

# Quels sont les meilleurs pronostiqueurs ? Comment le savoir ?

## I- Analyse probabiliste

### 1- Avec l'intervalle de fluctuation de seconde :

Comme mentionné dans une section du précédent article, si on considère qu'on a une chance sur trois (en prenant les éventualités Gagné, Nul, Perdu comme équiprobables) de pronostiquer correctement le résultat

(G,N,P) d'un match, on peut utiliser l'intervalle de fluctuation vu en seconde  $[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}]$  avec ici  $n=36$

le nombre de matchs pronostiqués et  $p=1/3$  ce qui donne entre  $\frac{1}{3} - \frac{1}{\sqrt{36}} = \frac{1}{6}$  et  $\frac{1}{3} + \frac{1}{\sqrt{36}} = \frac{1}{2}$  comme

bornes de l'intervalle de fluctuation à 95 %, c.a.d que si on répond au hasard, on a environ 95 % de chance d'avoir entre 6 et 18 bons résultats parmi les 36 pronostiqués.

Or, parmi les 67 grilles collectées pour cette expérience, deux seulement sont en dehors (au-dessus) de cet intervalle (pour des candidats extérieurs au lycée).

Vu ainsi, les résultats des pronostiqueurs ne semblent guère brillant, changeons de point de vue, le score moyen est d'environ 14,6 et le score médian de 15 assez largement au-dessus de la moyenne théorique de 12 bons pronostics.

### 2-Avec la loi binomiale (première)

En reprenant les hypothèses précédentes, le nombre de bons résultats (G,N,P) pronostiqués lors des 36 matchs peut-être comparé à la loi binomiale  $B(36;1/3)$  qui correspond au nombre de bons résultats lors de la répétition de manière indépendante, à 36 reprises, d'une épreuve à deux issues dont la probabilité de succès est  $1/3$ .

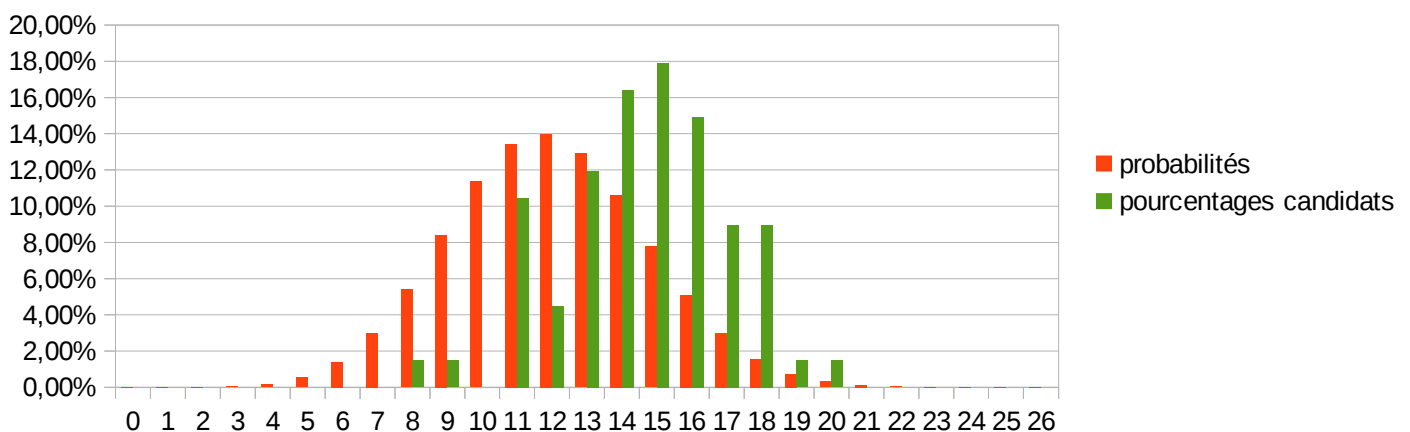
On peut alors calculer son intervalle de fluctuation à 95 % qui est légèrement plus petit que celui de seconde, donnant ici entre 7 et 18 bons résultats.

On observe une fois de plus que seuls deux pronostiqueurs sortent de cet intervalle par le haut.

C'est très peu, puisqu'il n'est censé contenir que 95 % des valeurs, sur 67 candidats, il est donc conforme au hasard, d'en observer deux en dehors de l'intervalle de fluctuation !

