Bayesian Hierarchical Modelling for the Prediction of Football Results

Gianluca Baio[1,2] Marta Blangiardo[3]

[1]Department of Epidemiology and Public Health,
University College London (UK)

[2]Department of Statistics,
University of Milano Bicocca (Italy)

[3]Department of Epidemiology and Public Health,
Imperial College London (UK)

Seminari del Dipartimento di Statistica
Firenze, 12 Marzo 2009

Outline of presentation

1. **Introduction**
2. Statistical framework
3. Bayesian modelling
4. Some results
5. Other possible applications
6. Conclusions

Outline of presentation

1. Introduction
2. **Statistical framework**
3. Bayesian modelling
4. Some results
5. Other possible applications
6. Conclusions

Outline of presentation

1. Introduction
2. Statistical framework
3. **Bayesian modelling**
4. Some results
5. Other possible applications
6. Conclusions

Outline of presentation

1. Introduction
2. Statistical framework
3. Bayesian modelling
4. **Some results**
5. Other possible applications
6. Conclusions

Outline of presentation

1. Introduction
2. Statistical framework
3. Bayesian modelling
4. Some results
5. **Other possible applications**
6. Conclusions

Outline of presentation

1. Introduction
2. Statistical framework
3. Bayesian modelling
4. Some results
5. Other possible applications
6. **Conclusions**

Background

- (Surprisingly) many examples of statistical models for football results exist in the literature
- Although the Binomial or Negative Binomial have been proposed in the late 1970s (Pollard et al 1977), the Poisson distribution has been widely accepted as a suitable model for the number of goals scored. The "classical model" typically assumes independence between the goals scored by the home and the away team (Maher 1982)
- However, some authors have shown empirical, although relatively low, levels of correlation between the two quantities and therefore have used correction factor to the independent Poisson model to improve the performance in terms of prediction (for instance Dixon & Coles 1997)
- Others like Lee (1997) and Karlis & Ntzoufras (2003) (K&N) argue that this assumption is unrealistic, as "*it is reasonable to assume that the two outcome variables are correlated since the two teams interact during the game*"

Background

- (Surprisingly) many examples of statistical models for football results exist in the literature
- Although the Binomial or Negative Binomial have been proposed in the late 1970s (Pollard et al 1977), the Poisson distribution has been widely accepted as a suitable model for the number of goals scored. The "classical model" typically assumes independence between the goals scored by the home and the away team (Maher 1982)

- However, some authors have shown empirical, although relatively low, levels of correlation between the two quantities and therefore have used correction factor to the independent Poisson model to improve the performance in terms of prediction (for instance Dixon & Coles 1997)
- Others like Lee (1997) and Karlis & Ntzoufras (2003) (K&N) argue that this assumption is unrealistic, as "*it is reasonable to assume that the two outcome variables are correlated since the two teams interact during the game*"

Bivariate Poisson (BP) model

- To account for correlation, K&N define a model with:

$$Z_1 \sim \text{Poisson}(\theta_1) \quad Z_2 \sim \text{Poisson}(\theta_2) \quad Z_3 \sim \text{Poisson}(\theta_3)$$

independently, whereby for $X = Z_1 + Z_3$ and $Y = Z_2 + Z_3$ we have

$$(X, Y) \sim \text{BP}(\theta_1, \theta_2, \theta_3)$$

- By properties of the BP:
  1. $\text{E}[X] = \theta_1 + \theta_3$
  2. $\text{E}[Y] = \theta_2 + \theta_3$
  3. $Cov(X, Y) = \theta_3$
  4. The distribution of $D = X - Y$, which is all is needed to determine the outcome (but not the exact result!) of the game, does not depend on the correlation parameter θ_3

Bivariate Poisson (BP) model

- To account for correlation, K&N define a model with:

$$Z_1 \sim \text{Poisson}(\theta_1) \quad Z_2 \sim \text{Poisson}(\theta_2) \quad Z_3 \sim \text{Poisson}(\theta_3)$$

independently, whereby for $X = Z_1 + Z_3$ and $Y = Z_2 + Z_3$ we have

$$(X, Y) \sim \text{BP}(\theta_1, \theta_2, \theta_3)$$

- By properties of the BP:
  1. $\text{E}[X] = \theta_1 + \theta_3$
  2. $\text{E}[Y] = \theta_2 + \theta_3$
  3. $Cov(X, Y) = \theta_3$
  4. The distribution of $D = X - Y$, which is all is needed to determine the outcome (but not the exact result!) of the game, does not depend on the correlation parameter θ_3

Estimation

- K&N consider the reasonable case of a BP regression model

$$(X_i, Y_i) \sim \text{BP}(\theta_{1i}, \theta_{2i}, \theta_{3i})$$
$$\log \theta_{ki} = \mathbf{w}_{ki}\boldsymbol{\beta}_k, \qquad k = 1, 2, 3$$

where $i = 1, \ldots, n$ denotes the observations, $\mathbf{w}_{ki}$ is a vector of explanatory variables and $\boldsymbol{\beta}_k$ is a vector of regression coefficients

- Also, they argue that parameter estimation is not straightforward in this case, because of the implied correlation θ_3, therefore:
  - EM-like algorithms are required for frequentist estimation
  - RJMCMC-like algorithms are required for Bayesian estimation

