Supplemental Methods

Negative Binomial model

In this paper we use the following parameterization for the Negative binomial distribution.

$$P(k|\mu, \theta) = \frac{\Gamma(k+\theta)}{\Gamma(k)\theta!} \left(\frac{\mu}{\mu+\theta}\right)^k \left(\frac{\theta}{\mu+\theta}\right)^\theta$$

where the variance of the distribution is given by:

$$Var = \mu + \frac{\mu^2}{\theta}$$

and hence the coefficient of variation is given by:

$$CV^2 = \frac{1}{\mu} + \frac{1}{\theta}$$

Computation of the default theta

We assume, a biological covariation of 20% for large means.

$$\begin{aligned} CV^2 &= \frac{1}{\mu} + \frac{1}{\theta} \\ \lim_{\mu \to \infty} CV^2 &= \frac{1}{\theta} \\ \theta &\approx \frac{1}{CV^2} \end{aligned}$$

and hence equate a default $\theta = 25$.

Autoencoder Gradient

We use L-BFGS to fit the autoencoder model as described in the Methods. To speed up the fitting we implemented the gradient as derived below.

The expectations μ are modeled by:

$$\begin{aligned} \mu_{ij} &= s_i e^{y_{ij} + \bar{x}_j} \\ \mathbf{Y} &= \mathbf{XWW}^T + \mathbf{b} \end{aligned}$$

where the matrix $\mathbf{X}$ is given by the matrix: $\log \frac{k_{ij}+1}{s_i} - \bar{x}_j$.

The negative binomial log likelihood is given by:

$$\begin{aligned} ll =& \sum_{ij} k_{ij} \log(\mu_{ij}) + \sum_{ij} \theta \log(\theta) - \sum_{ij} (k_{ij} + \theta) \log(\mu_{ij} + \theta) \\ &+ \sum_{ij} \log(\Gamma(\theta + k_{ij})) - \sum_{ij} \log(\Gamma(\theta) k_{ij}!) \end{aligned}$$

For the derivation of the gradient only the first and third term need to be considered, as all other terms are independent of μ.

Computing the derivative of the first term with respect to the matrix $\mathbf{W}$ by subsituting the autoencoder model for μ. Here the operations $\log[\mathbf{A}]$ and $\exp[\mathbf{A}]$ are understood to be element-wise for a matrix or vector $\mathbf{A}$,

$$\begin{aligned} &\frac{d}{dw_{ab}} \sum_{ij} k_{ij} \log(\mu_{ij}) \\ &= \frac{d}{dw_{ab}} \sum_{ij} k_{ij} \log\left[\exp\left[\mathbf{X}\mathbf{W}\mathbf{W}^T + \mathbf{b}\right]\right] \\ &= \frac{d}{dw_{ab}} \sum_{ij} k_{ij} \left(\mathbf{X}\mathbf{W}\mathbf{W}^T + \mathbf{b}\right) \\ &= \frac{d}{dw_{ab}} \sum_{ij} k_{ij} \left(\sum_{lm} x_{il} w_{lm} w_{jm} + b_j\right) \\ &= \sum_{ij} k_{ij} \left(x_{ia} w_{jb} + \delta_{aj} \sum_{l} x_{il} w_{lb}\right) \\ &= \sum_{ij} x_{ia} k_{ij} w_{jb} + \sum_{il} k_{ia} x_{il} w_{lb}. \end{aligned}$$

Which can be written as:

$$\mathbf{K}^T\mathbf{X}\mathbf{W} - \mathbf{X}^T\mathbf{K}\mathbf{W}$$

Equivalently the derivative of the third term is:

$$-\mathbf{L}^T\mathbf{X}\mathbf{W} - \mathbf{X}^T\mathbf{L}\mathbf{W}$$

where the components of the matrix $\mathbf{L}$ are computed by:

$$l_{ij} = \frac{(k_{ij} + \theta)\mu_{ij}}{\theta + \mu_{ij}}$$

The combined result is then:

$$\frac{dll}{d\mathbf{W}} = \mathbf{K}^T\mathbf{X}\mathbf{W} + \mathbf{X}^T\mathbf{K}\mathbf{W} - \mathbf{L}^T\mathbf{X}\mathbf{W} - \mathbf{X}^T\mathbf{L}\mathbf{W}$$

The derivative of the first term with respect to the bias $\mathbf{b}$ is computed as:

$$\begin{aligned}
&\frac{d}{db_a}\sum_{ij} k_{ij}\log(\mu_{ij}) \\
&= \frac{d}{db_a}\sum_{ij} k_{ij}\left(\sum_{lm} x_{il}w_{lm}w_{jm} + b_j\right) \\
&= \sum_i k_{ia}
\end{aligned}$$

Equivalently for the third therm the derivative is $-\sum_i l_{ia}$ and so the derivative of the loglikelihood with respect to the bias is:

$$\frac{dll}{db_a} = \sum_i k_{ia} - l_{ia}$$