Revue de statistique appliquée

P. Roger
M.-H. Broihanne

Les joueurs de loto français choisissent-ils leurs numéros au hasard ?

Revue de statistique appliquée, tome 54, n° 3 (2006), p. 83-98
<http://www.numdam.org/item?id=RSA_2006__54_3_83_0>

© Société française de statistique, 2006, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

Numdam
Article numérisé dans le cadre du programme Numérisation de documents anciens mathématiques
http://www.numdam.org/

Rev. Statistique Appliquée, 2006, **LIV** (3), 83-98

LES JOUEURS DE LOTO FRANÇAIS CHOISISSENT-ILS LEURS NUMÉROS AU HASARD?

P. ROGER, M.-H. BROIHANNE

Université Louis Pasteur, Laboratoire de Recherche en Gestion et en Économie
Pôle Européen de Gestion et d'Économie
61, avenue de la Forêt Noire, F-67085 Strasbourg Cedex
roger@cournot.u-strasbg.fr, mhb@cournot.u-strasbg.fr

RÉSUMÉ

Nous montrons dans cet article que les joueurs du loto français ne choisissent pas leurs numéros au hasard. Ils utilisent des stratégies ou heuristiques communes qui conduisent à une distribution de probabilité des grilles jouées différente d'une distribution uniforme. En particulier, il existe une nette préférence pour les «petits» numéros. Ce point est montré en utilisant un estimateur simple des numéros réellement joués fondé sur l'existence des rangs de gain incluant le numéro complémentaire. Le résultat obtenu n'est pas surprenant car il se conforme aux observations faites dans d'autres pays européens ainsi qu'aux Etats-Unis. En revanche, la structuration des gains du loto français autorise une procédure d'estimation beaucoup plus simple que celle utilisée dans d'autres pays.

Mots-clés : *loto, numéro complémentaire.*

ABSTRACT

In this paper, we show that french lotto players don't select numbers randomly. They share some heuristics to choose numbers. This leads to a probability distribution of actually played numbers very different from the uniform distribution. In fact, the data show a clear preference for low numbers. To prove this point, we use a very simple estimator of the probability distribution of played numbers, based on the specific configuration of gains in the French lotto. The result is not surprising; several empirical works show the same kind of preferences in foreign European countries or in the United States. However, the estimator we propose in the present paper cannot be applied to these foreign games, due to the structuration of their prizes.

Keywords : *lotto, bonus number.*

1. Introduction

Le 9 février 2005, le numéro 53 est sorti à la loterie de Venise [1] , ce qui a enfin ramené le calme en Italie. Ce numéro n'était pas sorti depuis 182 tirages et, en dépit d'interventions télévisées de statisticiens expliquant l'indépendance des tirages successifs, et du Ministre de l'Économie invitant les Italiens à la modération, des sommes considérables ont été misées sur ce numéro, au prix, dans certains cas, de détournements de fonds. Le journal italien *La Repubblica* attribuait 4 morts violentes à cette longue série de tirages sans apparition du numéro 53.

Victor Hugo, dans «Le dernier jour d'un condamné» met les paroles suivantes dans la bouche du gardien de prison.

«*Voici. Je suis un pauvre gendarme. Le service est lourd, la paye est légère; mon cheval est à moi et me ruine. Or je mets à la loterie pour contre-balancer. Il faut bien avoir une industrie. Jusqu'ici il ne m'a manqué pour gagner que d'avoir les bons numéros. J'en cherche partout de sûrs; je tombe toujours à côté. Je mets le 76; il sort le 77. J'ai beau les nourrir, ils ne viennent pas*».

Ces deux exemples illustrent le fait que les joueurs n'attribuent pas forcément aux différents numéros la même probabilité de sortie puisque, dans le premier, ils lient la probabilité de sortie d'un numéro au nombre de tirages sans apparition de ce numéro et, dans le second, le gendarme évoque des «numéros sûrs».

En France, comme dans beaucoup d'autres pays, les jeux de hasard constituent un phénomène de société dont l'importance économique est considérable. La Française des Jeux réalise, avec le seul loto, un chiffre d'affaires annuel de l'ordre de 2 milliards d'euros, alors que son chiffre d'affaires total était de 8,55 milliards d'euros en 2004, ce qui la place parmi les 50 plus grandes entreprises françaises.

L'attractivité du loto réside dans la simplicité du jeu et dans la promesse d'un jackpot très élevé à chaque tirage. Le jeu a été créé pour remplacer la Loterie Nationale déclinante et pour faire face à la concurrence du tiercé. En effet, le tiercé présente le caractère d'une loterie active puisque le parieur choisit des chevaux alors que la Loterie Nationale était une loterie passive, les numéros étant déjà pré-imprimés sur le billet. Le loto présente ainsi un bon compromis en étant un jeu de hasard mais en laissant le joueur choisir les numéros sur lesquels il souhaite miser.

Le premier tirage a eu lieu le mercredi 19 mai 1976 ; sur le rythme d'un tirage hebdomadaire le mercredi dans les premières années, il est passé à 2 tirages hebdomadaires (mercredi et samedi) en février 1984 puis à 4 tirages par semaine en 1990. Même si l'engouement pour ce jeu s'est quelque peu ralenti ces dernières années, de 20 à 30 millions de grilles sont cependant jouées à chaque tirage.

