

Solutions Chapter 13 - Random Walk Metropolis

Alexandra Hotti

December 2019

13. Ticked off

Problem 13.1.1

You specify a $\text{beta}(1; 1)$ distribution as a prior. Use independent sampling to estimate the prior predictive distribution (the same as the posterior predictive except using sampling from the prior in the first step rather than the posterior), and show that its mean is approximately 50.

Answer: The mean lyme disease count of the prior predictive is: 50.764.

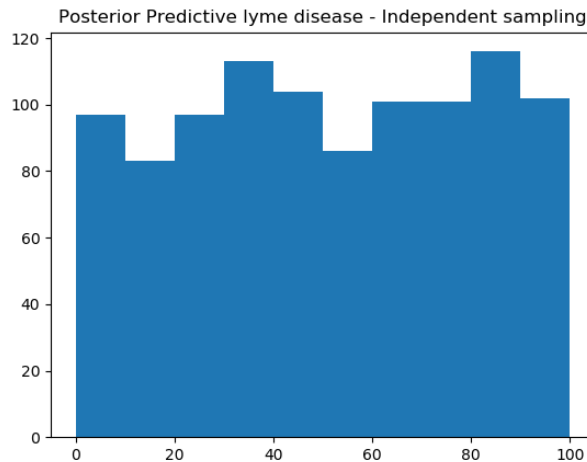


Figure 1: The Prior Predictive distribution estimated with Independent Sampling.

Problem 13.1.2

In a single sample you find that there are 6 ticks that test positive for *Borrelia*. Assuming a $\text{beta}(1; 1)$ prior graph the posterior distribution, and find its mean.

Answer: The beta distribution is a conjugate prior to the binomial likelihood. This gives us the following posterior expression

$$\text{Beta}(\alpha + \sum_{i=1}^n X_i, \beta + \sum_{i=1}^n N_i - \sum_{i=1}^n X_i) = \text{Beta}(1 + 6, 1 + 100 - 6) = \text{Beta}(7, 95) \quad (1)$$

The mean of this distribution is 0.068627.

Problem 13.1.3

Generate 100 independent samples from this distribution using your software's inbuilt (pseudo-)random number generator. Graph this distribution. How does it compare to the PDF of the exact posterior? (Hint: in R the command is `rbeta`"; in Matlab it is `betarnd`; in Mathematica it is `RandomVariate[BetaDistribution...]`"; in Python it is `numpy.random.beta`".)

Answer:

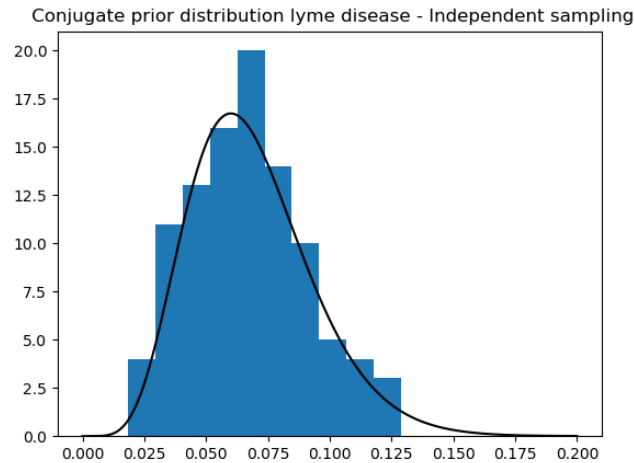


Figure 2: Distribution estimate over disease percentage in population using samples from a conjugate prior (blue), vs actual pdf (black).

Problem 13.1.5

Estimate the variance of the posterior using independent sampling for a sample size of 100. How does your sample estimate compare with the exact solution?

Answer: The actual pdf variance 0.00062056. Estimated variance from 100 samples.

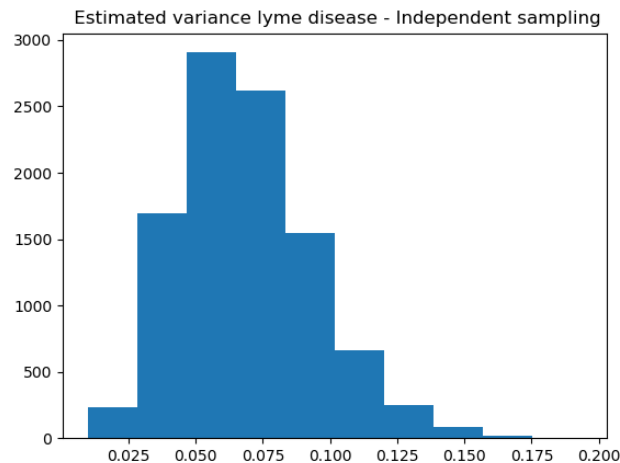


Figure 3

Problem 13.1.6.

