STAT 222 Homework #2 **Mary Silva**
UCSC 30 April 2020

1. Posterior Inference for one sample problems using DP priors

Consider data $= (y_1, ..., y_n)$, and the following DP-based nonparametric model:

$$Y_i|G \overset{iid}{\sim} G, \quad i = 1, ..., n$$
$$G \sim DP(\alpha, G_0)$$

with $G_0 = N(m, s^2)$ for fixed m, s^2, and α.

The objective here is to use simulated data to study posterior inference results for G under different prior choices for M and G_0, different underlying distributions that generate the data, and different sample sizes. In particular, consider:

- two data generating distributions: 1) a N(0; 1) distribution, and 2) the mixture of normal distributions which yields a bimodal c.d.f. with heavy right tail,

$$0.5N(-2.5,\ 0.5^2) + 0.3N(0.5,\ 0.7^2) + 0.2N(1.5,\ 2^2)$$

- sample sizes n = 20, n = 200, and n = 2000.

SOLUTION: In order to simulate from the posterior distribution, Ferguson's definition of the DP is used. In the first case we draw samples from a $N(0, 1)$ distribution, with a sample size of 20, 200, and 2000. For each simulation I will generate 10 posterior samples. The hyperparameters m and s^2 are chosen to be the sample mean and sample standard deviation of each data sample.

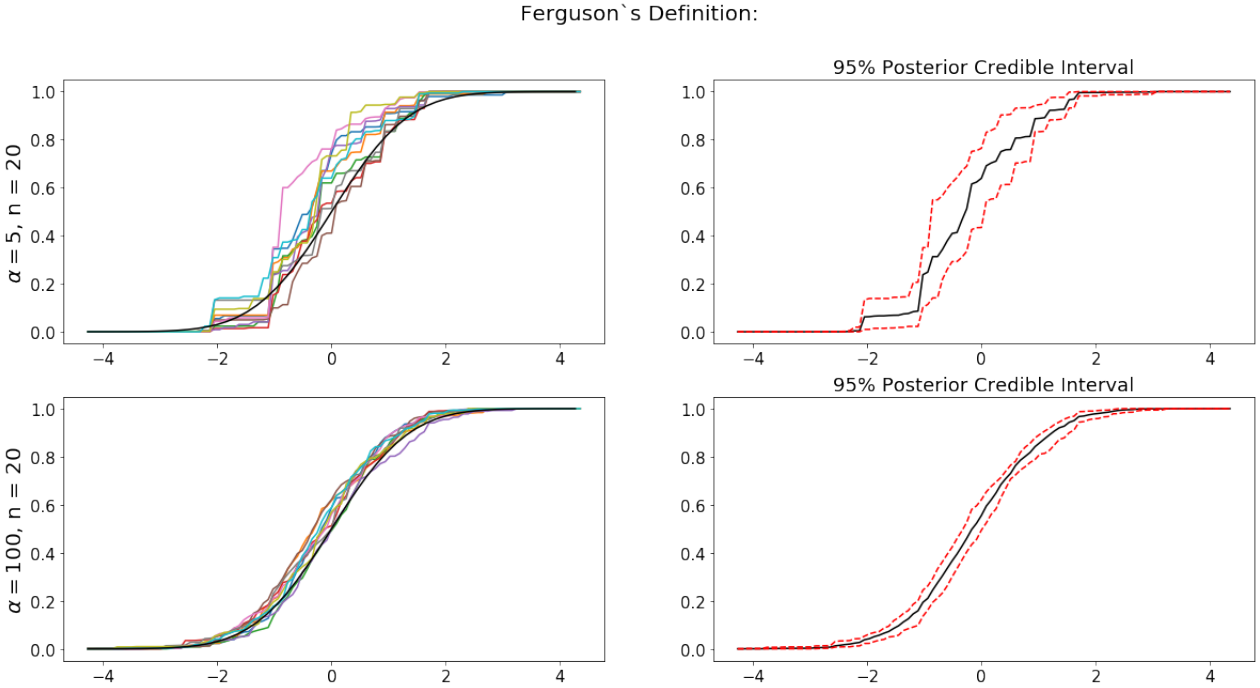

Figure 1: Ferguson's definition of the DP drawb from a N(0,1) distribution with a sample size of 20 and $\alpha = 5$ (top) and $\alpha = 100$ (bottom). The right is the corresponding 95% posterior credible interval.

[1]

[1] Code available at `https://github.com/msilva00/BNP_Homework/blob/master/HW2/HW2_Prob1.ipynb`.

