

Modelli statistici per la previsione di risultati calcistici

Implementazione in del modello Dixon-Coles ed applicazioni al campionato di Serie A 2021-2022

Relatore:
Egidi Leonardo

Laureando:
Fantuzzi Giulio

Università degli studi di Trieste

5 settembre 2023

Una distribuzione di probabilità per il numero di goal

- In una partita, ogni volta che una squadra è in possesso di palla ha l'opportunità di costruire un'azione in attacco e di segnare un goal.
- Indichiamo con p la probabilità che un'azione si traduca in goal e con n il numero di azioni costruite in una partita.
- Assumendo che p sia costante e che le azioni siano indipendenti tra loro, il numero (aleatorio) di goal segnati nella partita (X) segue:

$$X \sim \text{Bin}(n, p) .$$

- L'approssimazione della distribuzione binomiale con la distribuzione di Poisson, in questo contesto, è particolarmente adatta:
 - ① il numero di prove (azioni offensive) è molto grande ;
 - ② la probabilità di successo (segnare un goal) è molto piccola.

$$X \sim \text{Pois}(\lambda) .$$

Una distribuzione di probabilità per il numero di goal

- In una partita, ogni volta che una squadra è in possesso di palla ha l'opportunità di costruire un'azione in attacco e di segnare un goal.
- Indichiamo con p la probabilità che un'azione si traduca in goal e con n il numero di azioni costruite in una partita.
- Assumendo che p sia costante e che le azioni siano indipendenti tra loro, il numero (aleatorio) di goal segnati nella partita (X) segue:

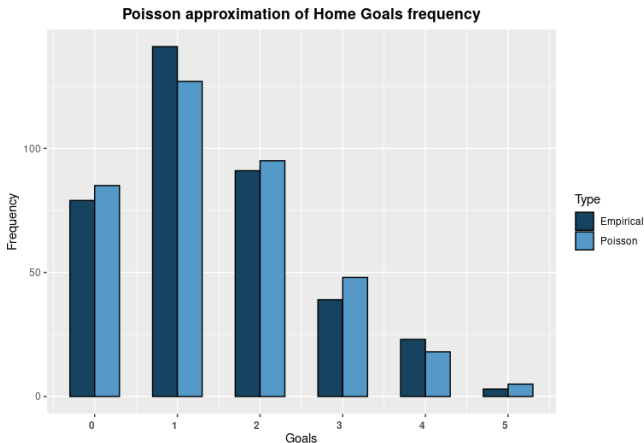
$$X \sim \text{Bin}(n, p) \ .$$

- L'approssimazione della distribuzione binomiale con la distribuzione di Poisson, in questo contesto, è particolarmente adatta:
 - ① il numero di prove (azioni offensive) è molto grande ;
 - ② la probabilità di successo (segnare un goal) è molto piccola.

$$X \sim \text{Pois}(\lambda) \ .$$

Una distribuzione di probabilità per il numero di goal

- Inizialmente l'idea di approssimare il numero di goal tramite una distribuzione di Poisson è stata più volte messa in discussione.
- Nonostante alcuni limiti della distribuzione di Poisson, l'adattamento ai dati empirici si è rivelato, in realtà, più che soddisfacente.



Il modello di Maher: formulazione del modello

- Consideriamo la partita tra le squadre i e j .
- I punteggi delle due squadre vengono descritti mediante due distribuzioni di Poisson tra loro indipendenti:

$$X_{ij} \sim \text{Pois}(\lambda_{ij}) \quad , \quad Y_{ij} \sim \text{Pois}(\mu_{ij}) \quad .$$

- Il risultato della partita $(x; y)$ è dunque una determinazione della coppia aleatoria $(X_{ij}; Y_{ij})$. Per l'ipotesi di indipendenza riesce:

$$Pr(X_{ij} = x; Y_{ij} = y) = Pr(X_{ij} = x) \cdot Pr(Y_{ij} = y) = \frac{(\lambda_{ij})^x}{x!} e^{-\lambda_{ij}} \cdot \frac{(\mu_{ij})^y}{y!} e^{-\mu_{ij}} .$$

Il modello di Maher: formulazione del modello

- Consideriamo la partita tra le squadre i e j .
- I punteggi delle due squadre vengono descritti mediante due distribuzioni di Poisson tra loro indipendenti:

$$X_{ij} \sim \text{Pois}(\lambda_{ij}) \quad , \quad Y_{ij} \sim \text{Pois}(\mu_{ij}) \quad .$$

- Il risultato della partita $(x; y)$ è dunque una determinazione della coppia aleatoria $(X_{ij}; Y_{ij})$. Per l'ipotesi di indipendenza riesce:

$$Pr(X_{ij} = x; Y_{ij} = y) = Pr(X_{ij} = x) \cdot Pr(Y_{ij} = y) = \frac{(\lambda_{ij})^x}{x!} e^{-\lambda_{ij}} \cdot \frac{(\mu_{ij})^y}{y!} e^{-\mu_{ij}} .$$

Il modello di Maher: formulazione del modello

- Consideriamo la partita tra le squadre i e j .
- I punteggi delle due squadre vengono descritti mediante due distribuzioni di Poisson tra loro indipendenti:

$$X_{ij} \sim \text{Pois}(\lambda_{ij}) \quad , \quad Y_{ij} \sim \text{Pois}(\mu_{ij}) \quad .$$

- Il risultato della partita $(x; y)$ è dunque una determinazione della coppia aleatoria $(X_{ij}; Y_{ij})$. Per l'ipotesi di indipendenza riesce:

$$Pr(X_{ij} = x; Y_{ij} = y) = Pr(X_{ij} = x) \cdot Pr(Y_{ij} = y) = \frac{(\lambda_{ij})^x}{x!} e^{-\lambda_{ij}} \cdot \frac{(\mu_{ij})^y}{y!} e^{-\mu_{ij}} .$$

- La media dei goal segnati da una squadra non è una misura caratteristica della singola squadra, ma è influenzata dalle caratteristiche di entrambe le squadre coinvolte nella partita.
- λ_{ij} e μ_{ij} vengono espressi in funzione di alcuni coefficienti di attacco e di difesa che riassumono le abilità delle squadre. In particolare:

$$\lambda_{ij} = \alpha_i \cdot \beta_j \quad ; \quad \mu_{ij} = \alpha_j \cdot \beta_i \quad ,$$

in cui:

- α : coefficienti di forza d'attacco;
- β : coefficienti di debolezza difensiva.

