Bayesian Model for Joint SNPs analysis

Gibbs version

The Bayesian model is designed to determine the association status of a gene, while considering known biological facts regarding the variables. Recall the determination rule that if all the SNPs on a gene have true relative risks equaled to 1, then the gene is considered as not associated with the disease. If there is least 1 SNP that has true relative risk higher than 1, then the gene is considered as associated with the disease.

Assume on a given gene, there are total J SNPs. Let G be a Bernoulli variable denoting gene associated statu swith probability b. The probability variable b ranges between 0 and 1 and follows a Beta distribution. Let H_j be a Bernoulli variable denoting associated status for each j^{th} SNP on the given gene.

Recall that R_j is the relative risk of the j^{th} SNP on a given gene. The prior distributions of these variables are as follows:

$$
\begin{aligned}
&f(A_j) \sim Beta(\alpha_A, \beta_A) \\
&f(M_j) \sim Beta(\alpha_M, \beta_M) \\
&R_j|H_j = 1 \sim Gamma(\alpha_{R_j}, \beta_{R_j}) \qquad where\ mode > 1 \\
&R_j|H_j = 0 \sim Gamma(\alpha_0, \beta_0) \qquad where\ mode = 1 \\
&H_j|G = 1 \sim Bernoulli(k \cdot p0) \\
&H_j|G = 0 \sim Bernoulli(p0) \\
&G \sim Bernoulli(b) \\
&b \sim Beta(\alpha_b, \beta_b)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
f(H_j|G) &= Bernoulli(H_j; k \cdot p0) \cdot G + Bernoulli(H_j; p0) \cdot (1 - G) \\
&= [(k\ p0)^{H_j}(1 - k\ p0)^{1-H_j}]^G \cdot [p0^{H_j}(1 - p0)^{1-H_j}]^{(1-G)} \\
&= (k\ p0)^{H_j G} \cdot (1 - k\ p0)^{(1-H_j)G} \cdot p0^{H_j(1-G)} \cdot (1 - p0)^{(1-H_j)(1-G)} \\
f(R_j|H_j) &= Gamma(\alpha_{R_j}, \beta_{R_j})^{H_j} \cdot Gamma(\alpha_0, \beta_0)^{(1-H_j)} \\
&= Gamma\Big(H_j\alpha_{R_j} + (1 - H_j)\alpha_0,\ H_j\beta_{R_j} + (1 - H_j)\beta_0\Big)
\end{aligned}
$$

- **Note**:
  - $p0$ is a small positive number indicating the percentage of risk SNPs giving a neutral gene.
  - However $k \cdot p0$ is fairly large, indicating the percentage of risk SNPs giving an associated

gene.

Let

$$
\begin{aligned}
&\mathbf{\Theta} = \{A_1, ..., A_J, M_1, ..., M_J, R_1, ..., R_J\} \\
&\mathbf{H} = \{H_1, ...H_J\} \\
&\mathbf{S} = \begin{bmatrix} S_{00(1)} & ... & S_{22(1)} \\ \vdots & & \vdots \\ S_{00(J)} & ... & S_{22(J)} \end{bmatrix}_{J\times 9}
\end{aligned}
$$

The joint posterior distribution is written as follows and the last equation is derived based on independence assumptions.

$$
\begin{aligned}
f(\mathbf{S}, \mathbf{\Theta}, \mathbf{H}, G, b) &= f(\mathbf{S}|\mathbf{\Theta}, \mathbf{H}, G, b) \cdot f(\mathbf{\Theta}, \mathbf{H}, G, b) \\
&= f(\mathbf{S}|\mathbf{\Theta}, \mathbf{H}, G, b) \cdot f(\mathbf{\Theta}|\mathbf{H}, G, b) \cdot f(\mathbf{H}, G, b) \\
&= f(\mathbf{S}|\mathbf{\Theta}, \mathbf{H}, G, b) \cdot f(\mathbf{\Theta}|\mathbf{H}, G, b) \cdot f(\mathbf{H}|G, b) \cdot f(G|b) \cdot f(b) \\
&= f(\mathbf{S}|\mathbf{\Theta}) \cdot f(\mathbf{\Theta}|\mathbf{H}) \cdot f(\mathbf{H}|G) \cdot f(G|b) \cdot f(b)
\end{aligned}
$$

