Année 2014-2015
Cité scolaire Paul Valéry
Mathématiques - F. Gaunard
`http://frederic.gaunard.com`

2^{nde}3 - Simulation & Echantillonnage: Annexe

Cette annexe de cours complète la dernière section du chapitre Probabilités avec des figures et des exercices.

---

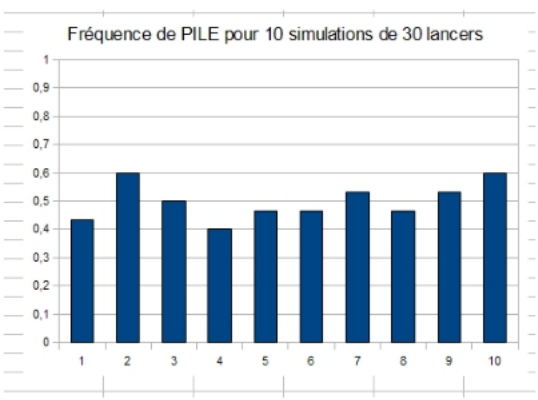
1. Simulation de n lancers

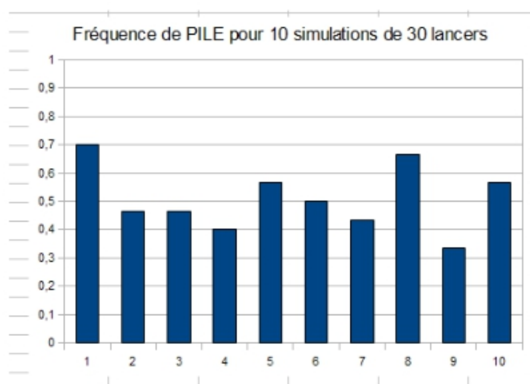
On utilise la fonction aléatoire du tableur pour *simuler* n lancers d'une pièce équilibrée. On s'intéresse à l'issue PILE. Ainsi, on dit que si l'ordinateur renvoie la valeur 1, cela correspond à PILE, s'il renvoie 0, cela correspond à FACE. Pour faire cela, on rentre dans la cellule la formule

=ALEA.ENTRE.BORNES(0;1)

(si on voulait simuler un lancer d'un dé à 6 faces, on aurait rentré ALEA.ENTRE.BORNES(1;6) par exemple). Par recopie vers le bas, on peut donc avoir une colonne de longueur n dont les cellules comportent les résultats de n lancers. Faisant cette opération en parallèle sur 10 colonnes, on obtient alors 10 échantillons de taille n. Ayant calculé la fréquence de l'apparition de PILE pour chaque échantillon (c'est à dire ici la moyenne des résultats de tous les lancers du même échantillon), il suffit ensuite d'utiliser la fonction diagramme du tableur pour afficher le diagramme à bâtons de la distribution des fréquences.

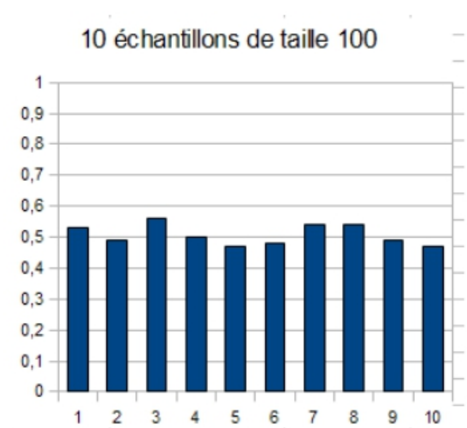
1.1. $n = 30$. On réalise l'opération précédente avec $n = 30$, deux fois. On obtient les résultats ci-dessous.





Remarque. On observe dans le premier cas une fluctuation de la fréquence des PILE entre $0,4$ et $0,6$ et dans le deuxième cas entre $0,35$ et $0,7$.

1.2. $n = 100$. Même chose mais cette fois avec $n = 100$. On voit notamment que les fréquences fluctuent cette fois entre $0,47$ et $0,55$: on se rapproche de la probabilité théorique de $0,5$.



2. Intervalles de fluctuation et de confiance

2.1. **Tester une hypothèse.** Dans une région où il y a autant d'hommes que de femmes, les entreprises sont tenues de respecter la parité. L'entreprise A a un effectif de 100 personnes dont 43 femmes alors que l'entreprise B a un effectif de 2500 personnes dont 1150 femmes.

Ainsi, on a 43% de femmes dans l'entreprise A et 46% dans l'entreprise B. On a alors l'impression que, en plus de ne pas être paritaires, les femmes sont mieux représentées dans l'entreprise B.

Si respecter la parité revient à ne pas tenir compte du caractère homme-femme, on peut a priori considérer l'ensemble des salariés de chacune des entreprises comme un échantillon prélevé au hasard dans la population de la région. Ainsi, on a deux *intervalles de fluctuation* pour chacun des échantillon.

