Appendix: Mathematical Derivation and Computational Complexity of the Bayesian Score

In this appendix, we first describe the mathematical derivation of the marginal likelihood of model M_h: $P(G|M_h)$ (Section 2). Then we show four equivalent formulas for solving $P(G|M_h)$ (Section 3). After that, we show how to obtain the marginal likelihood of model M_l from the solution to the marginal likelihood of model M_h (Section 4). Finally, we show that the overall computational complexity for computing the Bayesian score is $O(min(N_{11}, N_{12}, N_{21}, N_{22}))$ (Section 5).

1 Definition and Notations

We want to evaluate rule $P \Rightarrow y$ with respect to a group of instances G where $G_P \subseteq G$. Let θ denote the probability of $Y = y$ in G, let θ_1 denote the probability of $Y = y$ in G_P and let θ_2 denote the probability of $Y = y$ in the instances of G not covered by P $(G \setminus G_P)$.

We define the following three models:

1. $\boldsymbol{M_e}$ is the model that conjectures that θ_1 is the **same** as θ_2.
2. $\boldsymbol{M_h}$ is the model that conjectures that θ_1 is **higher** than θ_2.
3. $\boldsymbol{M_l}$ is the model that conjectures that θ_1 is **lower** than θ_2.

Let α and β be the beta parameters for the prior distribution on θ. Let α_1 and β_1 be the beta parameters for the prior distribution on θ_1 and let α_2 and β_2 be the beta parameters for the prior distribution on θ_2. Let N_{*1} and N_{*2} be the number of instances in G with $Y = y$ and with $Y \neq y$, respectively. Let N_{11} and N_{12} be the number of instances in G_P with $Y = y$ and with $Y \neq y$, respectively and let N_{21} and N_{22} be the number of instances in $G \setminus G_P$ with $Y = y$ and with $Y \neq y$, respectively.

We define the Bayesian score of rule $P \Rightarrow y$ with respect to group G as follows:

$$BS(P \Rightarrow y, G) = Pr(M_h|G) = \frac{Pr(G|M_h) \cdot Pr(M_h)}{Pr(G|M_e) \cdot Pr(M_e) + Pr(G|M_h) \cdot Pr(M_h) + Pr(G|M_l) \cdot Pr(M_l)} \tag{1}$$

Evaluating Equation 1 requires evaluating the marginal likelihood for models M_e, M_h and M_l. Evaluating the marginal likelihood of M_e is easy and is given by the following well known closed-form solution:

$$Pr(G|M_e) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+N_{*1}+\beta+N_{*2})} \cdot \frac{\Gamma(\alpha+N_{*1})}{\Gamma(\alpha)} \cdot \frac{\Gamma(\beta+N_{*2})}{\Gamma(\beta)} \tag{2}$$

where Γ is the gamma function.

In the rest of this appendix, we describe how to obtain a closed-form solution for the marginal likelihood of M_h and M_l and then analyse the overall computational complexity of the Bayesian score (evaluating Equation 1).

2 Derivation of the Closed-form Solution for Model M_h

The marginal likelihood of model M_h ($Pr(G|M_h)$) is defined as follows:

$$= \frac{1}{k} \int_{\theta_1=0}^{1} \int_{\theta_2=0}^{\theta_1} {\theta_1}^{N_{11}} \cdot (1-\theta_1)^{N_{12}} \cdot {\theta_2}^{N_{21}} \cdot (1-\theta_2)^{N_{22}} \cdot beta(\theta_1;\alpha_1,\beta_1) \cdot beta(\theta_2;\alpha_2,\beta_2) d\theta_2 d\theta_1$$

$$= \frac{1}{k} \underbrace{\int_{\theta_1=0}^{1} {\theta_1}^{N_{11}} \cdot (1-\theta_1)^{N_{12}} \cdot beta(\theta_1;\alpha_1,\beta_1)}_{f_1} \underbrace{\int_{\theta_2=0}^{\theta_1} {\theta_2}^{N_{21}} \cdot (1-\theta_2)^{N_{22}} \cdot beta(\theta_2;\alpha_2,\beta_2) d\theta_2}_{f_2} d\theta_1 \tag{3}$$

We first show how to solve the integral over θ_2 in closed form, which is denoted by f_2 in Equation 3. We then expand the function denoted by f_1, multiply it by the solution to f_2, and solve the integral over θ_1 in closed form to complete the integration.

We use the regularized incomplete beta function to solve the integral given by f_2. Using the notation in the expression denoted by f_2, the incomplete beta function is as follows:

$$\int_{\theta_2=0}^{\theta_1} {\theta_2}^{a-1} \cdot (1-\theta_2)^{b-1} d\theta_2 = \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(a+b)} \cdot \sum_{j=a}^{a+b-1} \frac{\Gamma(a+b)}{\Gamma(j+1)\cdot\Gamma(a+b-j)} \cdot \theta_1^j \cdot (1-\theta_1)^{a+b-1-j} \tag{4}$$

where a and b should be natural numbers.

