Bayesian Modelling of Volleyball Sets

V. Palaskas[1], I. Ntzoufras[2], S. Drikos[3]

[1]Department of Statistics
AUEB Sports Analytics Group

[2]Department of Statistics
AUEB Sports Analytics Group

[3]AUEB Sports Analytics Group

Mathsport International Conference, July 2019

Overview

Introduction

- Objectives of this work:

Introduction

- Objectives of this work:
  1. To propose and compare models based on the multinomial and Skellam distributions for the set difference.

Introduction

- Objectives of this work:
  1. To propose and compare models based on the multinomial and Skellam distributions for the set difference.
  2. To implement Bayesian variable selection in skill events.

Introduction

- Objectives of this work:
    1. To propose and compare models based on the multinomial and Skellam distributions for the set difference.
    2. To implement Bayesian variable selection in skill events.
    3. Finally, to compare the out-of-sample prediction accuracy between models with and without skills (as additional covariates).

Greek A1 professional men's League regular season 2016-2017

- The data concern the Greek A1 professional men's League regular season 2016-2017. All data have been collected by Dr. Sotiris Drikos.
- For each regular season game ($n = 132$):
  1. **Set score**
  2. **Skill events**

Skills with their actions and outcomes (1)

Main skills	Action	Outcome	Ordinal Levels
Serve	Failed execution	Loss of a point	1
	Very good execution	The ball is in the field of the team served	2
	Perfect execution	Win of a point	3
Pass	Failed execution	Loss of a point	1
	Poor execution	The ball was in the possession of the opponent	2
	Very good execution	The ball is in the possession of the team with good perspective.	3
	Perfect execution	The ball is in the possession of the team with excellent perspective.	4

Skills with their actions and outcomes (2)

Main skills	Action	Outcome	Ordinal Levels
Attack of type 1,2	Failed execution	Loss of a point	1
	Blocked attack	Loss of a point	2
	Perfect Execution	Win of a point	3
Block	Failed execution	Loss of a point	1
	Violation in the net	Loss of a point	2
	Perfect execution	Win of a point	3
Setting	Failed execution	Loss of a point	1

Volleyball Rules

1. The possible results for a team in a volleyball game are either **win** or **loss**.

Volleyball Rules

1. The possible results for a team in a volleyball game are either **win** or **loss**.
2. Each team can reach minimum the **zero** and maximum the **three** winning sets. As a result, the set difference is defined in support $\{-3, -2, -1, 1, 2, 3\}$.

Volleyball Rules

1. The possible results for a team in a volleyball game are either **win** or **loss**.
2. Each team can reach minimum the **zero** and maximum the **three** winning sets. As a result, the set difference is defined in support $\{-3, -2, -1, 1, 2, 3\}$.
3. The **exact** set difference in a volleyball game determines the points earned by two competing teams.

Bayesian models for the set difference.

- We mainly focus on the ordered logistic as well as the truncated version of Skellam model.

Bayesian models for the set difference.

- We mainly focus on the ordered logistic as well as the truncated version of Skellam model.
- We implement Bayesian modelling via the probabilistic programming language of Stan.

Bayesian models for the set difference.

- We mainly focus on the ordered logistic as well as the truncated version of Skellam model.
- We implement Bayesian modelling via the probabilistic programming language of Stan.
- Low informative priors $\rightarrow N(0, 10^4)$ (except for rare cases)

Ordered logistic model with team abilities

We have defined the ordinal response of the set difference (Y)

$$Y_k = k - 3 - \delta,$$

$$\textit{where } \delta = \begin{cases} 1, & \text{k= 1, 2, 3} \\ 0, & \text{k= 4, 5, 6} \end{cases}$$

for $k = 1, \ldots, K$ possible set differences (here $K = 6$). The model using cumulative probabilities of set differences (γ_{ik}) is specified by

$$\log \frac{\gamma_{ik}}{1 - \gamma_{ik}} = c_k - (gen\_abil_{ht_i} - gen\_abil_{at_i}),\ for\ k = 1, \ldots, K - 1$$

c_k: constant parameters, ht_i, at_i: home and away teams, gen_abil_g: general ability of g team, for $g = 1, \ldots, G$; G: number of teams (here $G = 12$). For the ability parameters, a sum to zero constraint is used.

Posterior summary statistics of ordered logistic model (1)

	Set differences					
Model	-3	-2	-1	1	2	3
Ordered logistic	5.47	17.04	16.73	19.83	29.01	11.92

Table: Percentages (%) of set differences when two equal teams are competing.