- NB: K&N's main interest lies in the estimation of the effects used to explain the number of goals scored (ie. the vector $\boldsymbol{\beta}_k$); prediction is only a byproduct of the model

Estimation

- K&N consider the reasonable case of a BP regression model

$$
\begin{aligned}
(X_i, Y_i) &\sim \mathsf{BP}(\theta_{1i}, \theta_{2i}, \theta_{3i}) \\
\log \theta_{ki} &= \mathbf{w}_{ki}\boldsymbol{\beta}_k, \qquad k = 1, 2, 3
\end{aligned}
$$

  where $i = 1, \ldots, n$ denotes the observations, $\mathbf{w}_{ki}$ is a vector of explanatory variables and $\boldsymbol{\beta}_k$ is a vector of regression coefficients

- Also, they argue that parameter estimation is not straightforward in this case, because of the implied correlation θ_3, therefore:
  - EM-like algorithms are required for frequentist estimation
  - RJMCMC-like algorithms are required for Bayesian estimation

- NB: K&N's main interest lies in the estimation of the effects used to explain the number of goals scored (ie. the vector $\boldsymbol{\beta}_k$); prediction is only a byproduct of the model

Estimation

- K&N consider the reasonable case of a BP regression model

$$(X_i, Y_i) \sim \mathsf{BP}(\theta_{1i}, \theta_{2i}, \theta_{3i})$$
$$\log \theta_{ki} = \mathbf{w}_{ki}\boldsymbol{\beta}_k, \qquad k = 1, 2, 3$$

  where $i = 1, \ldots, n$ denotes the observations, $\mathbf{w}_{ki}$ is a vector of explanatory variables and $\boldsymbol{\beta}_k$ is a vector of regression coefficients

- Also, they argue that parameter estimation is not straightforward in this case, because of the implied correlation θ_3, therefore:
  - EM-like algorithms are required for frequentist estimation
  - RJMCMC-like algorithms are required for Bayesian estimation

- NB: K&N's main interest lies in the estimation of the effects used to explain the number of goals scored (ie. the vector $\boldsymbol{\beta}_k$); prediction is only a byproduct of the model

Bayesian hierarchical modelling

- Although K&N provide (somewhere else) a Bayesian estimation using the BP model, their framework is intrinsically frequentist

- Alternative: the use of a full **Bayesian hierarchical model** can account for the correlation between the observed pair of counts
  - The structure associated with this model allows for more information to be included, avoiding the need for more complicated estimation algorithms (standard Gibbs sampling is sufficient)
  - Assuming two *conditionally independent* Poisson variables for the number of goals scored and a hierarchical structure, correlation is taken into account since the observable variables are mixed at an upper level

- Moreover, as we are framed in a Bayesian context, the main purpose is the prediction of a new game under the model, naturally handled using the (posterior) predictive distribution

Bayesian hierarchical modelling

- Although K&N provide (somewhere else) a Bayesian estimation using the BP model, their framework is intrinsically frequentist
- Alternative: the use of a full **Bayesian hierarchical model** can account for the correlation between the observed pair of counts
  - The structure associated with this model allows for more information to be included, avoiding the need for more complicated estimation algorithms (standard Gibbs sampling is sufficient)
  - Assuming two *conditionally independent* Poisson variables for the number of goals scored and a hierarchical structure, correlation is taken into account since the observable variables are mixed at an upper level
- Moreover, as we are framed in a Bayesian context, the main purpose is the prediction of a new game under the model, naturally handled using the (posterior) predictive distribution

Bayesian hierarchical modelling

- Although K&N provide (somewhere else) a Bayesian estimation using the BP model, their framework is intrinsically frequentist

- Alternative: the use of a full **Bayesian hierarchical model** can account for the correlation between the observed pair of counts
  - The structure associated with this model allows for more information to be included, avoiding the need for more complicated estimation algorithms (standard Gibbs sampling is sufficient)
  - Assuming two *conditionally independent* Poisson variables for the number of goals scored and a hierarchical structure, correlation is taken into account since the observable variables are mixed at an upper level

- Moreover, as we are framed in a Bayesian context, the main purpose is the prediction of a new game under the model, naturally handled using the (posterior) predictive distribution

The model (1)

- We consider a league made by T teams playing each other twice in a season (one at home and one away)
- During the campaign a total of $g = 1, \ldots, G = T \times (T - 1)$ games are played

- We indicate the number of goals scored by the home and by the away team in the $g-$th game of the season as y_{g1} and y_{g2} respectively, with:

$$y_{gj} \mid \theta_{gj} \sim \text{Poisson}(\theta_{gj}),$$

  where the parameters $\boldsymbol{\theta} = (\theta_{g1}, \theta_{g2})$ represent the *scoring intensity* for the g-th game and for the team playing at home ($j = 1$) or away ($j = 2$), respectively

The model (1)

- We consider a league made by T teams playing each other twice in a season (one at home and one away)
- During the campaign a total of $g = 1, \ldots, G = T \times (T-1)$ games are played