Create a proposal function for this problem that takes as input a current value of θ , along with a step size, and outputs a proposed value. For a proposal distribution here we use a normal distribution centred on the current θ value with a standard deviation (step size) of 0.1. This means you will need to generate a random θ from a normal distribution using your statistical software's inbuilt random number generator. (Hint: the only slight modification you need to make here is to ensure that we don't get $\theta < 0$ or $\theta > 1$ is to use periodic boundary conditions. To do this we use modular arithmetic. In particular we set $\theta_{proposed} = \text{mod}(\theta \text{ proposed}; 1)$. The command for this in R is `x%%1`; in Matlab the command is `mod(x, 1)`; in Mathematica it is `Mod[x, 1]`; in Python it is `x % 1`

Answer:

$$r = \begin{cases} 1 & \text{if } p(\theta_{proposed}|data) \geq p(\theta_{current}|data) \\ \frac{p(\theta_{proposed}|data)}{p(\theta_{current}|data)} & \text{if } p(\theta_{proposed}|data) < p(\theta_{current}|data) \end{cases} \quad (2)$$

$$r = \begin{cases} 1 & \text{if } p(\theta_{proposed}|data) \geq p(\theta_{current}|data) \\ \frac{p(data|\theta_{proposed})p(\theta_{proposed})}{p(data|\theta_{current})p(\theta_{current})} & \text{if } p(\theta_{proposed}|data) < p(\theta_{current}|data) \end{cases} \quad (3)$$

Problem 13.1.10

For a sample size of 100 from your Metropolis sampler compare the sampling distribution to the exact posterior. How does the estimated posterior compare with that obtained via independent sampling using the same sample size?

Answer:

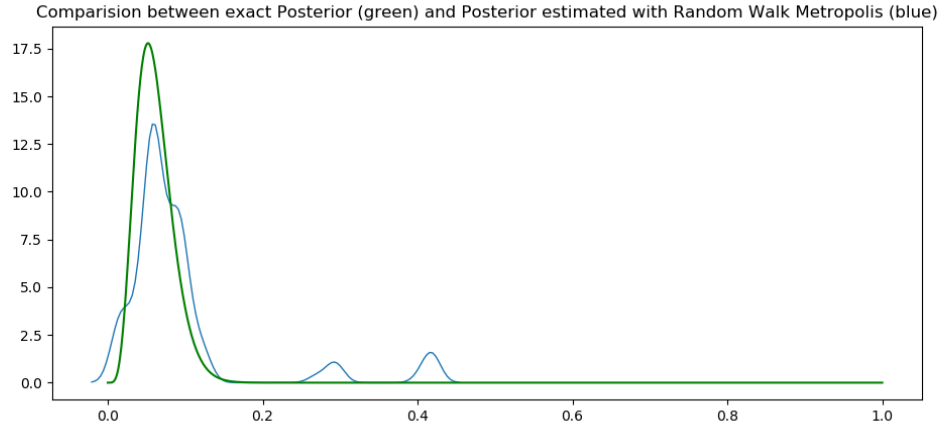


Figure 4: Comparison between the exact conjugate posterior and the posterior estimated using random walk metropolis with 100 samples and a step size of $\sigma = 0.1$

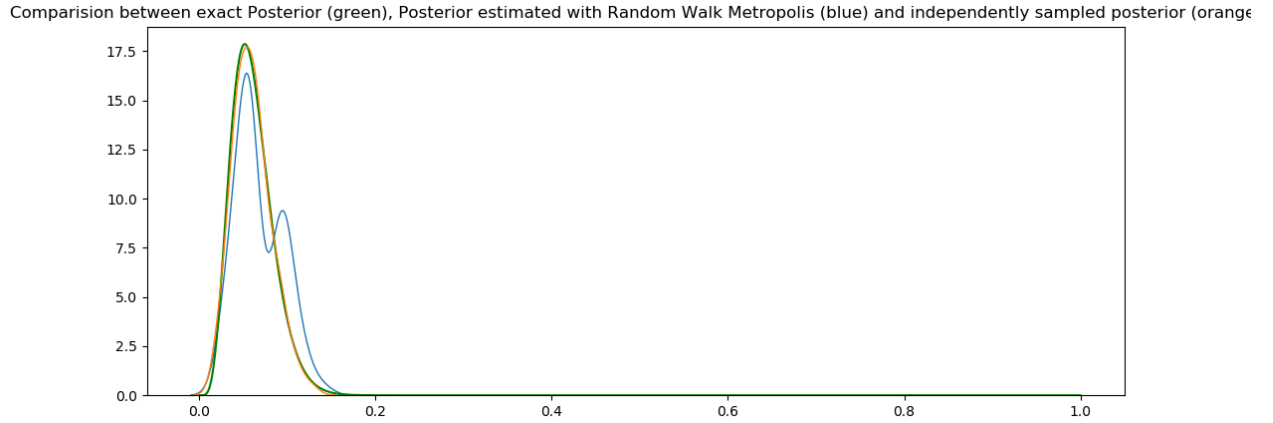


Figure 5: Comparison between the exact conjugate posterior, the posterior estimated using random walk metropolis with 100 samples and a step size of $\sigma = 0.1$ and the posterior estimated via independent samples.

Problem 13.1.11.

Run 1000 iterations, where in each iteration you run a single chain for 100 iterations. Store the results in a 1000 x 100 matrix. For each iterate calculate the sample mean. Graph the resultant distribution of sample means. Determine the accuracy of the MCMC at estimating the posterior mean?

