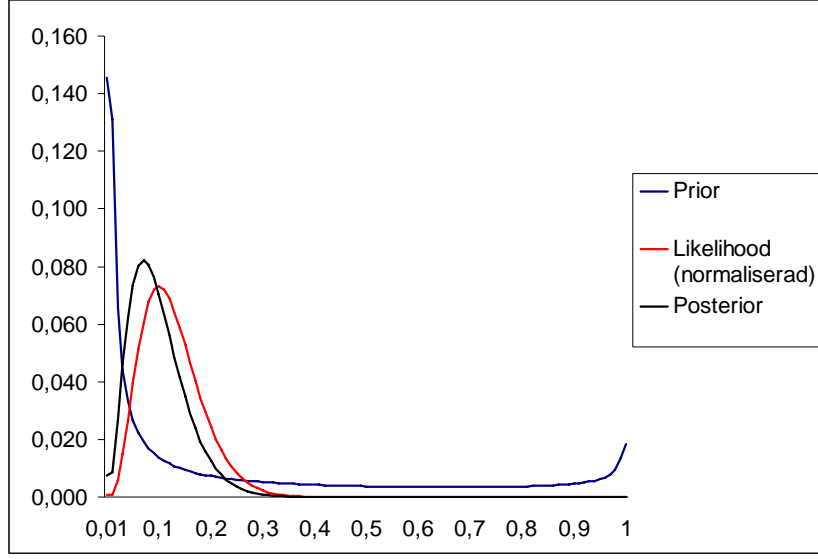
FIGURE 3. Prior, likelihood and posterior for the Jeffreys prior with negative binomial data.

proper. The results are given in Figure 3. The posteriors for the three priors (uniform, Jeffreys for binomial and Jeffreys for negative binomial) are very similar.

Jeffreys prior has the strange property that it depends on the data (the way the sample was obtained). This is awkward, since θ is the same quantity in both 1a and 1b. My beliefs about θ a priori should not depend on how data was collected to analyze θ.

Question 1d. *"Do a Bayesian test of the hypothesis: $H_0 : \theta = 0.05$. Assume a uniform prior, and two other priors (to learn about the sensitivity with respect to the prior)."*

Solution 1d. A Bayesian test of $H_0 : \theta = 0.05$ can be performed using the Bayes factor comparing the two models:

- $M_0 : x_1, ..., x_{30}|\theta \overset{iid}{\sim} Bern(0.05)$
- $M_1 : x_1, ..., x_{30}|\theta \overset{iid}{\sim} Bern(\theta)$, where $\theta \sim Beta(\alpha, \beta)$ a priori.

The Bayes factor is the ratio of the marginal likelihood of models M_1 and M_2:

$$
\begin{aligned}
BF_{01} &= \frac{p(x_1, ..., x_n|M_0)}{p(x_1, ..., x_n|M_1)} = \frac{p(x_1, ..., x_n|\theta = 0.05)}{\int p(x_1, ..., x_n|\theta)p(\theta)d\theta} \\
&= \frac{0.05^s(1 - 0.05)^f}{\int \theta^s(1 - \theta)^f \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1 - \theta)^{\beta-1}d\theta}.
\end{aligned}
$$

Now we need to evaluate the integral in the denominator

$$
\frac{1}{B(\alpha, \beta)} \int \theta^{s+\alpha-1}(1 - \theta)^{f+\beta-1}d\theta.
$$

How do we do this? Well, $\theta^{s+\alpha-1}(1-\theta)^{f+\beta-1}$ is essentially the $Beta(s+\alpha, f+\beta)$ density. We therefore know that

$$\int \frac{1}{B(s+\alpha, f+\beta)}\theta^{s+\alpha-1}(1-\theta)^{f+\beta-1}d\theta = 1$$

and therefore that

$$\int \theta^{s+\alpha-1}(1-\theta)^{f+\beta-1}d\theta = B(s+\alpha, f+\beta).$$

The Bayes factor can therefore be written

$$BF_{01} = \frac{0.05^s(1-0.05)^f}{\frac{B(s+\alpha, f+\beta)}{B(\alpha, \beta)}}.$$

Note that this result is derived in the notes for Lecture 9. Your have the result there. Using the uniform prior ($\alpha = \beta = 1$) we get the result $BF_{01} = 3.94$. This is on the borderline between 'not worth more than a bare mention' and 'substantial' evidence in favor of H_0 according to Kass and Raftery's Section 3.2 in their JASA paper. If we change the prior to the $Beta(1, 99)$ prior we the following summary from bernoullimodellen.xls:

Prior (Beta)
a   1
b   99

Posterior (Beta)
a+y   4
b+n-y   126

Data
Number of trials (n)   30
Number of successes (y)   3

Prior and posterior summaries
Prior mean   0,01
Prior standard dev.   0,009900495
ML estimate   0,1
Posterior mean   0,030769231
Posterior standard dev.   0,01508816

The Bayes factor is now $BF_{01} = 13.92$, even stronger in favor of H_0. The $Beta(1, 99)$ prior was too tightly centered around very small θ, see the summary above. The null hypothesis is then preferred since the data (likelihood) is centered over larger θ. For example, the ML estimate is $\hat{\theta} = 0.1$.

Finally lets use the prior $Beta(10, 100)$, which is more located in the region where the likelihood is substantial. We would expect the BF to favor the alternative hypothesis now, and indeed, $BF_{01} = 0.6211$, which favors M_1. Note that in order to use the scale in Kass and Raftery that we used earlier, one has to invert this number , i.e. $1/0.6211 = 1.61$, so the evidence is rather weak. Overall, it is hard to find a prior which gives strong evidence in favor of the alternative hypothesis. This suggest that data support the null hypothesis strongly.
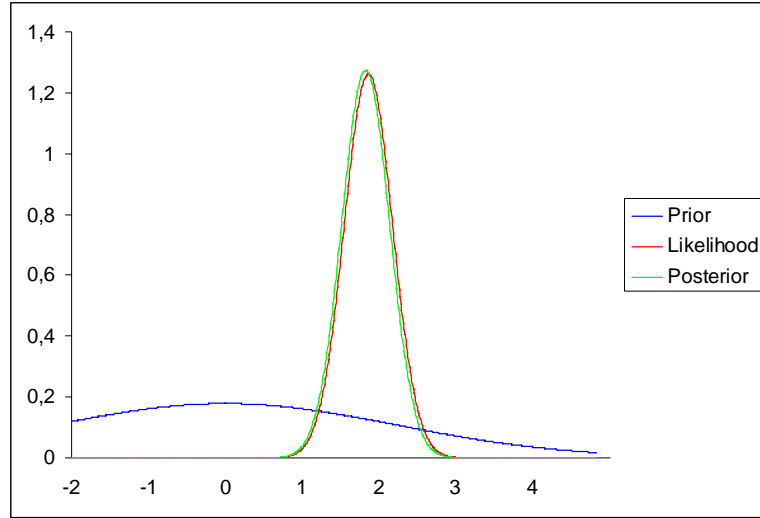
FIGURE 4. Prior-to-Posterior. Normal model. Exercise 2a.