La caractéristique essentielle du loto réside dans le fait qu'il s'agit d'un jeu de pari mutuel. Cela signifie que les gains sont définis comme un pourcentage fixe des mises et sont répartis sur l'ensemble des gagnants [2] . Puisque le joueur est actif

[1] Cette loterie fonctionne sur le principe de la loterie génoise, c'est-à-dire que les joueurs misent sur des sous-ensembles de 1 à 5 numéros choisis parmi 90. Ce type de loterie existe depuis le 16ème siècle (voir Bradley (2001), Willmann (1999), Roger (2005)).

[2] Contrairement à la loterie italienne évoquée plus haut dans laquelle les gains individuels sont fixés indépendamment du nombre de gagnants.

en choisissant ses numéros, la question de la distribution de probabilité des numéros effectivement choisis par les joueurs est importante. Or les organisateurs de loteries de ce type, considérant sans doute que cette information est stratégique, ne la rendent pas publique.

Au-delà de l'aspect statistique du problème, cette question a des conséquences financières pour le joueur. En effet, le loto étant fondé sur le principe du pari mutuel, avec 50,485 % des mises redistribuées, jouer des nombres populaires, s'il en existe, est une stratégie sous-optimale puisqu'elle diminue la rentabilité espérée du joueur.

Dans cet article nous montrons sur le cas du loto français que, d'une part, il existe une nette préférence pour les petits numéros et que, d'autre part, les spécificités du loto français autorisent l'estimation de la distribution des numéros joués par une méthode bien plus simple que celle employée en général (voir Farrell *et al.* (2000)), à savoir une méthode du maximum de vraisemblance couplée avec une simulation de Monte Carlo. En effet, dans la plupart des pays, une seule combinaison contenant le numéro complémentaire est gagnante, à savoir l'obtention de 5 numéros plus le complémentaire. En France, trois numéros corrects et le complémentaire assurent un gain. En conséquence, il existe un grand nombre de gagnants de ce type et ils constituent un large échantillon de joueurs dont on est sûr qu'ils ont joué le numéro complémentaire. Cela permet de construire un estimateur des probabilités de jouer tel ou tel numéro.

Les résultats que nous obtenons par cette méthode simple mais originale confirment ceux obtenus dans d'autres pays à quelques différences marginales près [3]. De plus, l'estimation basée sur les propriétés du numéro complémentaire est obtenue instantanément alors que celle s'appuyant sur la méthode du maximum de vraisemblance et la simulation de Monte Carlo de Farrell *et al.* demande des heures, voire des jours de calcul quand la base de données comporte un nombre important de tirages [4] .

Nous montrons que la préférence pour les petits numéros est marquée chez les joueurs du loto de la Française des Jeux en considérant près de huit années de données, soit plus de 1500 tirages. Dans un premier temps, nous procédons à une analyse de régression dans laquelle la variable dépendante est la déviation standardisée du nombre de gagnants (du gros lot) par rapport au nombre espéré, sous l'hypothèse que ce nombre de gagnants suit une loi de Poisson (hypothèse vérifiée si les joueurs choisissent leurs numéros au hasard). La variable indépendante de la régression est la moyenne des numéros sortis au tirage officiel. Nous mettons ainsi en évidence une relation décroissante entre les deux variables. Le nombre de gagnants est lié négativement à la moyenne des numéros tirés, même après prise en compte du nombre total de grilles jouées.

La seconde analyse montre que si l'on réalise une typologie des tirages selon le nombre de gagnants, les distributions des nombres sortis dans les différentes catégories ne sont pas identiques, confirmant la tendance apparue dans la première analyse.

Enfin, à l'aide des effectifs de gagnants (à 3 numéros) ayant joué le bon numéro complémentaire, on estime la distribution des nombres effectivement joués. Celle-

---

3 Voir Ziemba *et al.* (1986) pour le Canada, Scott et Gulley (1995) pour différents lotos U.S., Papachristou et Karamanis (1998) pour la Grèce et Farrell *et al.* (2000) pour la Grande-Bretagne.

4 Roger-Broihanne (2006) mentionnent un temps de calcul de plusieurs jours pour 3638 tirages.

ci n'a rien d'uniforme. 33 numéros sur 49 se situent en dehors de l'intervalle de confiance à 99 % induit par l'hypothèse de choix de grilles au hasard. On constate que les numéros inférieurs ou égaux à 13 sont les plus populaires et les numéros supérieurs à 29 les plus impopulaires. Jouer des petits numéros implique donc le risque, en cas de gain, de devoir partager le jackpot et diminue ainsi l'espérance de rentabilité du jeu.

L'article est organisé comme suit. La section 2 rappelle brièvement les règles du jeu, en particulier celles qui concernent la répartition des gains entre les différents rangs. La section 3 présente les différentes méthodes de test de l'existence de numéros préférés et d'estimation de la distribution de probabilité des nombres effectivement joués. La section 4 décrit la base de données construite pour l'analyse et récapitule les résultats empiriques obtenus sur plus de 1500 tirages. La section 5 conclut l'article en proposant des prolongements de la recherche.

2. Les règles du jeu

Jouer au loto est à la portée de tous puisqu'il suffit de cocher 6 numéros sur une grille en comportant 49. Les joueurs peuvent choisir leurs numéros eux-mêmes ou laisser l'ordinateur de la Française des Jeux opérer ce choix. Il s'agit du Système Flash.

Au tirage officiel, ce sont en fait 7 nombres qui sont tirés au hasard sans remise, les 6 premiers constituant la grille principale et le dernier étant appelé «numéro complémentaire», numéro qui joue un rôle particulier dans la détermination des gains.