Answer:

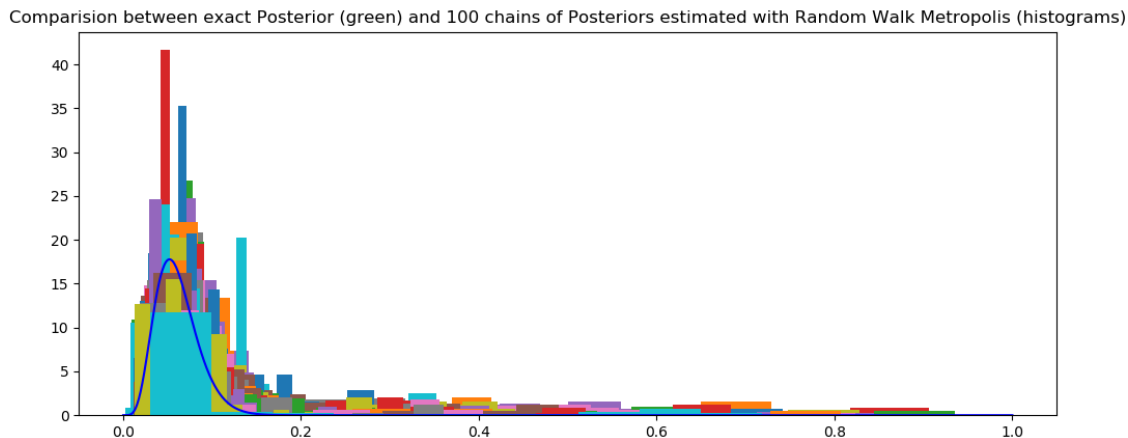


Figure 6: Comparison between the exact conjugate posterior and the posterior estimated using random walk metropolis with 100 samples in 100 chains (histograms) and a step size of $\sigma = 0.1$ and the posterior estimated via independent samples.

Problem 13.1.12.

Graph the distribution of the sample means for the second 50 observations of each chain. How does this result compare with that of the previous question? Why is there a difference?

Answer: Now we have allowed the chains time to converge to the posterior, and hence by discarding the initial 50 iterations we have reduced the effect of the random starting position. Since we are now using converged chains the estimator of the mean is unbiased.

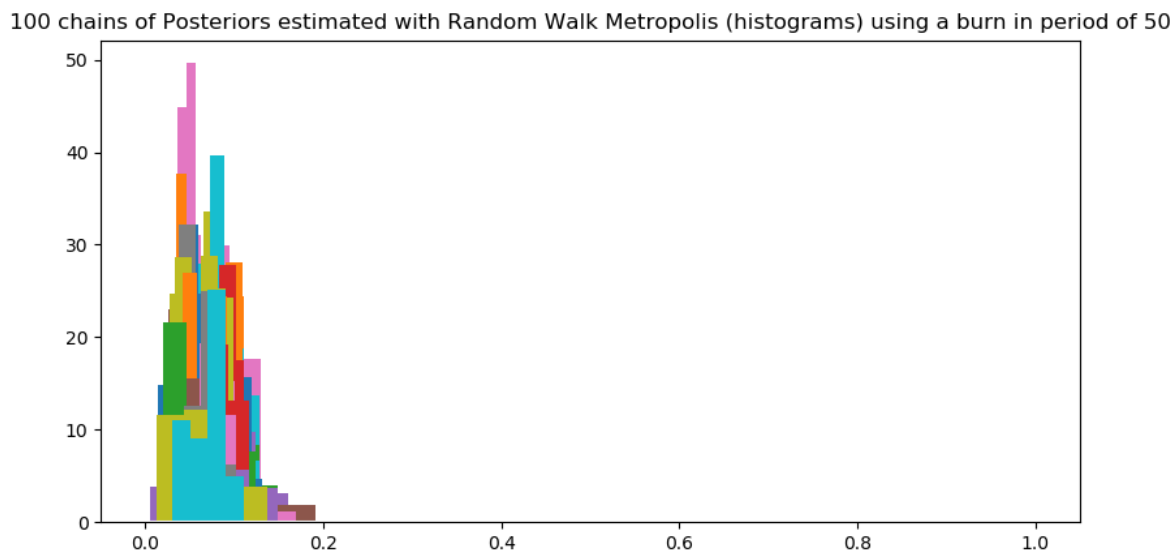


Figure 7: The posterior estimated using random walk metropolis with 100 samples in 100 chains (histograms), a burn in of 50 iterations and a step size of $\sigma = 0.1$ and the posterior estimated via independent samples.

13.1.13.

Decrease the standard deviation (step size) of the proposal distribution to 0.01. For a sample size of 200, how the posterior for a step size of 0.01 compare to that obtained for 0.1?

Answer: When we use the larger step size of 0.1 we are able to find and explore the posterior faster and more accurately. When we use this smaller step size sometimes we manage to get a good posterior estimate, however, sometimes we get stuck somewhere in the density space and can not get away before the number of iterations is up. This is caused by the autocorrelation in the dependent sampler. A small step size means that the algorithm is unable to move far from the previous iteration. Therefore, we need to increase the number of iterations when we use such a small step size, see Figure 8.

