

# CENTIPEDE 2.0

ANIL RAJ AND HEEJUNG SHIM

## 1. OVERVIEW

CENTIPEDE aims to infer motif sites bound by transcription factors based on the DNase I cleavage patterns measured from DNase-Seq assays <sup>1</sup>. The model relies on two assumptions: (1) sites bound by transcription factors have higher DNase I sensitivity than unbound sites, and (2) each transcription factor has a characteristic DNase I cleavage profile at bound sites.

Given a putative binding site, CENTIPEDE models the number of reads mapped to each base pair as a mixture of two distributions, where the mixing proportions capture the probability of the factor being bound. Specifically, conditional on being bound, the total number of reads are modelled as drawn from a negative binomial distribution and the read profile, conditional on the total number of reads, are modelled as drawn from a multinomial distribution.

$$p(X_n, T_n^X | Z_n = 1) = p(X_n | Z_n = 1, T_n^X) p(T_n^X | Z_n = 1) \quad (1.1)$$

$$= \text{mult}(\pi; T_n^X) \text{NegBin}(\alpha, \tau) \quad (1.2)$$

$$p(X_n, T_n^X | Z_n = 0) = p(X_n | Z_n = 0, T_n^X) p(T_n^X | Z_n = 0) \quad (1.3)$$

$$= \text{mult}(\pi_o; T_n^X) \text{NegBin}(\alpha_o, \tau_o), \quad (1.4)$$

where  $X_n \in \mathbb{N}_0^L$ ,  $T_n^X = \sum_l X_{nl}$ ,  $\pi \in \mathbb{S}^L$ ,  $\mathbb{S}^L$  is the standard  $L$ -simplex,  $\pi_o = \frac{1}{L} \mathbf{1}^L$ ,  $\mathbf{1}^L$  is the  $L$ -dimensional vector of ones,  $\alpha, \alpha_o \in \mathbb{R}^+$ , and  $\tau, \tau_o \in [0, 1]$ .  $L$  is the length of the site around the binding motif <sup>2</sup>.

A key limitation of this generative model is that it does not appropriately model the correlation between number of DNase I reads mapped to two different positions around the motif, conditional on the site being bound <sup>3</sup>. One way to model additional correlation structure between different positions relative to the motif would be to place a logistic-normal prior on the multinomial parameter  $\pi$ ; the covariance matrix in the logistic-normal distribution can then be estimated from the data. However, the non-conjugacy of this prior makes exact inference intractable, leading us to use approximate techniques like variational inference to compute the posterior probabilities of the latent variables. This approach will be derived and discussed in the second half of this document.

Alternately, we propose to replace the multinomial part of the CENTIPEDE model with a multiscale generative model from which a multiscale transformation of the DNase I read counts are assumed to be drawn. Here, we derive the likelihood of the data and the posterior probability of

---

*Date:* May 18, 2013.

<sup>1</sup>**Q: I'm a bit confused with how CENTIPEDE works. It pre-selects a set of motif matches by using a known motif or k-mer and run CENTIPEDE model? So DNase I cleavage patterns provides information on actual binding conditional on motif match?**

<sup>2</sup>**Q: In CENTIPEDE, do they provide how to select  $L$  in advance. Or results are not sensitive to choice of  $L$ ?**

<sup>3</sup>**Comment: Where does the correlation come from? If it's because underlying signals are smooth (i.e. spatially structured), I'd argue that a multi-scale poisson model can model spatially structured signals. I'd like to use 'spatially structured or smooth' rather than 'correlation'. I think that a logistic-normal prior can be considered to model spatially structured signals**

the parameters given the data. See Kolaczyk (1999) for a detailed discussion of this inhomogenous Poisson process model <sup>4</sup>.

Alternately, we propose to replace the multinomial part of the CENTIPEDE model with a hierarchical multi-scale poisson model. Here, we describe the model and derive maximum likelihood estimators for parameters. See Kolaczyk (1999) for a detailed discussion of this inhomogenous multi-scale poisson process model.

## 2. POISSON-BINOMIAL MODEL

Specifically, keeping Kolaczyk's notation for the parameters, let  $Y_{njk}$  be the multiscale transformation of the data  $X_{nl}$ , where  $j \in \{0, \dots, J-1\}$ ,  $J = \log_2 L$ , and,  $k \in \{0, \dots, 2^j - 1\}$ . Note that, the range of the  $k$  index depends on the value of the  $j$  index. For each site  $n$ , the transformation is given as

$$Y_{J-1,k} = X_{2k} + X_{2k+1} \quad (2.1)$$

$$Y_{j,k} = Y_{j+1,2k} + Y_{j+1,2k+1}, \quad j \in \{0, \dots, J-2\}. \quad (2.2)$$

Conditional on the total number of reads at a site, the likelihood function of  $Y_n$  <sup>5</sup> factorizes as follows:

$$p(Y_n | Z_n = 1, T_n^X, R_n) = \prod_{j,k} p(Y_{njk} | Z_n = 1, T_{njk}, R_{jk}), \quad (2.3)$$

where

$$R_{jk} = \gamma_{jk} \delta\left(\frac{1}{2}\right) + (1 - \gamma_{jk}) \delta(B_{jk}) \quad (2.4)$$

$$p(\gamma_{jk}) = \text{Ber}(\pi_j), \quad (2.5)$$

$$(2.6)$$

and,  $\delta(a)$  is the Dirac delta function centered at  $a$ .

Conditional on the total number of reads at a site, the likelihood function of parameter  $R = (R_{jk})$  factorizes as follows:

$$p(X_n | Z_n = 1, T_n^X, R) = \prod_{j,k} p(Y_{njk} | Z_n = 1, T_{njk}, R_{jk}), \quad (2.7)$$

where

$$R_{jk} = \gamma_{jk} \delta\left(\frac{1}{2}\right) + (1 - \gamma_{jk}) \delta(B_{jk}) \quad (2.8)$$

$$p(\gamma_{jk}) = \text{Ber}(\pi_j), \quad (2.9)$$

$$(2.10)$$

and,  $\delta(a)$  is the Dirac delta function centered at  $a$ .

For each scale  $j$ , the product is only over even values of the index  $k$ . For ease of notation, we introduce a variable  $T_{njk} = Y_{n,j-1,k/2}$ , the total number of reads at a coarser resolution from which

---

<sup>4</sup>Comment: Ummm. I think we should change this paragraph, but I wasn't sure how it should be corrected. First, precisely speaking, there is no transformed data. Instead of transforming data, parameters are transformed (re-parameterized)? Kolaczyk (1999) called his model 'A Bayesian multiscale model', but I'm not sure if we want to use 'Bayesian' in our approach. What about 'a hierarchical multi-scale poisson model'? I made changes in the next paragraph...