- We indicate the number of goals scored by the home and by the away team in the $g-$th game of the season as y_{g1} and y_{g2} respectively, with:

$$y_{gj} \mid \theta_{gj} \sim \text{Poisson}(\theta_{gj}),$$

  where the parameters $\boldsymbol{\theta} = (\theta_{g1}, \theta_{g2})$ represent the *scoring intensity* for the g-th game and for the team playing at home $(j = 1)$ or away $(j = 2)$, respectively

The model (2)

We model $\boldsymbol{\theta}$ according to a widely used formulation (see Karlis & Ntzoufras 2003 and the references therein), assuming a log-linear random effect model:

$$\begin{aligned} \log \theta_{g1} &= \mathit{home} + \mathit{att}_{h(g)} + \mathit{def}_{a(g)} \\ \log \theta_{g2} &= \mathit{att}_{a(g)} + \mathit{def}_{h(g)} \end{aligned}$$

where

- The parameter *home* represents the advantage for the team hosting the game and we assume that this effect is constant for all the teams and throughout the season
- The scoring intensity is determined jointly by the attack and defense ability of the two teams involved, represented by the parameters *att* and *def*, respectively
- The nested indexes $h(g), a(g) = 1, \ldots, T$ identify the team that is playing at home (away) in the g-th game of the season

The model (2)

We model $\boldsymbol{\theta}$ according to a widely used formulation (see Karlis & Ntzoufras 2003 and the references therein), assuming a log-linear random effect model:

$$\begin{aligned}\log \theta_{g1} &= \mathit{home} + \mathit{att}_{h(g)} + \mathit{def}_{a(g)} \\ \log \theta_{g2} &= \mathit{att}_{a(g)} + \mathit{def}_{h(g)}\end{aligned}$$

where

- The parameter *home* represents the advantage for the team hosting the game and we assume that this effect is constant for all the teams and throughout the season
- The scoring intensity is determined jointly by the attack and defense ability of the two teams involved, represented by the parameters *att* and *def*, respectively
- The nested indexes $h(g), a(g) = 1, \ldots, T$ identify the team that is playing at home (away) in the g-th game of the season

The model (2)

We model $\boldsymbol{\theta}$ according to a widely used formulation (see Karlis & Ntzoufras 2003 and the references therein), assuming a log-linear random effect model:

$$\begin{aligned}\log\theta_{g1} &= \mathit{home} + \mathit{att}_{h(g)} + \mathit{def}_{a(g)}\\ \log\theta_{g2} &= \mathit{att}_{a(g)} + \mathit{def}_{h(g)}\end{aligned}$$

where

- The parameter *home* represents the advantage for the team hosting the game and we assume that this effect is constant for all the teams and throughout the season
- The scoring intensity is determined jointly by the attack and defense ability of the two teams involved, represented by the parameters *att* and *def*, respectively
- The nested indexes $h(g), a(g) = 1, \ldots, T$ identify the team that is playing at home (away) in the g-th game of the season

The model (2)

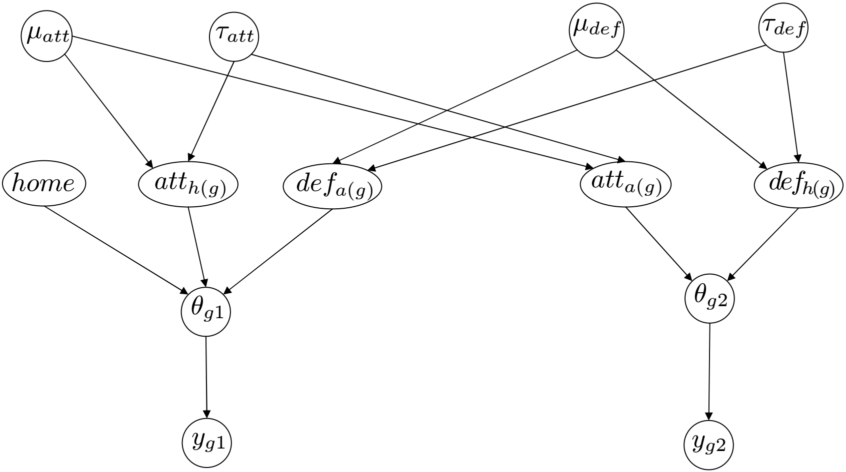
We model $\boldsymbol{\theta}$ according to a widely used formulation (see Karlis & Ntzoufras 2003 and the references therein), assuming a log-linear random effect model:

$$
\begin{aligned}
\log \theta_{g1} &= \mathit{home} + \mathit{att}_{h(g)} + \mathit{def}_{a(g)} \\
\log \theta_{g2} &= \mathit{att}_{a(g)} + \mathit{def}_{h(g)}
\end{aligned}
$$

where

- The parameter *home* represents the advantage for the team hosting the game and we assume that this effect is constant for all the teams and throughout the season
- The scoring intensity is determined jointly by the attack and defense ability of the two teams involved, represented by the parameters *att* and *def*, respectively
- The nested indexes $h(g), a(g) = 1, \ldots, T$ identify the team that is playing at home (away) in the g-th game of the season

Graphical representation of the model



Underlying assumptions

- The result of a game is specific to the teams involved
  - Only the scoring intensities directly influence the observed number of goals scored in each single game