Si on compare les répartitions des nombres de résultats trouvés par les pronostiqueurs à la distribution de la loi binomiale, on peut être un peu plus optimiste, le diagramme des pronostiqueurs est nettement décalé vers la droite : Si très peu de pronostiqueurs s'éloignent significativement d'un résultat obtenu totalement au hasard, le groupe des pronostiqueurs pris comme ensemble réalise des scores assez nettement supérieurs.

Nombre de bons résultats (G,N,P) comparé à la loi  $B(36;1/3)$



## II-Comparaison statistique avec le simulateur aléatoire

Les 67 pronostiqueurs ont-ils fait mieux que le simulateur aléatoire élaboré dans l'article précédent (et qui ne tenait compte que des statistiques des nombres de buts marqués lors des éditions précédentes et du classement FIFA (avec bonus pour le pays hôte)) ?

### 1- Le protocole

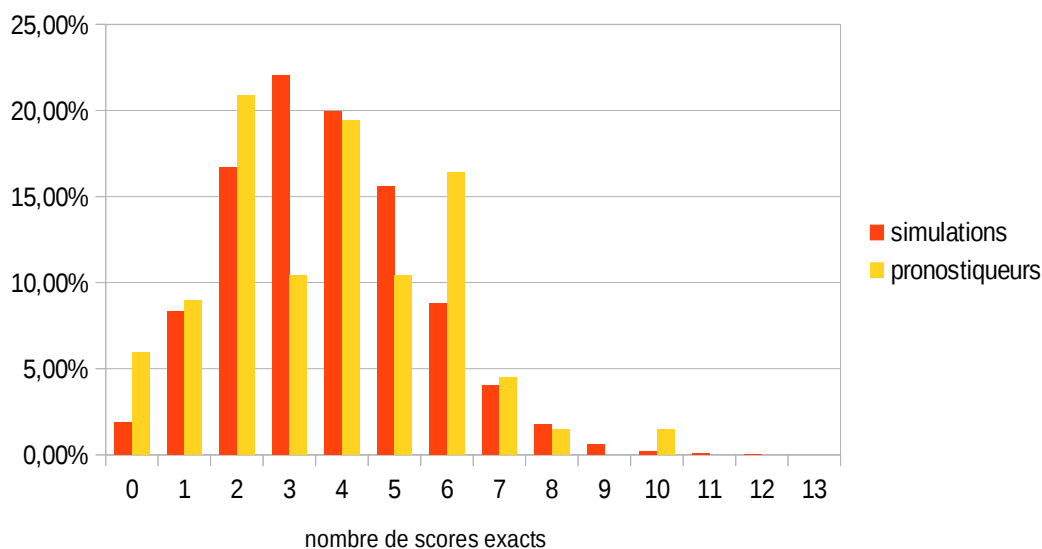
Pour cela j'ai réalisé 2880 simulations, en retenant comme classement, celui obtenu en appliquant la différence de but globale et non la différence de but particulière qui figurait dans le règlement mais dont la programmation est trop complexe. Un seul groupe voit son classement modifié, le groupe E où l'Italie prend la première place du groupe (ayant battu la Belgique avec laquelle elle est ex-aequo) alors que le fichier distribué donne la Belgique première en raison d'une meilleure attaque (1 but de plus, même différence de but).

Quelques modifications peuvent également avoir affecté vos classements (mais vous aviez rempli vos grilles avec ce classement) en raison de l'application de la différence de but particulière, mais aussi en cas d'égalité parfaite (point, différence de but globale, attaque) entre deux équipes (ou plus et là même le fichier modifié ne fonctionne plus) ; le règlement prévoyant de recourir aux classements du fair-play (puis FIFA) que je vous n'aviez pas à pronostiquer, le fichier désigne comme premier l'équipe inscrite en premier dans le groupe.

### 2- Nombre de scores exacts

Les paris sportifs portent souvent sur le score exact d'un match, et dans le score total des pronostiqueurs qui sera étudié plus loin, un score exact rapportait 5points.

