Martin Lavrinenko: **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

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

2. 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.

3. 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

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

Weight can be defined as

$$w = \frac{a}{a + n} \tag{22}$$

4. As the number of observations increases (n) the weight of the prior will continually decrease as n is in the denomerator. That is, as we observe more data the influence of the prior decreases and the influence of likelihood increases. Eventually, the likelihood will converge to match the posterior distribution.

(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

1. The posterior distribution for θ_1 can be found by first determining the prior

$$\propto \theta_1^{a_1-1} * \theta_2^{a_2-1} * \cdots * \theta_6^{a_6-1} \tag{23}$$

The likelihood function can be defined as

$$\propto \theta_1^{y_1} * \theta_2^{y_2} * \cdots * \theta_6^{y_6} \tag{24}$$

The posterior can be found by multiplying the exponents of the like θ terms together

$$\propto \theta_1^{y_1+a_1-1} * \theta_2^{y_2+a_2-1} * \cdots * \theta_6^{y_6+a_6-1} \tag{25}$$

Our hyper-parameters can now defined as

$$D(y_1 + a_1, y_2 + a_2, \ldots, y_6 + a_6) \tag{26}$$

Now we must substitute the observed value of each value for the sample size of n = 100. As stated in the initial problem α_i is equal to 2.

$$D(19 + 2, 12 + 2, 17 + 2, 18 + 2, 20 + 2, 14 + 2) \tag{27}$$

Therefore the Dirichlet distribution for n=100 is

$$D(21, 14, 19, 20, 22, 16) \tag{28}$$

Using the following property we can extract the marginal distribution for θ_1

$$(\theta_1, \theta_2, \ldots, \theta_6) \propto D(\alpha_1, \alpha_2, \ldots, \alpha 6) \tag{29}$$

The marginal distribution of θ_1 is

$$\theta_1 \propto \beta(\alpha_1, S) \tag{30}$$

S is defined as

$$S = \sum_{i \neq 1} \alpha_i \tag{31}$$

Therefore, the beta distribution of θ_1 when n=100 is

$$\theta_1 \propto \beta(21, 91) \tag{32}$$

The same can also be done for n=1,000

$$D(190 + 2, 120 + 2, 170 + 2, 180 + 2, 200 + 2, 140 + 2) \tag{33}$$

Therefore the Dirichlet distribution for n=1,000 is

$$D(192, 122, 172, 182, 202, 142) \tag{34}$$

Using the same steps for n=100 we can define the following beta distribution for n=1,000

$$\theta_1 \propto \beta(192, 820) \tag{35}$$

Visual depiction of the distribution of θ_1 in respect to n=100 and 1000 below

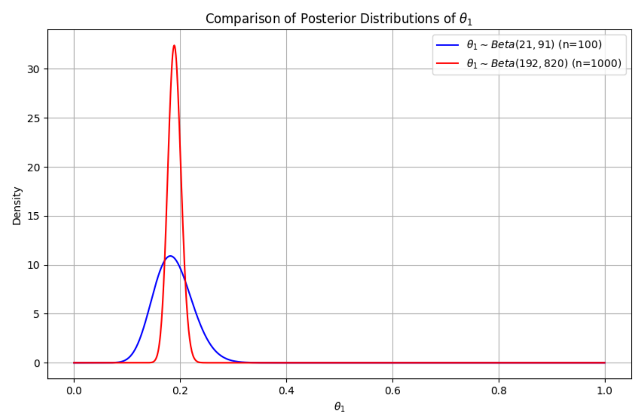

Figure 1: Caption

As the sample size increases, the density of θ_1 becomes more concentrated around 0.19.