

Jeux de grattage

RÉSUMÉ. On étudie ici un jeu de hasard commercialisé par une société de jeux : on essaie de déterminer si la répartition des lots est effectuée selon un hasard total ou non et si une connaissance, même partielle, de cette dernière peut être exploitée pour remporter des gains plus grands.

17.1 Imaginons une observation troublante, méritant enquête...

On considère un jeu de grattage, commercialisé à l'échelle d'un pays. Une tranche d'émission du jeu consiste en la confection de 500 000 tickets, répartis en 10 000 carnets de 50 tickets. Chaque ticket coûte 3 €. Par contrat avec la puissance publique, la société de jeux annonce le nombre et la répartition des sommes gagnantes parmi chacune des tranches d'émission, ce que nous reportons au tableau 17.1.

Valeur				Nbre				Total			
40 000 €	×	1	=	40 000 €	10 €	×	9 000	=	90 000 €		
20 000 €	×	1	=	20 000 €	6 €	×	28 000	=	168 000 €		
10 000 €	×	2	=	20 000 €	4 €	×	25 000	=	100 000 €		
1 000 €	×	5	=	5 000 €	3 €	×	47 500	=	142 500 €		
500 €	×	18	=	9 000 €							
200 €	×	800	=	160 000 €							
100 €	×	850	=	85 000 €							
50 €	×	2 020	=	101 000 €							
20 €	×	4 000	=	80 000 €							
7 697				520 000 €				109 500			
								500 500 €			

TABLE 17.1. Répartition des lots (valeur et nombre d'occurrences) dans une tranche d'émission. Un ticket coûte 3 €.

On suppose que les habitants du pays sont suffisamment nombreux et jouent suffisamment pour que dix tranches environ soient émises par an. On imagine alors qu'un passionné de

statistique veuille mener une enquête et déterminer plus précisément comment sont répartis les lots à l'intérieur d'une tranche, et notamment, s'ils sont répartis selon un hasard uniforme ou si la société de jeux s'assure qu'il n'y a pas plus d'un lot significatif, i.e., de valeur supérieure ou égale à 20 €, par carnet de 50 tickets.

Allons un pas plus loin dans la description de cette situation fictive et considérons que notre détective statisticien ait acheté 100 carnets, en ayant pris soin d'étaler ses achats tout au long d'une année. Sur ces 100 carnets,

- 33 ne contenaient que des lots inférieurs ou égaux à 10 €, appelés dans la suite, lots mineurs ;
- 67 comportaient un unique lot dit significatif ;
- aucun ne permettait de gagner deux lots significatifs.

Ses conclusions en sont d'une part, que le hasard de répartition des lots significatifs n'est pas uniforme mais que la société émet des contraintes techniques fortes sur leur répartition en carnets ; et que d'autre part, il existe des stratégies de joueurs informés permettant une espérance de gain strictement positive. Ces stratégies requièrent l'observation du grattage des premiers tickets dans un carnet. Nous allons maintenant expliquer comment il parvient à ces conclusions à partir de ses observations et du tableau de répartition 17.1.

17.2 Construction et exploitation d'un modèle statistique

Puisque l'on a affaire à dix tranches d'émission, la population que nous considérons a été formée de 100 000 carnets. Nous avons prélevé un échantillon de 100 carnets et avons classé ces carnets en trois catégories, selon qu'ils contenaient aucun, un, ou deux lots significatifs. En fait, pour que l'analyse qui suit soit plus robuste, nous estimons que l'ensemble des valeurs possibles aurait été constitué par $\llbracket 0, 50 \rrbracket$, les entiers entre 0 et 50.

Si l'on note $c_j \in \llbracket 0, 50 \rrbracket$, pour $j = 1, \dots, 100$, les observations effectuées, on peut donc les modéliser comme la réalisation d'un 100-échantillon $C_1, \dots, C_{100}$ de variables aléatoires indépendantes et identiquement distribuées selon une loi ν (inconnue) sur $\llbracket 0, 50 \rrbracket$.

