TP 4 : Etude statistique du Loto

L'objectif de ce TP est d'aborder le jeu du Loto d'un point de vue statistique. Cette étude visera deux objectifs :

1. La prédiction à partir des tirages précédents de certains numéros potentiellement intéressants pour le prochain tirage
2. La réalisation d'une simulation du jeu pour obtenir un apperçu des gains pour un joueur donné

Le principe du Loto est le suivant. Le joueur choisit 6 numéros sur une grille qui en compte 49 allant de 1 à 49. La "Française Des Jeux", qui organise le jeu, réalise un tirage de 6 numéros parmi les 49 numéros suivis d'un numéro complémentaire, soit 7 numéros à chaque tirage. Lorsqu'un joueur trouve exactement les 6 premiers numéros du tirage, on dit qu'il est gagnant au rang 1. S'il trouve 5 numéros parmi les 6 premiers et que son 6^{e} numéro correspond au numéro complémentaire, on dit qu'il est gagnant au rang 2. Si au contraire, il n'a pas le numéro complémentaire, on dit qu'il est gagnant de rang 3, et ainsi de suite. A chaque rang correspond un gain en argent, le rang 1 étant celui qui rapporte le plus.

1 Les numéros du Loto

Nous allons commencer par étudier la distribution des numéros et tenter une prédiction des numéros sur lesquels il vaut mieux miser et ceux qu'il vaut mieux éviter.

1.1 La "Forme" des numéros

On appelle "Forme" d'un numéro le nombre de fois que ce numéro est sorti au cours des 20 derniers tirages. Nous désignerons par X la variable aléatoire donnant le nombre d'apparitions d'un numéro donné au bout de 20 tirages.

Q1. Déterminez le nombre total de tirages à 7 numéros possibles. Calculez le nombre de tirages possibles lorsqu'un numéro est fixé. Déterminez la probabilité p d'apparition d'un numéro donné lors d'un tirage.

Q2. Téléchargez le fichier `Loto-1.txt` à l'adresse
`http://epoc.isima.fr/~diarrassouba/enseignements.htm`. Ce fichier donne pour chaque numéro le nombre d'apparitions au cours des 20 derniers tirages. Lisez les données qu'il contient et tracez l'histogramme correspondant aux nombres d'apparition. Vous laisserez "R" calculer le nombre de classes de l'histogramme. De quelle loi semble se rapprocher cet histogramme.

Q3. Sachant que la probabilité $P(Y = k) = C_n^k p^k (1-p)^{n-k}$ où Y suit une loi binomiale de paramètres n et p, écrivez une fonction `bin(n,p,k)` qui calcule la probabilité $P(Y = k)$.

Déterminez grâce à cette fonction les probabilités $P(Y = k)$ pour $k = 0, ..., 10$ avec $n = 20$ et p étant la valeur calculée à la question 1. Calculez les effectifs théoriques correspondants à chaque valeur de k.

Q4. Déterminez à partir des données du fichier, les effectifs observés pour chaque valeur de $k = 0, ..., 10$ et comparez-les aux effectifs obtenus à la question précédente. Que peut-on en conclure quant à distribution de la variable X ?

Q5. Dites quels sont, d'après les données observées, les numéros qui sont en "bonne forme" c'est-à-dire qui sont apparus au moins 4 fois sur les 20 derniers tirages.

1.2 Les apparitions depuis le début

Nous allons maintenant déterminer la distribution des numéros depuis le début du jeu. Téléchargez le fichier `Loto-2.txt` qui donne pour chaque numéro le nombre total d'apparitions ainsi que l'écart (nous verrons plutard de quoi il s'agit) des numéros depuis le premier tirage jusqu'au tirage du 11 Octobre 2006. Le nombre de tirages correspondant est $N = 4479$. Nous désignerons par Y la variable aléatoire donnant le nombre d'apparitions d'un numéro depuis le début du jeu. Lisez les données contenues dans le fichier `Loto-2.txt`. Dans la suite, on désignera par `y` l'échantillon donné par ce fichier.