Dans le premier $n = 100$, $p = 0,5$ donc l'intervalle est $[0,4; 0,6]$, on constate alors que $0,43$ est dans l'intervalle de fluctuation au seuil 95%, on ne peut donc finalement que valider, avec 95% de chances, le fait que l'entreprise est bien paritaire.

Par contre dans le second, $n = 2500$ et $p = 0,5$ donc l'intervalle obtenu est $[0,48; 0,52]$ et il se trouve que $0,46$ n'y est pas! Contrairement à la premiere impression, on doit donc rejeter, au risque 5%, l'hypothèse selon laquelle l'entreprise B est paritaire!

2.2. **Estimer une proportion inconnue.** Lorsqu'on *ne connait pas* la proportion réelle d'un certain caractère dans la population (par exemple lors d'un vote) on peut utiliser un *intervalle de confiance*, construit à l'aide des fréquences observées (ou dans le cas d'un vote, sur les intentions de vote relevées lors d'un sondage sur un échantillon de la population).

Par exemple, au premier tour d'une élection, trois candidats A,B,C s'affrontent. Les deux candidats placés en tête à l'issue du premier tour, participeront à un éventuel second tour (sauf en cas de majorité absolue de l'un des candidats).

Le sondage, réalisé sur un échantillon de 1000 personnes ayant le droit de vote donne les estimations suivantes: 30% des votes pour le candidat A, 28% pour le candidat B et 24% pour le candidat C. A première vue, il semblerait que les candidats A et B se retrouveront au premier tour. Estimons les intevalles de confiance, au seuil 95%, pour chacun des candidats.

On les note I_A, I_B et I_C respectivement. Comme $n = 1000$, ici la fluctuation autour de l'estimation sera de $1/\sqrt{1000} \simeq 0,032$. Ainsi, $I_A = [0,268; 0,332]$, $I_B = [0,248; 0,312]$ et $I_C = [0,208; 0,272]$.

Or, le jour du vote le candidat A comptabilise 29,9% des voix, le candidat B 26,2% et enfin le candidat C 26,9%. Par conséquent, ce sont les candidats A et C qui participeront au second tour. On pourrait alors penser que les sondages réalisés avant l'election ne sont pas valides, mais en regardant les intervalles de confiance obtenus, on constate que chacun des résultats appartient bien à l'intervalle correspondant.

Il apparaît que les estimations obtenues ne sont peut être pas assez précises, cela vient de la taille de l'échantillon qui est trop petit et permet une trop grande fluctuation.

3. Exercices

Exercice 3.1. Selon un sondage auprès de 1900 personnes, 71% déclarent consommer de l'eau du robinet et 52% de l'eau en bouteille (certaines personnes consomment les deux). Déterminer les intervalles de confiance au seuil 95% liés à ces résultats.

Exercice 3.2. La répartition de chaque groupe sanguin dans le monde est indiquée ci-dessous.

Groupe sanguin	O	A	B	AB
Monde	45%	40%	11%	4%

On effectue une étude sur les groupes sanguins dans l'ethnie rueese des Oyirad. Pour cela, on réalise un prélèvement sanguin sur 1000 personnes de cette ethnie, choisies au hasard. Les résultats de ces analyses sanguines sont regroupées ci-dessous.

Groupe sanguin	O	A	B	AB
Oyirad	26%	23%	41%	11%

Au vu de ces résultats, peut-on dire que la répartition des groupes sanguins chez les Oyirad est inhabituelle? Argumenter.

Exercice 3.3. Le mildiou est une maladie qui attaque environ 60% des plants de radis lorsque ceux-ci ne sont pas traités. Un institut spécialisé fait des tests avec quatre produits de traitement en vente libre, sur des échantillons de 105 plants. Les résultats obtenus sont consigués dans le tableau suivant.

Produit de traitement	A	B	C	D
Pourcentage de plants malade	13,9	25,9	45,4	51,2

1. Déterminer l'intervalle de fluctuation au seuil 95% lié à la proportion de plants malades du mildiou.
2. Quels sont les produits que l'on peut supposer efficaces au seuil 95%?

Exercice 3.4. On considère l'intervalle de fluctuation au seuil 95%: $[0,49875; 0,56125]$. Déterminer la taille de l'échantillon et la proportion de la valeur du caractère étudié.

Exercice 3.5. Deux candidats sont présents au deuxième tour de l'élection municipale de Maths City, où le nombre de votants est de 10000. Trois sondages ont été faits la semaine précédent le scrutin. Les résultats sont les suivants.

	Candidat A	**Candidat B**
Sondage 1: 800 personnes	55%	45%
Sondage 2: 950 personnes	52%	48%
Sondage 3: 1000 personnes	53%	47%

On suppose que toutes les personnes interrogées ont répondu honnêtement. Le candidat A est-il certain de l'emporter? Argumenter.