17.2.1 Rejet de l'hypothèse de hasard uniforme

Si les lots significatifs étaient distribués totalement au hasard dans une émission, un carnet pris au hasard en contiendrait un nombre $C \sim \text{Bin}(50, p_0)$, pour le paramètre $p_0 = 7\,697/500\,000$. En particulier, on se serait attendu à ce qu'environ $q_{\text{ref}} = 18\%$ des carnets tirés dans l'expérience aient contenu deux lots significatifs ou plus.

On montre alors que les observations contredisent gravement cette hypothèse : un test d'ajustement approprié, prenant q_{ref} comme valeur de référence, admet une p -valeur plus petite que 10^{-8} et permet donc de rejeter clairement l'hypothèse de répartition des lots selon un hasard uniforme. On peut même voir qu'il aurait suffi de beaucoup moins d'observations, d'une vingtaine seulement, pour effectuer cette infirmation. Comme les tickets ne coûtent pas trop cher, l'expérience est donc facilement reproductible.

17.2.2 Inférence sur ν

On veut alors déterminer ν plus précisément, ou, tout du moins, déterminer les trois valeurs

$$(a, b, q) = \left(\nu\{0\}, \nu\{1\}, \nu(\llbracket 2, 50 \rrbracket) \right).$$

Le test mené précédemment conduit, par la dualité entre tests et intervalles de confiance, à proposer l'encadrement $q \in [0, 3 \text{ \%}]$, qui est la réalisation d'un intervalle de confiance de niveau 95 %.

On veut maintenant minorer q . On peut voir, pour s'échauffer, que la valeur $q_{\text{ref}} = 0$ n'est pas compatible avec les observations non plus. En effet, si $q = q_{\text{ref}} = 0$, alors, au vu du tableau 17.1, on a $b = b_{\text{ref}} = b_0 = 7\,697/10\,000$ et $a = a_{\text{ref}} = 1 - b_0$. Mais, en considérant les observations

$$\mathbb{I}_{\{C_1=0\}}, \dots, \mathbb{I}_{\{C_{100}=0\}},$$

on peut construire un test rejetant cette hypothèse $q_{\text{ref}} = 0$ avec une p -valeur de 1.5 %. La solution n'est donc pas aussi simple qu'il n'y paraît.

En fait, il faut calibrer le triplet (a, b, q) de telle sorte que diverses contraintes soient vérifiées et que le triplet d'observations (33, 67, 0) ne soit pas atypique pour lui. Pour énoncer les contraintes de manière simple, on suppose que $\nu(\llbracket 3, 50 \rrbracket) = 0$. On a alors le système

$$\begin{cases} a, b, q & \geq 0 \\ a + b + q & = 1, \\ b + 2q & = b_0, \\ q & \leq 0.03, \end{cases}$$

En ajoutant en plus des inégalités sur a et b obtenues par la réalisation d'intervalles de confiance, on peut montrer que ce système admet une solution.

On en déduit par exemple la modélisation $(a, b, q) = (0.256, 0.718, 0.026)$, qui est satisfaisante à tous les égards et qui, à de petites variations près, est la seule à convenir d'un point de vue statistique, sous l'hypothèse qu'un carnet contient au plus deux lots significatifs.

17.3 Sur l'existence et la pertinence de stratégies de joueurs informés

On peut se demander si l'inférence précédente, dont le résultat est somme toute assez précis, est exploitable en termes stratégiques. On suppose, à cet effet, qu'il existe un buraliste qui, à réception d'un carnet de 50 tickets, ne le découpe pas pour mettre les tickets individuels sur un présentoir, mais accroche le carnet avec une pince à linge et le débite au fur et à mesure des ventes ; et que de plus, il arrive à suivre le résultat des gains de ses clients. Ainsi, à tout moment, lorsqu'il reste x tickets dans le carnet, il peut déterminer s'il y avait eu un lot significatif dans les $50 - x$ tickets déjà vendus.

