DEMONSTRATIO MATHEMATICA
Vol. XXXIV No 2 2001

Bolesław Kopociński

UNFINISHED LEAGUE SEASON OF FOOTBALL

Dedicated to Professor Kazimierz Urbanik

1. Polish football league of the 1939 season

The Polish football (soccer) league of the 1939 season was interrupted because of the aggression of the neighbor countries on Poland. At that time there were still left 4 – 7 games, different for the participating teams, to be played. The leader of the table of the league at the stopping time (see Table 1) was Ruch Chorzów with possession of 18 points from 14 games, preceding Wisła Cracow with possession of 16 points from 12 games and Pogoń Lwów with possession of 16 points from 13 games. In the preceding 1938 season the champion of the league was Ruch (27 pt) preceding Warta Poznań (21 pt).

In this note we attempt to answer the question of the expected final of the league table in the 1939 season and who has the maximum chance for league championship. We take into account the expected position of the team in the league table based on the number of remissing games to the end of the season, the strengths of teams, the home-field advantage and random elements. The statistical data used in the note are from [1]. The mathematical method used in the note is the linear model of mathematical statistics (see [4]).

2. Poisson model of the league

Let us consider a model of the football league described in [3]. In this model it is assumed that the result in goals of the game depends upon tree components: the difference of strengths of teams, the home-field advantage, and random events. Strictly speaking, it is assumed that the outcome of

Mathematical Subject Classification: 62F10, 62P99.
Key words and phrases: football, model of league, estimation.

a game in goals between team i and team j with strenghts m_i and m_j is X_{ij}, Y_{ij} (on the first place is a home team) is equal to

$$X_{ij} = \Pi_1(ar_{ij} + bm_i + c), \quad Y_{ij} = \Pi_2(ar_{ji} + c), \tag{1}$$

where $\Pi_1(\lambda_1), \Pi_2(\lambda_2)$ are independent random variables with Poisson distribution, $\lambda_1 = ar_{ij}+bm_i+c$, $\lambda_2 = ar_{ji}+c$ are means of these random variables, $r_{ij} = \max(m_i - m_j, 0)$ is specific function of difference of the strengths, a, b, c are parameters of the model. The strengths $m = (m_1, \dots, m_n)$ characterize the participating teams (for our data set $n = 10$), the parameters a, b, c characterize the Polish football at this time.

In [3] there are analyzed a few examples of leagues, among others the Polish 1938 season. It is assumed that the strength of the team is equal to the number of finally scored points. Taking into account the observed outcomes of all games of the season, the parameters of the model were estimated as: $a = 0.09956$, $b = 0.07959$, $c = 1.16706$.

The parameters a, b, c, which characterize the football as a branch of sport, are rather stable from the season to season. The troublesome problem lies in the estimation of the strength. In this note assuming that a, b, c are given for the 1939 season, we estimate the strengths of the teams. As result, having a complete description of the model, we present the comparison of scores of the played games of the 1939 season before the war and the expected values obtained from the model. Next we estimate the expected final status of the league table. By simulation 10 000 times the missing part of the season we estimate the probabilities of league championship by the main teams of the league.

3. Strength of the team

Let $W = (w_{ij})$ denote the index of game: $w_{ij} = 1$, if the game is played; $w_{ij} = 0$, otherwise. Let $D^W = (D^W_{ij}) = ((X_{ij} - Y_{ij})w_{ij})$. Let us introduce the notations

$$w_{i\cdot} = \sum_{j=1}^{n} w_{ij}, \quad w_{\cdot j} = \sum_{i=1}^{n} w_{ij},$$

$$D_{i\cdot} = \sum_{j=1}^{n} D^W_{ij}, \quad D_{\cdot j} = \sum_{i=1}^{n} D^W_{ij},$$

$$\overline{m} = \frac{1}{n}\sum_{i=1}^{n} m_i, \quad \overline{D} = \frac{1}{n(n-1)}\sum_{i=1}^{n} D_{ij}.$$

In the vector description let

$$W_1 = (w_{1\cdot}, \dots, w_{n\cdot}), \quad W_2 = (w_{\cdot 1}, \dots, w_{\cdot n}),$$
$$D_1 = (D_{1\cdot}, \dots, D_{n\cdot}), \quad D_2 = (D_{\cdot 1}, \dots, D_{\cdot n}).$$

THEOREM 1. *Let a, b, c be given. The last squares estimator m on the basis D^W has the form*

$$\hat{m} = BA^{-1}, \tag{2}$$

where $A = (a+b)^2 I(W_1) + a^2 I(W_2) - a(a+b)(W + W^T)$, $B = (a+b)D_1 - aD_2$, and $I(W_k)$ is the diagonal matrix with the vector W_k on the main diagonal, $k = 1, 2$.