Il existe 7 possibilités de gain (appelées «rangs») à chaque tirage, et ce depuis le 15/10/1997 [5] . Un joueur peut gagner s'il a trouvé les 6 numéros de la grille principale ou de 3 à 5 bons numéros de cette grille, avec ou sans le numéro complémentaire. Nous noterons 3 + C, 4 + C et 5 + C les rangs de gain incluant le numéro complémentaire [6] .

Le tableau 1 récapitule la répartition des gains entre les différents rangs. Rappelons cependant que cette répartition porte sur 50,485 % des mises compte tenu du prélèvement à la source. On constate sur ce tableau que les rangs 4 et 5 (resp. 6 et 7) sont regroupés car une règle stipule que les gagnants au rang 4 (resp. 6) touchent un montant double de celui encaissé par les gagnants au rang 5 (resp. 7). Le montant dévolu aux gagnants d'un rang est donc proportionnel au nombre de joueurs mais pas au nombre de gagnants, c'est l'illustration du principe de pari mutuel. En conséquence, il est plus rentable de jouer des grilles «impopulaires», s'il en existe.

---

[5] Le jeu existe depuis 1976 mais le nombre de rangs de gain a évolué au cours du temps mais semble stabilisé à 7 depuis cette date.

[6] 3 + C (resp. 4 + C et 5 + C) correspond à avoir 3 (resp. 4 et 5) bons numéros de la grille principale, plus le numéro complémentaire.

TABLEAU 1
La répartition des gains

Rang	Numéros corrects	1er tirage	2nd tirage
1	6	29,1%	45,4%
2	5+C	3,1%	3,1%
3	5	10,3%	10,3%
4 et 5	4+C et 4	13,6%	13,6%
6 et 7	3+C et 3	27,6%	27,6%
Report		16,3%	

Source : www.fdjeux.com

Par ailleurs, au premier tirage, les sommes dévolues aux gagnants ne sont pas toutes redistribuées. 16,3 % apparaissent sous la mention «Report» dans le tableau 1. Cela signifie que 16,3 % de la part des mises qui devraient être redistribuées lors de ce tirage sont en fait affectées aux gagnants du second tirage. L'existence d'un tel report était, au départ, justifiée par le fait que la participation aux deux tirages n'était pas obligatoire à la création de ce second tirage. Le report constituait donc, pour les joueurs, une incitation à cette double participation. Par la suite, la participation aux deux tirages d'un même jour est devenue obligatoire mais le report est resté. On peut l'interpréter comme la volonté de l'organisateur de maintenir, dans les résultats publiés, un jackpot élevé (au moins au second tirage) car l'existence de celui-ci est un argument essentiel dans le marketing de ce type de loterie. On constate d'ailleurs que le pourcentage dévolu au premier rang est nettement plus élevé au second tirage.

Une grille coûte aujourd'hui 0,3 euros mais la mise minimale est de 1,2 euros, correspondant à deux grilles valides pour les deux tirages d'un même jour. Avant le passage à l'euro et depuis 1996 la mise était de 2 Frs, soit 0,304 euros.

3. Tests de l'existence de numéros préférés

Les données disponibles (voir section 4) permettent de connaître, outre les numéros sortis à chaque tirage, le nombre de grilles gagnantes à chaque rang et les gains individuels correspondants. Compte tenu des informations contenues dans le tableau 1, il est possible de déduire le montant total joué et donc le nombre de grilles validées par les joueurs.

3.1. Le caractère aléatoire des tirages officiels

À titre de précaution, avant l'analyse des numéros sélectionnés par les joueurs, il faut s'assurer que les tirages officiels réalisés par la Française des Jeux sont bien aléatoires. Le test du χ^2 ne peut être utilisé tel quel car si les tirages successifs sont indépendants, les numéros d'un même tirage ne le sont pas, puisque chaque tirage est opéré sans remise. On utilise alors la statistique suivante [7] :

$$\chi_J^2 = \frac{N-1}{N-n} \sum_{i=1}^{N} \frac{(O_i - E)^2}{E}$$

où N désigne le cardinal de l'ensemble des nombres à l'intérieur duquel se fait le tirage ($N = 49$), n est le nombre de numéros tirés dans la grille principale ($n = 6$), O_i est la fréquence du numéro i sur l'ensemble des tirages considérés (en nombre m), E est la fréquence théorique définie par $E = m\frac{n}{N}$.

χ_J^2 est asymptotiquement distribuée selon une loi de χ^2 à $N - 1$ degrés de liberté (ddl).

3.2. Numéros préférés et nombre de gagnants

3.2.1. Neutralisation du volume des ventes

L'existence de numéros préférés peut être testée à partir de l'intuition suivante. Sous l'hypothèse d'un choix aléatoire de numéros par les joueurs, le nombre de gagnants au premier rang doit être déterminé par la quantité de grilles vendues et pas par les numéros sortis au tirage officiel. Or les études mentionnées plus haut ont montré que, dans plusieurs pays, les joueurs ont tendance à sélectionner des petits numéros. De ce fait, un moyen simple de tester cette hypothèse consiste à établir une relation entre le nombre de gagnants au premier rang et la moyenne des numéros sortis au tirage officiel. Notons N_6^k et X^k ces quantités pour le tirage k; N_6 et X désignent les vecteurs à m composantes contenant les valeurs de ces variables pour les m tirages.

