A Probability Model for Analyzing the Points Achieved by a Team in a Football Match. An Application to the Spanish Football League

December 10th, 2016

E.Gómez–Déniz, N. Dávila Cárdenes and J.M. Pérez–Sánchez

University of Las Palmas de G.C. and University de Granada

Contents

Contents

Contents

Contents

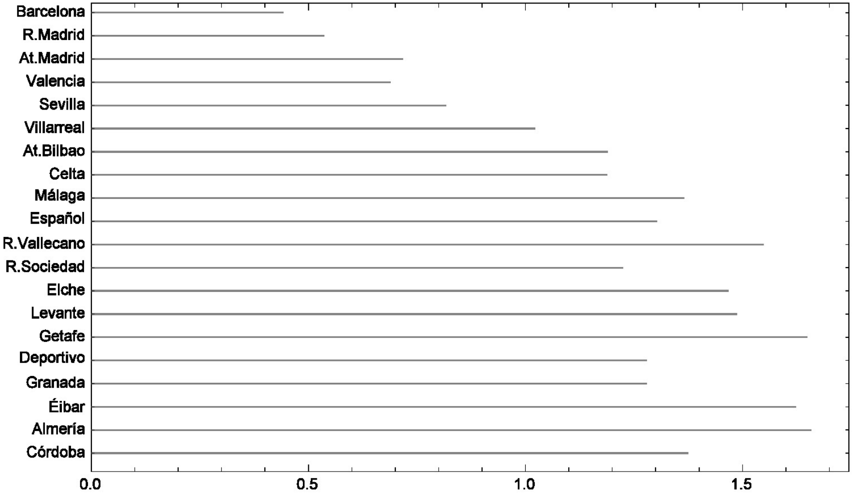
Contents

Contents

Motivation

- The aim of this work is to develop a suitable probability model for studying the points achieved by a team in a football match, say a random variable X.
- For this purpose, we use weighted and censored (truncated) distributions to build a discrete probability distribution with truncated support in the set $\{0, 1, 2, 3\}$ and with mass zero in $x = 2$. We test its performance using data from the Spanish Football League (First division) during the session 2013–14.
- Empirical analysis shows that the sequence of points for the teams with a lot of points, and therefore those fighting for the title of the competition are characterized to be under–dispersed (variance lower than the mean) while the teams with less points show over-dispersion (variance larger than the mean).
- Let X the random variable which give us the sequence of points achieved by a team in a competition, then the index of dispersion is defined as $\mathrm{ID} = var(X)/E(X)$.

Motivation


Barcelona
R.Madrid
At.Madrid
Valencia
Sevilla
Villarreal
At.Bilbao
Celta
Málaga
Español
R.Vallecano
R.Sociedad
Elche
Levante
Getafe
Deportivo
Granada
Éibar
Almería
Córdoba
0.0
0.5
1.0
1.5

Background

- In order to model the sequence of points achieved by a team in a football match, we need a probability function with bounded support and which takes the value zero when $x = 2$. To get this:
- Firstly, in order to do that the support of the distribution be 0, 1, 2 and 3 we will use right truncated or censored distributions, i.e. no observation to the right of 3 (see Moore, 1954; Patil, 1962 or Goswin, 2016, among others, for analyzing truncated or censored distribution).
- Secondly in order to get that the probability at 2 should be zero we will use weighted distributions (see for intance Fisher, 1934, Patil and Rao, 1978 and Harandi and Alamtsaz, 2013). Recall that for a distribution with probability function $f_\theta(x)$, X with support in $\mathcal{X}$, depending on a vector of parameteres $\theta \in \Theta$, we can construct a new distribution with probability function

$$g_\theta(x) = \frac{w(x)}{E_{f_\theta(x)}(x)} f_\theta(x), \tag{1}$$

where it is assumed that $E_{f_\theta(x)}(w(x)) < \infty$, and w is a weighted function depending on x.

