Université de Picardie Jules Verne **2010-2011**
UFR des Sciences - LAMFA

PAF - Formation Enseignement des Mathématiques - 28 janvier 2011

Mathématiques : statistiques et simulation

Dans ce document, je vous propose une deuxième liste d'exercices.

Exemple de loi d'échantillonnage

Exercice 1. On considère une population $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ de $N = 5$ personnes, où $\omega_1, \omega_4, \omega_5$ sont des femmes et ω_2, ω_3 sont des hommes. On s'intéresse au caractère statistique C "sexe des individus" à deux modalités : M : masculin et $\overline{M}$: féminin.

1) Quelle est la proportion p d'hommes dans cette population ?

2) On munit la population Ω de l'équiprobabilité (correspondant au choix aléatoire d'une personne dans la population), et on considère une variable aléatoire X de loi de Bernoulli $\mathcal{B}(p)$,

$$\text{c'est-à-dire } X : \begin{cases} \Omega & \longrightarrow \quad \mathbb{R} \\ \omega_i & \mapsto \quad \begin{cases} 1 \text{ si } \omega_i \text{ présente la modalité } M \\ 0 \text{ sinon} \end{cases} \end{cases}$$

On choisit successivement au hasard, avec remise, $n = 2$ personnes parmi les N personnes de la population.

Combien d'échantillon de taille 2 peut-on ainsi constituer ?

3) On définit alors la variable aléatoire S_n qui à tout échantillon de taille $n = 2$ associe le nombre de personnes de l'échantillon présentant la modalité M,

$$\text{c'est-à-dire } S_n : \begin{cases} \Omega^n = \Omega^2 & \longrightarrow \quad \{0,1,\ldots,n\} = \{0,1,2\} \\ (\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_n}) = (\omega_{i_1}, \omega_{i_2}) & \mapsto \quad \sum_{k=1}^{n} X(\omega_{i_k}) = X(\omega_{i_1}) + X(\omega_{i_2}) \end{cases}.$$

A l'aide d'un tableur (voir copie d'écran ci-dessous)

a) Compléter le tableau représentant l'application S_n.

b) En déduire la loi de probabilité de S_n.

c) Vérifier que S_n suit bien la loi Binomiale $\mathcal{B}(n,p)$.

d) Vérifier sur cet exemple que $E(S_n) = np$ et $Var(S_n) = np(1-p)$.

e) Déterminer la loi de probabilité de la fréquence d'échantillon $F_n = \dfrac{S_n}{n}$.

f) Vérifier sur cet exemple que $E(F_n) = p$ et $Var(F_n) = \dfrac{p(1-p)}{n}$.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Exercice 1**												
2													
3	N_1								Loi de probabilité de S_n				
4	N_2												
5	N								x_i	0	1	2	
6	p		n						p_i				
7													
8	Représentation de S_n								Loi Binomiale B(n,p)				
9													
10	Ω		ω_1	ω_2	ω_3	ω_4	ω_5		x_i	0	1	2	
11		X	0	1	1	0	0		p_i				
12	ω_1	0											
13	ω_2	1											
14	ω_3	1											
15	ω_4	0											
16	ω_5	0											
17													

Fluctuation d'échantillonnage

Exercice 2. Référendum

1) Lors d'un référendum, un sondage aléatoire simple pratiqué sur 1000 personnes a donné 55 % pour le Oui et 45 % pour le Non. Peut-on prévoir le résultat du référendum ?

2) Si, pour un référendum, on sait que "oui" se situe autour de 50 %, combien de personnes faudrait-il interroger pour que la proportion de "Oui" soit connue à 1 % près ? (en plus ou en moins)

Exercice 3. Contrôle qualité

1) Une personne est chargée de contrôler la qualité des pièces produites par une machine. Sur 1500 pièces prises au hasard dans la production, il en trouve 30 défectueuses. Quelle est la proportion de pièce défectueuse dans la production ?

2) Un grossiste en fourniture de bureau revend des rouleaux de ruban adhésif transparent et affirme que seulement 0,8 % des rouleaux présente un défaut de jaunissement du papier. Un client achète 500 rouleaux et constate que 6 rouleaux, soit 1,2% des rouleaux, jaunissent le papier. Le client peut-il faire une réclamation auprès du grossiste ?

