BAYESIAN INFERENCE

Amanpreet Singh Saini
2023MAS1001

Department of Mathematics
IIT ROPAR

Supervisor - **Dr. Deepak Singh**
November 28, 2024

Contents

What is Binomial Distribution?

Binomial Distribution

The Binomial distribution models the number of successes in a fixed number of independent trials, where each trial has two possible outcomes (success or failure), and the probability of success is the same across all trials.

- **Parameters:**
  - n: Number of trials
  - p: Probability of success in a single trial
- **Probability Mass Function (PMF):**

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1}$$

where k is the number of successes, n is the number of trials, and p is the probability of success.

What is Beta Distribution?

Beta Distribution

The Beta distribution is a continuous probability distribution that can take values [0,1] and is often used to model the probability of success in a random process where the probability itself is unknown and needs to be inferred.

- **Parameters:**
  - α: Shape parameter
  - β: Shape parameter
- **Probability Density Function (PDF):**

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \tag{2}$$

where $B(\alpha,\beta)$ is the Beta function.

What is Beta-Binomial Distribution?

Beta-Binomial Distribution

The Beta-Binomial distribution is a discrete probability distribution that models the number of successes in a fixed number of trials, where the probability of success varies between trials according to a Beta distribution.

Limitation in Binomial : Success probability is constant across all trials. But in Beta-Binomial , it follows Beta distribution

Suppose Y_i counts the number of "successes" in n_i independent trials for individual i, with success probabilities θ_i , $i = 1, 2, \dots k$.

If the probabilities vary according to a Beta distribution, then $Y_1, Y_2, \dots, Y_k$ are described by a BBM:

$$Y_i \mid \theta_i \overset{\text{indep}}{\sim} \text{Bin}\left(n_i, \theta_i\right) \quad \text{and} \quad \theta_i \mid \alpha, \beta \overset{\text{iid}}{\sim} \text{Beta}(\alpha, \beta).$$

Probability Density Function

Probability Mass function of Y_i is given as:

$$\Pr(Y_i = y_i|\theta_i) = \binom{n_i}{y_i} \theta_i^{y_i} (1-\theta_i)^{n_i - y_i} \tag{3}$$

And as $\theta_i \sim$ Beta(α,β)
$\therefore$ Probability Density function for θ_i is given as:

$$f(\theta_i|\alpha,\beta) = \frac{\theta_i^{\alpha-1}(1-\theta_i)^{\beta-1}}{B(\alpha,\beta)} \quad \text{where} \quad B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{4}$$

To find the marginal probability of $Y_i = y_i$,we integrate out θ_i

$$\Pr(Y_i = y_i|\alpha,\beta) = \int_0^1 \Pr(Y_i = y_i|\theta_i) f(\theta_i|\alpha,\beta)\, d\theta_i. \tag{5}$$

Substituting the Binomial and Beta distributions into this integral:

$$\Pr(Y_i = y_i|\alpha,\beta) = \int_0^1 \binom{n_i}{y_i} \theta_i^{y_i} (1-\theta_i)^{n_i - y_i} \frac{\theta_i^{\alpha-1}(1-\theta_i)^{\beta-1}}{B(\alpha,\beta)}\, d\theta_i. \tag{6}$$

$$\Pr(Y_i = y_i|\alpha, \beta) = \binom{n_i}{y_i} \frac{1}{B(\alpha, \beta)} \int_0^1 \theta_i^{y_i+\alpha-1} (1-\theta_i)^{n_i-y_i+\beta-1} \, d\theta_i. \tag{7}$$

The integral now has the form of a Beta function $B\left(y_i + \alpha, n_i - y_i + \beta\right)$, replacing it

$$\Pr(Y_i = y_i|\alpha, \beta) = \binom{n_i}{y_i} \frac{B(y_i + \alpha, n_i - y_i + \beta)}{B(\alpha, \beta)}. \tag{8}$$

Finally, expressing the Beta functions in terms of Gamma functions gives us:

$$\Pr(Y_i = y_i|\alpha, \beta) = \binom{n_i}{y_i} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+y_i)\Gamma(\beta+n_i-y_i)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n_i)}. \tag{9}$$

This expression represents Probability density function for Beta-Binomial distribution.

Likelihood function

The marginal likelihood function $L(\alpha, \beta)$ for the Beta-Binomial Model (BBM) when all sample sizes n_i = n.