<sup>5</sup>Comment: I think this part is a bit tricky. Likelihood function is about parameters. I made changes in the next paragraph, but I'm not sure if this is the best. Moreover, I understand why you use  $R_n$  instead of  $R$ . But, then, we should use  $R_{njk}$  instead of  $R_{jk}$ . For now, I start with  $R$  and later mention that we allow each site to have own  $R_n$ ?

reads at the finer resolution are drawn. Therefore, at a given scale and location,  $Y_{njk}$  is effectively drawn from a mixture of a binomial distribution, with parameter  $\frac{1}{2}$  and a binomial distribution with parameter  $B_{jk}$ .

We will derive and explore the following three variations of the above model <sup>6</sup>:

- (1) There is NO prior on  $B_{jk}$ , i.e., point estimates for  $B_{jk}$  will be computed by maximizing the likelihood of the model.
- (2) There is a prior on  $B_{jk}$ ,  $B_{jk} \sim \text{beta}(\mu_j, \mu_j)$ .
- (3) Overdispersion is accounted for by allowing for site-specific values for  $R$ , drawn from a beta distribution with scale and location dependent precisions.

Specifically, the three variations are <sup>7</sup>:

$$\begin{aligned}
 (1) \quad & Y_{njk}|T_{njk}, R_{jk} \sim \text{binom}(R_{jk}; T_{njk}) \\
 & R_{jk} = \gamma_{jk} \delta\left(\frac{1}{2}\right) + (1 - \gamma_{jk}) \delta(B_{jk}) \\
 & \gamma_{jk} \sim \text{Ber}(\pi_j) \\
 \text{Thus, } & Y_{njk}|T_{njk}, \pi_j, B_{jk} \sim \pi \text{binom}\left(\frac{1}{2}; T_{njk}\right) + (1 - \pi) \text{binom}(B_{jk}; T_{njk}) \\
 (2) \quad & Y_{njk}|T_{njk}, R_{jk} \sim \text{binom}(R_{jk}; T_{njk}) \\
 & R_{jk} = \gamma_{jk} \delta\left(\frac{1}{2}\right) + (1 - \gamma_{jk}) \delta(B_{jk}) \\
 & \gamma_{jk} \sim \text{Ber}(\pi_j) \\
 & B_{jk} \sim \text{beta}(\mu_j, \mu_j) \\
 \text{Thus, } & Y_{njk}|T_{njk}, \pi_j, B_{jk} \sim \pi \text{binom}\left(\frac{1}{2}; T_{njk}\right) + (1 - \pi) \text{BetaBinom}(\mu_j, \mu_j; T_{njk}) \\
 (3) \quad & Y_{njk}|T_{njk}, R_{njk} \sim \text{binom}(R_{njk}; T_{njk}) \\
 & R_{njk} \sim \text{beta}(R_{jk}, \tau_{jk}) \\
 & R_{jk} = \gamma_{jk} \delta\left(\frac{1}{2}\right) + (1 - \gamma_{jk}) B_{jk} \\
 & \gamma_{jk} \sim \text{Ber}(\pi_j) \\
 \text{Thus, } & Y_{njk}|T_{njk}, \pi_j, B_{jk} \sim \pi \text{BetaBinom}\left(\frac{1}{2} \tau_{jk}, \frac{1}{2} \tau_{jk}; T_{njk}\right) \\
 & + (1 - \pi) \text{BetaBinom}(B_{jk} \tau_{jk}, (1 - B_{jk}) \tau_{jk}; T_{njk})
 \end{aligned}$$

In the framework of CENTIPEDE, we now have two sets of latent variables: one specifying whether a factor is bound at a site or not, and the other specifying the degree of smoothness in DNase I cleavage rates at different scales and locations around a specific motif. Let us assume that the latent variable  $Z_n$  is observed to be 1 for a set of sites. Maximum likelihood estimates for the parameters in each model can be calculated by maximizing the lower bound on the likelihood,

---

<sup>6</sup>I'm not sure if we want to use prior/posterior in the description for the model. So I would make changes.. (1) no variation (2) Assume  $B_{jk}$  is sampled from a underlying distribution  $\text{beta}(\mu_j, \mu_j)$ . (3) To account for overdispersion across sites, we allow for site-specific values for  $R$ , drawn from a beta distribution with scale and location dependent mean and variance.

<sup>7</sup>Comment: After mean/precision conversation with you: in your notation for beta distribution, two parameters (mean/precision) indicate  $\alpha/(\alpha + \beta)$  and  $\alpha + \beta$ . I wouldn't use precision for  $\alpha + \beta$  because precision is  $1/\text{variance}$ . Also I would define what each parameter means for each distribution like  $\alpha/(\alpha + \beta)$  and  $\alpha + \beta$  for beta and  $\alpha$  and  $\beta$  for betabinomial.

obtained by proposing a family of posterior distributions  $q(\gamma_{jk})$ , using the EM algorithm <sup>8</sup>.

$$q(\gamma_{jk}) = \text{Ber}(\tilde{\gamma}_{jk}) \quad (2.11)$$

**2.1. Model 1.** For this model, the lower bound to the log-likelihood can be derived as follows:

$$\mathcal{L} = \sum_{n,j,k} \log p(Y_{njk} | \pi_j, B_{jk}; T_{jk}) \quad (2.12)$$

$$= \sum_{n,j,k} \log \sum_{\gamma_{jk}} p(Y_{njk} | \gamma_{jk}, B_{jk}; T_{njk}) p(\gamma_{jk} | \pi_j) \quad (2.13)$$

$$\geq \sum_{n,j,k} \sum_{\gamma_{jk}} q(\gamma_{jk}) \left( \log p(Y_{njk} | \gamma_{jk}, B_{jk}; T_{njk}) + \log \frac{p(\gamma_{jk} | \pi_j)}{q(\gamma_{jk})} \right) \quad (2.14)$$

$$= \sum_{n,j,k} \mathbf{E}_{q(\gamma)} [\log p(Y_{njk} | \gamma_{jk}; T_{njk})] + \mathbf{E}_{q(\gamma)} \left[ \log \frac{p(\gamma_{jk} | \pi_j)}{q(\gamma_{jk})} \right] \quad (2.15)$$

$$= \sum_{n,j,k} \mathbf{E}_{q(\gamma)} \left[ \gamma_{jk} \log p(Y_{njk} | \frac{1}{2}; T_{njk}) + (1 - \gamma_{jk}) \log p(Y_{njk} | B_{jk}; T_{njk}) \right] + \mathbf{E}_{q(\gamma)} \left[ \log \frac{p(\gamma_{jk} | \pi_j)}{q(\gamma_{jk})} \right] \quad (2.16)$$

$$= \sum_{n,j,k} \tilde{\gamma}_{jk} \log p(Y_{njk} | \frac{1}{2}; T_{njk}) + (1 - \tilde{\gamma}_{jk}) \log p(Y_{njk} | B_{jk}; T_{njk}) + \mathbf{E}_{q(\gamma)} \left[ \log \frac{p(\gamma_{jk} | \pi_j)}{q(\gamma_{jk})} \right]. \quad (2.17)$$