Question 2a. *"Let $x_1, ..., x_{10} \overset{iid}{\sim} N(\theta, 1)$. Let the sample mean be $\bar{x} = 1.873$. Assume that $\theta \sim N(0, 5)$ apriori. Compute the posterior distribution of θ."*

Solution 2a. Remember the prior-to-posterior mapping for this model (Lecture 1):

$$
\begin{aligned}
\theta &\sim N(\mu_0, \tau_0^2) \overset{x_1,...,x_n}{\Longrightarrow} \theta | x \sim N(\mu_n, \tau_n^2) \\
\frac{1}{\tau_n^2} &= \frac{n}{\sigma^2} + \frac{1}{\tau_0^2} \\
\mu_n &= w\bar{x} + (1-w)\mu_0 \\
w &= \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}.
\end{aligned}
$$

Here we have $n = 10$, $\sigma^2 = 1$, $\mu_0 = 0$, $\tau_0^2 = 5$. This gives us

$$
\begin{aligned}
w &= \frac{\frac{10}{1}}{\frac{10}{1} + \frac{1}{5}} = \frac{50}{51} = 0.98039 \\
\mu_n &= \frac{50}{51} \cdot 1.873 + \frac{1}{51} \cdot 0 = 1.8363 \\
\tau_n^2 &= \left(\frac{10}{1} + \frac{1}{5}\right)^{-1} = \frac{5}{51}.
\end{aligned}
$$

From normalmodellen.xls you get the graph in Figure 4.

Question 2b. *"Assume now that you have a second sample $y_1, ..., y_{10} \overset{iid}{\sim} N(\theta, 2)$, where θ is the same quantity as in 2a. The sample mean in this second sample is $\bar{y} = 0.582$. Compute the posterior distribution of θ using both samples (the x's and the y's) under the assumption that the two samples are indepedent."*

Answer 2b. The easiest way to do this is use the posterior from the first sample as a prior for the second sample. That is, for this second sample we use the prior $\theta \sim N(1.836\,3, \frac{5}{51})$.
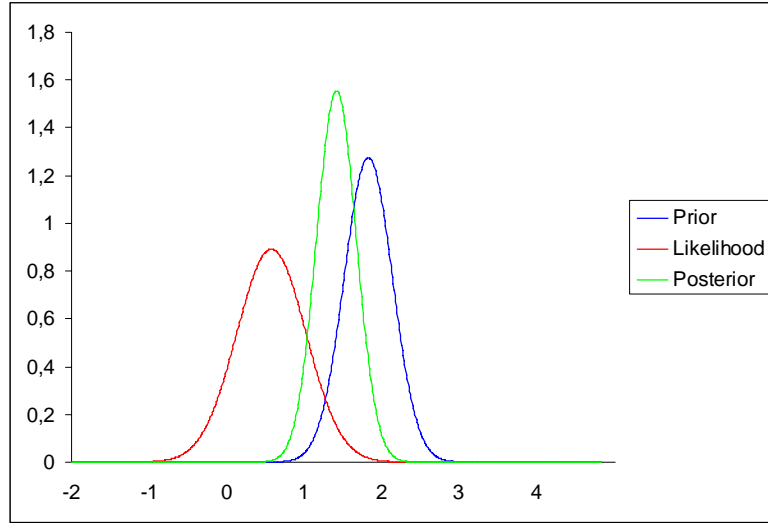
FIGURE 5. Prior-to-Posterior. Normal model. Exercise 2b.

We then get the posterior

$$w = \frac{\frac{10}{2}}{\frac{10}{2} + \frac{1}{\frac{5}{51}}} = \frac{25}{76}$$

$$\mu_n = \frac{25}{76} \cdot 0.582 + (1 - \frac{25}{76}) \cdot 1.8363 = 1.4237$$

$$\tau_n^2 = \left(\frac{10}{2} + \frac{1}{\frac{5}{51}}\right)^{-1} = \frac{5}{76}.$$

Figure 5. Shows the results.

Exercise 2c. *"You have now managed to obtain a third sample $z_1, ..., z_{10} \overset{iid}{\sim} N(\theta, 3)$, with mean $\bar{z} = 1.221$. Unfortunatly, the measuring device for this latter sample was defective: any measurement above 3 was recorded as 3. There were 2 such measurements. Compute the posterior distribution based on all three samples (x, y and z)."*

Solution 2c. This is more complicated. Let us do as before, using the posterior from the first two samples (obtained in 2b) as the prior. The prior is therefore $\theta \sim N(1.4237, \frac{5}{76})$. Now, the mean of the eight measurements which were correctly recorded is $(1.221 \cdot 10 - 3 \cdot 2)/8 = 0.77625$. Let's first use these 8 observations to update the posterior. We then add the information from the two measurements that where truncated at 3. The 8 correctly recorded observations gives the following updating of the $N(1.4237, \frac{5}{76})$ prior:

$$w = \frac{\frac{8}{3}}{\frac{8}{3} + \frac{1}{\frac{5}{76}}} = \frac{10}{67}$$

$$\mu_n = \frac{10}{67} \cdot 0.77625 + (1 - \frac{10}{67}) \cdot 1.4237 = 1.3271$$

$$\tau_n^2 = \left(\frac{8}{3} + \frac{1}{\frac{5}{76}}\right)^{-1} = \frac{15}{268} = 0.05597.$$
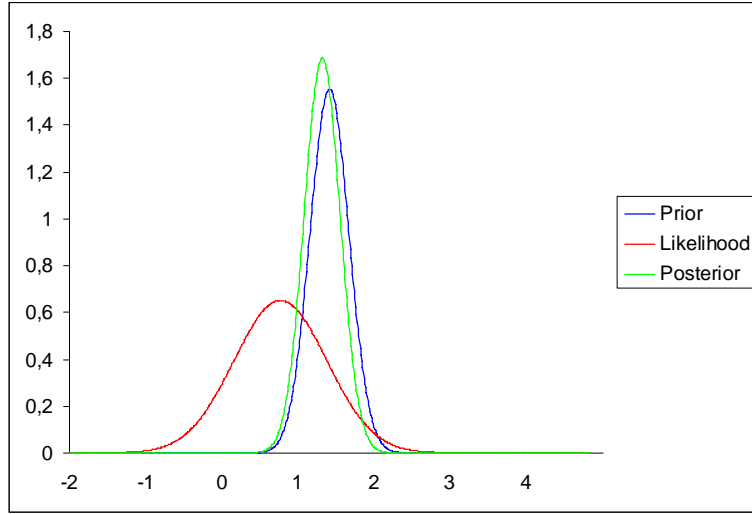
FIGURE 6. Prior-to-Posterior. Normal model. Exercise 2b, but using only the eight correctly recorded observations (and the 20 observations from 2a and 2b).

Note that most of the weight is now given to the prior (which is the posterior from 2a and 2b combined). This is reasonable since the 8 new observations have relatively large variance ($\sigma^2 = 3$). Figure 6 displays the prior-to-posterior mapping.

OK, now we are ready for the final piece of information: the two truncated observations. We do not know their exact values, but we do know that they were at least equal to or larger than 3. This is important information which we cannot ignore. The likelihood of these two observations (let's call them z_1 and z_2) is

$$
\begin{aligned}
p(z_1, z_2|\theta) &= \Pr(z_1 \geq 3)\Pr(z_2 \geq 3) = \left[1 - \Phi(\frac{3-\theta}{\sqrt{3}})\right]\left[1 - \Phi(\frac{3-\theta}{\sqrt{3}})\right] \\
&= \left[1 - \Phi(\frac{3-\theta}{\sqrt{3}})\right]^2.
\end{aligned}
$$

This likelihood function is graphed in Figure 7.