Q1. Déterminez les numéros qui sont apparus le plus souvent et le moins souvent ainsi que leurs nombres respectifs d'apparitions.

Q2. Triez les numéros par ordre décroissant d'apparition (utilisez la fonction `sort()` en précisant `index.return = TRUE`) et donnez les 10 numéros les plus fréquents depuis le début du jeu.

Q3. Tracez l'histogramme associé aux nombres d'apparitions. En appliquant le théorème central limite, dites vers quelle loi la variable Y doit s'approcher. Déterminez les paramètres théoriques μ et σ de cette loi.

Q4. Nous supposerons que la variable Y suit une loi normale dont la moyenne μ et l'écart-type σ sont ceux calculés à la question 3. Calculez avec la fonction `pnorm()` les probabilités suivantes : $P(Y \leq x_1)$, $P(Y \geq x_2)$ et $P(x_1 \leq Y \leq x_2)$ avec $x_1 = \mu - \sigma$ et $x_2 = \mu + \sigma$ (attention au sens des inégalités et à la valeur renvoyée par `pnorm()`). Déterminez les numéros dont les nombres d'apparitions sont respectivement $\leq x_1$ et $\geq x_2$.

1.3 Les écarts des numéros

Les numéros déterminés précédemment sont-ils les plus intéressants pour un joueur ? Pour en avoir une idée, nous tenons aussi compte des écarts des numéros c'est-à-dire du nombre de tirages successifs sans sortie. Prenons comme référence le tirage N. Un écart de 1 signifie que le numéro est sorti à l'avant dernier tirage mais n'est pas sorti au tirage N. Un écart de 2 signifie que le numéro est sorti au tirage $N - 2$ mais pas aux tirages $N - 1$ et N. Un écart de 0 signifie que le numéro est sorti au dernier tirage (le tirage N). On modélisera les écarts par une variable aléatoire discrète E.

Dans la suite, on déterminera la distribution de E et on s'en servira pour le choix de quelques numéros intéressants. L'écart pour chaque numéro est donné par la colonne "ET" du fichier `Loto-2.txt`.

Q1. Déterminez les écarts minimum et maximum observés depuis le dernier tirage. Tracez l'histogramme associé aux écarts. Comment évoluent les effectifs lorsque l'écart augmente ?

Q2. Sachant que les tirages sont indépendants, déterminez la probabilité théorique $P(E = k)$. Calculez avec la formule que vous obtenez la probabilité $P(E = k)$, pour $k = 0, ..., 25$.

Q3. Calculez les effectifs théoriques à partir des probabilités calculées à la question précédente et comparez-les aux effectifs observés dans le fichier pour chaque valeur de $k = 0, ..., 25$.

Q4. Quels sont les numéros qui sont restés au moins 15 fois de suite sans sortir ?

Q5. Au vu des probabilités déterminées à la question 3, que pensez-vous des chances de sortie au prochain tirage pour un numéro qui est resté longtemps sans sortir ?

2 Simulation des gains du Loto

Nous allons maintenant réaliser une simulation du Loto et estimer les gains probables. Pour recevoir une partie des gains au Loto, il faut que le joueur trouve au moins trois numéros parmi les 6 principaux et éventuellement le numéro complémentaire. La "Française Des Jeux" reverse à l'ensemble des gagnants 50.485% des sommes totales jouées. Cette somme est distribuée aux différents gagnants en fonction de leurs rangs. Le gain total obtenu dans chaque rang est ensuite divisé par le nombre de gagnants de ce rang. Pour simplifier, nous supposerons que le nombre $M = C_{49}^{6} = 13\ 983\ 816$ de grilles sont jouées à chaque tirage à raison de 0.60 € par grille, et que le nombre de gagnants dans chaque rang est égal au nombre de grilles possibles dans ce rang. Les gains par grille sont alors les suivants :