- La media dei goal segnati da una squadra non è una misura caratteristica della singola squadra, ma è influenzata dalle caratteristiche di entrambe le squadre coinvolte nella partita.
- λ_{ij} e μ_{ij} vengono espressi in funzione di alcuni coefficienti di attacco e di difesa che riassumono le abilità delle squadre. In particolare:

$$\lambda_{ij} = \alpha_i \cdot \beta_j \quad ; \quad \mu_{ij} = \alpha_j \cdot \beta_i \quad ,$$

in cui:

- α : coefficienti di forza d'attacco;
- β : coefficienti di debolezza difensiva.

- Dopo aver osservato un sufficiente numero di partite (N) è possibile ricavare i coefficienti di attacco (α) e di difesa (β) di tutte le squadre (n), attraverso metodi di stima di massima verosimiglianza.

$$L(\alpha_i, \beta_i, i = 1, \dots, n) = \prod_{k=1}^N \frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \cdot \frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} ;$$

$$l(\alpha_i, \beta_i, i = 1, \dots, n) = \sum_{k=1}^N \log \left(\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \right) + \log \left(\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right) ;$$

con $\lambda_k = \alpha_{i(k)}\beta_{j(k)}$ e $\mu_k = \alpha_{j(k)}\beta_{i(k)}$, in cui $i(k)$ e $j(k)$ denotano gli indici della squadra di casa e trasferta nella k-esima partita.

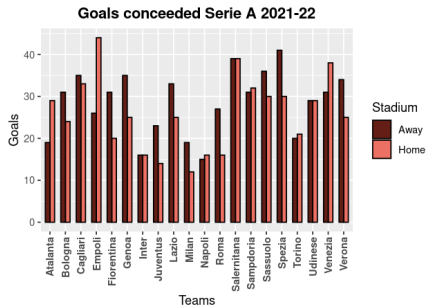
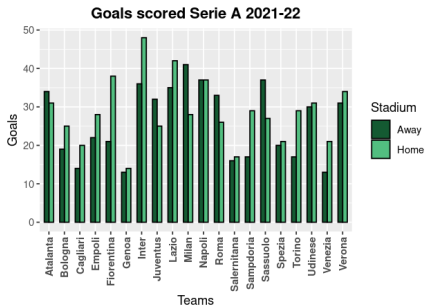
- ❶ Il modello tende a sottostimare i risultati di pareggio tra le squadre ed in generale i risultati con pochi goal (0-0, 1-0, 0-1, 1-1).
- ❷ L'ipotesi di indipendenza tra i goal segnati dalle due squadre non appare così ragionevole: spesso le squadre aumentano la propria intensità in attacco se sono in svantaggio, mentre si richiudono in fase difensiva per proteggere il proprio vantaggio.
- ❸ Il modello considera indifferentemente la squadra che gioca in casa con quella che gioca in trasferta: assenza di *home effect*.

- ❶ Il modello tende a sottostimare i risultati di pareggio tra le squadre ed in generale i risultati con pochi goal (0-0, 1-0, 0-1, 1-1).
- ❷ L'ipotesi di indipendenza tra i goal segnati dalle due squadre non appare così ragionevole: spesso le squadre aumentano la propria intensità in attacco se sono in svantaggio, mentre si richiudono in fase difensiva per proteggere il proprio vantaggio.
- ❸ Il modello considera indifferentemente la squadra che gioca in casa con quella che gioca in trasferta: assenza di *home effect*.

- ① Il modello tende a sottostimare i risultati di pareggio tra le squadre ed in generale i risultati con pochi goal (0-0, 1-0, 0-1, 1-1).
- ② L'ipotesi di indipendenza tra i goal segnati dalle due squadre non appare così ragionevole: spesso le squadre aumentano la propria intensità in attacco se sono in svantaggio, mentre si richiudono in fase difensiva per proteggere il proprio vantaggio.
- ③ Il modello considera indifferentemente la squadra che gioca in casa con quella che gioca in trasferta: assenza di *home effect*.

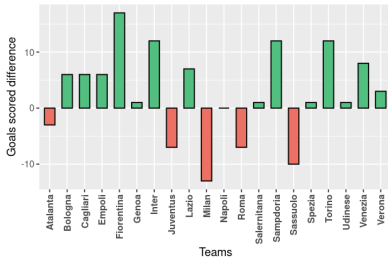
Home effect

- *Home effect*: beneficio associato al proprio stadio di casa.
- Le squadre, quando giocano nel proprio stadio, vincono più partite e segnano più goal rispetto a quando sono in trasferta.
- Numerosi fattori: maggior supporto dei propri tifosi, assenza di fatica dovuta al viaggio, familiarità con il proprio campo, ...

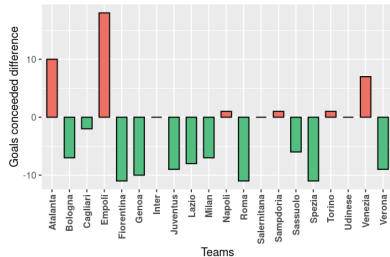


Home effect

Goals scored difference Home vs Away

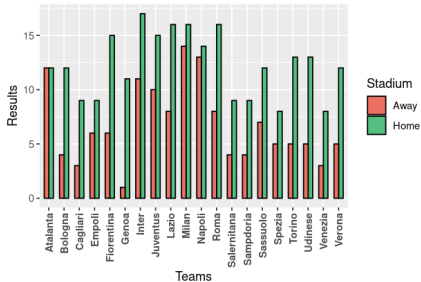


Goals conceded difference Home vs Away

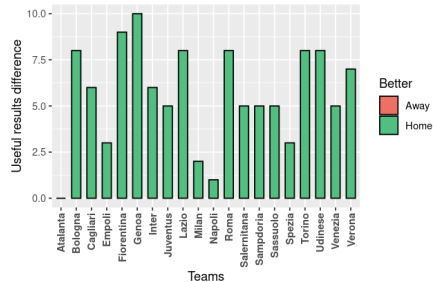


Home effect

Useful results Serie A 2021-22



Useful results difference Home vs Away



Il modello di Dixon e Coles (versione statica)

- Si parte ancora dalle variabili $X_{ij} \sim \text{Pois}(\lambda_{ij})$ e $Y_{ij} \sim \text{Pois}(\mu_{ij})$.
- Tenendo conto dell'*home effect*, si specificano i parametri come:

$$\lambda_{ij} = \exp\{\alpha_i + \beta_j + \gamma\} \quad ; \quad \mu_{ij} = \exp\{\alpha_j + \beta_i\} \quad .$$

