Une expérience de modélisations et simulations :
La phase finale de l'euro 2016 de football

I-La réalisation d'un fichier tableur

A chaque compétition sportive, de nombreux sites proposent des fichiers tableurs complétés pour ceux qui voudraient se livrer à un concours de pronostics.
Leur réalisation fait généralement appel à des compétences allant bien au-delà du B2I et de l'algorithmique de lycée, mais il est néanmoins possible d'en réaliser un en n'utilisant pratiquement que les opérations classiques et les commande SI et OU.
Le fichier ci-joint vous permet, en complétant uniquement les scores des match (cases jaunes) de la feuille 1 pour vos pronostics, puis au fur et à mesure que les résultats sont connus, les scores réels en feuille 2, d'obtenir un score en tant que pronostiqueur pour la phase de poule de l'euro 2016.
Pour cela le barème choisi est le suivant : 5pts par score exact, 2pts par résultat juste (match gagné ou non) et 3pts par équipe bien classée, plus 3pts par équipe qualifiée.

Ainsi en feuille 3 pour savoir si le score pronostiqué est le bon on entre la formule :
=SI(ET(pronostics.B3=résultats.B3;résultats.C3=pronostics.C3);C1;0)
qui teste si on a à la fois le bon nombre de buts pour la première équipe, **et** pour la seconde ; et dans ce cas affiche le nombre de points prévus, 5, mis en mémoire dans la case C1. On notera la présence de $ pour bloquer cette référence, puisqu'on va bien sûr étirer la formule pour les autres matchs.

Pour savoir si l'on a le bon résultat, la formule choisie est :
=SI((résultats.B3-résultats.C3)*(pronostics.B3-pronostics.C3)>0;D1;SI(ET(pronostics.B3- pronostics.C3=0 ;résultats.B3-résultats.C3=0);D1;0))
Elle teste d'abord si la même équipe a gagné dans les pronostics et la réalité, en vérifiant que le produit des différences entre les scores des deux équipes est strictement positif, puis si cette différence est égale à 0 dans la réalité et les pronostics (match nul), et dans ce cas affiche le nombre de points prévus par le barème, stocké en D1.

Pour les points de classements, on peut utiliser
=SI(résultats.P3=pronostics.P3;E1;0)
Elle attribue les points prévus si le classement réel est celui pronostiqué.

Pour les points des qualifiés, cela se complique car si les deux premiers sont qualifiés automatiquement, c'est aussi le cas des quatre meilleurs troisièmes !
J'ai donc opté pour la création d'une liste de qualifiés (qui comme le classement de chaque groupe, puis celui des meilleurs troisièmes, utilisent la commande Rang) pour vos pronostics et pour le classement réel ; liste pour laquelle on teste pronostic par pronostic, si l'équipe figure bien dans la liste des qualifiés réels.
La commande « recherche » permet de faire cela plus simplement mais outre qu'elle n'est pas utilisée au lycée, elle semblait avoir des difficultés pour reconnaître certains pays, j'ai donc opté pour la commande :
=SI(OU(pronostics.S12=résultats.S12;pronostics.S12=résultats.S13;pronostics.S12=résultats.S14;pronostics.S12=résultats.S15;pronostics.S12=résultats.S16;pronostics.S12=résultats.S17;pronostics.S12=résultats.S18;pronostics.S12=résultats.S19;pronostics.S12=résultats.S20;pronostics.S12=résultats.S21;pronostics.S12=résultats.S22;pronostics.S12=résultats.S23;pronostics.S12=résultats.S24;pronostics.S12=résultats.S25;pronostics.S12=résultats.S26;pronostics.S12=résultats.S27);I1;0)

Le score final est donc la somme de tous les points gagnés :
=SOMME(C3:I44)

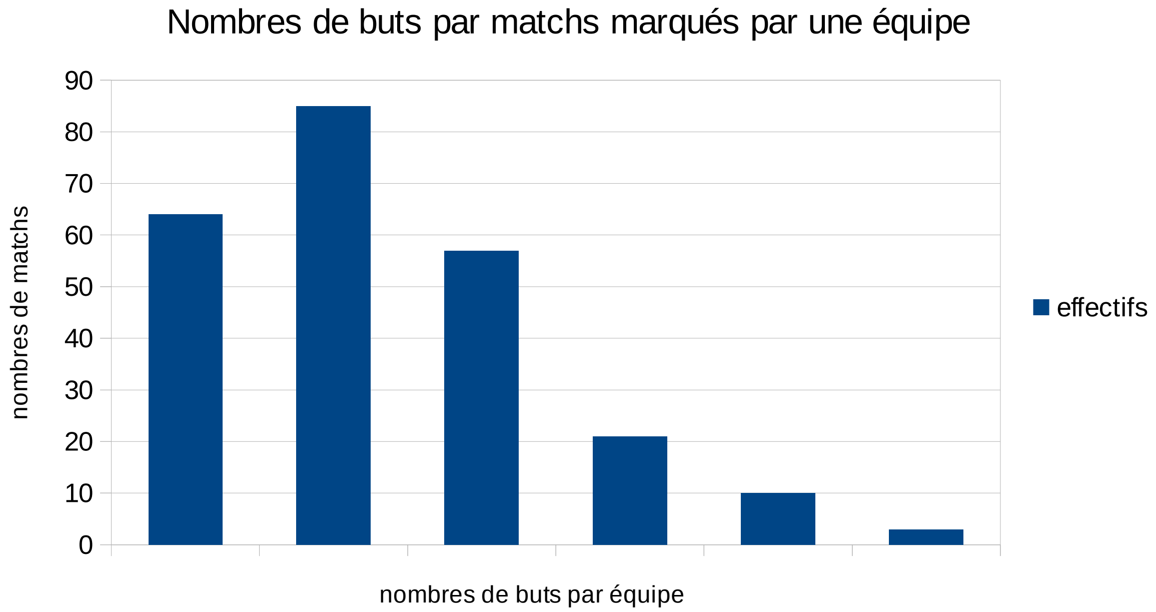
II-Comment savoir si on a un bon score ?

