Modélisation de matches de football

HOANG Lê nguyên

On se propose de modéliser les matchs de football. L'ensemble des équipes sera paramétré par une variable θ dont on précisera la dimension ultérieurement. Une variable aléatoire permettra alors de déterminer l'ensemble des résultats, à partir des paramètres. A priori, en tant qu'observateurs des matchs de football, seul l'ensemble des résultats nous est accessible. On essaiera alors d'estimer à partir de celui-ci des paramètres, qu'on notera $\widehat{\theta}$. On pourra alors émettre des pronostics en appliquant à nouveau notre loi de probabilité. On écrit alors :

$$\theta \quad \longmapsto \quad \text{resultats} \quad \longmapsto \quad \widehat{\theta} \quad \longmapsto \quad \text{pronostics}$$

Remarquons que la variable aléatoire calculant les résultats et les pronostics dépend des paramètres et de l'organisation des matchs considérés. Par ailleurs l'application recalculant les paramètres devra ne dépendre que de l'ensemble des résultats. L'objectif de ce tipe est de déterminer les fonctions recherchées de façon cohérentes, et d'en étudier leurs propriétés.

1 Résultats de l'étude

1.1 La coupe du monde 2006

Voyons immédiatement un intérêt pratique de notre étude. on va pronostiquer les résultats de la coupe du monde 2006 suivant notre modèle. On se donne l'ensemble des équipes internationales, et on considère l'ensemble des scores des matchs internationaux précédents la coupe du monde, entre 2004 et 2006. L'application numérique décrite dans la partie 2 nous donne une valeur des paramètres correspondant à un maximum local strict de la log-vraisemblance (voir partie 2). En lançant 100 000 simulations, on a les résultats présentés dans le tableau 1 en annexe.

Les résultats de la coupe du monde sont alors très en accord avec ces pronostics. En effet, le vainqueur, l'Italie, est associé à une probabilité de 0.1, tandis que la France est crédité de 0.14. De plus, les huit quart-definalistes sont dans les treize favoris, alors que quatorze des seize équipes étant en huitième de finale sont dans les seize premiers.

1.2 Simulation d'un championnat

On rappelle que dans un championnat, une victoire rapporte 3 points, tandis qu'un match nul en rapporte 1, et qu'une défaite n'en rapporte pas. On considèrera alors le championnat de France de ligue 1 de football 2004/2005. Ainsi on considère l'ensemble des 20 équipes du championnat, et l'ensemble des matchs du championnat. On s'appuie alors sur les vrais résultats, et on applique alors la fonction de calcul des paramètres pour obtenir les niveaux des équipes. On remarque alors l'algorithme conduit directement à un maximum local strict de la log-vraisemblance.

On simule alors des championnats, pour obtenir les moyennes suivantes, sur 10 000 simulations (voir tableau 2 en annexe). On constate alors un écart assez important entre le classement réel et le classement simulé. Cependant, on peut rappeler que l'espérance calculée pour une équipe i est en fait la valeur suivante : $2\sum_{j\neq i} 3\mathbb{P}(X_{ij} > X_{ji}) + \mathbb{P}(X_{ij} = X_{ji})$.

On peut alors se dire qu'un tel classement ne représente pas vraiment le niveau des équipes, c'est pourquoi on réalise un classement à la différence de buts (tableau 3 en annexe). Le résultat calculé pour une équipe i est alors : $2\sum_{j\neq i}\sum_{k,l=0}^{+\infty} k\mathbb{P}(X_{ij} = k) - l\mathbb{P}(X_{ij} = l) = 2\sum_{j\neq i} E(X_{ij}) - E(X_{ji}) = 2\widehat{n_i}\sum_{j\neq i}\widehat{d_j} - 2\widehat{d_i}\sum_{j\neq i}\widehat{n_j}$ où les $\widehat{n_i}$ et $\widehat{d_i}$ sont les paramètres recalculés. On obtient finalement $\widehat{n_i}\dfrac{x_i}{\widehat{n_i}} - \widehat{d_i}\dfrac{y_i}{\widehat{n_i}} = x_i - y_i$, après avoir utilisé le fait que les paramètres calculés sont dans $(dL)^{-1}(0)$. Il est donc normal de trouver des espérances très proches à l'issue des simulations, puisqu'en théories on aurait égalité.

1.3 Biais, consistance

L'objectif de cette partie et d'étudier les propriétés du modèle statistique. Pour des applications numériques aisées, on fixera à 5 le nombre d'équipes.

Définition 1 *On se donne des paramètres θ. Soit l'algorithme suivant :*
- On choisit aléatoirement un ensemble de t matchs.
- On simule les matchs, puis on recalcule selon notre algorithme de nouveaux paramètres.
On définit alors une variable aléatoire $X_{\theta,t}$ dont la valeur est l'ensemble des paramètres recalculés.

Comme on le verra plus tard, $n_i d_j$ est l'espérance de X_{ij}. Or en général le nombre de buts moyens d'une équipe dans un match est dans $[0, 2\,;\, 5]$, on se limitera à des valeurs de n_i et d_j dans $[0, 2\,;\, 5]$. Ces valeurs seront dites raisonnables.