The estimator $\hat{m}$ is unbiased, with the covariance matrix

$$C_{\hat{m}} = A^{-1} C_B A^{-1},$$

where $C_B = (c_{ij})$,

$$c_{ii} = \sum_{j=1}^{n} \Big((a+b)^2 \mathrm{Var}(D_{ij}^W) + a^2 \mathrm{Var}(D_{ji}^W) \Big),$$

$$c_{ij} = -a(a+b)(\mathrm{Var}(D_{ij}^W) + \mathrm{Var}(D_{ji}^W)),$$

$\mathrm{Var}(D_{ij}^W) = w_{ij}(a|m_i - m_j| + bm_i + 2c)$, $1 \le i, j \le 1$.

COROLLARY 1. *If a, b, c are given and $w_{ij} = 1$ for $1 \le i, j \le n$, $i \ne j$, then*

$$\hat{m}_i = \Big(n((a+b)^{-2} + a^2) - b^2 \Big)^2 \Big((a+b)D_{i\cdot} - aD_{\cdot i} + 2a(a+b)n\overline{D}/b \Big), \quad 1 \le i \le n.$$

Note the point strengths considered in [3]: provide 2 points for the winning team and provide one point for each team in case of a tie; the condition $\overline{m} = 2(n-1)$ is then satisfied. This condition is not guaranteed for the estimator $\hat{m}$, hence we postulate it in the next theorem.

THEOREM 2. *In the class of the estimators satisfying the condition $\overline{m} = 2(n-1)$ the regression estimator has the form*

$$\hat{m} = (B + \lambda e)A^{-1}, \tag{3}$$

where $e = (1, \dots, 1)$, $\lambda = (eA^{-1}B^T - 2n(n-1))/(eA^{-1}e^T)$.

Proof of Theorem 1. From (1) we obtain the expected value $\mathrm{E}(D_{ij}) = a(m_i - m_j) + bm_i$. Let us consider the function

$$\begin{aligned} L(m|W, D^W) &= \sum_{(i,j)\varepsilon W} \Big(a(m_i - m_j) + bm_i - D_{ij} \Big)^2 \\ &= \sum_{j=2}^{n} \Big((a+b)m_1 - am_j - D_{1j} \Big)^2 \mathbf{1}_{(1,j)\varepsilon W} \\ &\quad + \sum_{i=2}^{n} \Big((a+b)m_i - am_1 - D_{i1} \Big)^2 \mathbf{1}_{(i,1)\varepsilon W} \\ &\quad + \sum_{2 \le i,j \le n} \Big((a+b)m_i - am_j - D_{ij} \Big)^2 \mathbf{1}_{(i,j)\varepsilon W}. \end{aligned}$$

By equating $\partial L/\partial m_1$ to zero we get the equation

$$(a+b)\sum_{j=2}^{n}\Big((a+b)m_1 - am_j - D_{1j}\Big)w_{1j}$$
$$-a\sum_{j=2}^{n}\Big((a+b)m_j - am_1 - D_{j1}\Big)w_{j1} = 0.$$

Hence we obtain the first equation of the following system of equations

$$((a+b)^2 w_{i\cdot} + a^2 w_{\cdot i})m_i - a(a+b)\sum_{j=1,\, j\neq i}^{n} m_j(w_{ij}+w_{ji}) = (a+b)D_{i\cdot} - aD_{\cdot i}.$$

We denote this system as $Am^T = B^T$. Because $A = A^T$, thus $m = BA^{-1}$. Its solution defines our estimator of the strength vector.

It is known that the regression estimators are unbiased. Now we consider the covariance matrix of $\hat{m}$:

$$\begin{aligned}\mathrm{E}((\hat{m}-m)^T(\hat{m}-m)) &= \mathrm{E}(((B-\mathrm{E}(B))A^{-1})^T(B-\mathrm{E}(B))A^{-1})\\ &= A^{-1}\mathrm{E}((B-\mathrm{E}(B))^T(B-\mathrm{E}(B)))A^{-1} = A^{-1}C_BA^{-1}.\end{aligned}$$

Since $D_{i\cdot}$, $D_{j\cdot}$, $i \neq j$, are mutually independent, $D_{\cdot i}$, $D_{\cdot j}$, $i \neq j$, are mutually independent, we have $\mathrm{Cov}(D_{i\cdot}, D_{j\cdot}) = \mathrm{Cov}(D_{\cdot i}, D_{\cdot j}) = \mathrm{Var}(D_{ij})$, $c_{ij} = \mathrm{Cov}((a+b)D_{i\cdot} - aD_{i\cdot}, (a+b)D_{\cdot j} - aD_{\cdot j})$. Hence

$$c_{ij} = -a(a+b)(w_{ij}\mathrm{Var}(D_{ij}) + w_{ji}\mathrm{Var}(D_{ji})),\quad i \neq j,$$
$$c_{ii} = \sum_{k=1}^{n}\Big((a+b)^2(w_{ik}\mathrm{Var}(D_{ik}) + a^2 w_{ki}\mathrm{Var}(D_{ki})\Big),\quad 1 \leq i,j \leq n.$$

