Correction TP 4 : Etude statistique du Loto

1 Les numéros du Loto

1.1 La "Forme" des numéros

Q1. Le nombre de tirages à 7 numéros possibles correspond au nombre de combinaisons de 7 éléments parmi 49, soit C_{49}^7. Un fois qu'un numéro est fixé, il reste 6 numéros à choisir parmi 48. Donc le nombre de tirages possibles avec un numéro fixé est C_{48}^6.

La probabilité p d'apparition d'un numéro lors d'un tirage est alors le nombre de tirages possible contenant ce numéro divisé par le nombre total de tirages possibles, soit $p = \frac{C_{48}^6}{C_{49}^7} = \frac{1}{7}$.

Remarque : utilisez la fonction `choose(n,k)` pour calculer le nombre de combinaisons à k éléments parmi n.

Q2. L'histogramme correspond à celui d'une loi binomial de paramètres $n = 20$ et $p = \frac{1}{7}$. En effet, à chaque tirage, l'apparaition ou non d'un numéro correspond à une épreuve de Bernouilli avec une probabilité $p = \frac{1}{7}$ de réussite et une probabilité $1 - p$ d'échec. Le nombre d'apparition des numéros au cours des 20 derniers tirages correspond à la somme des résultats de ces 20 épreuves de Bernouilli. Il s'agit donc d'une loi Binomiale de paramêmtre $n = 20$ et $p = \frac{1}{7}$.

Q3. La probabilité $P(Y = k) = C_n^k p^k (1-p)^{n-k}$ pour une loi binomiale de paramètres n et p. La fonction `bin(n,p,k)` s'écrit :

```
bin ← function(n,p,k)
{
x = choose(n,k)*p^ 2*(1-p)^(n-k) ;
return(x)
}
```

k	$P(Y = k)$	Eff Théor	Eff Obs
k = 0	0.045	2.45	4
k = 1	0.152	7.48	6
k = 2	0.242	11.85	11
k = 3	0.242	11.85	12
k = 4	0.171	8.39	9
k = 5	0.091	4.47	3
k = 6	0.038	1.86	3
k = 7	0.012	0.62	1
k = 8	0.003	0.16	0
k = 9	0.0007	0.037	0
k = 10	0.00014	0.0068	0

Q4. On voit que les effectifs observés sont proches des effectifs théoriques. Comme on l'a vu précédemment la loi X suit une loi binomiale.

Q5. Les numéros sortis au moins 4 fois sur les 20 derniers tirages sont :

```
x$No[x$NbApp >= 4]
4 13 15 20 21 27 30 34 35 36 38 41 43 44 46 49
```

1.2 Les apparitions depuis le début

Q1.

```
y$No[y$TOT == min(y$TOT)]
17
y$TOT[y$TOT == min(y$TOT)]
588
y$No[y$TOT == max(y$TOT)]
45
y$TOT[y$TOT == max(y$TOT)]
649
```

Q2.

```
 z = sort(y$TOT,index.return=TRUE)
 z$ix
[1] 17 29 33 47 3 8 12 14 11 48 21 25 2 13 10 41 22 26 24 28 43 44 9 20 23
[26] 40 15 18 27 32 37 5 49 35 30 46 39 42 31 19 1 38 6 7 36 4 16 34 45
```

Q3. D'après le théorème centrale limite, la loi variable Y converge vers une loi normale de moyenne $\mu = Np$ et d'écart-type $\sigma = \sqrt{Np(1-p)}$. Comme $N \gg 30$, on supposera que Y suit une loi normale de moyenne $\mu = Np = 639.8571$ et d'écart-type $\sigma = 23.41899$.

Q4.

```
> mu = N*p
> sigma = sqrt(N*p*(1-p))
> x1 = mu - sigma = 616.4382
> x2 = mu + sigma = 663.2761
> P1 = pnorm(x1,mean=mu,sd=sigma)
> P1
> [1] 0.1586553 > P2 = 1 - pnorm(x2,mean=mu,sd=sigma)
> P2
> [1] 0.1586553
> P3 = 1 - P1 - P2
> P3
> [1] 0.6826895

> y$TOT[y$TOT <= x1]
[1] 615 588 596 602 613
> y$TOT[y$TOT >= x2]
[1] 664 672 667 670 678 680 671 665 689
```

1.3 Les écarts des numéros

Q1.

```
> min(y$ET)
0
> max(y$ET)
25
```

Plus les écarts augmentent, plus les effectifs diminuent.