Conjecture 1 *Pour des valeurs paramètres raisonnables,* $\lim_{t \to +\infty} \mathbb{P}(\text{pseudo} - \text{unicite}) = 1$

On a en effet les résultats numériques suivants, pour des paramètres choisis aléatoirement dans l'ensemble des valeurs raisonnables, et sur 1000 simulations :

Nombre de matchs t	Cas où $H(n^1, d^1)$ est négative
50	544
100	651
200	784
400	903
1000	959

Définition 2 *On définit pour une valeur de t et de paramètre le biais, égal à* $E(X_{\theta,t}) - \theta$. *Le modèle statistique est dit biaisé si le biais est non nul.*
Il est consistant si pour tout des paramètres raisonnables, $\lim_{t \to +\infty} \text{Var}(X_{\theta,t}) = 0$.

Conjecture 2 *Le modèle proposé est globalement biaisé, mais il est consistant.*

On fixe la valeur des paramètres, et on regarde $X_{\theta,t}$, pour $t \in \{50, 1000, 10000\}$, après 1000 simulations, pour obtenir les résultats des tableaux 4,5 et 6. On peut alors remarquer que le biais est globalement positif, et qu'il semble tendre vers 0, quand t tend vers $+\infty$.

2 Modèle mathématique

2.1 Loi probabiliste

Définition 3 *On appellera variable aléatoire associée à l'attaque de i contre j la variable aléatoire qui au match entre i et j associe le nombre de buts marqués par l'équipe i. On la notera* X_{ij}. *Un score sera un ensemble de couples* (x_{tij}, x_{tji}), *représentant le score dans le match entre i et j, à la date t.*

Notons que X_{ii} n'est en fait pas défini. On supposera dorénavant pour simplifier que $\forall (i,j) \neq (k,l), X_{ij}$ et X_{kl} sont indépendants. En particulier, cela veut dire que X_{ij} et X_{ji} sont indépendants. Cette hypothèse est très discutable puisqu'a priori le fait qu'une équipe soit forte implique une certaine domination de sa part, ce qui laisse alors moins de chance à l'adversaire de marquer. Elle nous permet de n'avoir à étudier que des termes X_{ij} pour modéliser un ensemble de matchs. Remarquons que pour parfaitement décortiquer un match, il faudrait également prendre en compte d'autres résultats tels que la domination générale, la possession de balle, le nombre d'occasions, etc... Mais cela nous mènerait à une étude trop complexe.

Une première idée serait de modéliser un match de football par la loi binomiale. Cependant, il est encore plus naturel de le modéliser par la convergence de la loi binomiale qu'est la loi de Poisson.

Définition 4 X_λ *suit une loi de Poisson de paramètre* λ *si* $\forall k \in \mathbb{N}, \mathbb{P}(X_\lambda = k) = \frac{\lambda^k}{k!} e^{-\lambda}$.

Proposition 1 $\forall \lambda \in \mathbb{R}_+^*, E(X_\lambda) = \text{Var}(X_\lambda) = \lambda$.

On simule alors les matchs de football en prenant un réel t au hasard dans $]0, 1[$. La valeur de la variable aléatoire X_λ est alors l'unique entier x tel que

$$t \in \Big] \sum_{k=0}^{x-1} \mathbb{P}(X_\lambda = k), \sum_{k=0}^{x} \mathbb{P}(X_\lambda = k)\Big] = \Big] \sum_{k=0}^{x-1} \frac{\lambda^k}{k!} e^{-\lambda}, \sum_{k=0}^{x} \frac{\lambda^k}{k!} e^{-\lambda}\Big].$$

On pose donc, pour tout match entre i et j, deux variables aléatoires $X_{ij} = X_{\lambda_{ij}}$ et $X_{ji} = X_{\lambda_{ji}}$, avec des λ_{ij} dépendant des paramètres.

2.2 Loi de log-vraisemblance

Une méthode souvent utilisée en statistique consiste à maximiser la probabilité de la vraisemblance en fonction des paramètres, c'est-à-dire à trouver, pour un ensemble de résultats donné, l'ensemble des paramètres tel que $\mathbb{P}(\text{Vraisemblance}_\theta)$ soit maximale, la vraisemblance correspondant aux résultats donnés. Pour des raisons purement techniques, notamment pour dériver aisément, on s'intéressera au logarithme de cette probabilité.

Définition 5 *On appellera log-vraisemblance générale l'application* $L_0 : \theta \longmapsto \ln \mathbb{P}(\text{Vraisemblance}_\theta)$, *et ensemble général des solutions l'ensemble* $L_0^{-1}(\sup L_0)$.

On aimerait pour pouvoir définir notre calcul des paramètres, avoir un ensemble général des solutions de cardinal 1. Pour cela, on va préciser la définition de θ. L'ensemble des paramètres intéressants pour caractériser $\mathbb{P}(\text{Vraisemblance}_\theta)$ est, comme on l'a vu, l'ensemble des λ_{ij}. Pour simplifier le problème, remarquons que λ_{ij} est l'espérance de buts de i contre j, qu'on imagine dépendant des capacités offensives de i, et des propriétés défensives de j, d'où la définition suivante :

Définition 6 *On pose $\theta = ((N_i),(D_i))$, et $\forall(i,j), \lambda_{ij} = N_i D_j$. Pour une équipe i, on appellera N_i niveau offensif de l'équipe i et D_i perméabilité défensive de l'équipe i. N et D seront les vecteurs généraux offensif et défensif. Par ailleurs, on notera par la suite m_{ij} le nombre de matchs entre i et j, $x_i = \sum_{\text{match de i}} x_{tij}$ et $y_i = \sum_{\text{match de i}} x_{tji}$ les nombres de buts marqués et encaissés par i.*