Exercice 4. Génétique

1) Dans un pays, sur 429 440 naissances, on a dénombré 221 023 garçons. Ce résultat est-il conforme à l'hypothèse selon laquelle il y a 50 % de naissances masculines (et donc 50% de naissance féminines) ?

2) Sur 200 plantes examinées, on en compte 134 d'un phénotype "A" et 66 d'un phénotype "a". Peut-on admettre la loi de Mendel, prévoyant les proportions respectives 3/4 et 1/4 des deux phénotypes ?

Exercice 5. Places de parking

Une enquête préliminaire à l'implantation d'un supermarché dans une petite ville a montré que, sur 820 familles interrogées au hasard, 270 envisageraient d'y aller le samedi après-midi. Sachant que la ville compte 2 000 familles, quel est le nombre de places de parking à prévoir pour que dans 95 % des cas il ne soit pas complet le samedi après-midi ?

Exercice 6. L'affaire Castaneda contre Partida

En Novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison pour cambriolage d'une résidence et tentative de viol. Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79,1% de la population du comté était d'origine mexicaine, sur les 870 personnes convoquées pour être jurés lors des 11 années précédentes, seules 339 d'entre elles étaient d'origine mexicaine.

Lors du procès, un statisticien produisit des arguments pour convaincre la Cour Suprême du bien fondé de la requête de l'accusé (dont les juges votèrent à 5 contre 4 en faveur de la requête).

En vous situant dans le rôle de ce statisticien, produisez à votre tour des calculs, des raisonnements, des graphiques... pour montrer que le hasard ne peut pas « raisonnablement » expliquer à lui seul la sous représentation des américains d'origine mexicaine dans les jurys de ce comté.

Exercice 7. Avril 2002

Voici un extrait d'article, publié dans le journal « Le Monde » par le statisticien Michel Lejeune, après le premier tour de l'élection présidentielle de 2002. « Pour les rares scientifiques qui savent comment sont produites les estimations, il était clair que l'écart des intentions de vote entre les candidats Le Pen et Jospin rendait tout à fait plausible le scénario qui s'est réalisé. En effet, certains des derniers sondages indiquaient 18 % pour Jospin et 14 % pour Le Pen. Si l'on se réfère à un sondage qui serait effectué dans des conditions idéales [...], on obtient sur de tels pourcentages une incertitude de plus ou moins 3 % étant donné la taille de l'échantillon [...]. »

1) Si l'on tient compte de l'incertitude liée au sondage, entre quels pourcentages pourraient se situer réellement (à 95 % de confiance) les deux candidats si le sondage donne 18 % pour l'un et 14 % pour l'autre ?

2) Représenter sur un même graphique les deux « fourchettes » calculées à la question précédente. Peut-on prévoir l'ordre des candidats ?

3) Au premier tour de l'élection présidentielle de 2002, L. Jospin a obtenu 16,18 % des voix et J.-M. Le Pen 16,86 %. Expliquer la phrase « l'écart des intentions de vote entre les candidats Le Pen et Jospin rendait tout à fait plausible le scénario qui s'est réalisé ».

Exercice 8. Effet d'un médicament

200 personnes sont atteintes de la même maladie. On a administré du sérum à la moitié d'entre elles (groupe 1), mais pas aux autres (groupe 2, dit groupe de contrôle) ; les deux groupes sont traités de la même façon. Le nombre de guérisons dans le groupe 1 est de 75, alors qu'il n'est que de 65 dans le groupe 2. Peut-on en déduire que le sérum aide à la guérison ?

Convergence de la fréquence d'échantillon vers la probabilité

Exercice 9. Approximation de π par la méthode de Monte-Carlo

Ce calcul d'une valeur approchée de π consiste à tirer au hasard de nombres x et y dans l'intervalle $[0;1]$, autrement dit à simuler des valeurs de deux variables aléatoires indépendantes X et Y. Si $x^2+y^2 \leq 1$, le point $M(x,y)$ appartient à un quart de disque de rayon 1 : la probabilité pour qu'il en soit ainsi est $P(X^2+Y^2 \leq 1) = \frac{\pi}{4}$ (résultat admis), ce qui correspond aussi au rapport des aires d'un quart de disque de rayon 1 et d'un carré de côté 1.