From Figure 1, changing α changes the posterior distributions, because the dataset is too small to condition the posteriors to become more similar to the normal distribution. It's like if we are forcing the normal behaviour of the posterior by increasing our confidence in the normal prior and decrease our confidence in the data. We also see, larger α leads to a smaller 95% posterior credible interval. This is due to the decrease of the variance as α gets larger.

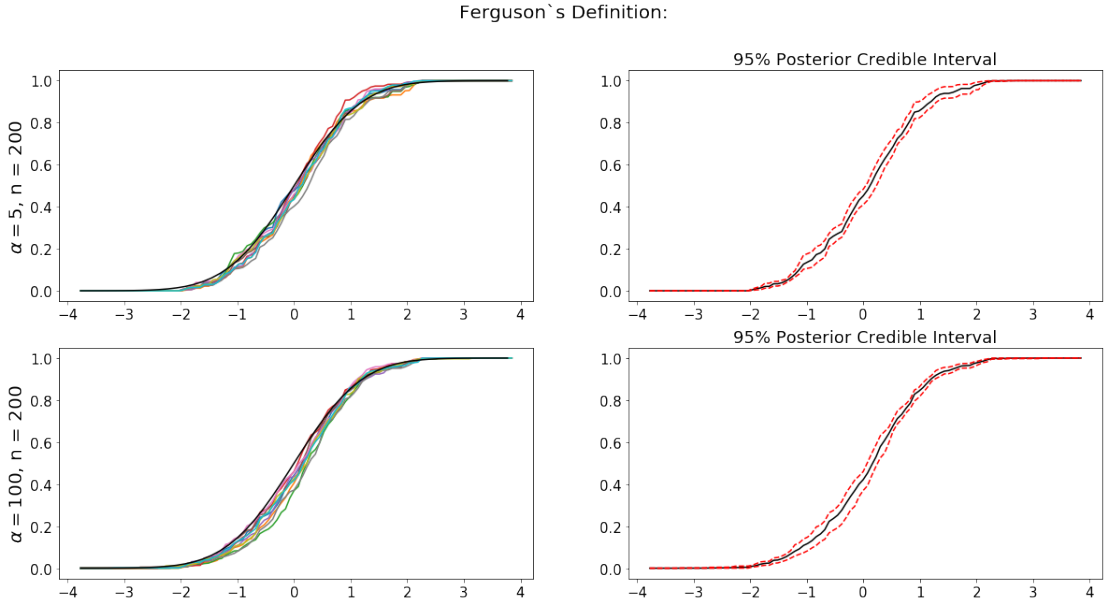

Figure 2: Ferguson's definition of the DP drawb from a N(0,1) distribution with a sample size of 200 and $\alpha = 5$ (top) and $\alpha = 100$ (bottom). The right is the corresponding 95% posterior credible interval.