The distributions in the first and second terms are binomial distributions and the relevant likelihood functions can be written as follows.

$$\mathcal{L}_{njk}^b = \log p(Y_{njk} | \frac{1}{2}; T_{njk}) = \mathcal{C}_{njk} + T_{njk} \log \left( \frac{1}{2} \right) \quad (2.18)$$

$$\mathcal{L}_{njk}^{bb}(B_{jk}) = \log p(Y_{njk} | B_{jk}; T_{njk}) = \mathcal{C}_{njk} + Y_{njk} \log(B_{jk}) + (T_{njk} - Y_{njk}) \log(1 - B_{jk}) \quad (2.19)$$

Thus,

$$\mathcal{L} = \sum_{n,j,k} \tilde{\gamma}_{jk} \mathcal{L}_{njk}^b + (1 - \tilde{\gamma}_{jk}) \mathcal{L}_{njk}^{bb}(B_{jk}) + \tilde{\gamma}_{jk} \log \frac{\pi_j}{\tilde{\gamma}_{jk}} + (1 - \tilde{\gamma}_{jk}) \log \frac{(1 - \pi_j)}{(1 - \tilde{\gamma}_{jk})} \quad (2.20)$$

Maximizing  $\mathcal{L}$  with respect to  $\tilde{\gamma}_{jk}$  while keeping other parameters fixed gives

$$\frac{\partial \mathcal{L}}{\partial \tilde{\gamma}_{jk}} = \frac{1}{N} \sum_n \mathcal{L}_{njk}^b - \mathcal{L}_{njk}^{bb}(B_{jk}) + \log \pi_j - \log(1 - \pi_j) - \log \tilde{\gamma}_{jk} + \log(1 - \tilde{\gamma}_{jk}) = 0 \quad (2.21)$$

$$\log \frac{\tilde{\gamma}_{jk}}{1 - \tilde{\gamma}_{jk}} = \log \frac{\pi_j}{1 - \pi_j} + \frac{1}{N} \sum_n \mathcal{L}_{njk}^b - \mathcal{L}_{njk}^{bb}(B_{jk}) \quad (2.22)$$

Maximizing  $\mathcal{L}$  with respect to  $\pi_j$  while keeping other parameters fixed gives

$$\frac{\partial \mathcal{L}}{\partial \pi_j} = \frac{1}{\pi_j} \sum_k \tilde{\gamma}_{jk} - \frac{1}{1 - \pi_j} \sum_k (1 - \tilde{\gamma}_{jk}) = 0 \quad (2.23)$$

$$\pi_j = \frac{1}{K} \sum_k \tilde{\gamma}_{jk} \quad (2.24)$$

---

<sup>8</sup>I won't go through your derivation. I trust you! I think it's ok to derive MLE by your way. But if we want to put this derivation into supplementary material, I would prefer to use typical EM derivation.

Maximizing  $\mathcal{L}$  with respect to  $B_{jk}$  while keeping other parameters fixed gives

$$\frac{\partial \mathcal{L}}{\partial B_{jk}} = \frac{1 - \gamma_{jk}}{B_{jk}} \sum_n Y_{nj k} - \frac{(1 - \gamma_{jk})}{1 - B_{jk}} \sum_n (T_{nj k} - Y_{nj k}) = 0 \quad (2.25)$$

$$B_{jk} = \frac{\sum_n Y_{nj k}}{\sum_n T_{nj k}} \quad (2.26)$$

**2.2. Model 2.** Following the derivation in the previous model, the lower bound to the log-likelihood can be written as follows:

$$\mathcal{L} \geq \sum_{n,j,k} \tilde{\gamma}_{jk} \log p(Y_{nj k} | \frac{1}{2}; T_{nj k}) + (1 - \tilde{\gamma}_{jk}) \log p(Y_{nj k} | \mu_j; T_{nj k}) + \mathbf{E}_{q(\gamma)} \left[ \log \frac{p(\gamma_{jk} | \pi_j)}{q(\gamma_{jk})} \right]. \quad (2.27)$$

The distribution in the first term is a binomial distribution and the distribution in the second is a “symmetric” beta-binomial distribution. Since the parameters of these distributions are fixed, we will replace the relevant likelihood functions as follows.

$$\mathcal{L}_{nj k}^b = \log p(Y_{nj k} | \frac{1}{2}; T_{nj k}) \quad (2.28)$$

$$\mathcal{L}_{nj k}^{bb} = \log p(Y_{nj k} | \mu_j; T_{nj k}) \quad (2.29)$$

Thus,

$$\mathcal{L} = \sum_{n,j,k} \tilde{\gamma}_{jk} \mathcal{L}_{nj k}^b + (1 - \tilde{\gamma}_{jk}) \mathcal{L}_{nj k}^{bb} + \tilde{\gamma}_{jk} \log \frac{\pi_j}{\tilde{\gamma}_{jk}} + (1 - \tilde{\gamma}_{jk}) \log \frac{(1 - \pi_j)}{(1 - \tilde{\gamma}_{jk})} \quad (2.30)$$

Maximum likelihood estimates for  $\gamma_{jk}$  and  $\pi_j$ , as derived earlier, can be written as

$$\log \frac{\tilde{\gamma}_{jk}}{1 - \tilde{\gamma}_{jk}} = \log \frac{\pi_j}{1 - \pi_j} + \frac{1}{N} \sum_n \mathcal{L}_{nj k}^b - \mathcal{L}_{nj k}^{bb} \quad (2.31)$$

$$\pi_j = \frac{1}{K} \sum_k \tilde{\gamma}_{jk} \quad (2.32)$$

**2.3. Model 3.** Again, as before, the lower bound to the log-likelihood can be written as follows:

$$\mathcal{L} \geq \sum_{n,j,k} \tilde{\gamma}_{jk} \log p(Y_{nj k} | \frac{1}{2}, \tau_{jk}; T_{nj k}) + (1 - \tilde{\gamma}_{jk}) \log p(Y_{nj k} | B_{jk}, \tau_{jk}; T_{nj k}) + \mathbf{E}_{q(\gamma)} \left[ \log \frac{p(\gamma_{jk} | \pi_j)}{q(\gamma_{jk})} \right]. \quad (2.33)$$