The proposed model and some properties

Let us to start with the classical Poisson distribution which probability function given by,

$$f_\theta(x) = \frac{\theta^x \exp(-\theta)}{x!}, \quad x = 0, 1, \ldots, \quad \theta > 0. \tag{2}$$

It is straightforward to see that

$$\sum_{i=4}^{\infty} f_\theta(x) = \kappa_1(\theta) = \frac{1}{6}\left[6 + \theta(6 + \theta(3 + \theta))\right] \exp(-\theta).$$

Thus

$$\frac{f_\theta(x)}{\kappa_1(\theta)}, \quad x = 0, 1, 2, 3$$

is a genuine probability function with support in $\{0, 1, 2, 3\}$.

The proposed model and some properties

Now, it is also simple to demonstrate that

$$\frac{1}{\kappa_1(\theta)} \frac{(x-2)^2 f_\theta(x)}{\sum_{x=0}^{3}(x-2)^2 f_\theta(x)} = \frac{24+\theta(6+\theta^2)}{6+\theta(6+\theta(3+\theta))}.$$

Thus, expression

$$g_\theta(x) = \kappa(\theta)(x-2)^2 \frac{\theta^x}{x!}, \quad x = 0, 1, 2, 3. \tag{3}$$

where

$$\kappa(\theta) = \frac{6}{24+\theta(6+\theta^2)}. \tag{4}$$

defines a genuine probability function with support in $\mathcal{X} = \{0, 1, 2, 3\}$ and with mass probability zero at $x = 2$.

The proposed model and some properties

Since the probability function (3) can be written as

$$g_\lambda(x) = q(x)\exp\left[-x\lambda - \log(\vartheta(\lambda))\right],$$

where $q(x) = (x-2)^2/x!$, $\lambda = -\log\theta$ and $\vartheta(\theta) = \exp(\lambda)\kappa(\exp(-\lambda))$, and where $-\infty < \lambda < \infty$, it can be seen that the probability function in (3) is a member of the natural exponential family of distributions. Furthermore, the probability function in (3) can also be rewritten as

$$p_x = \frac{q(x)\theta^x}{\vartheta(\theta)},$$

it is also a power series distribution (see Johnson and Kotz, 2005, p.75). Thus, we have another distribution, together with the Bernoulli, binomial, geometric, negative binomial, Poisson and logarithmic series, within this interesting class of distributions.

The proposed model and some properties

Bardwell (1960) – see also Amidi (1976)– discussed discrete probability functions $f_\theta(x)$ which fit the relation

$$\frac{df_\theta(x)}{d\theta} = B(\theta)\,[x - D(\theta)]\,f_\theta(x). \tag{5}$$

It is shown that in this case the mean is $\mu = D(\theta)$ and the variance is $\mu_2 = (d\mu/d\theta)(1/B(\theta))$. It is also shown that in this case

$$\mu_i = \mu_2 \left[\frac{d\mu_{i-1}}{d\theta} \frac{1}{d\mu/d\theta} + (i-1)\mu_{i-2} \right], \tag{6}$$

where μ_i is the ith moment about the mean.
Expression (5) is verified for the probability function (3). In this case we have

$$B(\theta) = \frac{1}{\theta}, \qquad D(\theta) = -\frac{\theta\, \kappa'(\theta)}{\kappa(\theta)}.$$

The proposed model and some properties

Additionally, see Noack (1950) and Johnson et al. (2005), the following recurrence relations between the moments about the origin and the central moments (about the mean), respectively, are satisfied,

$$\begin{aligned} \mu'_{r+1} &= \theta \frac{d\mu'_r}{d\theta} + \mu'_1 \mu'_r, \\ \mu_{r+1} &= \theta \left(\frac{d\mu_r}{d\theta} + r \frac{d\mu'_1}{d\theta} \mu_{r-1} \right). \end{aligned}$$

These moments can also be obtained from the probability generating function, which is given by

$$G_X(z) = \frac{z\theta(6 + z^2\theta^2) + 24}{24 + \theta(\theta^2 + 6)}, \quad |z| \leq 1. \tag{7}$$