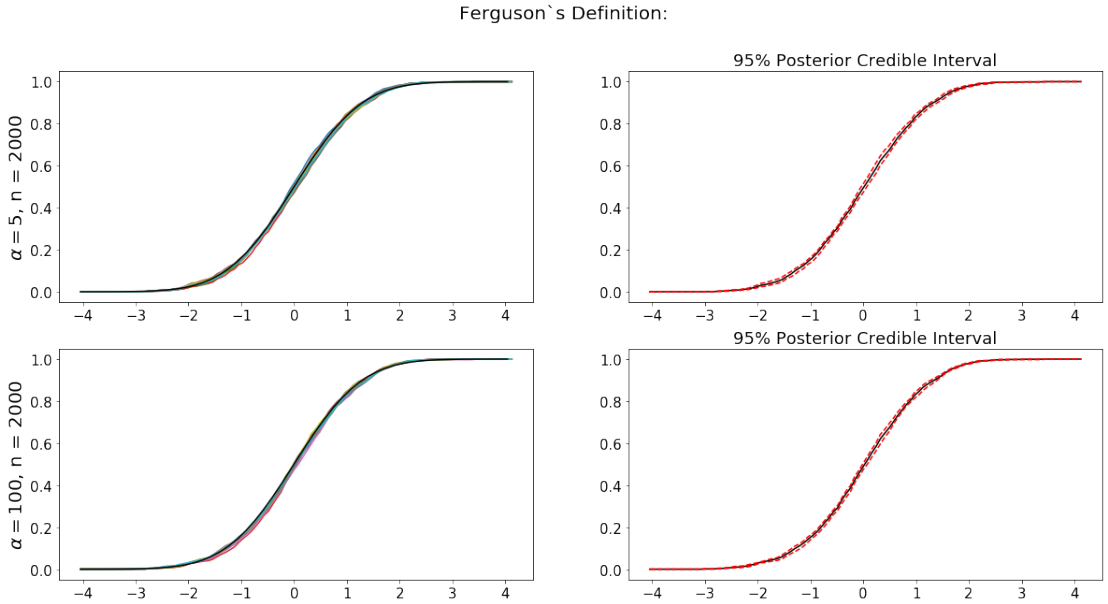

Figure 3: Ferguson's definition of the DP drawb from a N(0,1) distribution with a sample size of 2000 and $\alpha = 5$ (top) and $\alpha = 100$ (bottom). The right is the corresponding 95% posterior credible interval.

Based on Figures 2-3, the effect of α is negligible because the simulated data is already producing a normal behavior to the posterior distribution. Also, the credible intervals don't change too much as α grows.

Now, we will use data that comes from a finite mixture of normal distributions. We simulate the "true" distribution and plot the density using 20000 samples from the mixture and use it to compare the results.

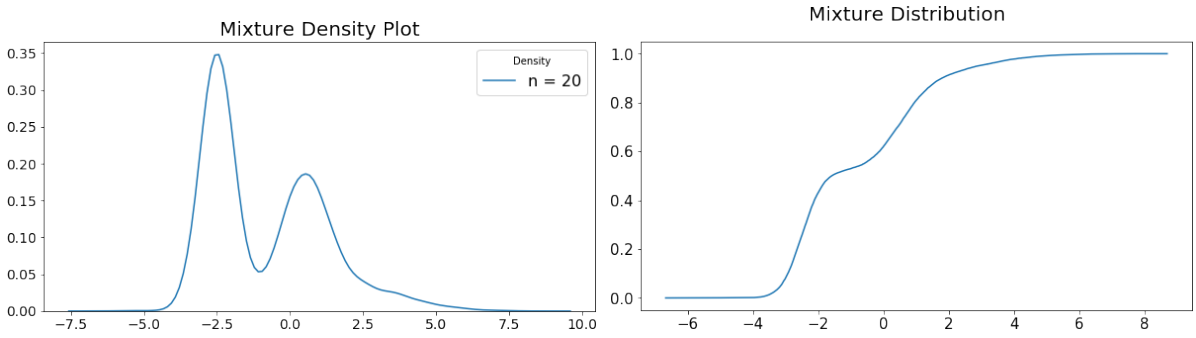

Figure 4: Density for 20000 samples from the mixture of normal.

Next, we simulate from the posterior distribution using Ferguson's definition of the DP for simulated data with sample sizes $n = 20, 200$, and 2000, respectively.