The distributions in the first and second terms are beta-binomial distributions and the relevant likelihood functions can be written as follows.

$$\begin{aligned} \mathcal{L}_{nj k}^b(\tau_{jk}) &= \log p(Y_{nj k} | \frac{1}{2}, \tau_{jk}; T_{nj k}) = \mathcal{C}_{nj k} + \log \Gamma(Y_{nj k} + 0.5\tau_{jk}) + \log \Gamma(T_{nj k} - Y_{nj k} + 0.5\tau_{jk}) \\ &\quad - \log \Gamma(T_{nj k} + \tau_{jk}) + \log \Gamma(\tau_{jk}) - 2 * \log \Gamma(0.5\tau_{jk}) \end{aligned} \quad (2.34)$$

$$\begin{aligned} \mathcal{L}_{nj k}^{bb}(B_{jk}, \tau_{jk}) &= \log p(Y_{nj k} | B_{jk}, \tau_{jk}; T_{nj k}) = \mathcal{C}_{nj k} + \log \Gamma(Y_{nj k} + B_{jk}\tau_{jk}) + \log \Gamma(T_{nj k} - Y_{nj k} + (1 - B_{jk})\tau_{jk}) \\ &\quad - \log \Gamma(T_{nj k} + \tau_{jk}) + \log \Gamma(\tau_{jk}) - \log \Gamma(B_{jk}\tau_{jk}) - \log \Gamma((1 - B_{jk})\tau_{jk}) \end{aligned} \quad (2.35)$$

Thus,

$$\mathcal{L} = \sum_{n,j,k} \tilde{\gamma}_{jk} \mathcal{L}_{nj k}^b(\tau_{jk}) + (1 - \tilde{\gamma}_{jk}) \mathcal{L}_{nj k}^{bb}(B_{jk}, \tau_{jk}) + \tilde{\gamma}_{jk} \log \frac{\pi_j}{\tilde{\gamma}_{jk}} + (1 - \tilde{\gamma}_{jk}) \log \frac{(1 - \pi_j)}{(1 - \tilde{\gamma}_{jk})} \quad (2.36)$$

Maximum likelihood estimates for  $\gamma_{jk}$  and  $\pi_j$  are the same as derived in the earlier two models. Since the remaining parameters  $B_{jk}$  and  $\tau_{jk}$  occur within  $\log \Gamma(\cdot)$  functions, closed form update

equations for these parameters cannot be derived. Instead, we'll maximize the likelihood with respect to these parameters using generalized convex optimization algorithms. The gradient of the likelihood with respect to each of these parameters can be derived as follows.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial B_{jk}} &= (1 - \tilde{\gamma}_{jk}) \sum_n [\tau_{jk} \Psi(Y_{nj}k + B_{jk} \tau_{jk}) - \tau_{jk} \Psi(T_{nj}k - Y_{nj}k + (1 - B_{jk}) \tau_{jk}) \\ &\quad - \tau_{jk} \Psi(B_{jk} \tau_{jk}) + \tau_{jk} \Psi((1 - B_{jk}) \tau_{jk})] \end{aligned} \quad (2.37)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \tau_{jk}} &= \sum_n \tilde{\gamma}_{jk} [0.5 \Psi(Y_{nj}k + 0.5 \tau_{jk}) + 0.5 \Psi(T_{nj}k - Y_{nj}k + 0.5 \tau_{jk}) \\ &\quad - \Psi(T_{nj}k + \tau_{jk}) + \Psi(\tau_{jk}) - \Psi(0.5 \tau_{jk})] \\ &\quad + (1 - \tilde{\gamma}_{jk}) [B_{jk} \Psi(Y_{nj}k + B_{jk} \tau_{jk}) + (1 - B_{jk}) \Psi(T_{nj}k - Y_{nj}k + (1 - B_{jk}) \tau_{jk}) \\ &\quad - \Psi(T_{nj}k + \tau_{jk}) + \Psi(\tau_{jk}) - B_{jk} \Psi(B_{jk} \tau_{jk}) - (1 - B_{jk}) \Psi((1 - B_{jk}) \tau_{jk})] \end{aligned} \quad (2.38)$$

**2.4. Posterior distribution of  $R_{jk}$  – Model 2.** <sup>9</sup> Following Kolaczyk (1999), the posterior distribution of  $R_{jk}$  can be computed as follows

$$p(R_{jk} | Y_{\cdot jk}, T_{\cdot jk}) = \frac{1}{\mathcal{Z}_{jk}} p(Y_{\cdot jk} | R_{jk}, T_{\cdot jk}) p(R_{jk}) \quad (2.39)$$

$$= \left( \prod_n p(Y_{nj}k | R_{jk}, T_{nj}k) \right) p(R_{jk}) \quad (2.40)$$

$$= \pi_j \left( \prod_n p(Y_{nj}k | R_{jk}, T_{nj}k) \right) \delta\left(\frac{1}{2}\right) + (1 - \pi_j) \left( \prod_n p(Y_{nj}k | R_{jk}, T_{nj}k) \right) \text{beta}(\mu_j, \mu_j), \quad (2.41)$$

where  $\mathcal{Z}_{jk}$  is the appropriate normalizing constant. Now,

$$p(Y_{nj}k | R_{jk}, T_{nj}k) = \Gamma_{nj}k R_{jk}^{Y_{nj}k} (1 - R_{jk})^{T_{nj}k - Y_{nj}k}, \quad (2.42)$$

where  $\Gamma_{nj}k$  is the normalizing constant of the binomial distribution. Thus,

$$p(Y_{\cdot jk} | R_{jk}, T_{\cdot jk}) = \prod_n \Gamma_{nj}k \prod_n R_{jk}^{Y_{nj}k} \prod_n (1 - R_{jk})^{T_{nj}k - Y_{nj}k} \quad (2.43)$$

$$= \Gamma_{jk} R_{jk}^{\sum_n Y_{nj}k} (1 - R_{jk})^{\sum_n T_{nj}k - Y_{nj}k}, \quad (2.44)$$

where  $\Gamma_{jk} = \prod_n \Gamma_{nj}k$ . The normalizing constant  $\mathcal{Z}_{jk}$  can now be computed as follows.

$$\mathcal{Z}_{jk} = \pi_j \int \Gamma_{jk} R_{jk}^{\sum_n Y_{nj}k} (1 - R_{jk})^{\sum_n T_{nj}k - Y_{nj}k} \delta\left(\frac{1}{2}\right) dR_{jk} \quad (2.45)$$

$$+ (1 - \pi_j) \int \Gamma_{jk} R_{jk}^{\sum_n Y_{nj}k} (1 - R_{jk})^{\sum_n T_{nj}k - Y_{nj}k} \text{beta}(\mu_j, \mu_j) dR_{jk} \quad (2.46)$$

$$= \Gamma_{jk} \left[ \pi_j \left(\frac{1}{2}\right)^{\sum_n T_{nj}k} + (1 - \pi_j) \frac{\Gamma(2\mu_j) \Gamma(\mu_j + \sum_n Y_{nj}k) \Gamma(\mu_j + \sum_n T_{nj}k - Y_{nj}k)}{\Gamma(\mu_j)^2 \Gamma(2\mu_j + \sum_n T_{nj}k)} \right] \quad (2.47)$$

---

<sup>9</sup>Ummm, actually we can calculate posterior distribution for  $R_{jk}$  in all three models. I re-wrote this section in the next section. BTW, we use this posterior mean to show mean curve (not for actual inference)?