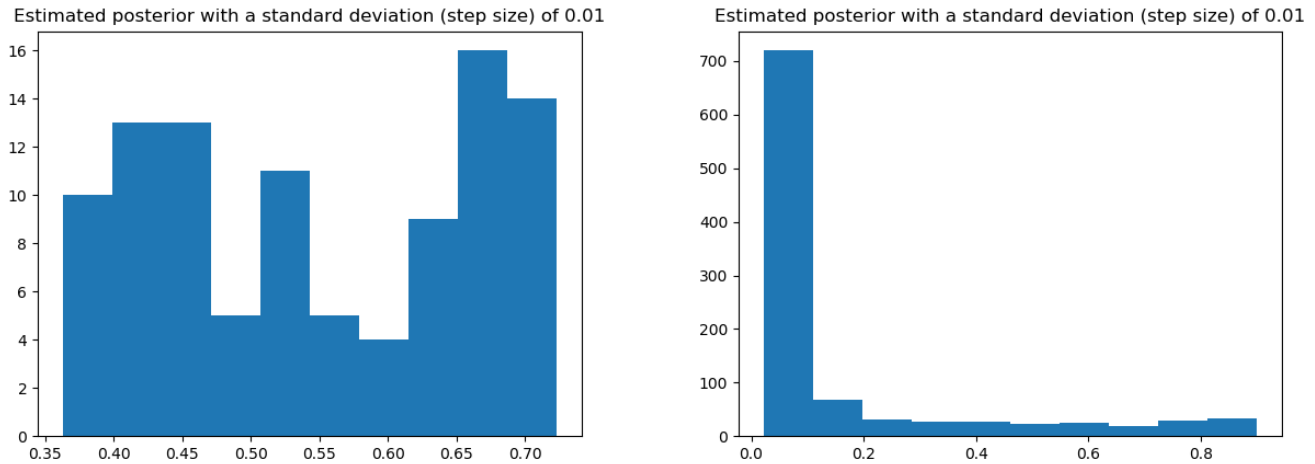


Figure 8: The posterior estimated using random walk metropolis with 100 samples and a smaller step size of $\sigma = 0.01$. To the left we use 100 iterations and to the right 1000 iterations are used.

Problem 13.1.14.

Increase the standard deviation (step size) of the proposal distribution to 1. For a sample size of 200, how does the posterior for a step size of 1 compare to that obtained for 0.1?

Answer: Now the sampler quickly finds the high density region fast. However, since we take such large steps it is poor at exploring this region. Thus, we get a high number of rejections every time we propose a new θ , which will likely be proposed as far away points from the high density regions. Overall we get sampled points around the high density region, however most of these points will lie at the same locations since we get stuck for a long time in high density locations.

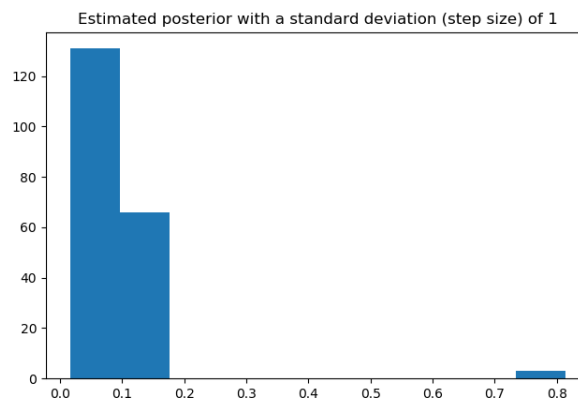


Figure 9: The posterior estimated using random walk metropolis with 200 samples and a larger step size of $\sigma = 1$.

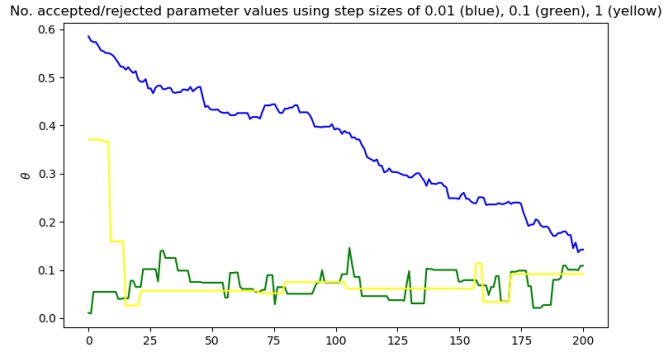


Figure 10: The evolution of the estimated θ parameter using three different step sizes σ . When we use a small step size, convergence is slow (blue) and when the step size is too large we reject most proposals once we find a high-density region (yellow). A middle-ground case is given by the green plot where we make fewer rejections and explore more of the parameter space.

Problem 13.1.15.

Suppose we collect data for a number of such samples (each of size 100), and find the following numbers of ticks that test positive for Borrelia: (3,2,8,25). Either calculate the new posterior exactly, or use sampling to estimate it. (Hint: in both cases make sure you include the original sample of 6.)

Answer: To the left in Figure 13 independent samples from the posterior estimated via RWM can be found. Evidently the distribution is not replicated well. These suggest that either the assumption of independence or identical distribution is violated. If one tick has the bacteria it will infect nearby animals and in doing so, make it more likely for other ticks to become infected; meaning independence is likely violated. Also, due to the contagious nature of the disease that there will be hotspots; meaning that the assumption of identical distribution is likely violated.

