STAT4610 Midterm Report

Fan Zixian 3035771610

November 13, 2023

Abstract

This scientific report aims to give a result for the STAT4610 Midterm project with detailed explanation. The related codes are not all included in the main parts. The full codes are in the attached jupyter notebook file.

1 Initial Presentations

From the clinical trial table, it shows that the first trial has zero number of observed toxicity, and the last trial has neither observed toxicity number nor treated patients.

For the Bayesian Model Averaging Equation, as we specify a discrete uniform distribution for the prior, we have

$$\begin{aligned} P(M_k|D) &= \frac{P(D|M_k)P(M_k)}{\sum_{k=1}^{3} P(D|M_k)P(M_j)} \\ &= \frac{P(D|M_k)\frac{1}{3}}{\sum_{k=1}^{3} P(D|M_k)\frac{1}{3}} \\ &= \frac{P(D|M_k)}{\sum_{k=1}^{3} P(D|M_k)} \end{aligned}$$

2 Question 1

Three possible approaches are required to be implemented to this problem, and they are presented one by one.

2.1 First approach

This approach is to simply generate the first sample, which follows the Standard Uniform Distribution, and then generate the later samples accordingly.

Under M_1 assumption, we have:

$$p_1 \sim Unif(0,1)$$
$$p_2 \sim Unif(p_1,1)$$
$$p_3 \sim Unif(p_2,1)$$
$$p_4 \sim Unif(p_3,1)$$
$$p_5 \sim Unif(p_4,1)$$

Under M_2 assumption, we have:

$$p_1 \sim Unif(0,1)$$
$$p_2 \sim Unif(0,p_1)$$
$$p_3 \sim Unif(0,p_2)$$
$$p_4 \sim Unif(0,p_3)$$
$$p_5 \sim Unif(0,p_4)$$

Under M_3 assumption, we have:

$$p_1 \sim Unif(0, 1)$$
$$p_2 \sim Unif(p_1, 1)$$
$$p_3 \sim Unif(p_2, 1)$$
$$p_4 \sim Unif(0, p_3)$$
$$p_5 \sim Unif(0, p_4)$$

In the functions, we set the random seed to be 4610 to make sure all the result are provided from the same seed, which is easier to compare. Afterwards, the generated probabilities from the three models are shown in 1:

```
The 5 probability samples generated from M1: [0.45571091466811897, 0.5612612802557018, 0.8624188213142363, 0.8657826909514685, 0.9154688051978277]
The 5 probability samples generated from M2: [0.45571091466811897, 0.08837298954128828, 0.06066068716651454, 0.001483158130220272, 0.0005490526134516267]
The 5 probability samples generated from M3: [0.45571091466811897, 0.5612612802557018, 0.8624188213142363, 0.021086201726927923, 0.007805933790902358]
```

Figure 1: The Five probabilities generated from the joint uniform under Method 1.

2.2 Second Approach

Under this approach, for M_1 and M_2, we first directly generate 5 samples, and sort them in the required order under the constraint similar to the First Approach. But for M_3, since the order is not constraint, we cannot specify the order inside it. For example, we cannot assume p_2 is greater than p_4 in the code, hence we can do it by the random shuffle as shown in 2:

```python
def Method2_M3_generator(Q1 = False):
    if Q1:
        np.random.seed(4610)
    five_samples = uniform.rvs(0, 1, size = 5)
    # We don't sort them
    p_3 = max(five_samples)

    remaining_numbers = [num for num in five_samples if num != p_3]
    random.shuffle(remaining_numbers)

    p_1 = remaining_numbers[0]
    p_2 = remaining_numbers[1]
    p_4 = remaining_numbers[2]
    p_5 = remaining_numbers[3]
    if (p_1 > p_2):
        temp = p_1
        p_1 = p_2
        p_2 = temp

    if (p_5 > p_4):
        temp = p_5
        p_5 = p_4
        p_4 = temp

    result = [p_1, p_2, p_3, p_4, p_5]
    return result
```

Figure 2: Code to get the probabilities under M3.