Note that when $\theta_1 = 1$ in Equation 4, the solution to the integral in that equation is simply the following:

$$\int_{\theta_2=0}^{1} {\theta_2}^{a-1} \cdot (1-\theta_2)^{b-1} d\theta_2 = \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(a+b)} \tag{5}$$

We now solve the integral given by f_2 in Equation 3 as follows:

$$
\begin{aligned}
f_2 &= \int_{\theta_2=0}^{\theta_1} {\theta_2}^{N_{21}} \cdot (1-\theta_2)^{N_{22}} \cdot beta(\theta_2; \alpha_2, \beta_2) d\theta_2 \\
&= \int_{\theta_2=0}^{\theta_1} {\theta_2}^{N_{21}} \cdot (1-\theta_2)^{N_{22}} \cdot \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\cdot\Gamma(\beta_2)} \cdot \theta_2^{\alpha_2-1} \cdot (1-\theta_2)^{\beta_2-1} d\theta_2 \\
&= \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\cdot\Gamma(\beta_2)} \int_{\theta_2=0}^{\theta_1} {\theta_2}^{N_{21}+\alpha_2-1} \cdot (1-\theta_2)^{N_{22}+\beta_2-1} d\theta_2 \\
&= \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\cdot\Gamma(\beta_2)} \int_{\theta_2=0}^{\theta_1} {\theta_2}^{a-1} \cdot (1-\theta_2)^{b-1} d\theta_2
\end{aligned}
$$

where $a = N_{21} + \alpha_2$ and $b = N_{22} + \beta_2$.
Using Equation 4, we get the following:

$$
f_2 = \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\cdot\Gamma(\beta_2)} \cdot \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(a+b)} \cdot \sum_{j=a}^{a+b-1} \frac{\Gamma(a+b)}{\Gamma(j+1)\cdot\Gamma(a+b-j)} \cdot \theta_1^j \cdot (1-\theta_1)^{a+b-1-j} \tag{6}
$$

We now turn to f_1, which can be expanded as follows:

$$
\begin{aligned}
f_1 &= \int_{\theta_1=0}^{1} {\theta_1}^{N_{11}} \cdot (1-\theta_1)^{N_{12}} \cdot beta(\theta_1; \alpha_1, \beta_1) \\
&= \int_{\theta_1=0}^{1} {\theta_1}^{N_{11}} \cdot (1-\theta_1)^{N_{12}} \cdot \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\cdot\Gamma(\beta_1)} \cdot \theta_1^{\alpha_1-1} \cdot (1-\theta_1)^{\beta_1-1} \qquad (7) \\
&= \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\cdot\Gamma(\beta_1)} \int_{\theta_1=0}^{1} {\theta_1}^{c-1} \cdot (1-\theta_1)^{d-1}
\end{aligned}
$$

where $c = N_{11} + \alpha_1$ and $d = N_{12} + \beta_1$.
Now we combine Equations 6 and 7 to solve Equation 3:

$$
\begin{aligned}
Pr(G|M_h) &= \frac{1}{k} \cdot f_1 \cdot f_2 \; d\theta_1 \\
&= \frac{1}{k} \cdot \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\cdot\Gamma(\beta_1)} \cdot \int_{\theta_1=0}^{1} {\theta_1}^{c-1} \cdot (1-\theta_1)^{d-1} \cdot \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\cdot\Gamma(\beta_2)} \cdot \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(a+b)} \\
&\quad \cdot \sum_{j=a}^{a+b-1} \frac{\Gamma(a+b)}{\Gamma(j+1)\cdot\Gamma(a+b-j)} \cdot \theta_1^j \cdot (1-\theta_1)^{a+b-1-j} d\theta_1 \\
&= \frac{1}{k} \cdot \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\cdot\Gamma(\beta_1)} \cdot \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\cdot\Gamma(\beta_2)} \cdot \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(a+b)} \cdot \int_{\theta_1=0}^{1} {\theta_1}^{c-1} \cdot (1-\theta_1)^{d-1} \\
&\quad \cdot \sum_{j=a}^{a+b-1} \frac{\Gamma(a+b)}{\Gamma(j+1)\cdot\Gamma(a+b-j)} \cdot \theta_1^j \cdot (1-\theta_1)^{a+b-1-j} d\theta_1 \\
&= \frac{1}{k} \cdot \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\cdot\Gamma(\beta_1)} \cdot \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\cdot\Gamma(\beta_2)} \cdot \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(a+b)} \\
&\quad \cdot \sum_{j=a}^{a+b-1} \frac{\Gamma(a+b)}{\Gamma(j+1)\cdot\Gamma(a+b-j)} \cdot \int_{\theta_1=0}^{1} {\theta_1}^{(c+j)-1} \cdot (1-\theta_1)^{(a+b+d-1-j)-1} d\theta_1 \\
&= \frac{1}{k} \cdot \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\cdot\Gamma(\beta_1)} \cdot \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\cdot\Gamma(\beta_2)} \cdot \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(a+b)} \\
&\quad \cdot \sum_{j=a}^{a+b-1} \frac{\Gamma(a+b)}{\Gamma(j+1)\cdot\Gamma(a+b-j)} \cdot \frac{\Gamma(c+j)\cdot\Gamma(a+b+d-1-j)}{\Gamma(a+b+c+d-1)} \\
&= \frac{1}{k} \cdot \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\cdot\Gamma(\beta_1)} \cdot \frac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\cdot\Gamma(\beta_2)} \cdot \sum_{j=a}^{a+b-1} \frac{\Gamma(a)\cdot\Gamma(b)}{\Gamma(j+1)\cdot\Gamma(a+b-j)} \cdot \frac{\Gamma(c+j)\cdot\Gamma(a+b+d-1-j)}{\Gamma(a+b+c+d-1)}
\end{aligned} \tag{8}
$$