Given the vector $Y = (Y_1, Y_2, \ldots, Y_k)$ of observed counts , where each Y_i represents the number of successes in n trials for each individual, the likelihood function for parameters α and β is:

$$L(\alpha, \beta) = \Pr(\mathbf{Y}|\alpha, \beta) = \prod_{i=1}^{k} \Pr(Y_i = y_i|\alpha, \beta) \tag{10}$$

By using equation (9) ,

$$L(\alpha, \beta) = \prod_{i=1}^{k} \left(\binom{n}{y_i} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+y_i)\Gamma(\beta+n-y_i)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)} \right) \tag{11}$$

Since each term $\binom{n}{y_i}$ is constant with respect to α and β, we can focus on the product of the remaining terms:

$$L(\alpha, \beta) = \left(\prod_{i=1}^{k} \binom{n}{y_i} \right) \frac{\Gamma(\alpha+\beta)^k}{\Gamma(\alpha)^k \Gamma(\beta)^k \Gamma(\alpha+\beta+n)^k} \prod_{i=1}^{k} \left(\Gamma(\alpha+y_i)\Gamma(\beta+n-y_i) \right) \tag{12}$$

The first product term $\prod_{i=1}^{k} \binom{n}{y_i}$ is simply a constant with respect to α and β , so the main expression is:

$$L(\alpha, \beta) = \prod_{i=1}^{k} \frac{\Gamma(\alpha+y_i)\Gamma(\beta+n-y_i)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)} \tag{13}$$

This is the marginal likelihood function for the parameters α and β when all trials have the same size n.

Polynomial Expansion

Now we approximate the likelihood function for the Beta-Binomial Model (BBM) by decomposing it into polynomial factors as

$$L(\alpha, \beta) = R_1(\alpha) R_2(\beta) R_3(\alpha + \beta) \tag{14}$$

Definitions and Setup

k_j: The number of y_i's equal to j $\implies \sum_{j=0}^{n} k_j = k$.
s_i : The cumulative count of observations with at least i successes.

$$s_i = k_i + k_{i+1} + \cdots + k_n$$

for example

$$\begin{aligned}
s_0 &= k_0 + k_1 + k_2 + \cdots + k_n, \\
s_1 &= k_1 + k_2 + \cdots + k_n, \\
&\vdots \\
s_n &= k_n.
\end{aligned}$$

adding all the s_i values,

$$
\begin{aligned}
s &= \sum_{i=0}^{n} s_i \\
&= (k_0 + k_1 + k_2 + \cdots + k_n) + (k_1 + k_2 + \cdots + k_n) + (k_2 + \cdots + k_n) + \cdots + k_n \\
&= 0 \cdot k_0 + 1 \cdot k_1 + 2 \cdot k_2 + \cdots + n \cdot k_n \\
&= \sum_{i=1}^{k} y_i
\end{aligned}
$$

$\implies$ s= The total number of successes,
Therefore the total number of failures = nk - s

$$L(\alpha,\beta) = R_1(\alpha)R_2(\beta)R_3(\alpha+\beta)$$

$$R_1(\alpha) = \prod_{i=1}^{k} \frac{\Gamma(\alpha+y_i)}{\Gamma(\alpha)} = \left(\frac{\Gamma(\alpha)}{\Gamma(\alpha)}\right)^{k_0} \left(\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}\right)^{k_1} \cdots \left(\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right)^{k_n} \tag{15}$$

$$= \alpha^{s_1}(\alpha+1)^{s_2}\dots(\alpha+n-1)^{s_n} = \sum_{j=s_1}^{s} a_j\alpha^j \tag{16}$$

$$R_2(\beta) = \prod_{i=1}^{k} \frac{\Gamma(\beta+n-y_i)}{\Gamma(\beta)} = \left(\frac{\Gamma(\beta)}{\Gamma(\beta)}\right)^{k_n} \left(\frac{\Gamma(\beta+1)}{\Gamma(\beta)}\right)^{k_{n-1}} \cdots \left(\frac{\Gamma(\beta+n)}{\Gamma(\beta)}\right)^{k_0} \tag{17}$$

$$= \beta^{k-s_n}(\beta+1)^{k-s_{n-1}}\dots(\beta+n-1)^{k-s_1} = \sum_{j=k-s_n}^{nk-s} b_j\beta^j \tag{18}$$

The coefficients $a_{s_i},\dots a_s$ and $b_{k-s_1},\dots,b_{nk-s}$ can be calculated by recursive polynomial multiplication.

