Solutions Chapter 9 - Conjugate priors

Alexandra Hotti

November 2019

9.1 The epidemiology of Lyme disease

Lyme disease is a tick-borne infectious disease spread by bacteria of species Borrelia, which are transmitted to ticks when they feed on animal hosts. Whilst fairly common in the US, this disease has recently begun to spread throughout Europe.

Imagine you are researching the occurrence of Lyme disease in the UK. As such, you begin by collecting samples of 10 ticks from fields and grasslands around Oxford, and counting the occurrence of the Borrelia bacteria.

Problem 9.1.1.

You start by assuming that the occurrence of Borrelia bacteria in one tick is independent of that in other ticks. In this case, why is it reasonable to assume a binomial likelihood?

Answer: Assuming that the disease status of ticks is independent (Which I of course know nothing about) and assuming the same disease prevalence in every field (identically-distributed with the same probability for every trial), then since we are counting occurrences (discrete data) and we are using a fixed sample size of 10 in every field we can use a binomial likelihood.

Problem 9.1.2.

Suppose the number of Borrelia-positive ticks within each sample i is given by the random variable X_i, and that the underlying prevalence (amongst ticks) of this disease is θ. Write down the likelihood for sample i.

Answer:

$$Pr(X_i|\theta) = \binom{10}{X_i}\theta^{X_i}(1-\theta)^{10-X_i} \tag{1}$$

The likelihood can be found in Figure 2. The θ_{MLE} is evidently located at $\theta = 0.1$.

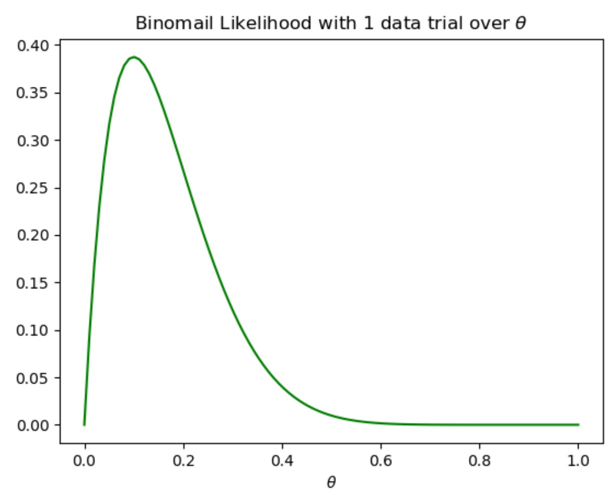

Figure 1: Binomial likelihood after 1 trial with a data sample of size 10 containing one disease occurrence.

0.1 Problem 9.1.4.

By numerical integration show that the area under the likelihood curve is about 0.09. Comment on this result.

Since we vary the parameter θ this is not a valid probability distribution and therefore the area under the graph does not integrate to 1.

0.2 Problem 9.1.5.

Assuming that $\theta = 10\%$, graph the probability distribution (also known as the sampling distribution). Show that, in contrast to the likelihood, this distribution is a valid probability distribution.

Answer: See Figure 2 for the distribution plot. By simply summing over the distribution for X_i, (i.e. by summing the bars) i get: $1.0000000000000004 \approx 1$. Also, we can see that all values are non negative in the bar chart and thus it is a valid distribution. This is to be expected since we vary the data and not the distribution parameter.

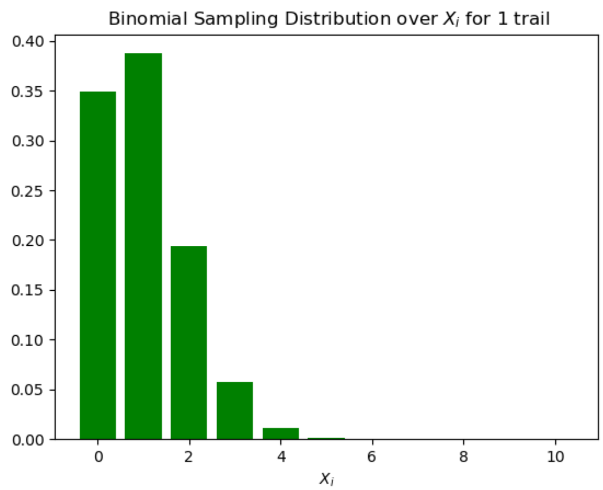

Figure 2: Binomial sampling distribution for X_i after 1 trial with a data sample of size 10.

Problem 9.1.7.

A colleague mentions that a reasonable prior to use for θ is a beta(a, b) distribution. Graph this for a = 1 and b = 1.