where

$$f(\mathbf{S}|\mathbf{\Theta}) = \prod_{j=1}^{J} P_{00(j)}^{S_{00(j)}} ... P_{22(j)}^{S_{22(j)}}$$

$$\begin{aligned}
f(\mathbf{\Theta}|\mathbf{H}) &= \prod_{j=1}^{J} f(A_j|\mathbf{H}) f(M_j|\mathbf{H}) f(R_j|\mathbf{H}) \\
&= \prod_{j=1}^{J} f(A_j) f(M_j) f(R_j|\mathbf{H}) \\
&= \prod_{j=1}^{J} f(A_j) f(M_j) f(R_j|H_j) \\
&= \prod_{j=1}^{J} f(A_j) f(M_j) \cdot Gamma\Big(R_j; \alpha_{R_j}, \beta_{R_j}\Big)^{H_j} Gamma\Big(R_j; \alpha_0, \beta_0\Big)^{(1-H_j)} \\
&= \prod_{j=1}^{J} f(A_j) f(M_j) \cdot \Big[{R_j}^{\alpha_{R_j}-1} e^{-R_j \beta_{R_j}}\Big]^{H_j} \Big[{R_j}^{\alpha_0 - 1} e^{-R_j \beta_0}\Big]^{(1-H_j)}
\end{aligned}$$

$$\begin{aligned}
f(\mathbf{H}|G) &= \prod_{j=1}^{J} f(H_j|G) \\
&= \prod_{j=1}^{J} \Big[(k\ p0)^{H_j} (1 - k\ p0)^{1-H_j}\Big]^{G} \cdot \Big[p0^{H_j} (1 - p0)^{1-H_j}\Big]^{(1-G)} \\
&= \Big[(k\ p0)^{\sum H_j} (1 - k\ p0)^{J - \sum H_j}\Big]^{G} \cdot \Big[p0^{\sum H_j} (1 - p0)^{J - \sum H_j}\Big]^{(1-G)}
\end{aligned}$$

$$f(G|b) = b^G (1-b)^{1-G}$$

$$f(b) = \frac{\Gamma(\alpha_b + \beta_b)}{\Gamma(\alpha_b)\Gamma(\beta_b)} b^{\alpha_b - 1} (1-b)^{\beta_b - 1}$$

1. Given initial values $\mathbf{\Theta^{(0)}}, \mathbf{H^{(0)}}, G^{(0)}, b^{(0)}$

In general, assuming $\mathbf{\Theta^{(t)}}, \mathbf{H^{(t)}}, G^{(t)}, b^{(t)}$, where $t = 0, 1, 2, ...$

2. Update b.

Giving the joint posterior distribution above, fixed other parameters, there is:

$$\begin{aligned}
f(b|others) &\propto f(G|b) f(b) \\
&\propto b^{G + \alpha_b - 1} (1-b)^{\beta_b - G}
\end{aligned}$$

So conditioning other parameters at time t, $\mathbf{b^{(t)}} \sim \mathbf{Beta}(\alpha_{\mathbf{new}}, \beta_{\mathbf{new}})$.