And the generated five probabilities are shown in 3:

```
The 5 probability samples generated from M1: [0.02445007136416011, 0.19392335512887104, 0.37019155426811023, 0.45571091466811897, 0.686416601739761]
The 5 probability samples generated from M1: [0.686416601739761, 0.45571091466811897, 0.37019155426811023, 0.19392335512887104, 0.02445007136416011]
The 5 probability samples generated from M1: [0.02445007136416011, 0.37019155426811023, 0.686416601739761, 0.45571091466811897, 0.19392335512887104]
```

Figure 3: The Five probabilities generated from the joint uniform under Method 2.

2.3 Third Approach

We first derive the prior full conditional distribution theoretically, which also means that we won't use any of the information from the observed y's. We assume $f(p_1, ..., p_5|M_k)$ to be the representation of parameters under model M_k. Just to mention here, for the patients with different treatment, if we already know the toxicity probability p_i and the number in each treatment group n_i, we have

$$y_i|p_i \sim Bin(n_i, p_i)$$

And the $p_i's$ follow the Uniform distribution under specific constraints. The next step is to derive the relationship of p_i under M_k with $1 \leq i \leq 5$ and $1 \leq k \leq 3$.

Then, for M_1, we have

$$\begin{aligned} f(p_1, ..., p_5|M_1) &= f(p_1)f(p_2|p_1)f(p_3|p_2)f(p_4|p_3)f(p_5|p_4) \\ &= 1 \cdot \frac{1}{1-p_1} \cdot \frac{1}{1-p_2} \cdot \frac{1}{1-p_3} \cdot \frac{1}{1-p_4} \end{aligned}$$

Hence, the prior full conditional distribution is

$$\begin{aligned} p_1|M_1, p_2, p_3, p_4, p_5 &\sim Unif(0, p_2) \\ p_2|M_1, p_1, p_3, p_4, p_5 &\sim Unif(p_1, p_3) \\ p_3|M_1, p_1, p_2, p_4, p_5 &\sim Unif(p_2, p_4) \\ p_4|M_1, p_1, p_2, p_3, p_5 &\sim Unif(p_3, p_5) \\ p_5|M_1, p_1, p_2, p_3, p_4 &\sim Unif(p_4, 1) \end{aligned}$$

that is,

$$\begin{aligned} f(p_1|M_1, p_2, p_3, p_4, p_5) &= \frac{1}{p_2} \\ f(p_2|M_1, p_1, p_3, p_4, p_5) &= \frac{1}{p_3 - p_1} \\ f(p_3|M_1, p_1, p_2, p_4, p_5) &= \frac{1}{p_4 - p_2} \\ f(p_4|M_1, p_1, p_2, p_3, p_5) &= \frac{1}{p_5 - p_3} \\ f(p_5|M_1, p_1, p_2, p_3, p_4) &= \frac{1}{1 - p_4} \end{aligned}$$

For M_2, we have

$$\begin{aligned} f(p_1, ..., p_5|M_2) &= f(p_1)f(p_2|p_1)f(p_3|p_2)f(p_4|p_3)f(p_5|p_4) \\ &= 1 \cdot \frac{1}{p_1} \cdot \frac{1}{p_2} \cdot \frac{1}{p_3} \cdot \frac{1}{p_4} \end{aligned}$$

Hence, the prior full conditional distribution is

$$\begin{aligned} p_1|M_2, p_2, p_3, p_4, p_5 &\sim Unif(p_2, 1) \\ p_2|M_2, p_1, p_3, p_4, p_5 &\sim Unif(p_3, p_1) \\ p_3|M_2, p_1, p_2, p_4, p_5 &\sim Unif(p_4, p_2) \\ p_4|M_2, p_1, p_2, p_3, p_5 &\sim Unif(p_5, p_3) \\ p_5|M_2, p_1, p_2, p_3, p_4 &\sim Unif(0, p_4) \end{aligned}$$

that is,

$$\begin{aligned} f(p_1|M_2, p_2, p_3, p_4, p_5) &= \frac{1}{1 - p_2} \\ f(p_2|M_2, p_1, p_3, p_4, p_5) &= \frac{1}{p_1 - p_3} \\ f(p_3|M_2, p_1, p_2, p_4, p_5) &= \frac{1}{p_2 - p_4} \\ f(p_4|M_2, p_1, p_2, p_3, p_5) &= \frac{1}{p_3 - p_5} \\ f(p_5|M_2, p_1, p_2, p_3, p_4) &= \frac{1}{p_4} \end{aligned}$$