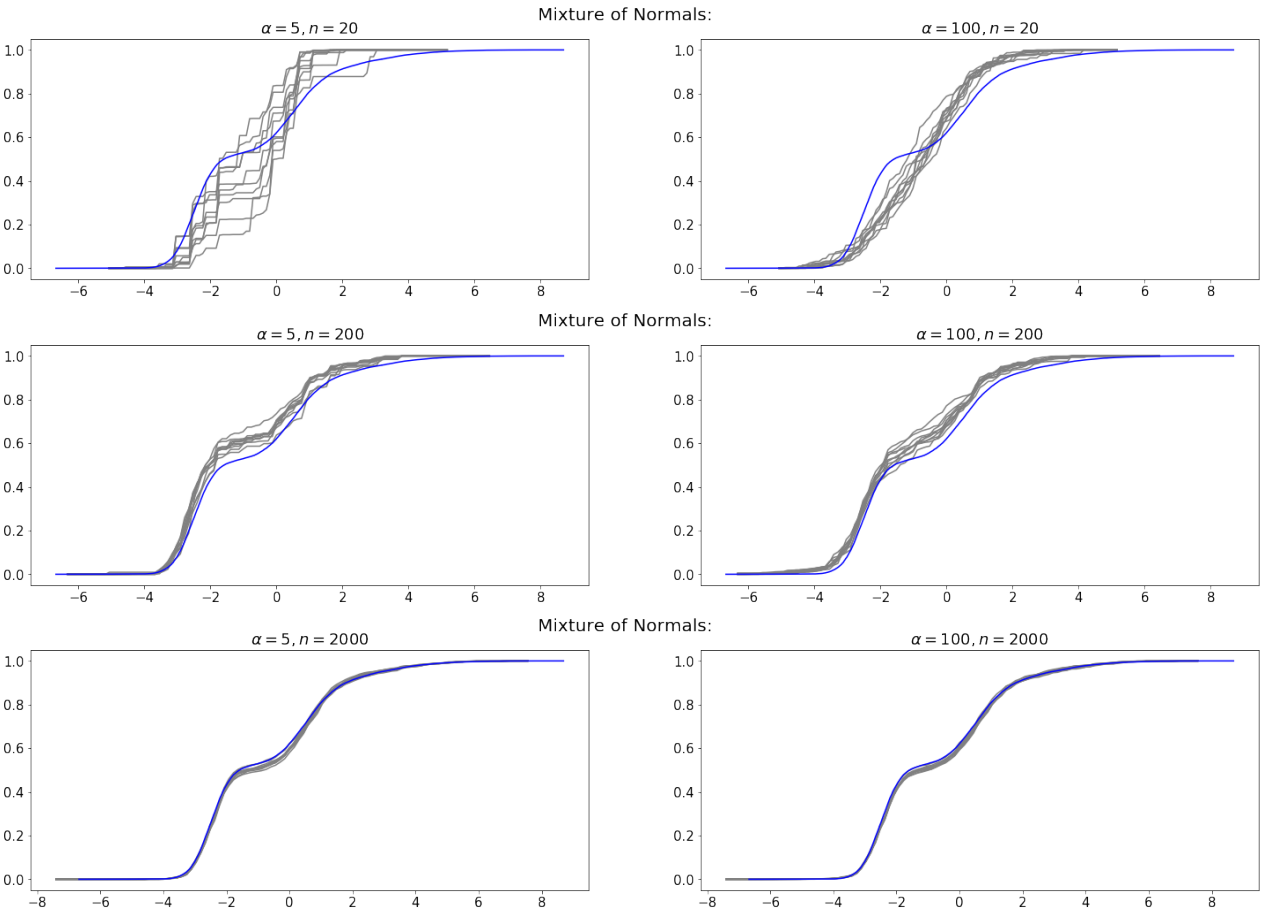

Figure 5: For each plot, we use sample datasets with $n = 20$(top), $n = 200$(middle), and $n = 2000$ (bottom) and use $\alpha = 5$(left) and $\alpha = 100$(right). The hyperparameters of the prior normal distribution are chosen based on the mean and sample standard deviation of each dataset. The grey lines represent posterior realizations from the DP, and the blue line represents the "true" mixture density (the 20000 samples from the mixture distribution).

2. Posterior inference for count data using MDP priors We consider again modeling a single distribution F, for count responses. The model for the data $\{y_1, ..., y_n\}$ is given by

$$y_i|F \overset{iid}{\sim} F;\ i = 1, ..., n$$
$$F|\alpha, \lambda \sim DP(\alpha, Pois(\lambda))$$

That is, we now have a DP prior for F, given random precision parameter α, and a random mean λ for the centering Poisson distribution. We assume independent gamma priors for α and λ. Again, we use simulated data under two different scenarios for the true data generating distribution:

(a) Poisson distribution with mean 5.

(b) Mixture of two Poisson distributions with means 3 and 11, and corresponding mixture weights given by 0.7 and 0.3.

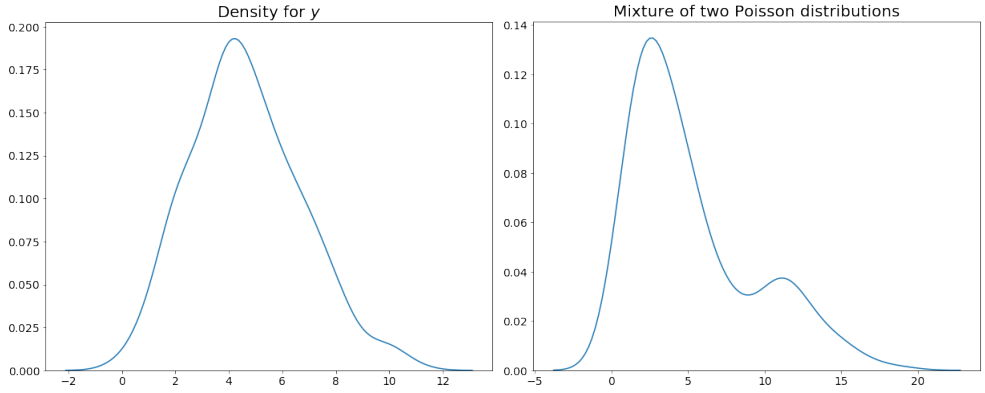

Figure 6: Simulated date for a Poisson distribution with mean 5 (left) and simulated data for a mixture of two Poisson distributions with means 3 and 11, and corresponding mixture weights given by 0.7 and 0.3.