1-Analyse des données pour bâtir un modèle

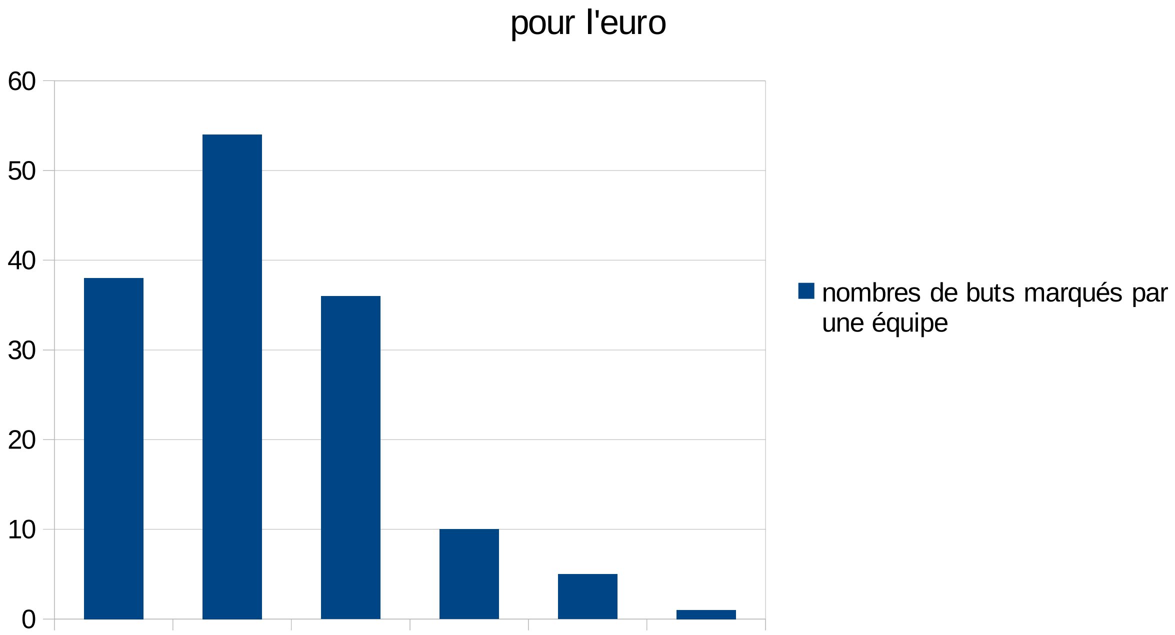
On peut évidemment se comparer aux autres parieurs, mais pour que cela soit significatif il faudrait un grand nombre de parieurs.

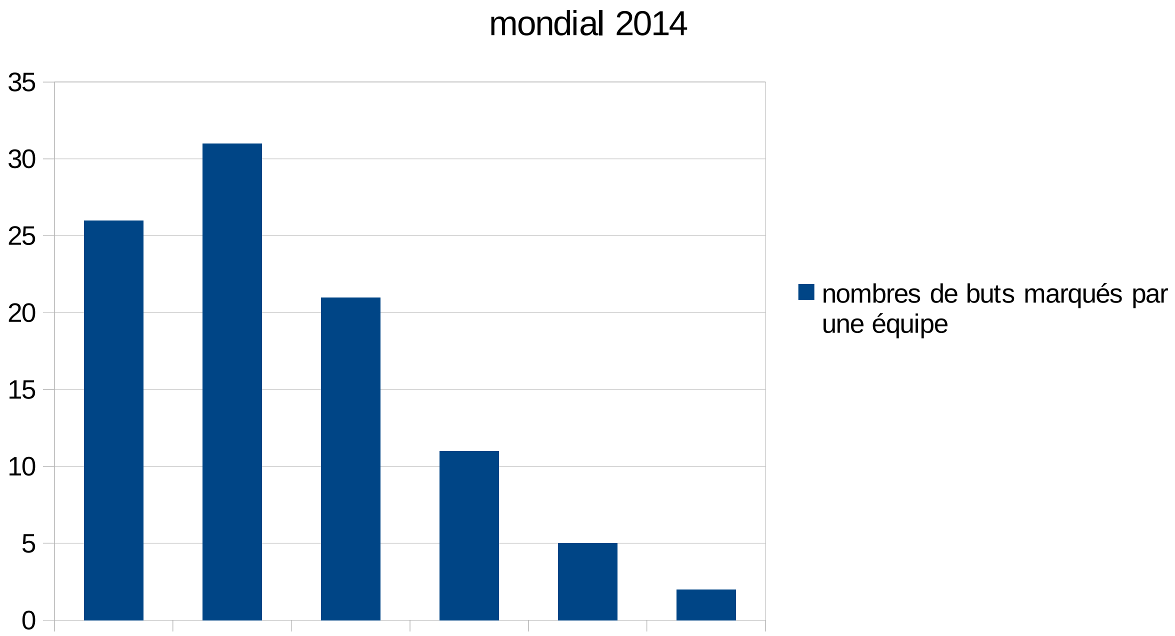
Il est bien plus simple de modéliser la situation puis de réaliser un grand nombre de simulation du modèle mis au point.