Proposition 2 $L_0(N,D) = \sum_{\text{match}} x_{tij} \ln(N_i D_j) - \ln x_{tij}! - N_i D_j + x_{tji} \ln(N_j D_i) - \ln x_{tji}! - N_j D_i$

Pour démontrer ce résultat, il suffit de se souvenir du fait que les X_{ij} sont indépendantes. Ainsi $\mathbb{P}(\text{Vraisemblance}_\theta) = \prod \mathbb{P}(X_{N_i D_j} = x_{tij})\, \mathbb{P}(X_{N_j D_i} = x_{tji}) = \prod \frac{(N_i D_j)^{x_{tij}}}{x_{tij}!} e^{-N_i D_j} \frac{(N_j D_i)^{x_{tji}}}{x_{tji}!} e^{-N_j D_i}$
d'où le résultat en passant au logarithme. La log-vraisemblance est donc de classe $\mathcal{C}^\infty$.

Proposition 3 *Soit $\mathcal{R}$ la relation d'équivalence sur l'ensemble des paramètres définie par $(N,D)\mathcal{R}(N',D') \iff \exists \mu/(N,D) = (\mu N', \frac{1}{\mu} D')$. Alors $(N,D)\mathcal{R}(N',D') \Longrightarrow L_0(N,D) = L_0(N,D)$.*

Ainsi, si $(N_0,D_0) \in L_0^{-1}(\sup L_0)$, alors $\{(N,D) |\ (N,D)\mathcal{R}(N_0,D_0)\} = \{(\mu N_0, \frac{1}{\mu} D_0)\} \subset L_0^{-1}(\sup L_0)$. On est alors amené à de nouvelles définitions réduisant l'ensemble général des solutions :

Définition 7 *Soit i_0 une équipe. On fixe son niveau offensif n_{i_0} dans $\mathbb{R}_+^*$. On appelle alors log-vraisemblance réduite l'application $L : (n,d) = ((n_i)_{i \neq i_0}, (d_i)) \longmapsto L_0((n_i),(d_i))$.*
n sera alors un vecteur offensif réduit, et $L^{-1}(\sup L)$ l'ensemble réduit des solutions.

Par la suite, on travaillera principalement sur ces objets réduits, donc on ne mentionnera plus l'adjectif "réduit".

Proposition 4 $\frac{\partial L_0}{\partial n_i} = \frac{\partial L}{\partial n_i} = \frac{x_i}{n_i} - \sum_{\text{quipe } j} m_{ij} d_j$ *et* $\frac{\partial L_0}{\partial d_i} = \frac{\partial L}{\partial d_i} = \frac{x_i}{d_i} - \sum_{\text{quipe } j} m_{ij} n_j$

2.3 Existence et pseudo-unicité des solutions

Définition 8 *Soit $\mathcal{G} = \big(\{(i,1)\} \cup \{(i,2)\}, A\big)$ où $A = \{((i,1),(j,2)) | \exists t/i$ a joué contre j à la date t et $x_{tij} \neq 0\}$. On appellera le $\mathcal{G}$ le graphe des attaques.*

En fait, deux points $(i,1)$ et $(j,2)$ de $\mathcal{G}$ sont liés si i a marqué contre j.

Proposition 5 *Si $\mathcal{G}$ est connexe, alors l'ensemble général des solutions est non vide.*

Pour démontrer ce résultat, on peut d'abord remarquer que l'ensemble général des solutions est non vide si et seulement si l'ensemble des solutions réduit est non vide, d'après la proposition 3. De plus on pourra remarquer que la log-vraisemblance est inférieure à, pour un match donné entre i et j à la date t, $\ln \mathbb{P}(X_{n_i d_j} = x_{ijt})$, qui est une fonction concave de $n_i d_j$ tendant vers $-\infty$ en 0 et en $+\infty$ si $x_{ijt} \neq 0$. Comme n_{i_0} est fixé, et qu'on dispose d'un chemin de i_0 à i pour tout i, par propagation, on montre que l'ensemble des points tels que la log-vraisemblance est plus grande qu'une valeur fixée est un compact, et le maximum de la log-vraisemblance est alors atteint. La démonstration complète figure en annexe.

On aimerait maintenant garantir que $\text{Card}\, L_0^{-1}(\sup L_0) = 1$. Cependant, la proposition 3 anéantit tout espoir d'un tel résultat. Elle incite toutefois à considérer l'espace des paramètres quotienté par $\mathcal{R}$ (voir proposition 3), ou plus exactement $\text{Card}\, \pi(L_0^{-1}(\sup L_0))$ où π est la projection sur l'espace quotient. Or l'application qui à (N,D) associe $(\frac{n_{i_0}}{N_{i_0}}(N_i)_{i \neq i_0}, \frac{N_{i_0}}{n_{i_0}} D)$ est en fait une projection, qui à tout paramètre associe le représentant de sa projection sur l'espace quotient ayant n_{i_0} pour niveau offensif, en enlevant ce terme. On a alors $\pi(L_0^{-1}(\sup L_0)) = L^{-1}(\sup L)$. On parlera alors de pseudo-unicité des solutions lorsque le cardinal de ces ensembles est inférieur ou égal à 1. Malheureusement, je n'ai pas réussi à dégager des conditions simples et suffisantes à cette unicité. Par ailleurs, on aimerait même garantir que $dL^{-1}(0)$ est réduit à un élément, ce qui permettrait d'assurer que notre algorithme conduise à l'unique élément de $L^{-1}(\sup L)$ (voir proposition 8). On a tout de même le résultat suivant :

Proposition 6 *Si f admet un unique point fixe, alors il y a pseudo-unicité.*

La fonction f sera définie à la définition 9. Pour démontrer cette proposition, il faut remarquer d'une part que $L^{-1}(\sup L) \subset dL^{-1}(0)$, et d'autre part que $dL^{-1}(0) = (f - Id)^{-1}(0)$. La première inclusion est triviale, et l'inclusion réciproque découle du fait que f maximise L selon les axes, et que selon les axes, L est concave.

