Explanatory or Predictive Models for Pair-Contest Success

Yuxi Jiang William Thomas James Thornton

University of Oxford
University of Warwick

April 26, 2019

Structured Bradley-Terry Model

- 'Ability', α_i, as a function of predictors
- Logistic probability function
- Generalised Linear Mixed Model

$$
\begin{aligned}
\lambda_i &= \log(\alpha_i) \\
&= \sum_{p=1}^{P} \beta_p x_{i,p} + b_i, \\
\eta_{i,j} &= \lambda_i - \lambda_j \\
\mathbb{P}(i\ beats\ j) &= \frac{1}{1 + e^{-\eta_{i,j}}}
\end{aligned}
$$

Penalized Quasi-Likelihood Approach

- $\mathbb{E}[\mathbf{y}|\mathbf{b}] = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b})$
- $\mathbb{E}(y_i|\mathbf{b}) = \mu_i^b$, $Var(y_i|\mathbf{b}) = \phi a_i v(\mu_i^b)$, where $v(.)$ is a specified variance function, a_i is a known constant and ϕ is a dispersion parameter. Assume that $\mathbf{b} \sim N(\mathbf{0}, \mathbf{D}(\theta))$.
- The integrated quasi-likelihood function is given by

$$e^{ql(\boldsymbol{\beta},\theta)} \propto |\mathbf{D}|^{-1/2} \int \exp\left\{-\frac{1}{2\phi}\sum_{i=1}^{n} d_i(y_i, \mu_i^b) - \frac{1}{2}\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b}\right\} \mathbf{db}$$

- We use Laplace's method of integral approximation and maximise the resulting expression to estimate $\boldsymbol{\beta}$ and θ.

Laplace Method Bias

- Laplace's method involves a normal approximation.
- Our data is Binomial and sparse.

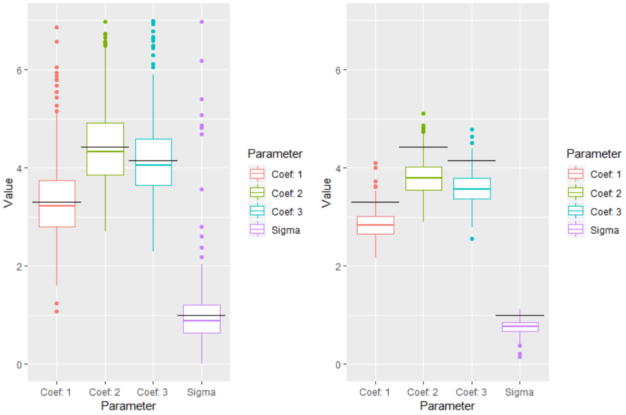


Figure: Summaries of parameters estimated from 500 contests, each with 1000 matches. Left: 10 players. Right: 100 players.

Sequential Reduction Method

- Cliques and Maximal cliques

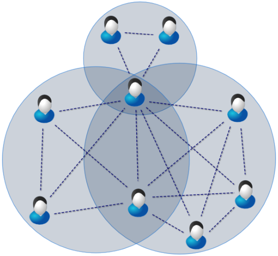


Figure: Source from:
http://c0de-x.com/how-to-find-maximal-cliques/cliques/

Sequential Reduction Method

- Conditional dependence structure of random effects
  - Posterior dependence graph
    - A vertex for each random effect
    - An edge between two vertices if in the design matrix for the random effects, the two columns contain at least a row that both have non-zero values
- Factorise the integrand of likelihood over maximal cliques

$$
\begin{aligned}
g(b_1, \ldots, b_m | \beta, D, W) &= \prod_{i=1}^{n} \prod_{j:j>i} f(W_{i,j} | \eta_{i,j}) \prod_{\ell=1}^{m} \varphi(b_\ell) \\
&= \prod_{C \in M(\mathcal{G})} g_C(\mathbf{b}_C)
\end{aligned}
$$