Following the same calculation, the posterior mean can be computed as follows.

$$\mathbf{E}[R_{jk}|Y_{\cdot jk}, T_{\cdot jk}] = \frac{\Gamma_{jk}}{Z_{jk}} \left[ \pi_j \left( \frac{1}{2} \right)^{\sum_n T_{njk}+1} + (1 - \pi_j) \frac{\Gamma(2\mu_j) \Gamma(\mu_j + \sum_n Y_{njk} + 1) \Gamma(\mu_j + \sum_n T_{njk} - Y_{njk})}{\Gamma(\mu_j)^2 \Gamma(2\mu_j + \sum_n T_{njk} + 1)} \right] \quad (2.48)$$

$$= \frac{\pi_j \left( \frac{1}{2} \right)^{\sum_n T_{njk}+1} + (1 - \pi_j) \frac{\Gamma(2\mu_j) \Gamma(\mu_j + \sum_n Y_{njk} + 1) \Gamma(\mu_j + \sum_n T_{njk} - Y_{njk})}{\Gamma(\mu_j)^2 \Gamma(2\mu_j + \sum_n T_{njk} + 1)}}{\pi_j \left( \frac{1}{2} \right)^{\sum_n T_{njk}} + (1 - \pi_j) \frac{\Gamma(2\mu_j) \Gamma(\mu_j + \sum_n Y_{njk}) \Gamma(\mu_j + \sum_n T_{njk} - Y_{njk})}{\Gamma(\mu_j)^2 \Gamma(2\mu_j + \sum_n T_{njk})}} \quad (2.49)$$

$$= \frac{\pi_j \left( \frac{1}{2} \right)^{\sum_n T_{njk}+1} + (1 - \pi_j) \frac{B(\mu_j + \sum_n Y_{njk} + 1, \mu_j + \sum_n T_{njk} - Y_{njk})}{B(\mu_j, \mu_j)}}{\pi_j \left( \frac{1}{2} \right)^{\sum_n T_{njk}} + (1 - \pi_j) \frac{B(\mu_j + \sum_n Y_{njk}, \mu_j + \sum_n T_{njk} - Y_{njk})}{B(\mu_j, \mu_j)}}, \quad (2.50)$$

where  $B(\cdot, \cdot)$  is the beta function.

**2.5. Posterior distribution of  $R_{jk}$ .** The posterior distribution of  $R_{jk}$  can be computed as follows.

$$p(R_{jk}|Y_{\cdot jk}, T_{\cdot jk}) = p(R_{jk}, \gamma_{jk} = 1|Y_{\cdot jk}, T_{\cdot jk}) + p(R_{jk}, \gamma_{jk} = 0|Y_{\cdot jk}, T_{\cdot jk}) \quad (2.51)$$

$$= p(R_{jk}|\gamma_{jk} = 1, Y_{\cdot jk}, T_{\cdot jk})p(\gamma_{jk} = 1|Y_{\cdot jk}, T_{\cdot jk}) + p(R_{jk}|\gamma_{jk} = 0, Y_{\cdot jk}, T_{\cdot jk})p(\gamma_{jk} = 0|Y_{\cdot jk}, T_{\cdot jk}) \quad (2.52)$$

$$(2.53)$$

where

$$p(\gamma_{jk} = 0|Y_{\cdot jk}, T_{\cdot jk}) = \frac{p(Y_{\cdot jk}|\gamma_{jk} = 0, T_{\cdot jk})(1 - \pi_j)}{p(Y_{\cdot jk}|\gamma_{jk} = 1, T_{\cdot jk})\pi_j + p(Y_{\cdot jk}|\gamma_{jk} = 0, T_{\cdot jk})(1 - \pi_j)} \quad (2.54)$$

$$= 1 - p(\gamma_{jk} = 1|Y_{\cdot jk}, T_{\cdot jk}). \quad (2.55)$$

$$(2.56)$$

We will derive  $p(Y_{\cdot jk}|\gamma_{jk} = 0, T_{\cdot jk})$ ,  $p(Y_{\cdot jk}|\gamma_{jk} = 1, T_{\cdot jk})$ ,  $p(R_{jk}|\gamma_{jk} = 0, Y_{\cdot jk}, T_{\cdot jk})$ , and  $p(R_{jk}|\gamma_{jk} = 1, Y_{\cdot jk}, T_{\cdot jk})$  for each model.

**2.5.1. Model 1.**

$$p(Y_{\cdot jk}|\gamma_{jk} = 0, T_{\cdot jk}) = \left[ \prod_n \binom{T_{njk}}{Y_{njk}} \right] (B_{jk})^{\sum_n Y_{njk}} (1 - B_{jk})^{\sum_n (T_{njk} - Y_{njk})} \quad (2.57)$$

$$p(Y_{\cdot jk}|\gamma_{jk} = 1, T_{\cdot jk}) = \left[ \prod_n \binom{T_{njk}}{Y_{njk}} \right] \left( \frac{1}{2} \right)^{\sum_n T_{njk}} \quad (2.58)$$

$$p(R_{jk}|\gamma_{jk} = 0, Y_{\cdot jk}, T_{\cdot jk}) = \delta(B_{jk}) \quad (2.59)$$

$$p(R_{jk}|\gamma_{jk} = 1, Y_{\cdot jk}, T_{\cdot jk}) = \delta\left(\frac{1}{2}\right) \quad (2.60)$$

$$(2.61)$$

Then, the posterior mean can be written as

$$\mathbf{E}[R_{jk}|Y_{\cdot jk}, T_{\cdot jk}] = \frac{\left(\frac{1}{2}\right)^{(1+\sum_n T_{njk})}\pi_j + (B_{jk})^{(1+\sum_n Y_{njk})}(1 - B_{jk})^{\sum_n (T_{njk} - Y_{njk})}(1 - \pi_j)}{\left(\frac{1}{2}\right)^{\sum_n T_{njk}}\pi_j + (B_{jk})^{\sum_n Y_{njk}}(1 - B_{jk})^{\sum_n (T_{njk} - Y_{njk})}(1 - \pi_j)}. \quad (2.62)$$

