THE LOG-NORMAL MODEL OF RELATIONAL DATA

Let Z_{jk} be the j[th] stimulus coordinate on the k[th] dimension, k=1,…,s, where s is the number of dimensions. Let d_{jm} be the Euclidean distance between stimulus j and stimulus m in the s-dimensional space:

$$d_{jm} = \sqrt{\sum_{k=1}^{s}(Z_{jk} - Z_{mk})^2} \qquad (1)$$

We assume that the observed distances, d^*_{jm}, are

$$d^*_{jm} = d_{jm} + \varepsilon_{jm} \qquad (2)$$

And that they are drawn from the log-normal distribution because distances are inherently positive:

$$\ln(d^*_{jm}) \sim N(\ln(d_{jm}), \sigma^2) \qquad (3)$$

That is

$$f(d^*_{jm}) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}} d^*_{jm}} e^{\left(-\frac{1}{2\sigma^2}\left(\ln(d^*_{jm}) - \ln(d_{jm})\right)^2\right)} \qquad (4)$$

The log-normal is a more realistic model of the noise process because, by definition of the log-normal, $d^*_{jm} > 0$ so that $(d_{jm} + \varepsilon_{jm}) > 0$ and $\sigma > 0$, where σ is a ***shape parameter***. The mean and variance are:

$$E\left[d_{jm}^{*}\right]=d_{jm}e^{\frac{1}{2}\sigma^2} \quad \text{and} \quad VAR\left(d_{jm}^{*}\right)=\left(e^{\sigma^2}-1\right)d_{jm}^{2}e^{\sigma^2} \qquad (5)$$

So that as $d_{jm} \to 0$ $E\left(d_{jm}^{*}\right) \to 0$ and $VAR\left(d_{jm}^{*}\right) \to 0$. The upshot is that the smaller the observed distance the smaller the variance of that distance because $d_{jm}^{*} = d_{jm} + \varepsilon_{jm}$.

Our likelihood function is:

$$\mathbf{L^{*}}(Z_{jk}|\mathbf{D^{*}})=\frac{1}{(2\pi\sigma^2)^{\frac{q(q-1)/2}{2}}}\left(\prod_{j=1}^{q-1}\prod_{m=j+1}^{q}\frac{1}{d_{jm}^{*}}\right)e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^{*}\right)-\ln\left(\sqrt{\sum_{k=1}^{s}\left(Z_{jk}-Z_{mk}\right)^2}\right)\right)^2} \qquad (6)$$

To implement a Bayesian model we use simple normal prior distributions for the stimuli coordinates:

$$\boldsymbol{\xi}(Z_{jk}) = \frac{1}{(2\pi\kappa^2)^{\frac{1}{2}}}e^{-\frac{Z_{jk}^2}{2\kappa^2}} \qquad (7)$$

and a uniform prior for the variance term:

$$\xi(\sigma^2)=\frac{1}{c},\ 0< c < b \qquad (8)$$

Where, empirically, b is no greater than 2.

Hence, our posterior distribution is:

$$\boldsymbol{\xi}(Z_{jk}|\mathbf{D^{*}}) \propto \prod_{j=1}^{q-1}\prod_{m=j+1}^{q}\left\{f_{jm}(Z_{jm}\,|\,d_{jm}^{*})\right\}\xi(Z_{11})\xi(Z_{12})...\xi(Z_{1s})\xi(Z_{21})...\xi(Z_{qs})\xi(\sigma^2) \qquad (9)$$

Taking the log of the right hand side and dropping the unnecessary constants:

$$\ell n\xi \propto -\frac{q(q-1)/2}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^{*}\right)-\ln\left(\sqrt{\sum_{k=1}^{s}\left(Z_{jk}-Z_{mk}\right)^2}\right)\right)^2$$

$$-\frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z_{jk}^2\right)-\ln(c)=$$

$$-\frac{q(q-1)/2}{2}\ln(\sigma^2)-\frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d_{jm}^{*}\right)-\ln\left(d_{jm}\right)\right)^2-\frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z_{jk}^2\right)-\ln(c) \quad (10)$$