Bayesian Modeling of Goal Arrival Times

Ilias Leriou*, Ioannis Ntzoufras* and Dimitris Karlis *

*Athens University of Economics and Business

Objectives

- Set the data in gap time form.
- Find a plausible parametric distribution for Bayesian survival modeling.
- Out of Sample prediction assessment.

Brief Review of the literature

1. Dixon and Robinson (1998)
   - Arrival times = two dimensional Poisson birth processes (Poisson regression creates proportional hazards models)
   - Analysis of 4000 games from English competitions. Rate increases during the game and it is influenced by score.
2. Del Corral (2008)
   - Analysis of first substitution time and their determinants in Spanish league for season 2004-5.
3. Nevo (2013)
   - Cox model for 1st & 2nd goal. 760 Premier League games (2 seasons, 2008-2010).

Data Layout

Let t_{1im} and t_{2im} be the event gap times for team 1 and team 2 respectively with $i = 1, 2, \ldots n$ and $m = 1, 2, \ldots, M$ the game indicator.

Game	t_1	t_2	c_1	c_2	Home Team	Away Team
1	50	NA	0	50	Benfica	PFC CSKA Moskva
1	NA	13	13	0	Benfica	PFC CSKA Moskva
1	NA	8	8	0	Benfica	PFC CSKA Moskva
1	NA	NA	23	23	Benfica	PFC CSKA Moskva
...	...	...	...	...	...	...

Initial Model Comparisons

Fixed Effects Model	WAIC	LOOIC
Weibull	4029.7	4032.9
Exponential	4040.3	4042.4
Log-Logistic	4064.4	4065.0

Final Model: Fixed Effects Weibull

The model's structure is presented below:

$$T_{ij} \sim Weibull(\gamma, \lambda_{ij}), j = 1, 2, i = 1, 2, \ldots, n$$

with

$$log\left[E(T_{i1})\right] = \mu + home + a_{HT_i} + d_{AT_i}$$

$$log\left[E(T_{i2})\right] = \mu + a_{AT_i} + d_{HT_i}$$

where

$$E(T_{ij}) = \lambda_{ij}^{-\frac{1}{\gamma}} \Gamma(1 + 1/\gamma)$$

with $i = 1, 2, \ldots, n$, $j = 1, 2$.

Weakly informative priors to parameters as follows:

$$a_k, d_k, \mu, home \sim Normal(0, 10^{-3})$$

while the following weakly informative Gamma prior was assigned to the positive parameter γ:

$$\gamma \sim Gamma(10^{-3}, 10^{-3})$$

STZ constrains for attacking and defensive parameters to allow for comparisons of the abilities of each team with the overall level of the fixed effects:

$$\sum_{k=1}^{K} a_k = 0, \quad \sum_{k=1}^{K} d_k = 0$$

Inference

	Mean	sd	2.5%	97.5%
μ	4.43	0.1	4.25	4.63
γ	1.13	0.05	1.04	1.23
home	-0.23	0.09	-0.41	-0.05

Additionally, the estimates regarding the attacking and defensive abilities seem to largely agree with the most component teams in the League.

Out-of-sample predictions: Quarter Finals

Teams	1	X	2	Actual Score
Sevilla FC - FC Bayern Munchen	0.16	0.09	0.75	$1-2$
Juventus Football Club - Real Madrid CF	0.37	0.19	0.44	$0-3$
FC Barcelona - AS Roma	0.66	0.31	0.03	$4-1$
Liverpool FC - Manchester City FC	0.49	0.15	0.36	$3-0$
Manchester City FC - Liverpool FC	0.46	0.16	0.38	$1-2$
AS Roma - FC Barcelona	0.05	0.37	0.59	$3-0$
Real Madrid CF - Juventus Football Club	0.58	0.17	0.25	$1-3$
FC Bayern Munchen - Sevilla FC	0.83	0.07	0.10	$0-0$

Out-of-sample prediction: Final Real Madrid - Liverpool

Posterior Median point wise prediction of the result: $3-2$

Winning team according to our model: Real Madrid

Model	1	X	2
Weibull	0.67	0.13	0.20
Double Poisson	0.58	0.27	0.15

Acknowledgements

The research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under the HFRI PhD Fellowship grant 186758/I2/01/11/2017.