Il faut déjà mettre au point un modèle, pour cela on peut analyser les résultats des années précédentes, c.a.d les nombres de buts par équipe et par match marqués lors de compétition similaires : Les phase de poules de la phase finale des euros 2004; 2008 et 2012 plus celle du mondial 2014 qui a l'avantage d'être la base de donnée la plus récente et surtout de présenter un premier tour avec 32 qualifiés, de niveaux plus disparates qu'à l'euro, or l'euro vient de passer d'une formule à 16 clubs pour les éditions précédentes à une formule à 24 clubs, donc avec des équipes de niveaux plus disparates que ce n'était le cas auparavant.



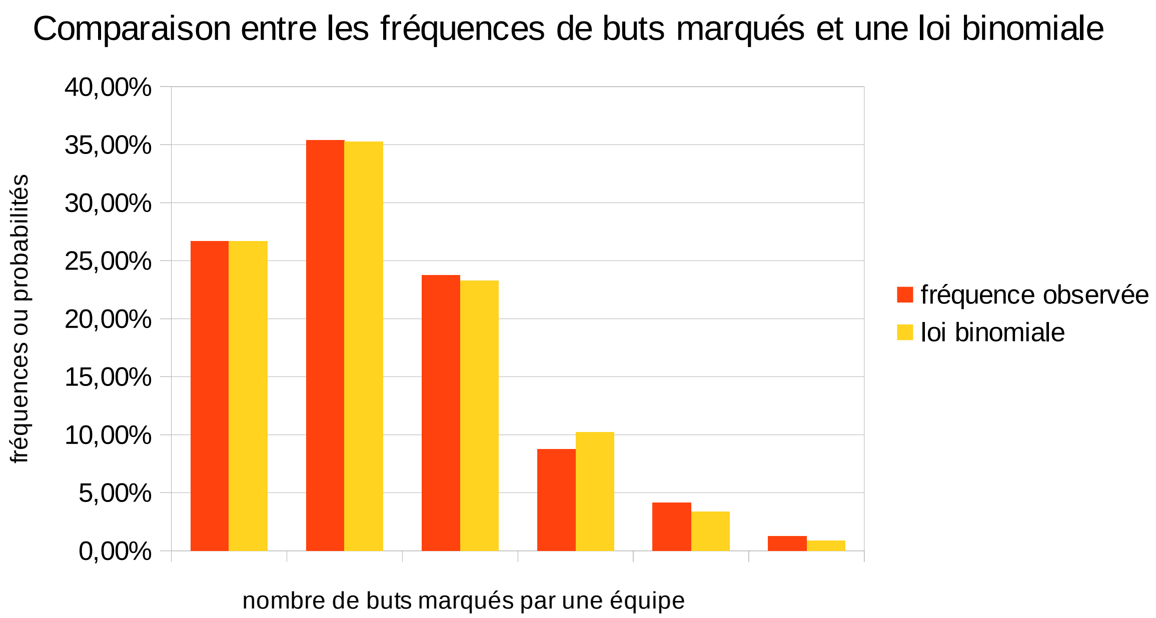
Ci-dessus l'ensemble des données récupérées,ci-dessous, la confirmation de plus petits scores à l'euro.





Le nombre moyen de buts marqués par une équipe est alors de 1,32 avec un écart-type de 1,15. (Un peu supérieurs dans les deux cas pour les données du mondial 2014 par rapport à celles des euros)

Si on observe les diagramme bâton, on observe une forme assez proche de la distribution d'une loi binomiale avec une probabilité faible (ce qui fausse sa quasi-symétrie habituelle), on peut donc envisager de modéliser le nombre de buts par matchs par une loi binomiale B(*n*;*p*).
Pour cela on sait que la moyenne est E=np et la variance (carré de l'écart-type) $V=np(1-p)$ on pourrait donc être satisfait si on avait un résultat proche d'un entier pour n, avec $np=1{,}32$ et $np(1-p)=1{,}15^2=1{,}3225$ mais cela donne $1{,}32(1-p)=1{,}3225$ soit $p\approx 0{,}002$ puis $n\approx 660$.
Une telle modélisation donne la distribution suivante (tronquée à 5 buts pour une équipe, la probabilité étant ensuite faible) :



avec *n*=660 et *p*=0,002

Le résultat est très proche, mais devoir simuler un tirage aléatoire à 660 étapes pour chaque score peut décourager, et va surtout ralentir considérablement les calculs.

On peut préférer créer une loi qui corresponde exactement à la base statistique utilisée (pas très robuste soit dit en passant, trois euros et un mondial, ça ne donne que 240 matchs, il aurait fallu revenir plus loin dans le temps, au risque que le style de jeu développé ayant évolué les statistiques ajoutées ne soient plus pertinentes, songez aux 13 buts de Just Fontaine en 1958, ou aux 9 de Michel Platini en 1984 alors qu'il n'y avait que 8 équipes) :

k	P(X=k)	P(X<=k)
0	26,7%	26,7%
1	35,4%	62,1%
2	23,8%	85,8%
3	8,8%	94,6%
4	4,2%	98,8%
5	1,3%	100,0%

Ainsi on simulera le hasard à l'aide du tableur avec la commande alea() qui donne un nombre au hasard entre 0 et 1, de la façon suivante :

=SI(ALEA()<F9;0;SI(ALEA()<F10;1;SI(ALEA()<F11;2;SI(ALEA()<F12 ;3;SI(ALEA()<F13.H28;4;5)))))

Avec F9 jusqu'à F14 qui correspondent à $P(X \leqslant k)$ pour la loi créée d'après nos statistiques.

On peut ainsi créer des milliers de simulations de pronostics qu'on comparera avec les résultats réels pour obtenir des milliers de scores de pronostiqueurs fictifs, ce qui permettra d'estimer si notre propre score est un bon score ou non.