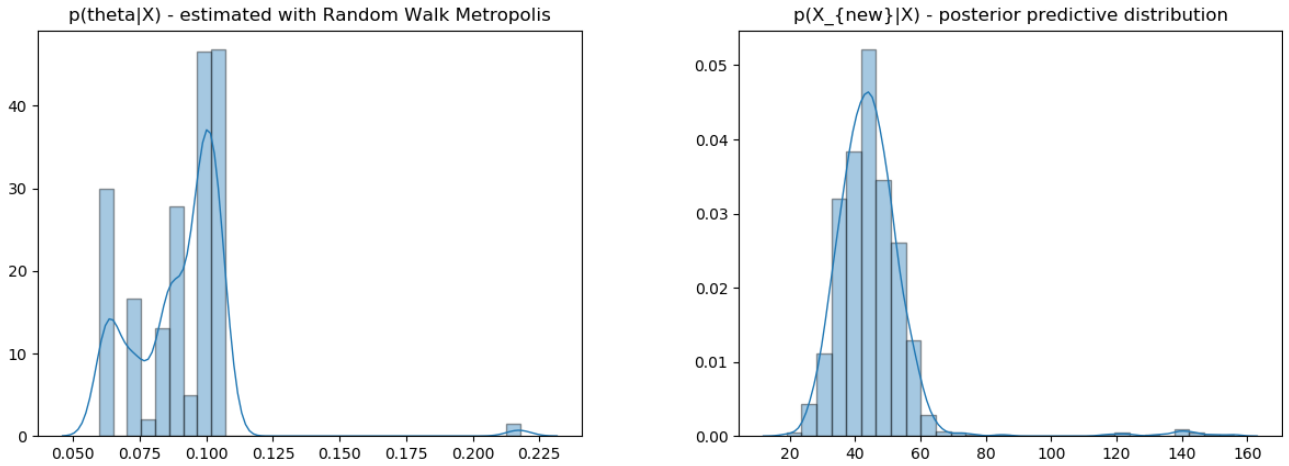


Figure 11: To the left is the distribution estimate created by sampling from the posterior distribution estimated via random walk metropolis. To the right is the posterior predictive distribution.

Problem 13.1.17.

A colleague suggests as an alternative you use a beta-binomial likelihood, instead of the existent binomial likelihood. This distribution has two uncertain parameters $\alpha > 0$ and $\beta > 0$ (the other parameter is the sample size; $n = 100$ in this case), where the mean of the distribution is $\frac{n\alpha}{\alpha+\beta}$. Your colleague and you decide to use weakly informative priors of the form: $\alpha \sim \Gamma(1, \frac{1}{8})$, $\beta \sim \Gamma(10, 1)$ (Here we use the parameterisation such that the mean of $\Gamma(a, b) = \frac{a}{b}$). Visualise the joint prior

in this case.

Answer: Note, if our conclusion regarding the ticks in the problem above holds we will have increased variance between data sets. Thus, we need to use the beta-binomial. Also, we no longer get a symmetrically bounded prior for α and β . Thus, we need to account for this by using the Random Walk Metropolis-Hastings algorithm.

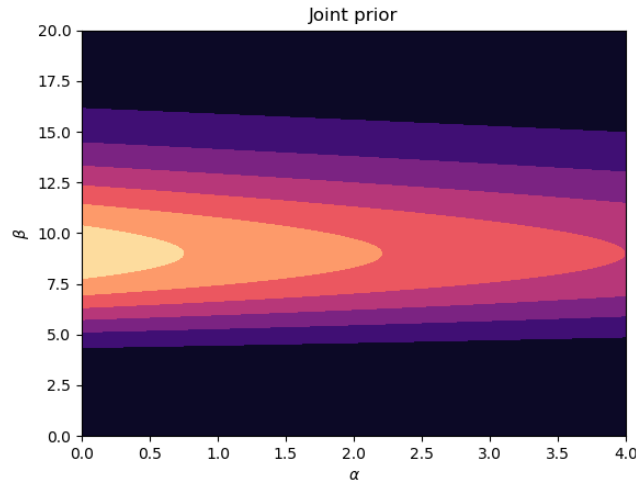


Figure 12: The joint prior between alpha and beta

Problem 13.1.19.

By using one of the the Random Walk Metropolis-Hastings algorithm, estimate the joint posterior distribution. Visualise the PDF of the joint posterior. How are α and β correlated here?

Answer:

First, the pmf of the beta-binomial likelihood gets the following form:

$$f(x|n, \alpha, \beta) = \binom{n}{x} \frac{B(x + \alpha, n - x + \beta)}{B(\alpha, \beta)} \quad (4)$$

where

$$\text{where } B(u, v) = \int_0^1 t^{u-1} (1-t)^{v-1} dt \text{ is the (complete) beta function} \quad (5)$$

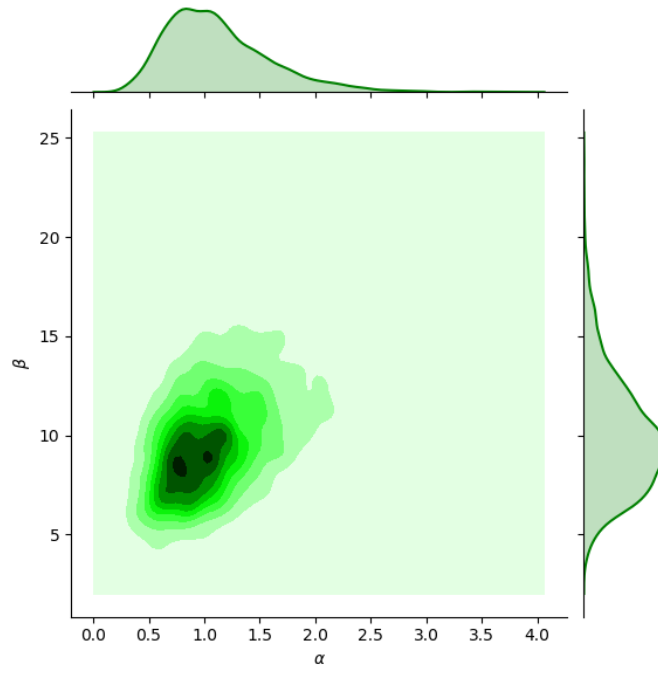


Figure 13: The joint posterior between alpha and beta estimated with Metropolis Hastings.