The posterior density from all the data combines this likelihood with the posterior from the 28 correctly recorded observations (10+10+8) that we obtained earlier. This 28-observation posterior is now playing the role of the prior. The full posterior is therefore proportional to

$$
\exp\left[-\frac{1}{2 \cdot 0.055\,97}(\theta - 1.3271)^2\right]\left[1 - \Phi(\frac{3-\theta}{\sqrt{3}})\right]^2.
$$

I have evaluated this expression in Excel. The result is given in Figure 8. Note how the two truncated observations has some, but not substantial, effect on the prior. The prior is pulled up towards larger θ-values. Compare Figure 8 with Figure 4, to see how much we have learned about θ from these 30 observations. Quite a bit.
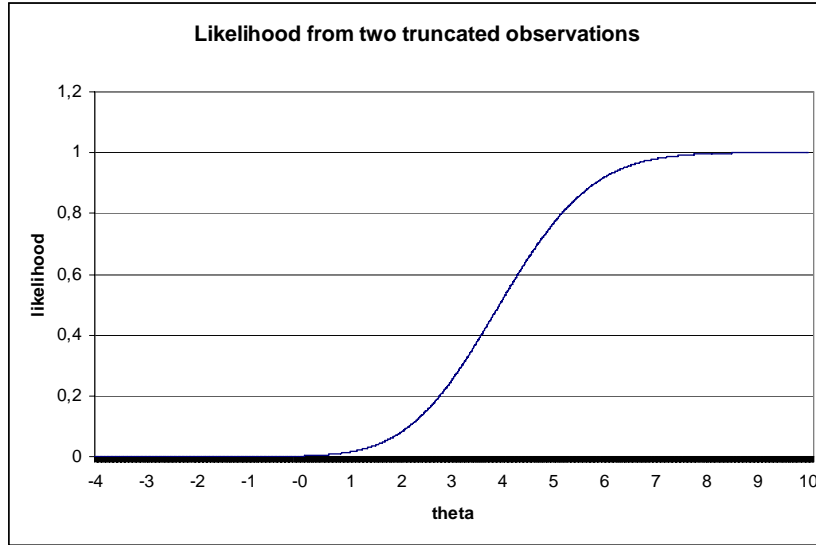
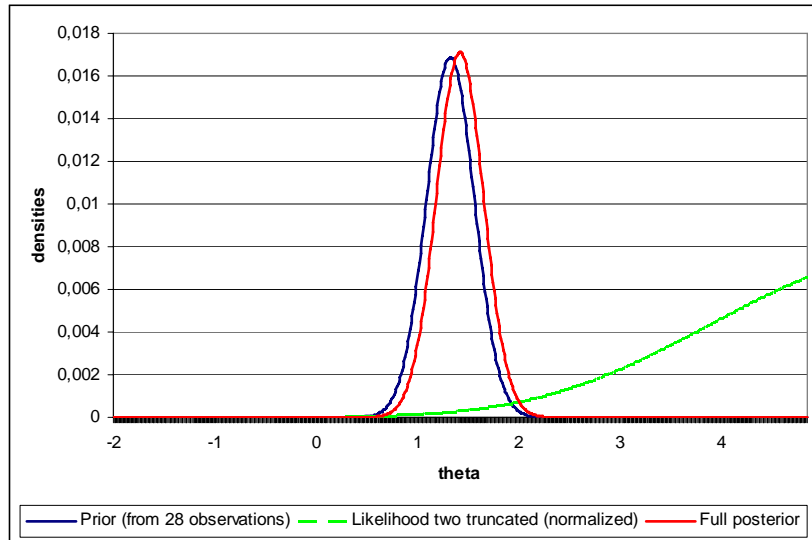FIGURE 7. Likelihood from the two truncated observations.

FIGURE 8. Prior-to-posterior mapping from adding the two truncated observations.

Question 3. *"You want to estimate the annual cost of your mobile phone bill. In the month of january 2007, you made 23 calls, with average duration of 114 seconds per call. The call rate is 5 SEK per minute (hence, every second costs 5/60 SEK, there are no initial costs. Same cost day and night). Assume an exponential distribution for the duration of the calls and that every month has 30 days. Use Bayesian methods to compute the probability that the annual call cost for the year 2007 exceeds 3000 SEK. Make whatever assumptions you find necessary/plausible."*

Solution 3. Let c_i denote the call cost in month i of 2007. Let n_i be the number of calls in this month and $\bar{d}_i$ the average duration of calls in the same month. The cost for month i

then be written
$$c_i = n_i \bar{d}_i.$$
The annual call expenditures, c^*, can then be written
$$c^* = 23 \cdot 114 \cdot (5/60) + \sum_{i=2}^{12} n_i \bar{d}_i \cdot (5/60),$$
where we note that the cost for January is known. Our aim is to obtain the predictive distribution of c^*. There are three basic steps to go through to get there:

  (1) We need to find the distribution of the $\bar{d}_i$ ($i = 2, ..., 12$). Part of the solution is to obtain the posterior distribution of the exponential parameter using the data for January.
  (2) We need to find the distribution of the n_i ($i = 2, ..., 12$). We will assume that the n_i came from a Poisson distribution, and part of the solution is to obtain the posterior distribution of the Poisson parameter.
  (3) We then use the distributions of the n_i and d_i to obtain the predictive distribution of c^*. Here we need to use simulation, but of the simplest kind.

First we assume that the sequences $\bar{d}_i$ and n_i are iid. We can therefore concentrate on $\bar{d}$ and n in a single month. We also assume that $\bar{d}$ and n are independent.

Step 1. Analyzing the durations.

Let d_j denote the duration of the jth call in a given month. By assumption, we know that $d_j \sim Exponential(\theta)$, and we shall also assume that the durations are independent. Assume for the moment that n is known. $\theta = 1/E(d_j)$ is the duration intensity (see page 55 in Gelman et al.), not the mean duration. Let us first derive the posterior distribution of θ using the 23 calls in January. The likelihood is obtained from the exponential density (using the formula on page 55 in Gelman et al.):
$$(0.1) \qquad p(\bar{d}|\theta) = \theta^n \exp(-n \cdot \bar{d} \cdot \theta) = \theta^{23} \exp(-23 \cdot 114 \cdot \theta).$$
Looking at this expression it is easy to see that a Gamma(α, β) density
$$p(\theta) \propto \theta^{\alpha-1} \exp(-\beta\theta).$$
is a conjugate prior for θ. It is also clear that this prior has the interpretation of carrying the information $\alpha - 1$ observations with average duration $\beta/(\alpha - 1)$ (Gelman et al. page 55). I had rather little prior information about this duration. I judge that my information is roughly equivalent to the information in a sample with 2 calls with average duration 60 seconds. Thus, $\alpha = 3$, and $\beta/(\alpha - 1) = 60$, which gives $\beta = 2 \cdot 60 = 120$. This prior is graphed in Figure 9 (remember $\theta = 1/E(d_j)$).

Now, multiplying the prior and the likelihood gives in general the posterior
$$\begin{aligned} p(\theta|\bar{d}) &\propto p(d_1|\theta)p(\theta) \propto \theta^n \exp(-n\bar{d}\theta)\theta^{\alpha-1}\exp(-\beta\theta) \\ &= \theta^{n+\alpha-1}\exp[-(n\bar{d}+\beta)\theta]. \end{aligned}$$
This we recognize as the $Gamma(n + \alpha, n\bar{d} + \beta)$ density. In our particular example we have
$$\theta|\bar{d} \sim Gamma(23 + 3, 23 \cdot 114 + 120) = Gamma(26, 2742).$$
Figure 10 plots the prior and posterior of the call duration intensity (θ).

Step 2. Analyzing the number of calls.