Si Q^k grilles ont été jouées au k-ème tirage, le nombre de gagnants au premier rang, noté N_6^k, suit, sous l'hypothèse de choix aléatoire, une loi de Poisson de paramètre $\lambda^k = Q^k p$ où p désigne la probabilité de gain au premier rang ($p = \frac{6!43!}{49} = \frac{1}{13983816} = 7,15 \times 10^{-8}$ pour le loto français). En effet, chaque grille jouée peut être vue comme un essai pour obtenir le gros lot. Dans ces conditions, le nombre de gagnants suit une loi binomiale de paramètres Q^k et p. Compte tenu de l'ordre de grandeur de ces paramètres, l'approximation par la loi de Poisson est justifiée.

Pour neutraliser l'ordre de grandeur des ventes, on centre et réduit le nombre de gagnants au premier rang en posant :

$$Z^k = \frac{N_6^k - \lambda^k}{\sqrt{\lambda^k}}$$

Le vecteur $Z' = \left(Z^1, ..., Z^m\right)$ donne ainsi les déviations standardisées du nombre de gagnants par rapport au nombre espéré.

[7] Voir Joe (1993) ou Stern et Cover (1989).

La mise en évidence d'une préférence pour les petits numéros peut alors être réalisée en régressant Z sur X à l'aide du modèle :

$$Z = aX + b\mathbf{1} + \varepsilon$$

où $\mathbf{1}$ est un vecteur de dimension m dont toutes les composantes sont égales à 1, a et b désignant les coefficients de la régression et ε la perturbation.

L'hypothèse nulle de choix aléatoire de numéros s'écrit alors $H_0 : a = 0$ puisqu'en cas de choix aléatoire, la moyenne des numéros tirés ne doit pas influencer le nombre de gagnants.

3.2.2. Typologie des tirages selon le nombre de gagnants

Une alternative pour réaliser ce test consiste à constituer des classes de tirages selon le nombre de gagnants au premier rang et à répartir les 49 numéros en sous-ensembles (par exemple 7 sous-ensembles de 7 numéros). Sous l'hypothèse nulle, il ne devrait pas y avoir de différence dans la répartition des numéros dans les différentes catégories. En revanche, si une préférence marquée existe pour les petits numéros, ceux-ci devraient être surreprésentés dans les tirages comptant plusieurs gagnants au premier rang et sous représentés dans la catégorie de tirages sans gagnant. Nous montrerons dans la section suivante que les petits nombres (inférieurs à 28) apparaissent plus souvent dans les tirages à plus de 3 gagnants et les nombres supérieurs à 29 dans les tirages sans gagnant. Il est bon de rappeler ici que le nombre moyen de grilles jouées à chaque tirage est proche de 25 millions, ce qui donne une espérance du nombre de gagnants au premier rang, en cas de choix aléatoire, proche de 2 (plus précisément, $25 \times 10^6 \times 7,15 \times 10^{-8} = 1,79$).

3.2.3. Les proportions de gagnants avec numéros complémentaires

Le tableau 2 rappelle les probabilités de gain pour les 7 rangs. On constate qu'en cas de choix aléatoire des numéros par les joueurs, la probabilité de gagner au rang 3+C sachant que l'on a 3 numéros corrects s'écrit :

$$p_c = \frac{C_{42}^2 \times C_6^3}{C_{42}^2 \times C_6^3 + C_{42}^3 \times C_6^3} = \frac{C_{42}^2}{C_{42}^2 + C_{42}^3} = \frac{C_{42}^2}{C_{43}^3} = 0,0697$$

En divisant la population des tirages selon le numéro complémentaire, on peut, pour chaque tirage, obtenir une estimation de cette probabilité. Notons, pour le tirage k, N_3^k et N_{3+C}^k les nombres de gagnants aux rangs 7 et 6 et $\hat{p}_c^k$ la proportion des gagnants au rang 6 dans l'ensemble des gagnants aux rangs 6 et 7. On a donc :

$$\hat{p}_c^k = \frac{N_{3+C}^k}{N_{3+C}^k + N_3^k}$$

Si c_k désigne le numéro complémentaire du tirage k, on estime la probabilité p_c^i de gain au rang 6 conditionnellement à un gain aux rangs 6 ou 7 lorsque le numéro

TABLEAU 2
Probabilités de gains

Rang	**Numéros gagnants**	**Combinaisons gagnantes**	**Probabilité de gain**
1	6	1	$7{,}15\times10^{-8}$
2	5+C	6	$4{,}29\times10^{-7}$
3	5	$C_{42}^1 \times C_6^5 = 252$	$1{,}8\times10^{-5}$
4	4+C	$C_{42}^1 \times C_6^4 = 630$	$4{,}5\times10^{-5}$
5	4	$C_{42}^2 \times C_6^4 = 12915$	$9{,}23\times10^{-4}$
6	3+C	$C_{42}^2 \times C_6^3 = 17220$	$1{,}23\times10^{-3}$
7		$C_{42}^3 \times C_6^3 = 229600$	$1{,}64\times10^{-2}$

complémentaire est i par :

$$\hat{p}_c^i = \frac{1}{m_i} \sum_{k=1}^{m_i} \hat{p}_c^k \mathbf{1}_{\{c_k=i\}}$$

où m_i désigne le nombre de tirages dont le numéro complémentaire est i. Il reste ensuite à tester les 49 hypothèses nulles $H_0 : p_c^i = p_c$ pour $i \in \{1, ..., 49\}$. Lorsque cette hypothèse nulle est rejetée, c'est le signe qu'un numéro est plus (ou moins) populaire que la moyenne.