Essayons maintenant de justifier les conjectures 1 et 2. On fixe des paramètres $((n_i)_{i \neq i_0}, d)$ raisonnables, et on prend un grand nombre de matchs. On suppose en plus le nombre d'équipes assez grand, et leurs niveaux offensifs suffisamment homogènes. Par ailleurs, on peut imaginer une répartition aléatoire des matchs, donc si le nombre de matchs est grand, $\forall i \neq j$, le nombre de match entre i et j est presque une constante indépendante de i et j.

On divise alors la log-vraisemblance par cette constante. Par ailleurs, en appliquant la loi des grands nombres, on peut supposer que le nombre de buts marqués par i contre j est l'espérance de la loi de probabilité associée, c'est-à-dire $n_i d_j$.

On remarque que $(n, d) \in dL^{-1}(0)$. Or d'après nos hypothèses, à i fixé, on peut considérer que, lorsqu'on fait une somme sur les $j \neq i$, celle-ci est à peu près égale à la somme sur tous les j. On a alors $(f_n \circ f_d(\widehat{n}))_i \approx \frac{n_i}{\sum n_j} \sum \widehat{n_j}$ Ainsi, en appliquant l'inégalité des accroissements finis à la norme subordonnée à la norme 1 $(\sum |\widehat{n_j}| + \sum |\widehat{d_j}|)$, on voit que $f_n \circ f_d$ est contractante sur l'espace des vecteurs offensifs, donc, d'après le théorème du point fixe, cette fonction admet un unique point fixe. Ainsi, f aussi, ce qui permet d'appliquer la proposition 6, et de vérifier que (n, d) est l'unique solution réduite.

3 Calcul numérique

3.1 Fonction d'itération

Dans cette partie, on ne cherchera pas à déterminer un élément de l'ensemble des solutions, dont la recherche est délicate. On se contentera alors de trouver un maximum local de la log-vraisemblance. On introduit alors des fonctions d'itéraions.

Définition 9 *Soient* $f_n : (d_i) \longmapsto (\frac{y_i}{\sum m_{ij} d_j})_{i \neq i_0}$ *et* $f_d : (n_i)_{i \neq i_0} \longmapsto (\frac{x_i}{\sum m_{ij} n_j})$. f_n *et* f_d *seront appelés fonctions d'itération des attaques et des défenses.*
$f : (n, d) = ((n_i)_{i \neq i_0}, (d_i)_i) \longmapsto (f_n \circ f_d(n), f_d(n))$ *sera la fonction d'itération globale, qu'on appellera aussi fonction d'itération.*

Proposition 7 $\forall (n, d), (L(f^k(n, d)))_{k \in \mathbb{N}}$ *est une suite croissante.*

En effet, comme la log-vraisemblance est concave selon les axes, et comme f_n et f_d associent à d et n l'élément complémentaire annulant les dérivées partielles par rapport à n et d, f est en fait une maximisation par rapport à d, puis à n.

Proposition 8 *Supposons que le graphe des attaques soit connexe et* $L_{|dL^{-1}(0)}$ *injective.*
Alors $\forall (n, d), (f^k(n, d))_{k \in \mathbb{N}}$ *converge vers un point de* $dL^{-1}(0)$.

Comme le graphe des attaques est connexe, la suite considérée vit alors dans un compact (voir démonstration de la proposition 5), et a au moins une valeur d'adhérence. On peut alors montrer que toutes valeurs d'adhérence est alors un élément de $dL^{-1}(0)$, car sinon l'image par f de la suite extraite convergerait vers une autre valeur, pour laquelle la log-vraisemblance serait strictement plus grande. Or $(L(f^k(n, d)))_{k \in \mathbb{N}}$ est une suite croissante, donc convergente, et puisque $L_{|dL^{-1}(0)}$ est injective, la suite n'admet qu'une seule valeur d'adhérence. Comme elle vit dans un compact, elle converge vers son unique valeur d'adhérence, qui est donc dans $dL^{-1}(0)$. La démonstration complète figure en annexe.