We not turn to the inference for the number of calls per month. We will assume the number of calls to be independent between month and that $n \sim$ Poisson(λ). We have a single observation (the number of calls in January). We know from Lecture 2 that the conjugate
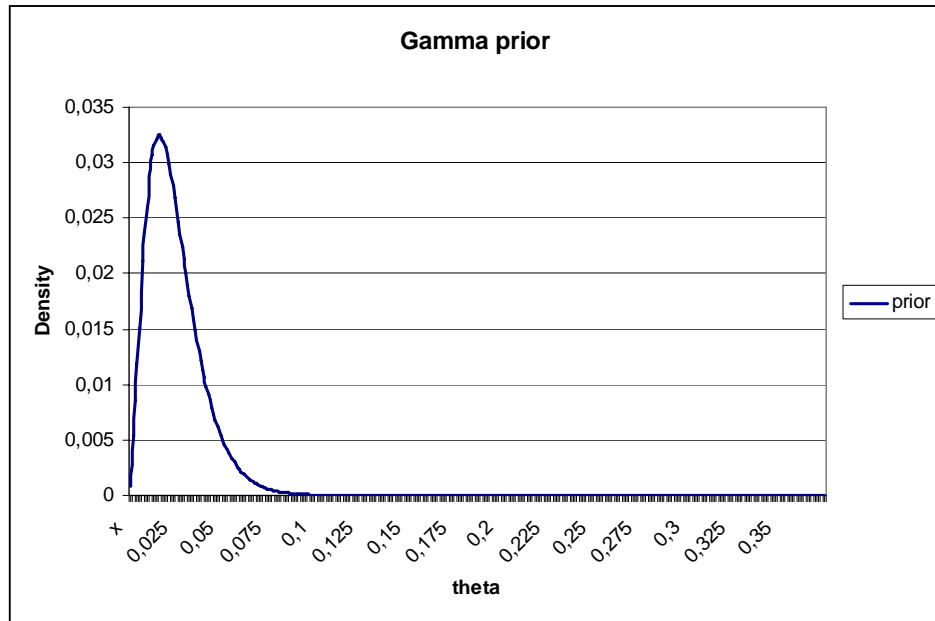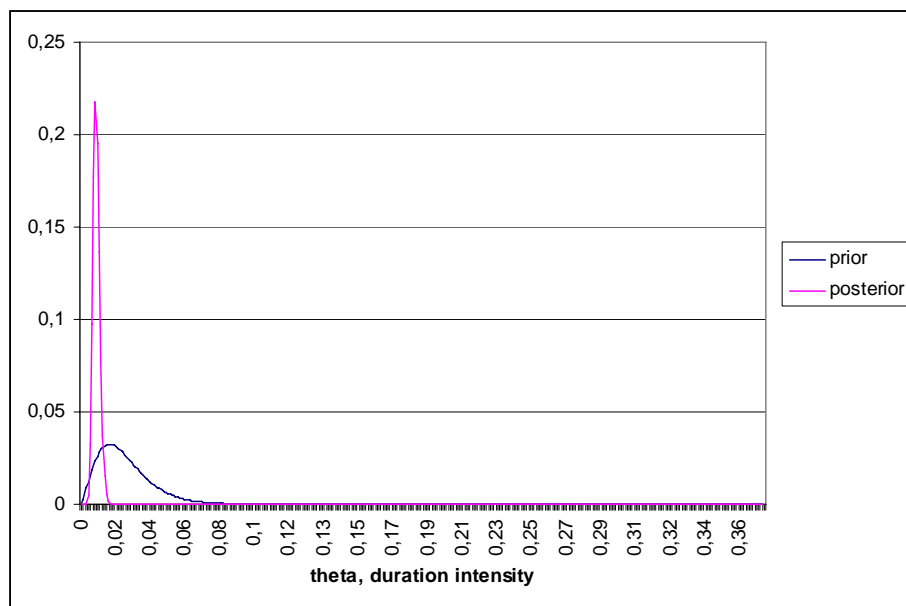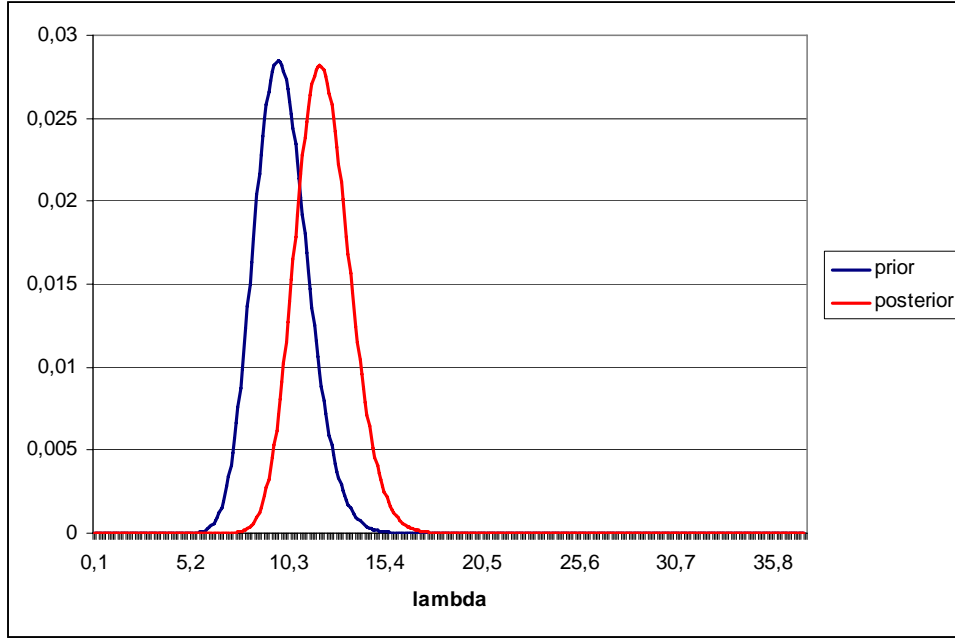
FIGURE 9

FIGURE 10. Prior and posterior for the call duration rate, θ.

prior for this model is the Gamma(a, b) density. This prior carries the same information as a data set with 'a counts in b observations'. I will use $a = 50$ and $b = 5$. This prior is plotted in Figure 11.

The posterior for λ is $Gamma(50 + 23, 5 + 1) = Gamma(73, 6)$ (see Lecture 2), which is also plotted in Figure 11.

FIGURE 11. Prior and posterior for the mean number of calls, λ.

Step 3. Putting it all together - the predictive density of c^*

OK, our aim was to get the distribution of the annual cost:

$$c^* = 23 \cdot 114 \cdot (5/60) + \sum_{i=2}^{12} n_i \bar{d}_i \cdot (5/60).$$

where we note that the cost for January is known. We can only obtain the of distribution c^* by simulation. We can simulate from the following sequence of distributions

$$
\begin{aligned}
\lambda | n_1 &\sim Gamma(73, 6) \\
n_i &\stackrel{iid}{\sim} Poisson(\lambda), \text{ for } i = 2, ..., 12 \\
\theta | \bar{d}_1 &\sim Gamma(26, 2742) \\
\bar{d}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} z_j, \text{ where } z_j \stackrel{iid}{\sim} \text{Exponential}(\theta).
\end{aligned}
$$

For each draw of n_i and $\bar{d}_i$ (for $i = 2, ..., 12$) we compute $c^* = 23 * 114 * (5/60) + \sum_{i=2}^{12} n_i \bar{d}_i$. The collection of c^* draws is a random sample from the predictive density of c^*. Note that if $n_i = 0$, then clearly $\bar{d}_i = 0$. Figure 12 gives the predictive histogram for c^*, obtained for 10000 draws. It is clear that $\Pr(c^* \geq 3000 | n_1, \bar{d}_1)$ is small. We can compute this probability by simply counting the number of draws where $c^* \geq 3000$, and then divide by the total number of draws. This gives $\Pr(c^* \geq 3000 | n_1, \bar{d}_1) = 0.00069$.