- Per tenere conto dell'interdipendenza tra goals:

$$\tau_{\lambda_{ij}, \mu_{ij}}(x, y) = \begin{cases} 1 - \lambda_{ij} \cdot \mu_{ij} \cdot \rho & x = y = 0 \\ 1 + \lambda_{ij} \cdot \rho & x = 0, y = 1 \\ 1 + \mu_{ij} \cdot \rho & x = 1, y = 0 \\ 1 - \rho & x = y = 1 \\ 1 & \text{altrimenti} \end{cases}$$

in cui ρ è un parametro di dipendenza tale che:

$$\max \left\{ -\frac{1}{\lambda_{ij}}, -\frac{1}{\mu_{ij}} \right\} \leq \rho \leq \min \left\{ \frac{1}{\lambda_{ij}\mu_{ij}}, 1 \right\} .$$

Il modello di Dixon e Coles (versione statica)

- Si parte ancora dalle variabili $X_{ij} \sim \text{Pois}(\lambda_{ij})$ e $Y_{ij} \sim \text{Pois}(\mu_{ij})$.
- Tenendo conto dell'*home effect*, si specificano i parametri come:

$$\lambda_{ij} = \exp\{\alpha_i + \beta_j + \gamma\} \quad ; \quad \mu_{ij} = \exp\{\alpha_j + \beta_i\} \quad .$$

- Per tenere conto dell'interdipendenza tra goals:

$$\tau_{\lambda_{ij}, \mu_{ij}}(x, y) = \begin{cases} 1 - \lambda_{ij} \cdot \mu_{ij} \cdot \rho & x = y = 0 \\ 1 + \lambda_{ij} \cdot \rho & x = 0, y = 1 \\ 1 + \mu_{ij} \cdot \rho & x = 1, y = 0 \\ 1 - \rho & x = y = 1 \\ 1 & \text{altrimenti} \end{cases}$$

in cui ρ è un parametro di dipendenza tale che:

$$\max \left\{ -\frac{1}{\lambda_{ij}}, -\frac{1}{\mu_{ij}} \right\} \leq \rho \leq \min \left\{ \frac{1}{\lambda_{ij}\mu_{ij}}, 1 \right\} .$$

Il modello di Dixon e Coles (versione statica)

- Si parte ancora dalle variabili $X_{ij} \sim \text{Pois}(\lambda_{ij})$ e $Y_{ij} \sim \text{Pois}(\mu_{ij})$.
- Tenendo conto dell'*home effect*, si specificano i parametri come:

$$\lambda_{ij} = \exp\{\alpha_i + \beta_j + \gamma\} \quad ; \quad \mu_{ij} = \exp\{\alpha_j + \beta_i\} \quad .$$

- Per tenere conto dell'interdipendenza tra goals:

$$\tau_{\lambda_{ij}, \mu_{ij}}(x, y) = \begin{cases} 1 - \lambda_{ij} \cdot \mu_{ij} \cdot \rho & x = y = 0 \\ 1 + \lambda_{ij} \cdot \rho & x = 0, y = 1 \\ 1 + \mu_{ij} \cdot \rho & x = 1, y = 0 \\ 1 - \rho & x = y = 1 \\ 1 & \text{altrimenti} \end{cases}$$

in cui ρ è un parametro di dipendenza tale che:

$$\max \left\{ -\frac{1}{\lambda_{ij}}, -\frac{1}{\mu_{ij}} \right\} \leq \rho \leq \min \left\{ \frac{1}{\lambda_{ij}\mu_{ij}}, 1 \right\} .$$

Il modello di Dixon e Coles (versione statica)

- Dopo aver osservato un sufficiente numero di partite, è possibile stimare a massima verosimiglianza i parametri del modello.

$$L(\alpha_i, \beta_i, \gamma, \rho) = \prod_{k=1}^N \tau_{\lambda_k, \mu_k}(x_k, y_k) \frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \cdot \frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} ;$$

$$l(\alpha_i, \beta_i, \gamma, \rho) = \sum_{k=1}^N \log[\tau_{\lambda_k, \mu_k}(x_k, y_k)] + \log \left[\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \right] + \log \left[\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right].$$

- Per garantire l'identificabilità del modello si introducono i vincoli:

$$\frac{1}{n} \sum_{i=1}^n \alpha_i = 1 \quad ; \quad \frac{1}{n} \sum_{i=1}^n \beta_i = 1 \quad ,$$

- o equivalentemente (utile per l'implementazione):

$$\sum_{i=1}^{n-1} \alpha_i = -\alpha_n \quad ; \quad \sum_{i=1}^{n-1} \beta_i = -\beta_n \quad .$$

Il modello di Dixon e Coles (versione statica)

- Dopo aver osservato un sufficiente numero di partite, è possibile stimare a massima verosimiglianza i parametri del modello.

$$L(\alpha_i, \beta_i, \gamma, \rho) = \prod_{k=1}^N \tau_{\lambda_k, \mu_k}(x_k, y_k) \frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \cdot \frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} ;$$

$$l(\alpha_i, \beta_i, \gamma, \rho) = \sum_{k=1}^N \log[\tau_{\lambda_k, \mu_k}(x_k, y_k)] + \log \left[\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \right] + \log \left[\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right].$$

- Per garantire l'identificabilità del modello si introducono i vincoli:

$$\frac{1}{n} \sum_{i=1}^n \alpha_i = 1 \quad ; \quad \frac{1}{n} \sum_{i=1}^n \beta_i = 1 \quad ,$$

- o equivalentemente (utile per l'implementazione):

$$\sum_{i=1}^{n-1} \alpha_i = -\alpha_n \quad ; \quad \sum_{i=1}^{n-1} \beta_i = -\beta_n \quad .$$

Il modello di Dixon e Coles (versione statica)

- Dopo aver osservato un sufficiente numero di partite, è possibile stimare a massima verosimiglianza i parametri del modello.

$$L(\alpha_i, \beta_i, \gamma, \rho) = \prod_{k=1}^N \tau_{\lambda_k, \mu_k}(x_k, y_k) \frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \cdot \frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} ;$$

$$l(\alpha_i, \beta_i, \gamma, \rho) = \sum_{k=1}^N \log[\tau_{\lambda_k, \mu_k}(x_k, y_k)] + \log \left[\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \right] + \log \left[\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right].$$

- Per garantire l'identificabilità del modello si introducono i vincoli:

$$\frac{1}{n} \sum_{i=1}^n \alpha_i = 1 \quad ; \quad \frac{1}{n} \sum_{i=1}^n \beta_i = 1 \quad ,$$

- o equivalentemente (utile per l'implementazione):

$$\sum_{i=1}^{n-1} \alpha_i = -\alpha_n \quad ; \quad \sum_{i=1}^{n-1} \beta_i = -\beta_n \quad .$$

- La staticità dei parametri costituisce un limite strutturale del modello: le squadre non hanno un rendimento costante nel tempo, ma vivono momenti più o meno prolifici in fase offensiva e difensiva.
- Idea generale: le prestazioni di una squadra sono correlate maggiormente ai risultati delle partite più recenti rispetto a quelli delle partite più distanti nel tempo.
- Si assume che i parametri siano localmente costanti nel tempo e che le informazioni passate abbiano meno valore di quelle più recenti. Inoltre, i parametri ad ogni istante di tempo t sono stimati sulla base delle partite giocate fino a quel momento.