Answer:

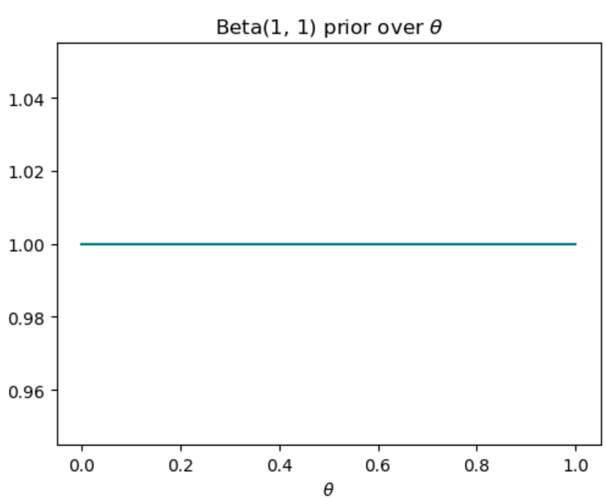

Figure 3: Beta(1,1) prior over θ.

Problem 9.1.8.

How does this distribution change as you vary a and b?

Answer: The mean of the distribution is computed as: $\frac{a}{a+b}$ such that when $a = b$ we get a distribution centered on $\theta = 0.5$. Large values of a and b makes the distribution more peaked around the mean and decreases the standard deviation. When $a > b$ the mean is shifted towards 1 and $a < b$ the mean is shifted towards 0.

Problem 9.1.9.

Prove that a $beta(a, b)$ prior is conjugate to the binomial likelihood, showing that the posterior distribution is given by a $beta(X + a, 10X + b)$ distribution.

Answer: The binomial likelihood is proportional to:

$$p(X|\theta) \propto \theta^X (1 - \theta)^{10-X} \tag{2}$$

The beta prior is proportional to:

$$p(\theta|a, b) \propto \theta^{a-1}(1 - \theta)^{b-1} \tag{3}$$

We easily see that this prior and likelihood have the same shape. Now, the posterior is proportional to:

$$p(\theta|X) \propto \theta^X (1 - \theta)^{10-X} \cdot \theta^{a-1}(1 - \theta)^{b-1} = \theta^{X+a-1}(1 - \theta)^{X-10+b-1} \tag{4}$$

which is a $beta(X + a, 10 - X + b)$ pdf!

0.3 Problem 9.1.10.

Graph the posterior for a = 1 and b = 1. How does the posterior distribution vary as you change the mean of the beta prior? (In both cases assume that X = 1.)

Answer:

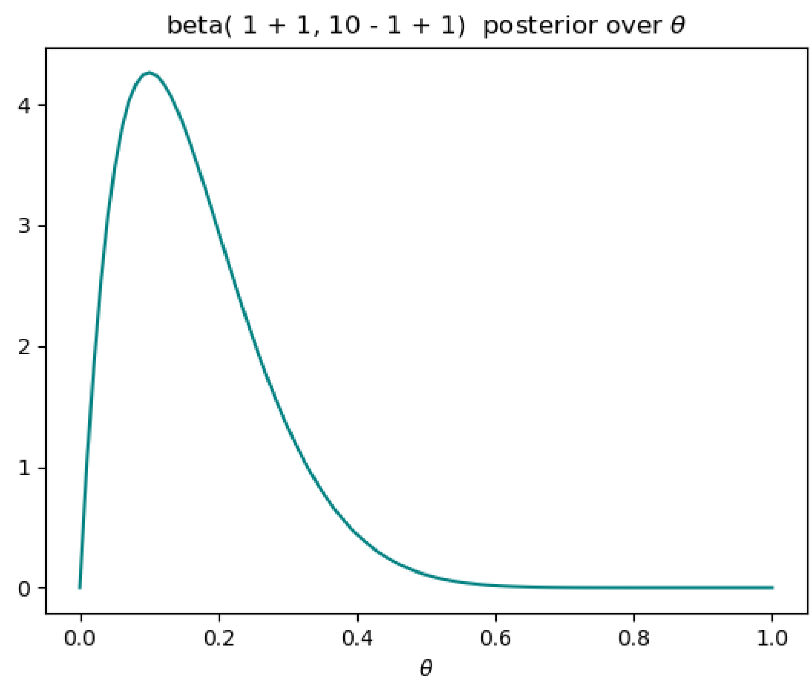

Figure 4: Beta(2,10) posterior over θ.

Problem 9.1.11.