The simulations can easily be done with R, Matlab, Minitab, SPSS, or even Excel if you get the appropriate add-in for simulating from the gamma and poisson distributions. I did it in Matlab.
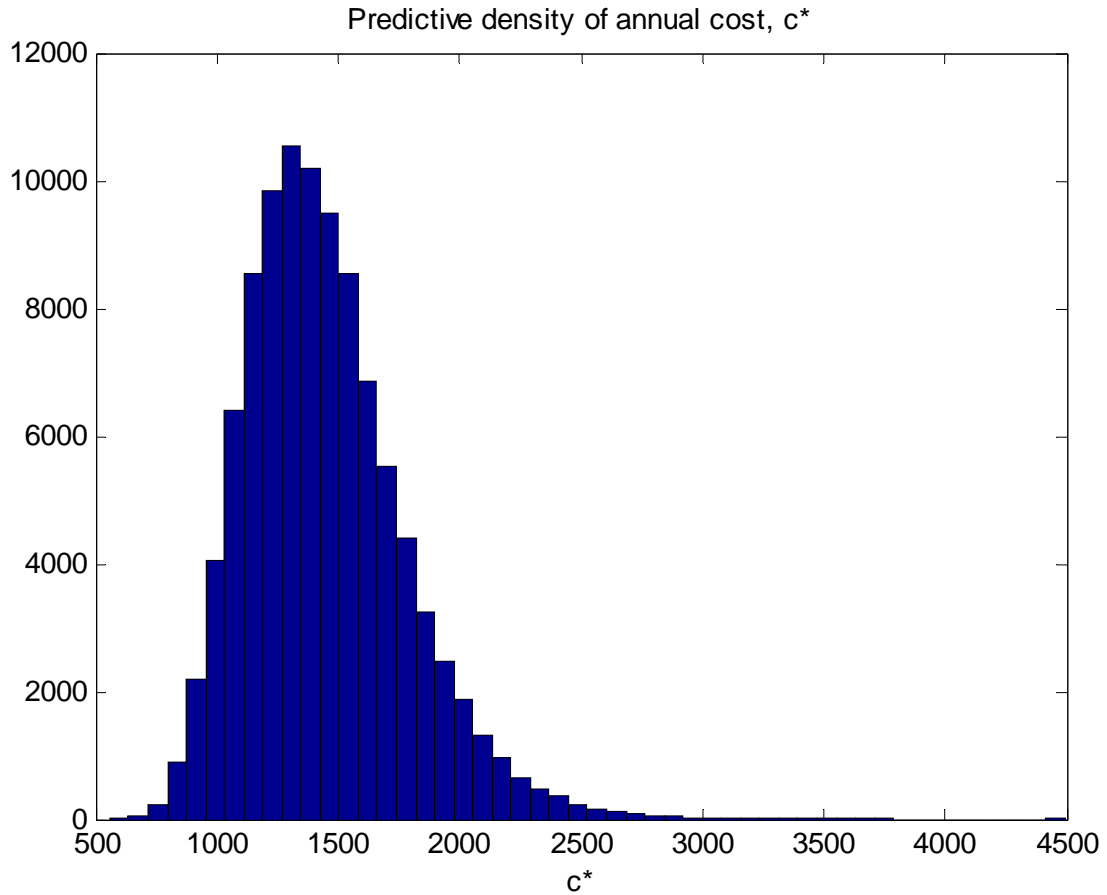
FIGURE 12

Question 4a. *"Assume that the data on the next page (can also be downloaded from the course web site) follows a quadratic regression with an intercept*

$$y_i = \beta_0 + x_i\beta_1 + x_i^2\beta_2 + \varepsilon_i,$$

where ε_i is assumed to be iid normal noise with variance σ^2. Use simulation methods to generate a sample from the joint posterior of $\beta_0, \beta_1, \beta_2$ and σ^2. Use a relatively vague conjugate prior (simple is good, do not complicate). Compute the predictive distribution for $x = 1.1$. [Hint: use BACC for the simulation]."

Solution 4a. I have used a modified version of my LinReg.r script for BACC. The modified script (Exam1Question4a.r) for obtaining posterior draws and graphs can be found on the course web page. Much of the things I do with simulations can actually be done analytically (no need for matrix algebra if you use simulations). The graphs are given below. Figure 13 displays the marginal posteriors and Figure 14, the draws across iterations. Convergence looks good.

The prediction for $x = 1.1$ is given by

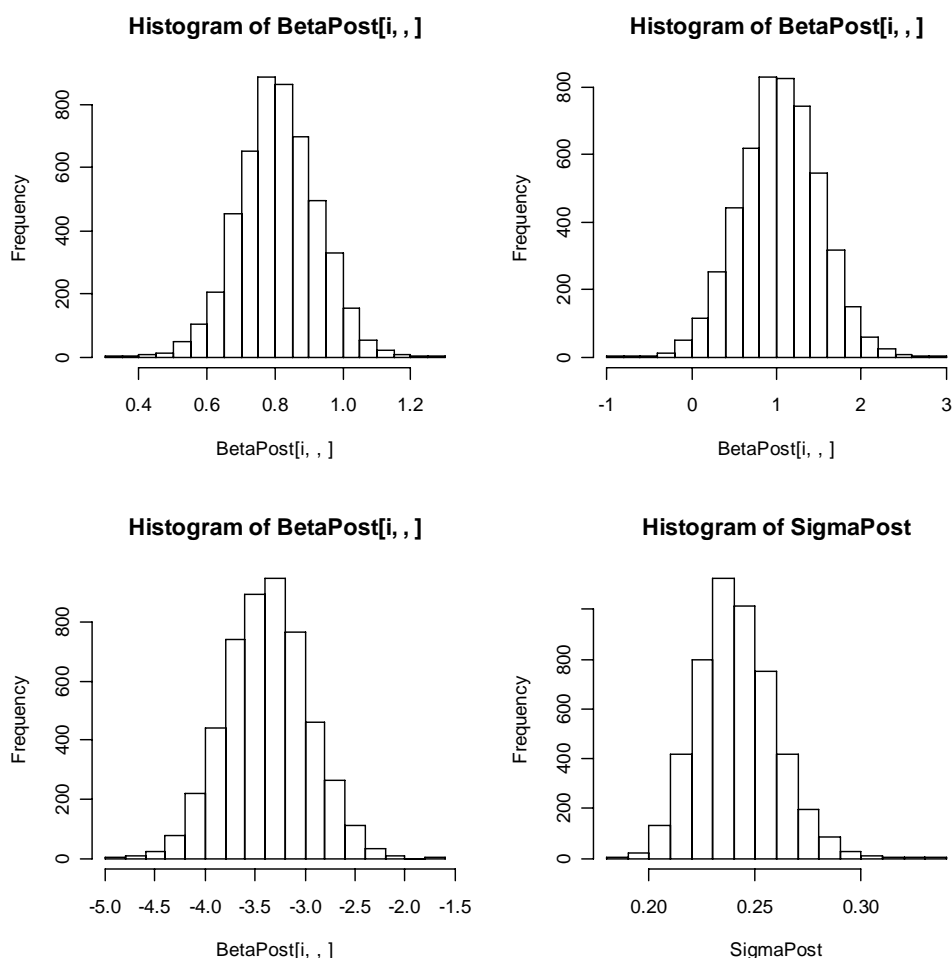$$\hat{y} = \beta_0 + 1.1 \cdot \beta_1 + 1.1^2 \cdot \beta_2 + \varepsilon.$$

FIGURE 13. Posterior distribution of the regression coefficients in the quadratic model.

This is a simple function of the three β coefficients and the regression error $\varepsilon \sim N(0, \sigma^2)$. To obtain the predictive distribution for $x = 1.1$ we there simply compute $\beta_0 + 1.1 \cdot \beta_1 + 1.1^2 \cdot \beta_2$ for each posterior draw that we get from BACC, add an error draw $\varepsilon \sim N(0, \sigma^2)$, and plot a histogram of the resulting $\hat{y}$ draws. Note that σ^2 changes from draw to draw. Figure 15 gives the results.