where

$$\alpha_{new} = G^{(t)} + \alpha_b \qquad \beta_{new} = \beta_b - G^{(t)} + 1$$

3. Update G.

$$\begin{aligned}
f(G|others) &= \frac{f(\mathbf{S}, \mathbf{\Theta}, \mathbf{H}, G, b)}{f(\mathbf{S}, \mathbf{\Theta}, \mathbf{H}, b)} \\
&= \frac{f(\mathbf{S}, \mathbf{\Theta}, \mathbf{H}, G, b)}{f(\mathbf{S}, \mathbf{\Theta}, \mathbf{H}, G=0, b) + f(\mathbf{S}, \mathbf{\Theta}, \mathbf{H}, G=1, b)} \\
&= \frac{f(\mathbf{S}|\mathbf{\Theta})\ f(\mathbf{\Theta}|\mathbf{H})\ f(\mathbf{H}|G)\ f(G|b)\ f(b)}{f(\mathbf{S}|\mathbf{\Theta})f(\mathbf{\Theta}|\mathbf{H})f(\mathbf{H}|0)f(0|b)f(b) + f(\mathbf{S}|\mathbf{\Theta})f(\mathbf{\Theta}|\mathbf{H})f(\mathbf{H}|1)f(1|b)f(b)} \\
&= \frac{f(\mathbf{H}|G)f(G|b)}{f(\mathbf{H}|0)f(0|b) + f(\mathbf{H}|1)f(1|b)} \\
&= \frac{\left[b \cdot (k\ p0)^{\sum H_j}(1-k\ p0)^{J-\sum H_j}\right]^G \cdot \left[(1-b)\cdot p0^{\sum H_j}(1-p0)^{J-\sum H_j}\right]^{(1-G)}}{b \cdot (k\ p0)^{\sum H_j}(1-k\ p0)^{J-\sum H_j} + (1-b)\cdot p0^{\sum H_j}(1-p0)^{J-\sum H_j}}
\end{aligned}$$

Therefore, $\mathbf{G^t} \sim \mathbf{Bernoulli(prob\_G)}$, conditioning on $b^{(t+1)}$ and other parameters at time (t). Where

$$prob\_G = \frac{b \cdot (k\ p0)^{\sum H_j}(1-k\ p0)^{J-\sum H_j}}{b \cdot (k\ p0)^{\sum H_j}(1-k\ p0)^{J-\sum H_j} + (1-b)\cdot p0^{\sum H_j}(1-p0)^{J-\sum H_j}}$$

4. Update H.

Similarly as above, we have

$$\begin{aligned}
f(H_l|others) &= \frac{f(\mathbf{S}|\mathbf{\Theta})\ f(\mathbf{\Theta}|\mathbf{H})\ f(\mathbf{H}|G)\ f(G|b)\ f(b)}{f(\mathbf{S}|\mathbf{\Theta})f(\mathbf{\Theta}|\mathbf{H_l}=\mathbf{0})f(\mathbf{H_l}=\mathbf{0}|G)f(G|b)f(b) + f(\mathbf{S}|\mathbf{\Theta})f(\mathbf{\Theta}|\mathbf{H_l}=\mathbf{1})f(\mathbf{H_l}=\mathbf{1}|G)f(G|b)f(b)} \\
&= \frac{f(\mathbf{\Theta}|\mathbf{H})f(\mathbf{H}|G)}{f(\mathbf{\Theta}|\mathbf{H_l}=\mathbf{0})f(\mathbf{H_l}=\mathbf{0}|G) + f(\mathbf{\Theta}|\mathbf{H_l}=\mathbf{1})\ f(\mathbf{H_l}=\mathbf{1}|G)} \\
&= \frac{\prod_{j=1}^J f(A_j)f(M_j)f(R_j|H_j)f(H_j|G)}{\left[\prod_{j=1}^J f(A_j)f(M_j)f(R_j|H_j)f(H_j|G)\right]_{H_l=0} + \left[\prod_{j=1}^J f(A_j)f(M_j)f(R_j|H_j)f(H_j|G)\right]_{H_l=1}} \\
&= \frac{f(R_l|H_l)f(H_l|G)}{f(R_l|0)f(0|G) + f(R_l|1)f(1|G)} \\
&= \frac{\left[Gamma(R_l;\alpha_{R_l},\beta_{R_l})\ (kp0)^G p0^{(1-G)}\right]^{H_l}\left[Gamma(R_l;\alpha_0,\beta_0)\ (1-kp0)^G(1-p0)^{(1-G)}\right]^{(1-H_l)}}{Gamma(R_l;\alpha_0,\beta_0)\ (1-kp0)^G(1-p0)^{(1-G)} + Gamma(R_l;\alpha_{R_l},\beta_{R_l})\ (kp0)^G p0^{(1-G)}}
\end{aligned}$$

Therefore, $\mathbf{H_l^t} \sim \mathbf{Bernoulli(prob\_H_l)}$, conditioning on $b^{(t+1)}, G^{(t+1)}$ and other parameters at time (t).

