Tennis Project

1 The Bradley-Terry Model

We denote D to be the data of pairwise comparisons (matches) among K individuals. To compare two elements i, j, the Bradley-Terry model suggests,

$$\mathbb{P}(i > j) = \frac{\lambda_i}{\lambda_i + \lambda_j}, \tag{1.1}$$

where $i > j$ means that element i beats element j and $\lambda = \{\lambda_i\}_{i=1}^{K}$ is parameters with $\lambda_i > 0$ that represents the skill of i.

Let w_{ij} denote the number of matches where i beats j, then the total number of matches element i wins is, $w_i = \sum_{j=1, j \neq i}^{K} w_{ij}$ and $n_{ij} = w_{ij} + w_{ji}$ is the total number of matches between individuals i and j. Using equation 1.1, we have log-likelihood,

$$l_D(\lambda) = \sum_{1 \leq i \neq j \leq K} [w_{ij} \log \lambda_i - w_{ij} \log(\lambda_i + \lambda_j)] \tag{1.2}$$

$$= \sum_{1 \leq i \neq j \leq K} w_{ij} \log \lambda_i - \sum_{1 \leq i < j \leq K} n_{ij} \log(\lambda_i + \lambda_j), \tag{1.3}$$

with notation $1 \leq i \neq j \leq K$ denotes the set $\{(i,j) \in \{1, \cdots, K\}^2$ such that $i \neq j\}$ and $1 \leq i < j \leq K$ denotes the set $\{(i,j) \in \{1, \cdots, K\}^2$ such that $i < j\}$. Trying to maximise this for λ takes a long time to converge so we want to introduce a latent variable, say z, so the complete log-likelihood will have a simpler form. This is possible as, $p(z, D|\lambda) = p(z|D, \lambda)p(D|\lambda)$ meaning we can sum the log-likehoods,2 $l_{z,D}(\lambda) = l_z(\lambda) + l_D(\lambda)$.

For each $1 \leq i < j \leq K$ and for each match $k = 1, \cdots n_{ij}$, let $Y_{ki} \sim \text{Exp}(\lambda_i)$ and $Y_{kj} \sim \text{Exp}(\lambda_j)$ where $\text{Exp}(\lambda)$ is the exponential distribution

with parameter λ Therefore,

$$\mathbb{P}(Y_{ki} < Y_{kj}) = \frac{\lambda_i}{\lambda_i + \lambda_j}, \tag{1.4}$$

we can think of Y_{ki} and Y_{kj} as 'clocks' where the individual whose clock goes off first as the winner. The more skilled a player is the more likely their clock will go off first. Now, for each i, j pair we introduce latent random variable,

$$Z_{ij} = \sum_{k=1}^{n_{ij}} \min(Y_{kj}, Y_{ki}) \sim \text{Gamma}(n_{ij}, \lambda_i + \lambda_j), \tag{1.5}$$

$$p(z|D, \lambda) = \prod_{1 \leq i < j \leq K | n_{ij} > 0} \text{Gamma}(z_{ij}; n_{ij}, \lambda_i + \lambda_j), \tag{1.6}$$

$$l_z(\lambda) = \sum_{1 \leq i < j \leq K} [n_{ij} \log(\lambda_i + \lambda_j) - (\lambda_i + \lambda_j)z_{ij} + (n_i j - 1)z_{ij} - \log \Gamma(n_{ij})]$$
$$\tag{1.7}$$

Z_{ij} has a gamma distribution because the minimum of two exponential is exponentially distribution, $\min(Y_{kj}, Y_{ki}) \sim \text{Exp}(\lambda_i + \lambda_j)$ and the sum of identically distributed exponential random variables has a gamma distribution with the shape parameter being the number of random variables in the sum and the rate parameter is the parameter of the exponential distribution.

Therefore, we have complete log-likelihood,

$$l_c(\lambda) = l_D(\lambda) + l_z(\lambda) \tag{1.8}$$

$$l_c(\lambda) = \sum_{i=1}^{K} w_i \log \lambda_i - \sum_{1 \leq i < j \leq K | n_{ij} > 0} [(\lambda_i + \lambda_j)z_{ij} - (n_i j - 1)z_{ij} + \log \Gamma(n_{ij})]. \tag{1.9}$$

If we also use a a prior,

$$p(\lambda) = \prod_{i=1}^{K} \text{Gamma}(\lambda_i; a, b), \tag{1.10}$$

with shape parameter $a > 0$ and rate parameter $b > 0$. We can then maximise using the EM algorithm, where for iteration, t,

$$\lambda^{(t)} = \underset{\lambda}{\text{argmax}}\, Q(\lambda, \lambda^{(t-1)}) \tag{1.11}$$

where

$$Q(\lambda, \lambda^{(t-1)}) = \mathbb{E}_{Z|D,\lambda^{(t-1)}}(l_c(\lambda)) + \log p(\lambda) \tag{1.12}$$

$$= \sum_{i=1}^{K}[(a - 1 + w_i)\log \lambda_i - b\lambda_i] - \sum_{1 \leq i < j \leq K} \frac{n_{ij}(\lambda_i + \lambda_j)}{(\lambda_i^{(t-1)} + \lambda_j^{(t-1)})} \tag{1.13}$$

Using above,

$$\lambda_i^{(t)} = \frac{a - 1 + w_i}{b + \sum_{j \neq i} \frac{n_{ij}}{\lambda_i^{(t-1)} + \lambda_j^{(t-1)}}}. \tag{1.14}$$

We can also sample from the posterior distribution $p(\lambda, z|D)$ using a data augmentation algorithm. We can update $Z|\lambda$ using equation 1.5 and $\lambda|Z$ using $p(\lambda|z, D) \propto p(\lambda)\exp(l_c(\lambda))$, where $p(\lambda)$ is the conjugate prior to,

$$\exp(l_c(\lambda)) \propto \prod_{i=1}^{K} \left(\lambda_i^{w_i} \exp\left[-\left(\sum_{i<j|n_{ij}>0} z_{ij} + \sum_{i>j|n_{ij}>0} z_{ji} \right) \lambda_i \right] \right). \tag{1.15}$$

The sampler at iteration t is,

- For $1 \leq i < j \leq K$ such that $n_{ij} > 0$,

$$Z_{ij}^{(t)}|D, \lambda^{(t-1)} \sim \text{Gamma}(n_{ij}, \lambda_i^{(t-1)} + \lambda_j^{(t-1)}). \tag{1.16}$$

- For $i = 1, \cdots, K$,

$$\lambda_i^{(t)}|D, Z^{(t)} \sim \text{Gamma}\left(a + w_i, b + \sum_{i<j|n_{ij}>0} Z_{ij} + \sum_{i>j|n_{ij}>0} Z_{ji} \right). \tag{1.17}$$