Note, I will use the shape and rate parameterization in the prior Gamma distributions. In other words,

$$\alpha \sim Gamma(1, 0.1)$$
$$\lambda \sim Gamma(5, 1)$$

The shape 1 and rate 0.1 used in the Gamma prior for α, corresponds to a mean of 10 and variance of 200. This allows weight to be given to higher values of α while keeping α low, which is important for the case where the data is not unimodal. Using a unimodal Poisson centering distribution in the DP with low α will allow us to sample more from the empirical distribution of the data and less from the centering distribution. The shape 1.25 and rate 0.25, used in the Gamma prior for λ, corresponds to a mean of 5 and variance of 5. These are reasonable because the simulated data ranges between 0 and 30.

We can explore the posterior distributions of F, α and λ by Gibbs sampling algorithm. At each step, we update F by noting that the DP is a conjugate prior. F_{new} is updated following a $DP(\alpha_{current} + n, \widetilde{F}_0(\cdot))$, where

$$\widetilde{F}_0(\cdot) = \frac{\alpha}{\alpha + n} F(\cdot|\lambda) + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{y_i}(\cdot)$$

To update the hyper parameters α and λ, we can use the following to compute the marginal likelihood for DP priors with discrete baseline distributions

$$L(\alpha, \lambda; \text{data}) \propto \frac{\alpha^{n^*}}{\alpha^{(n)}} \prod_{j=1}^{n^*} f_0(y_j^*|\lambda) \left\{\alpha f_0(y_j^*|\lambda) + 1\right\}^{(n_j - 1)}$$

where $f_0(\cdot|\lambda)$ is the p.m.f of $F_0(\cdot|\lambda)$, n^* is the number of distinct values in $(y_1, ..., y_n)$, $\{y_j^* : j = 1, ..., n^*\}$ are the distinct values in $(y_1, ..., y_n)$, and $n_j = |\{i : y_i = y_j^*\}|$, for $j = 1, ..., n$. Using this marginalized likelihood, we can use a Metropolis within Gibbs step to update our parameters in blocks using a multivariate normal proposal distribution. We compute the acceptance ratio based on the ratio of the product of the likelihood and prior evaluated at the proposed candidate and the previous value for the hyper parameters.

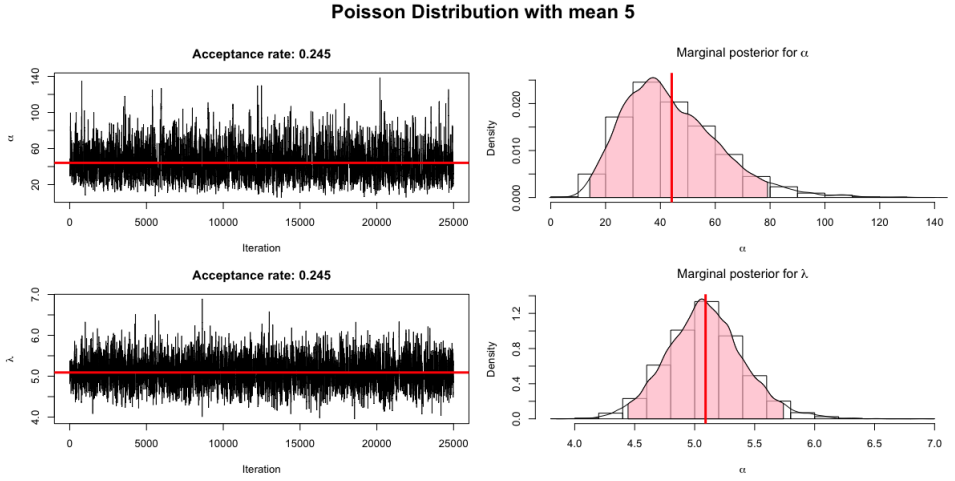

Figure 7: Left: Convergence trace plots of the marginal posterior distributions for α and λ with corresponding acceptance rates. Right: Kernel density estimates for the marginal posteriors, with mean (red line) and 95% HPD intervals, represented by the pink shaded region.