where $a = N_{21} + \alpha_2$, $b = N_{22} + \beta_2$, $c = N_{11} + \alpha_1$ and $d = N_{12} + \beta_1$.

We can solve for k by solving Equation 3 (without the k term) with $N_{11} = N_{12} = N_{21} = N_{22} = 0$. Doing so is equivalent to applying Equation 8 (without the k term) with $a = \alpha_2$, $b = \beta_2$, $c = \alpha_1$ and $d = \beta_1$.

3 Four Equivalent Solutions for Model M_h

In the previous section, we showed the full derivation of the closed-form solution to the marginal likelihood of model M_h. It turned out that there are four equivalent solutions to Equation 3. These solutions are derived by redefining which class map to the values 1 and 2 and by redefining which regions map to θ_1 and θ_2.

Let us use the notations introduced in the previous section: $a = N_{21} + \alpha_2$, $b = N_{22} + \beta_2$, $c = N_{11} + \alpha_1$ and $d = N_{12} + \beta_1$. Also, let us define C as follows:

$$C = \frac{1}{k} \cdot \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1) \cdot \Gamma(\beta_1)} \cdot \frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2) \cdot \Gamma(\beta_2)} \tag{9}$$

The marginal likelihood of M_h (Equation 3) can be obtained by solving any of the following four equations:

$$C \cdot \sum_{j=a}^{a+b-1} \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(j+1) \cdot \Gamma(a+b-j)} \cdot \frac{\Gamma(c+j) \cdot \Gamma(a+b+d-j-1)}{\Gamma(a+b+c+d-1)} \tag{10}$$

Which is the solution we derived in the previous section.

$$C \cdot \sum_{j=d}^{d+c-1} \frac{\Gamma(c) \cdot \Gamma(d)}{\Gamma(j+1) \cdot \Gamma(c+d-j)} \cdot \frac{\Gamma(b+j) \cdot \Gamma(c+d+a-j-1)}{\Gamma(a+b+c+d-1)} \tag{11}$$

$$C \cdot \left(\frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)} \cdot \frac{\Gamma(c) \cdot \Gamma(d)}{\Gamma(c+d)} - \sum_{j=b}^{a+b-1} \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(j+1) \cdot \Gamma(a+b-j)} \cdot \frac{\Gamma(d+j) \cdot \Gamma(a+b+c-j-1)}{\Gamma(a+b+c+d-1)} \right) \tag{12}$$

$$C \cdot \left(\frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)} \cdot \frac{\Gamma(c) \cdot \Gamma(d)}{\Gamma(c+d)} - \sum_{j=c}^{c+d-1} \frac{\Gamma(c) \cdot \Gamma(d)}{\Gamma(j+1) \cdot \Gamma(c+d-j)} \cdot \frac{\Gamma(a+j) \cdot \Gamma(c+d+b-j-1)}{\Gamma(a+b+c+d-1)} \right) \tag{13}$$

4 Derivation of the Closed-form Solution for Model M_l

The marginal likelihood of model M_l ($Pr(G|M_l)$) is defined as follows:

$$= \frac{1}{k} \underbrace{\int_{\theta_2=0}^{1} {\theta_2}^{N_{21}} \cdot (1-\theta_2)^{N_{22}} \cdot beta(\theta_2; \alpha_2, \beta_2)}_{f_1} \underbrace{\int_{\theta_1=0}^{\theta_2} {\theta_1}^{N_{11}} \cdot (1-\theta_1)^{N_{11}} \cdot beta(\theta_1; \alpha_1, \beta_1) d\theta_1}_{f_2} d\theta_2 \tag{14}$$