En simulant un grand nombre n de points, et en désignant par k le nombre de ces points situés dans le quart de cercle, on s'attend (loi des grands nombres) à ce que la fréquence $f = \frac{k}{n}$ approche la probabilité $P(X^2+Y^2 \leq 1)$. Ainsi, $4f$ approchera π. Mais attention, la convergence est lente !

On peut observer une telle simulation sur http://jpq.pagesperso-orange.fr/proba/montecarlo/

Utiliser une calculatrice ou un tableur pour obtenir une approximation de π.

Cette question a été proposée sous forme de devoir à des élèves de seconde (Journées Inter Académiques - Amiens décembre 2009 - Atelier Probabilités et Statistiques).

On considère le quart de cercle de centre O et de rayon 1 cm inscrit dans le carré de côté 1 cm et on lance une fléchette au hasard dans le carré.

Soit M le point à l'intérieur du carré correspondant à l'endroit où s'est plantée la fléchette. Soit $(x;y)$ les coordonnées du point M.

1) Donner une condition sur x et y pour que le point M soit à l'intérieur du cercle.
2) En utilisant la touche random de la calculatrice, simuler le calcul de x^2+y^2 avec x et y deux nombres aléatoires appartenant à $[0;1[$.

Noter, à chaque fois, si la fléchette est à l'intérieur du quart de cercle ou à l'extérieur.

En répétant 50 fois cette expérience, reproduire et compléter le tableau suivant :

Fléchette	A l'intérieur du quart de cercle	A l'extérieur du quart de cercle
Effectif		
Fréquence		

3) Calculer l'aire du quart de disque de rayon 1 cm, notée A_1, et l'aire du carré, notée A_2.

On démontre que la "fréquence théorique" d'apparition de la fléchette dans le quart de cercle est $f = \frac{A_1}{A_2}$. Quel nom porte cette fréquence théorique ? A quoi correspond-elle ?

Calculer cette fréquence théorique, comparer et commenter avec le résultat de la simulation.
4) Avec un niveau de confiance de 95%, donner un encadrement de f en utilisant les résultats du 2).
5) Avec un niveau de confiance de 95%, en déduire un encadrement de π.
6) Combien de lancers faudrait-il simuler pour obtenir un encadrement de π (toujours avec un niveau de confiance de 95%) à 10^{-2} près ?
7) Une simulation avec le logiciel Algobox donne les effectifs suivants :

Fléchette	A l'intérieur du quart de cercle	A l'extérieur du quart de cercle
Effectif	156 843	43 157

grâce à l'algorithme suivant :

```
1 VARIABLES
2       x EST_DU_TYPE NOMBRE
3       y EST_DU_TYPE NOMBRE
4       n EST_DU_TYPE NOMBRE
5       i EST_DU_TYPE NOMBRE
6       p EST_DU_TYPE NOMBRE
7 DEBUT_ALGORITHME
8       n PREND_LA_VALEUR 0
9       POUR i ALLANT_DE 1 A 200000
10          DEBUT_POUR
11              x PREND_LA_VALEUR random()
12              y PREND_LA_VALEUR random()
13              SI (x*x+y*y<=1) ALORS
14                  DEBUT_SI
15                      n PREND_LA_VALEUR n+1
16                  FIN_SI
17              FIN_POUR
18      AFFICHER "Le nombre intérieur au cercle est :"
19      AFFICHER n
20      p PREND_LA_VALEUR 200000-n
21      AFFICHER "Le nombre extérieur au cercle est : "
22      AFFICHER p
23 FIN_ALGORITHME
```

Donner la distribution des fréquences.

Donner un encadrement de π (toujours avec un niveau de confiance de 95%).

8) a) Expliquer le fonctionnement des lignes 9, 13, 15 et 20.

b) Modifier l'algorithme ci-contre afin qu'il réalise autant de simulations qu'un nombre qui sera entré par l'utilisateur. De plus, lorsque le point M obtenu est dans le quart de cercle, vous l'afficherez en bleu et lorsqu'il est en dehors du quart de cercle, vous l'afficherez en rouge. Joignez une capture d'écran.

c) Modifier de nouveau le programme afin de déterminer le plus petit entier i de simulations nécessaires pour trouver une fréquence proche de π à 10^{-2} près.