4. Étude empirique

4.1. Les données

On peut télécharger l'historique des résultats depuis l'origine du jeu sur le site de la Française des Jeux [8] . Les informations fournies sont la date, le jour de la semaine, le numéro du tirage de l'année, le numéro de tirage du jour (1 ou 2), les nombres de gagnants et les gains individuels à chaque rang. Le tableau 1, indiquant les clés de répartition des mises, permet de reconstituer le nombre de grilles vendues, en prenant toutefois la précaution de tenir compte du changement de devise en 2002 et du taux de prélèvement «à la source».

En effet, la grille coûtait 2 Frs avant le passage à l'euro (soit 0,304 euros) alors que son prix a été arrondi à 0,3 euros après cette date. Il est important de noter que

[8] www.fdjeux.com ou www.loto.fr pour le site du loto.

le tableau 1 est resté stable depuis 1997 mais ces clés de répartition ont changé à de nombreuses reprises depuis 1976. Pour reconstituer les ventes sur toute la période, il faut alors tenir compte de l'évolution du taux de prélèvement et des clés de répartition des gains. Toutes ces informations figurent sur le site $www.sojah.com$ qui contient tous les textes législatifs (anciens et actuels) sur les jeux de hasard en France.

Nous avons procédé à l'estimation des ventes à partir du pourcentage des mises redistribuées aux rangs 4 et 5, soit 13,6 %, plutôt qu'à partir des rangs 6 et 7 pour deux raisons. La première est liée aux arrondis. En effet, les gains sont arrondis au dixième d'euro inférieur; il est donc préférable de choisir un rang avec un nombre plus faible de gagnants. La seconde raison est plus importante; pendant la période d'étude, les règles concernant le report des sommes non gagnées au premier rang ont été modifiées. En ce qui concerne le premier tirage, ces sommes sont maintenant redistribuées aux rangs 6 et 7, ce qui complique quelque peu le calcul. Notons cependant que cette précaution conduit encore à une légère sous-estimation des ventes. En effet, le nombre de gagnants est de l'ordre de 30 000 aux rangs 4 et 5, ce qui conduit en moyenne à une sous-estimation de 1 500 euros des sommes redistribuées à ce rang. Cela conduit à une sous-estimation du nombre de grilles vendues de l'ordre de $\frac{1500}{(0,136 \times 0,50485 \times 0,3)} = 72823$, c'est-à-dire environ 0,3 %.

Nous avons, dans cet article, retenu les 1534 tirages officiels réalisés depuis le 15 octobre 1997 jusqu'au 16 février 2005. Le choix de la date de début de période est justifié par l'apparition des 7 rangs de gain à cette date.

Sur cette période, en moyenne, il y a eu 1,7 gagnants au premier rang et ils ont empoché 952 000 euros chacun. 426 tirages, soit plus de 25 %, se sont soldés sans gagnant au premier rang et 446 avec un seul gagnant. La distribution du nombre de ces gagnants du gros lot (6 numéros corrects), tout en étant très concentrée, fait apparaître quelques «points aberrants» qui expliquent l'écart entre moyenne et médiane. En effet, le nombre maximum de gagnants est de 38 et 8 tirages sur les 1534 ont eu plus de 10 gagnants.

Concernant les grilles jouées, la moyenne sur l'ensemble de la période est de 24,5 millions avec un minimum à 17,9 millions et un maximum à 60 millions. Cette variabilité des ventes est en partie due au fait que la Française des Jeux peut, de temps à autre, annoncer une «Super-Cagnotte» attractive qui va stimuler la demande de grilles. Cela survient en général après un ou plusieurs tirages sans gagnant. C'est en particulier cette grande variabilité des ventes qui impose de neutraliser le volume des ventes dans l'analyse de la relation entre moyenne du tirage et nombre de gagnants (section 3.2.1). Par ailleurs, on peut mentionner, mais ce n'est pas l'objet de cet article, que le 13 février 2004 a débuté le premier loto européen appelé Euromillions, dont le tirage a lieu le vendredi, chaque grille coûtant 2 euros. Notre période d'analyse comporte donc 1323 tirages avant cette date et 212 tirages après cette date. Cet événement a bien sûr pour conséquence de modifier le comportement d'achat des joueurs de loto puisque certains ont opéré une substitution entre les deux jeux. En particulier, le nombre moyen de grilles vendues est de 24,88 millions avant le 13 février 2004 et tombe à 22,47 millions après. Un test usuel de comparaison de moyennes montre que cette différence est largement significative.

4.2. Les tirages officiels

La figure 1 représente les fréquences d'apparition des 49 numéros dans les 1534 tirages. La fréquence théorique est de 187,84 ($= \frac{6}{49} \times 1534$) et les fréquences observées varient entre 167 et 215. Toutefois la statistique χ^2_J à 48 ddl vaut 33,79 pour une valeur critique de 65,16 au seuil de 5 %. On ne peut donc rejeter l'hypothèse de tirages officiels aléatoires, ce qui est plutôt rassurant pour les joueurs. Le même calcul effectué sur les numéros complémentaires donne une valeur de la statistique de 49 avec une fréquence théorique de $31{,}30 = \frac{1534}{49}$ et des fréquences observées variant de 21 à 41. On ne peut, là non plus, rejeter l'hypothèse de tirage aléatoire pour les numéros complémentaires.