We expand R_3 using the polynomial approximation

$$R_3(\alpha+\beta) = \prod_{i=1}^{k} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)} = \left(\prod_{j=1}^{n} \frac{1}{\alpha+\beta+n-j} \right)^{k} \tag{19}$$

$$= (\alpha+\beta+n)^{-nk} \prod_{j=1}^{n} \left(\frac{1}{1-\frac{j}{\alpha+\beta+n}} \right)^{k} \tag{20}$$

$$\approx (\alpha+\beta+n)^{-nk} \left(1 + \frac{u_1}{\alpha+\beta+n} + \cdots + \frac{u_m}{(\alpha+\beta+n)^m} \right)^{k} \tag{21}$$

$$= \sum_{l=0}^{km} \frac{c_l}{(\alpha+\beta+n)^{nk+l}}. \tag{22}$$

The coefficients $u_i, \ldots, u_m$ and $c_1, \ldots, c_{km}$ can be calculated by recursive polynomial multiplication.
Note: The integer m may be chosen as large as necessary for desired accuracy

$$
\begin{aligned}
L(\alpha, \beta) &= R_1(\alpha) R_2(\beta) R_3(\alpha + \beta) && (23) \\
&\approx \left(\sum_{i=s_1}^{s} a_i \alpha^i \right) \left(\sum_{j=k-s_n}^{nk-s} b_j \beta^j \right) \left(\sum_{l=0}^{km} \frac{c_l}{(\alpha + \beta + n)^{nk+l}} \right) && (24) \\
&= \sum_{i=s_1}^{s} \sum_{j=k-s_n}^{nk-s} \sum_{l=0}^{km} a_i b_j c_l \frac{\alpha^i \beta^j}{(\alpha + \beta + n)^{nk+l}}. && (25)
\end{aligned}
$$

PRIOR AND POSTERIOR DENSITIES

Prior Distribution

Let us use non-informative prior distribution given as

$$p(\alpha, \beta) \propto h(\alpha, \beta) = (\alpha + \beta + \gamma)^{-c}, \alpha > 0, \beta > 0 \tag{26}$$

$$h(\alpha, \beta) = \frac{1}{(\alpha + \beta + \gamma)^c} \tag{27}$$

$$= \frac{1}{(\alpha + \beta + n)^c} + \frac{c(n - \gamma)}{(\alpha + \beta + n)^{c+1}} + \frac{c(c+1)(n-\gamma)^2}{2(\alpha + \beta + n)^{c+2}} + \cdots \tag{28}$$

$$\approx \sum_{t=0}^{m} \frac{v_t}{(\alpha + \beta + n)^{c+t}}, \quad \text{where} \quad v_t = \frac{\Gamma(c+t)(n-\gamma)^t}{\Gamma(c)t!} \tag{29}$$

Posterior Distribution

$$f(\alpha, \beta \mid Y) \propto \mathrm{L}(\alpha, \beta) h(\alpha, \beta) \tag{30}$$

$$\approx \sum_{i=s_1}^{s} \sum_{j=k-s_n}^{nk-s} \sum_{l=0}^{(k+1)m} a_i^* b_j^* c_l^* \frac{\alpha^i \beta^j}{(\alpha + \beta + n)^{nk+c+l}}. \tag{31}$$

Computation

```
Fully Bayesian Approach Results:
--------------------------------------------------------------
Number of Observations (k): 500
Number of Trials per Observation (n): 100
Prior Alpha (a): 2
Prior Beta (b): 3
Posterior Alpha: 1.989365
Posterior Beta: 3.029162
Computation Time: 256.8079 seconds
--------------------------------------------------------------
```

```
Polynomial Expansion Method Results:
--------------------------------------------------------------
Number of Observations (k): 500
Number of Trials per Observation (n): 100
Prior Alpha (a): 2
Prior Beta (b): 3
Approximated Posterior Alpha: 1.971
Approximated Posterior Beta: 3.0295
Computation Time: 1.311302e-05 seconds
--------------------------------------------------------------
```

Conclusion

Accuracy:

The posterior probability calculated using the polynomial expansion method closely approximates the results of the fully Bayesian approach.

Efficiency:

The polynomial expansion method is significantly faster than the fully Bayesian approach, making it ideal for scenarios requiring quick computations.

This implementation ensures that the polynomial expansion method provides posterior parameters that are almost identical to the fully Bayesian approach, but with a dramatic reduction in computation time.

References

- **Research Paper**
  1. Everson, Philip J., and Eric T. Bradlow. "Bayesian inference for the beta-binomial distribution via polynomial expansions." Journal of Computational and Graphical Statistics, 11(1) (2002): 202-207.
- **Books**
  1. Donovan, T. M., Mickey, R. M. (2019). Bayesian Statistics for Beginners
- **Youtube**
  1. 3Blue1Brown
  2. Learn Statistics with Brian
  3. Bary Van Veen


THANK YOU