For M_3, we have

$$\begin{aligned} f(p_1, ..., p_5|M_3) &= f(p_1)f(p_2|p_1)f(p_3|p_2)f(p_4|p_3)f(p_5|p_4) \\ &= 1 \cdot \frac{1}{1-p_1} \cdot \frac{1}{1-p_2} \cdot \frac{1}{p_3} \cdot \frac{1}{p_4} \end{aligned}$$

Hence, the prior full conditional distribution is

$$
\begin{aligned}
p_1|M_3, p_2, p_3, p_4, p_5 &\sim Unif(0, p_2) \\
p_2|M_3, p_1, p_3, p_4, p_5 &\sim Unif(p_1, p_3) \\
p_3|M_3, p_1, p_2, p_4, p_5 &\sim Unif(max(p_2, p_4), 1) \\
p_4|M_3, p_1, p_2, p_3, p_5 &\sim Unif(p_5, p_3) \\
p_5|M_3, p_1, p_2, p_3, p_4 &\sim Unif(0, p_4)
\end{aligned}
$$

that is,

$$
\begin{aligned}
f(p_1|M_3, p_2, p_3, p_4, p_5) &= \frac{1}{p_2} \\
f(p_2|M_3, p_1, p_3, p_4, p_5) &= \frac{1}{p_3 - p_1} \\
f(p_3|M_3, p_1, p_2, p_4, p_5) &= \frac{1}{1 - max(p_2, p_4)} \\
f(p_4|M_3, p_1, p_2, p_3, p_5) &= \frac{1}{p_3 - p_5} \\
f(p_5|M_3, p_1, p_2, p_3, p_4) &= \frac{1}{p_4}
\end{aligned}
$$

In order to use Gibbs sampler to draw samples, in the first question, we just start with the joint uniform prior, and go one step following the Gibbs sampler. The generates five probabilities are shown in 4:

```
The 5 probability samples generated from M1: [0.2557728913931253, 0.5345810764823737, 0.7702898321424881, 0.8637158972060577, 0.9154688051978277]
The 5 probability samples generated from M1: [0.5038113683135906, 0.1372701527280495, 0.061125780933146315, 0.0020187863680903276, 0.0005490526134516267]
The 5 probability samples generated from M1: [0.2557728913931253, 0.5345810764823737, 0.8624188213142363, 0.02870127987957881, 0.007805933790902358]
```

Figure 4: The Five probabilities generated from the joint uniform under Method 3.

3 Question 2

In this question, we generate 100,000 samples from the three models in Question 1, and plot the probabilities' distribution accordingly.

3.1 Method 1

Under method 1, we have the following plots of 5, 6, 7:

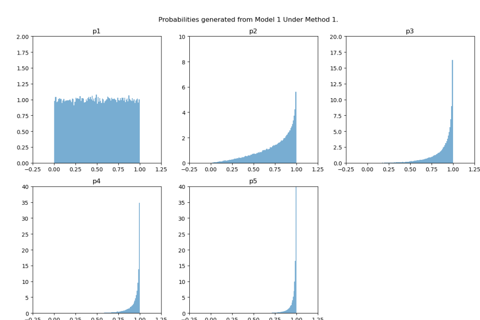

Figure 5: The density distribution of Model 1 simulated by Method 1.

From the plots, we can see that p_1 always follows Standard Uniform Distribution, but the skewness and kurtosis are increasing quickly in the following probabilities under M_1 and M_2. In the plots of M_3, the density distributions are still wiggly skewed.