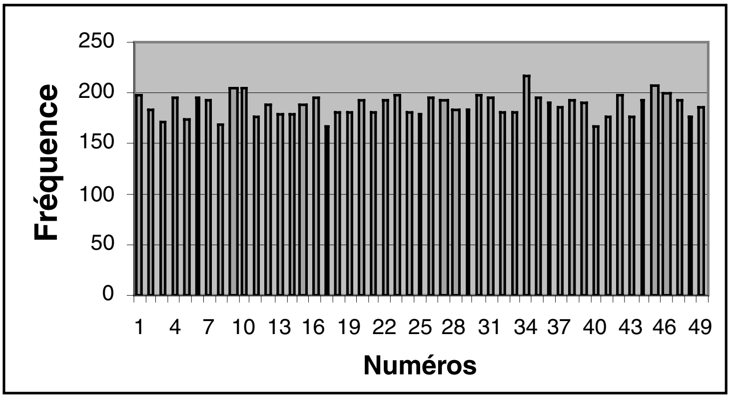

FIGURE 1
Fréquences des numéros dans la grille principale

4.3. Régression du nombre de gagnants sur la moyenne des tirages

Le tableau 3 présente les résultats de la régression réalisée sur les 1534 tirages dans le modèle $Z = aX + b\mathbf{1} + \varepsilon$ où Z est la déviation standardisée du nombre de gagnants et X désigne la moyenne des numéros tirés à chaque tirage. Le coefficient a a le signe attendu si les petits numéros sont préférés et ce coefficient est très largement significatif puisque la valeur du t de Student associée est de $-9{,}72$. Le R^2 de la régression est égal à 0,058.

Cette analyse de régression sur l'ensemble des tirages est cependant délicate à interpréter plus avant car il est bien connu que les coefficients de régression sont fortement influencés par les valeurs extrêmes. Or il existe quelques tirages pour lesquels le nombre de gagnants est très élevé puisque le maximum est de 38 gagnants le 17 octobre 2001. Si on se limite aux tirages pour lesquels le nombre de gagnants est inférieur ou égal à 10, on élimine 8 tirages de la base de données mais les résultats sont similaires. Le R^2 passe à 0,076 et le coefficient de X devient $-0{,}063$ avec une valeur du t de Student égale à $-11{,}19$. On peut donc bien attribuer les résultats obtenus au

TABLEAU 3
Probabilités de gains

	Coefficient	**Écart-type**	**t de Student**
b	$1,7$	$0,182$	$9,32$
a	$-0,069$	$0,007$	$-9,72$
R^2	$0,058$	Fisher : 94,60	

fait que le nombre de gagnants est significativement plus important quand la moyenne du tirage est faible.

4.4. Typologie selon le nombre de gagnants

Nous avons distingué les tirages selon le nombre de gagnants au premier rang en nous appuyant sur l'intuition suivante : si les joueurs cochent de manière préférentielle de petits numéros, ceux-ci devraient être associés aux tirages dans lesquels il y a un grand nombre de gagnants.

Le tableau 4 distingue 3 catégories de tirages correspondant à 0 gagnant, 1 ou 2 gagnants ou plus de 3 gagnants. Ces trois catégories sont notées 0, 1-2 et 3+; le nombre moyen de grilles vendues par tirage varie très peu entre les trois catégories puisqu'il oscille entre 24,5 millions (1-2) et 25,7 millions (3+).

Notons S_j, $j = 1, ..., 7$ les sous-ensembles de 7 numéros compris entre $7(j-1)+1$ et $7j$ et le nombre de sorties des numéros dans la catégorie de tirages k. Sous l'hypothèse de choix aléatoire, l'espérance du nombre de sorties dans chaque sous-ensemble est indépendante du nombre de gagnants au premier rang (et donc de la catégorie k), et égale à $6/7$. En effet, sur un ensemble de m tirages, $6m$ nombres sont tirés et $6m/7$ en moyenne appartiennent à chaque sous-ensemble de 7 numéros. Compte tenu de cette remarque, les résultats sont présentés dans le tableau 4 sous la forme $\frac{7\sum_{k=1}^{m} n_j^k}{6m}$, les sous-ensembles j apparaissant en lignes et les catégories de gagnants en colonne. La figure 2 représente les colonnes du tableau 4; les triangles correspondent à la catégorie 3+, les carrés à la catégorie 1-2 et les losanges à celle sans gagnant. Les nombres inférieurs (supérieurs) à 1 traduisent ainsi une sous (sur) - représentation du sous-ensemble de numéros considérés. La dernière ligne du tableau indique la valeur du χ^2 à 6 ddl pour chaque catégorie de gagnants, l'effectif théorique étant calculé par rapport à une distribution uniforme sur les sous-ensembles de numéros. Il apparaît que dans la catégorie sans gagnant ainsi que dans la catégorie à plus de 3 gagnants, l'hypothèse nulle peut être rejetée avec des valeurs calculées de 53,47 et 64,14. Les données du tableau montrent aussi que les rejets ont des causes « symétriques », à savoir une sur-représentation des numéros élevés dans la catégorie sans gagnant et une sous représentation de ces mêmes numéros dans la catégorie à plus de 3 gagnants. Par contre, l'hypothèse nulle ne peut être rejetée dans la catégorie intermédiaire, même au seuil de 10 %, car le χ^2 calculé vaut 3,24. Ce résultat est sans doute lié au fait que le nombre de grilles jouées est en moyenne de 25 millions, ce qui correspond à un nombre espéré de gagnants de 1,79 (cf. section 3.2.2) sous l'hypothèse d'un choix aléatoire. Ces tests confirment bien l'impression «visuelle» de la figure 2.