Q2. Un ecart de k signifie que le numéro est sortie est sorti au tirage $N-k$ et n'est pas sortie pendant les tirages $N-k-1$ jusqu'à N (k tirages). La probabilité d'apparition au tirage $N-k$ est p et celle de non sortie est $1-p$. Donc $P(E=k)=p(1-p)^k$.

```
> k = c(0 :25)
> k
> pasc(p,k)
[1] 0.142857143 0.122448980 0.104956268 0.089962516 0.077110728 0.066094909
[7] 0.056652780 0.048559525 0.041622450 0.035676386 0.030579759 0.026211222
[13] 0.022466762 0.019257225 0.016506192 0.014148165 0.012126999 0.010394570
[19] 0.008909632 0.007636827 0.006545852 0.005610730 0.004809197 0.004122169
[25] 0.003533288 0.003028532
```

Q3.

```
> pasc(p,k)*49
[1] 7.0000000 6.0000000 5.1428571 4.4081633 3.7784257 3.2386506 2.7759862
[8] 2.3794167 2.0395001 1.7481429 1.4984082 1.2843499 1.1008713 0.9436040
[15] 0.8088034 0.6932601 0.5942229 0.5093339 0.4365719 0.3742045 0.3207467
```

Les effectifs observés s'obtiennent avec la commande `length(y$ET[y$ET == 0])`

```
7 5 6 3 6 2 4 1 4 1 1 0 1 0 1 2 0 0 1 1 0 1 1 0 0 1
```

Q4.

```
y$No[y$ET >= 15]
```

Q5. En observant les tendances, on voit qu'un numéro a peu de chance de rester longtemps sans sortir. En particulier, en théorie moins d'un numéro sur 49 restera 13 fois sans sortir. Donc un joueur peu considérer que les numéros ayant un grand écart aura plus de chance de sortir au prochain tirage. Un autre joueur peut aussi considérer qu'un numéro avec un grand écart est à éviter.

2 Simulation des gains du Loto

Q1.

```
function(v)
{
if(v <= 1)
return(1)
else if(v <= 7)
return(2)
else if(v <= 259)
return(3)
else if(v <= 889)
return(4)
else if(v <= 13804)
return(5)
else if(v <= 31024)
return(6)
else if(v <= 260624)
return(7)
else if(v <= 13983816)
return(8)
else return(0)
}
```

Q3.

Q4.

```
function(n,vGain)
{
total = 0 ;
for(i in 1 :n)
{
num_grille = runif(1,min=1,max=13983816) ;
rg = Rang(num_grille) ;
total = total + vGain[rg] ;
}
return(total) ;
}
```

Q5.

```
function(N,n,vGain)
{
res = rep(0,times=N) ;
```

```
for(i in 1 :N)
{
res[i] = Loto(n,vGain) ;
}
return(res) ;
}
```

Q6. `> simu = SimulationLoto(2000,1000,gains).`

Q7. `> summary(simu).`

Q8.

```
> simu2 = simu[simu <= 2500]
> length(simu2)
> hist(simu2,breaks=c(20*c(0 :50),1000 + 300*c(1 :5)))
```

La grande partie des gagnant a obtenu un gain compris entre 0 et une valeur ≤ 1000 avec un pic entre ces deux valeurs. Très peu de joueurs ont obtenu un gain ≥ 2500. Aucun gagant de rang 1 n'a été obtenu.

Q8. Intervalle de confiance à 95% :

```
> p1_c = length(simu[simu <= 600]) / 2000
z = abs(qnorm(0.025))
> p1_c - z*sqrt(p1_c*(1-p1_c)/2000)
> p2_c + z*sqrt(p1_c*(1-p1_c)/2000)

> p2_c = length(simu[simu > 600]) / 2000
> p2_c - z*sqrt(p2_c*(1-p2_c)/2000)
> p2_c + z*sqrt(p2_c*(1-p2_c)/2000)
```

Q9. Pour rentabiliser son invetissement dans le jeu du Loto il faudrait que les gains obtenus avec 1000 grilles soit > 600 € (1000*0.6 €). Même si plusieurs simulations peuvent donnés des résultats différents les chances d'obtenir un gain ≥ 600 € restent faibles.