3.2 Etude de la hessienne

Proposition 9 *On considère p équipes. Soit $H(n, d)$ la hessienne de L en (n, d). Alors* $H(n, d) =$

$$\left(\begin{array}{ccc|ccc}
-\frac{x_2}{n_2^2} & & 0 & 0 & & -m_{i-p+1,j} \\
 & \ddots & & & \ddots & \\
0 & & -\frac{x_p}{n_p^2} & -m_{i-p+1,j} & & 0 \\
\hline
0 & & -m_{i,j-p+1} & -\frac{y_1}{d_1^2} & & 0 \\
 & \ddots & & & \ddots & \\
-m_{i,j-p+1} & & 0 & 0 & & -\frac{y_p}{d_p^2}
\end{array}\right)$$

La démonstration de ce résultat découle immédiatement de la proposition 8. Pour étudier la hessienne, voyons deux résultats sur les matrices symétriques :

Proposition 10 *Soit M une matrice symétrique réelle positive. Soit x_0 un vecteur pris "aléatoirement". On définit alors la suite x_n par $x_{n+1} = \dfrac{Mx_n}{||Mx_n||_2}$. Alors (x_n) converge vers un vecteur propre de M de norme 1, dont la valeur propre associée est la valeur propre maximale de M.*

Proposition 11 *Soit M une matrice symétrique, $(l_1, l_2 \dots l_k)$ ses lignes. Soit $R = \max\limits_{1 \leq i \leq k} ||l_i||_\infty$. Alors $\mathrm{Sp}\, M \subset [-R, R]$.*

3.3 Description de l'algorithme

On a maintenant tout ce qu'il faut pour un calcul numérique des paramètres. On se donne un ensemble de p équipes et de scores. On supposera alors le graphe des attaques connexes, ainsi que $L_{|dL^{-1}(0)}$ est injective. A défaut de pouvoir déterminer un maximum global de L, on cherchera à avoir un maximum local.

Pour cela, on se donne $\epsilon > 0$, $\alpha > 0$, une valeur fixée de n_{i_0} et une valeur (n^0, d^0), raisonnable. On itère alors la fonction d'itération à ce vecteur. On arrête lorsqu'on atteint un point (n^1, d^1) vérifiant $|||dL(n^1, d^1)|||_\infty < \epsilon$, dont l'existence découle de la proposition 8.

On cherche ensuite à voir si $H(n^1, d^1)$ est définie négative. Pour cela, on pose $A = H(n^1, d^1) + RI_{2p-1}$, où R est la valeur vue à la proposition 11. A est alors forcément positive d'après la proposition 11. On applique alors la proposition 10, en prenant un vecteur x_0 aléatoirement et en calculant (x_k) jusqu'à ce que $||x_{k+1} - x_k||_2 < \epsilon$. On obtient alors la valeur propre maximale de A, égale à $\dfrac{||Ax_k||_2}{||x_k||_2}$. La plus grande valeur propre de $H(n^1, d^1)$ est alors $\beta = \dfrac{||Ax_k||_2}{||x_k||_2} - R$. Or $L((n,d) + \alpha x_k) = L(n,d) +^t (\alpha x_k) H(n'^1, d^1)(\alpha x_k) + o(\alpha^2) = L(n,d) + \beta\alpha^2 + o(\alpha^2)$. Ainsi le terme correctif du premier ordre s'écrit $\beta\alpha^2$. Pour qu'il soit significatif, on se contentera de le prendre supérieur à 0.01. En prenant $\alpha = 0.1$, on compare alors β à 1 :

Si $\beta > 1$, alors, selon la droite passant par (n^1, d^1) de direction x_k, L est localement convexe autour de (n^1, d^1). On peut alors espérer maximiser L en prenant les points $(n^1, d^1) \pm \alpha x_k$. Si en ces points, L est plus grande on réapplique alors récursivement l'algorithme décrit avec ces points pour point de départ, et on prend des deux valeurs obtenues celle pour laquelle L est maximale.
Sinon ou si $\beta \leq 1$, alors on admet qu'on a presque atteint un maximum local. Ainsi on retourne (n^1, d^1).

En fixant $\epsilon = 10^{-5}, n_{i_0} = 1$, $\alpha = 0.1$, $n_i^0 = 1$ et $d_i^0 = 1$ pour tout i, on définit alors la fonction de calcul de paramètres en rajoutant la valeur n_{i_0} pour obtenir un élément de K.

4 Conclusion

En conclusion, notre étude a permis de dégager une loi probabiliste tout à fait cohérente et en accord avec les constatations expérimentales. En outre, malgré de grosses difficultés théoriques pour le définir, le modèle statistique semble parfaitement convenir pour des tests expérimentaux, comme le montrent nos différentes simulations. Néanmoins, on peut se demander comment affiner la modélisation, notamment dans le calcul des paramètres. On peut en effet se dire que le niveau des équipes varie au cours du temps, auquel cas, on pourrait introduire une pondération de la log-vraisemblance. Ainsi, au lieu de chercher à maximiser la probabilité de la vraisemblance, on peut essayer de trouver le maximum de la fonction :

$$L' : (n,d) \longmapsto \prod_{\text{match}} (\mathbb{P}(X_{n_i d_j} = x_{ijt})\, \mathbb{P}(X_{n_j d_i} = x_{jit}))^{\alpha(t)}$$

où $\alpha : \mathbb{N} \longmapsto \mathbb{R}$ est une loi de pondération, prenant de grandes valeurs lorsque le match est récent, ou qu'il est important. Les matchs amicaux peuvent ainsi avoir une importance moindre. Pour finir, on peut remarquer que ce modèle convient également pour la simulation de matchs de rugby, de baskets ou de handball, quoique l'ensemble des valeurs raisonnables est alors différent. Cependant, d'autres sports ou jeux tels que le tennis et le volley doivent être modélisés différemment, par exemple par la loi binomiale. La log-vraisemblance est alors nettement modifiée, ce qui conduirait à une toute autre étude.