You now collect a larger dataset (encompassing the previous one) that has a sample size of 100 ticks in total; of which you find 7 carry Borrelia. Find and graph the new posterior using the conjugate prior rules for a beta(1, 1) prior and binomial likelihood.

Answer: As we increase the amount of data in the likelihood the posterior variance decreases and its peaks increases as our certainy increases.

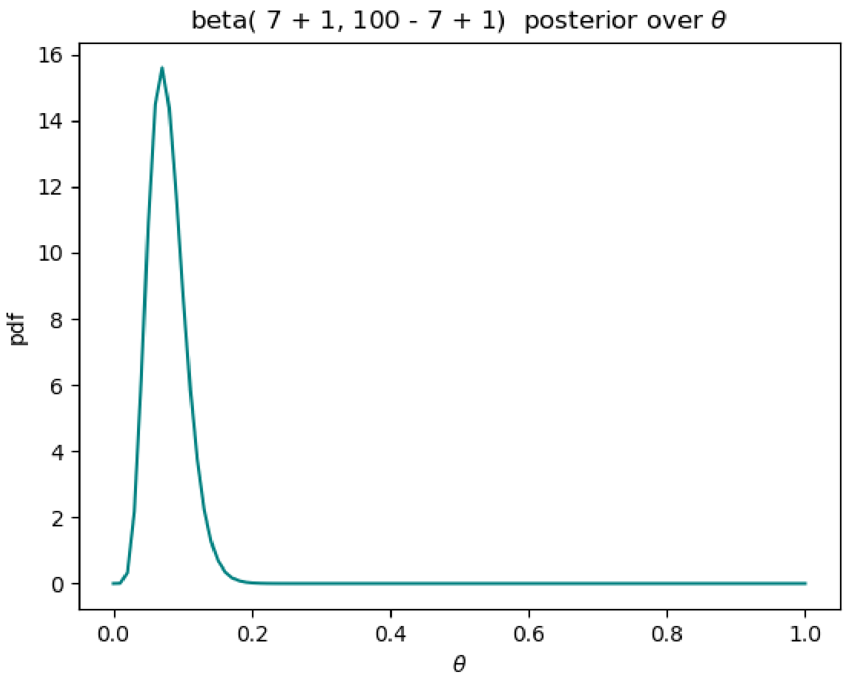

Figure 5: Beta(8, 94) posterior over θ.

Problem 9.1.13.

Now we will use sampling to estimate the posterior predictive distribution for a sample size of 100, using the posterior distribution obtained from the entire sample of 200 ticks (11 of which were disease-positive). To do this we will first sample a random value of from the posterior: so $\theta_i \sim p(\theta|X)$. We then sample a random value of the data X by sampling from the binomial sampling distribution $X_i \sim Binomial(100, \theta_i)$. We repeat this process a large number of times to obtain samples from this distribution. Follow the previous rules to produce 10,000 samples from the posterior predictive distribution, which we then graph using a histogram.

Answer:
In general the posterior predictive is given by:

$$p(x'|x) = \int p(x'|\theta, x)p(\theta|x)d\theta = \int p(x'|\theta)p(\theta|x)d\theta \tag{5}$$

Where the first factor is the posterior and the second the likelihood.

To estimate the posterior predictive, I will use the following scheme:

1. Define $p(\theta|x) = beta(11 + 1, 200 - 1 + 1)$
2. Sample $\theta_i \sim beta(11 + 1, 200 - 1 + 1)$
3. Use θ_i to define the likelihood of the new data: $Binomial(100, \theta_i)$.
4. Sample the number of disease positive ticks from this likelihood: $X_i^{'} \sim Binomial(100, \theta_i)$

5. Store X_i' in a plot of counts over disease occurrences.

The results can be found in Figure 6.

