Your Name: Yaw Appaw **Problem Set 1 Due: Feb. 18, 2025**

(1) The Poisson distribution has probability mass function

$$p(y_i|\theta) = \frac{\theta^{y_i} e^{-\theta}}{y_i!}, \qquad \theta > 0, \qquad y_i = 0, 1, \ldots \tag{1}$$

and let $y_1, \ldots, y_n$ be random sample from this distribution.

1. Show that the gamma distribution $\mathcal{G}(\alpha, \beta)$ is a conjugate prior distribution for the Poisson distribution.
2. Show that $\bar{y}$ is the MLE for θ.
3. Write the mean of the posterior distribution as a weighted average of the mean of the prior distribution and the MLE.
4. What happens to the weight on the prior mean as n becomes large?

Solution

1. The Gamma prior is given by

$$\pi(\theta|\alpha, \beta) \propto \theta^{\alpha-1} e^{-\frac{\theta}{\beta}}. \tag{2}$$

   The likelihood function is given by

$$f(y|\theta) \propto \theta^{\sum y_i} e^{-\theta n} \tag{3}$$

   The posterior funtion can be found by multiplying the prior and the likelihood

$$\pi(\theta| \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}) \propto \theta^{\alpha-1+\sum y_i} e^{-\theta(\frac{1}{\beta}+n)} \tag{4}$$

1. To show that $\bar{y}$ is the MLE for θ, we must first define the following likelihood function

$$P(y_1|\theta) * P(y_2|\theta) * \ldots * P(y_n|\theta) \tag{5}$$

   Once we multiply the likelihood functions together we have the following

$$\propto \frac{\theta^{y_1} e^{-\theta}}{y_1!} * \frac{\theta^{y_2} e^{-\theta}}{y_2!} * \ldots * \frac{\theta^{y_n} e^{-\theta}}{y_n!} \tag{6}$$

We will then take the log of this expression

$$\propto \log \frac{\theta^{y_1} e^{-\theta}}{y_1!} * \log \frac{\theta^{y_2} e^{-\theta}}{y_2!} * \ldots * \log \frac{\theta^{y_n} e^{-\theta}}{y_n!} \tag{7}$$

As we are only concerned with maximizing in respect to theta we can then drop the numerator as it has nothing to do with θ

$$\propto \log \theta^{y_1} e^{-\theta} \cdot \log \theta^{y_2} e^{-\theta} \cdot \cdots \cdot \log \theta^{y_n} e^{-\theta} \tag{8}$$

By applying the log rule we have the following $\log A^B = B * \log A$

$$\propto y_1 \log \theta - \theta + y_2 \log \theta - \theta + \cdots + y_n \log \theta - \theta \tag{9}$$

After taking the first derivative we are left with the following

$$\propto \frac{y_1}{\theta} - 1 + \frac{y_2}{\theta} - 1 + \cdots + \frac{y_n}{\theta} - 1 \tag{10}$$

This can now be rewritten in $\sum$ form

$$\frac{\sum y_i}{\theta} - n \tag{11}$$

Which can be rewritten as

$$\frac{\sum y_i}{\theta} = n \tag{12}$$

Which can be rewritten as

$$\theta = \frac{\sum y_i}{n} \tag{13}$$

The above equation can be interpreted as follows: As the number of observations becomes increasingly large, $\bar{y}$ will converge to θ by the law of large numbers.

1. To show the mean of the posterior distribution as a weighted average of the mean of the prior distribution and the MLE we can use the posterior function as found in part 1.

$$\pi(\theta| \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}) \propto \theta^{\alpha - 1 + \sum y_i} e^{-\theta(\frac{1}{\beta} + n)} \tag{14}$$

As this follows the format for a Γ distribution we can define the following Γ distribution. Using the exponents from the above posterior function as the hyper parameters.

$$\propto \gamma(\alpha + \sum y_i, -\theta(\frac{1}{\beta} + n)) \tag{15}$$

The mean of a Γ distribution is defined as the first hyper-parameter over the 2nd hyper-parameter.Thus, the expected value of θ given y can be expressed as.

$$E(\theta|y) \propto \frac{\alpha + \sum y_i}{\frac{1}{\beta} + n} \tag{16}$$

The equation must now be rewritten to include the MLE proof.

$$E(\theta|y) \propto \frac{\alpha + n * \frac{\sum y_i}{n}}{\frac{1}{\beta} + n} \tag{17}$$

After expanding you will be left with

$$E(\theta|y) \propto \frac{\alpha}{\frac{1}{\beta} + n} + \frac{n}{\frac{1}{\beta} + n} * \frac{\sum y_i}{n} \tag{18}$$

The weight of the first hyper parameter can be simplified as

$$w = \frac{\alpha}{\frac{1}{\beta} + n} \tag{19}$$

The weight of the second hyper parameter is just 1-w

$$1 - w = \frac{n}{\frac{1}{\beta} + n} \tag{20}$$

The conditional probability of θ given y is now

$$E(\theta|y) \propto w * E(\theta) + (1 - w) * \hat{\theta} \tag{21}$$

(2) Consider the following two sets of data obtained after tossing a die 100 and 1000 times, respectively:

n	1	2	3	4	5	6
100	19	12	17	18	20	14
1000	190	120	170	180	200	140

Suppose you are interested in θ_1, the probability of obtaining a one spot. Assume your prior for all the probabilities is a Dirichlet distribution, where each $\alpha_i = 2$. Compute the posterior distribution for θ_1 for each of the sample sizes in the table. Plot the resulting distribution and compare the results. Comment on the effect of having a larger sample.

Solution

article pgfplots amsmath pgfplotstable tikz

Question 2

February 19, 2025

1 Posterior Calculation

We assume a Dirichlet prior with parameters $\alpha_i = 2$ for all die faces. Given the observed counts, the posterior distribution for θ_1 follows:

$$\theta_1 \sim \text{Beta}(\alpha_{\text{posterior}}, \beta_{\text{posterior}})$$

where

$$\alpha_{\text{posterior}} = \alpha_{\text{prior}} + x_1, \quad \beta_{\text{posterior}} = \sum \alpha_i^{\text{posterior}} - \alpha_{\text{posterior}}.$$

For the cases where $n = 100$ and $n = 1000$:

$$\theta_1^{(100)} \sim \text{Beta}(21, 81), \quad \theta_1^{(1000)} \sim \text{Beta}(192, 810).$$

2 Posterior Distributions

The following graph visualizes the posterior distributions for both sample sizes.

[width=12cm, height=8cm, xlabel=θ_1 (Probability of rolling a 1), ylabel=Density, legend pos=north east, domain=0:0.4, samples=100, grid=major] [color=blue, thick, dashed] expression[domain=0:0.4] gamma(21+81) / (gamma(21) * gamma(81)) * x$^(21-1) * (1-x)^(81-1)$; n = 100;
[color=red, thick] expression[domain=0:0.4] gamma(192+810) / (gamma(192) * gamma(810)) * x$^(192-1) * (1-x)^(810-1)$; n = 1000;

3 Conclusion

As expected, increasing the sample size results in a sharper posterior distribution, reducing uncertainty about the true probability of rolling a 1.