La stratégie proposée est de la forme : s'il reste x tickets dans un carnet dans lequel aucun lot significatif n'est encore tombé, alors on achète ces x tickets. Nous allons déterminer s'il existe une valeur de x telle que cette stratégie a une espérance de gain, rapportée aux mises jouées, strictement positive, et le cas échéant, si une telle situation arrive couramment ou au contraire, rarement. L'un dans l'autre, cela nous donnera une estimation des gains finaux qu'un joueur informé par le buraliste peut espérer.

17.3.1 Etude préliminaire : gains et pertes moyens et typiques

On peut distinguer trois types de tickets : les tickets perdants, les tickets porteurs d'un lot mineur, et ceux associés à un lot significatif. Le tableau 17.1 précise la proportion de chacun. L'espérance du gain d'un ticket, conditionnellement au fait qu'il est au mieux porteur d'un lot mineur, est $m \approx 1$ €. L'espérance d'un ticket portant un gain significatif est quant à elle de 68 €.

En fait, les gains supérieurs ou égaux à 500 € sont tellement rares et perturbent tant le calcul d'une valeur typique de gain significatif qu'on préfère ne pas tenir compte d'eux. On préfère considérer le gain moyen M d'un ticket conditionnellement au fait qu'il porte un gain significatif entre 20 € et 200 € ; ce gain vaut $M \approx 55$ €.

17.3.2 Espérance de solde de gain strictement positive pour x suffisamment petit

On se place ici dans la situation où il reste x tickets à vendre dans un carnet et où aucun lot significatif n'est encore sorti. Le client informé achète alors les x tickets restants, ce qui lui coûte $3x$ € et on veut déterminer l'espérance du solde de son gain (les sommes gagnées moins les sommes mises).

Concernant l'espérance de ses gains, on va mener les calculs en distinguant selon que le carnet contient aucun, un ou deux lots significatifs. Dans tous les cas, on émet l'hypothèse que le ou les lot(s) significatif(s) est (sont) placé(s) au hasard dans le carnet. Ainsi, tout carnet contenant zéro lot significatif arrive à l'étape où il reste x tickets à vendre sans qu'aucun lot ne soit encore sorti, de même qu'une fraction $x/50$ des carnets contenant exactement un tel lot et une fraction d_x de ceux avec deux lots significatifs, où il faut préciser d_x . Les proportions initiales entre ces trois catégories étant a , b et q , on en déduit les proportions réalisées conditionnellement au fait qu'aucun lot significatif ne soit sorti dans les $50 - x$ premiers tickets.

Avec les notations précédentes, l'espérance du gain dans cette situation, déduction faite de la mise, est alors de la forme

$$e_x = \frac{1}{r_x} \left(axm + b(x/50)(M + (x-1)m) + qd_x(2M + (x-2)m) \right)$$

pour $x \geq 2$, où r_x est un facteur de normalisation à préciser. L'espérance du solde vaut quant à lui $s_x = e_x - 3x$.

17.3.3 Mise en œuvre de la stratégie

En pratique, on pourrait recommander la valeur $x = 4$. Une proportion significative (supérieure à 30 %) de l'ensemble des carnets arrive au moment où il reste 4 tickets à vendre sans qu'un lot significatif ne soit encore sorti. L'espérance du gain, mise déduite, est légèrement supérieure à 2 € dans ce cas. Le client informé par un buraliste vendant plusieurs carnets par semaine, disons 20, peut alors espérer un gain de 12 € au moins par semaine.

17.4 Suggestions de développements

17.4.1 Points du texte à développer en priorité

On détaillera la construction du modèle statistique et mettra en évidence les approximations éventuellement réalisées à cet effet. On mettra en œuvre de façon soigneuse le test du paragraphe 17.2.1 et le calcul, par dualité, des intervalles de confiance du paragraphe 17.2.2.

17.4.2 Détermination des gains typiques

Pour le calcul de M , une manière de montrer que l'on peut attendre des gains allant jusqu'à 200 €, mais pas au-delà, est d'étudier les lots obtenus par un joueur grattant deux tickets par semaine : quelle est la loi de son gain maximal en un an, en dix ans ?