5 Annexes

5.1 Tableaux

Equipe	Probabilité de victoire	Classement réel
Portugal	0.20116	Italie
France	0.14497	France
Italie	0.10242	Allemagne
Brésil	0.09100	Portugal
Pays-Bas	0.05882	Ukraine
Espagne	0.05512	Brésil
Suisse	0.05149	Argentine
Argentine	0.04740	Angleterre
Allemagne	0.04401	Australie
Angleterre	0.03667	Espagne
Pologne	0.02766	Suède
Mexique	0.02147	Pays-Bas
Ukraine	0.02053	Suisse
Suède	0.01550	Ghana
Croatie	0.01515	Mexique
Australie	0.01121	Paraguay

Tableau 1 : simulation de la coupe du monde 2006

	Classement réel	Points	Classement simulé	Espérance de points
1	Lyon	79	Lyon	78.4769
2	Lille	67	Lille	69.8886
3	Monaco	63	Monaco	65.3052
4	Rennes	55	Saint-Etienne	62.5159
5	Marseille	55	Rennes	58.2949
6	Saint-Etienne	53	Lens	57.4477
7	Lens	52	Marseille	56.6919
8	Auxerre	52	Auxerre	54.1817
9	PSG	51	Sochaux	53.7110
10	Sochaux	50	Strasbourg	52.3657
11	Strasbourg	48	PSG	52.1002
12	Nice	46	Bordeaux	49.6557
13	Toulouse	46	AC Ajaccio	49.5205
14	AC Ajaccio	45	Nantes	48.2616
15	Bordeaux	44	Nice	47.7746
16	Metz	44	Toulouse	47.5144
17	Nantes	43	Metz	43.8151
18	Caen	42	Bastia	41.1371
19	Bastia	41	Caen	37.4839
20	Istres	32	Istres	33.3504

Tableau 2 : Simulation par points de la ligue 1 2004/2005

	Classement réel	Différence de buts	Classement simulé	Espérance de diff. de buts
1	Lyon	+34	Lyon	+33.8966
2	Lille	+23	Lille	+23.0362
3	Monaco	+17	Monaco	+17.0146
4	Saint-Etienne	+13	Saint-Etienne	+12.8847
5	Rennes	+7	Rennes	+7.1049
6	Lens	+6	Lens	+5.9734
7	Marseille	+5	Marseille	+5.0069
8	Auxerre	+1	Auxerre	+1.1432
9	Sochaux	+1	Sochaux	+0.9963
10	PSG	-1	PSG	-1.0126
11	Strasbourg	-1	Strasbourg	-1.0230
12	AC Ajaccio	-4	Bordeaux	-4.0269
13	Bordeaux	-4	AC Ajaccio	-4.0579
14	Nantes	-5	Nantes	-4.9156
15	Nice	-7	Nice	-7.0813
16	Toulouse	-7	Toulouse	-7.1347
17	Metz	-12	Metz	-11.9339
18	Bastia	-16	Bastia	-15.9923
19	Caen	-24	Caen	-23.8641
20	Istres	-26	Istres	-26.0476

Tableau 3 : Simulation par différences de buts de la ligue 1 2004/2005

Equipe	n_i	$E(\widehat{n}_i)$	$\mathrm{Var}(\widehat{n}_i)$	$\sigma(\widehat{n}_i)$	$\widehat{d}_i$	$E(\widehat{d}_i)$	$\mathrm{Var}(\widehat{d}_i)$	$\sigma(\widehat{d}_i)$
1	1.0	1.0	0.0	0.0	0.7624	0.7672	1.2024	1.0965
2	1.9968	2.1386	0.5844	0.7645	0.6180	0.6224	1.2498	1.1179
3	1.6596	1.7739	1.0009	1.0004	0.9563	0.9637	1.3467	1.1605
4	0.4828	0.5159	1.0544	1.0268	1.1915	1.1961	1.4954	1.2228
5	0.5790	0.6193	1.1349	1.0653	1.4246	1.4230	1.6687	1.2917

Equipe	n_i	$E(\widehat{n}_i)$	$\mathrm{Var}(\widehat{n}_i)$	$\sigma(\widehat{n}_i)$	$\widehat{d}_i$	$E(\widehat{d}_i)$	$\mathrm{Var}(\widehat{d}_i)$	$\sigma(\widehat{d}_i)$
1	1.0	1.0	0.0	0.0	0.7624	0.7641	0.1830	0.4277
2	1.9968	2.0245	0.0842	0.2902	0.6180	0.6220	0.1938	0.4402
3	1.6596	1.6772	0.1452	0.3811	0.9563	0.9589	0.2149	0.4636
4	0.4828	0.4871	0.1549	0.3936	1.1915	1.1949	0.2439	0.4939
5	0.5790	0.5889	0.1674	0.4091	1.4246	1.4202	0.2850	0.5339

Equipe	n_i	$E(\widehat{n}_i)$	$\mathrm{Var}(\widehat{n}_i)$	$\sigma(\widehat{n}_i)$	$\widehat{d}_i$	$E(\widehat{d}_i)$	$\mathrm{Var}(\widehat{d}_i)$	$\sigma(\widehat{d}_i)$
1	1.0	1.0	0.0	0.0	2.9894	3.0146	0.0046	0.0682
2	1.6026	1.6035	0.0006	0.0253	1.1617	1.1561	0.0050	0.0710
3	2.5798	2.6126	0.0021	0.0467	1.9378	1.9300	0.0058	0.0765
4	0.5668	0.5669	0.0023	0.0481	1.0417	1.0372	0.0061	0.0786
5	1.1865	1.1868	0.0027	0.0521	1.0719	1.0672	0.0065	0.0808