$$(2.63)$$

2.5.2. *Model 2.*

$$p(Y_{\cdot jk} | \gamma_{jk} = 0, T_{\cdot jk}) = \left[ \prod_n \binom{T_{njk}}{Y_{njk}} \right] \frac{B(\mu_j + \sum_n Y_{njk}, \mu_j + \sum_n (T_{njk} - Y_{njk}))}{B(\mu_j, \mu_j)} \quad (2.64)$$

$$p(Y_{\cdot jk} | \gamma_{jk} = 1, T_{\cdot jk}) = \left[ \prod_n \binom{T_{njk}}{Y_{njk}} \right] \left( \frac{1}{2} \right)^{\sum_n T_{njk}} \quad (2.65)$$

$$p(R_{jk} | \gamma_{jk} = 0, Y_{\cdot jk}, T_{\cdot jk}) = \frac{\mu_j + \sum_n Y_{njk}}{2\mu_j + \sum_n T_{njk}} \quad (2.66)$$

$$p(R_{jk} | \gamma_{jk} = 1, Y_{\cdot jk}, T_{\cdot jk}) = \delta \left( \frac{1}{2} \right) \quad (2.67)$$

$$(2.68)$$

Then, the posterior mean can be written as

$$\mathbf{E} [R_{jk} | Y_{\cdot jk}, T_{\cdot jk}] = \frac{\left( \frac{1}{2} \right)^{(1 + \sum_n T_{njk})} \pi_j + \frac{B(1 + \mu_j + \sum_n Y_{njk}, \mu_j + \sum_n (T_{njk} - Y_{njk}))}{B(\mu_j, \mu_j)} (1 - \pi_j)}{\left( \frac{1}{2} \right)^{\sum_n T_{njk}} \pi_j + \frac{B(\mu_j + \sum_n Y_{njk}, \mu_j + \sum_n (T_{njk} - Y_{njk}))}{B(\mu_j, \mu_j)} (1 - \pi_j)}, \quad (2.69)$$

$$(2.70)$$

where  $B(\cdot, \cdot)$  is the beta function.

2.5.3. *Model 3.*

$$p(Y_{\cdot jk} | \gamma_{jk} = 0, T_{\cdot jk}) = \prod_n \int \binom{T_{njk}}{Y_{njk}} R_{njk}^{Y_{njk}} (1 - R_{njk})^{T_{njk} - Y_{njk}} \frac{R_{njk}^{B_{jk}\tau_{jk} - 1} (1 - R_{njk})^{(1 - B_{jk})\tau_{jk} - 1}}{B(B_{jk}\tau_{jk}, (1 - B_{jk})\tau_{jk})} dR_{njk} \quad (2.71)$$

$$= \prod_n \binom{T_{njk}}{Y_{njk}} \frac{B(B_{jk}\tau_{jk} + Y_{njk}, (1 - B_{jk})\tau_{jk} + T_{njk} - Y_{njk})}{B(B_{jk}\tau_{jk}, (1 - B_{jk})\tau_{jk})} \quad (2.72)$$

$$p(Y_{\cdot jk} | \gamma_{jk} = 1, T_{\cdot jk}) = \prod_n \binom{T_{njk}}{Y_{njk}} \frac{B(\frac{1}{2}\tau_{jk} + Y_{njk}, \frac{1}{2}\tau_{jk} + T_{njk} - Y_{njk})}{B(\frac{1}{2}\tau_{jk}, \frac{1}{2}\tau_{jk})} \quad (2.73)$$

$$p(R_{jk} | \gamma_{jk} = 0, Y_{\cdot jk}, T_{\cdot jk}) = \delta(B_{jk}) \quad (2.74)$$

$$p(R_{jk} | \gamma_{jk} = 1, Y_{\cdot jk}, T_{\cdot jk}) = \delta \left( \frac{1}{2} \right) \quad (2.75)$$

$$(2.76)$$

Then, the posterior mean can be written as

$$\mathbf{E} [R_{jk} | Y_{\cdot jk}, T_{\cdot jk}] = \frac{\frac{1}{2}\pi_j \prod_n \frac{B(\frac{1}{2}\tau_{jk} + Y_{njk}, \frac{1}{2}\tau_{jk} + T_{njk} - Y_{njk})}{B(\frac{1}{2}\tau_{jk}, \frac{1}{2}\tau_{jk})} + B_{jk}(1 - \pi_j) \prod_n \frac{B(B_{jk}\tau_{jk} + Y_{njk}, (1 - B_{jk})\tau_{jk} + T_{njk} - Y_{njk})}{B(B_{jk}\tau_{jk}, (1 - B_{jk})\tau_{jk})}}{\pi_j \prod_n \frac{B(\frac{1}{2}\tau_{jk} + Y_{njk}, \frac{1}{2}\tau_{jk} + T_{njk} - Y_{njk})}{B(\frac{1}{2}\tau_{jk}, \frac{1}{2}\tau_{jk})} + (1 - \pi_j) \prod_n \frac{B(B_{jk}\tau_{jk} + Y_{njk}, (1 - B_{jk})\tau_{jk} + T_{njk} - Y_{njk})}{B(B_{jk}\tau_{jk}, (1 - B_{jk})\tau_{jk})}} \quad (2.77)$$

$$(2.78)$$

## 3. CENTIPEDE-PBM

Each of the above three models can be straightforwardly incorporated into CENTIPEDE's learning framework by first deriving how the Poisson-Binomial model modifies the likelihood function. The key change is restricted to the multinomial part of CENTIPEDE's likelihood function. Here,



we derive the form of the change for Model 1, and then apply the change directly to Models 2 and 3 <sup>10</sup>.