- The scoring intensities (and therefore the observed goals) depend on the teams involved through the *att* and *def* parameters, which are considered to be *exchangeable*
  - We assume a common random process from which each is drawn
  - It is possible to include expert knowledge to inform their distribution

- The $\frac{G}{2}$ observations for all the home (away) games inform each other
  - This happens through the exchangeable model for *att* and *def*, which for all the teams depend on common (random) hyper-parameters (μ_{att}, τ_{att}) and (μ_{def}, τ_{def})
  - This produces a level of correlation among all the home (away) scoring intensities

Underlying assumptions

- The result of a game is specific to the teams involved
  - Only the scoring intensities directly influence the observed number of goals scored in each single game

- The scoring intensities (and therefore the observed goals) depend on the teams involved through the *att* and *def* parameters, which are considered to be *exchangeable*
  - We assume a common random process from which each is drawn
  - It is possible to include expert knowledge to inform their distribution

- The $\frac{G}{2}$ observations for all the home (away) games inform each other
  - This happens through the exchangeable model for *att* and *def*, which for all the teams depend on common (random) hyper-parameters (μ_{att}, τ_{att}) and (μ_{def}, τ_{def})
  - This produces a level of correlation among all the home (away) scoring intensities

Underlying assumptions

- The result of a game is specific to the teams involved
  - Only the scoring intensities directly influence the observed number of goals scored in each single game
- The scoring intensities (and therefore the observed goals) depend on the teams involved through the *att* and *def* parameters, which are considered to be *exchangeable*
  - We assume a common random process from which each is drawn
  - It is possible to include expert knowledge to inform their distribution
- The $\frac{G}{2}$ observations for all the home (away) games inform each other
  - This happens through the exchangeable model for *att* and *def*, which for all the teams depend on common (random) hyper-parameters (μ_{att}, τ_{att}) and (μ_{def}, τ_{def})
  - This produces a level of correlation among all the home (away) scoring intensities

The data for the Italian Serie A 1991-92

g	*home team*	*away team*	$h(g)$	$a(g)$	y_{g1}	y_{g2}
1	Verona	Roma	18	15	0	1
2	Napoli	Atalanta	13	2	1	0
3	Lazio	Parma	11	14	1	1
4	Cagliari	Sampdoria	4	16	3	2
...	...	...	...	...	...	...
303	Sampdoria	Cremonese	16	5	2	2
304	Roma	Bari	15	3	2	0
305	Inter	Atalanta	9	2	0	0
306	Torino	Ascoli	17	1	5	2

The data for the Italian Serie A 1991-92

g	*home team*	*away team*	$h(g)$	$a(g)$	y_{g1}	y_{g2}
1	Verona	Roma	18	15	0	1
2	Napoli	Atalanta	13	2	1	0
3	Lazio	Parma	11	14	1	1
4	Cagliari	Sampdoria	4	16	3	2
...	...	...	...	...	...	...
303	Sampdoria	Cremonese	16	5	2	2
304	Roma	Bari	15	3	2	0
305	Inter	Atalanta	9	2	0	0
306	Torino	Ascoli	17	1	5	2

For instance, Sampdoria are always associated with the index 16, whether they play away, as for $a(4)$, or at home, as for $h(303)$

Prior specification of the model

- The variable *home* is modelled as a fixed effect, assuming a standard minimally informative prior distribution:

$$home \sim \text{Normal}(0, 0.0001)$$

- Conversely, for each $t = 1, \ldots, T$, the team-specific effects are modelled as exchangeable from a common distribution:

$$att_t \mid \mu_{att}, \tau_{att} \sim \text{Normal}(\mu_{att}, \tau_{att}) \qquad def_t \mid \mu_{def}, \tau_{def} \sim \text{Normal}(\mu_{def}, \tau_{def})$$

with $\sum_{t=1}^{T} att_t = 0,$ and $\sum_{t=1}^{T} def_t = 0$ to ensure identifiability

- Finally, the hyper-priors of the attack and defense effects are modelled independently using again a flat prior distribution:

$$\mu_{att} \sim \text{Normal}(0, 0.0001), \qquad \mu_{def} \sim \text{Normal}(0, 0.0001),$$
$$\tau_{att} \sim \text{Gamma}(0.1, 0.1), \qquad \tau_{def} \sim \text{Gamma}(0.1, 0.1)$$

Prior specification of the model

- The variable *home* is modelled as a fixed effect, assuming a standard minimally informative prior distribution:

$$\textit{home} \sim \text{Normal}(0, 0.0001)$$

- Conversely, for each $t = 1, \ldots, T$, the team-specific effects are modelled as exchangeable from a common distribution:

$$\textit{att}_t \mid \mu_{\textit{att}}, \tau_{\textit{att}} \sim \text{Normal}(\mu_{\textit{att}}, \tau_{\textit{att}}) \quad \textit{def}_t \mid \mu_{\textit{def}}, \tau_{\textit{def}} \sim \text{Normal}(\mu_{\textit{def}}, \tau_{\textit{def}})$$

with $\sum_{t=1}^{T} att_t = 0,$ and $\sum_{t=1}^{T} def_t = 0$ to ensure identifiability