2-Affinage du modèle retenu :

Bien sûr les bons pronostiqueurs disposent d'informations non évaluées jusqu'à présent, comme la forme de l'équipe, l'absence de joueurs cadres, la difficulté du groupe, les confrontations récentes entre des équipes… Certaines de ses données peuvent s'intégrer à la modélisation, et c'est ce qu'on va tenter de faire sur deux points précis :

1- Les équipes n'ont pas le même niveau, et les groupes ont été construits, selon un système de « chapeau » avec têtes de séries, qui implique que dans un même groupe on retrouve généralement une équipe favorite et une autre supposée faible.

Pour intégrer un peu ce paramètre, on va utiliser les coefficients FIFA de chaque équipe afin de créer pour chaque match un coefficient à appliquer à la probabilité de marquer pour l'augmenter dans le cas de l'équipe la plus forte et pour la diminuer dans le cas de l'équipe la plus faible.

Belgique	1352
Allemagne	1309
Espagne	1277
Portugal	1184
Angleterre	1069
Autriche	1067
Turquie	983
Suisse	974
Italie	959
Hongrie	925
Roumanie	922
France	907
Ukraine	880
Croatie	856
Pays de Galles	839
Irlande du Nord	825
Pologne	821
Russie	821
République Tchèque	810
Eire	792
Slovaquie	784
Islande	724
Suède	713
Albanie	632

Pour chaque match, chaque équipe se voit attribuer un coefficient de minoration/majoration que j'ai choisi de calculer ainsi : On calcule le rapport entre le plus grand et le plus petit des coefficients FIFA des équipes qui s'affrontent : =MIN(E28:F28)/MAX(A28:F28)-1, par exemple pour Albanie-Suisse, le rapport des coefficients est $c=\frac{632}{974}-1\approx-0{,}35$ et ensuite on multiplie les seuils de la loi de probabilité par $1-\frac{c}{2}$ pour l'équipe la plus faible et par $1+\frac{c}{2}$ pour l'équipe la plus forte.

Ainsi, pour Albanie-Suisse, on multiplie les seuils de la loi de probabilité du nombre de buts marqués par 1,176 environ pour l'Albanie, ce nombre rendant moins probable qu'elle marque des buts (les seuils augmentent), et ceux de la suisse par 0,824, rendant plus probable qu'elle marque des buts.

		coefficients FIFA		c	minoration	majoration
Albanie	Suisse	632	974	-0,351129363	1,1755646817	0,8244353183

2- On considère en général que l'équipe qui reçoit dispose d'un avantage, j'ai choisi de matérialisé celui-ci en rajoutant 100pts au total FIFA de la France, ses coefficients sont donc calculés sur une base FIFA de 1007pts.

La formule tableur utilisée pour simuler le nombre de buts marqués dans un match devient alors :
=SI(ALEA()<$Feuille4.$F$9*$Feuille4.H42;0;SI(ALEA()<$Feuille4.$F$10*$Feuille4.H42;1;SI(ALEA()<$Feuille4.$F$11*$Feuille4.H42;2;SI(ALEA()<$Feuille4.$F$12*$Feuille4.H42;3;SI(ALEA()<$Feuille4.$F$13*$Feuille4.H42;4;5)))))
où la feuille 4 contient les coefficient de majoration ou minoration associés à chaque équipe pour chaque match, et la table avec les seuils de la loi de probabilité déjà mentionnée.

On obtient ainsi une liste de matchs avec leur résultats simulés aléatoirement selon cette modélisation.

MATCH																		troisièmes		
Groupe A					Pays	Points	G	N	P	buts marqués	buts encaissés	différence de buts	points fictifs	rang	classement					rang
France	0	1	Roumanie		France	4	1	1	1	2	2	0	40000,0002	3	Roumanie			France	40000,0002	2
Albanie	1	3	Suisse		Albanie	0	0	0	3	1	6	-5	-4,9999	4	Suisse			Russie	30000,0003	3
France	1	0	Albanie		Roumanie	7	2	1	0	5	2	3	70003,0005	1	France	40000,0002		Irlande Du Nord	29999,0003	4
Roumanie	2	2	Suisse		Suisse	5	1	2	0	6	4	2	50002,0006	2	Albanie			Turquie	40000,0003	1
Suisse	1	1	France															Eire	29997,0002	5
Roumanie	2	0	Albanie															Portugal	19999,0004	6
GROUPE B																				
Angleterre	2	2	Russie		Angleterre	3	0	3	0	5	5	0	30000,0005	2	Slovaquie					
Pays De Galles	1	2	Slovaquie		Pays De Galles	2	0	2	1	3	4	-1	19999,0003	4	Angleterre			Qualifiés		
Angleterre	2	2	Pays De Galles		Russie	3	0	3	0	3	3	0	30000,0003	3	Russie	30000,0003		Roumanie		
Russie	1	1	Slovaquie		Slovaquie	5	1	2	0	4	3	1	50001,0004	1	Pays De Galles			Suisse		
Slovaquie	1	1	Angleterre															Slovaquie		
Russie	0	0	Pays De Galles															Angleterre		
GROUPE C																		Ukraine		
Allemagne	0	2	Ukraine		Allemagne	1	0	1	2	1	4	-3	9997,0001	4	Ukraine			Pologne		
Pologne	2	1	Irlande Du Nord		Pologne	5	1	2	0	5	4	1	50001,0005	2	Pologne			Espagne		
Allemagne	1	1	Pologne		Ukraine	7	2	1	0	6	3	3	70003,0006	1	Irlande Du Nord	29999,0003		Croatie		
Ukraine	2	1	Irlande Du Nord		Irlande Du Nord	3	1	0	2	3	4	-1	29999,0003	3	Allemagne			Belgique		
Irlande Du Nord	1	0	Allemagne															Italie		
Ukraine	2	2	Pologne															Autriche		
GROUPE D																		Hongrie		
Espagne	2	1	République Tchèque		Espagne	9	3	0	0	5	2	3	90003,0005	1	Espagne			Turquie		
Turquie	2	2	Croatie		Turquie	4	1	1	1	3	3	0	40000,0003	3	Croatie			France		
Espagne	1	0	Turquie		République Tchèque	0	0	0	3	1	4	-3	-2,9999	4	Turquie	40000,0003		Russie		
République Tchèque	0	1	Croatie		Croatie	4	1	1	1	4	4	0	40000,0004	2	République Tchèque			Irlande Du Nord		
Croatie	1	2	Espagne																	
République Tchèque	0	1	Turquie																	
GROUPE E																				
Belgique	2	2	Italie		Belgique	7	2	1	0	10	3	7	70007,001	1	Belgique					
Eire	1	0	Suède		Eire	3	1	0	2	2	5	-3	29997,0002	3	Italie					
Belgique	4	1	Eire		Italie	7	2	1	0	5	3	2	70002,0005	2	Eire	29997,0002				
Italie	2	1	Suède		Suède	0	0	0	3	1	7	-6	-5,9999	4	Suède					
Suède	0	4	Belgique																	
Italie	1	0	Eire																	
GROUPE F																				
Portugal	0	0	Islande		Portugal	2	0	2	1	4	5	-1	19999,0004	3	Autriche					
Hongrie	1	1	Autriche		Hongrie	5	1	2	0	5	3	2	50002,0005	2	Hongrie					
Portugal	2	2	Hongrie		Islande	1	0	1	2	1	4	-3	9997,0001	4	Portugal	19999,0004				
Islande	1	2	Autriche		Autriche	7	2	1	0	6	4	2	70002,0006	1	Islande					
Autriche	3	2	Portugal																	
Islande	0	2	Hongrie																	

