
Modeling goal arrival times in Association Football
A Bayesian Analysis of Euro 2012 Data


Contents
• Motivation and short review
• The model
• Analysis of Euro 2012 Data
– Interpretation of the results
– GOF/Prediction
• Issues and Future work

Motivation

AIM

To investigate the possible modeling of goal arrival times using Bivariate distributions

Short review

- Dixon & Robinson (1998). *The Statistician.*

  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.

- Yu & Zhong (2005). *Econometric Th. & Appl*

  Weibull model for English Premiership data from 2001 – 2004. No significant effect for Beckham. No online covariates. No joint modeling

Short review (2)

- Thomas (2007). *J. Quant. Anal. in Sports*
  Analysis of inter-arrival times of goals in ice hockey using Weibull & Plateau-Hazard distributions
- Corral et al. (2007). *J. Sports Economics*
  Analysis of first substitution time and their determinants in Spanish league for season 2004-5
- Nevo & Ritov (2012). *arXiv:1207.6796v1*.
  Cox model for 1st & 2nd goal. 760 Premier League games (2 seasons, 2008-2010)

The model

Normal random effects

⇒For each game and

⇒For each group/phase.

Joint Modeling

Bivariate Weibull Marshall-Oilkin distribution for goal arrival times

The model

The Bivariate Weibull Marshall-Oilkin distribution

If $(X_1,X_2) \sim MO(r, \lambda_0, \lambda_1, \lambda_2)$ then

$$S(x_1, x_2) = S_w(x_1, r, \lambda_1) S_w(x_2, r, \lambda_2) S_w(\max\{x_1, x_2\}, r, \lambda_0)$$

where $S_w(x, r, \lambda)$ is the survival function of the Weibull distribution with shape parameter r and scale parameter λ; mean = $\lambda^{-1/k}\Gamma(1+1/r)$.

The model

The BW Marshall-Oilkin density function

If $(X_1,X_2) \sim MO(r, \lambda_0, \lambda_1, \lambda_2)$ then

$$f(x_1, x_2) = \begin{cases} f_w(x_1, r, \lambda_1) f_w(x_2, r, \lambda_0 + \lambda_2) & \text{if } x_1 < x_2 \\ f_w(x_2, r, \lambda_2) f_w(x_1, r, \lambda_0 + \lambda_1) & \text{if } x_2 < x_1 \\ \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} f_w(x_1, r, \lambda_0 + \lambda_1 + \lambda_2) & \text{if } x_1 = x_2 \end{cases}$$

where $f_w(x, r, \lambda)$ is the density function of the Weibull distribution with shape parameter r and scale parameter λ; mean = $\lambda^{-1/k}\Gamma(1+1/r)$.

The model

Covariates on the mean

a) Offline covariates

1. Team effects
2. UEFA coefficient (points) of the team
3. UEFA coefficient difference
4. UEFA ranking
5. UEFA ranking difference

UEFA coefficient

Co-hosts Poland (coefficient 23,806, rank 28) and Ukraine (coefficient 28,029, rank 15) .

The model

Covariates on the mean

b) Online covariates

1. Indicator for one goal difference
2. [Different] Effect for goal difference>2 (abs values)
3. Interaction of the goal difference parameters with uefa coefficient difference (points)
4. Sum of scored goals
5. Remaning time (at previous occasion)
6. Time from previous goal

The model

Covariates on the mean

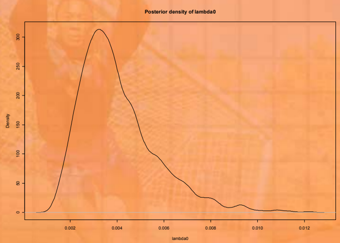
$$
\begin{aligned}
\log(\mu_{ij}) = m &+ game.re + round.re + att_i + def_j \\
&+ \beta_1 \Delta goal_1 + \beta_1 \Delta goal_1 \times \Delta uefa_{ij} \\
&+ \beta_3 \Delta goal_2 + \beta_4 \Delta .goal_2 \times \Delta uefa.c_{ij} \\
&+ \beta_5 (goals_i + goals_j) + \beta_6 rem.t + \beta_7 prev.sc.t \\
&+ \beta_8 uefa_i + \beta_9 \Delta Uefa_{ij} \\
&+ \beta_{10} uefa.rank_i + \beta_{11} \Delta uefa.rank_{ij}
\end{aligned}
$$