TABLEAU 4
Fréquences relatives de sous-ensembles de 7 numéros

Nombres	**0 gagnant**	**1-2 gagnants**	**3+ gagnants**
***N* tirages**	426	751	357
1 à 7	0,860	0,977	1,206
8 à 14	0,772	0,974	1,265
15 à 21	0,926	1,000	0,990
22 à 28	0,997	0,991	1,042
29 à 35	1,137	1,044	0,863
36 à 42	1,115	1,036	0,748
43 à 49	1,194	0,977	0,886
χ^2	53,47	3,24	64,14

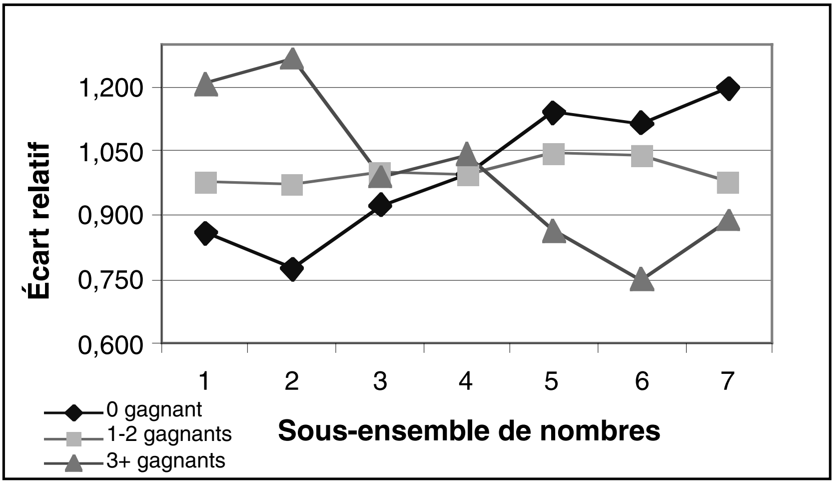

FIGURE 2
Fréquences relatives des sous-ensembles de numéros selon le nombre de gagnants

4.5. Estimation de la distribution des nombres joués

Afin d'affiner l'analyse, nous testons maintenant l'hypothèse de «popularité» ou d'impopularité de certains numéros en nous appuyant sur le numéro complémentaire. Rappelons que l'estimateur de la proportion de gagnants à 3+C dans l'ensemble des gagnants à 3 ou 3+C était donné, dans la section 3.2.3 par :

$$\hat{p}_c^i = \frac{1}{m_i} \sum_{k=1}^{m_i} \hat{p}_c^k \mathbf{1}_{\{c_k = i\}}$$

Sur les 1534 tirages, les 49 numéros sont sortis en tant que numéro complémentaire entre 21 et 41 fois. En cas de choix aléatoire des numéros par les joueurs, la valeur théorique de p_c est égale à 0,0697 (voir section 3.2.3) pour chacun des 49 numéros. Nous avons donc réalisé 49 tests de l'hypothèse nulle $p_c^i = p_c$. Les résultats sont résumés sur la figure 3. Les deux courbes en traits pleins encadrant la valeur théorique représentent les intervalles de confiance à 99 % pour chacun des numéros. Les valeurs observées se situent sur la courbe comportant des triangles comme identificateurs. Pour 33 numéros sur 49, l'hypothèse nulle est rejetée au seuil de 99 % dans un test bilatéral [9]. Conformément aux remarques faites dans les analyses précédentes, les petits numéros, inférieurs à 13 sont significativement plus joués que les autres (à l'exception du numéro 2) alors que les numéros supérieurs à 29 le sont significativement moins (à l'exception du 45 et du 49).

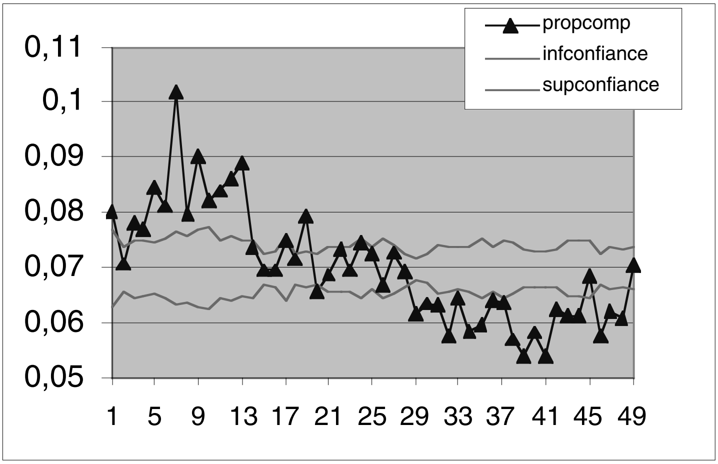

FIGURE 3
Intervalles de confiance des proportions de gagnants à 3 + C

Il s'ensuit qu'en termes de rentabilité espérée d'une grille, jouer 5-7-9-11-12-13 est la plus mauvaise stratégie, même si, comme nous l'avons montré dans la

[9] Les résultats chiffrés sont disponibles sur demande.

section 4.2, les tirages officiels sont faits au hasard et que, par conséquent, chaque grille a la même probabilité d'être tirée. A l'inverse, la grille 32-38-39-40-41-47 est celle dont on peut attendre la rentabilité espérée la plus élevée.

Bien évidemment, ces commentaires doivent être relativisés puisqu'ils supposent que le comportement (et les préférences de numéros) des joueurs est stable dans le temps.