Pour afficher automatiquement le nombre de matchs gagnés, on peut utiliser :
=SI($B3>$C3;1;0)+SI($B5>$C5;1)+SI($C7>$B7;1)
Et pour le nombre de matchs nul :
=SI($B3=$C3;1;0)+SI($B5=$C5;1)+SI($C7=$B7;1)
Le nombre matchs perdus étant égal à trois moins les nombres précédents.
Le nombre de points marqués s'en déduit facilement (=3*H3+I3), et la somme des buts marqués ou encaissés se résume à une somme.

Pour la suite, une simplification est utilisée qui peut conduire à des décalages avec le règlement officiel. En effet celui-ci prévoit d'utiliser la différence de buts particulière pour classer les équipes ayant un même nombre de points, et seulement en cas d'égalité entre celles-ci de recourir à la différence de but générale, or la différence de but particulière s'avère trop compliquée à programmer (le site l'équipe qui propose un concours similaire a le même problème), et elle intervient assez rarement (souvent les équipes qui ont le même nombre de points ont fait match nul entre elles), le classement des équipe se fait donc selon la différence de buts générale.

Pour cela on utilise une astuce en recourant à un système de points fictifs : plutôt que d'essayer de trier les équipes selon le nombre de points, puis en cas d'égalité selon la différence de buts, puis en cas d'égalité par le nombre de buts marqués, on attribue 10000 points (fictifs) par point réellement obtenu, 1 point (fictif) par but en plus dans la différence de but et 0,0001 point fictif par but marqué, ce qui élimine les ex-æquo (sauf ex-aequo parfaits, qu'il aurait fallu classer selon le classement du fair-play, puis en dernier recours le classement FIFA) et classe les équipes dans le bon ordre (sauf si une équipe marque 10000 buts…).

On utilise la commande rang pour classer les équipes de chaque groupe, avec une seconde colonne pour départager d'éventuels ex-aequo parfaits (selon l'ordre d'inscription et non le fair-play) puis le même système est utilisé pour classer les troisièmes et ne conserver que les quatre meilleurs, ce qui permet de compléter la liste des qualifiés.

3-Résultats des simulations :

Le tableur permet alors de réaliser des milliers de fiches « pronostics » de joueurs fictifs utilisant la modélisation précédente.

On compare ensuite leurs grilles de pronostics à celles des résultats réels de l'euro et on obtient autant de scores de pronostiqueurs que l'on a fait de simulations.

Les résultats de l'euro 2016 n'étant pas connus, on peut se faire une idée du score moyen d'un pronostiqueur qui joue au hasard (selon la modélisation précédente), en comparant les résultats des simulations à une grille de résultats fictifs, qu'on peut par exemple simuler selon la loi binomiale B(660;0,002) dont on a vu qu'elle donnait des résultats proches de ceux observés précédemment, ou ici, pour limiter les calculs par la loi B(0,132;10) qui donne la même espérance et un écart-type légèrement inférieur mais permet de limiter les répétitions.