Posterior summary statistics of ordered logistic model (1)

-

	Set differences					
Model	-3	-2	-1	1	2	3
Ordered logistic	5.47	17.04	16.73	19.83	29.01	11.92

Table: Percentages (%) of set differences when two equal teams are competing.

- When two **equal** teams are competing:

Posterior summary statistics of ordered logistic model (1)

- **Set differences**

Model	-3	-2	-1	1	2	3
Ordered logistic	5.47	17.04	16.73	19.83	29.01	11.92

Table: Percentages (%) of set differences when two equal teams are competing.

- When two **equal** teams are competing:
  1. It is more likely for the home team to win (incorporation of a common home effect).

Posterior summary statistics of ordered logistic model (1)

-

	Set differences					
Model	-3	-2	-1	1	2	3
Ordered logistic	5.47	17.04	16.73	19.83	29.01	11.92

Table: Percentages (%) of set differences when two equal teams are competing.

- When two **equal** teams are competing:
  1. It is more likely for the home team to win (incorporation of a common home effect).
  2. The set differences $-3, 3$ are unlikely to be observed.

Posterior summary statistics of ordered logistic model (1)

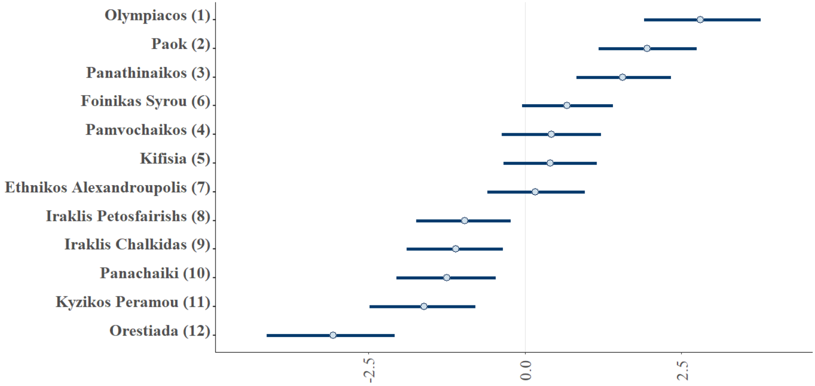

Figure: 95% Posterior intervals of the "net" general ability of all teams. The points are the posterior means; within brackets the observed ranking.

Skellam's distribution

Skellam's (or Poisson Difference) distribution with

$$\begin{cases} \textit{Mean} & E(Z) = \lambda_1 - \lambda_2 \\ \textit{Variance} & Var(Z) = \lambda_1 + \lambda_2 \end{cases}$$

and density function

$$f_{Sk}(z|\lambda_1, \lambda_2) = e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{\frac{z}{2}} I_{|z|}\left(2\sqrt{\lambda_1\lambda_2}\right)$$

where

$$I_r(x) = \left(\frac{x}{2}\right)^r \sum_{k=0}^{\infty} \frac{\left(\frac{x^2}{4}\right)^k}{k!\Gamma(r+k+1)}$$

is the modified Bessel function of order r (for more details see Abramowitz and Stegun 1965).

Zero-Deflated and Truncated Skellam (ZDTS)

$$Z \sim ZDTS(\lambda_1, \lambda_2)$$

- Density function

$$f_{ZDTS}(z|\lambda_1, \lambda_2) = \begin{cases} \frac{f_{Sk}(z|\lambda_1,\lambda_2)}{\sum_{z=-3}^{-1} f_{Sk}(z|\lambda_1,\lambda_2) + \sum_{z=1}^{3} f_{Sk}(z|\lambda_1,\lambda_2)}, & z \in \{-3,-2,-1,1,2,3\} \\ 0, & z \notin \{-3,-2,-1,1,2,3\} \end{cases}$$

Zero-Deflated and Truncated Skellam (ZDTS)

$$Z \sim ZDTS(\lambda_1, \lambda_2)$$

- Density function

$$f_{ZDTS}(z|\lambda_1, \lambda_2) = \begin{cases} \frac{f_{Sk}(z|\lambda_1,\lambda_2)}{\sum_{z=-3}^{-1} f_{Sk}(z|\lambda_1,\lambda_2)+\sum_{z=1}^{3} f_{Sk}(z|\lambda_1,\lambda_2)}, & z \in \{-3,-2,-1,1,2,3\} \\ 0, & z \notin \{-3,-2,-1,1,2,3\} \end{cases}$$