Il modello di Dixon e Coles (versione dinamica)

- Dixon e Coles apportarono una modifica sostanziale al modello

$$L(\alpha_i, \beta_i, \gamma, \rho) = \prod_{k \in A_t} \left[\tau_{\lambda_k, \mu_k}(x_k, y_k) \frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \cdot \frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right]^{\phi(t-t_k)} ;$$

$$l(\alpha_i, \beta_i, \gamma, \rho) = \sum_{k \in A_t} \phi(t - t_k) \left[\log[\tau_{\lambda_k, \mu_k}(x_k, y_k)] + \log \left(\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \right) + \log \left(\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right) \right]$$

in cui:

- t_k è la data della k -esima partita;
 - $A_t \doteq \{k : t_k \leq t\}$;
 - $\phi(\cdot)$ è una funzione non-crescente che serve a pesare i dati in base alla loro collocazione temporale.
- Massimizzare la (log-)verosimiglianza alla data t consente ancora di ricavare le stime dei parametri, con l'accortezza che adesso le stime saranno riferite solamente a quella precisa data.

Il modello di Dixon e Coles (versione dinamica)

- Dixon e Coles apportarono una modifica sostanziale al modello

$$L(\alpha_i, \beta_i, \gamma, \rho) = \prod_{k \in A_t} \left[\tau_{\lambda_k, \mu_k}(x_k, y_k) \frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \cdot \frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right]^{\phi(t-t_k)} ;$$

$$l(\alpha_i, \beta_i, \gamma, \rho) = \sum_{k \in A_t} \phi(t - t_k) \left[\log[\tau_{\lambda_k, \mu_k}(x_k, y_k)] + \log \left(\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \right) + \log \left(\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right) \right]$$

in cui:

- t_k è la data della k -esima partita;
 - $A_t \doteq \{k : t_k \leq t\}$;
 - $\phi(\cdot)$ è una funzione non-crescente che serve a pesare i dati in base alla loro collocazione temporale.
- Massimizzare la (log-)verosimiglianza alla data t consente ancora di ricavare le stime dei parametri, con l'accortezza che adesso le stime saranno riferite solamente a quella precisa data.

Modello dinamico di Dixon e Coles: la funzione $\phi(t)$

- La prima soluzione proposta da Dixon e Coles:

$$\phi(t) = \begin{cases} 1 & t \leq t_0 \\ 0 & t > t_0 \end{cases}$$

in cui t_0 è una soglia fissata a priori.

- La seconda soluzione proposta da Dixon e Coles:

$$\phi(t) = e^{-\xi t}$$

in cui ξ rappresenta un fattore di liscio.

Modello dinamico di Dixon e Coles: la funzione $\phi(t)$

- La prima soluzione proposta da Dixon e Coles:

$$\phi(t) = \begin{cases} 1 & t \leq t_0 \\ 0 & t > t_0 \end{cases}$$

in cui t_0 è una soglia fissata a priori.

- La seconda soluzione proposta da Dixon e Coles:

$$\phi(t) = e^{-\xi t}$$

in cui ξ rappresenta un fattore di liscio.

La stima dei parametri (modello di Maher)

Squadra	α	s.e.	IC (95%)
Udinese	0.641	0.082	[0.633 ; 0.650]
Napoli	0.628	0.073	[0.621 ; 0.635]
Inter	0.594	0.065	[0.588 ; 0.601]
Lazio	0.582	0.066	[0.576 ; 0.589]
Sampdoria	0.576	0.085	[0.567 ; 0.584]
Milan	0.571	0.069	[0.564 ; 0.578]
Verona	0.540	0.067	[0.533 ; 0.547]
Atalanta	0.532	0.066	[0.525 ; 0.539]
Sassuolo	0.525	0.066	[0.518 ; 0.531]
Juventus	0.523	0.069	[0.516 ; 0.530]
Empoli	0.516	0.073	[0.509 ; 0.524]
Bologna	0.494	0.075	[0.487 ; 0.502]
Torino	0.487	0.072	[0.480 ; 0.495]
Roma	0.485	0.063	[0.478 ; 0.491]
Spezia	0.481	0.075	[0.473 ; 0.488]
Fiorentina	0.469	0.061	[0.463 ; 0.475]
Cagliari	0.457	0.078	[0.449 ; 0.465]
Salernitana	0.451	0.079	[0.443 ; 0.459]
Genoa	0.448	0.086	[0.439 ; 0.457]
Venezia	0.421	0.072	[0.413 ; 0.428]

Tabella: Stime dei coefficienti d'attacco

Squadra	β	s.e.	IC (95%)
Milan	0.363	0.065	[0.357 ; 0.370]
Atalanta	0.395	0.057	[0.390 ; 0.401]
Spezia	0.405	0.048	[0.400 ; 0.410]
Inter	0.446	0.079	[0.438 ; 0.454]
Juventus	0.469	0.077	[0.461 ; 0.477]
Verona	0.481	0.063	[0.475 ; 0.488]
Torino	0.490	0.077	[0.482 ; 0.498]
Roma	0.491	0.075	[0.484 ; 0.499]
Cagliari	0.493	0.060	[0.487 ; 0.499]
Fiorentina	0.506	0.071	[0.499 ; 0.513]
Sampdoria	0.516	0.065	[0.510 ; 0.523]
Genoa	0.517	0.067	[0.511 ; 0.524]
Bologna	0.523	0.071	[0.516 ; 0.530]
Napoli	0.528	0.095	[0.518 ; 0.537]
Sassuolo	0.547	0.067	[0.540 ; 0.554]
Lazio	0.567	0.075	[0.559 ; 0.574]
Venezia	0.573	0.069	[0.566 ; 0.580]
Empoli	0.576	0.069	[0.569 ; 0.583]
Udinese	0.580	0.076	[0.572 ; 0.587]
Salernitana	0.585	0.066	[0.579 ; 0.592]

Tabella: Stime dei coefficienti di difesa

La stima dei parametri (modello Dixon-Coles statico)