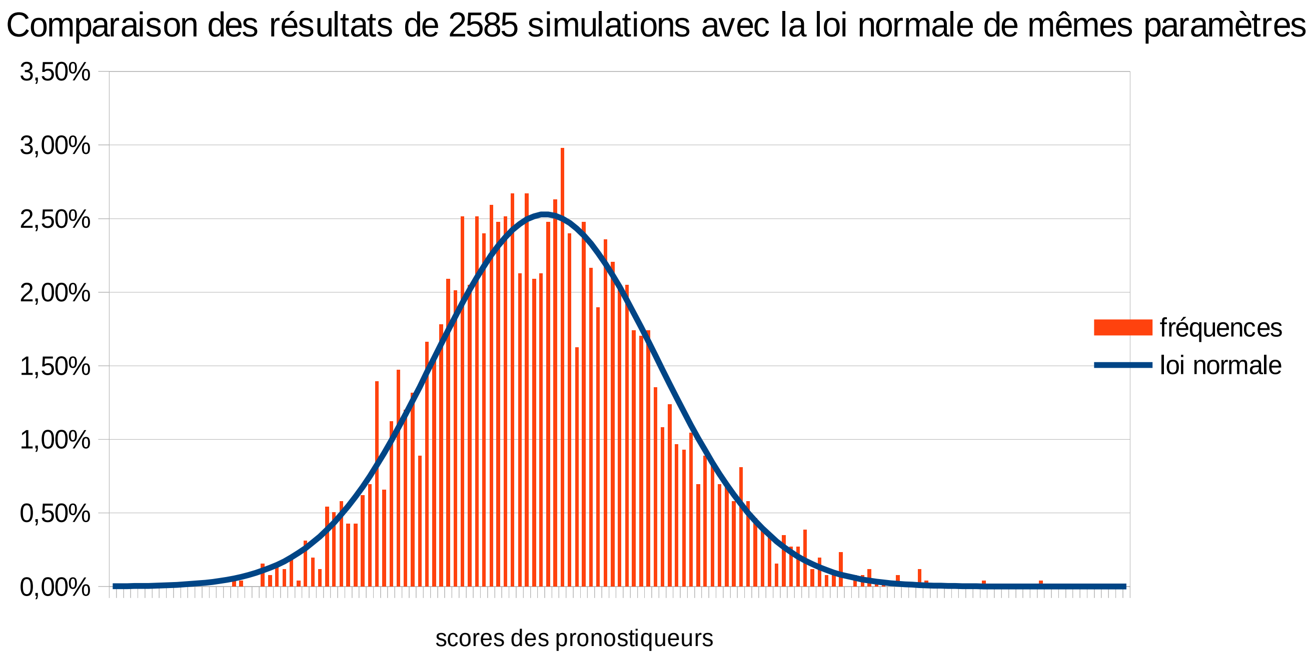
Pour cela on utilise l'algorithme de simulation d'une loi binomiale de paramètres N et p, qui figure aux programmes de toutes les classes de premières :

```
S prend la valeur 0
Pour I allant de 1 à N
    Si « nombre aléatoire »<p
        alors S prend la valeur S+1
    fin de Si
Fin de pour
Afficher S
```

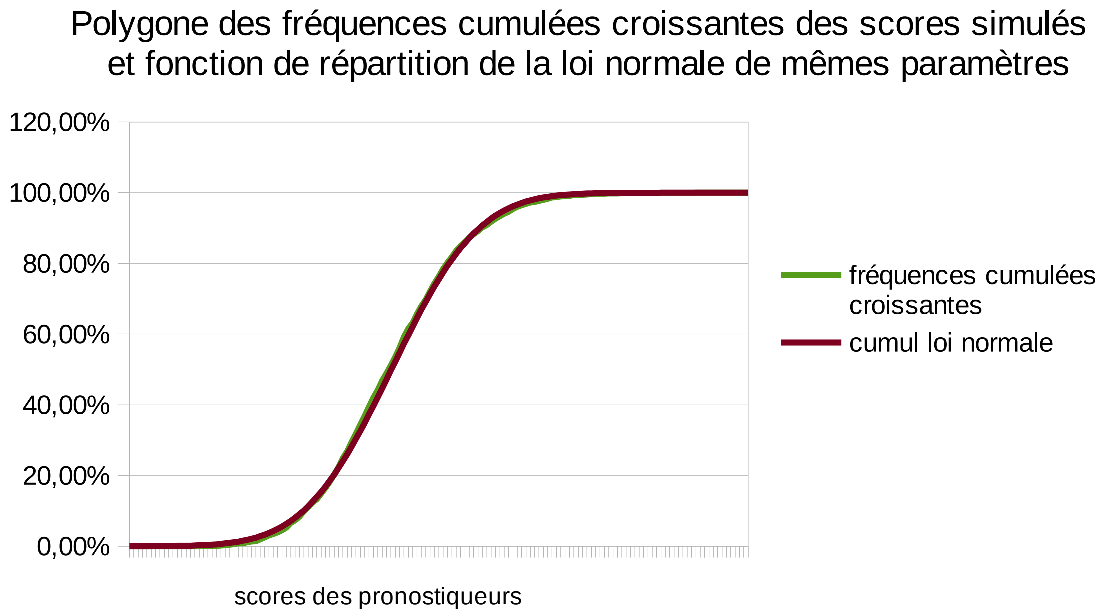
En version tableur, cet algorithme se traduit (sans la boucle, on répète à la main, d'où l'intérêt de prendre n=10 et non n=660...) :

```
=SI(ALEA()<0,132;1)+SI(ALEA()<0,132;1)+SI(ALEA()<0,132;1)+SI(ALEA()<0,132;1)+SI(ALEA()<0,132;1)+SI(ALEA()<0,132;1)+SI(ALEA()<0,132;1)+SI(ALEA()<0,132;1)+SI(ALEA()<0,132;1)+SI(ALEA()<0,132;1)
```

Sur 2585 simulations, le score moyen est alors de 80,6 avec un écart-type de 15,8 environ, et une répartition des fréquences qui est relativement proche de celle associée à la densité de probabilité de la loi normale de mêmes paramètres.



La comparaison des fréquences cumulées avec la fonction de répartition de la loi normale est encore plus nette à ce sujet puisqu'on ne parvient pas à les distinguer :



4-Qu'est-ce qu'un bon pronostiqueur ?

Un bon pronostiqueur c'est un pronostiqueur qui ayant analysé finement toutes les informations sur les équipes, réalise un meilleur score qu'un pronostiqueur qui joue au hasard, même dans le cas où ce hasard est un peu orienté (par les coefficients FIFA).