Tableau 4,5,6 : Calculs de biais

5.2 Démonstrations complètes

Démonstration de la proposition 5 D'après la renormalisation vue avant la proposition 6, $\mathrm{Im}L_0 = \mathrm{Im}L$. Il vient alors $\sup L_0 = \sup L = M$. On a donc $L_0^{-1}(\sup L_0) \neq \emptyset \iff \exists (N,D)/L_0(N,D) = M \iff \exists (n,d)/L(n,d) = M \iff L^{-1}(\sup L) \neq \emptyset$.

Soit $l \in ImL$, et i une équipe. Puisque $\mathcal{G}$ est connexe et que ses arrêtes ne relient que des éléments de type $(j,1)$ et $(k,2)$, $\exists i_1, i_2 \dots i_p = i/\{((i_0,1),(i_1,2)), \dots ((i_{p-1},2),(i_p,1))\} \subset A$, ce qui signifie que i_0 a marqué contre i_1 dans un match à la date t_0, i_2 contre i_1 à t_1...

Or $L(n,d) = \sum \ln \mathbb{P}(X_{n_i d_j} = x_{tij}) + \ln \mathbb{P}(X_{n_j d_i} = x_{tji}) \leq \ln \mathbb{P}(X_{n_i d_j} = x_{ijt})$ si i a joué contre j à la date t, puisque le logarithme d'une probabilité est toujours négatif. De plus, si $x_{ijt} \neq 0$, $\ln \mathbb{P}(X_{n_i d_j} = x_{ijt}) = x_{tij}\ln(n_i d_j) - \ln(x_{tij}!) - n_i d_j \to 0$ si $n_i d_j$ tend vers 0 ou vers $+\infty$. En particulier, comme $x_{i_0 j_1 t_0} \neq 0, \exists \alpha_1, \beta_1 \in \mathbb{R}_+^*/d_{j_1} \notin [\alpha_1, \beta_1] \Longrightarrow L(n,d) < l$.

Supposons $\exists \alpha_k, \beta_k \in \mathbb{R}_+^*/d_{i_k} \notin [\alpha_k, \beta_k] \Longrightarrow L(n,d) < l$. (1)
Or $x_{i_{k+1} j_k t_k} \neq 0 \Longrightarrow \big(\exists \gamma_k, \delta_k \in \mathbb{R}_+^*/n_{i_{k+1}} d_{i_k} \notin [\gamma_k, \delta_k] \Longrightarrow L(n,d) < l\big)$. (2)
On pose alors $\alpha_{k+1} = \dfrac{\gamma_k}{\beta_k}$ et $\beta_{k+1} = \dfrac{\delta_k}{\alpha_k}$. Supposons $n_{i_{k+1}} < \alpha_{k+1}$.
Si $d_{i_k} \leq \beta_k$, alors $n_{i_{k+1}} d_{j_k} < \gamma_k$, et on est dans le cas (2). Sinon, on est dans le cas (1).
En traitant de même le cas $n_{i_{k+1}} > \beta_{k+1}$, on a $n_{i_{k+1}} \notin [\alpha_{k+1}, \beta_{k+1}] \Longrightarrow L(n,d) < l$.

En combinant ce qu'on a vu, on a $\exists \alpha_i^1, \beta_i^1 \in \mathbb{R}_+^*/n_i \notin [\alpha_i^1, \beta_i^1] \Longrightarrow L(n,d) < B$.
Une démonstration similaire permettrait d'affirmer le même résultat pour les d_i. Ainsi :
$\exists \alpha_i^1, \beta_i^1, \alpha_i^2, \beta_i^2/(n_i, d_i) \notin [\alpha_i^1, \beta_i^1] \times [\alpha_i^2, \beta_i^2] \Longrightarrow \forall (n_j)_{j \notin \{i_0, i\}}, (d_j)_{j \neq \{i\}}, L'(n', d) < B$.

Ainsi si $F = \prod [\alpha_i^1, \beta_i^1] \times \prod [\alpha_i^2, \beta_i^2]$, alors $\sup L = \sup\limits_F L$. Or s'il y a un nombre fini d'équipes, F est un produit cartésien fini de compacts, donc F est compact. Ainsi $\exists (n,d) \in F/L(n,d) = \sup L$, d'où $(n,d) \in L^{-1}(\sup L)$, qui est donc non vide, et $L_0^{-1}(\sup L_0) \neq \emptyset$.

Démonstration de la proposition 8 Soit A l'ensemble des valeurs d'adhérence de $(f^k(n,d))_{k \in \mathbb{N}}$. Comme $(L(f^k(n,d)))_{k \in \mathbb{N}}$ est croissante, $\forall k, f^k(n,d) \in L^{-1}([L(n,d), \sup L])$, qui en tant qu'image réciproque d'un fermé de $\mathbb{R}$ est fermé dans l'espace des paramètres. Or comme le graphe des attaques est connexe, et comme on l'a vu dans la démonstration de la proposition 5, $L^{-1}([L(n,d)), \sup L])$ est borné, donc compact. Ainsi A est non vide. Comme L est continue, l'ensemble des valeurs d'adhérence de $(L(f^k(n,d)))$ est $L(A)$. Or $(L(f^k(n,d)))$ est croissante majorée, donc converge, donc $\exists l/L(A) = \{l\}$.