Comparaison entre les nombres de scores exacts des pronostiqueurs et des simulations



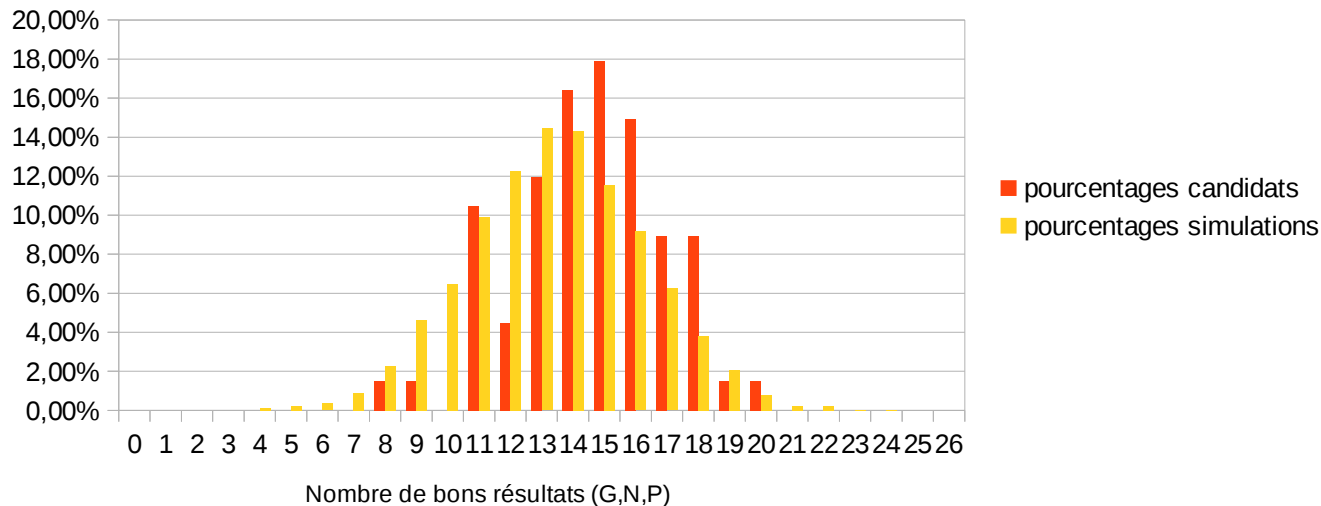
Visuellement, difficile de dire qui des pronostiqueurs ou du simulateur a réussi les meilleures performances, le simulateur (sur 2880 simulations) s'avère plus régulier que les pronostiqueurs (67 seulement), mais les graphiques semblent proches ce que confirme des scores moyens égaux (à 0,01près):3,695 pour le simulateur contre 3,687 pour les candidats !

Un candidat (extérieur au lycée, le grand gagnant du concours, on le verra plus loin) se distingue avec 10 scores exacts, le meilleur total parmi les 2880 simulations étant 12.

### 3- Nombres de bons pronostics

Un bon pronostic rapportait 2points dans le score total de pronostiqueur, et l'on a vu plus haut que les pronostiqueurs avaient fait un peu mieux que le hasard complet (une chance sur trois pour G,N,P). En-est-il de même en comparaison des 2880 simulations ?

Comparaison entre les nombres de bons résultats (G,N,P)  
des candidats et des simulations



La réponse est cette fois positive, le décalage qu'il y avait entre les résultats des pronostiqueurs et les probabilités de la loi binomiale  $B(36;1/3)$  est moins net entre les pronostiqueurs et le simulateur aléatoire, mais il reste visible et le nombre moyen de bon résultats vient le confirmer : 14,6 pour les pronostiqueurs contre 13,4 pour le simulateur aléatoire.

### 4-Analyse globale des scores totaux

Si on ajoute trois points par équipe correctement classée et trois points par équipe qualifiée, on obtient le score total des pronostiqueurs, lequel peut varier selon si on applique la différence de but générale, ou la différence de buts particulière pour les ex-aequo aux points (et plus rarement selon le tirage au sort pour les ex-aequo « parfaits »), une correction manuelle peut être faite pour les 67 pronostiqueurs, mais pas pour les 2880 simulations, l'analyse se basera donc sur l'application de la différence de buts générale.

La médiane pour ceux qui voudraient se situer par rapports aux autres candidats est de 107 avec des quartiles à 87 et 117, en d'autres termes, vous êtes dans la première moitié du classement si vous avez plus de 107 points et dans le quart supérieur si vous avez plus de 117 points.

