

## SUPPLEMENTAL MATERIAL 1

for the paper “Position dependencies in transcription factor binding sites”  
A. Tomovic and E. J. Oakeley

### Derivation 1. Derivation of formula (14) for calculating the Bayesian factor BF

In order to test dependencies between positions in transcription factor binding sites by Bayesian hypothesis testing we have to calculate the Bayes factor  $BF(H_0; H_1)$ , which is similar to (Minka, 2003; Zhou and Liu, 2004) except that we use different notation and make a small modification.

$$BF(H_0; H_1) = \frac{P(B_i, B_j | H_0)P(H_0)}{P(B_i B_j | H_1)P(H_1)} \quad (1)$$

If we assume that  $P(H_0) = P(H_1) = 0.5$  then (1) will be

$$BF(H_0; H_1) = \frac{P(B_i, B_j | H_0)}{P(B_i B_j | H_1)} \quad (2)$$

Under the null hypothesis, we have  $P(B_i, B_j) = P(B_i)P(B_j)$ , and (2) will be

$$BF(H_0; H_1) = \frac{P(B_i | H_0)P(B_j | H_1)}{P(B_i B_j | H_1)} \quad (3)$$

Then, using the fact that:

$$P(B_i | H_0) = \int P(B_i, \bar{p} | H_0) \quad (4)$$

Where  $\bar{p}$  is a vector of  $[P(a,i), P(c,i), P(g,i), P(t,i)]$

A conjugate prior for  $\bar{p}$  is the Dirichlet distribution:

$$P(\bar{p} | \alpha) \sim Dir(\alpha_a, \alpha_c, \alpha_g, \alpha_t) = \frac{\Gamma(\sum_{b_i} \alpha_{b_i})}{\prod_{b_i} \Gamma(\alpha_{b_i})} \prod_{b_i} P(b_i, i)^{\alpha_{b_i} - 1} \quad (5)$$

Where  $P(b_i, i) > 0$  and  $\sum P(b_i, i) = 1$ . Given a Dirichlet prior, the joint distribution of  $B_i$  and  $\bar{p}$  is:

$$P(B_i, \bar{p} | \alpha) = \frac{\Gamma(\sum_{b_i} \alpha_{b_i})}{\prod_{b_i} \Gamma(\alpha_{b_i})} \prod_{b_i} P(b_i, i)^{N(b_i, i) + \alpha_{b_i} - 1} \quad (6)$$

and the posterior is:

$$P(\bar{p} | B_i, \alpha) \sim Dir(N(b_i, i) + \alpha_{b_i}) \quad (7)$$

and, finally, we can calculate:

$$P(B_i | H_0) = \int P(B_i, \bar{p} | H_0) = \frac{\Gamma(\sum_{b_i} \alpha_{b_i})}{\Gamma(n + \sum_{b_i} \alpha_{b_i})} \prod_{b_i} \frac{\Gamma(N(b_i, i) + \alpha_{b_i})}{\Gamma(\alpha_{b_i})} \quad (8)$$

and likewise for  $P(B_j | H_0)$ .

Then, we need to calculate  $P(B_i B_j | H_1)$  and this is:

$$P(B_i B_j | H_1) = \int_{\hat{p}} P(B_i B_j, \hat{p} | H_1) \quad (9)$$

Where  $\hat{p}$  is a vector of [P(a,a,i,j), P(a,c,i,j), ... ,P(t,t,i,j)]

A conjugate prior for  $\hat{p}$  is the Dirichlet distribution:

$$P(\hat{p} | \alpha) \sim Dir(\alpha_{aa}, \alpha_{ac}, \dots, \alpha_{tt}) = \frac{\Gamma(\sum_{b_i, b_j} \alpha_{b_i b_j})}{\prod_{b_i, b_j} \Gamma(\alpha_{b_i b_j})} \prod_{b_i, b_j} P(b_i, b_j, i)^{\alpha_{b_i b_j} - 1} \quad (10)$$

Where  $P(b_i, b_j, i, j)$  and  $\sum_{b_i, b_j} P(b_i, b_j, i, j) = 1$ . Given a Dirichlet prior, the joint distribution of  $B_i B_j$

and  $\hat{p}$  is:

$$P(B_i B_j, \hat{p} | \alpha) = \frac{\Gamma(\sum_{b_i, b_j} \alpha_{b_i b_j})}{\prod_{b_i, b_j} \Gamma(\alpha_{b_i b_j})} \prod_{b_i, b_j} P(b_i, b_j, i, j)^{N(b_i, b_j, i, j) + \alpha_{b_i b_j} - 1} \quad (11)$$