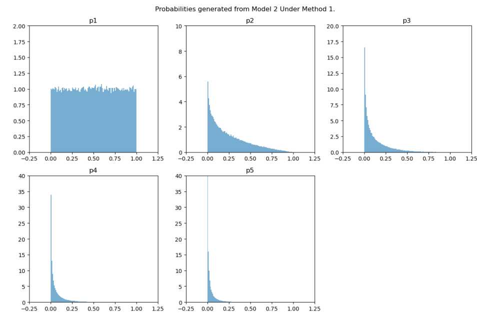

Figure 6: The density distribution of Model 2 simulated by Method 1.

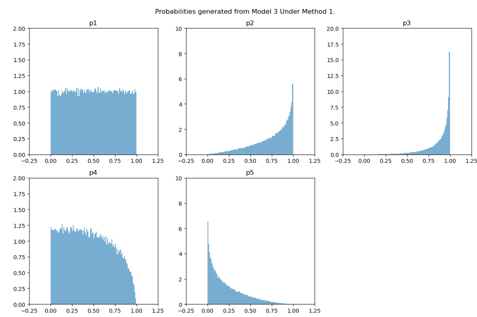

Figure 7: The density distribution of Model 3 simulated by Method 1.

3.2 Method 2

Under method 2, we have the following plots of 8, 9, 10:

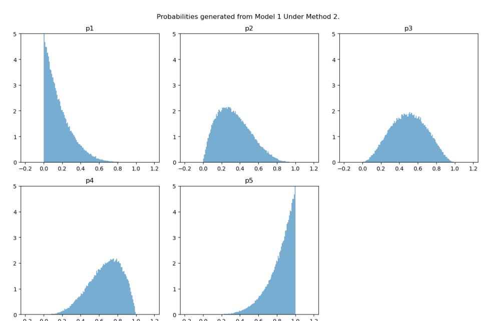

Figure 8: The density distribution of Model 1 simulated by Method 2.

From the plots, we can see that p_3 under M_1 and M_2 are more bell-shaped, and the other probabilities just skewed to it. For the distributions of M_3, p_1 and p_5 have similar shape, and p_2 and p_4 have similar shape. Compared to Method 1, the density distributions are much more smoother.

3.3 Method 3

Under method 2, we have the following plots of 11, 12, 13:

It is found that the density distributions generated by Method 3 is similar to the distribution generated by Method 2.

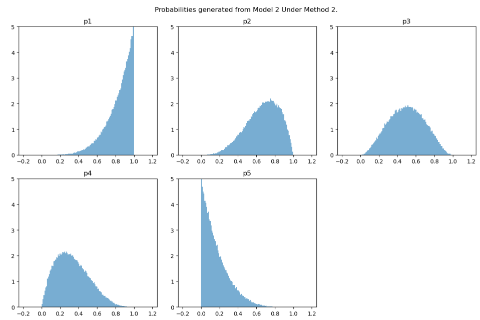

Figure 9: The density distribution of Model 2 simulated by Method 2.

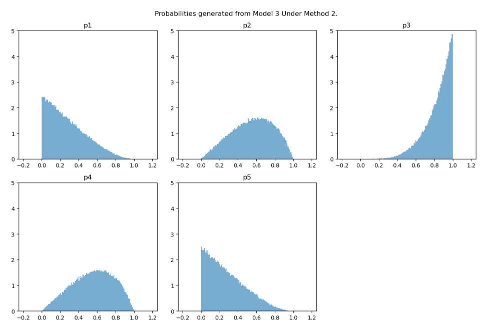

Figure 10: The density distribution of Model 3 simulated by Method 2.

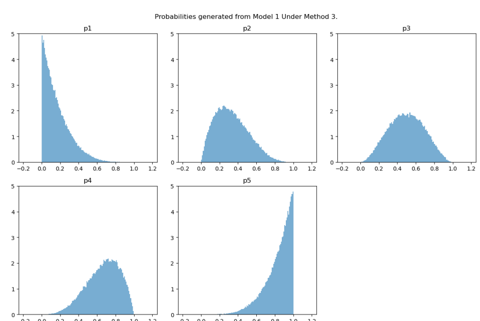

Figure 11: The density distribution of Model 1 simulated by Method 3.