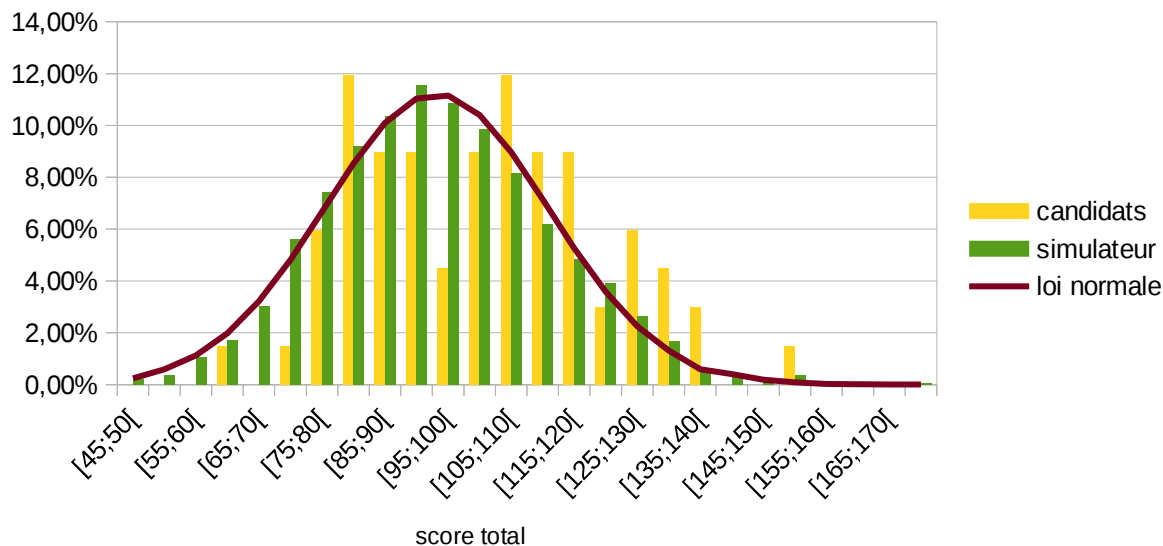
Le score moyen des 67 pronostiqueurs est de 102,8 avec un écart-type de 18,4.

Le score moyen sur 2880 simulations est de 95,1 avec un écart-type de 17,7.

On peut donc, comme on l'a déjà observé dans la partie probabilité, remarquer que globalement les pronostiqueurs ont réalisé de meilleurs scores que le simulateur dont une des simulation obtient le meilleur score:172pts (contre 150 points pour le meilleur pronostiqueur), et une autre le plus mauvais: 46pts (contre 62points pour le plus mauvais pronostiqueurs, lequel a joué avec un système qui pouvait a priori paraître intéressant : ne mettre que des scores de 2-0, il y en a eu très peu et plus de nuls que prévus).

Plus globalement le grand nombre de simulations donne une courbe proche de la distribution d'une loi normale de mêmes paramètres : La loi  $N(95,1 ; 17,7^2)$ . On peut alors comparer les résultats des pronostiqueurs, ceux du simulateur et cette loi normale sur un même graphique (après avoir regroupé les résultats par classe, vu le faible nombre de pronostiqueurs) qui confirme l'impression générale : Des résultats globalement meilleurs que le simulateur.

Comparaison des scores des candidats et du simulateur, avec une loi normale



### III- Qui mérite le titre de « top pronostiqueur » ?

#### 1-Rappel du protocole

Comme indiqué dans l'article précédent, pourront se prévaloir du titre de « top pronostiqueur » ceux dont le résultat (score total) ne pourra pas « raisonnablement » être imputé au hasard (au sens : « pas meilleur que le générateur aléatoire ») c.a.d ceux dont le score total sort de l'intervalle de fluctuation à 95 % de la loi normale associée aux résultats des simulations. Celui-ci est centré autour de l'espérance (ici 95,1) avec une marge de plus ou moins deux écart-type (ici 17,7) ce qui donne [59,7;130,5].

Heureusement personne n'a obtenu moins de 59,7 points, tous ceux qui ont obtenu 131points ou plus méritent donc d'être cités et félicités.

Ils sont cinq (ce qui fait plus que les 2,5 % qu'on devrait trouver en moyenne au-dessus de l'intervalle de fluctuation si les scores des pronostiqueurs suivaient la loi normale, ce qui confirme une fois de plus que globalement les pronostiqueurs ont fait un peu mieux que le hasard)

#### 2- Comparaison avec des professionnels

Parmi les 67 pronostiqueurs, il en est deux qui ne savent pas qu'ils ont concouru avec nous, deux spécialistes anglais du pronostic sportif : Mark Lawrenson le spécialiste de la BBC et Paul Merson de la chaîne concurrente Sky.

Leurs pronostics étant disponibles sur internet, j'ai choisi de les intégrer à notre concours.

Ils ne sont pas parmi les top pronostiqueurs...