- Mean:

$$E(Z_{ZDTS}) = (\lambda_1 - \lambda_2) \frac{I_1 + \frac{2I_2*(\lambda_1+\lambda_2)}{(\lambda_1\lambda_2)^{\frac{1}{2}}} + \frac{3I_3*\left(\lambda_1^2+\lambda_1\lambda_2+\lambda_2^2\right)}{\lambda_1\lambda_2}}{I_1 * (\lambda_1 + \lambda_2) + \frac{I_2*\left(\lambda_1^2+\lambda_2^2\right)}{(\lambda_1\lambda_2)^{\frac{1}{2}}} + \frac{I_3*\left(\lambda_1^3+\lambda_2^2\right)}{\lambda_1\lambda_2}}$$

where $I_r = I_r\left(2\sqrt{\lambda_1\lambda_2}\right)$.

Zero-Deflated and Truncated Skellam (ZDTS) model

In terms of convenience, a similar structure for parameters λ_1, λ_2 was adopted as in simple Skellam model (see Karlis and Ntzoufras 2006).

$$\log(\lambda_{1i}) = \mu + home + att_{ht_i} + def_{at_i}$$
$$\log(\lambda_{2i}) = \mu + att_{at_i} + def_{ht_i}$$

μ: constant, *home*: common home effect, ht_i, at_i: home and away team in i game, att_g, def_g: "net" attacking and defensive abilities of g team, for $g = 1, \ldots, G$; G: number of teams in the data (here $G = 12$). For both attacking and defensive abilities, a sum to zero constraint is used.

ZDTS: Priors specification

In order to avoid **numeric overflow problem** for model parameters λ_1, λ_2, we adopt the specific priors in order to be consistent with our distributional assumptions

$$\begin{aligned} \mu &\sim N(0, 0.37^2) \\ home &\sim N(0, 0.37^2) \\ att &\sim N(0, 1^2) \\ def &\sim N(0, 1^2) \end{aligned} \quad (1)$$

Posterior summary statistics of ZDTS model

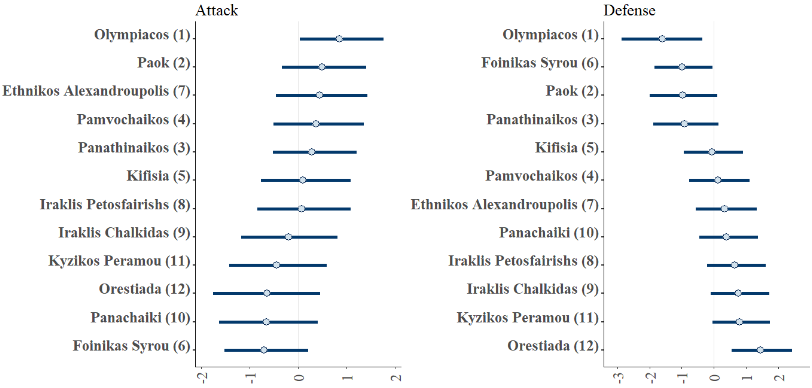

Figure: 95% Posterior intervals of both "net" attacking and defensive abilities of all teams. The points are the posterior means; within brackets the observed ranking.

Goodness-of-fit comparison (Frequencies)

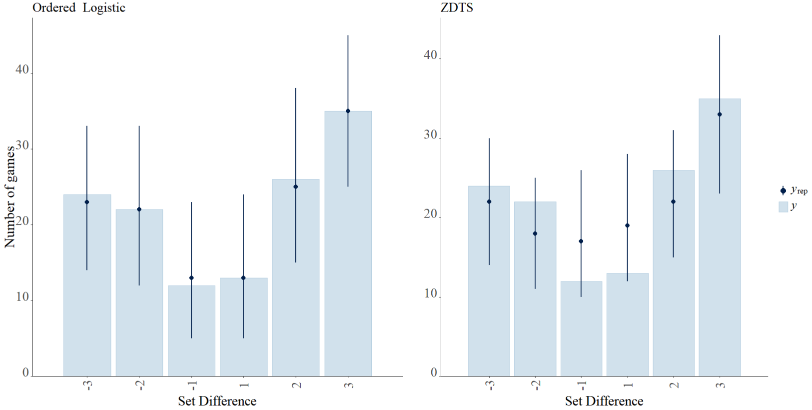

Figure: 95% Posterior intervals of predicted frequencies of set differences. y_{rep} and y are the generated (median) and observed quantities, respectively.