The proposed model and some properties

From (7) we can get the mean and second order moment about the origin, among others, which are given by

$$
\begin{aligned}
E(X) &= \frac{3\theta(2+\theta^2)}{24+\theta(6+\theta)}, &(8)\\
E(X^2) &= \frac{3\theta(2+3\theta^2)}{24+\theta(6+\theta^2)}. &(9)
\end{aligned}
$$

Using (8) and (9) we get the variance, given by

$$
var(X) = \frac{24\theta\left[6+\theta^2(9+\theta)\right]}{\left[24+\theta(6+\theta^2)\right]^2}. \qquad (10)
$$

Furthermore, relation connecting the cumulants $k_{[r]}$ and the moments μ_r about the origin can be obtained using expression (8) in Noack (1950). Relations between factorial–cumulants and cumulants can also be given using results in Khatri (1959). See also Johnson and Kotz (2005), p.77. Since the probability function (3) is a member of the family discussed by Bardwell (1960) we have the following result.

The proposed model and some properties

Theorem

If X follows the probability function (3), then it is verified that:

1. *Mean deviation* $= E|X-\mu| = -2\lambda(1-\lambda^2)\frac{\partial}{\partial\theta}G_\theta([\mu])$.
2. $f_\theta(x) = \eta \exp\left\{\int \frac{x-\mu}{\mu_2}\frac{d\mu}{d\theta}\, d\theta\right\}$.

Here, $G_\theta([\mu]) = \sum_{x=0}^{[\mu]} f_\theta(x)$, *where* $[\cdot]$ *represents the integer part and* η *is a constant of normalization.*

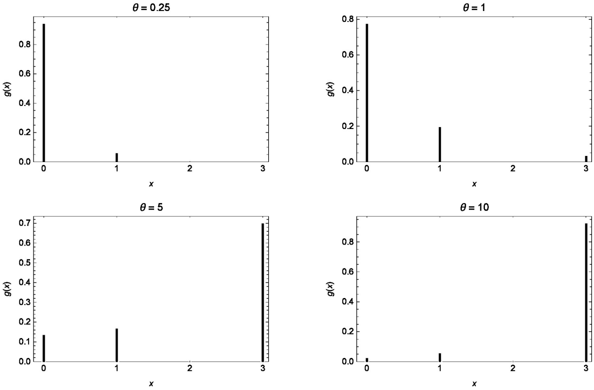
Demostración.

The results follow directly, using Theorems 2 and 3 in Bardwell (1960). □

The proposed model and some properties

According to Bardwell (1960), a characteristic of the family of functions satisfying (5) is that each has a unique maximum. That is, the modal value is achieved at $x_{\text{mode}} = [\mu] - 1$.

Figure shows the suitability of the new distribution proposed to model the number of points.



The proposed model and some properties

The cumulative distribution function can be written in terms of the exponential integral function given by

$$E_n(z) = \int_1^{\infty} \frac{\exp(-zt)}{t^n}\, dt,$$

and results

$$G(x) = \Pr(X \leq x) = \frac{6\theta^{x+1}\left[3 - x - \theta + \exp(\theta)(4 - \theta(\theta - 3))E_{\theta}(-x)\right]}{x!\left[\theta(\theta^2 + 6) + 24\right]}.$$

Some computations provide that

$$var(X) - E(X) = \frac{3\theta^2(48\theta - \theta^4 + 12)}{\left[24 + \theta(6 + \theta^2)\right]^2},$$

from which it is simple to verify that the probability function accommodate for overdispersion (variance larger than the mean) when $0{,}250081 < \theta < 3{,}54676$ and for infradispersion when $0 < \theta < 0{,}250081$ and when $\theta > 3{,}54676$.

Estimation and simulation

Two methods of estimation of the parameter of the distribution are studied here. Using (8) it is simple to see that the estimator of θ is the real solution of the equation

$$3\theta^3 - \bar{x}\theta^2 - 6\bar{x}\theta - 24\bar{x} = 0,$$

where $\bar{x}$ is the sample mean.