Squadra	α	s.e.	IC (95%)
Inter	0.504	0.112	[0.493 ; 0.515]
Lazio	0.440	0.117	[0.428 ; 0.451]
Napoli	0.358	0.119	[0.346 ; 0.370]
Milan	0.291	0.123	[0.279 ; 0.304]
Verona	0.253	0.127	[0.240 ; 0.266]
Atalanta	0.246	/	/
Sassuolo	0.245	0.128	[0.232 ; 0.258]
Udinese	0.181	0.131	[0.167 ; 0.194]
Fiorentina	0.135	0.133	[0.122 ; 0.149]
Roma	0.133	0.133	[0.120 ; 0.147]
Juventus	0.083	0.136	[0.069 ; 0.097]
Empoli	-0.017	0.145	[-0.031 ; -0.002]
Sampdoria	-0.124	0.152	[-0.139 ; -0.109]
Torino	-0.153	0.152	[-0.168 ; -0.138]
Bologna	-0.177	0.155	[-0.192 ; -0.161]
Spezia	-0.247	0.161	[-0.263 ; -0.231]
Cagliari	-0.460	0.179	[-0.478 ; -0.442]
Venezia	-0.463	0.179	[-0.481 ; -0.445]
Salernitana	-0.494	0.182	[-0.512 ; -0.475]
Genoa	-0.736	0.204	[-0.756 ; -0.715]

Tabella: Stime dei coefficienti d'attacco

Squadra	β	s.e.	IC (95%)
Milan	-0.584	0.188	[-0.603 ; -0.565]
Napoli	-0.579	0.188	[-0.598 ; -0.560]
Inter	-0.522	0.185	[-0.540 ; -0.503]
Juventus	-0.388	0.171	[-0.406 ; -0.371]
Torino	-0.283	0.161	[-0.300 ; -0.267]
Roma	-0.214	0.157	[-0.230 ; -0.198]
Atalanta	-0.074	/	/
Fiorentina	-0.028	0.144	[-0.042 ; -0.013]
Bologna	0.050	0.138	[0.036 ; 0.064]
Udinese	0.126	0.135	[0.113 ; 0.140]
Genoa	0.133	0.132	[0.120 ; 0.146]
Lazio	0.144	0.135	[0.130 ; 0.157]
Verona	0.150	0.133	[0.136 ; 0.163]
Sampdoria	0.201	0.129	[0.188 ; 0.214]
Cagliari	0.270	0.124	[0.258 ; 0.283]
Sassuolo	0.271	0.126	[0.259 ; 0.284]
Venezia	0.283	0.123	[0.271 ; 0.296]
Empoli	0.313	0.122	[0.301 ; 0.326]
Spezia	0.317	0.121	[0.305 ; 0.329]
Salernitana	0.412	0.116	[0.401 ; 0.424]

Tabella: Stime dei coefficienti di difesa

Parametro	Stima	s.e.	IC(95 %)
γ	0.305	0.044	[0.301 ; 0.309]
ρ	-0.071	0.077	[-0.079 ; -0.063]

Tabella: Stime dei coefficienti γ e ρ

La stima dei parametri (modello Dixon-Coles dinamico)

- Assumendo $\phi(t) = e^{-\xi t}$, la log-verosimiglianza del modello risultava:

$$l(\alpha_i, \beta_i, \gamma, \rho) = \sum_{k \in A_t} e^{-\xi(t-t_k)} \left[\log[\tau_{\lambda_k, \mu_k}(x_k, y_k)] + \log \left(\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \right) + \log \left(\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right) \right]$$

- Non ci si può più limitare ad inserire la funzione nell'algoritmo di ottimizzazione (funzione *optim()* di R). Infatti:

$$\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \in [0, 1] \implies \log \left[\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \right] \leq 0 \quad ,$$

$$\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \in [0, 1] \implies \log \left[\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right] \leq 0 \quad .$$

- Se ρ è molto piccolo, inoltre, si ha $\tau_{\lambda_k, \mu_k}(x_k, y_k) \approx 1$, per cui:

$$\log[\tau_{\lambda_k, \mu_k}(x_k, y_k)] \approx 0 \quad .$$

La stima dei parametri (modello Dixon-Coles dinamico)

- Definiamo ora la quantità:

$$P_k = \underbrace{\log [\tau_{\lambda_k, \mu_k}(x_k, y_k)]}_{\approx 0} + \underbrace{\log \left(\frac{(\lambda_k)^{x_k}}{x_k!} e^{-\lambda_k} \right)}_{\leq 0} + \underbrace{\log \left(\frac{(\mu_k)^{y_k}}{y_k!} e^{-\mu_k} \right)}_{\leq 0} \leq 0 \quad \forall k$$

- Possiamo così riscrivere la log-verosimiglianza nella forma:

$$l(\alpha_i, \beta_i, \gamma, \rho) = \sum_{k \in A_t} e^{-\xi(t-t_k)} \cdot P_k$$

- La log-verosimiglianza è allora una funzione crescente di ξ :

$$\begin{aligned} \frac{\partial l}{\partial \xi} &= \frac{\partial}{\partial \xi} \left(\sum_{k \in A_t} e^{-\xi(t-t_k)} \cdot P_k \right) \\ &= \sum_{k \in A_t} \frac{\partial}{\partial \xi} \left(e^{-\xi(t-t_k)} \cdot P_k \right) \\ &= \sum_{k \in A_t} \underbrace{e^{-\xi(t-t_k)}}_{\geq 0} \cdot \underbrace{P_k}_{\leq 0} \cdot \underbrace{(t_k - t)}_{\leq 0} \geq 0 \quad \forall \xi \end{aligned}$$

Verosimiglianza profilo per la stima di ξ

- 1 Si fissa un valore di ξ ;
- 2 si parte da una giornata di campionato (i);
- 3 si stimano (a max verosimiglianza) gli altri parametri del modello, usando come training-set tutte le giornate fino alla ($i - 1$);
- 4 si valuta poi l'accuratezza delle previsioni del modello in relazione alla i -esima giornata (test-set);
- 5 si passa alla giornata $i + 1$ e si ri-stimano i parametri, questa volta usando come training-set tutte le giornate fino alla i -esima;
- 6 si continua allo stesso modo fino all'ultima giornata di campionato;
- 7 si cambia il valore di ξ e si ripete la procedura.

Verosimiglianza profilo per la stima di ξ

- 1 Si fissa un valore di ξ ;
- 2 si parte da una giornata di campionato (i);
- 3 si stimano (a max verosimiglianza) gli altri parametri del modello, usando come training-set tutte le giornate fino alla ($i - 1$);
- 4 si valuta poi l'accuratezza delle previsioni del modello in relazione alla i -esima giornata (test-set);
- 5 si passa alla giornata $i + 1$ e si ri-stimano i parametri, questa volta usando come training-set tutte le giornate fino alla i -esima;
- 6 si continua allo stesso modo fino all'ultima giornata di campionato;
- 7 si cambia il valore di ξ e si ripete la procedura.

