

Inference on Sports with Beta-Binomial Model

Andrew White

November 19, 2013

1 Model

We're going to use a beta-binomial. The data will be a matrix of wins x_{ij} and a matrix of total games played between opponent i and j , n_{ij} . The probability of a win is p_{ij} which comes from a beta distribution with parameters (α_i, α_j) . The α s are the team strenghts, and are in units of psuedo-wins.

To generate samples from this model given a schedule, first sample the probability that team i beats team j (p_{ij}) from $\text{Beta}(\alpha_i, \alpha_j)$. Next, sample an outcome of the game given p_{ij} .

To sample parameters for this model, we'll use Markov Chain Monte Carlo. First, we sample each individual p_{ij} . Those are independent and distriubted according:

$$\Pr(p_{ij}|X, N, \vec{\alpha}) = \frac{\Gamma(\alpha_i + x_{ij}, \alpha_j + n_{ij} - x_{ij})}{\Gamma(\alpha_i + x_{ij})\Gamma(\alpha_j + n_{ij} - x_{ij})} p_{ij}^{x_{ij} + \alpha_i - 1} (1 - p_{ij})^{n_{ij} - x_{ij} + \alpha_j - 1}$$

Then we use a Metropolis-hastings move for the α vector. Stochastically, we choose an index k and then randomly perturb α_k until all of them have been sampled. The likelihood ratio is evaluated as:

$$\Pr(\vec{\alpha}|\dots) = \prod_{p_{ij} \in j > i} \frac{\Gamma(\alpha_i + \alpha_j + n_{ij})}{\Gamma(\alpha_i + x_{ij})\Gamma(\alpha_j + n_{ij} - x_{ij})} p_{ij}^{x_{ij} + \alpha_i - 1} (1 - p_{ij})^{n_{ij} - x_{ij} + \alpha_j - 1}$$

$$\frac{\Pr(\alpha'_k|\dots)}{\Pr(\alpha_k|\dots)} = \left[\prod_{p_{kj} \in j > k} \frac{\Gamma(\alpha'_k + \alpha_j + n_{kj})\Gamma(\alpha_k + x_{kj})}{\Gamma(\alpha_k + \alpha_j + n_{kj})\Gamma(\alpha'_k + x_{kj})} p_{kj}^{\alpha'_k - \alpha_k} \right] \cdot \left[\prod_{p_{ik} \in k > i} \frac{\Gamma(\alpha'_k + \alpha_i + n_{ik})\Gamma(\alpha_k + n_{ik} - x_{ik})}{\Gamma(\alpha_k + n_{ik} + \alpha_i)\Gamma(\alpha'_k + n_{ik} - x_{ik})} (1 - p_{ik})^{\alpha'_k - \alpha_k} \right] \quad (1)$$

Data: X (times team i beat team j), N (games played between i and j),
 m (number of teams)
Result: $\vec{\alpha}$
Initialize: $\alpha_i = \sum_j x_{ij}$, $p_{ij} = 0.5$;
while *True* **do**
 for i in $0 \dots m - 1$ **do**
 for j in $i + 1 \dots m - 1$ **do**
 Sample p_{ij} from $\text{Beta}(\alpha_i + x_{ij}, \alpha_j + n_{ij} - x_{ij})$ Set $p_{ji} = 1 - p_{ij}$;
 end
 end
 for sample k in $0 \dots m - 1$ **do**
 sample ϵ uniformly from $[-\eta, \eta]$;
 propose $\alpha'_k = \alpha_k + \epsilon$;
 Accept with probability $\min\left(\frac{\text{Pr}(\alpha'_k | \dots)}{\text{Pr}(\alpha_k | \dots)}, 1\right)$ (Eq. 1);
 end
 Histogram $\vec{\alpha}$;
end

Algorithm 1: Sampling parameters for the model

Data: N (games played between i and j), $\vec{\alpha}$, m (number of teams)
Result: X (the win matrix), $f(X)$ (a function of the result of the games)
while *True* **do**
 Initialize $X = 0$;
 for i in $0 \dots m - 1$ **do**
 for j in $i + 1 \dots m - 1$ **do**
 Sample p_{ij} from $\text{Beta}(\alpha_i, \alpha_j)$;
 Sample x_{ij} from $\text{Bin}(p_{ij}, n_{ij})$;
 Set $x_{ji} = n_{ij} - x_{ij}$;
 end
 end
 Histogram X and/or $f(X)$;
end

Algorithm 2: Generating realizations of the win matrix from the model

2 Algorithm

3 Logarithm Transformations

3.1 Equation 2

$$\begin{aligned} \ln \left(\frac{\Pr(\alpha'_k | \dots)}{\Pr(\alpha_k | \dots)} \right) = & \sum_{p_{kj} \in j > k} \ln \Gamma(\alpha'_k + \alpha_j + n_{kj}) + \ln \Gamma(\alpha_k + x_{kj}) + (\alpha'_k - \alpha_k) \ln p_{kj} - \\ & \ln \Gamma(\alpha_k + \alpha_j + n_{kj}) - \ln \Gamma(\alpha'_k + x_{kj}) + \\ & \sum_{p_{ik} \in k > i} \ln \Gamma(\alpha'_k + \alpha_k + n_{ik}) + \ln \Gamma(\alpha_k + n_{ik} - x_{ik}) + (\alpha'_k - \alpha_k) \ln(1 - p_{kk}) - \\ & \ln \Gamma(\alpha_k + \alpha_k + n_{ik}) - \ln \Gamma(\alpha'_k + n_{ik} - x_{ik}) \quad (2) \end{aligned}$$

4 Stirling's Approximation

$$\begin{aligned} \frac{\Gamma(x + \alpha)}{\Gamma(x + \beta)} & \approx x^{\beta - \alpha} \\ \ln \left(\frac{\Gamma(x + \alpha)}{\Gamma(x + \beta)} \right) & \approx (\beta - \alpha) \ln x \\ \ln \left(\frac{\Pr(\alpha'_k | \dots)}{\Pr(\alpha_k | \dots)} \right) & \approx \sum_{p_{kj} \in j > k} (\alpha'_k - \alpha_k) [\ln(\alpha_j + n_{kj}) + \ln(x_{kj}) + \ln p_{kj}] + \\ & \sum_{p_{ik} \in k > i} (\alpha'_k - \alpha_k) [\ln(\alpha_i + n_{ik}) + \ln(x_{ik}) + \ln(1 - p_{ik})] \quad (3) \end{aligned}$$