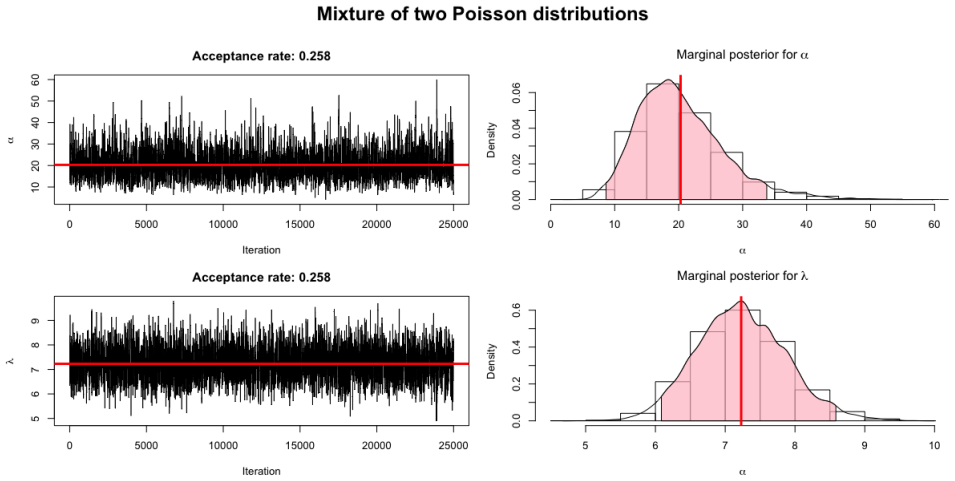

Figure 8: Left: Convergence trace plots of the marginal posterior distributions for α and λ with corresponding acceptance rates. Right: Kernel density estimates for the marginal posteriors, with mean (red line) and 95% HPD intervals, represented by the pink shaded region.

We obtain point estimates for the two underlying data generating probability mass functions through the posterior predictive distribution. To obtain a posterior distribution $F(y)|\text{data}$, we take our MCMC samples from the joint $p(\alpha, \lambda) \propto \pi(\alpha)\pi(\lambda)L(\alpha, \lambda; \text{data})$ In both cases, the DP recovered the underlying distribution.

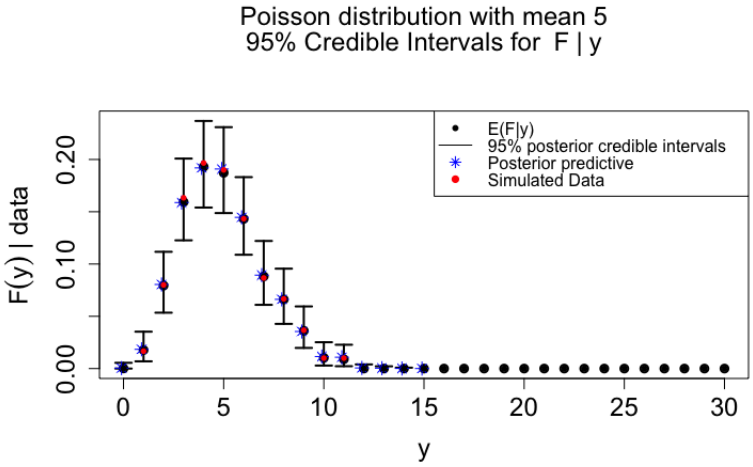

Figure 9: The result is that our posterior distribution very closely compares to the data.

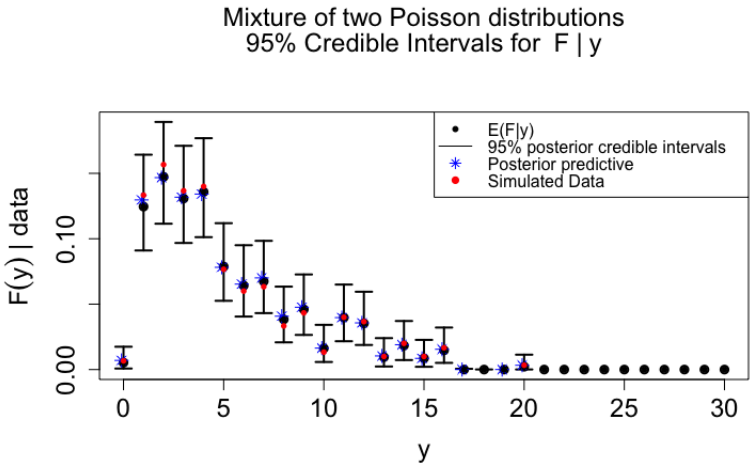

Figure 10: Again, the result is that our posterior distribution very closely compares to the data. As with problem 1, the variability can be reduced at each point by increasing n.