A Supplemental Note on the paper "Modeling within-motif dependence for TFBS predictions" by Zhou and Liu

Qing Zhou

In this note, we give the details for calculating the BF in equation (5) of Zhou and Liu (2004). Since correlated pairs are non-overlapping, it suffices to illustrate the calculation by considering only two positions of a motif. Suppose we observe di-nucleotides $\mathbf{X} = \{X_{n1}X_{n2}\}_{n=1}^{N}$. Let $\mathbf{X}_k = \{X_{nk}\}_{n=1}^{N}$ for $k = 1, 2$. Denote the marginal counts of $\mathbf{X}_k$ by $N_k = [N_k(A), \cdots, N_k(T)]$ and the joint counts of $\mathbf{X}$ by $N_{12} = [N_{12}(A, A), \cdots, N_{12}(T, T)]$. Let H_0 denote the hypothesis that the two positions are independent, and let H_1 denote that they are correlated. Then the Bayes factor $BF(H_1; H_0)$ is defined as

$$BF(H_1; H_0) = \frac{P(\mathbf{X}|H_1)}{P(\mathbf{X}|H_0)}, \tag{1}$$

where $P(\mathbf{X}|H_0) = P(\mathbf{X}_1|H_0)P(\mathbf{X}_2|H_0)$ by the independence assumption. Then one can calculate

$$\begin{aligned} P(\mathbf{X}|H_1) &= \int_{\Theta_{12}} P(\mathbf{X}|\Theta_{12})\pi(\Theta_{12}|H_1)d\Theta_{12} \qquad (2) \\ &= \frac{\Gamma(\sum_{i,j}\alpha_{12}(i,j))}{\prod_{i,j}\Gamma(\alpha_{12}(i,j))} \cdot \frac{\prod_{i,j}\Gamma(N_{12}(i,j)+\alpha_{12}(i,j))}{\Gamma(N+\sum_{i,j}\alpha_{12}(i,j))}, \end{aligned}$$

where $\pi(\Theta_{12}|H_1) = Dir(\alpha_{12}(A, A), \cdots, \alpha_{12}(T, T))$ is the prior distribution for Θ_{12} under H_1. Similarly one can calculate, for $k = 1, 2$,

$$P(\mathbf{X}_k|H_0) = \frac{\Gamma(\sum_j \alpha_k(j))}{\prod_j \Gamma(\alpha_k(j))} \cdot \frac{\prod_j \Gamma(N_k(j) + \alpha_k(j))}{\Gamma(N + \sum_j \alpha_k(j))}, \tag{3}$$

where α_k is the paramter for the prior Dirichlet distributions under H_0. We recommend to set $\alpha_1(i) = \sum_j \alpha_{12}(i, j)$ and $\alpha_2(j) = \sum_i \alpha_{12}(i, j)$ in the prior distributions. Thus $BF(H_1; H_0)$ in equation (1) can be calculated by plugging in equations (2) and (3).