Il est permis de s'interroger sur la raison qui pousse les organisateurs (dans tous les pays) à ne pas dévoiler directement ces informations en diffusant pour chaque tirage l'histogramme des numéros joués [10] . Une explication possible est que ceux-ci souhaitent éviter les «mouvements de foule», à savoir la concentration des jeux sur les grilles peu jouées aux tirages précédents. En effet, les joueurs savent que les gains sont partagés du fait du principe du pari mutuel et pourraient chercher à sélectionner les combinaisons peu jouées. La plupart d'entre eux adoptant le même raisonnement, l'effet obtenu serait, pour les joueurs, contraire à l'objectif recherché. Ce phénomène serait générateur de déceptions pour les clients du produit phare de la Française des Jeux. En cas de gain, ils encaisseraient des sommes faibles et la concentration des mises sur certaines combinaisons engendrerait un pourcentage élevé de tirages sans gagnant au premier rang. Ce raisonnement fait fi de l'hypothèse de rationalité parfaite si souvent posée dans les modèles économiques mais l'anecdote «italienne» relatée au début de cet article illustre le caractère peut-être irréaliste de cette hypothèse dans le contexte des jeux de hasard [11] .

5. Conclusion

Dans cet article, nous avons montré que les joueurs du loto de la Française des Jeux ne choisissent pas leurs numéros au hasard et ont une prédilection pour les «petits» numéros. Ce résultat a été obtenu à partir d'outils statistiques élémentaires et d'informations publiques, même si le présent travail a nécessité l'organisation de ces informations provenant de plusieurs sources.

Il est donc surprenant de ne pas trouver ce genre d'analyse sur les multiples sites web consacrés à ce jeu alors que la plupart d'entre eux diffusent des informations comme les écarts, c'est-à-dire le nombre de tirages pendant lequel un numéro donné n'est pas sorti. Or, cette information est sans valeur compte tenu de l'indépendance des tirages successifs.

Cette étude peut être prolongée dans plusieurs directions. Une question naturelle consiste à se demander s'il est possible, par un choix judicieux de numéros, de transformer un jeu largement défavorable (compte tenu du prélèvement à la source) en un jeu dont l'espérance de rentabilité est positive. Les méthodes employées par Ziemba *et al.* (1986) ou Papachristou et Karamanis (1998) pourraient être utilisées à cette fin.

[10] Simon (1999) a obtenu de la société Camelot, organisatrice du loto britannique, l'histogramme des numéros joués pour un seul tirage et montre que l'hypothèse de choix au hasard des grilles peut être rejetée.

[11] Voir Broihanne *et al.* (2004) pour des exemples sur les marchés financiers.

Deux autres questions intéressantes sont, d'une part, de déterminer l'influence des Super cagnottes sur la demande de grilles et, d'autre part, d'analyser l'existence d'une cannibalisation éventuelle entre les différents jeux de hasard.

La première analyse reste cependant difficile sur le loto français car les sommes non gagnées à un tirage donné ne sont pas systématiquement reportées sur le tirage suivant, contrairement à ce qui se passe sur la plupart des lotos étrangers [12] .

Nous avons en revanche montré par un test élémentaire que la moyenne des ventes a baissé depuis le démarrage de l'Euromillions, ce qui incite à poursuivre une étude plus détaillée du phénomène de substitution.

Remerciements

Nous remercions les rapporteurs du comité de rédaction pour les remarques et suggestions ayant permis d'améliorer le texte.

Références

[1] BEENSTOCK M. et HAITOVSKY Y. (2001), Lottomania and other Anomalies in the Market for Lotto, *Journal of Economic Psychology,* 22, 721-744.

[2] BRADLEY R. E. (2001), Euler and the Genoese Lottery, WP, Adelphi University.

[3] BROIHANNE M.-H., MERLI M. et ROGER P. (2004), *Finance Comportementale*, Economica.

[4] FARRELL L., HARTLEY R., LANOT G. et WALKER I. (2000), The Demand for Lotto : The Role of Conscious Selection, *Journal of Business & Economic Statistics*, 18, n°2, 228-241.

[5] JOE H. (1993), Tests of Uniformity for Sets of Lotto Numbers, *Statistics & Probability Letters,* 16, n°3, 169-251.

[6] PAPACHRISTOU G. et KARAMANIS D. (1998), Investigating Efficiency of Betting Markets : Evidence from the Greek 6/49 Lotto, *Journal of Banking and Finance,* 22, 1597-1615.

[7] ROGER P. (2005), *Lotomania : approche scientifique du jeu et du comportement des joueurs,* Editions Village Mondial.

[8] ROGER P. et BROIHANNE M.-H. (2006), Efficiency of Betting Markets and Rationality of Players : Evidence from the French 6/49 Lotto, *Journal of Applied statistics*, à paraître.

[12] Voir Beenstock et Haitovski (2001).

[9] SCOTT F.A. et GULLEY O. D. (1995), Testing for Efficiency in Lotto Markets, *Economic Inquiry*, 33, 175-188.

[10] SIMON J. (1999), An Analysis of the Distribution of Combinations Chosen by UK National Lottery Players, *Journal of Risk and Uncertainty*, 17 :3, 243-276.

[11] STERN H. et COVER T. (1989), Maximum Entropy and the Lottery, *Journal of the American Statistical Association,* 84, n°408, 980-985.

[12] WILLMAN G. (1999), The History of Lotteries, Mimeo, Stanford University.

[13] ZIEMBA W.T., BRUMELLE S.L., GAUTHIER A. et SCHWARTZ S. (1986), *Dr Z's 6/49 Lotto Guidebook*, Dr Z Investments, Vancouver.