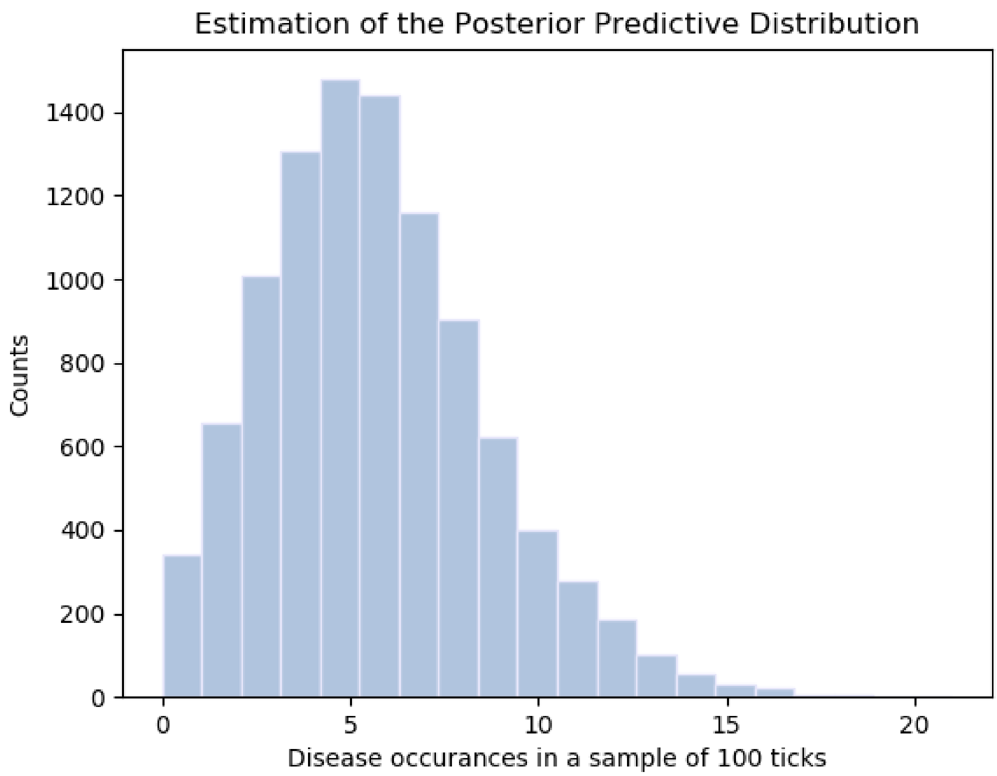

Figure 6: Posterior predictive estimation from 10,000 samples.

Problem 9.1.14.

Does our model fit the data?

Answer: The previous data of sample size 100 had $X_1 = 4$ and $X_2 = 7$. Which are both contained near the peak of the posterior predictive distribution.

Problem 9.1.15.

Indicate whether you expect this model to hold across future sampling efforts.

Answer: One Borrelia infected tick would make it more likely that another tick catches Borrlia (TBE?). Therefore, it might not be possible to assume that the disease status of one tick is independent of the other. An alternative likelihood could be the Beta-Binomial. The Beta-binomial allows for varying probability across trial. Meaning that if one field has Borrelia amongst it ticks, we can account for that these ticks likely will infect each other. So that we can get a larger variance between different trials and fields. One way of figuring out if this holds is to calculate the variance and see which model fits best.

9.2 Epilepsy

In the data file *conjugate_epil.csv* there is a count of seizures for 112 patients with epilepsy who took part in a study [11]. Assume a) the underlying rate of seizures is the same across all patients, and b) the event of a seizure occurring is independent of any other seizures occurring.

Problem 9.2.1.

Under these assumptions what model might be appropriate for this data?

Answer: Since we have counts of discrete events and events that occur at a common given rate independently from each other: A poisson distribution.

Problem 9.2.2.

Write down the likelihood for the data.

Answer:

$$\mathcal{L} = \prod_{i}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta} \tag{6}$$

Problem 9.2.3.

Show that a gamma prior is conjugate to this likelihood.

Answer:
The gamma is proportional to:

$$p(\theta) \propto \theta^{a-1} e^{-b\theta} \tag{7}$$

The posterior is thus proportional to:

$$p(\theta|\bar{x}) \propto \theta^{a-1} e^{-b\theta} \cdot \prod_{i}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta} \propto e^{-b\theta - \sum_i^n \theta} \cdot \theta^{a-1+\sum_i^n x_i} = e^{-\theta(b+n)} + \theta^{a-1+\sum_i^n x_i} \tag{8}$$

Which is a gamma distribution with parameters $a' = a + \sum_{i=1}^{n} x_i$ and $b' = (b+n)$

Problem 9.2.4.

Assuming a $\Gamma(4, 0.25)$ (with a parameterisation such that it has mean of 16) prior. Find the posterior distribution, and graph it.