C'est donc un pronostiqueur dont le score s'éloigne des scores obtenus lors des simulations précédentes.

Avec les résultats obtenus lors des 2585 simulations , on a vu que la loi normale de paramètres $\mu=80{,}6$ et $\sigma=15{,}8$ fournissait une assez bonne approximation des scores obtenus.

On sait calculer l'intervalle de fluctuation à 95 % pour cette lois de probabilité, en effet $P(\mu-2\sigma<X<\mu+2\sigma)\approx 0{,}95$ ce qui donne ici environ 95 % de chance, si on joue selon la modélisation précédente, d'avoir un score entre $80{,}6-2\times 15{,}8=49$ et 112 (environ).

On peut donc considérer sur les bases de ces simulations, <u>qui ne tiennent pas encore compte des résultats réels de l'euro 2016</u>, qu'un bon pronostiqueur est un pronostiqueur dont le score dépassera 112 points (à ajuster selon les résultats réels de l'euro, et l'intervention du hasard lors des nouvelles simulations).

III-Un peu de probabilités

1-Avec l'intervalle de fluctuation de seconde

Une autre façon de se comparer, plus élémentaire, mais cette fois-ci sur une base probabiliste, consiste à ne garder que les points liés aux résultats des matchs, et de comparer son score à celui obtenu en choisissant totalement au hasard le résultat.

On a donc une probabilité $p=\frac{1}{3}$ d'avoir un bon résultat. On quantifier le hasard en estimant quels sont les scores probables de pronostiqueurs jouant au hasard; en effet, la fréquence des bons résultats lors des 36 matchs de poules doit se trouver dans l'intervalle de fluctuation à 95%. La formule vue en seconde, et applicable ici puisque n=36 est (de peu) au dessus de 30 et que les produits $n\,p=36\times\frac{1}{3}=12$ et $n(1-p)=36\times\frac{2}{3}=24$ dépassent 5, donne : $p-\frac{1}{\sqrt{n}}=\frac{1}{3}-\frac{1}{\sqrt{36}}=\frac{1}{6}$ et $p+\frac{1}{\sqrt{n}}=\frac{1}{3}+\frac{1}{\sqrt{36}}=\frac{1}{2}$ pour les bornes inférieure et supérieure.

En jouant au hasard, on a donc environ 95% de chance d'avoir entre $\frac{36}{6}=6$ et $\frac{36}{2}=18$ bons résultats.

Si on en obtient plus on peut donc considérer que ce score n'est pas dû qu'au hasard.

2- Avec une loi binomiale

Le nombre de bons pronostics suit la loi binomiale $B\left(36;\frac{1}{3}\right)$ puisqu'il y a 36 matchs de poules et qu'à chaque match, si on joue totalement au hasard entre Gagné, Nul et Perdu, on a une chance sur trois d'avoir le bon résultat.

On connaît pour cette loi les probabilités

score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
P(X=k)	0,0%	0,0%	0,0%	0,0%	0,2%	0,5%	1,4%	3,0%	5,4%	8,4%	11,4%	13,4%	14,0%	12,9%	10,6%	7,8%	5,1%	3,0%	1,6%	0,8%	0,3%	0,1%	0,0%
P(X<=k)	0,0%	0,0%	0,0%	0,0%	0,2%	0,8%	2,2%	5,1%	10,5%	19,0%	30,3%	43,8%	57,8%	70,7%	81,3%	89,1%	94,2%	97,2%	98,8%	99,5%	99,8%	99,9%	100,0%

L'intervalle de fluctuation à 95 % (en ne laissant de chaque côté qu'au plus 2,5 % des résultats les moins probables) est alors l'intervalle [a;b] avec a et b les plus petits entiers tels que $P(X\leqslant a)\geqslant 0{,}025$ et $P(X\leqslant b)\geqslant 0{,}975$ soit ici [7;18].

Un pronostiqueur pourra donc considérer que son score n'est pas dû qu'à la chance si sur les 36 matchs du premier tour il a réussi à trouver 19 résultats corrects ou plus.

On remarque que l'intervalle de fluctuation de seconde est un peu moins précis.

On notera surtout que <u>les hypothèses prises ici sont simplifiées à l'extrême</u> (notamment l'équiprobabilité entre G,N et P, qui n'est pas observée dans la réalité,seulement 20.1% de matchs nuls sur la période étudiée) et qu'<u>il est donc a priori facile de considérer sur ce seul critère qu'on est un bon pronostiqueur!</u>

IV-Et la phase finale ?

1-Pourquoi s'arrêter au stade de la phase de poule ?

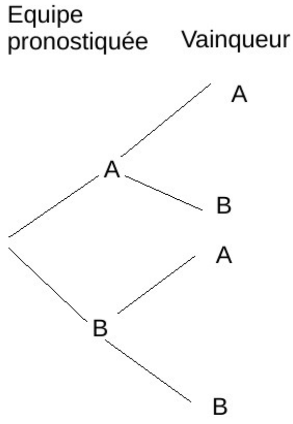
La répartition des meilleurs troisièmes dans le tableau final répond à un processus complexe (https://fr.wikipedia.org/wiki/Championnat_d'Europe_de_football_2016#Distribution_des_.C3.A9quipes_class.C3.A9es_meilleures_troisi.C3.A8mes_de_groupes_en_huiti.C3.A8mes_de_finale)

qu'il n'est pas facile de traduire en version tableur avec des formules simples (si, ou…) et surtout, cette phase est source d'une très grande variabilité : Certains pronostiqueurs n'auront pas qualifiés les bonnes équipes, ou pas à la bonne place, ainsi rapidement un grand nombre de pronostiqueur se retrouve avec des pronostics non pertinents donc aucune chance de marquer des points, tandis que ceux qui avaient réussis leurs pronostics lors de la première phase voient leurs chances de marquer des points rester conséquentes, ce qui augmente la dispersion des scores finaux et donc rend la prédictibilité des résultats plus compliquée à obtenir (plus de simulations pour avoir un intervalle de fluctuation raisonnable).