Verosimiglianza profilo per la stima di ξ

- 1 Si fissa un valore di ξ ;
- 2 si parte da una giornata di campionato (i);
- 3 si stimano (a max verosimiglianza) gli altri parametri del modello, usando come training-set tutte le giornate fino alla ($i - 1$);
- 4 si valuta poi l'accuratezza delle previsioni del modello in relazione alla i -esima giornata (test-set);
- 5 si passa alla giornata $i + 1$ e si ri-stimano i parametri, questa volta usando come training-set tutte le giornate fino alla i -esima;
- 6 si continua allo stesso modo fino all'ultima giornata di campionato;
- 7 si cambia il valore di ξ e si ripete la procedura.

Verosimiglianza profilo per la stima di ξ

- 1 Si fissa un valore di ξ ;
- 2 si parte da una giornata di campionato (i);
- 3 si stimano (a max verosimiglianza) gli altri parametri del modello, usando come training-set tutte le giornate fino alla ($i - 1$);
- 4 si valuta poi l'accuratezza delle previsioni del modello in relazione alla i -esima giornata (test-set);
- 5 si passa alla giornata $i + 1$ e si ri-stimano i parametri, questa volta usando come training-set tutte le giornate fino alla i -esima;
- 6 si continua allo stesso modo fino all'ultima giornata di campionato;
- 7 si cambia il valore di ξ e si ripete la procedura.

Verosimiglianza profilo per la stima di ξ

- ① Si fissa un valore di ξ ;
- ② si parte da una giornata di campionato (i);
- ③ si stimano (a max verosimiglianza) gli altri parametri del modello, usando come training-set tutte le giornate fino alla ($i - 1$);
- ④ si valuta poi l'accuratezza delle previsioni del modello in relazione alla i -esima giornata (test-set);
- ⑤ si passa alla giornata $i + 1$ e si ri-stimano i parametri, questa volta usando come training-set tutte le giornate fino alla i -esima;
- ⑥ si continua allo stesso modo fino all'ultima giornata di campionato;
- ⑦ si cambia il valore di ξ e si ripete la procedura.

Verosimiglianza profilo per la stima di ξ

- ① Si fissa un valore di ξ ;
- ② si parte da una giornata di campionato (i);
- ③ si stimano (a max verosimiglianza) gli altri parametri del modello, usando come training-set tutte le giornate fino alla ($i - 1$);
- ④ si valuta poi l'accuratezza delle previsioni del modello in relazione alla i -esima giornata (test-set);
- ⑤ si passa alla giornata $i + 1$ e si ri-stimano i parametri, questa volta usando come training-set tutte le giornate fino alla i -esima;
- ⑥ si continua allo stesso modo fino all'ultima giornata di campionato;
- ⑦ si cambia il valore di ξ e si ripete la procedura.

Verosimiglianza profilo per la stima di ξ

- 1 Si fissa un valore di ξ ;
- 2 si parte da una giornata di campionato (i);
- 3 si stimano (a max verosimiglianza) gli altri parametri del modello, usando come training-set tutte le giornate fino alla ($i - 1$);
- 4 si valuta poi l'accuratezza delle previsioni del modello in relazione alla i -esima giornata (test-set);
- 5 si passa alla giornata $i + 1$ e si ri-stimano i parametri, questa volta usando come training-set tutte le giornate fino alla i -esima;
- 6 si continua allo stesso modo fino all'ultima giornata di campionato;
- 7 si cambia il valore di ξ e si ripete la procedura.

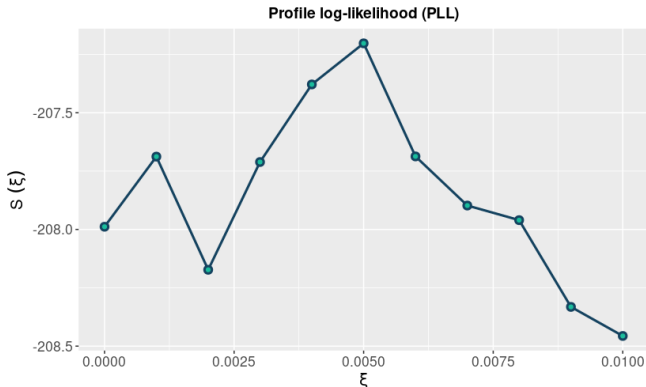
- Il valore di ξ associato alla migliore capacità predittiva al modello è:

$$\xi^* = \operatorname{argmax}_{\xi \in \mathbb{R}^+} S(\xi), \text{ con } S(\xi) = \sum_{k=1}^N (\delta_k^H \log p_k^H + \delta_k^D \log p_k^D + \delta_k^A \log p_k^A).$$

Verosimiglianza profilo per la stima di ξ

- Il valore di ξ associato alla migliore capacità predittiva al modello è:

$$\xi^* = \operatorname{argmax}_{\xi \in \mathbb{R}^+} S(\xi), \text{ con } S(\xi) = \sum_{k=1}^N (\delta_k^H \log p_k^H + \delta_k^D \log p_k^D + \delta_k^A \log p_k^A).$$



- Date due squadre i e j , è possibile associare a ciascun risultato $(x; y)$ una misura di probabilità, data da:

$$Pr(X_{ij} = x; Y_{ij} = y) = \tau_{\lambda_{ij}, \mu_{ij}}(x, y) \cdot \frac{(\alpha_i \cdot \beta_j \cdot \gamma)^x}{x!} e^{-\alpha_i \cdot \beta_j \cdot \gamma} \cdot \frac{(\alpha_j \cdot \beta_i)^y}{y!} e^{-\alpha_j \cdot \beta_i}$$

- Gli esiti di una partita possono essere descritti dalla matrice A_{ij} , in cui l'entrata a_{ij} contiene la probabilità congiunta che la squadra di casa segni i goal mentre la squadra in trasferta ne segni j .

$$A_{ij} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0g} \\ a_{10} & a_{11} & \cdots & a_{1g} \\ \vdots & \vdots & \ddots & \vdots \\ a_{g0} & a_{g2} & \cdots & a_{gg} \end{bmatrix}$$

- Date due squadre i e j , è possibile associare a ciascun risultato $(x; y)$ una misura di probabilità, data da:

$$Pr(X_{ij} = x; Y_{ij} = y) = \tau_{\lambda_{ij}, \mu_{ij}}(x, y) \cdot \frac{(\alpha_i \cdot \beta_j \cdot \gamma)^x}{x!} e^{-\alpha_i \cdot \beta_j \cdot \gamma} \cdot \frac{(\alpha_j \cdot \beta_i)^y}{y!} e^{-\alpha_j \cdot \beta_i}$$

- Gli esiti di una partita possono essere descritti dalla matrice A_{ij} , in cui l'entrata a_{ij} contiene la probabilità congiunta che la squadra di casa segni i goal mentre la squadra in trasferta ne segni j .