Ordered: Goodness-of-fit (Ranking)

Posterior Predictive Table			Observed Final Table		
Pred.(Obs.) Ranking	Teams	Mean Post. Points	Obs. Ranking	Teams	Obs. Points
1 (1)	Olympiacos	58.87	1	Olympiacos	62
2 (2)	Paok	53.00	2	Paok	53
3 (3)	Panathinaikos	49.20	3	Panathinaikos	50
4 (6)	Foinikas Syrou	40.28	4	Pamvochaikos	39
5 (5)	Kifisia	36.89	5	Kifisia	37
6 (4)	Pamvochaikos	35.06	6	Foinikas Syrou	37
7 (7)	Ethnikos Alexan.	34.81	7	Ethnikos Alexan.	36
8 (8)	Iraklis Petosfairishs	22.51	8	Iraklis Petosfairish	28
9 (9)	Iraklis Chalkidas	21.17	9	Iraklis Chalkidas	16
10 (10)	Panachaiki	19.70	10	Panachaiki	16
11 (11)	Kyzikos Peramou	16.34	11	Kyzikos Peramou	14
12 (12)	Orestiada	5.70	12	Orestiada	7

M.A.E $= 2.41$

ZDTS: Goodness-of-fit (Ranking)

Posterior Predictive Table			Observed Final Table		
Pred.(Obs.) Ranking	Teams	Mean Post. Points	Obs. Ranking	Teams	Obs. Points
1 (1)	Olympiacos	59.98	1	Olympiacos	62
2 (2)	Paok	52.97	2	Paok	53
3 (3)	Panathinaikos	49.87	3	Panathinaikos	50
4 (6)	Foinikas Syrou	38.04	4	Pamvochaikos	39
5 (4)	Pamvochaikos	37.88	5	Kifisia	37
6 (5)	Kifisia	36.17	6	Foinikas Syrou	37
7(7)	Ethnikos Alexan	35.64	7	Ethnikos Alexan.	36
8 (8)	Iraklis Petosfairishs	24.88	8	Iraklis Petosfairishs	28
9 (9)	Iraklis Chalkidas	19.54	9	Iraklis Chalkidas	16
10 (10)	Panachaiki	18.82	10	Panachaiki	16
11 (11)	Kyzikos Peramou	15.67	11	Kyzikos Peramou	14
12 (12)	Orestiada	6.55	12	Orestiada	7

M.A.E $= 1.51$

Goodness-of-fit model comparison

Comparison	Deviation	
	Ordered logistic	ZDTS
1. Frequencies of set difference results ($\lvert Q \rvert = 6$)	0.50	4.04
2. Relative frequencies of set difference results ($\lvert Q \rvert = 6$) (%)	0.38	3.06
3. Expected set difference ($\lvert Q \rvert = 132$)	0.92	0.90
4. Expected points ($\lvert Q \rvert = 12$)	2.41	1.51
5. Expected Total set difference ($\lvert Q \rvert = 12$)	2.19	0.98

Table: Deviations between observed and predicted measures for the regular season.

Bayesian variable selection

- Problem of handling discrete parameters$\rightarrow$ interconnected algorithm between RStan and R.

Bayesian variable selection

- Problem of handling discrete parameters$\rightarrow$ interconnected algorithm between RStan and R.
- Ordered logistic: the skill events differences (**17 in total**).

Bayesian variable selection

- Problem of handling discrete parameters$\rightarrow$ interconnected algorithm between RStan and R.
- Ordered logistic: the skill events differences (**17 in total**).
- ZDTS: both home and away team skill events (**34 in total**).

Bayesian model comparison via information criteria

	Ordered logistic		
	With team abilities	With team abilities and skills	With skills
AIC	380	254.6	246.5
BIC	429.0	326.7	284.0
DIC	377.5	250.4	245.2
WAIC	379.7	258.8	247.5

	(ZDTS)		
	With team abilities	With team abilities and skills	With skills
AIC	391.6	318.3	316.3
BIC	466.5	422	351
DIC	379.5	328.9	364.3
WAIC	379.6	308.5	331.8

