Polya Distribution

1 Introduction

Assume,
$n_i = n(x_i)$
$n_k(x_i) = n_{ik}$

$$p(x_1, x_2, ..., x_m \mid \alpha) = \prod_{i=1}^{m} \left(\frac{n_i!}{\prod_k n_{ik!}} \frac{\Gamma(\sum_k \alpha_k)}{\Gamma(n_i + \sum_k \alpha_k)} \prod_k \frac{\Gamma(n_{ik} + \alpha_k)}{\Gamma(\alpha_k)} \right) \tag{1}$$

The log likelihood is:

$$log\ p(x_1, x_2, ..., x_m \mid \alpha) \tag{2}$$
$$= \sum_{i=1}^{m} log \left(\frac{n_i!}{\prod_k n_{ik}!} \frac{\Gamma(\sum_k \alpha_k)}{\Gamma(n_i + \sum_k \alpha_k)} \prod_k \frac{\Gamma(n_{ik} + \alpha_k)}{\Gamma(\alpha_k)} \right) \tag{3}$$
$$= \sum_{i=1}^{m} \left(log(n_i!) - \sum_k log(n_{ik}!)! + log(\Gamma(\sum_k \alpha_k)) - log(\Gamma(n_i + \sum_k \alpha_k)) + \sum_k log(\Gamma(n_{ik} + \alpha_k)) - \sum_k log(\Gamma(\alpha_k)) \right) \tag{4}$$

Now, we differentiate in terms of α_k:

$$\frac{d\ log\ p(D \mid \alpha)}{d\ \alpha_k} = \sum_{i=1}^{m} \left(\psi(\sum_k \alpha_k) - \psi(n_i + \sum_k \alpha_k) + \psi(n_{ik} + \alpha_k) - \psi(\alpha_k) \right) \tag{5}$$
$$\frac{d^2\ log\ p(D \mid \alpha)}{d\ \alpha_k^2} = \sum_{i=1}^{m} \left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) + \psi'(n_{ik} + \alpha_k) - \psi'(\alpha_k) \right) \tag{6}$$
$$\frac{d^2\ log\ p(D \mid \alpha)}{d\ \alpha_k\ d\ \alpha_j} = \sum_{i=1}^{m} \left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) \right) \tag{7}$$

Now, for beta binomial case (k = 2), we have two parameters α_1 and α_2. Hence the FIM is:

$$FIM = -E \begin{bmatrix} \frac{d^2\ log\ p(D \mid \alpha)}{d\ \alpha_1^2} & \frac{d^2\ log\ p(D \mid \alpha)}{d\ \alpha_1\ d\ \alpha_2} \\ \frac{d^2\ log\ p(D \mid \alpha)}{d\ \alpha_2\ d\ \alpha_1} & \frac{d^2\ log\ p(D \mid \alpha)}{d\ \alpha_2^2} \end{bmatrix} \tag{8}$$

$$\begin{bmatrix} \sum_{i=1}^{m} \left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) + \psi'(n_{i1} + \alpha_1) - \psi'(\alpha_1) \right) & \sum_{i=1}^{m} \left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) \right) \\ \sum_{i=1}^{m} \left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) \right) & \sum_{i=1}^{m} \left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) + \psi'(n_{i2} + \alpha_2) - \psi'(\alpha_2) \right) \end{bmatrix} \tag{9}$$

$$FIM_{11} = -E\left[\sum_{i=1}^{m}\left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) + \psi'(n_{i1} + \alpha_1) - \psi'(\alpha_1)\right)\right] \tag{10}$$

$$= -\sum_{i=1}^{m}\left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) + E\left[\psi'(n_{i1} + \alpha_1)\right] - \psi'(\alpha_1)\right) \tag{11}$$

$$= -m * \left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) + E\left[\psi'(n_{i1} + \alpha_1)\right] - \psi'(\alpha_1)\right) \tag{12}$$

$$FIM_{12} = FIM_{21} = -E\left[\sum_{i=1}^{m}\left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k)\right)\right] \tag{13}$$

$$= -\sum_{i=1}^{m} E\left[\left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k)\right)\right] \tag{14}$$

$$= -m * \left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k)\right) \tag{15}$$

$$FIM_{22} = -E\left[\sum_{i=1}^{m}\left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) + \psi'(n_{i2} + \alpha_2) - \psi'(\alpha_2)\right)\right] \tag{16}$$

$$= -\sum_{i=1}^{m}\left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) + E\left[\psi'(n_{i2} + \alpha_2)\right] - \psi'(\alpha_2)\right) \tag{17}$$

$$= -m * \left(\psi'(\sum_k \alpha_k) - \psi'(n_i + \sum_k \alpha_k) + E\left[\psi'(n_{i2} + \alpha_2)\right] - \psi'(\alpha_2)\right) \tag{18}$$

2 Maximum Likelihood Estimation

To find the MLE of the parameters we start by taking log likelihood of the equation (1).