Figure 16 displays the fitted values and the actual data. The figure plots $y = \hat{\beta}_0 + x\hat{\beta}_1 + x^2\hat{\beta}_2$ against all the x's in the data. $\hat{\beta}$ denotes the posterior mean of the parameter. See my script Exam1Question4a.r for details. The fit in of the quadratic model is far from acceptable.
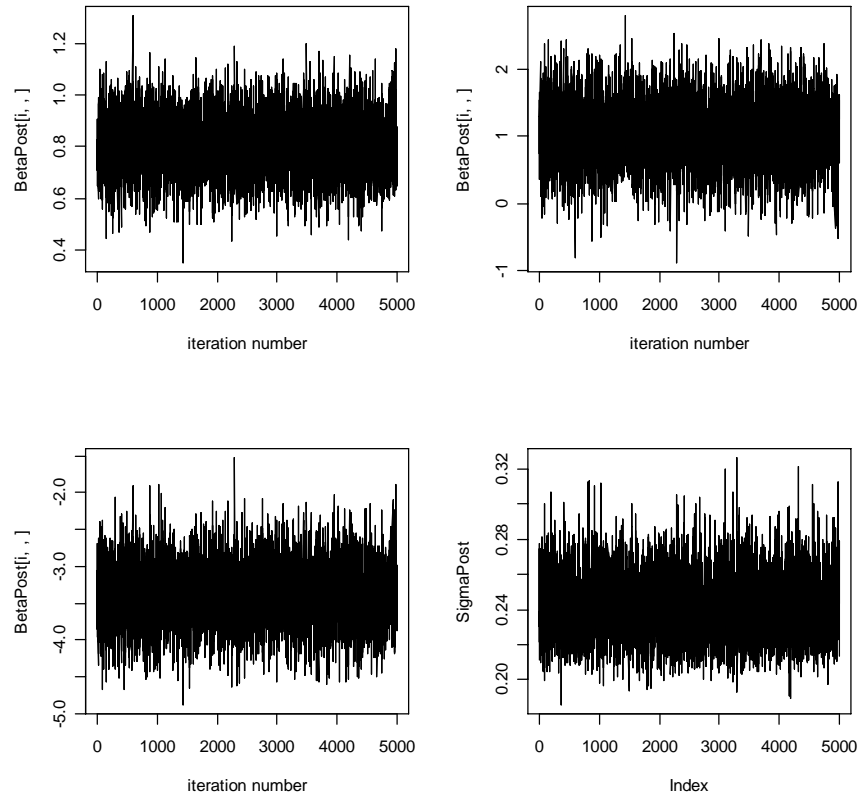
FIGURE 14. MCMC trajectories for the regression coefficients in the quadratic model.

Question 4b. *"Fit a linear spline with 10 fixed knots to the data. Compare the fit of the spline to the quadratic fit in 4.b. Compute the predictive distribution for $x = 1.1$. [Hint: construct the linear splines in R, then use BACC on the constructed spline covariates]."*

Solution 4b. The script Exam1Question4a.r takes you through the solution. Some graphs are given below. The spline fit is substantially better than the quadratic fit.

Histogram of yfit
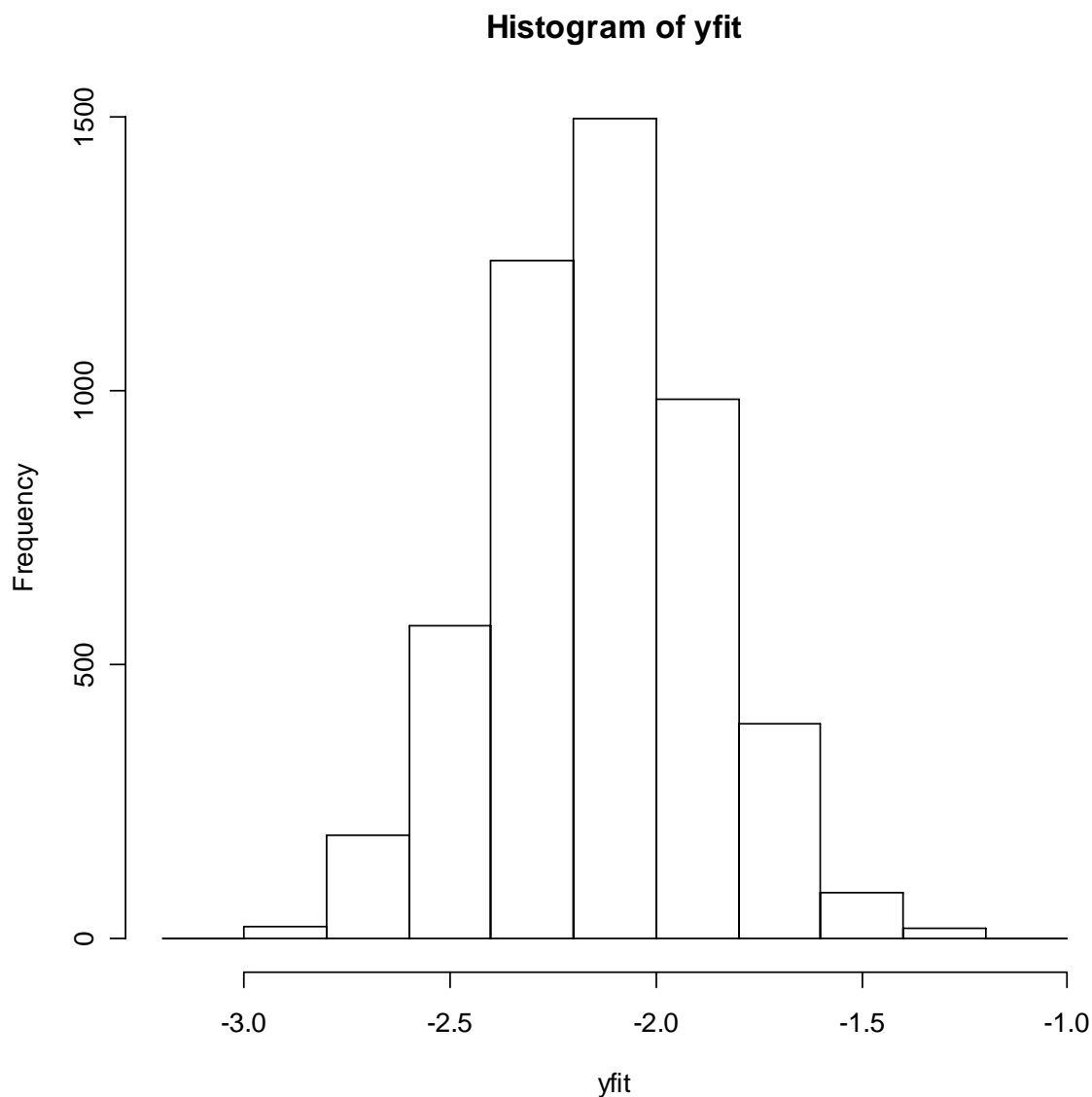
FIGURE 15. Predictive distribution of y for $x = 1.1$. Quadratic model.

Question 4c. *"Consider the following alternative model for the regression data:*

$$y_i = \sin(x_i\alpha_1) + \varepsilon_i,$$

where $\varepsilon_i \overset{iid}{\sim} N(0, 0.1^2)$. Compute the posterior distribution of α_1. Use a uniform prior between -10 and 10. Compute the predictive distribution for $x = 1.1$. [Hint: evaluate the posterior on a grid. Alternatively, use the Metropolis algorithm]."