By solving the integral given by f_2, we get:

$$f_2 = \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1) \cdot \Gamma(\beta_1)} \int_{\theta_1=0}^{\theta_2} {\theta_1}^{c-1} \cdot (1-\theta_1)^{d-1} d\theta_2$$

$$= \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1) \cdot \Gamma(\beta_1)} \cdot \frac{\Gamma(c) \cdot \Gamma(d)}{\Gamma(c+d)} \cdot \sum_{j=c}^{c+d-1} \frac{\Gamma(c+d)}{\Gamma(j+1) \cdot \Gamma(c+d-j)} \cdot \theta_2^j \cdot (1-\theta_2)^{c+d-1-j}$$

where, same as before, $c = N_{11} + \alpha_1$ and $d = N_{12} + \beta_1$.
By solving f_1, we get:

$$f_1 = \frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2) \cdot \Gamma(\beta_2)} \int_{\theta_2=0}^{1} {\theta_2}^{a-1} \cdot (1-\theta_2)^{b-1}$$

Now we can solve Equation 14:

$$Pr(G|M_l) = C \cdot \sum_{j=c}^{c+d-1} \frac{\Gamma(c) \cdot \Gamma(d)}{\Gamma(j+1) \cdot \Gamma(c+d-j)} \cdot \frac{\Gamma(a+j) \cdot \Gamma(c+d+b-1-j)}{\Gamma(a+b+c+d-1)} \quad (15)$$

Where C is defined by Equation 9 in the previous section.

Notice that Equation 15 (the solution to $Pr(G|M_l)$) can be obtained from Equation 13 (one of the four solutions to $Pr(G|M_h)$) as follows:

$$Pr(G|M_l) = C \cdot \frac{\Gamma(a) \cdot \Gamma(b)\Gamma(c) \cdot \Gamma(d)}{\Gamma(a+b) \cdot -\Gamma(c+d)} - Pr(G|M_h) \quad (16)$$

It turned out that no matter which formula we used to solve $Pr(G|M_h)$, we can use Equation 16 to obtain $Pr(G|M_h)$.

5 Computational Complexity

Since we require that N_{11}, N_{12}, N_{21}, N_{22}, α_1, β_1, α_2 and β_2 be natural numbers, the gamma function simply becomes a factorial function: $\Gamma(x) = (x-1)!$. Since such numbers can become very large, it is convenient to use the logarithm of the gamma function and express Equations 2, 10, 11, 12, 13 and 16 in logarithmic form. The logarithm of the integer gamma function can be pre-computed and efficiently stored in an array as follows:

$lnGamma[1] = 0$
For $i = 2$ to n
  $lnGamma[i] = lnGamma[i-1] + ln(i-1)$

We then can use $lnGamma$ in solving the above equations. However, Equations 10, 11, 12 and 13 include a sum, which makes the use of the logarithmic form more involved. To deal with this issue, we can define function $lnAdd$, which takes two arguments x and y that are in logarithmic form and returns $ln(e^x + e^y)$. It does so in a way that preserves a good deal of numerical precision that could be lost if $ln(e^x + e^y)$ were calculated in a direct manner. This is done as follows:

$$lnAdd(x, y) = x + ln(1 + e^{(y-x)})$$

Now that we introduced functions $lnGamma$ and $lnAdd$, it is straightforward to evaluate Equations 2, 10, 11, 12, 13 and 16 in logarithmic form in order to preserve numerical precision.

Let us now analyze the overall computational complexity for computing the Bayesian score for a specific rule, as defined by Equation 1. Doing so requires computing $Pr(M_e|G)$, $Pr(M_h|G)$ and $Pr(M_l|G)$. $Pr(M_e|G)$ can be computed in $O(1)$ using Equation 2. $Pr(M_h|G)$

can be computed by applying Equation 10, Equation 11, Equation 12 or Equation 13. The computational complexity of these equations are $O(N_{22}+\beta_2)$, $O(N_{11}+\alpha_1)$, $O(N_{21}+\alpha_2)$, $O(N_{12}+\beta_1)$. Therefore, $Pr(M_h|G)$ can be computed in $O(min(N_{11}+\alpha_1, N_{12}+\beta_1, N_{21}+\alpha_2, N_{22}+\beta_2)$. $Pr(M_l|G)$ can be computed from $Pr(M_h|G)$ in $O(1)$ using Equation 16. By assuming that α_1, β_1, α_2, β_2 are bounded from above, the overall time complexity for computing the Bayesian score is $\boldsymbol{O(min(N_{11}, N_{12}, N_{21}, N_{22})}$.