Finally, the MLE are somewhat easy to derive since we are in the exponential family. Let now $\underline{x} = (x_1, x_2, \ldots, x_n)$ be a random sample obtained from model (3), then the log–likelihood function is proportional to

$$\ell(\theta) \propto -n \log\left[24 + \theta(6 + \theta^2)\right] - n\bar{x}\log\theta$$

The likelihood equation obtained from

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{(6 + 3\theta^2)n}{24 + \theta(6 + \theta^2)} + \frac{n\bar{x}}{\theta},$$

provides the maximum likelihood estimator of θ, which is solution of the equation

$$\Psi(\theta) = (3 - \bar{x})\theta^3 + 6(1 - \bar{x})\theta - 24\bar{x} = 0. \tag{11}$$

Estimation and simulation

Since $\Psi'(\theta) = 3\theta^2(3-\bar{x}) + 6(1-\bar{x}) > 0$ we conclude that the solution is unique. Now, the following result is easily established.

Proposition:

The unique maximum likelihood estimator $\widehat{\theta}$ of θ is consistent and asymptotically normal and therefore

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, I^{-1}(\theta)),$$

where $I(\theta)$ is the Fisher's information about θ.

By using Corollary 3.11 in Lehmann and Casella (1998), we conclude that the maximum likelihood estimator of θ is asymptotically efficient.

Numerical Application

In the first case, without covariates, the maximum likelihood method was used to estimate the parameters of the distribution in (3), from the corresponding likelihood equation given in (11).

Home observed and expected points:

Points	Observed Counts	Expected Counts
0	116	110.57
1	85	93.15
3	179	176.28
Total	380	380

$\widehat{\theta} = 3{,}3697\,(0{,}131)$
$\chi^2_1 = 1{,}021$
p-value $= 31{,}22\,\%$
AIC $= 802{,}401$

Away observed and expected points:

Points	Observed Counts	Expected Counts
0	179	166.14
1	85	104.29
3	116	109.57
Total	380	380

$\widehat{\theta} = 2{,}5108\,(0{,}103)$
$\chi^2_1 = 4{,}939$
p-value $= 2{,}62\,\%$
AIC $= 806{,}512$

In order to analyze the factors involved in the total points achieved by a team in a competition, we consider four groups of variables. Those related to the statistics of the game, one group of variables directly associated to the match, non-sport variables and those related to the referee.

Variable	Description
Game statistics	
hthg	Half time home team goals
htag	Half time away team goals
hs	Home team shots.
as	Away team shots.
hst	Home team shots on target
ast	Away team shots on target
hf	Number of home team fouls.
af	Number of away team fouls.
hc	Home team corners
ac	Away team corners
hy	Home team yellow cards.
ay	Away team yellow cards.
hr	Home team red cards.
ar	Away team red cards.

Variable	Description
Match variables	
Derby	Match played between teams from the same city or region or between the strongest teams in the league.
Category	Depending on the historical participation data of the teams in the con
Extra games	
budh	Logarithm of home team budget
buda	Logarithm of away team budget
stah	Logarithm of the capacity of the home stadium.
staa	Logarithm of the capacity of the away stadium.
Referee	
agereferee	Logarithm of the age of the referee.
International	Scored as 0 if the referee has no international experience and 1 if he does.
acient	Logarithm of the age of experience in the first division