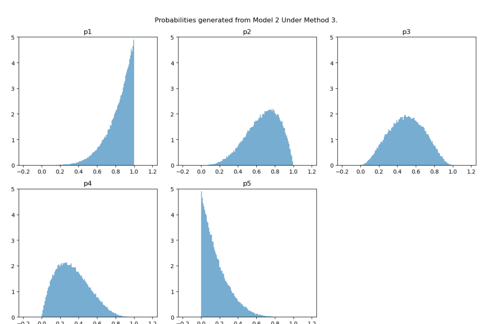

Figure 12: The density distribution of Model 2 simulated by Method 3.

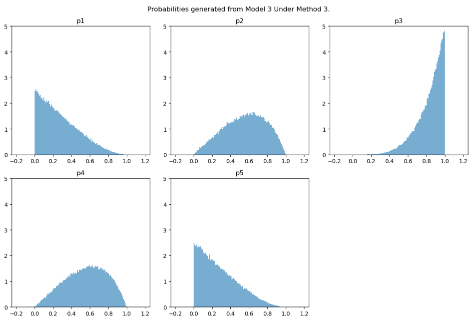

Figure 13: The density distribution of Model 3 simulated by Method 3.

4 Question 3

4.1 Posterior Full Conditional Distribution

From the problem statement, it is shown that the marginal likelihood under model M_k is

$$P(D|M_k) \propto \int f(p_1, ..., p_5|M_k) \prod_{j=1}^{5} \{p_j^{y_j}(1-p_j)^{n_j-y_j}\} dp_1 \cdots dp_5$$

, where D represents the observed data y's. And we have the likelihood to be

$$f(y|p_1, ..., p_5, M_k) = \prod_{j=1}^{5} \{p_j^{y_j}(1-p_j)^{n_j-y_j}\}$$

With the toxicity data, we can further derive the likelihood to be

$$\begin{aligned} f(y|p_1, ..., p_5, M_k) &= (p_1^0(1-p_1)^3) \times (p_2^1(1-p_2)^6) \times (p_3^4(1-p_3)^{12}) \times (p_4^3(1-p_4)^6) \times (p_5^0(1-p_5)^0) \\ &= (1-p_1)^3 p_2(1-p_2)^6 p_3^4(1-p_3)^{12} p_4^3(1-p_4)^6 \end{aligned}$$

Based on the Bayes's theorem and suggestion from [FP08], we have the posterior joint conditional distribution to be

$$\begin{aligned} P(M_k|D) &= \frac{P(D|M_k)P(M_k)}{\sum_{k=1}^{3} P(D|M_k)P(M_j)} \\ &= \frac{P(D|M_k)\frac{1}{3}}{\sum_{k=1}^{3} P(D|M_k)\frac{1}{3}} \\ &= \frac{P(D|M_k)}{\sum_{k=1}^{3} P(D|M_k)} \\ &= \frac{\int f(p_1, ..., p_5|M_k) \prod_{j=1}^{5} \{p_j^{y_j}(1-p_j)^{n_j-y_j}\} dp_1 \cdots dp_5}{\sum_{k=1}^{3} \int f(p_1, ..., p_5|M_k) \prod_{j=1}^{5} \{p_j^{y_j}(1-p_j)^{n_j-y_j}\} dp_1 \cdots dp_5} \end{aligned}$$

Then we aim to derive the posterior full conditional distribution based on the obtained likelihood and the prior full conditional distribution from Question 2.