Paul Merson avec 4 scores exacts, 15 bons résultats et 109 points au total fait à peine mieux que le score moyen des pronostiqueurs (dont certains qui ont répondu par amour des maths...) 3,7 scores exacts, 14,6 bons résultats et 102,8 points.

Son confrère de la BBC fait un peu mieux avec 7 scores exacts 15 bons résultats et 116pts (en remplissant une équipe du classement de son dernier groupe un peu au hasard, il n'avait prédit que des matches nuls!)

Ils se classent respectivement 18ième et 25ième sur 67.

### 3- La liste des « Top pronostiqueurs »

1- Thomas B (candidat hors lycée) 153pts (ou 156ps si on applique la Différence de Buts Particulière) dont 10 scores exacts et 20 bons résultats

2- Dimitri Hesnard 137pts (ou 134 avec DBP) dont 5 scores exacts et 17 bons résultats.

3- Leonard Bertrand 135pts (ou 132 avec DBP) dont 8 scores exacts et 16 bons résultats.

4- Audrey Brocard 132pts (ou 129 avec DBP) dont 6 scores exacts et 18 bons résultats.

5- David Lebourg (du collège de Cany-Barville) 131pts (mais seulement 125pts avec la DBP) dont 6 scores exacts et 16 bons résultats.

Hervé B (candidat hors lycée) initialement huitième ex-aequo avec 125 points (dont 6 scores exacts et 19 bons résultats) est le seul candidat à passer au-dessus de 131pts avec l'application de la différence de buts particulière (131pts).

### Conclusion :

La libéralisation en France des paris sportifs, dont les entreprises recrutent d'anciennes gloires sportives et figurent parmi les principaux sponsors du football, a renforcé spectaculairement l'engouement pour les concours de pronostics, avec souvent un enjeu financier.

Ainsi des sites comme l'Equipe ou même le Monde

[http://www.lemonde.fr/les-decodeurs/article/2016/06/03/attaque-defense-experience-les-forces-et-faiblesses-des-24-equipes-de-l-euro-2016-de-football\\_4933442\\_4355770.html](http://www.lemonde.fr/les-decodeurs/article/2016/06/03/attaque-defense-experience-les-forces-et-faiblesses-des-24-equipes-de-l-euro-2016-de-football_4933442_4355770.html)

proposent selon la mode anglo-saxonne des articles sur les chances de voir telle équipe battre telle autre voire de participer à un concours en ligne, en relayant souvent les offres des offices de paris en lignes.

Même les scientifiques (comme l'école polytechnique fédérale de Lausanne <http://www.kickoff.ai/previous-matches> ou ce mathématicien relayé par le Monde [http://www.lemonde.fr/football/article/2015/12/12/euro-2016-comment-le-tableau-final-favorise-la-france-par-julien-guyon-mathematicien\\_4830555\\_1616938.html](http://www.lemonde.fr/football/article/2015/12/12/euro-2016-comment-le-tableau-final-favorise-la-france-par-julien-guyon-mathematicien_4830555_1616938.html))

et les banques <http://mobile2.tdg.ch/articles/57572001ab5c373524000001> y vont de leur étude.

Aussi fines que soient leurs analyses, il ne faut pas perdre de vue qu'un résultat sportif (en particulier au football) reste un événement aléatoire, qui comme tout événement aléatoire peut seulement être plus ou moins bien modélisé, pour essayer d'en quantifier plus ou moins bien l'incertitude, notamment via l'utilisation des intervalles de fluctuations.

Notre modeste expérience (échantillon de 67 personnes) a ainsi permis de souligner trois points précis :

-Trouver un score exact relève essentiellement de la chance (c'est le seul domaine où les pronostiqueurs dans leur ensemble ne font pas mieux que le générateur aléatoire)

-Le pronostiqueurs obtiennent **globalement** des résultats meilleurs que le plus pur hasard (nombres de bons résultats) mais aussi que le générateur aléatoire.

- Très peu de pronostiqueurs obtiennent des résultats significativement meilleurs (au sens de « hors de l'intervalle de fluctuation ») que le hasard (2 sur 67 au niveau des bons résultats) ou que le générateur aléatoire ( 4 ou 5 sur 67 selon la règle appliquée pour les ex-aequo).

Et même pour ceux-là, leur faut-il conclure qu'ils gagneraient de l'argent en pariant sur des sites spécialisés ? Les scores médiocres des deux spécialistes anglais du pronostic sportif devraient les inciter à la modération.