- Finally, the hyper-priors of the attack and defense effects are modelled independently using again a flat prior distribution:

$$\mu_{\textit{att}} \sim \text{Normal}(0, 0.0001), \qquad \mu_{\textit{def}} \sim \text{Normal}(0, 0.0001),$$

$$\tau_{\textit{att}} \sim \text{Gamma}(0.1, 0.1), \qquad \tau_{\textit{def}} \sim \text{Gamma}(0.1, 0.1)$$

Prior specification of the model

- The variable *home* is modelled as a fixed effect, assuming a standard minimally informative prior distribution:

$$\mathit{home} \sim \mathrm{Normal}(0, 0.0001)$$

- Conversely, for each $t = 1, \ldots, T$, the team-specific effects are modelled as exchangeable from a common distribution:

$$\mathit{att}_t \mid \mu_{\mathit{att}}, \tau_{\mathit{att}} \sim \mathrm{Normal}(\mu_{\mathit{att}}, \tau_{\mathit{att}}) \qquad \mathit{def}_t \mid \mu_{\mathit{def}}, \tau_{\mathit{def}} \sim \mathrm{Normal}(\mu_{\mathit{def}}, \tau_{\mathit{def}})$$

with $\sum_{t=1}^{T} att_t = 0,$ and $\sum_{t=1}^{T} def_t = 0$ to ensure identifiability

- Finally, the hyper-priors of the attack and defense effects are modelled independently using again a flat prior distribution:

$$\mu_{\mathit{att}} \sim \mathrm{Normal}(0, 0.0001), \qquad \mu_{\mathit{def}} \sim \mathrm{Normal}(0, 0.0001),$$
$$\tau_{\mathit{att}} \sim \mathrm{Gamma}(0.1, 0.1), \qquad \tau_{\mathit{def}} \sim \mathrm{Gamma}(0.1, 0.1)$$

Features of Bayesian model

- Hierarchical model allows "borrowing strength" (the different "experiments" will be related and the estimation will be more precise)
- Bayesian model allows to compute the posterior predictive distribution, i.e. to simulate future occurrences from the model, given the posterior distributions of all relevant parameters
- Estimation is performed using standard Gibbs sampling
  - Easy to implement
  - Relatively fast to run
- The hierarchical structure will typically produce a *shrinkage*, so that extreme results will be more difficult to predict

Features of Bayesian model

- Hierarchical model allows "borrowing strength" (the different "experiments" will be related and the estimation will be more precise)

- Bayesian model allows to compute the posterior predictive distribution, i.e. to simulate future occurrences from the model, given the posterior distributions of all relevant parameters

- Estimation is performed using standard Gibbs sampling
  - Easy to implement
  - Relatively fast to run

- The hierarchical structure will typically produce a *shrinkage*, so that extreme results will be more difficult to predict

Features of Bayesian model

- Hierarchical model allows "borrowing strength" (the different "experiments" will be related and the estimation will be more precise)
- Bayesian model allows to compute the posterior predictive distribution, i.e. to simulate future occurrences from the model, given the posterior distributions of all relevant parameters
- Estimation is performed using standard Gibbs sampling
  - Easy to implement
  - Relatively fast to run
- The hierarchical structure will typically produce a *shrinkage*, so that extreme results will be more difficult to predict

Features of Bayesian model

- Hierarchical model allows "borrowing strength" (the different "experiments" will be related and the estimation will be more precise)
- Bayesian model allows to compute the posterior predictive distribution, i.e. to simulate future occurrences from the model, given the posterior distributions of all relevant parameters
- Estimation is performed using standard Gibbs sampling
  - Easy to implement
  - Relatively fast to run
- The hierarchical structure will typically produce a *shrinkage*, so that extreme results will be more difficult to predict