And the posterior is:

$$P(\hat{p} | B_i B_j, \alpha) \sim Dir(N(b_i, b_j, i, j) + \alpha_{b_i b_j}) \quad (12)$$

so we can calculate:

$$P(B_i B_j | H_1) = \int_{\hat{p}} P(B_i B_j, \hat{p} | H_1) = \frac{\Gamma(\sum_{b_i, b_j} \alpha_{b_i b_j})}{\Gamma(n + \sum_{b_i, b_j} \alpha_{b_i b_j})} \prod_{b_i, b_j} \frac{\Gamma(N(b_i, b_j, i, j) + \alpha_{b_i b_j})}{\Gamma(\alpha_{b_i b_j})} \quad (13)$$

and thus BF can be calculated:

$$BF(H_0; H_1) = \frac{\Gamma(\sum_b \alpha_b)}{\Gamma(n + \sum_b \alpha_b)} \left( \prod_b \frac{\Gamma(N(b, i) + \alpha_b)}{\Gamma(\alpha_b)} \right) \left( \frac{\Gamma(\sum_b \alpha_b)}{\Gamma(n + \sum_b \alpha_b)} \right)^* \quad (14)$$

$$* \left( \prod_b \frac{\Gamma(N(b, j) + \alpha_b)}{\Gamma(\alpha_b)} \right) / \left( \prod_{b_i, b_j} \frac{\Gamma(N(b_i, b_j, i, j) + \alpha_{b_i b_j})}{\Gamma(\alpha_{b_i b_j})} \frac{\Gamma(\sum_{b_i, b_j} \alpha_{b_i b_j})}{\Gamma(n + \sum_{b_i, b_j} \alpha_{b_i b_j})} \right)$$

Because we choose  $\alpha_{b_i} = \sum_{b_j} \alpha_{b_i b_j}$ , we have:

$$BF(H_0; H_1) = \frac{\Gamma(\sum_{b_i, b_j} \alpha_{b_i b_j})}{\Gamma(n + \sum_{b_i, b_j} \alpha_{b_i b_j})} \prod_{b_i} \frac{\Gamma(N(b_i, i) + \alpha_{b_i})}{\Gamma(\alpha_{b_i})} \prod_{b_j} \frac{\Gamma(N(b_j, j) + \alpha_{b_j})}{\Gamma(\alpha_{b_j})} \prod_{b_i, b_j} \frac{\Gamma(\alpha_{b_i b_j})}{\Gamma(N(b_i, b_j, i, j) + \alpha_{b_i b_j})} \quad (15)$$

The calculation should include only bases  $b_i, b_j$  for which  $N(b_i, i) \neq 0$  and  $N(b_j, j) \neq 0$ .

**Derivation 2. Derivation of formula (15): the relationship between BF and mutual information**

It is possible to show (like in (Minka, 2003)) that there is a relationship between the Bayes factor BF and mutual information  $M_{ij}$  if we choose a uniform prior, i.e.  $\alpha_k = 1$

Using the fact that  $\Gamma(1) = 1$ , and the approximation

$$\frac{\Gamma(k)}{\Gamma(n+k)} \approx \frac{\Gamma(k)}{\Gamma(n+1)n^{k-1}} \approx \frac{1}{\Gamma(n+1)}$$

and Stirling's approximation that  $\log \Gamma(x+1) \approx x \log x - x$ , we get:

$$\begin{aligned} \log_2(BF(H_0; H_1)) &\approx n \sum_{b_i} \frac{N(b_i, i)}{n} \log_2 \frac{N(b_i, i)}{n} + n \sum_{b_j} \frac{N(b_j, j)}{n} \log_2 \frac{N(b_j, j)}{n} - n \sum_{b_i, b_j} \frac{N(b_i, b_j, i, j)}{n} \log_2 \frac{N(b_i, b_j, i, j)}{n} \\ &= -n \sum_{b_i, b_j} \frac{N(b_i, b_j, i, j)}{n} \log_2 \frac{N(b_i, b_j, i, j)}{n} \frac{n}{N(b_i, i)} \frac{n}{N(b_j, j)} = \\ &= -n \sum_{b_i, b_j} P(b_i, b_j, i, j) \log_2 \frac{P(b_i, b_j, i, j)}{P(b_i, i)P(b_j, j)} = \\ &= -nM_{ij} \end{aligned}$$

Mutual information and the Bayes factor become more closely related as the sample size  $n$  gets higher (because of the approximation of Stirling's formula).

**Analytic formula for calculating the hypothetical minimum and maximum for  $S_{old}$  and  $S_{new}$**

$S_{old}^{min}$  and  $S_{old}^{max}$  are hypothetically the minimum and maximum for  $S_{old}$ , and  $S_{new}^{min}$  and  $S_{new}^{max}$  are hypothetically the minimum and maximum for  $S_{new}$  calculated by :

$$S_{old}^{min} = \sum_{i=1}^k \min_b W_{b,i}$$

$$S_{old}^{max} = \sum_{i=1}^k \max_b W_{b,i}$$

$$S_{new}^{min} = \sum_{i=1}^{k_1} \min_b W_{b,i} + \sum_{i=1}^{k_2} \min_{b_1, b_2} W_{b_1, b_2, j_i, j_{i+1}} + \dots + \sum_{i=1}^{k_m} \min_{b_1, \dots, b_{i+m-1}} W_{b_1, \dots, b_{i+m-1}, j_1, \dots, j_{i+m-1}}$$

$$S_{new}^{max} = \sum_{i=1}^{k_1} \max_b W_{b,i} + \sum_{i=1}^{k_2} \max_{b_1, b_2} W_{b_1, b_2, j_i, j_{i+1}} + \dots + \sum_{i=1}^{k_m} \max_{b_1, \dots, b_{i+m-1}} W_{b_1, \dots, b_{i+m-1}, j_1, \dots, j_{i+m-1}}$$

**REFERENCES**

Minka, T. (2003) Bayesian inference, entropy, and the multinomial distribution. Technical report (Microsoft research).

Zhou, Q. and Liu, J.S. (2004) Modeling within-motif dependence for transcription factor binding site predictions, *Bioinformatics*, 20, 909-916.