Answer:

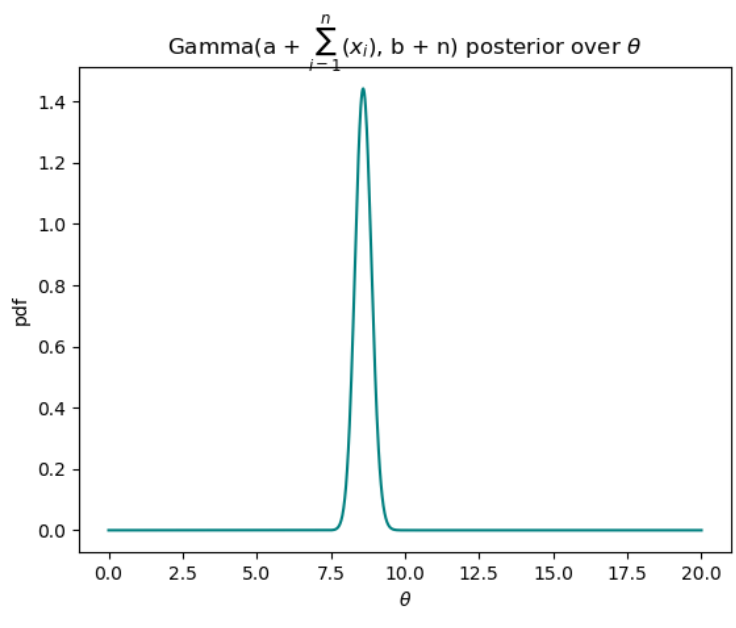

Figure 7: Posterior for Epilepsy counts in a study with 112 individuals. Since we have a conjugate gamma prior to the poisson likelihood the posterior can be computed in a closed form.

Problem 9.2.5.

Find or look-up the posterior predictive distribution, and graph it.

Answer:

$$p(\lambda|x^{'}, \mathbf{x}) = \frac{p(x^{'}|\lambda, x)p(\lambda|\mathbf{x})}{p(x^{'}|\mathbf{x})} \tag{9}$$

Now solving for the posterior predictive:

$$p(x^{'}|\mathbf{x}) = \frac{p(x^{'}|\lambda)p(\lambda|\mathbf{x})}{p(\lambda|x^{'}, \mathbf{x})} = \frac{\mathcal{L} \cdot posterior_{\mathbf{x}}}{posterior_{\mathbf{x}\&x^{'}}} \tag{10}$$

Now the likelihood in eq. (10) is given by the Poisson distribution, the posterior in the numerator is distributed as $gamma(N\bar{x} + \alpha, N + \beta)$ and lastly the posterior in the numerator is distributed as: $gamma(N\bar{x} + \alpha + x^{'}, N + \beta + 1)$. Meaning that we simply add the value of $x^{'}$ to the first parameter in the gamma distribution and then add 1 to the second parameter as we have one extra data point.

Now, we use this information to rewrite the posterior predictive distribution:

$$p(x^{'}|\mathbf{x}) = \frac{\frac{\lambda^{x^{'}} e^{-\lambda}}{x^{'}!} \cdot \frac{(N+\beta)^{N\bar{x}+\alpha}}{\Gamma(N\bar{x}+\alpha)} \lambda^{N\bar{x}+\alpha-1} \cdot e^{-(N+\beta)\lambda}}{\frac{(N+1+\beta)^{N\bar{x}+x^{'}+\alpha}}{\Gamma(N\bar{x}+\alpha+x^{'})} \lambda^{N\bar{x}+\alpha+x^{'}-1} e^{-(N+1+\beta)\lambda}} \tag{11}$$

Now, since the posterior predictive distribution has no dependence on λ, we know that all of these terms should cancel in eq. (12). Rewriting this expression gives us:

$$p(x^{'}|\mathbf{x}) = \frac{\frac{1}{x^{'}!} \cdot \frac{(N+\beta)^{N\bar{x}+\alpha}}{\Gamma(N\bar{x}+\alpha)}}{\frac{(N+1+\beta)^{N\bar{x}+x^{'}+\alpha}}{\Gamma(N\bar{x}+\alpha+x^{'})}} = \frac{(N+\beta)^{N\bar{x}+\alpha}\Gamma(N\bar{x}+\alpha+x^{'})}{x^{'}!\Gamma(N\bar{x}+\alpha)(N+1+\beta)^{N\bar{x}+x^{'}+\alpha}} =$$

$$(\frac{N+\beta}{N+\beta+1})^{N\bar{x}+\alpha}(\frac{1}{N+\beta+1})^{x^{'}} \cdot \frac{\Gamma(N\bar{x}+x^{'}+\alpha)}{x^{'}!\Gamma(N\bar{x}+\alpha)} \tag{12}$$

Now, we see that this is the negative binomial. Where $x^{'} \sim NB(N\bar{x}+\alpha, \frac{N+\beta}{N+\beta+1})$