Vignettes Panini

Arnaud Legrand, Jean-Marc Vincent

RICM4, Probabilités et Simulations

Ce devoir est à faire en binôme. Pour effectuer ce devoir, vous utiliserez `R` et `Rstudio`. Vous publierez vos observations sur `http://rpubs.com` et vous enverrez l'url à `arnaud.legrand@imag.fr` avant le 6 novembre à minuit en indiquant la chaîne [RICM4-PS] dans le sujet du message.

Votre petit cousin collectionne les vignettes autocollantes des footballeurs célèbres. Il achète au marchand de journaux des pochettes de 10 vignettes à un prix de 2 euros. L'album de la collection complète contient $M = 300$ places distinctes pour les vignettes autocollantes (prix d'achat de l'album 4 euros). Il souhaite savoir combien de pochettes il devra acheter pour avoir toute la collection. Saurez vous lui répondre ?
Dans un premier temps, on simplifie le problème en supposant que l'on achète les vignettes une par une et on note T le nombre total de vignettes achetées pour avoir toute la collection. À la fin, il restera plein de doubles ! On supposera également que l'éditeur des vignettes est honnête et qu'il imprime les vignettes dans les mêmes proportions (il n'y a pas de vignettes plus rares que d'autres).

Approche expérimentale

Q1) Écrire un simulateur en `R` générant T en fonction de M.

Q2) Étudier expérimentalement la valeur moyenne de T en fonction de M. On prendra des échantillons de l'ordre de 100 ou 1000 et une dizaine de valeurs M entre 10 et 200. Une représentation graphique sera appréciée... Que remarquez vous ?

Modélisation

On modélise la séquence des vignettes achetées par une suite de variables aléatoires X_1, X_2, ..., X_n à valeurs dans $\{1, \ldots, M\}$.

Q3) Quelles sont les hypothèses statistiques sur la suite des variables X_n que l'énoncé suggère ?

Q4) Lorsque la collection comporte déjà $i-1$ vignettes, on note Y_i le nombre de vignettes à acheter pour avoir une vignette qui n'est pas dans la collection. Donner, en la justifiant, la loi de Y_i (vous pourrez calculer la probabilité $P(Y_i = k)$ et vous en déduirez l'espérance et la variance de Y_i). Les variables Y_i sont-elles indépendantes ? En déduire $\mathbb{E}(T)$ et $Var(T)$ ainsi que des équivalents quant M est grand.

Synthèse

Q5) Vérifier graphiquement que les résultats obtenus par simulation sont cohérents avec vos calculs.

Q6) Au fait, combien votre petit cousin devra-t-il dépenser en moyenne sachant que chaque paquet de 2 euros contient 10 vignettes et que l'album comporte 300 places ?

Extension au cas non uniforme (ou au marchand malhonnête)

Q7) On s'intéresse maintenant au cas où la loi de probabilité d'obtenir chaque vignette n'est plus uniforme. Adapter votre simulateur afin d'étudier les deux cas suivants :
- La probabilité d'obtenir la carte i est égale à $p_i = \frac{1}{\alpha i}$ avec $\alpha = \sum_{k=1}^{M} \frac{1}{i}$.
- La probabilité d'obtenir la carte i est égale à $p_i = \frac{1}{\alpha i^2}$ avec $\alpha = \sum_{k=1}^{M} \frac{1}{i^2}$.

Que remarquez-vous quant à l'influence de cette distribution sur la dépendance de T par rapport à M ?

Q8) Selon vous, que se passera-t-il pour une simulation avec $p_i = \frac{1}{\alpha 2^i}$ et $\alpha = \sum_{k=1}^{M} \frac{1}{2^i}$?