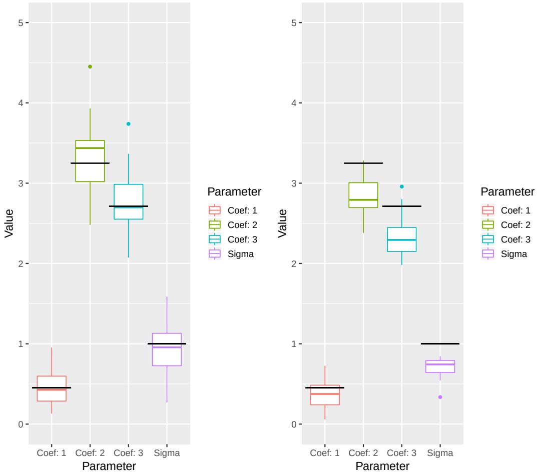
  - The Hammersley-Clifford theorem [Besag(1974)]

Sequential Reduction Method

1. Factorise the likelihood over maximal cliques
2. For each of the random effects
    1. Integrate over the random effect by integrating over the maximal cliques that contains the random effect
    2. Factorise the marginal density obtained over the maximal cliques of the new posterior dependence graph

Sequential Reduction vs PQL Results

- 100 players and the simulation is repeated 20 times for sequential reduction (left) and PQL (right).



MCMC Approach with Flat Prior Bayesian Formulation

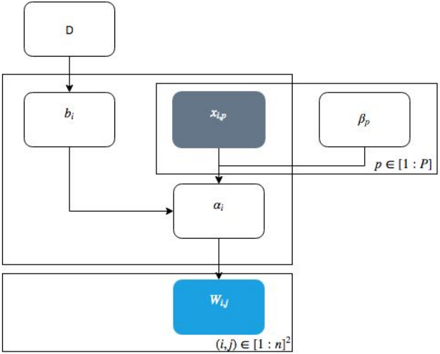


Figure: Plate diagram showing the dependency on observed win 'counts': $\{W_{ij}\}_{i,j}$ based on covariates $\{x_{i,j}\}_{i,j}$ and the unknowns: β_p, α_i, b_i and D

- Flat priors
- Gibbs sampling with nested rejection sampling
- [Zeger and Karim(1991)]

MCMC vs PQL Results

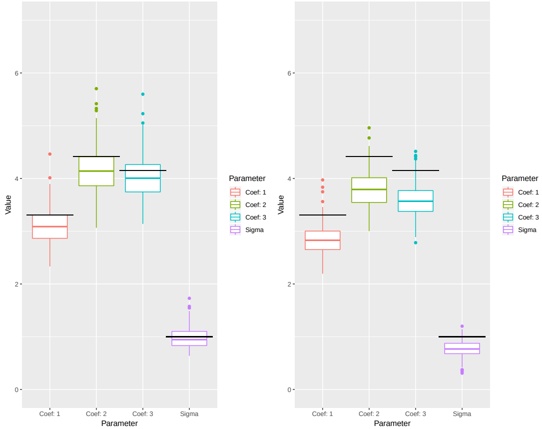


Figure: #players = 100, simulated data with 1000 matches.
Left: Boxplot of realised states in a single Markov Chain (of length 500) for Bayesian formulation detailed above. Right: $S = 500$ simulated PQL estimates. Horizontal black lines indicate the 'true' values from the data generating process.

Conclusion

- The PQL approach suffers from significant bias when faced with sparse, binary data
- Sequential reduction can also be computationally expensive in some scenarios
- The MCMC described approach is fairly complicated and computationally expensive
- Further work is required to identify an improved method

References

Julian Besag.
Spatial interaction and the statistical analysis of lattice systems.
Journal of the Royal Statistical Society. Series B (Methodological), pages 192–236, 1974.

Scott L Zeger and M Rezaul Karim.
Generalized linear models with random effects; A Gibbs sampling approach.
Journal of the American statistical association, 86(413):79–86, 1991.

Likelihood of Binomial GLMM

$$L(\beta, D) = \int \prod_{i=1}^{n} \prod_{j:j>i} f(W_{i,j}|\eta_{i,j})\varphi_D(b_i)\mathbf{db}$$

$$f(W_{i,j}|\eta_{i,j}) = \binom{W_{i,j} + W_{j,i}}{W_{i,j}} \left(\frac{1}{1 + e^{-\eta_{i,j}}}\right)^{W_{i,j}} \left(\frac{1}{1 + e^{\eta_{i,j}}}\right)^{W_{j,i}}$$

$$\varphi_D(b_i) = \frac{1}{|D|^{-\frac{1}{2}}} e^{-\frac{1}{2}b_i D^{-1} b_i}$$