$$\mathcal{L} = \sum_n \log \sum_{Z_n} p(X_n | Z_n, T_n^X) p(T_n^X | Z_n) p(Z_n | S_n, \beta) \quad (3.1)$$

$$\geq \sum_n \sum_{Z_n} q(Z_n) \log p(X_n | Z_n, T_n^X) + \sum_{Z_n} q(Z_n) \log p(T_n^X | Z_n) + \mathbf{E}_{q(Z_n)} \left[ \log \frac{p(Z_n)}{q(Z_n)} \right] \quad (3.2)$$

$$= \sum_n \tilde{z}_n \log p(X_n | Z_n = 1, T_n^X) + (1 - \tilde{z}_n) \log p(X_n | Z_n = 0, T_n^X) \\ + \sum_{Z_n} q(Z_n) \log p(T_n^X | Z_n) + \mathbf{E}_{q(Z_n)} \left[ \log \frac{p(Z_n)}{q(Z_n)} \right] \quad (3.3)$$

$$= \sum_n \tilde{z}_n \log \left[ \prod_{j,k} \sum_{\gamma_{jk}} p(Y_{nj k} | \gamma_{jk}, B_{jk}; T_{nj k}) p(\gamma_{jk} | \pi_j) \right] + (1 - \tilde{z}_n) \log \prod_{j,k} p(Y_{nj k} | \frac{1}{2}; T_{nj k}) \\ + \sum_{Z_n} q(Z_n) \log p(T_n^X | Z_n) + \mathbf{E}_{q(Z_n)} \left[ \log \frac{p(Z_n)}{q(Z_n)} \right] \quad (3.4)$$

$$\geq \sum_{n,j,k} \tilde{z}_n \sum_{\gamma_{jk}} q(\gamma_{jk}) \left( \log p(Y_{nj k} | \gamma_{jk}, B_{jk}; T_{nj k}) + \log \frac{p(\gamma_{jk} | \pi_j)}{q(\gamma_{jk})} \right) + (1 - \tilde{z}_n) \log p(Y_{nj k} | \frac{1}{2}; T_{nj k}) \\ + \sum_{Z_n} q(Z_n) \log p(T_n^X | Z_n) + \mathbf{E}_{q(Z_n)} \left[ \log \frac{p(Z_n)}{q(Z_n)} \right] \quad (3.5)$$

$$= \sum_{n,j,k} \tilde{z}_n \left( \tilde{\gamma}_{jk} \log p(Y_{nj k} | \frac{1}{2}; T_{nj k}) + (1 - \tilde{\gamma}_{jk}) \log p(Y_{nj k} | B_{jk}; T_{nj k}) + \mathbf{E}_{q(\gamma)} \left[ \log \frac{p(\gamma_{jk} | \pi_j)}{q(\gamma_{jk})} \right] \right) \\ + (1 - \tilde{z}_n) \log p(Y_{nj k} | \frac{1}{2}; T_{nj k}) + \sum_{Z_n} q(Z_n) \log p(T_n^X | Z_n) + \mathbf{E}_{q(Z_n)} \left[ \log \frac{p(Z_n)}{q(Z_n)} \right] \quad (3.6)$$

$$= \sum_{n,j,k} \tilde{z}_n \left( \tilde{\gamma}_{jk} \mathcal{L}_{nj k}^b + (1 - \tilde{\gamma}_{jk}) \mathcal{L}_{nj k}^{bb}(B_{jk}) + \tilde{\gamma}_{jk} \log \frac{\pi_j}{\tilde{\gamma}_{jk}} + (1 - \tilde{\gamma}_{jk}) \log \frac{(1 - \pi_j)}{(1 - \tilde{\gamma}_{jk})} \right) \\ + (1 - \tilde{z}_n) \mathcal{L}_{nj k}^b + \sum_{Z_n} q(Z_n) \log p(T_n^X | Z_n) + \mathbf{E}_{q(Z_n)} \left[ \log \frac{p(Z_n)}{q(Z_n)} \right] \quad (3.7)$$

$$(3.8)$$

The last two terms include the negative binomial contribution and the KL-divergence between the prior and posteriors, exactly as in CENTIPEDE. Thus, the modified likelihood terms for the three models, and the relevant modified update equations, can be written as follows.

**3.1. Model 1.** The modified likelihood terms include

$$\mathcal{L} = \sum_{n,j,k} \tilde{z}_n \left( \tilde{\gamma}_{jk} \mathcal{L}_{nj k}^b + (1 - \tilde{\gamma}_{jk}) \mathcal{L}_{nj k}^{bb}(B_{jk}) + \tilde{\gamma}_{jk} \log \frac{\pi_j}{\tilde{\gamma}_{jk}} + (1 - \tilde{\gamma}_{jk}) \log \frac{(1 - \pi_j)}{(1 - \tilde{\gamma}_{jk})} \right) + (1 - \tilde{z}_n) \mathcal{L}_{nj k}^b, \quad (3.9)$$

<sup>10</sup> **Again, I wont go through your derivation. I trust you!**

where

$$\mathcal{L}_{njk}^b = \log p(Y_{njk} | \frac{1}{2}; T_{njk}) = \mathcal{C}_{njk} + T_{njk} \log \left( \frac{1}{2} \right) \quad (3.10)$$

$$\mathcal{L}_{njk}^{bb}(B_{jk}) = \log p(Y_{njk} | B_{jk}; T_{njk}) = \mathcal{C}_{njk} + Y_{njk} \log(B_{jk}) + (T_{njk} - Y_{njk}) \log(1 - B_{jk}) \quad (3.11)$$

In the update equation for  $\tilde{z}$ , the likelihood functions of the multinomial parameters can be replaced by the following terms.

$$\log \frac{\tilde{z}_n}{1 - \tilde{z}_n} = \sum_{j,k} (1 - \tilde{\gamma}_{jk}) (\mathcal{L}_{njk}^{bb}(B_{jk}) - \mathcal{L}_{njk}^b) + \tilde{\gamma}_{jk} \log \frac{\pi_j}{\tilde{\gamma}_{jk}} + (1 - \tilde{\gamma}_{jk}) \log \frac{(1 - \pi_j)}{(1 - \tilde{\gamma}_{jk})} + \text{remaining terms} \quad (3.12)$$

A similar modification applied for the other two models. The update equations for  $\tilde{\gamma}$ ,  $B$  and  $\pi$  can be modified as follows:

$$\log \frac{\tilde{\gamma}_{jk}}{1 - \tilde{\gamma}_{jk}} = \log \frac{\pi_j}{1 - \pi_j} + \frac{\sum_n \tilde{z}_n (\mathcal{L}_{njk}^b - \mathcal{L}_{njk}^{bb}(B_{jk}))}{\sum_n \tilde{z}_n} \quad (3.13)$$

$$\pi_j = \frac{1}{K} \sum_k \tilde{\gamma}_{jk} \quad (3.14)$$

$$B_{jk} = \frac{\sum_n \tilde{z}_n Y_{njk}}{\sum_n \tilde{z}_n T_{njk}} \quad (3.15)$$

