

Additional file 7 – derivation of moment estimators for m_i and s_i

Assume the read count $y_{i,j}$ for the i -th CRE and the j -th bulk DNase-seq sample is generated by a Poisson distribution:

$$y_{i,j} \sim \text{Poisson}(L_j \mu_{i,j})$$

$$\log(\mu_{i,j}) = m_i + s_i \delta_{i,j}$$

Further assume $\delta_{i,j} \sim N(0, 1)$. According to the law of total expectation and total variance for conditional probability:

$$\begin{aligned} \mathbb{E}\left(\frac{y_{i,j}}{L_j}\right) &= \frac{1}{L_j} \mathbb{E}[\mathbb{E}(y_{i,j} | \delta_{i,j})] \\ &= \mathbb{E}[\mu_{i,j}] \\ &= \mathbb{E}[e^{m_i + s_i \delta_{i,j}}] \\ &= e^{m_i} \mathbb{E}[e^{s_i \delta_{i,j}}] \\ &= e^{m_i + \frac{1}{2}s_i^2} \end{aligned}$$

$$\begin{aligned} \mathbb{E}\left(\frac{y_{i,j}}{L_j}\right)^2 &= \frac{1}{L_j^2} \mathbb{E}[\mathbb{E}(y_{i,j}^2 | \delta_{i,j})] \\ &= \frac{1}{L_j^2} \mathbb{E}[Var(y_{i,j} | \delta_{i,j}) + \mathbb{E}(y_{i,j} | \delta_{i,j})^2] \\ &= \frac{1}{L_j^2} \mathbb{E}[L_j \mu_{i,j} + L_j^2 \mu_{i,j}^2] \\ &= \frac{1}{L_j} \mathbb{E}[\mu_{i,j}] + \mathbb{E}[\mu_{i,j}^2] \\ &= \frac{1}{L_j} e^{m_i + \frac{1}{2}s_i^2} + \mathbb{E}[e^{2m_i + 2s_i \delta_{i,j}}] \\ &= \frac{1}{L_j} e^{m_i + \frac{1}{2}s_i^2} + e^{2m_i + 2s_i^2} \\ &= \frac{1}{L_j} e^{m_i + \frac{1}{2}s_i^2} + \left[e^{m_i + \frac{1}{2}s_i^2}\right]^2 e^{s_i^2} \\ &= \frac{1}{L_j} \mathbb{E}\left(\frac{y_{i,j}}{L_j}\right) + \left[\mathbb{E}\left(\frac{y_{i,j}}{L_j}\right)\right]^2 e^{s_i^2} \end{aligned}$$

Thus we have:

$$e^{s_i^2} = \frac{\mathbb{E}\left(\frac{y_{i,j}}{L_j}\right)^2 - \frac{1}{L_j} \mathbb{E}\left(\frac{y_{i,j}}{L_j}\right)}{\left[\mathbb{E}\left(\frac{y_{i,j}}{L_j}\right)\right]^2}$$

and

$$s_i^2 = \log \left(\frac{\mathbb{E}\left(\frac{y_{i,j}}{L_j}\right)^2 - \frac{1}{L_j} \mathbb{E}\left(\frac{y_{i,j}}{L_j}\right)}{\left[\mathbb{E}\left(\frac{y_{i,j}}{L_j}\right)\right]^2} \right)$$

After replacing $E\left(\frac{y_{i,j}}{L_j}\right)^2$ and $E\left(\frac{y_{i,j}}{L_j}\right)$ by their sample estimates $\sum_j \left(\frac{y_{i,j}}{L_j}\right)^2 / J$ and $\sum_j \left(\frac{y_{i,j}}{L_j}\right) / J$, respectively, we have:

$$\tilde{s}_i = \sqrt{\log \left(\frac{\sum_j (y_{i,j}/L_j)^2 / J - \sum_j (y_{i,j}/L_j)^2 / J}{(\sum_j (y_{i,j}/L_j) / J)^2} \right)}$$

Since

$$m_i = \log E\left(\frac{y_{i,j}}{L_j}\right) - \frac{1}{2}s_i^2$$

We have

$$\tilde{m}_i = \log \left(\frac{\sum_j (y_{i,j}/L_j)}{J} \right) - \tilde{s}_i^2 / 2$$