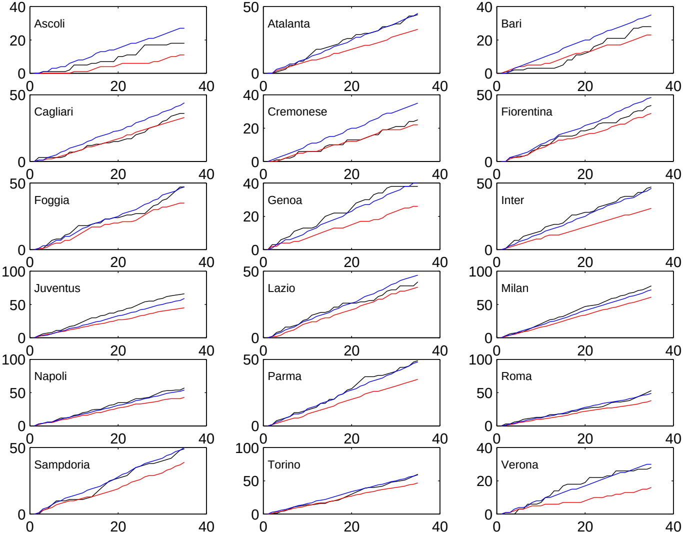
Results (1) – estimation of the effects

teams		*attack effect*				*defense effect*		
	mean	*2.5%*	*median*	*97.5%*	*mean*	*2.5%*	*median*	*97.5%*
Ascoli	-0.2238	-0.5232	-0.2165	0.0595	0.4776	0.2344	0.4804	0.6987
Atalanta	-0.1288	-0.4050	-0.1232	0.1321	-0.0849	-0.3392	-0.0841	0.1743
Bari	-0.2199	-0.5098	-0.2213	0.0646	0.1719	-0.0823	0.1741	0.4168
Cagliari	-0.1468	-0.4246	-0.1453	0.1255	-0.0656	-0.3716	-0.0645	0.2109
Cremonese	-0.1974	-0.4915	-0.1983	0.0678	0.1915	-0.0758	0.1894	0.4557
Fiorentina	0.1173	-0.1397	0.1255	0.3451	0.0672	-0.1957	0.0656	0.3372
Foggia	0.3464	0.1077	0.3453	0.5811	0.3701	0.1207	0.3686	0.6186
Genoa	-0.0435	-0.3108	-0.0464	0.2149	0.1700	-0.0811	0.1685	0.4382
Inter	-0.2077	-0.4963	-0.2046	0.0980	-0.2061	-0.5041	-0.2049	0.0576
Juventus	0.1214	-0.1210	0.1205	0.3745	-0.3348	-0.6477	-0.3319	-0.0514
Lazio	0.0855	-0.1626	0.0826	0.3354	0.0722	-0.1991	0.0742	0.3145
Milan	0.5226	0.2765	0.5206	0.7466	-0.3349	-0.6788	-0.3300	-0.0280
Napoli	0.2982	0.0662	0.2956	0.5267	0.0668	-0.2125	0.0667	0.3283
Parma	-0.1208	-0.3975	-0.1200	0.1338	-0.2038	-0.5136	-0.2031	0.0859
Roma	-0.0224	-0.2999	-0.0182	0.2345	-0.1358	-0.4385	-0.1300	0.1253
Sampdoria	-0.0096	-0.2716	-0.0076	0.2436	-0.1333	-0.4484	-0.1317	0.1346
Torino	0.0824	-0.1821	0.0837	0.3408	-0.4141	-0.7886	-0.4043	-0.1181
Verona	-0.2532	-0.5601	-0.2459	0.0206	0.3259	0.1026	0.3254	0.5621
	mean	*2.5%*	*median*	*97.5%*				
home	0.2124	0.1056	0.2128	0.3213				

Results (1) – estimation of the effects

teams	*attack effect*				*defense effect*			
	mean	*2.5%*	*median*	*97.5%*	*mean*	*2.5%*	*median*	*97.5%*
Ascoli	-0.2238	-0.5232	-0.2165	0.0595	0.4776	0.2344	0.4804	0.6987
Atalanta	-0.1288	-0.4050	-0.1232	0.1321	-0.0849	-0.3392	-0.0841	0.1743
Bari	-0.2199	-0.5098	-0.2213	0.0646	0.1719	-0.0823	0.1741	0.4168
Cagliari	-0.1468	-0.4246	-0.1453	0.1255	-0.0656	-0.3716	-0.0645	0.2109
Cremonese	-0.1974	-0.4915	-0.1983	0.0678	0.1915	-0.0758	0.1894	0.4557
Fiorentina	0.1173	-0.1397	0.1255	0.3451	0.0672	-0.1957	0.0656	0.3372
Foggia	0.3464	0.1077	0.3453	0.5811	0.3701	0.1207	0.3686	0.6186
Genoa	-0.0435	-0.3108	-0.0464	0.2149	0.1700	-0.0811	0.1685	0.4382
Inter	-0.2077	-0.4963	-0.2046	0.0980	-0.2061	-0.5041	-0.2049	0.0576
Juventus	0.1214	-0.1210	0.1205	0.3745	-0.3348	-0.6477	-0.3319	-0.0514
Lazio	0.0855	-0.1626	0.0826	0.3354	0.0722	-0.1991	0.0742	0.3145
Milan	0.5226	0.2765	0.5206	0.7466	-0.3349	-0.6788	-0.3300	-0.0280
Napoli	0.2982	0.0662	0.2956	0.5267	0.0668	-0.2125	0.0667	0.3283
Parma	-0.1208	-0.3975	-0.1200	0.1338	-0.2038	-0.5136	-0.2031	0.0859
Roma	-0.0224	-0.2999	-0.0182	0.2345	-0.1358	-0.4385	-0.1300	0.1253
Sampdoria	-0.0096	-0.2716	-0.0076	0.2436	-0.1333	-0.4484	-0.1317	0.1346
Torino	0.0824	-0.1821	0.0837	0.3408	-0.4141	-0.7886	-0.4043	-0.1181
Verona	-0.2532	-0.5601	-0.2459	0.0206	0.3259	0.1026	0.3254	0.5621
	mean	*2.5%*	*median*	*97.5%*				
home	0.2124	0.1056	0.2128	0.3213				