Solution 4c. Since the prior is uniform, the posterior is proportional to the likelihood:

$$p(y_1, ..., y_n | x, \alpha_1) \propto \exp\left[-\frac{1}{2 \cdot 0.1^2} (y_i - \sin(x_i\alpha_1))^2 \right].$$
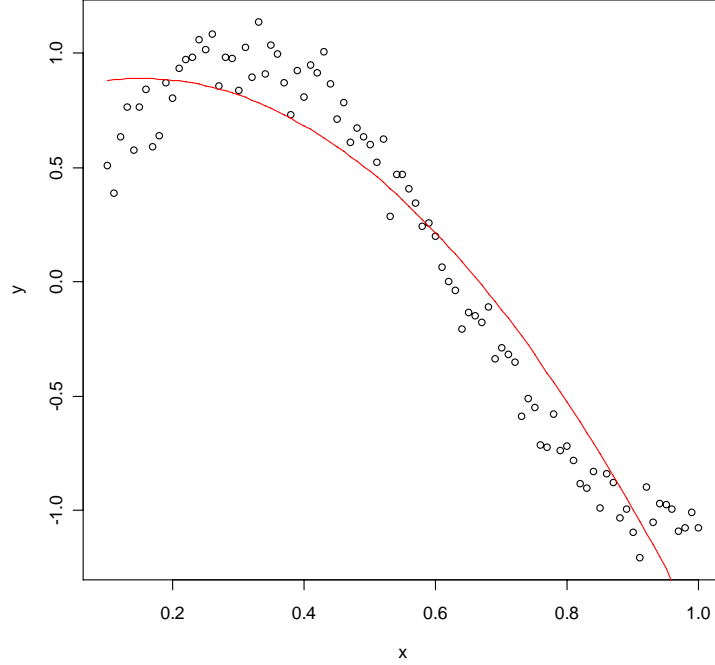
FIGURE 16. Actual data (points) and model fitted values (red line) from quadratic model. Posterior mean is used to estimate the parameters.

All we need to do is to evaluate this expression on a grid, and plot the resulting curve. The script Exam1Question4a.r does the job. The result is given in Figure 19.

To obtain the predictive density we use that

$$p(y|x = 1.1, Data) = \int p(y|x = 1.1, \alpha_1)p(\alpha_1|Data)d\alpha_1.$$

We have $p(\alpha_1|Data)$ represented over the grid: $p(\alpha_1 = z_1|Data), ..., p(\alpha_1 = z_N|Data)$, where $z_1, ..., z_n$ are the grid points. We can therefore approximate:

$$\int p(y|x = 1.1, \alpha_1)p(\alpha_1|Data)d\alpha_1 \approx \sum_{i=1}^{N} p(y|x = 1.1, \alpha_1 = z_i)p(\alpha_1 = z_i|Data).$$

We need to evaluate this sum for many different values of y. That the sum needs to be evaluate over a y-grid: $y_1, ..., y_m$. See my script for details. The predictive density is displayed in Figure 20.

Figure 21 plots the fit of the sine-model using the posterior mode of α_1 (this a bonus-figure, I did not ask for it). The fit is really good. This comes as no surprise to me: the actual data was in fact simulated from the sine model with $\alpha_1 = 5$! The true mean at $x = 1.1$ is therefore $\sin(5 \cdot 1.1) = -0.70554$.
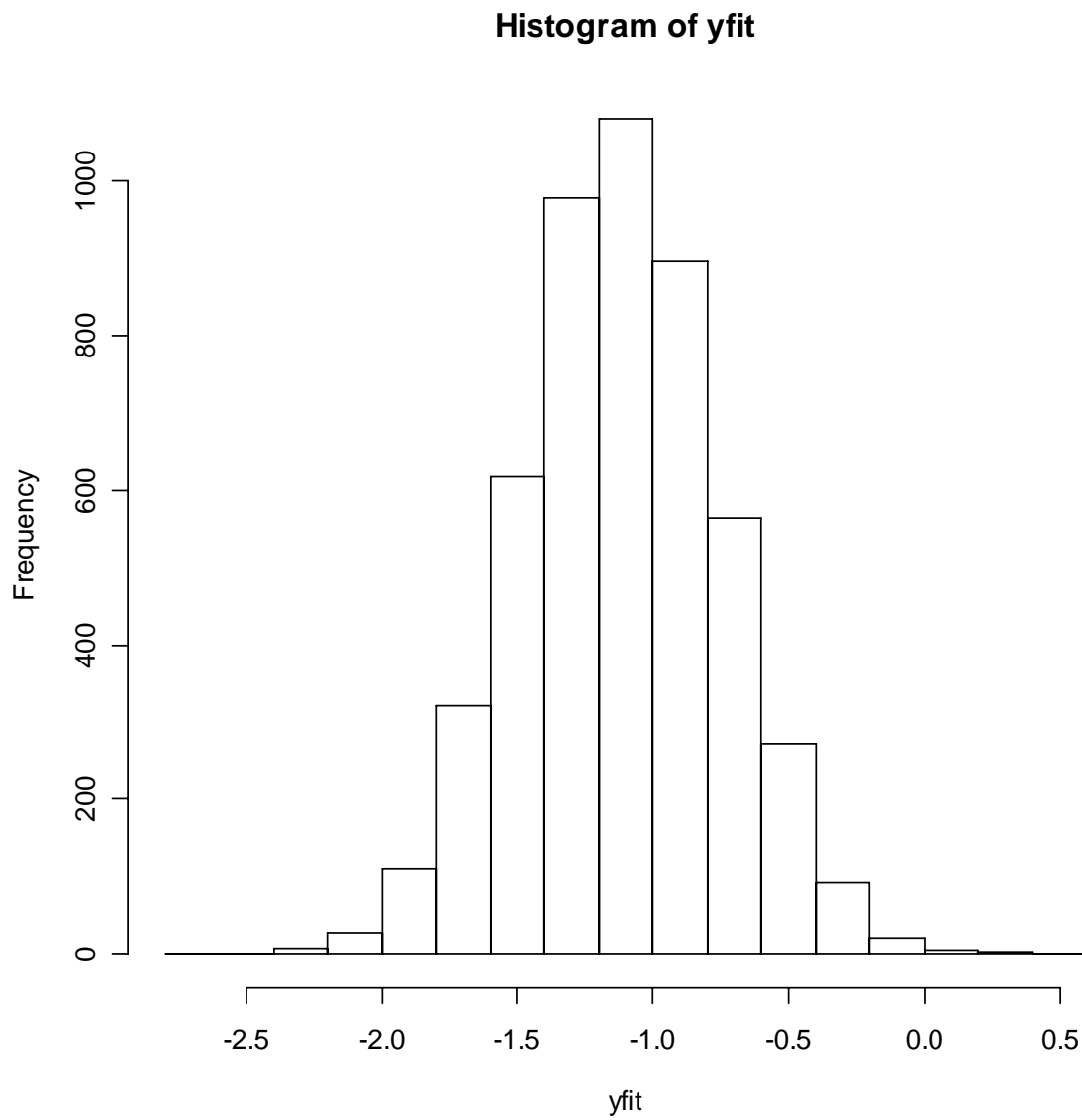
Histogram of yfit