Where

$$prob\_H_l = \frac{Gamma(R_l; \alpha_{R_l}, \beta_{R_l})\ (kp0)^G p0^{(1-G)}}{Gamma(R_l; \alpha_0, \beta_0)\ (1-kp0)^G (1-p0)^{(1-G)} + Gamma(R_l; \alpha_0, \beta_0)\ (1-kp0)^G (1-p0)^{(1-G)}}$$

5. Update R_l.

For j from 1 to J, update $R_j^{(t)}$ conditioned on $b^{(t+1)}, G^{(t+1)}, \mathbf{H^{(t+1)}}, R_1^{(t+1)}, ..., R_{j-1}^{(t+1)}$ and other variables at state t.

i) Proposal function:

$$f(R_j^*) \sim Gamma(shape = 1 + 10R_j^{(t)}, rate = 10)$$

By contructing shape using $R_j^{(t)}$, the mode of proposal function will be $(shape - 1)/rate = R_j^{(t)}$ and variance is $shape/rate^2 = 0.1R_j^{(t)} + 0.01$. And there are:

$$f(R_j^*|R_j^{(t)}) = dGamma(R_j^*; shape = 1 + 10R_j^{(t)}, rate = 10)$$
$$f(R_j^{(t)}|R_j^*) = dGamma(R_j^{(t)}; shape = 1 + 10R_j^*, rate = 10)$$

ii) Acceptance rate:

$$\begin{aligned} r &= \frac{lik(R_j^*)/f(R_j^*|R_j^{(t)})}{lik(R_j^{(t)})/f(R_j^{(t)}|R_j^*)} \\ &\propto \frac{[P_{00(j)}^{S_{00(j)}} ... P_{22(j)}^{S_{22(j)}}]|_{R_j=R_j^*} \cdot f(R_j^*|H_j)}{[P_{00(j)}^{S_{00(j)}} ... P_{22(j)}^{S_{22(j)}}]|_{R_j=R_j^{(t)}} \cdot f(R_j^{(t)}|H_j)} \cdot \frac{f(R_j^{(t)}|R_j^*)}{f(R_j^*|R_j^{(t)})} \end{aligned}$$

The proportion in each cell is relatively small and may be too small to calculate when taking the power of sample sizes. Thus to calculate the ratio of proportions first, then to take power will be better.

iii) Update or maintain values.
Define acceptance indicator: $I = rBernoulli(r, min(1))$. If acceptance indicator equals to 1, update value as R_j^*. Otherwise stay at the current value $R_j^{(t)}$.

$$R_j^{(t+1)} = R_j^{*I} {R_j^{(t)}}^{1-I}$$

6. Update M_j.

For j from 1 to J, update $M_j^{(t)}$ conditioned on $b^{(t+1)}, G^{(t+1)}, \mathbf{H^{(t+1)}}, R_1^{(t+1)}, ..., R_J^{(t+1)}, M_1^{(t+1)}, ..., M_{j-1}^{(t+1)}$ and other variables at state t.

i) Proposal function:

$$f(R_j^*) \sim Beta(shape = \frac{1 + 100M_j^{(t)}}{1 - M_j^{(t)}}, rate = 102)$$