Results (2) – prediction

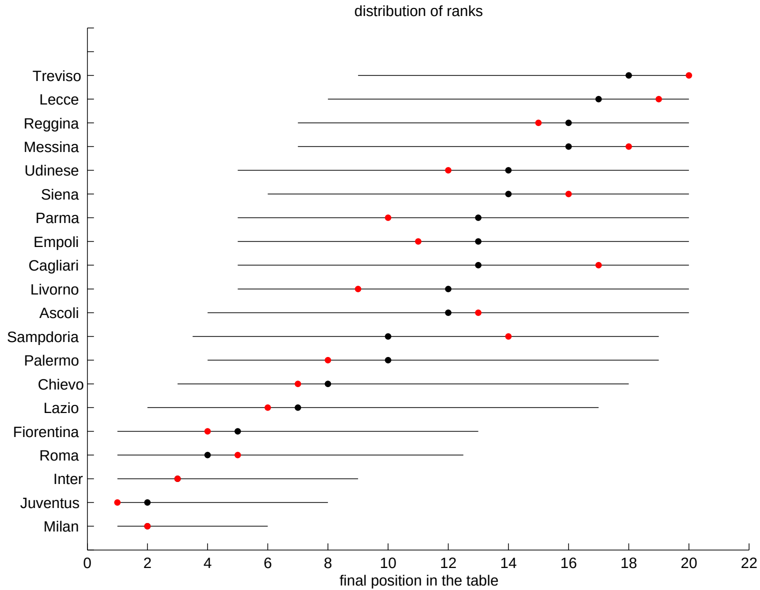


— Observed cumulative points through the season
— Cumulative points predicted from the Bayesian hierarchical model for each week
— Cumulative points predicted from the Bivariate Poisson model for each week

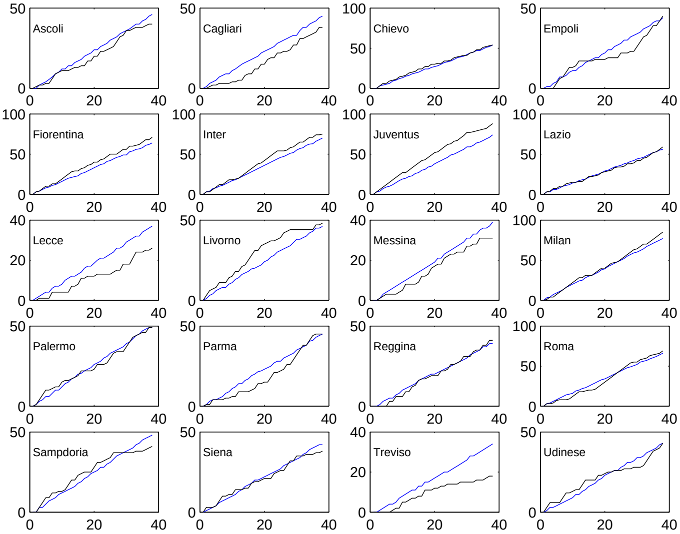
Observed *vs.* simulated table for the season 2005-2006

team		*Observed results*				*Simulated results (medians)*			
		points	*scored*	*conceded*	*goal diff*	*points*	*scored*	*conceded*	*goal diff*
1.	Juventus	91	71	24	47	76	67	31	36
2.	Milan	88	85	31	54	79	79	35	44
3.	Inter	76	68	30	38	72	64	35	29
4.	Fiorentina	74	66	41	25	66	63	43	20
5.	Roma	69	70	42	28	67	66	43	23
6.	Lazio	62	57	47	10	58	55	47	8
7.	Chievo	54	54	49	5	55	52	48	4
8.	Palermo	52	50	52	-2	51	49	51	-2
9.	Livorno	49	37	44	-7	48	39	45	-6
10.	Parma	45	46	60	-14	46	46	57	-11
11.	Empoli	45	46	62	-16	45	46	58	-12
12.	Udinese	43	40	54	-10	47	43	51	-8
13.	Ascoli	43	43	53	-14	45	41	52	-11
14.	Sampdoria	41	47	51	-4	50	47	50	-3
15.	Reggina	41	39	65	-26	40	41	61	-20
16.	Siena	39	42	60	-13	46	43	53	-10
17.	Cagliari	39	42	55	-18	44	43	57	-14
18.	Messina	31	34	58	-24	39	37	56	-19
19.	Lecce	29	30	57	-27	38	33	55	-22
20.	Treviso	21	24	56	-32	35	29	54	-25

Posterior distribution of ranks

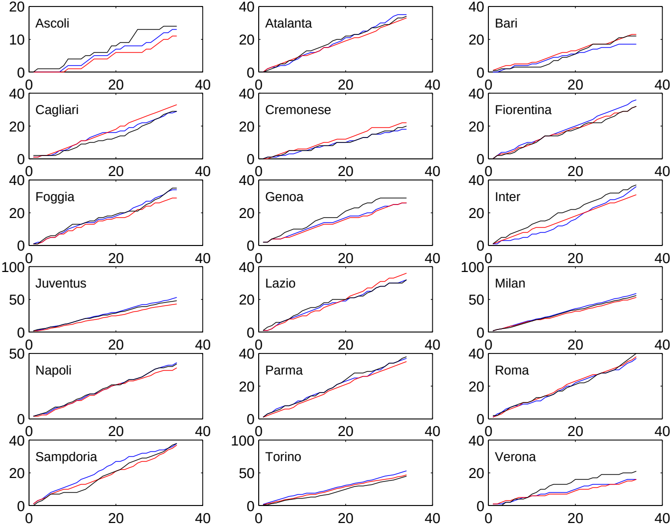


Posterior predictive validation of the model



— Observed cumulative points through the season
— Cumulative points predicted from the model for each week