The model

Other details

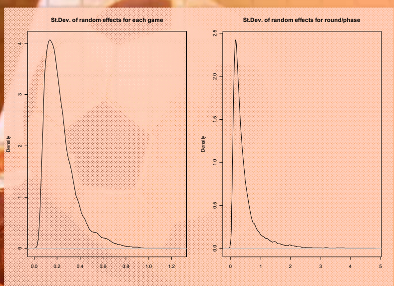
- Bayesian approach
- Flat priors
- Bayesian variable selection

Results for Euro 2012



Dependence

	mean	sd	2.5%	97.5%
$\log(\lambda_0)$	-5.591	0.388	-6.349	-4.816
λ_0	0.004	0.002	0.002	0.008



	mean	sd	2.5%	97.5%
s.game	0.229	0.137	0.071	0.607
s.round	0.438	0.441	0.077	1.718

Results for Euro 2012

Online Covariates

		mean	sd	2.5%	97.5%	min(p0,1-p0)	
Goal1	b[2]	0.597	0.276	0.076	1.159	0.011	**
Goal1xDUEFA	b[3]	-0.040	0.040	-0.121	0.037	0.156	
Goal2	b[4]	0.695	0.296	0.179	1.324	0.005	***
Goal2xDUEFA	b[5]	-0.070	0.038	-0.150	0.000	0.025	**
Goal sum	b[6]	-0.171	0.327	-0.847	0.409	0.326	
Rem. Time	b[7]	-0.031	0.021	-0.079	0.005	0.043	*
Prev. goal	b[8]	-0.025	0.014	-0.057	0.001	0.028	*

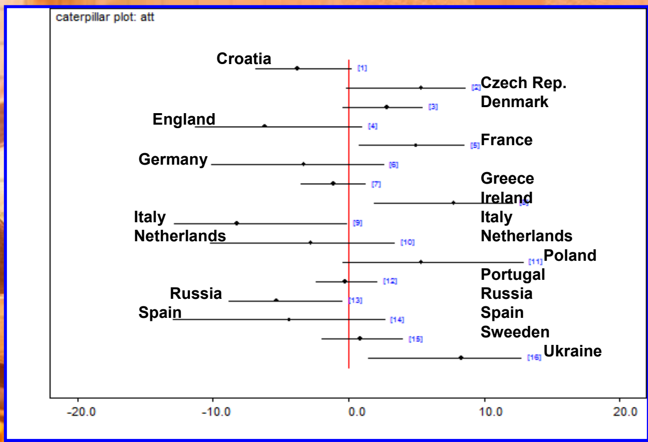
Results for Euro 2012

Offline Parameters

		mean	sd	2.5%	97.5%	min(p0,1-p0)
UEFA	b[9]	-0.957	0.693	-1.944	0.798	0.077
UEFA diff	b[10]	-0.189	0.534	-1.070	0.736	0.389
UEFA Rank	b[11]	-1.395	1.407	-3.850	0.292	0.192
UEFA DRank	b[12]	-0.607	0.772	-1.751	0.836	0.302
Shape param	r	0.932	0.132	0.675	1.201	----

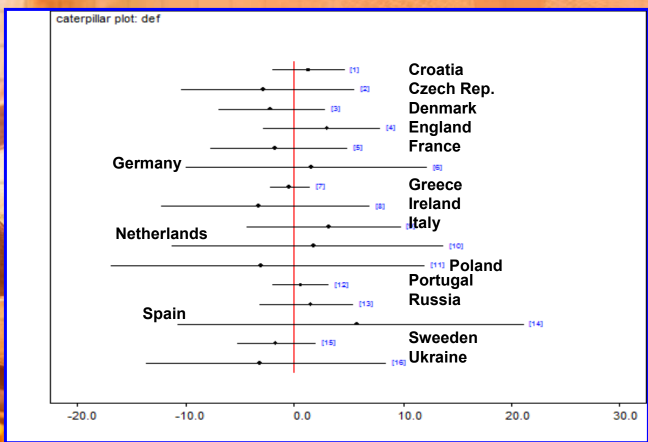
Results for Euro 2012

Team relative attacking strengths and their effect on expected arrival times



Results for Euro 2012

Team relative defensive strengths and their effect on expected arrival times



Results for Euro 2012

Posterior medians of team relative strengths
and their effect on expected arrival times