Par construction de f, $\forall i, \forall k \in \mathbb{N}^*$, $\dfrac{\partial L}{\partial n_i}(f^k(n,d)) = 0$. On fixe i.
Soit $(\widehat{n}, \widehat{d}) \in A$. Alors on peut trouver $\phi : \mathbb{N} \longmapsto \mathbb{N}$ strictement croissante tel que $(n_k, d_k) = f^{\phi(k)}(n,d) \to (\widehat{n}, \widehat{d})$. On pose $(f_d(\widehat{n_k}))_i = d_{ki}^*$, et $\Lambda_k : u \to L(n_k, d_k)$ avec u à la place de d_{ki}, et de même Λ, en remplaçant les n_k et d_k par $\widehat{n}$ et $\widehat{d}$.
On a alors $|\Lambda'(u) - \Lambda_k'(u)| \leq \sum\limits_j m_{ij} |\widehat{n_j} - n_{kj}| \leq (\mathrm{Card}\, E - 1) \max\limits_{j \in E}(m_{ij}) ||\widehat{n} - n_k||$.
Comme $n_k \to \widehat{n}$, Λ_k' converge alors uniformément vers Λ'.
De plus, on sait que f maximalise L d'abord selon le vecteur défensif, donc en particulier selon d_i. On a donc $L(f(n_k, d_k)) \geq \Lambda_k(d_{ki}^*) = \Lambda_k(d_{ki}) + \displaystyle\int_{d_{ki}}^{d_{ki}^*} \Lambda_k'(u) du$.

Cependant, on a $\left| \displaystyle\int_{d_{ki}}^{d_{ki}^*} \Lambda_k' - \int_{\widehat{d_i}}^{\widehat{d_i^*}} \Lambda' \right| \leq \left| \int_{d_{ki}}^{\widehat{d_i}} |\lambda_k'| \right| + \left| \int_{\widehat{d_i^*}}^{d_{ki}^*} |\lambda_k'| \right| + \left| \int_{\widehat{d_i}}^{\widehat{d_i^*}} |\Lambda_k' - \Lambda'| \right|$.
Soit $\epsilon > 0$. Comme $d_{ki} \to \widehat{d_i}, d_{ki}^* \to \widehat{d_i^*}$ et Λ_k' converge uniformément vers Λ, $\exists K \in \mathbb{N}$ tel que pour tout $k \geq K$, $|d_{ki} - \widehat{d_i}| \leq \epsilon, |d_{ki}^* - \widehat{d_i^*}| \leq \epsilon$ et $||\Lambda_k' - \Lambda'||_\infty \leq \epsilon$. Ainsi, sur $[\widehat{d_i} - \epsilon, \widehat{d_i} + \epsilon] \cup [\widehat{d_i^*} - \epsilon, \widehat{d_i^*} + \epsilon], \sup |\Lambda_k'| \leq \sup |\Lambda'| + \epsilon = M$. On a alors pour $k \geq K$,

$$\left| \int_{d_{ki}}^{d_{ki}^*} \Lambda_k' - \int_{\widehat{d_i}}^{\widehat{d_i^*}} \Lambda' \right| \leq \int_{\widehat{d_i} - \epsilon}^{\widehat{d_i} + \epsilon} M + \int_{\widehat{d_i^*} - \epsilon}^{\widehat{d_i^*} + \epsilon} M + \left| \int_{\widehat{d_i}}^{\widehat{d_i^*}} \epsilon \right| \leq (4M + |\widehat{d_i^*} - \widehat{d_i}|)\epsilon.$$

Comme $\Lambda_k(d_{ki}) = L(n_k, d_k) \to l$, on a alors $\Lambda_k(d_{ki}^*) \to l + \displaystyle\int_{\widehat{d_i}}^{\widehat{d_i^*}} \Lambda'$. Par définition de f, $\Lambda'(\widehat{d_i^*}) = 0$, donc si $\Lambda'(\widehat{d_i}) \neq 0$, on a alors $\widehat{d_i} \neq \widehat{d_i^*}$. Rappelons que Λ est concave (voir proposition 7), ainsi on vérifie aisément par disjonction de cas sur le signe de $\Lambda'(\widehat{d_i})$ que $\displaystyle\int_{\widehat{d_i}}^{\widehat{d_i^*}} \Lambda' > 0$. Ainsi $\exists K'/k \geq K' \Longrightarrow L(f(n_k, d_k)) \geq \Lambda_k(d_{ki}^*) \geq l + \dfrac{1}{2}\displaystyle\int_{\widehat{d_i}}^{\widehat{d_i^*}} \Lambda'$, et on a alors $L(f(n_k, d_k)) > l$ alors que c'est une suite croissante convergeant vers l.

Ainsi, on a $A \subset L^{-1}(l) \cap dL^{-1}(0) = (L_{|dL^{-1}(0)})^{-1}(l)$, donc, comme $L_{|dL^{-1}(0)}$ est injective, A possède au plus un élément, or il est non vide, donc $(f^k(n, d))_{k\in\mathbb{N}}$ a une unique valeur d'adhérence et vit dans un compact, donc $(f^k(n, d))_{k\in\mathbb{N}}$ converge.