FIGURE 17. Predictive distribution of y for $x = 1.1$. Spline model.
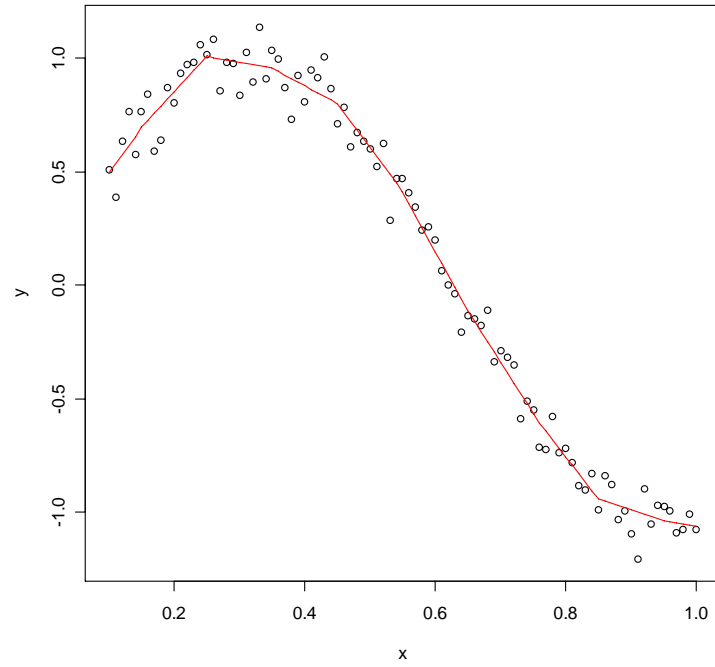
FIGURE 18. Actual data (points) and model fitted values (red line) from spline model. Posterior mean is used to estimate the parameters.
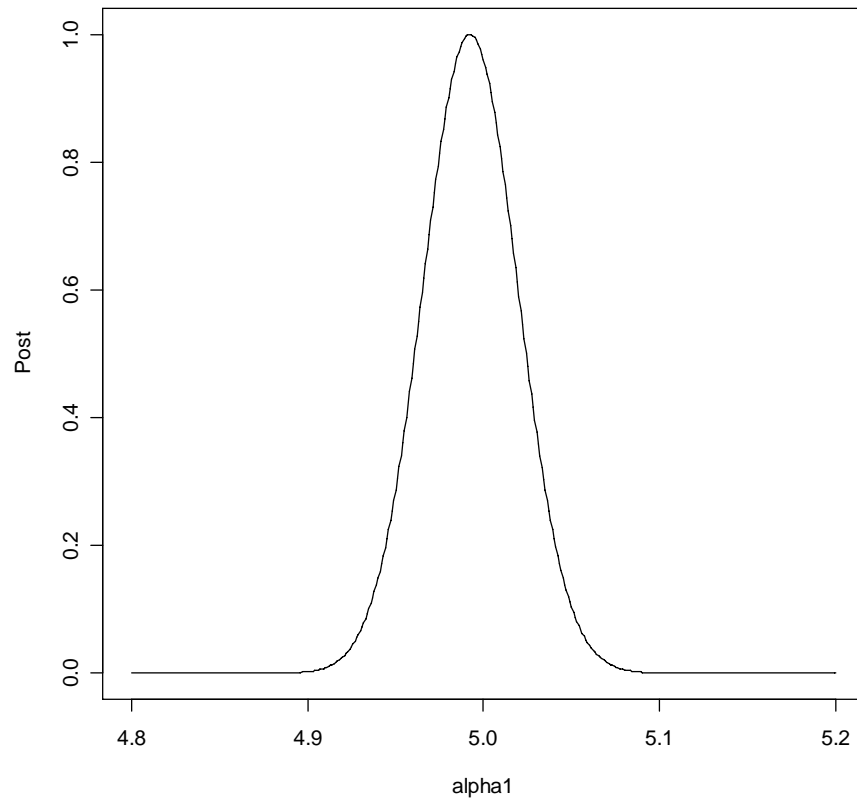
MATTIAS VILLANI

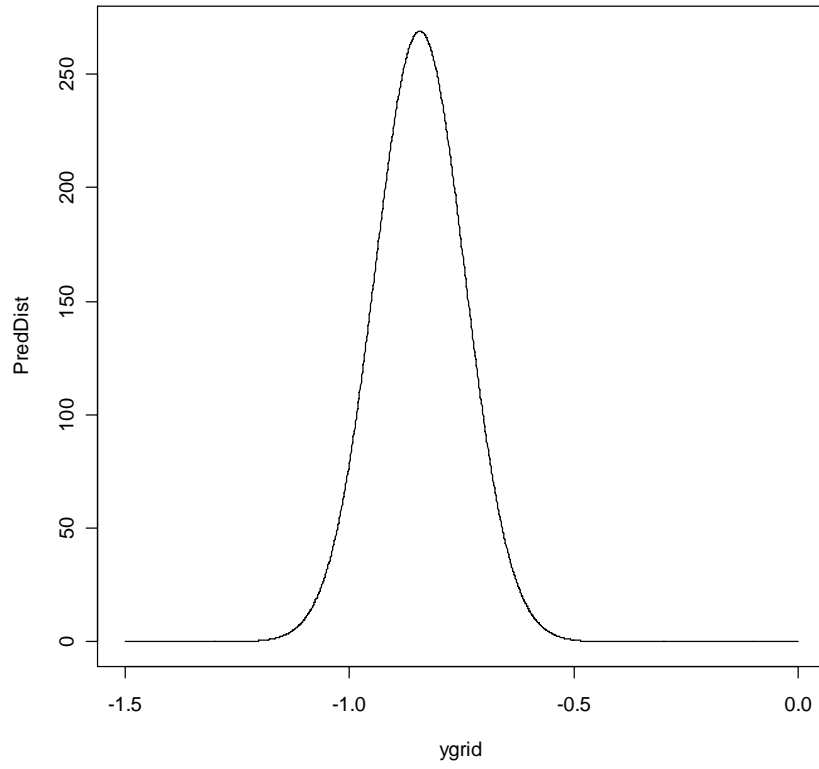FIGURE 19. Question 4c. Posterior density for α_1.

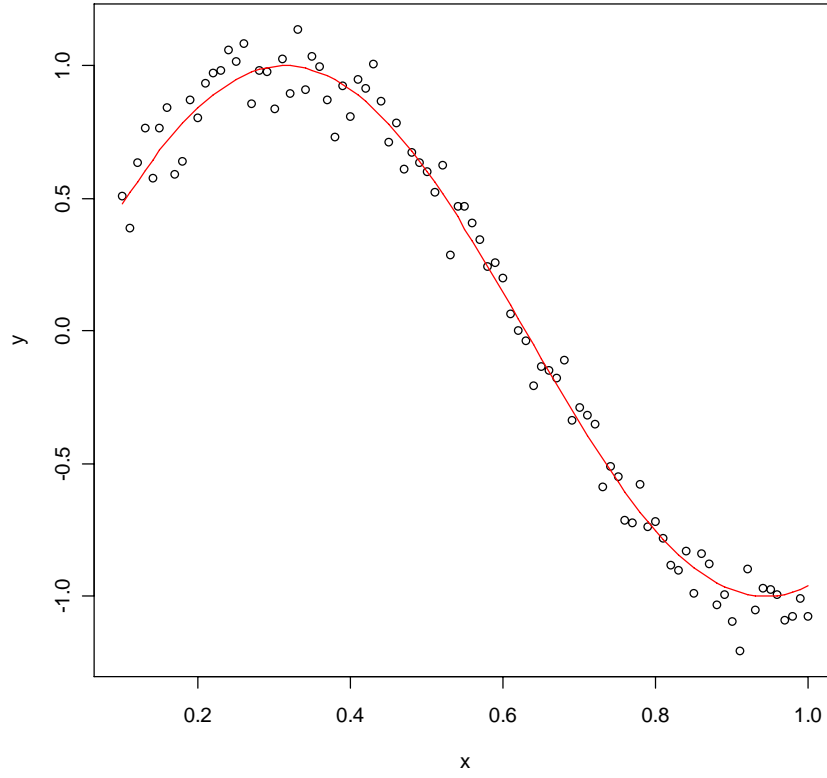FIGURE 20. Predictive distribution of y for $x = 1.1$. Sine model.

FIGURE 21. Actual data (points) and model fitted values (red line) from sine model. Posterior mode is used to estimate the parameters.