4.1.1 Model 1

$$f(p_1|p_2, p_3, p_4, p_5, M_1, D) \propto 1_{\{0<p_1<p_2\}} \cdot p_1^{y_1}(1-p_1)^{n_1-y_1}$$

And,

$$\begin{aligned} p_1|p_2, p_3, p_4, p_5, M_1, D &\propto trancatedBeta(y_1+1, n_1-y_1+1, 0, p_2) \\ &\propto trancatedBeta(1, 4, 0, p_2) \end{aligned}$$

$$f(p_2|p_1,p_3,p_4,p_5,M_1,D) \propto 1_{\{p_1<p_2<p_3\}} \cdot p_2^{y_2}(1-p_2)^{n_2-y_2}$$

And,

$$\begin{aligned} p_2|p_1,p_3,p_4,p_5,M_1,D &\propto trancatedBeta(y_2+1,n_2-y_2+1,p_1,p_3) \\ &\propto trancatedBeta(2,6,p_1,p_3) \end{aligned}$$

$$f(p_3|p_1,p_2,p_4,p_5,M_1,D) \propto 1_{\{p_2<p_3<p_4\}} \cdot p_3^{y_3}(1-p_3)^{n_3-y_3}$$

And,

$$\begin{aligned} p_3|p_1,p_2,p_4,p_5,M_1,D &\propto trancatedBeta(y_3+1,n_3-y_3+1,p_2,p_4) \\ &\propto trancatedBeta(5,9,p_2,p_4) \end{aligned}$$

$$f(p_4|p_1,p_2,p_3,p_5,M_1,D) \propto 1_{\{p_3<p_4<p_5\}} \cdot p_4^{y_4}(1-p_4)^{n_4-y_4}$$

And,

$$\begin{aligned} p_4|p_1,p_2,p_3,p_5,M_1,D &\propto trancatedBeta(y_4+1,n_4-y_4+1,p_3,p_5) \\ &\propto trancatedBeta(4,4,p_3,p_5) \end{aligned}$$

$$f(p_5|p_1,p_2,p_3,p_4,M_1,D) \propto 1_{\{p_4<p_5<1\}} \cdot p_5^{y_5}(1-p_5)^{n_5-y_5}$$

And,

$$\begin{aligned} p_5|p_1,p_2,p_3,p_4,M_1,D &\propto trancatedBeta(y_5+1,n_5-y_5+1,p_4,1) \\ &\propto trancatedBeta(1,1,p_4,1) \end{aligned}$$

4.1.2 Model 2

$$f(p_1|p_2,p_3,p_4,p_5,M_2,D) \propto 1_{\{p_2<p_1<1\}} \cdot p_1^{y_1}(1-p_1)^{n_1-y_1}$$

And,

$$\begin{aligned} p_1|p_2,p_3,p_4,p_5,M_2,D &\propto trancatedBeta(y_1+1,n_1-y_1+1,p_2,1) \\ &\propto trancatedBeta(1,4,p_2,1) \end{aligned}$$

$$f(p_2|p_1,p_3,p_4,p_5,M_2,D) \propto 1_{\{p_3<p_2<p_1\}} \cdot p_2^{y_2}(1-p_2)^{n_2-y_2}$$

And,

$$\begin{aligned} p_2|p_1,p_3,p_4,p_5,M_2,D &\propto trancatedBeta(y_2+1,n_2-y_2+1,p_3,p_1) \\ &\propto trancatedBeta(2,6,p_3,p_1) \end{aligned}$$

$$f(p_3|p_1,p_2,p_4,p_5,M_2,D) \propto 1_{\{p_4<p_3<p_2\}} \cdot p_3^{y_3}(1-p_3)^{n_3-y_3}$$

And,

$$\begin{aligned} p_3|p_1,p_2,p_4,p_5,M_2,D &\propto trancatedBeta(y_3+1,n_3-y_3+1,p_4,p_2) \\ &\propto trancatedBeta(5,9,p_4,p_2) \end{aligned}$$

$$f(p_4|p_1,p_2,p_3,p_5,M_2,D) \propto 1_{\{p_5<p_4<p_3\}} \cdot p_4^{y_4}(1-p_4)^{n_4-y_4}$$

And,

$$p_4|p_1,p_2,p_3,p_5,M_2,D \propto trancatedBeta(y_4+1,n_4-y_4+1,p_5,p_3)$$
$$\propto trancatedBeta(4,4,p_5,p_3)$$

$$f(p_5|p_1,p_2,p_3,p_4,M_2,D) \propto 1_{\{0<p_5<p_4\}} \cdot p_5^{y_5}(1-p_5)^{n_5-y_5}$$