By constructing shape using $M_j^{(t)}$, the mode of proposal function (Beta) will be $(shape - 1)/(shape + rate - 2) = M_j^{(t)}$ and variance will be $\frac{shape \cdot rate}{(shape+rate)^2(shape+rate+1)} = \frac{102(1+100M_j^{(t)})(1-M_j^{(t)})^2}{(103-2M_j^{(t)})^2(104-3M_j^{(t)})}$. Then there are

$$f(M_j^*|M_j^{(t)}) = dBeta(M_j^*; shape = \frac{1 + 100M_j^{(t)}}{1 - M_j^{(t)}}, rate = 102)$$
$$f(M_j^{(t)}|M_j^*) = dBeta(M_j^{(t)}; shape = \frac{1 + 100M_j^*}{1 - M_j^*}, rate = 102)$$

ii) Acceptance rate:

$$\begin{aligned} r &= \frac{lik(M_j^*)/f(M_j^*|M_j^{(t)})}{lik(M_j^{(t)})/f(M_j^{(t)}|M_j^*)} \\ &\propto \frac{[P_{00(j)}^{S_{00(j)}} ... P_{22(j)}^{S_{22(j)}}]|_{M_j = M_j^*}}{[P_{00(j)}^{S_{00(j)}} ... P_{22(j)}^{S_{22(j)}}]|_{M_j = M_j^{(t)}}} \cdot \frac{f(M_j^{(t)}|M_j^*)}{f(M_j^*|M_j^{(t)})} \end{aligned}$$

iii) Update or maintain values.

Define acceptance indicator: $I = rBernoulli(r, min(1))$. If acceptance indicator equals to 1, update value as M_j^*. Otherwise stay at the current value $M_j^{(t)}$.

$$M_j^{(t+1)} = {M_j^*}^I {M_j^{(t)}}^{1-I}$$

7. Update A_j.

For j from 1 to J, update $A_j^{(t)}$ conditioned on $b^{(t+1)}, G^{(t+1)}, \mathbf{H^{(t+1)}}, R_1^{(t+1)}, ..., R_J^{(t+1)}, M_1^{(t+1)}, ..., M_J^{(t+1)}, A_1^{(t+1)}, ..., A$ and other variables at state t.

i) Proposal function:

$$f(A_j^*) \sim Beta(shape = \frac{1 + 100A_j^{(t)}}{1 - A_j^{(t)}}, rate = 102)$$

By constructing shape using $A_j^{(t)}$, the mode of proposal function (Beta) will be $(shape - 1)/(shape + rate - 2) = A_j^{(t)}$ and variance will be $\frac{shape \cdot rate}{(shape+rate)^2(shape+rate+1)} = \frac{102(1+100A_j^{(t)})(1-A_j^{(t)})^2}{(103-2A_j^{(t)})^2(104-3A_j^{(t)})}$. Then there are

$$f(A_j^*|A_j^{(t)}) = dBeta(A_j^*; shape = \frac{1+100A_j^{(t)}}{1-A_j^{(t)}}, rate = 102)$$
$$f(A_j^{(t)}|A_j^*) = dBeta(A_j^{(t)}; shape = \frac{1+100A_j^*}{1-A_j^*}, rate = 102)$$

ii) Acceptance rate:

$$\begin{aligned} r &= \frac{lik(A_j^*)/f(A_j^*|A_j^{(t)})}{lik(A_j^{(t)})/f(A_j^{(t)}|A_j^*)} \\ &\propto \frac{[P_{00(j)}^{S_{00(j)}}...P_{22(j)}^{S_{22(j)}}]|_{A_j=A_j^*}}{[P_{00(j)}^{S_{00(j)}}...P_{22(j)}^{S_{22(j)}}]|_{A_j=A_j^{(t)}}} \cdot \frac{f(A_j^{(t)}|A_j^*)}{f(A_j^*|A_j^{(t)})} \end{aligned}$$

iii) Update or maintain values.

Define acceptance indicator: $I = rBernoulli(r, min(1))$. If acceptance indicator equals to 1, update value as A_j^*. Otherwise stay at the current value $A_j^{(t)}$.

$$A_j^{(t+1)} = A_j^{*I} A_j^{(t)^{1-I}}$$