Comparison with the BP model – Serie A 1991-1992



— Observed cumulative points through the season
— Cumulative points predicted from the Bayesian hierarchical model for each week
— Cumulative points predicted from the Bivariate Poisson model for each week

More serious (??) applications

- K&N (2006) discuss the use of the BP model to analyse paired count data in medicine
  - Before and after treatment measurements, with specific interest in the difference (ie. treatment effect)
- Karlis & Meligkotsidou (2005) use multivariate Poisson regression models with covariance structure in social science analysis
  - Data on the counts of (related) types of crimes using a Bayesian framework
- Tunaru (2002) works with a Bayesian multivariate Poisson-logNormal model to analyse accident data
  - The data include accidents between 1984 and 1991 in 150+ carriageways in Kent

More serious (??) applications

- K&N (2006) discuss the use of the BP model to analyse paired count data in medicine
  - Before and after treatment measurements, with specific interest in the difference (ie. treatment effect)

- Karlis & Meligkotsidou (2005) use multivariate Poisson regression models with covariance structure in social science analysis
  - Data on the counts of (related) types of crimes using a Bayesian framework

- Tunaru (2002) works with a Bayesian multivariate Poisson-logNormal model to analyse accident data
  - The data include accidents between 1984 and 1991 in 150+ carriageways in Kent

More serious (??) applications

- K&N (2006) discuss the use of the BP model to analyse paired count data in medicine
  - Before and after treatment measurements, with specific interest in the difference (ie. treatment effect)

- Karlis & Meligkotsidou (2005) use multivariate Poisson regression models with covariance structure in social science analysis
  - Data on the counts of (related) types of crimes using a Bayesian framework

- Tunaru (2002) works with a Bayesian multivariate Poisson-logNormal model to analyse accident data
  - The data include accidents between 1984 and 1991 in 150+ carriageways in Kent

Conclusions

- The Bayesian hierarchical model provides a (relatively) simple and effective method of predicting correlated counts, overcoming some of the limitations of multivariate Poisson models
  - *Computational* issues: BP needs specific estimation algorithms
  - *Theoretical* problems: BP can only deal with positive levels of correlations for the observed counts (Aitchinson & Ho 1989)

- The application of this model to football results prediction produces reasonable findings, even if the fit can be improved

- It would be interesting to include "expert" information in the definition of the prior distributions, and perhaps to analyse a time series (ie. many seasons)
  - Problems with different teams playing in the same league from one year to the next (relegation and promotions)

Conclusions

- The Bayesian hierarchical model provides a (relatively) simple and effective method of predicting correlated counts, overcoming some of the limitations of multivariate Poisson models
  - *Computational* issues: BP needs specific estimation algorithms
  - *Theoretical* problems: BP can only deal with positive levels of correlations for the observed counts (Aitchinson & Ho 1989)

- The application of this model to football results prediction produces reasonable findings, even if the fit can be improved

- It would be interesting to include "expert" information in the definition of the prior distributions, and perhaps to analyse a time series (ie. many seasons)
  - Problems with different teams playing in the same league from one year to the next (relegation and promotions)

Conclusions

- The Bayesian hierarchical model provides a (relatively) simple and effective method of predicting correlated counts, overcoming some of the limitations of multivariate Poisson models
  - *Computational* issues: BP needs specific estimation algorithms
  - *Theoretical* problems: BP can only deal with positive levels of correlations for the observed counts (Aitchinson & Ho 1989)

- The application of this model to football results prediction produces reasonable findings, even if the fit can be improved

- It would be interesting to include "expert" information in the definition of the prior distributions, and perhaps to analyse a time series (ie. many seasons)
  - Problems with different teams playing in the same league from one year to the next (relegation and promotions)

Some references

Pollard, R., Benjamin, P. and Reep, C. (1977)
Sport and the negative binomial distribution,
in S. Ladany and R. Machol, eds, 'Optimal Strategies in Sports', North Holland, New York, NY, pp. 188–195

Maher, M. (1982)
Modelling association football scores
Statistica Neerlandica **36**, 109–118

Dixon, M. and Coles, S. (1997)
Modelling association football scores and inefficiencies in the football betting market
Journal of the Royal Statistical Society C **46**, 265–280

Lee, A. (1997)
Modeling scores in the Premier League: is Manchester United really the best?
Chance **10**, 15–19

Karlis D. and Ntzoufras I. (2003)
Analysis of sports data by using bivariate Poisson models
Journal of the Royal Statistical Society D **52**, *381–393*

Karlis D. and Ntzoufras I. (2006)
Bayesian analysis of the differences of count data
Statistics in Medicine **25**, *1885–1905*

Karlis D. and Meligkotsidou L. (2005)
Multivariate Poisson regression with covariance structure
Statistics and Computing **15**, *255–265*

Tunaru, R. (2002)
Hierarchical Bayesian Models for Multiple Count Data
Austrian Journal of Statistics **31**, *221–229*

Aitchinson, J. and Ho, C. (1989)
The multivariate Poisson-log normal distribution
Biometrika **76**, *643–653*

Thank You!