Ordered logistic with skills

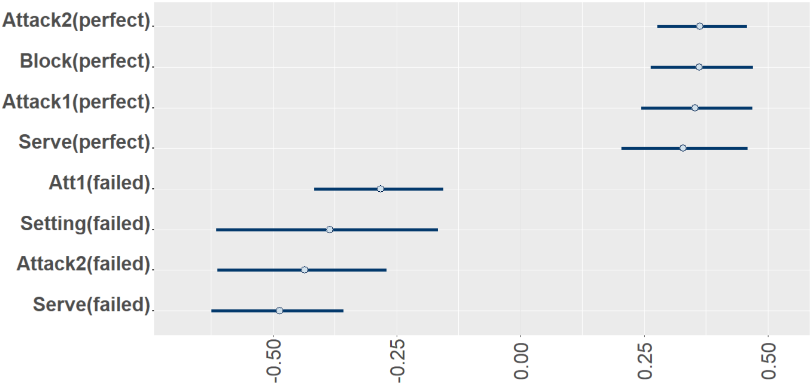

Figure: 95% Posterior intervals of skill events differences. The points are the posterior means.

ZDTS with team abilities and skills

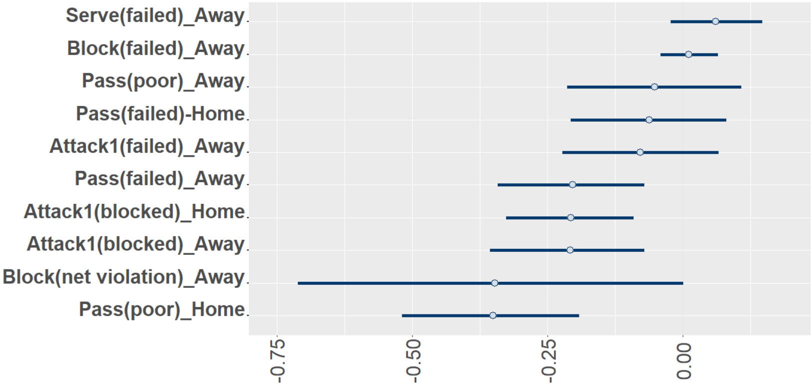

Figure: 95% Posterior intervals of both teams' skill events. The points in the above plot are posterior means.

Out-of-sample prediction: Split-Half season

Comparison	Deviation	
	Ordered Logistic	ZDTS
1. Frequencies of set difference results ($\|Q\| = 6$)	2.12	2.17
2. Relative frequencies of set difference results ($\|Q\| = 6$) (%)	3.22	3.29
3. Expected set difference ($\|Q\| = 66$)	1.04	1.10
4. Expected points ($\|Q\| = 12$)	1.91	2.71
5. Expected Total set difference ($\|Q\| = 12$)	3.52	4.57

Table: Deviations between predicted and observed measures for the second half of regular season (as test sample).

Out-of-sample prediction: Play-off

Comparison	Qualification probability	
	Ordered Logistic	ZDTS
1. Quarter finals (1-8) (%)	66.63	73.91
2. Semi finals (1-4) (%)	66.83	84.81
3. Finals (1-2) (%)	1.05	0.15

Table: Qualification probabilities in each play-off round of both models with team abilities; within brackets the number of teams in the specific play-off round.

Further research

- Model for the point difference in each set.

Further research

- Model for the point difference in each set.
- Use of one constant parameter instead of using one constant parameter for each cumulative logit.

Thank you for your attention!!!

GVS implementation (Rstan+R)

Require: Use of R

Random initialization of the parameters γ, β $(\beta^{(0)}, \gamma^{(0)})$.

for `(t in 1:T)` **do**

    Data input (**including** $\boldsymbol{\gamma^{(t-1)}}$) needed for running the model through Rstan.

Require: Use of Rstan

    Run the model through Rstan for one iteration in order to update the $\beta^{(t)}$ from the full conditional posterior distributions. `// Initial values` $\beta^{(t-1)}$

Require: Use of R

    for `(j in 1:p)` **do**

        Calculation of the O_j and then update the $\gamma_j^{(t)}$ from Bernoulli$\left(\frac{O_j}{1+O_j}\right)$

    end for

    Store the values of $\gamma_j^{(t)}$ in a $T \times p$ matrix

end for

return The posterior inclusion probabilities $\hat{\gamma}_j = \frac{\sum_{t=1}^{T} I\left(\gamma_j^{(t)}=1\right)}{T}$ for j variable, $j = 1, \ldots, p$.

Linear relationship between $E(Z_{ZDTS})$ and $E(Z_{Sk})$