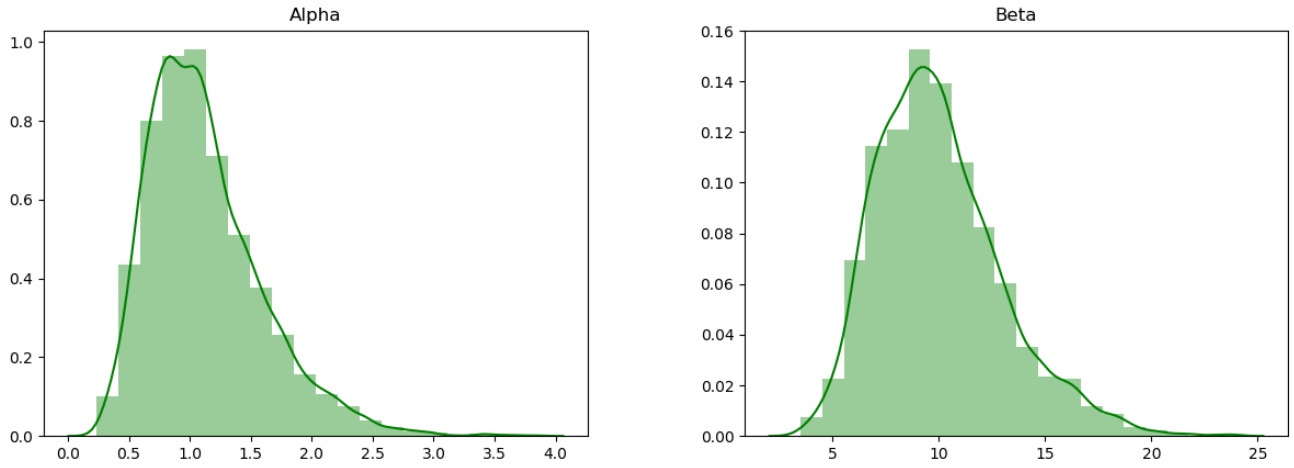


Figure 14: The posteriors of alpha and beta estimated with Metropolis Hastings

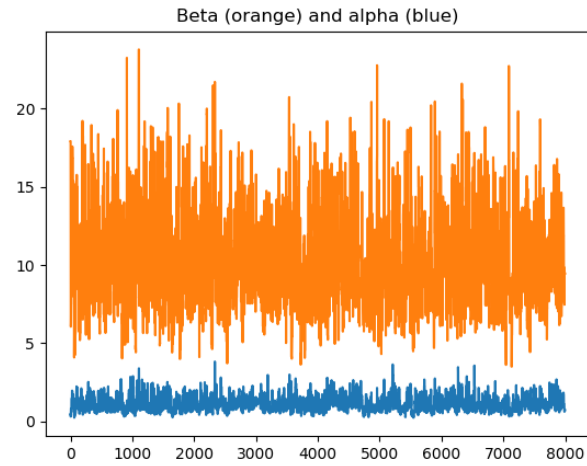


Figure 15: The acceptance of proposal values of alpha and beta estimated with Metropolis Hastings with 8000 iterations.

Problem 13.1.20.

Construct 80% credible intervals for the parameters of the beta-binomial distribution.

Answer: Credible intervals of 80% around the highest density regions

$$0.43526715 \geq \alpha \leq 1.5495638$$

$$5.77234643 \geq \beta \leq 13.11710476$$

Problem 13.1.21.

Carry out appropriate posterior predictive checks using the new model. How does it fare?

Answer:

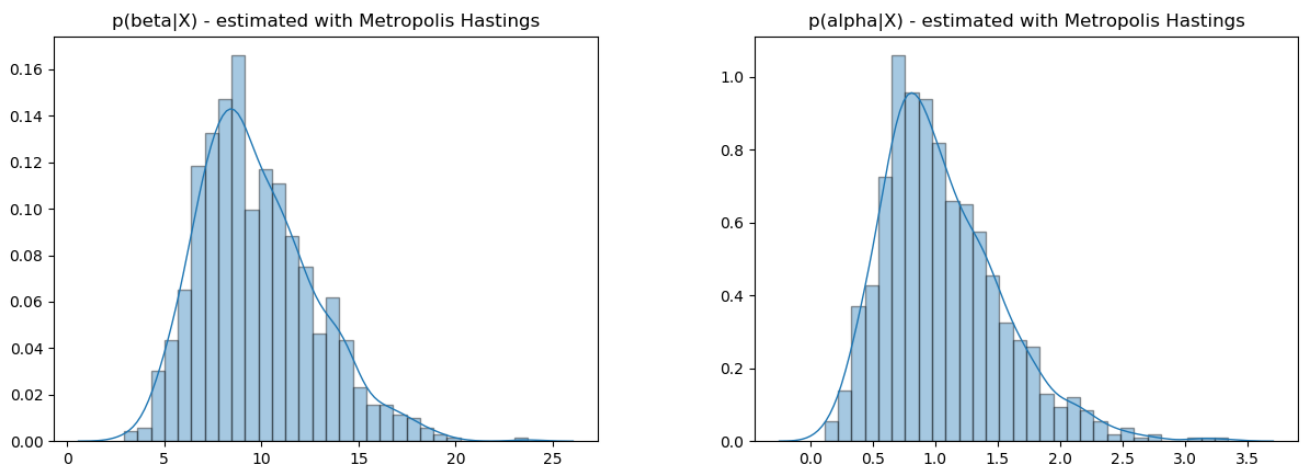


Figure 16

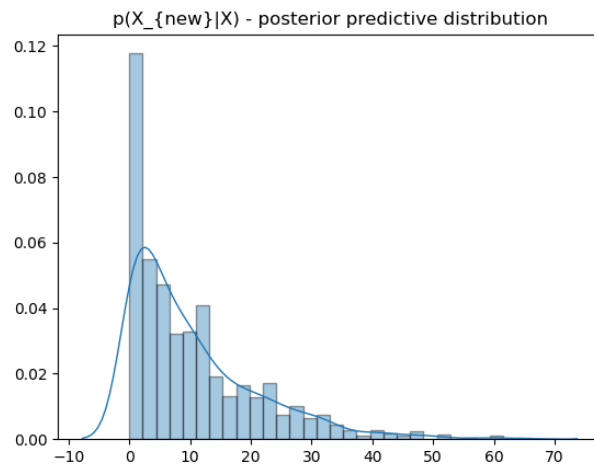


Figure 17: Posterior predictive - Expected number of disease-positive ticks in samples of 100

Now both the maximum (25) and the minimum (2) number of ticks are contained in the posterior predictive distribution, see Figure 17.

13.3 Malarial mosquitoes

Problem 13.3.1.

Using the data create a function that returns the likelihood.

Problem 13.3.2.

Find the maximum likelihood estimates of (ψ, μ) . (Hint 1: this may be easier if you create a function that returns the log-likelihood, and maximise this instead. Hint 2: use R's optim function.)

Answer: The MLE is estimated to be at: $\psi = 0.041206030150753775$ and $\mu = 0.09723618090452261$

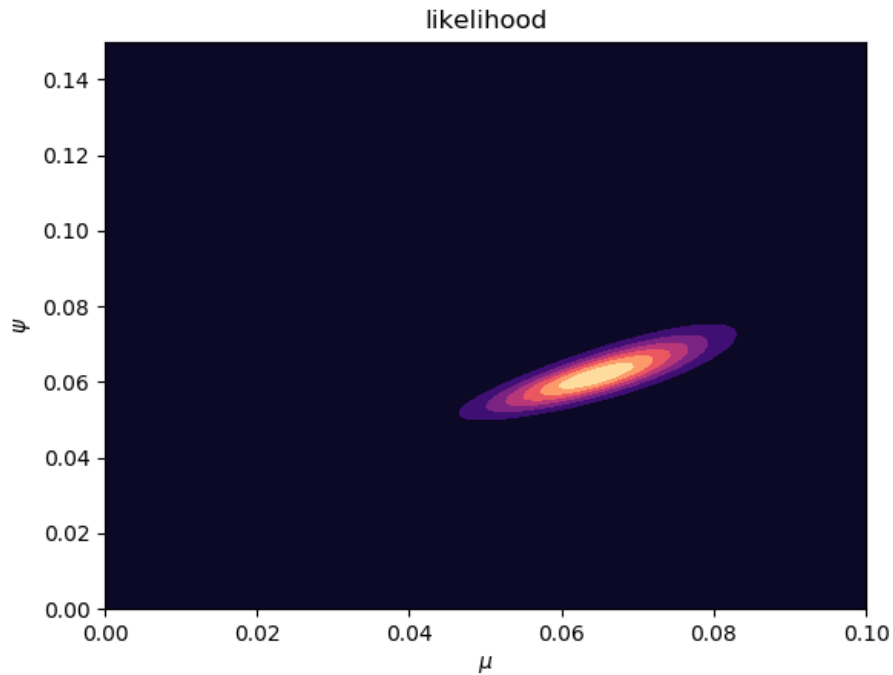


Figure 18: The likelihood distribution for ψ and μ .

Problem 13.3.8.

Create a Metropolis-Hastings sampler by combining your proposal and acceptreject functions.

Problem 13.3.9.

Use your sampler to estimate the posterior mean of ψ and μ for a sample size of 4000 (discard the first 50 observations.) (Hint: if possible, do this by running 4 chains in parallel.)

Answer:

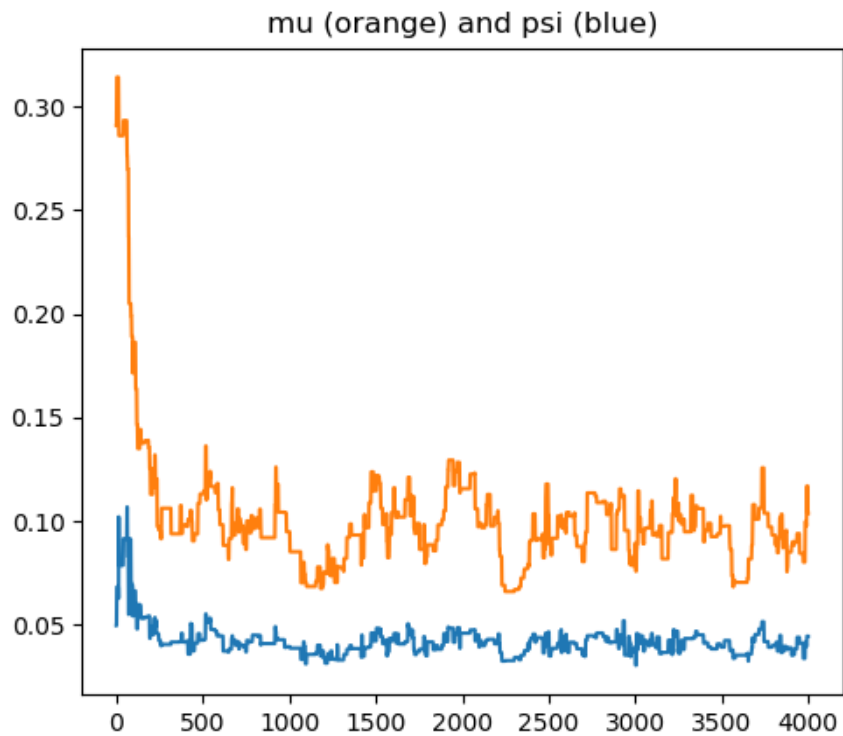


Figure 19: The acceptance chain for ψ and μ .

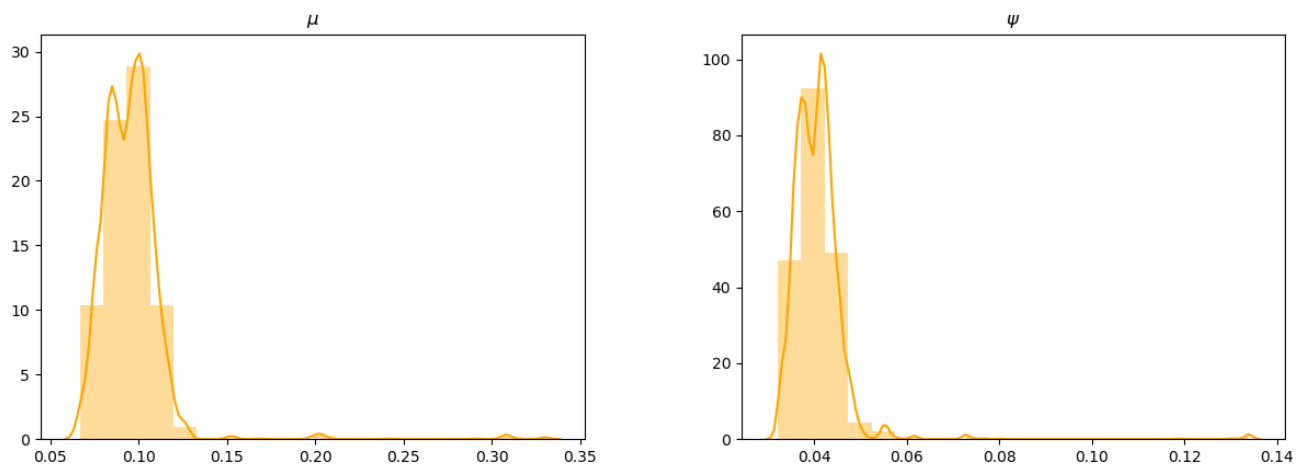


Figure 20: The 1 dimensional posterior estimates for ψ and μ . Estimated using Metropolis Hastings.

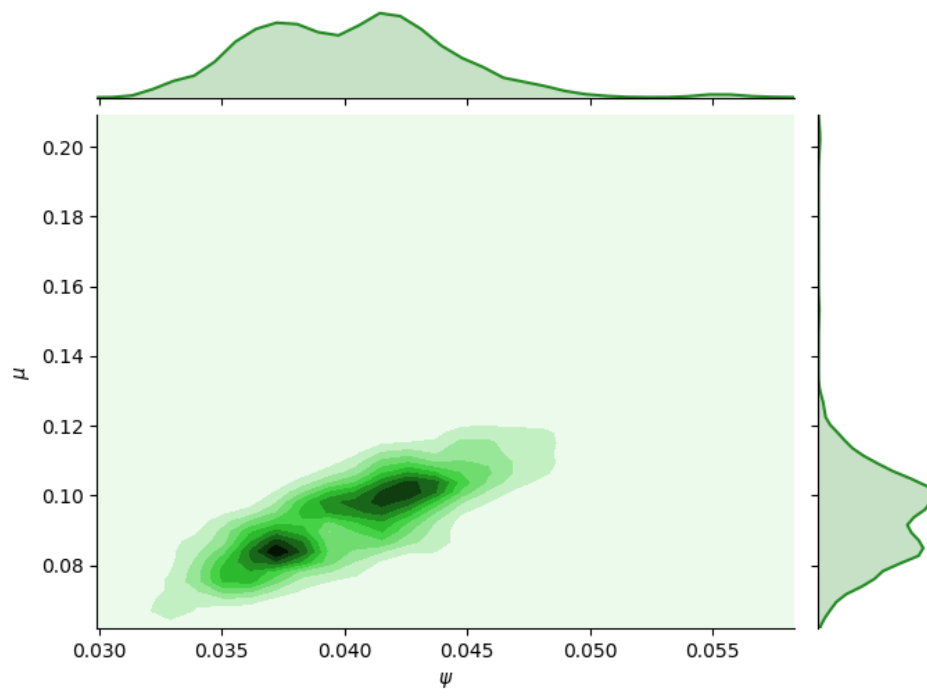


Figure 21: The joint posterior estimate for ψ and μ . Estimated using Metropolis Hastings.

Credible intervals of 80% around the highest density regions

$$0.08466733 \geq \mu \leq 0.11324098$$

$$0.03625231 \geq \psi \leq 0.04594355$$