6 bons numéros	: gain au rang 1 $\rightarrow$ 1 923 070 €
5 bons numéros + complémentaire	: gain au rang 2 $\rightarrow$ 21 885 €
5 bons numéros	: gain au rang 3 $\rightarrow$ 1 731€
4 bons numéros + complémentaire	: gain au rang 4 $\rightarrow$ 81 €
4 bons numéros	: gain au rang 5 $\rightarrow$ 40.62 €
3 bons numéros + complémentaire	: gain au rang 6 $\rightarrow$ 9.54 €
3 bons numéros	: gain au rang 7 $\rightarrow$ 4.42 €
0-1-2 bons numéros	: gain au rang 8 $\rightarrow$ 0 €

Pour simuler chaque tirage du Loto, nous allons attribuer à chaque grille un numéro compris entre 1 et M, et diviser l'intervalle $[0, M]$ en 8 tranches. Chaque tranche représentera un rang et sera de longueur le nombre de grilles correspondant au rang. Les tranches sont les suivantes :

0-1	: rang 1 (1 grille possible)
2-7	: rang 2 (6 grilles possibles)
8-259	: rang 3 (252 grilles possibles)
260-889	: rang 4 (630 grilles possibles)
890-13804	: rang 5 (12 915 grilles possibles)
13805-31024	: rang 6 (17 220 grilles possibles)
31025-260624	: rang 7 (229 600 grilles possibles)
260625-13983816	: rang 8 (13 983 816 grilles possibles)

Les tirages se font ensuite en tirant au hasard un nombre compris entre 1 et M et suivant une loi discrète uniforme (les grilles sont équipobables).

Q1. Ecrivez une fonction `Rang(m)` qui prend en paramètre un entier `m` et retourne le rang (entier entre 1 et 8) associé à la grille numéro `m`.

Q2. Construisez un vecteur `gains` de longueur 8 et qui contient les valeurs de gains associés à chaque rang. Exemple : gains[1] est la valeur du gain au rang 1.

Q3. Ecrivez une fonction `Loto(n,vGain)` qui réalise `n` tirages du Loto et retourne le gain total obtenu. Le paramêtre `vGain` sera le vecteur `gains` créé à la question précédente.

Q4. La fonction `Loto(n,vGain)` sert à réaliser `n` tirages du Loto pour un joueur donné. Ecrivez une fonction `SimulationLoto(N,n,vGain)` qui réalise, pour `N` joueurs, une série de `n` tirages du Loto et renvoie dans un vecteur de taille `N` contenant les gains totaux obtenus pour chacun des `N` joueurs.

Q5. Réalisez une simulation du jeu pour `N = 2000` joueurs jouant chacun `n = 1000` grilles (soit 20 grilles par semaine pendant 1 an). Stockez le résultat dans un vecteur que vous nommerez `simuLoto`.

Q6. Déterminez avec la fonction `summary()` les gains minimum et maximum, les gains au $\frac{1}{4}$ et au $\frac{3}{4}$ de la plage des valeurs. Déterminez aussi le gain moyen sur `N` joueurs.

Q7. Déterminez l'ensemble des gains ≤ 2500. Combien de gains y a-t-il ? Tracez l'histogramme correspondant en précisant comme vecteur de classes le vecteur
`c(20*c(0 :50),1000 + 300*c(1 :5))` (c'est-à-dire le vecteur (0,20,40,...,980,1000,1300,1500, ...,2200,2500)). Comment sont répartis les gains ?

Q8. Déterminez, sur l'ensemble de la simulation, des intervalles de confiance à 95% pour les proportions suivantes :

- la proportion de joueurs ayant obtenus un gain inférieur à 600 €
- la proportion de joueurs ayant obtenus un gain ≥ 600€.

Q9. Avec les résultats de cette simulation, que pensez-vous des chances de rentabiliser l'investissement pour une personne jouant 1000 grilles ?