$$A_{ij} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0g} \\ a_{10} & a_{11} & \cdots & a_{1g} \\ \vdots & \vdots & \ddots & \vdots \\ a_{g0} & a_{g2} & \cdots & a_{gg} \end{bmatrix}$$

Previsioni di probabilità

- Date due squadre i e j , è possibile associare a ciascun risultato $(x; y)$ una misura di probabilità, data da:

$$Pr(X_{ij} = x; Y_{ij} = y) = \tau_{\lambda_{ij}, \mu_{ij}}(x, y) \cdot \frac{(\alpha_i \cdot \beta_j \cdot \gamma)^x}{x!} e^{-\alpha_i \cdot \beta_j \cdot \gamma} \cdot \frac{(\alpha_j \cdot \beta_i)^y}{y!} e^{-\alpha_j \cdot \beta_i}$$

- Gli esiti di una partita possono essere descritti dalla matrice A_{ij} , in cui l'entrata a_{ij} contiene la probabilità congiunta che la squadra di casa segni i goal mentre la squadra in trasferta ne segni j .

$$A_{ij} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0g} \\ a_{10} & a_{11} & \cdots & a_{1g} \\ \vdots & \vdots & \ddots & \vdots \\ a_{g0} & a_{g2} & \cdots & a_{gg} \end{bmatrix}$$

$$Pr(H) = \sum_{i>j} a_{ij}$$

Previsioni di probabilità

- Date due squadre i e j , è possibile associare a ciascun risultato $(x; y)$ una misura di probabilità, data da:

$$Pr(X_{ij} = x; Y_{ij} = y) = \tau_{\lambda_{ij}, \mu_{ij}}(x, y) \cdot \frac{(\alpha_i \cdot \beta_j \cdot \gamma)^x}{x!} e^{-\alpha_i \cdot \beta_j \cdot \gamma} \cdot \frac{(\alpha_j \cdot \beta_i)^y}{y!} e^{-\alpha_j \cdot \beta_i}$$

- Gli esiti di una partita possono essere descritti dalla matrice A_{ij} , in cui l'entrata a_{ij} contiene la probabilità congiunta che la squadra di casa segni i goal mentre la squadra in trasferta ne segni j .

$$A_{ij} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0g} \\ a_{10} & a_{11} & \cdots & a_{1g} \\ \vdots & \vdots & \ddots & \vdots \\ a_{g0} & a_{g2} & \cdots & a_{gg} \end{bmatrix}$$

$$Pr(H) = \sum_{i>j} a_{ij}$$

$$Pr(D) = \sum_{i=j} a_{ij} = tr(A_{ij})$$

- Date due squadre i e j , è possibile associare a ciascun risultato $(x; y)$ una misura di probabilità, data da:

$$Pr(X_{ij} = x; Y_{ij} = y) = \tau_{\lambda_{ij}, \mu_{ij}}(x, y) \cdot \frac{(\alpha_i \cdot \beta_j \cdot \gamma)^x}{x!} e^{-\alpha_i \cdot \beta_j \cdot \gamma} \cdot \frac{(\alpha_j \cdot \beta_i)^y}{y!} e^{-\alpha_j \cdot \beta_i}$$

- Gli esiti di una partita possono essere descritti dalla matrice A_{ij} , in cui l'entrata a_{ij} contiene la probabilità congiunta che la squadra di casa segni i goal mentre la squadra in trasferta ne segni j .

$$A_{ij} = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0g} \\ a_{10} & a_{11} & \cdots & a_{1g} \\ \vdots & \vdots & \ddots & \vdots \\ a_{g0} & a_{g2} & \cdots & a_{gg} \end{bmatrix}$$