Note that $w_{ii} = 0$, $\mathrm{Var}(D_{ij}) = \mathrm{Var}(X_{ij})+ \mathrm{Var}(Y_{ij}) = a|m_i + m_j| + bm_i + 2c$, $1 \leq i,j \leq n$. □

Proof of Corollary 1. If $w_{ij} = 1$, $1 \leq i,j \leq n$, $w_{ii} = 0$, then $w_{i\cdot} = w_{\cdot j} = (n-1)$,

$$\sum_{j=1}^{n} m_j w_{ij} = n\overline{m} - m_i,\quad \sum_{i=1}^{n} m_i w_{ij} = n\overline{m} - m_j,$$

The i-th equation of the system (2) has the form

$$((a+b)^2 + a^2)m_i - 2a(a+b)(n\overline{m} - m_i) = (a+b)D_{i\cdot} - aD_{\cdot i}.$$

By adding these equations side by side, we obtain $\overline{m} = \overline{D}/b$ because under the assumption of Corollary 1 is $\sum D_{\cdot i} = \sum D_{i\cdot}$. Hence

$$\hat{m}_i = ((n(a+b)^2 + a^2) - b^2)^{-2}((a+b)D_{\cdot i} - aD_{i\cdot} + 2a(a+b)n\overline{D}/b).\quad \square$$

Proof of Theorem 2. Using the Lagrange method search of the minimum L under the condition $n\overline{m} = em^T = 2n(n-1)$ we obtain the system of equations $Am^T = B^T - \lambda e^T$. The solution is $m^T = A^{-1}(B - \lambda e)^T$. The condition $em^T = eA^{-1}(B - \lambda e)^T$ enables to calculate λ. □

4. Numerical results

Table 1 shows the league scores of the 1939 season in the stop of games and its description from the model. The expected number of scored points and goals are a fair fit to the outcome results. The measure *chi-square* of accuracy is shown in the least row of Table 1. Note that it has a limited usage to statistical inference because the standard assumptions of the *chi-square* distribution are not satisfied here. The estimators of the strengths of the teams are shown in the least column of Table 1. The standard errors of the estimators, calculated from Theorem 1, are large (from 3.7 to 4.2).

Tables 2 and 3 show the expected results of the model for the remaining games of the season and the expected final table of the league. The presented final order of teams is probable, but it is different than the transient league Table in the stopping time. The differences between Warta, Wisła and Pogoń are small. Because the result of the game depends upon the random element, the other orders in the Table are also possible but probably differ.

The formula (1) enables the simulation of result of games and construction of the league scores. By drawing of 10 000 Tables it is possible to estimate the probability the league championship for every team exactly up to 1%. In the result Ruch has a 81% chance for championship of the league, Wisła – a 8% chance, and others teams in total a 11 % chance. The championship and vice-championship for the pair Ruch - Warta has a chance of 32%, for the pair Ruch - Wisła – a chance of 23%, for pair Ruch - Pogoń – a chance of 18%.

In addition we present the result of one numerical experiment. Recall that in [3], having the results of the Polish league of the 1938 season, using points from table of the league as strengths of teams, we have estimated a, b, c. Using the obtained parameters from Corollary 1 for this season one may calculate the strengths of the teams. It is proved that the new strengths do not differ from the initial point strengths. We omit the details.

Table 1. The Polish football league of the 1939 season, the description of played games, estimation of the strength of the teams.

	Games	Points scored		Goals scored		Goals lost		Straingth (2)
		pt	Ept	b_1	Eb_1	b_2	Eb_2	$\hat{m}$
Ruch Chorzów	14	18	20.4	50	50.0	23	23.3	25.1
Wisła Cracow	12	16	15.8	31	31.3	20	18.9	16.6
Pogoń Lwów	13	16	14.3	27	29.5	22	25.1	14.0
AKS Chorzów	12	15	15.4	30	29.8	14	19.0	14.9
Warta Poznań	12	15	16.4	34	36.1	20	20.2	19.4
Cracovia	13	14	10.3	23	20.7	32	31.0	4.6
Polonia Warsaw	12	12	12.5	28	24.6	28	23.9	12.3
Garbarnia Cracow	13	10	9.8	17	20.6	32	32.5	4.5
Warszawianka	11	5	6.4	16	14.0	29	29.3	0.6
Union-Touring Łódź	12	3	2.6	15	10.1	51	43.5	−8.1
		$\chi^2 = 2.34$		$\chi^2 = 4.31$		$\chi^2 = 3.81$		

Notations: pt – points scored, b_1 – goals scored, b_2 – goals lost, E – the expected values.

