STAMATICS
Mini Project 2

Tejesh Vaish (190908)

June 6, 2021

1 Problem Statement

We observe multinomial data with parameters n, $\mathbf{x}$ and $\mathbf{p}$ (K dimensional) such that

$$\mathbf{x} \ = \ (x_1, \ldots, x_K) \ \sim Multi(p_1, \ldots, p_K), \qquad x_i \in \{0, \ldots, n\} \ and \ \sum_{i=1}^{n} x_i = n$$

$$Pr(\mathbf{x} = (x_1, \ldots, x_K) \mid \mathbf{p}) = \frac{n!}{x_1! \ldots x_K!} \prod_{i=1}^{K} p_i^{x_i}, \qquad \sum_{i=1}^{K} p_i = 1$$

We are also given the MLE of p_i as

$$\hat{p}_i = \frac{x_i}{n}$$

Now, we have to estimate $\mathbf{p}$ using Bayesian method taking Dirichlet as the prior distribution of $\mathbf{p}$ with $\alpha_i > 0$ as parameters.

$$Prior Distribution : \mathbf{p} \ = \ (p_1, \ldots, p_K) \ \sim Dir(\alpha_1, \ldots, \alpha_K), \qquad p_i \in (0,1) \ and \ \sum_{i=1}^{n} p_i = 1$$

$$f(\mathbf{p} = (p_1, \ldots, p_K) \mid \alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} p_i^{\alpha_i - 1}, \qquad (\alpha_i > 0)$$

(Here, f is the probability density function.)

2 Posterior Distribution of p

We need to calculate $f(\mathbf{p}|\mathbf{x})$ (the posterior distribution of $\mathbf{p}$). By applying Bayes theorem to probability distribution function, we know

$$f(\mathbf{p}|\mathbf{x}) = \frac{f(\mathbf{x}|\mathbf{p}) \cdot f(\mathbf{p})}{f(\mathbf{x})}$$

Here, $f(\mathbf{x})$ is the normalising constant and $f(\mathbf{x}|\mathbf{p})$ is proportional to the Likelihood function $Pr(\mathbf{x}|\mathbf{p})$ which gives us the following proportionality relation:

$$\begin{aligned} f(\mathbf{p}|\mathbf{x}) &\propto Pr(\mathbf{x}|\mathbf{p}) \cdot f(\mathbf{p}) \\ &\propto \Big(\prod_{i=1}^{K} p_i^{x_i}\Big)\Big(\prod_{i=1}^{K} p_i^{\alpha_i - 1}\Big) \\ &\propto \Big(\prod_{i=1}^{K} p_i^{x_i+\alpha_i-1}\Big) \end{aligned}$$

The above expression is that of Dirichlet distribution, so we get the posterior distribution as

$$Posterior Distribution: \ \mathbf{p}|\mathbf{x} \ = \ (p_1^{'}, \ldots, p_K^{'}) \ \sim Dir(\alpha_1 + x_1, \ldots, \alpha_K + x_K)$$

Thus, posterior distribution of $\mathbf{p}$ is also Dirichlet with updated parameters, which are updated according to data available.

3 Posterior Mean of p

Let us first calculate the prior mean of $\mathbf{p}$ which is given by

$$E[\mathbf{p}] = \int \mathbf{p} \cdot f(\mathbf{p}) d\mathbf{p}$$

Since f is a probability density function, we know that $\int f(\mathbf{p})d\mathbf{p} = 1$ for entire space of $\mathbf{p}$. Also, because of the constraint that $\sum_{i=1}^{K} p_i = 1$, $\mathbf{p}$ will be integrated for only $K-1$ dimensions (as K^{th} dimension is dependent).
Let $\sum_{i=1}^{K} \alpha_i = m$, so $E[p_i]$ will be given by

$$E[p_i] = \int \cdots \int p_i \cdot \frac{\Gamma(m)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} p_i^{\alpha_i - 1} \ dp_1 \ldots dp_{K-1}$$

Let us change the parameters as $\alpha_j^{'} = \alpha_j$ for $j \neq i$ and $\alpha_j^{'} = \alpha_j + 1$ for $j = i$. This will give $m^{'} = m + 1$. Putting these values in above equation will give

$$E[p_i] = \frac{\Gamma(m)}{\Gamma(m')} \cdot \frac{\Gamma(a_i + 1)}{\Gamma(a_i)} \int \cdots \int \frac{\Gamma(m^{'})}{\prod_{i=1}^{K} \Gamma(\alpha_i^{'})} \prod_{i=1}^{K} p_i^{\alpha_i^{'} - 1} \ dp_1 \ldots dp_{K-1}$$