Les matchs qui se jouent aux prolongations ou aux tirs au buts sont également source de complication dans le barème des points attribués.

2- Un modèle probabiliste pour la phase finale.

En général les concours de pronostics permettent aux pronostiqueurs de repartir, après la phase des poules avec les équipes qualifiées (ce qui réduit la variabilité des pronostics, mais ne la ramène pas au niveau de la phase de groupe, puisque les erreurs de pronostics lors des huitièmes de finale se répercutent sur les pronostics des quarts de finale et ainsi de suite…)
Dans la suite, on considère seulement le vainqueur de chaque match, et on considère qu'on a une chance sur deux de prédire son nom (On pourrait faire une étude statistique avec un modèle plus complexe, avec des points attribués pour avoir prédit le vainqueur, mais aussi les prolongations voire les tirs au but, et tenant compte de la force des équipes opposées, selon que l'on a un match entre le 1er et le 3^{ieme} du groupe, ou entre le 1er et un 2^{ième}, ou entre deux 2^{ièmes}, comme le prévoit le règlement des huitièmes de finales, mais la traduction probabiliste de cette situation est trop complexe :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).



Probabilité d'un bon pronostic : 2/4=0,5

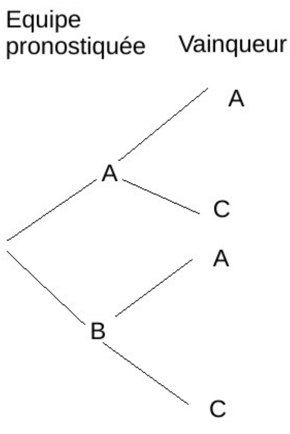
On a donc une chance sur deux pour chacun des huitièmes de finales d'avoir le bon qualifié.

La loi donnant le nombre de bon qualifiés (et donc de bons pronostics) est donc la loi binomiale B(8;0,5).

Mais cela se complique fortement au niveau des quarts de finale :

- Si on n'a eu aucun bon pronostic, on n'en aura plus puisque toute nos équipes sont éliminés.

- Si on a eu un seul bon pronostic, il y a une chance sur quatre qu'on ait un deuxième bon pronostic (que cette équipe gagne et qu'on l'ait donnée gagnante) puisqu'il y a une chance sur deux qu'on ai pronostiqué la victoire de cette équipe, puis une chance sur deux (événements indépendants) qu'elle gagne son match.



Probabilité d'un bon pronostic : 1/4=0,25

- Si on a eu deux bons pronostics, soit, ce sont deux équipes qui s'affrontent et là on obtient une chance sur deux d'avoir un bon demi-finaliste ; soit ce sont deux équipes qui ne s'affrontent pas et on se retrouve dans la situation précédente mais pour deux matchs.

Et ainsi de suite... On constate alors qu'il ne suffit pas de connaître le nombre de bons pronostics en huitième de finale, mais qu'il faut tenir compte match par match de la situation : 0; 1 ou 2 équipes correctement pronostiquées.
Une solution consisterait à réaliser un arbre avec les choix possibles selon la situation match par match mais comme il en reste 8+4+2+1=15 on obtient en bout de branches 2^{15} =32768 possibilités (un peu moins en réalité, puisqu'on peut « élaguer » les branches où on n'a plus un seul pronostic valable).
Il est plus simple (mais très long quand même) de procéder d'abord pour un quart de finale en relevant à chaque étape le couple (nombre de pronostics encore valable ; somme des pronostics réussis) , et de croiser cela avec un second quart de finale (il y a indépendance) puis d'en déduire les probabilité des scores et des nombres de pronostics réussis pour une demi-finale, puis de croiser avec la même chose pour la seconde demi-finale, avant de finir par les calculs pour le résultat de la finale, le tout selon les probabilités expliquées ci-dessus (voir le fichier tableur pour les détails).
Après trois pages de calculs, on obtient :

nombre de bons pronostics	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Probabilité*32768	128	768	2176	3984	5400	5816	5192	3953	2608	1500	752	326	120	36	8	1
en %	0,39%	2,34%	6,64%	12,16%	16,48%	17,75%	15,84%	12,06%	7,96%	4,58%	2,29%	0,99%	0,37%	0,11%	0,024%	0,003%
cumul en %	0,39%	2,73%	9,38%	21,53%	38,01%	55,76%	71,61%	83,67%	91,63%	96,21%	98,50%	99,50%	99,86%	99,97%	99,997%	100,00%

Soit un score moyen de 5,3125 pour un écart-type d'environ 2,2.

On peut alors chercher quel est l'intervalle de fluctuation à au moins 95 % de cette loi de probabilité, la ligne des $P(X\leqslant k)$ baptisée « cumul » donne ce résultat avec la méthode vue en première pour l'intervalle de fluctuation de la loi binomiale : On cherche *a* tel que $P(X\leqslant a)\geqslant 0{,}025$ et *b* tel que $P(X\leqslant b)\geqslant 0{,}975$, ici on récupère *a*=1 et *b*=10, un pronostiqueur pourra donc affirmer avec un risque inférieur à 5 % qu'il ne doit pas son résultat au hasard, s'il a obtenu entre 11 et 15 bons pronostics. (Si on considère que ce seuil correspond plutôt à c tel que $P(X\leqslant c)\geqslant 0{,}95$ alors c=9 et on peut considérer qu'à 10 bons pronostics, ce n'est pas que du hasard.)