Table 2. The Polish football league of the 1939 season, the expected results of the remaining games, final table.

	Remaining games				Final table		
	Games	Ept	Eb_1	Eb_2	Ept	Eb_1	Eb_2
Ruch Chorzów	4	7.4	18.3	4.0	25.4	68.3	27.0
Warta Poznań	6	8.5	17.9	8.7	23.5	51.9	28.7
Wisła Cracow	6	7.4	16.3	11.3	23.4	47.3	31.3
Pogoń Lwów	5	7.0	12.8	7.0	23.0	39.8	29.0
AKS Chorzów	6	6.6	14.1	12.3	21.6	44.1	26.3
Polonia Warsaw	6	7.5	14.7	10.0	19.5	42.7	38.0
Cracovia	5	3.2	6.9	12.8	17.2	29.9	44.8
Garbarnia Cracow	5	3.5	6.9	11.4	13.5	23.9	43.4
Warszawianka	7	3.6	9.2	21.2	8.6	25.2	50.2
Union-Touring Łódź	6	1.2	5.1	23.3	4.2	20.1	74.3

Table 3. The Polish football league of the 1939 season, the description of the games realised, the estimation of strength of the teams, modified strengths.

	Games	Points scored		Goals scored		Goals lost		Straingth (3)
		pt	Ept	b_1	Eb_1	b_2	Eb_2	$\hat{m}$
Ruch Chorzów	14	18	19.2	50	54.5	23	28.1	33.1
Wisła Cracow	12	16	15.5	31	34.9	20	22.4	24.2
Pogoń Lwów	13	16	14.1	27	33.4	22	29.2	22.0
AKS Chorzów	12	15	15.1	30	33.5	14	22.6	22.5
Warta Poznań	12	15	15.9	34	39.5	20	23.4	26.4
Cracovia	13	14	10.7	23	24.7	32	35.1	12.5
Polonia Warsaw	12	12	12.5	28	28.4	28	27.6	20.0
Garbarnia Cracow	13	10	10.4	17	24.5	32	36.5	11.8
Warszawianka	11	5	7.0	16	17.5	29	32.5	8.0
Union-Touring Łódź	12	3	3.5	15	13.8	51	47.2	−0.5
		$\chi^2 = 2.34$		$\chi^2 = 5.82$		$\chi^2 = 8.24$		

Table 4. The Polish league of the 1939 season, expected results of the games remaining, final table, modified strengths.

	Remaining games				Final table		
	Games	Ept	Eb_1	Eb_2	Ept	Eb_1	Eb_2
Ruch Chorzów	4	7.48	20.4	4.6	25.48	70.4	27.6
Pogoń Lwów	5	7.11	14.9	8.1	23.11	41.9	30.1
Wisła Cracow	6	7.08	18.0	13.3	23.08	49.0	33.3
Warta Poznań	6	7.89	18.6	11.1	22.89	52.6	31.1
AKS Chòrzów	6	6.50	15.9	14.2	21.50	45.9	28.2
Polonia Warsaw	6	7.47	16.6	11.8	19.47	44.6	39.8
Cracovia	5	3.78	8.9	13.8	17.78	31.9	45.8
Garbarnia Cracow	5	3.48	8.1	13.5	13.48	25.1	45.5
Warszawianka	7	3.86	10.9	23.8	8.86	26.9	52.8
Union-Touring Łódź	6	1.35	6.9	25.0	4.35	21.9	76.0

References

[1] J. Jeleń, A. Konieczny, Z. Dobrowolny, G. Stański, Z. Łagódka, A. Wowra, K. Marcinek, *Liga gra po pięćdziesiątce,* Wyd. II, Wydawnictwo Sport i Turystyka, Warsaw 1987.

[2] M. E. Glickman and H. S. Stern, *A state-space model for National Football League scores*, J. Amer. Statist. Assoc. **93**, 441(1998), 25–35.

[3] B. Kopociński, *Components of the game result in a football league*, Appl. Math. **28**, 1 (2001), 1–18.

[4] C. R. Rao, *Linear Statistical Inference and its Applications*, PWN, Warsaw 1982.

MATHEMATICAL INSTITUTE
WROCŁAW UNIVERSITY
Pl. Grunwaldzki 2/4
50-384 WROCŁAW, POLAND
E-mail: ibk@math.uni.wroc.pl

Received June 14, 2000.