Now, we investigate the effect of covariates to account for the points achieved by the teams playing in home and later away. For the sake of convenience, we rewrite (3) in another form, so that covariates may be introduced into the model. By equating (8) to μ we get, after some computations, that

$$\theta = \varphi(\mu) = \frac{\sqrt[3]{2}\left[\sqrt[3]{2}\,(3-\mu(\mu-4)) - \psi(\mu)^{2/3}\right]}{(\mu-3)\psi(\mu)^{1/3}}, \tag{12}$$

where

$$\psi(\mu) = 6\mu(\mu-3)^2 + \sqrt{2(\mu-3)^3(\mu(3+19\mu(\mu-3)-1)}.$$

A common specification for the mean parameter μ is in terms of exponential functions, ensuring the non–negativity of μ. That is,

$$\mu_i = \frac{3\exp\left(\beta^\top \mathbf{x}\right)}{1+\exp\left(\beta^\top \mathbf{x}\right)}, \tag{13}$$

where $\mathbf{x}$ is the vector of covariates and $\beta = (\beta_1, \ldots, \beta_q)^\top$ is an unknown vector of regression coefficients. Expression (13) ensures that the mean is a positive–valued function with support in $[0, 3]$.

The log–likelihood of the model with covariates is proportional to,

$$\ell(\beta_1,\ldots,\beta_q) \propto \sum_{i=1}^{n} \left[\kappa(\varphi(\mu_i)) + x_i \log\varphi(\mu_i)\right]$$

The normal equations are

$$\frac{\partial\ell}{\partial\beta_j} = \sum_{i=1}^{n} \frac{\partial}{\partial\beta_j}\kappa(\varphi(\mu_{ij})) + \sum_{i=1}^{n} \frac{x_{ij}}{\varphi(\mu_{ij})}\frac{\partial}{\partial\beta_j}\varphi(\mu_{ij}) = 0, \qquad (14)$$

for $j = 1, 2, \ldots, q$.

The second partial derivatives are given by

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta_j^2} &= \sum_{i=1}^{n} \frac{\partial^2}{\partial \beta_j^2} \kappa(\varphi(\mu_{ij})) + \sum_{i=1}^{n} \frac{x_{ij}}{\varphi(\mu_{ij})} \left[\frac{\partial^2}{\partial \beta_j^2} \varphi(\mu_{ij}) - \left(\frac{1}{\varphi(\mu_{ij})} \frac{\partial}{\partial \beta_j} \varphi(\mu_{ij}) \right)^2 \right] \\
\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_l} &= \sum_{i=1}^{n} \frac{\partial^2}{\partial \beta_j \beta_l} \kappa(\varphi(\mu_{ij})) + \sum_{i=1}^{n} \frac{x_{ij}}{\varphi(\mu_{ij})} \left[\frac{\partial^2}{\partial \beta_j \beta_l} \varphi(\mu_{ij}) - \left(\frac{1}{\varphi(\mu_{ij})} \frac{\partial}{\partial \beta_j} \varphi(\mu_{ij}) \right) \right.
\end{aligned}
$$

for $j = 1, 2, \ldots, q$, $l = 1, 2, \ldots, q$ and $j \neq l$.

These normal equations cannot be solved explicitly, but must be addressed either by a numerical method or by directly maximising the log–likelihood function. In this study, the FindMaximum function of Mathematica software package v.11.0 was used, although the same results can be obtained by other methods, too, such as Newton, PrincipalAxis or QuasiNewton (all of which are available in this package), or by other packages such as R, Matlab or WinRats. Finally, the standard errors of the parameter estimates were obtained by inverting the Hessian matrix.

Results for home matches:

Variable	Estimate	Standard Error	$\lvert t\rvert$-statistic	$\Pr > \lvert t\rvert$
hthg	1.5058	0.1328	11.3403	0.0000
htag	−1.8142	0.1794	10.1146	0.0000
hs	−0.0265	0.0322	0.8236	0.4107
as	0.1175	0.0361	3.2584	0.0012
hst	0.3849	0.0584	6.5989	0.0000
ast	−0.4438	0.0746	5.9462	0.0000
hf	−0.0172	0.0256	0.6701	0.5032
af	0.0600	0.0247	2.4310	0.0155
hc	−0.0019	0.0410	0.0461	0.9633
ac	−0.0272	0.0458	0.5945	0.5526
hy	0.0279	0.0646	0.4312	0.6665
ay	−0.0439	0.0217	2.0190	0.0442
hr	−0.9340	0.1873	4.9868	0.0000
ar	−0.0412	0.0698	0.5894	0.5559
derby	0.2440	0.0534	4.5673	0.0000
category	0.0182	0.0600	0.3029	0.7621
budh	−0.0286	0.0400	0.7137	0.4759
buda	−0.3020	0.0544	5.5514	0.0000
stah	0.1431	0.0145	9.8958	0.0000
staa	0.3078	0.0665	4.6264	0.0000
international	−0.2811	0.0439	6.3969	0.0000
acient	−0.1106	0.0494	2.2369	0.0259
agereferee	−0.1296	0.0529	2.4477	0.0148
intercept	−0.0072	0.0079	0.9147	0.3609
AIC = 549,220				