Team		mean	sd	2.5%	97.5%		mean	sd	2.5%	97.5%
Croatia	att[1]	-3.96	2.09	-7.79	0.22	def[1]	1.71	1.58	-1.49	4.64
Czech Rep.	att[2]	5.15	2.25	-0.24	9.02	def[2]	-3.99	3.86	-10.41	3.37
Denmark	att[3]	2.73	1.57	-0.55	5.42	def[3]	-2.76	2.41	-6.94	1.88
England	att[4]	-6.41	3.83	-14.35	1.00	def[4]	3.72	2.62	-1.97	7.86
France	att[5]	4.83	1.91	0.70	8.49	def[5]	-2.74	3.10	-7.70	2.80
Germany	att[6]	-2.16	4.92	-10.08	7.63	def[6]	3.06	6.01	-5.59	12.26
Greece	att[7]	-1.34	1.46	-4.49	1.25	def[7]	-0.40	0.96	-2.34	1.39
Ireland	att[8]	7.50	2.64	1.75	12.59	def[8]	-4.56	4.92	-12.29	4.23
Italy	att[9]	-8.58	3.70	-15.61	-0.09	def[9]	4.19	3.46	-2.94	9.78
Netherlands	att[10]	-1.65	4.93	-10.18	8.91	def[10]	3.52	6.63	-6.47	13.70
Poland	att[11]	4.08	5.18	-8.40	12.89	def[11]	-5.22	7.50	-16.91	6.66
Portugal	att[12]	-0.62	1.49	-3.58	2.12	def[12]	0.41	1.21	-1.92	2.90
Russia	att[13]	-5.64	2.62	-10.48	-0.45	def[13]	1.87	2.17	-2.85	5.42
Spain	att[14]	-2.72	6.41	-12.89	11.69	def[14]	8.17	8.07	-4.46	21.15
Sweeden	att[15]	0.77	1.53	-2.08	4.07	def[15]	-2.11	1.75	-5.25	1.41
Ukraine	att[16]	8.04	2.78	1.45	12.72	def[16]	-4.87	5.55	-13.65	4.77

Results for Euro 2012

Posterior medians of team relative strengths (compared to an overall average) and their effect on expected arrival times

Model with covariates

	Att	Def
Croatia	1/53	6
Czech Rep.	253	1/50
Denmark	18	1/15
England	1/470	67
France	138	1/16
Germany	1/22	9
Greece	1/3	1
Ireland	1942	1/99
Italy	1/6829	97
Netherlands	1/10	19
Poland	83	1/57
Portugal	1	2
Russia	1/2	8
Spain	1/263	819
Sweeden	2	1/9
Ukraine	3722	1/136

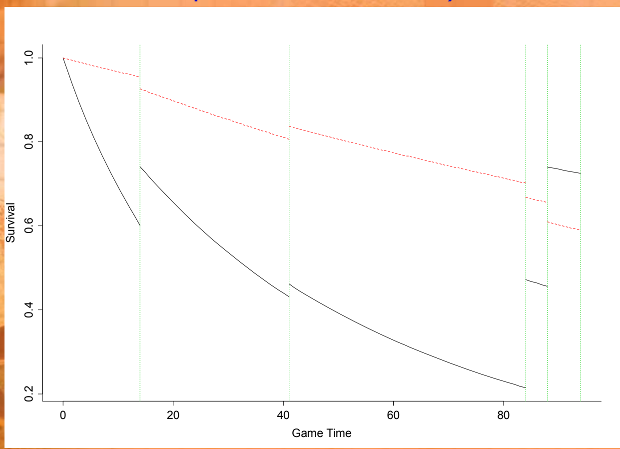
Model without covariates

	Att	Def
Croatia	1.51	0.75
Czech R.	0.84	1.90
Denmark	1.55	1.45
England	0.84	0.60
France	0.69	1.08
Germany	2.07	1.21
Greece	1.00	1.80
Ireland	0.68	2.05
Italy	1.48	0.65
Netherlnds	0.70	1.35
Poland	0.41	1.09
Portugal	0.97	0.55
Russia	1.21	1.52
Spain	2.36	0.10
Sweeden	1.82	2.52
Ukraine	0.44	1.20

Results for Euro 2012

Survival plot for the final of Euro 2012

(based on the actual score vs. the posterior medians of the expected arrival times)



Issues for further investigation

- Bivariate inverse Gaussian distribution
- More distributions?
- Implementation on other data
- Other covariates?
- Appropriate prior for Bayesian variable selection
- How to handle response

Here we have used a competing risk approach (the arrival goal time for the scoring team was considered as censoring time for the opponent)