### 3.2. Model 2.

$$\mathcal{L} = \sum_{n,j,k} \tilde{z}_n \left( \tilde{\gamma}_{jk} \mathcal{L}_{njk}^b + (1 - \tilde{\gamma}_{jk}) \mathcal{L}_{njk}^{bb} + \tilde{\gamma}_{jk} \log \frac{\pi_j}{\tilde{\gamma}_{jk}} + (1 - \tilde{\gamma}_{jk}) \log \frac{(1 - \pi_j)}{(1 - \tilde{\gamma}_{jk})} \right) + (1 - \tilde{z}_n) \mathcal{L}_{njk}^b, \quad (3.16)$$

where

$$\mathcal{L}_{njk}^b = \log p(Y_{njk} | \frac{1}{2}; T_{njk}) \quad (3.17)$$

$$\mathcal{L}_{njk}^{bb} = \log p(Y_{njk} | \mu_j; T_{njk}). \quad (3.18)$$

Update equations for  $\tilde{\gamma}$  and  $\pi$  can be given as

$$\log \frac{\tilde{\gamma}_{jk}}{1 - \tilde{\gamma}_{jk}} = \log \frac{\pi_j}{1 - \pi_j} + \frac{\sum_n \tilde{z}_n (\mathcal{L}_{njk}^b - \mathcal{L}_{njk}^{bb})}{\sum_n \tilde{z}_n} \quad (3.19)$$

$$\pi_j = \frac{1}{K} \sum_k \tilde{\gamma}_{jk} \quad (3.20)$$

### 3.3. Model 3.

$$\mathcal{L} = \sum_{n,j,k} \tilde{z}_n \left( \tilde{\gamma}_{jk} \mathcal{L}_{njk}^b(\tau_{jk}) + (1 - \tilde{\gamma}_{jk}) \mathcal{L}_{njk}^{bb}(B_{jk}, \tau_{jk}) + \tilde{\gamma}_{jk} \log \frac{\pi_j}{\tilde{\gamma}_{jk}} + (1 - \tilde{\gamma}_{jk}) \log \frac{(1 - \pi_j)}{(1 - \tilde{\gamma}_{jk})} \right) + (1 - \tilde{z}_n) \mathcal{L}_{njk}^o, \quad (3.21)$$

where

$$\begin{aligned}\mathcal{L}_{njk}^b(\tau_{jk}) &= \mathcal{C}_{njk} + \log \Gamma(Y_{njk} + 0.5\tau_{jk}) + \log \Gamma(T_{njk} - Y_{njk} + 0.5\tau_{jk}) \\ &\quad - \log \Gamma(T_{njk} + \tau_{jk}) + \log \Gamma(\tau_{jk}) - 2 * \log \Gamma(0.5\tau_{jk})\end{aligned}\quad (3.22)$$

$$\begin{aligned}\mathcal{L}_{njk}^{bb}(B_{jk}, \tau_{jk}) &= \mathcal{C}_{njk} + \log \Gamma(Y_{njk} + B_{jk}\tau_{jk}) + \log \Gamma(T_{njk} - Y_{njk} + (1 - B_{jk})\tau_{jk}) \\ &\quad - \log \Gamma(T_{njk} + \tau_{jk}) + \log \Gamma(\tau_{jk}) - \log \Gamma(B_{jk}\tau_{jk}) - \log \Gamma((1 - B_{jk})\tau_{jk}) \mathcal{L}_{njk}^o(\tau_{jk}^o) = \mathcal{C}_{njk} + \log \Gamma(Y_{njk} + 0.5\tau_{jk}^o) \\ &\quad - \log \Gamma(T_{njk} + \tau_{jk}^o) + \log \Gamma(\tau_{jk}^o) - 2 * \log \Gamma(0.5\tau_{jk}^o)\end{aligned}\quad (3.23)$$

$$(3.24)$$

Update equations for  $\tilde{\gamma}$  and  $\pi$  can be given as

$$\log \frac{\tilde{\gamma}_{jk}}{1 - \tilde{\gamma}_{jk}} = \log \frac{\pi_j}{1 - \pi_j} + \frac{\sum_n \tilde{z}_n \left( \mathcal{L}_{njk}^b(\tau_{jk}) - \mathcal{L}_{njk}^{bb}(B_{jk}, \tau_{jk}) \right)}{\sum_n \tilde{z}_n}\quad (3.25)$$

$$\pi_j = \frac{1}{K} \sum_k \tilde{\gamma}_{jk}\quad (3.26)$$

The gradient of the likelihood with respect to  $B$  and  $\tau$  can be derived as follows.

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial B_{jk}} &= \tau_{jk}(1 - \tilde{\gamma}_{jk}) \sum_n \tilde{z}_n [\Psi(Y_{njk} + B_{jk}\tau_{jk}) - \Psi(T_{njk} - Y_{njk} + (1 - B_{jk})\tau_{jk}) \\ &\quad - \Psi(B_{jk}\tau_{jk}) + \Psi((1 - B_{jk})\tau_{jk})]\end{aligned}\quad (3.27)$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \tau_{jk}} &= \sum_n \tilde{z}_n \tilde{\gamma}_{jk} [0.5\Psi(Y_{njk} + 0.5\tau_{jk}) + 0.5\Psi(T_{njk} - Y_{njk} + 0.5\tau_{jk}) \\ &\quad - \Psi(T_{njk} + \tau_{jk}) + \Psi(\tau_{jk}) - \Psi(0.5\tau_{jk})] \\ &\quad + \tilde{z}_n(1 - \tilde{\gamma}_{jk}) [B_{jk}\Psi(Y_{njk} + B_{jk}\tau_{jk}) + (1 - B_{jk})\Psi(T_{njk} - Y_{njk} + (1 - B_{jk})\tau_{jk}) \\ &\quad - \Psi(T_{njk} + \tau_{jk}) + \Psi(\tau_{jk}) - B_{jk}\Psi(B_{jk}\tau_{jk}) - (1 - B_{jk})\Psi((1 - B_{jk})\tau_{jk})]\end{aligned}\quad (3.28)$$

#### 4. INFERENCE

Inference on binding sites can be performed by using

$$\frac{\text{P(binding site|data)}}{\text{P(not binding site|data)}},\quad (4.1)$$

where

$$\text{P(not binding site|data)} = \text{P}(Z_n = 0|X_n, T_n^X) + \text{P}(Z_n = 1, \gamma_{jk} = 1 \forall j, k|X_n, T_n^X)\quad (4.2)$$

$$\text{P(binding site|data)} = 1 - \text{P(not binding site|data)}.\quad (4.3)$$

Here,

$$\text{P}(Z_n = 1, \gamma_{jk} = 1 \forall j, k|X_n, T_n^X) = \text{P}(\gamma_{jk} = 1 \forall j, k|Z_n = 1, X_n, T_n^X)\text{P}(Z_n = 1|X_n, T_n^X)\quad (4.4)$$

$$= \prod_{jk} [\text{P}(\gamma_{jk} = 1|Z_n = 1, X_n, T_n^X)]\text{P}(Z_n = 1|X_n, T_n^X).\quad (4.5)$$