$$log\ p(x_1, x_2, ..., x_m \mid \alpha) \tag{19}$$

$$= \sum_{i=1}^{m}\left(log(n_i!) - \sum_k log(n_{ik}!)! + log(\Gamma(\sum_k \alpha_k)) - log(\Gamma(n_i + \sum_k \alpha_k)) + \sum_k log(\Gamma(n_{ik} + \alpha_k)) - \sum_k log(\Gamma(\alpha_k))\right) \tag{20}$$

Then from this α_k can be found using iterative method. One method suggested in [Minkas 2012] is using fixed point iteration. The idea is to guess an initial α_k, find a function that bounds F from below which is tight at α_k , then optimize this function to arrive at α_k^{new} which converges the function.
In his paper [Minkas 2012] come up with the final fixed point iteration using the following bounds. First equation (20) can be bounded using the following bounds:

$$log\Gamma(z) - log\Gamma(z+n) \geq log\Gamma(\hat{z}) - log\Gamma(\hat{z}+n) + [\Psi(\hat{z}+n) - \Psi(\hat{z})](\hat{z} - z) \tag{21}$$

$$log\Gamma(z+n) - log\Gamma(z) \geq log\Gamma(\hat{z}+n) - log\Gamma(\hat{z}) + \hat{z}[\Psi(\hat{z}+n) - \Psi(\hat{z})](logz - log\hat{z}) \tag{22}$$

$$\tag{23}$$

[Guo and Qi 1976]
Then substituting equation (21) and (22) in equation (20) simplified and differentiating with α_k gives.

$$\frac{d\ log\ p(D \mid \alpha)}{d\ \alpha_k} = \sum_{i=1}^{m}\left(\frac{\alpha_k\psi(\sum_k \alpha_k) - \psi(n_i + \sum_k \alpha_k) + \psi(n_{ik} + \alpha_k) - \psi(\alpha_k)}{\alpha_k^{new}}\right) \tag{24}$$

$$\tag{25}$$

Finally, equation (25) can be set to zero to solve α_k^{new}

$$\alpha_k^{new} = \alpha_k \frac{\sum_{i=1}^{m} \Psi(n_{ik} + \alpha_k) - \Psi(\alpha_k)}{\sum_i \Psi(n_i + \sum_k \alpha_k) - \Psi(\sum_k \alpha_k)} \tag{26}$$

$$\tag{27}$$

We can also simplify equation (27) using the following gamma simplifications.

$$\Psi(n+x) - \Psi(x) = \frac{d}{dx}(log\frac{\Gamma(n+x)}{\Gamma(x)}) \tag{28}$$

$$= \frac{d}{dx}(\sum_{i=0}^{n-1}(log(x+i)) \tag{29}$$

$$= \sum_{i=0}^{n-1} \frac{1}{(x+i)} \tag{30}$$

$$\tag{31}$$

Then using the above simplification equation (27) can be reduced to:

$$\alpha_k^{new} = \alpha_k \frac{\sum_{i=1}^{m} \sum_{j=0}^{(n_{jk}-1)} \frac{1}{\alpha_k + j}}{\sum_{i=1}^{m} \sum_{j=0}^{(n_j-1)} \frac{1}{\sum_k \alpha_k + j}} \tag{32}$$

$$\tag{33}$$

3 Method of Moments

We know for Beta-Binomial:
$E[X] = \frac{n\alpha}{\alpha+\beta}$
$E[X^2] = \frac{n\alpha(n+n\alpha+\beta)}{(\alpha+\beta)(1+\alpha+\beta)}$
Now, the first order moment from the data:
$m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$
The second order moment:
$m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$
Equating first and second order moments with sample moments:

$$m_1 = \frac{n\alpha}{\alpha+\beta} \tag{34}$$

$$m_2 = \frac{n\alpha(n+n\alpha+\beta)}{(\alpha+\beta)(1+\alpha+\beta)} \tag{35}$$

From 34 we have:

$$\beta = \frac{\alpha(n-m_1)}{m_1} \tag{36}$$

Dividing 35 by 34 we have:

$$\frac{m_2}{m_1} = \frac{n+n\alpha+\beta}{(1+\alpha+\beta)} \tag{37}$$

Replacing β in 37 by 36 we have:

$$\frac{m_2}{m_1} = \frac{nm_1 + nm_1\alpha + n\alpha - m_1\alpha}{m_1 + m_1\alpha + n\alpha - \alpha m_1} \tag{38}$$

Solving for α from 38:

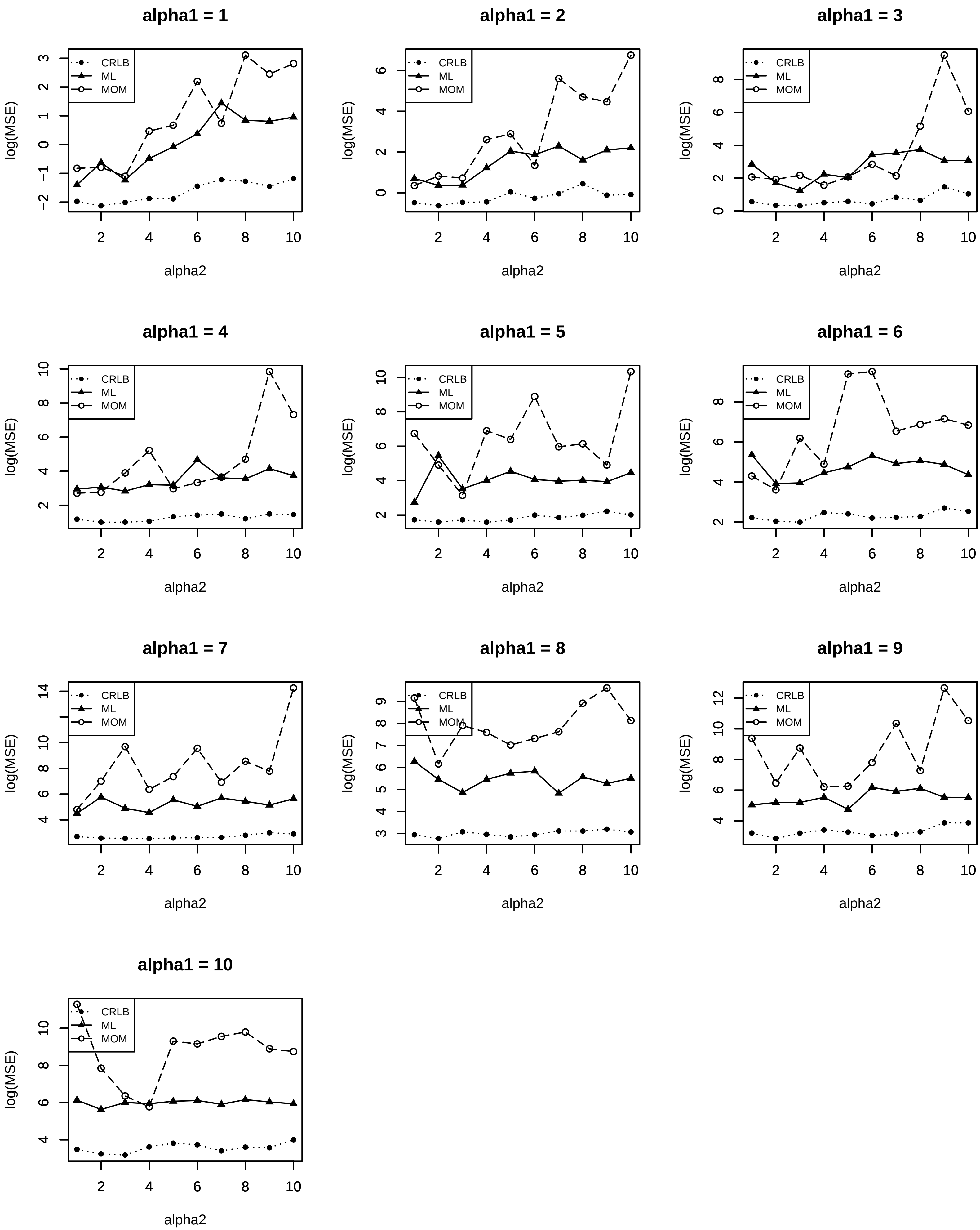
$$\alpha = \frac{nm_1 - m_2}{n(\frac{m_2}{m_1} - m_1 - 1) + m_1} \tag{39}$$

Putting value of α into 36 we get:

$$\beta = \frac{(n - m_1)(n - \frac{m_2}{m_1})}{n(\frac{m_2}{m_1} - m_1 - 1) + m_1} \tag{40}$$

4 Result

The graphs compare the log MSE of the three methods discussed above. In our experiment, we took 10 by 10 grid to generate the data sample from Beta binomial distribution. For each generated data we tried to estimate the parameter for the beta binomial. These plot compares MSE of CRLB, ML and MOM for a given α_1 to the range of α_2. The x-axis shows the range of α_2 corresponding to the MSE on the y-axis. The result shows CRLB has the smallest MSE value compared to all estimation methods for all experiments. Our empirical result agrees to the analytical explanation of CRLB, which claims CRLB is the lower bound for variance of the estimator. Furthermore, the lower bound variance for the examples generated from larger α_k parameters tends to have bigger value compared to data generated from smaller parameters.
In the other hand, MLE achieves lower MSE than MOM for most examples generated. However, in few experiments there exist a situation where MOM beats MLE. For example, we can see some points in the plots of α_1 less than 6 where MOM has smaller MSE value than MLE. Analytically, we expect MLE to outperform MOM for large sample size as MLE is asymptotically efficient. However, in the smaller sample size dataset there could be a situation where MOM could beat MLE . The result of some plot in our experiment shows MOM could also achieve better estimation than MLE. This is because our experiments is based on small sample size (i.e. n=20 document size).


alpha1 = 1
alpha1 = 2
alpha1 = 3
alpha1 = 4
alpha1 = 5
alpha1 = 6
alpha1 = 7
alpha1 = 8
alpha1 = 9
alpha1 = 10
CRLB
ML
MOM
log(MSE)
alpha2