And,

$$p_5|p_1,p_2,p_3,p_4,M_2,D \propto trancatedBeta(y_5+1,n_5-y_5+1,0,p_4)$$
$$\propto trancatedBeta(1,1,0,p_4)$$

4.1.3 Model 3

$$f(p_1|p_2,p_3,p_4,p_5,M_3,D) \propto 1_{\{0<p_1<p_2\}} \cdot p_1^{y_1}(1-p_1)^{n_1-y_1}$$

And,

$$p_1|p_2,p_3,p_4,p_5,M_3,D \propto trancatedBeta(y_1+1,n_1-y_1+1,0,p_2)$$
$$\propto trancatedBeta(1,4,0,p_2)$$

$$f(p_2|p_1,p_3,p_4,p_5,M_3,D) \propto 1_{\{p_1<p_2<p_3\}} \cdot p_2^{y_2}(1-p_2)^{n_2-y_2}$$

And,

$$p_2|p_1,p_3,p_4,p_5,M_3,D \propto trancatedBeta(y_2+1,n_2-y_2+1,p_1,p_3)$$
$$\propto trancatedBeta(2,6,p_1,p_3)$$

$$f(p_3|p_1,p_2,p_4,p_5,M_3,D) \propto 1_{\{max\{p_2,p_4\}<p_3<1\}} \cdot p_3^{y_3}(1-p_3)^{n_3-y_3}$$

And,

$$p_3|p_1,p_2,p_4,p_5,M_3,D \propto trancatedBeta(y_3+1,n_3-y_3+1,\{max\{p_2,p_4\},1)$$
$$\propto trancatedBeta(5,9,\{max\{p_2,p_4\},1)$$

$$f(p_4|p_1,p_2,p_3,p_5,M_3,D) \propto 1_{\{p_5<p_4<p_3\}} \cdot p_4^{y_4}(1-p_4)^{n_4-y_4}$$

And,

$$p_4|p_1,p_2,p_3,p_5,M_3,D \propto trancatedBeta(y_4+1,n_4-y_4+1,p_5,p_3)$$
$$\propto trancatedBeta(4,4,p_5,p_3)$$

$$f(p_5|p_1,p_2,p_3,p_4,M_3,D) \propto 1_{\{0<p_5<p_4\}} \cdot p_5^{y_5}(1-p_5)^{n_5-y_5}$$

And,

$$p_5|p_1,p_2,p_3,p_4,M_3,D \propto trancatedBeta(y_5+1,n_5-y_5+1,0,p_4)$$
$$\propto trancatedBeta(1,1,0,p_4)$$

4.2 Power Priori

Moreover, as mentioned by [FP08], it is hard to calculate the integration in the theoretical equation and compute the $P(D|M_k)$, especially for high dimensional parameters, which is the case of our problem as it has 5 probabilities but not many data points. Therefore, [FP08] proposed a new approach called power posterior to estimate the integration with the form of

$$p_t(\theta|y) \propto p(y|\theta)^t p(\theta), t \in [0,1]$$

So we can transform our problem accordingly to

$$p_t(p_1, ..., p_5|D, M_k) \propto p(D|p_1, ..., p_5)^t p(p_1, ..., p_5|M_k)$$

And what we are going to do is to estimate the marginal likelihood using the form

$$log\{p(y)\} = E_{\theta,t|y}[\frac{log\{p(y|\theta)\}}{p(t)}]$$

, where the expectation is with respect to the joint distribution

$$p(\theta, t|y) = p(\theta|t, y)p(t) = \frac{p(y|\theta)^t p(\theta)}{z(y|t)} p(t)$$

In order to have the approximation of the marginal likelihood, [FP08] also stated the formula for discrete cases of the log likelihood as

$$log\{p(y)\} \approx \sum_{i=0}^{n-1} (t_{i+1} - t_i) \frac{E_{\theta|y,t_{i+1}}[log\{p(y|\theta)\}] + E_{\theta|y,t_i}[log\{p(y|\theta)\}]}{2}$$

As the sampling is in the Gibbs sequence, we have the following equation

$$E_{p_1,...p_5|D,M_k,t_i}[logp(D|p_1, ..., p_5)] \approx \frac{1}{n} \sum_{j=1}^{n} logp(D|(p_1, ..., p_5)^j)$$

, where $(p_1, ..., p_5)^j$ stands for the sample generated from Gibbs sampler at the t_j.