Significant factors for Home matches:

- Positive:
  - At $\alpha = 1\,\%$: Half time home teams goals (hthg), away team shots (as), home team shots on target (hst), derby, capacity of the stadium (stah), capacity of the away stadium (staa).
  - At $\alpha = 5\,\%$: Number of away team fouls (af).
- Negative:
  - At $\alpha = 1\,\%$: Half time away team goals (htag), away team shots on target (ast), home team red cards (hr), away team budget (buda), international experience of the referee,
  - At $\alpha = 5\,\%$: Away team yellow cards (ay), experience in the first division of the referee (acient), age of the referee (agereferee).

Results for away matches:

Variable	Estimate	Standard Error	$\lvert t \rvert$-statistic	Pr $> \lvert t \rvert$
hthg	−1.4554	0.1412	10.3087	0.0000
htag	1.7338	0.1684	10.2958	0.0000
hs	0.0471	0.0331	1.4242	0.1552
as	−0.1107	0.0362	3.0546	0.0024
hst	−0.3791	0.0599	6.3324	0.0000
ast	0.4848	0.0699	6.9332	0.0000
hf	0.0045	0.0262	0.1715	0.8639
af	−0.0449	0.0239	1.8835	0.0604
hc	−0.0245	0.0407	0.6022	0.54745
ac	0.0513	0.0449	1.1441	0.2533
hy	−0.0251	0.0530	0.4744	0.6355
ay	0.0457	0.0149	3.0633	0.0023
hr	1.1392	0.2029	5.6120	0.0000
ar	0.0664	0.0142	4.6619	0.0000
derby	−0.2159	0.0892	2.4198	0.0160
category	-0.0392	0.0614	0.6388	0.5233
budh	−0.0081	0.0431	0.1869	0.8519
buda	0.3259	0.0649	5.0175	0.0000
stah	−0.1538	0.0173	8.8973	0.0000
staa	−0.4523	0.1184	3.8185	0.0001
international	0.1065	0.0603	1.7674	0.0780
acient	0.0691	0.0528	1.3089	0.1914
agereferee	0.0916	0.0466	1.9637	0.0503
intercept	0.0165	0.0078	2.1234	0.0344
AIC $=$ 552,529				

Significant factors for Away matches:

- Positive:
  - At $\alpha = 1\,\%$: Half time away team goals (htag), away team shots on target (ast), away team yellow cards (ay), home team red cards (hr), away red cards (ar), away team budget (buda).
  - At $\alpha = 5\,\%$: age of the referee (agereferee).
  - At $\alpha = 10\,\%$: international experience of the referee.
- Negative:
  - At $\alpha = 1\,\%$: Half time home teams goals (hthg), away team shots (as), home team shots on target (hst) , capacity of the stadium (stah), capacity of the away stadium (staa).
  - At $\alpha = 5\,\%$: Derby.
  - At $\alpha = 10\,\%$: Number of away team fouls (af).

Conclusions

- We have developed a new probability model for studying the points achieved by a team in a football match.
- We have used weighted and truncated distributions to build a discrete probability distribution with truncated support in the set $\{0, 1, 2, 3\}$ with mass zero in $x = 2$.
- We have tested the probability model by using data from the Spanish Football League (First Division) during season 2013–14.
- The new models with covariates improve to the models without covariates in terms of Akaike information criterium.