$$Pr(H) = \sum_{i>j} a_{ij}$$

$$Pr(D) = \sum_{i=j} a_{ij} = tr(A_{ij})$$

$$Pr(A) = \sum_{i<j} a_{ij}$$

Previsioni di probabilità

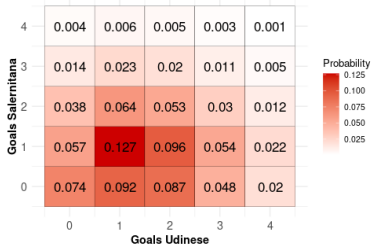
Match: Inter - Sampdoria



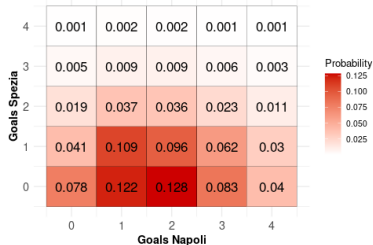
Match: Sassuolo - Milan



Match: Salernitana - Udinese

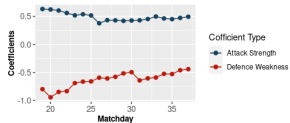


Match: Spezia - Napoli

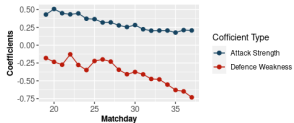


Evoluzione delle abilità delle squadre

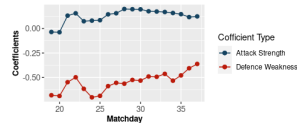
Inter abilities over time



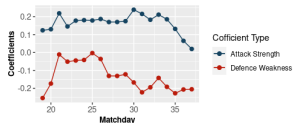
Milan abilities over time



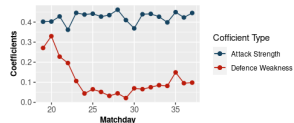
Juventus abilities over time



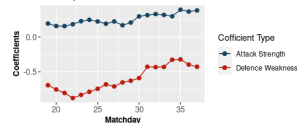
Roma abilities over time



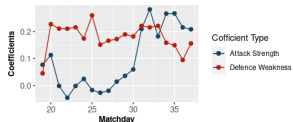
Lazio abilities over time



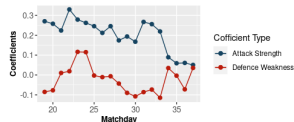
Napoli abilities over time



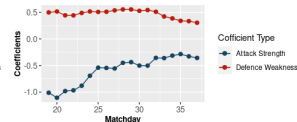
Udinese abilities over time



Fiorentina abilities over time

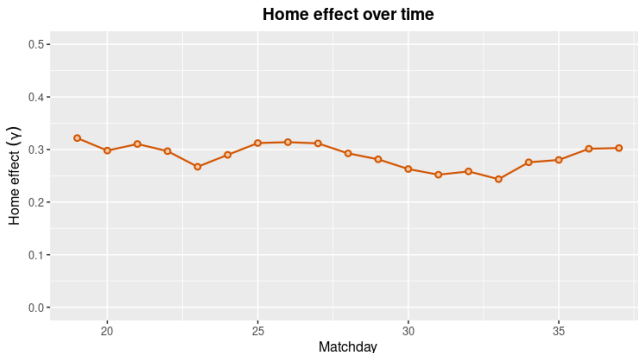


Salernitana abilities over time



Evoluzione dell'*home effect*

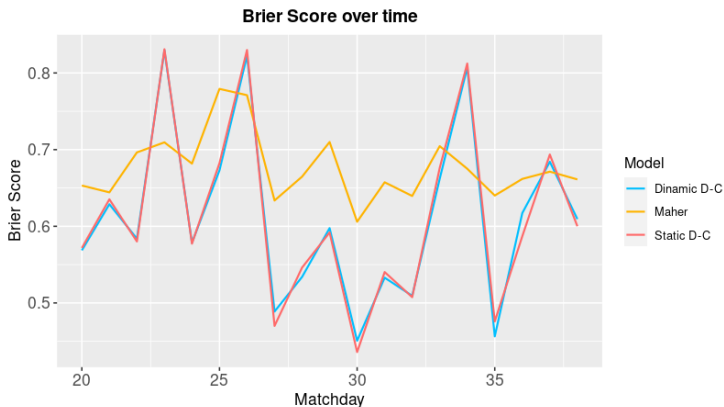
- Il parametro γ è dinamico e può cambiare nel corso del tempo;
- anche se la serie storia suggerisce stazionarietà (in media e varianza).



Valutazione e confronto tra modelli: il *Brier Score*

$$BS = \frac{1}{M} \sum_{m=1}^M \sum_i (p_{mi} - d_{mi})^2 \quad \text{con } i \in \{H, D, A\}$$

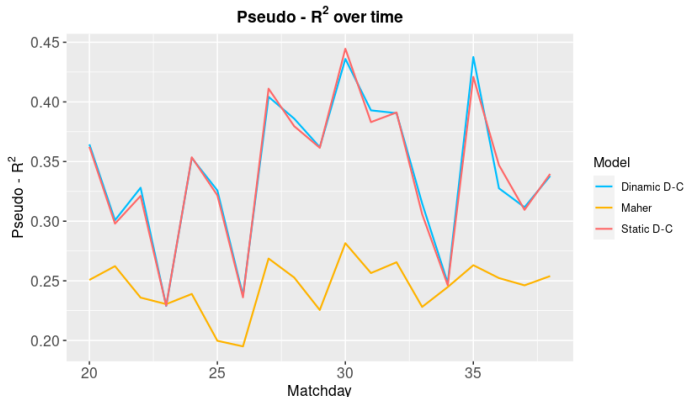
Model	Maher model	Static Dixon-Coles	Dinamic Dixon-Coles ($\xi = 0.005$)
Brier Score	0.67686	0.61288	0.61234



Valutazione e confronto tra modelli: lo *Pseudo-R*²

$$\text{Pseudo-}R^2 = \left(\prod_{m=1}^M p_m \right)^{\frac{1}{M}}$$

Model	Maher model	Static Dixon-Coles	Dinamic Dixon-Coles ($\xi = 0.005$)
Pseudo-R^2	0.24385	0.33465	0.33604



Matrici di confusione: H-D-A

Dixon-Coles model (static)

Predicted values	H	23	33	55
	D	2	3	1
	A	40	18	15
		A	D	H
		Actual values		

Dixon-Coles model (dinamic)

Predicted values	H	22	32	54
	D	4	4	1
	A	39	18	16
		A	D	H
		Actual values		

Model metrics	Static Dixon-Coles	Dinamic Dixon-Coles
Precision(PPV)	0.514	0.493
NPV	0.769	0.765
Sensitivity	0.482	0.478
Overall accuracy	0.516	0.511
Balanced accuracy	0.615	0.612
F1-score	0.428	0.432

Matrici di confusione: 1X - 2

Dixon-Coles model (static)

Predicted values	Actual values	
	FALSE	TRUE
TRUE	45	115
FALSE	20	10

Dixon-Coles model (static)

Predicted values	Actual values	
	FALSE	TRUE
TRUE	27	81
FALSE	44	38

Model metrics	Static Dixon-Coles	Dinamic Dixon-Coles
Precision(PPV)	0.719	0.726
NPV	0.667	0.667
Sensitivity	0.920	0.912
Accuracy	0.711	0.716
F1-score	0.807	0.809

- Il modello D-C presenta svariati ambiti di applicazione ed è piuttosto semplice, sia dal punto di vista teorico che implementativo.
- Nell'ambito dei modelli *goal-based* è un prezioso riferimento, a partire da cui vengono spesso sviluppate nuove proposte.
- La verosimiglianza profilo è stata molto onerosa in termini di tempo.
 - Parallelizzazione del codice?
 - Altre strutture dati?
 - Altri linguaggi di programmazione?
- Chiaro il confronto del modello Dixon-Coles rispetto a quello di Maher; meno quello tra le sue versioni statica e dinamica.

- Il modello D-C presenta svariati ambiti di applicazione ed è piuttosto semplice, sia dal punto di vista teorico che implementativo.
- Nell'ambito dei modelli *goal-based* è un prezioso riferimento, a partire da cui vengono spesso sviluppate nuove proposte.
- La verosimiglianza profilo è stata molto onerosa in termini di tempo.
 - Parallelizzazione del codice?
 - Altre strutture dati?
 - Altri linguaggi di programmazione?
- Chiaro il confronto del modello Dixon-Coles rispetto a quello di Maher; meno quello tra le sue versioni statica e dinamica.

- Il modello D-C presenta svariati ambiti di applicazione ed è piuttosto semplice, sia dal punto di vista teorico che implementativo.
- Nell'ambito dei modelli *goal-based* è un prezioso riferimento, a partire da cui vengono spesso sviluppate nuove proposte.
- La verosimiglianza profilo è stata molto onerosa in termini di tempo.
 - Parallelizzazione del codice?
 - Altre strutture dati?
 - Altri linguaggi di programmazione?
- Chiaro il confronto del modello Dixon-Coles rispetto a quello di Maher; meno quello tra le sue versioni statica e dinamica.

- Il modello D-C presenta svariati ambiti di applicazione ed è piuttosto semplice, sia dal punto di vista teorico che implementativo.
- Nell'ambito dei modelli *goal-based* è un prezioso riferimento, a partire da cui vengono spesso sviluppate nuove proposte.
- La verosimiglianza profilo è stata molto onerosa in termini di tempo.
 - Parallelizzazione del codice?
 - Altre strutture dati?
 - Altri linguaggi di programmazione?
- Chiaro il confronto del modello Dixon-Coles rispetto a quello di Maher; meno quello tra le sue versioni statica e dinamica.

Grazie per l'attenzione!