It is also essential to choose the hyperparameter of n and c for the problem. Here we select c to be 5, and n to be 30.

From the python result, we have generated the corresponding log likelihood, and then we take the exponent of them to get the final approximated result:

$$P(D|M_1) = 2.1121724693115795 \times 10^{-8}$$
$$P(D|M_2) = 1.0157455951473162 \times 10^{-11}$$
$$P(D|M_3) = 8.25767461646286 \times 10^{-12}$$

5 Question 4

We also ignore the constant terms of combination in all the result below

5.1 Model 1

The Integration of M_1 is

$$\begin{aligned} P(D|M_1) &= \int_1^1 f(p_1, ..., p_5|M_k)(1-p_1)^3 p_2(1-p_2)^5 p_3^4 (1-p_3)^8 p_4^3 (1-p_4)^3 dp_1 \cdots dp_5 \\ &= \int_0^1 \int_{p_1}^1 \int_{p_2}^1 \int_{p_3}^1 \int_{p_4}^1 (1-p_1)^3 p_2 (1-p_2)^5 p_3^4 (1-p_3)^8 p_4^3 (1-p_4)^3 dp_5 \cdots dp_1 \end{aligned}$$

And python gives the numerical answer to be $7.38222302275044 \times 10^{-10}$

5.2 Model 2

The Integration of M_2 is

$$\begin{aligned} P(D|M_2) &= \int_1^1 f(p_1,...,p_5|M_k)(1-p_1)^3 p_2(1-p_2)^5 p_3^4(1-p_3)^8 p_4^3(1-p_4)^3 dp_1 \cdots dp_5 \\ &= \int_0^1 \int_0^{p_1} \int_0^{p_2} \int_0^{p_3} \int_0^{p_4} (1-p_1)^3 p_2(1-p_2)^5 p_3^4(1-p_3)^8 p_4^3(1-p_4)^3 dp_5 \cdots dp_1 \end{aligned}$$

And python gives the numerical answer to be $5.50779957961358 \times 10^{-12}$

5.3 Model 3

The Integration of M_3 is

$$\begin{aligned} P(D|M_3) &= \int_1^1 f(p_1,...,p_5|M_k)(1-p_1)^3 p_2(1-p_2)^5 p_3^4(1-p_3)^8 p_4^3(1-p_4)^3 dp_1 \cdots dp_5 \\ &= \int_0^1 \int_{p_1}^{1} \int_{p_2}^{1} \int_0^{p_3} \int_0^{p_4} (1-p_1)^3 p_2(1-p_2)^5 p_3^4(1-p_3)^8 p_4^3(1-p_4)^3 dp_5 \cdots dp_1 \end{aligned}$$

And python gives the numerical answer to be $2.82264736718615 \times 10^{-10}$

6 Question 5

Based on all the findings above, the final $P(M_k|D)$ can be calculated. By using the gibbs sampler, we have:

$$P(M_1|D) = 0.9991289021396487$$

$$P(M_2|D) = 0.00048048196635357897$$

$$P(M_3|D) = 0.0003906158939976176$$

By using the numerical method to calculate the integration, we have:

$$P(M_1|D) = 0.719518534129861$$

$$P(M_2|D) = 0.00536825271681933$$

$$P(M_3|D) = 0.275113213153320$$

Both of the methods prefer M_1 most, however, the gibbs sampler prefers M_1 more, but the numerical method is more neutral. Given that the M_1 assumption is more reasonable, we prefer the gibbs sampler method more. It is because gibbs sampler performs better in high-dimensional cases by sampling from various conditional distribution. And in this question, the conditionality is very important for the probability cases, hence gibbs sampler can provide more reliable samples.

References

[FP08] N. Friel and A. N. Pettitt. Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(3):589–607, 2008.