The integral is integrating f with updated parameters so it will still give 1 and prior mean will be

$$E[p_i] = \frac{\alpha_i}{m}$$

As the posterior distribution differs from prior distribution only in terms of parameters α_i, so posterior mean will be given by

$$
\begin{aligned}
E[p_i] &= \frac{\alpha_i + x_i}{\sum_{i=1}^{k}(\alpha_i + x_i)} \\
\implies E[p_i] &= \frac{\alpha_i + x_i}{\sum_{i=1}^{k} x_i + \sum_{i=1}^{k} \alpha_i} \\
\implies E[p_i] &= \frac{\alpha_i + x_i}{n + m}
\end{aligned}
$$

We can represent the posterior mean as a convex combination of prior mean and MLE of p_i ($= \frac{x_i}{n}$) as follows :

$$
\begin{aligned}
E[p_i] &= \frac{\alpha_i}{n+m} + \frac{x_i}{n+m} \\
\implies E[p_i] &= \frac{m}{n+m} \cdot (\frac{\alpha_i}{m}) + \frac{n}{m+n} \cdot (\frac{x_i}{n}) \\
\implies E[p_i] &= \beta \cdot (\frac{\alpha_i}{m}) + (1-\beta) \cdot (\frac{x_i}{n}) \qquad\qquad (\beta > 0)
\end{aligned}
$$

The posterior mean is a weighted average between the prior mean and the data mean, so as n increases, posterior mean comes closer to MLE of p_i given by data.

4 IMDB Rating System

We have to prove that the rating used by IMDB can be derived from the model used above.

$$
Rating = \frac{n}{n+m} R + \frac{m}{n+m} C
$$

Denote the prior probability parameters as $\mathbf{p}$ and posterior probability parameters as $\mathbf{p}$'. The Dirichlet parameters are α_i ($i \in \{1 \dots 10\}$) and there sum as m. The number of voters giving rating i to a particular movie are x_i and their sum(i.e total votes for a movie) is n. To get the *Rating*, we use the posterior mean $\mathbf{p}$' as follows:

$$
\begin{aligned}
Rating &= \sum_{i=1}^{10} p_i^{'} \cdot i \\
&= \sum_{i=1}^{10} \left(\frac{\alpha_i + x_i}{n+m}\right) \cdot i \\
&= \frac{1}{n+m} \sum_{i=1}^{10} (x_i \cdot i) + \frac{1}{n+m} \sum_{i=1}^{10} (\alpha_i \cdot i) \\
&= \frac{n}{n+m} \sum_{i=1}^{10} \frac{x_i}{n} \cdot i + \frac{m}{n+m} \sum_{i=1}^{10} \frac{\alpha_i}{m} \cdot i
\end{aligned}
$$

As given, R is the average rating of the movie based on votes, so we know that

$$R = \sum_{i=1}^{10} \frac{x_i}{n} \cdot i$$

By looking at the rating formula, we can conclude that C is average prior rating given by

$$C = \sum_{i=1}^{10} \frac{\alpha_i}{m} \cdot i$$

Putting the given data ($C = 5.5$, $m = 2500$), we can say that α_i follow the above linear relation and final rating is given by

$$Rating = \frac{n}{n+m} R + \frac{m}{n+m} C$$

Using the above formula for sorting the movies gives us the following movies as "Top 10" :

IMDB ID
(1) tt5074352
(2) tt8108198
(3) tt8291224
(4) tt1954470
(5) tt4430212
(6) tt3322420
(7) tt2356180
(8) tt0073707
(9) tt2283748
(10) tt2338151

5 References

[1] http://www.mas.ncl.ac.uk/ nlf8/teaching/mas2317/notes/chapter2.pdf

[2] http://www.mas.ncl.ac.uk/ nmf16/teaching/mas3301/week6.pdf

[3] https://dvats.github.io/assets/course/mth511/notes/W12L26_notes.pdf

[4] http://users.cecs.anu.edu.au/ ssanner/MLSS2010/Johnson1.pdf