Contrôle : Statistique inférentielle

Patrick Tardivel, Xavier Dupuis, Université de Bourgogne

16/05/2024

Remarques : Calculatrice et formulaire autorisés. Durée 3 heures. Les exercices sont notés sur 16 et la clarté de la rédaction est un pourcentage d'augmentation compris entre 0% et 25%. La note sur 20 est donnée par la formule : (note sur 16)×(1−pourcentage).

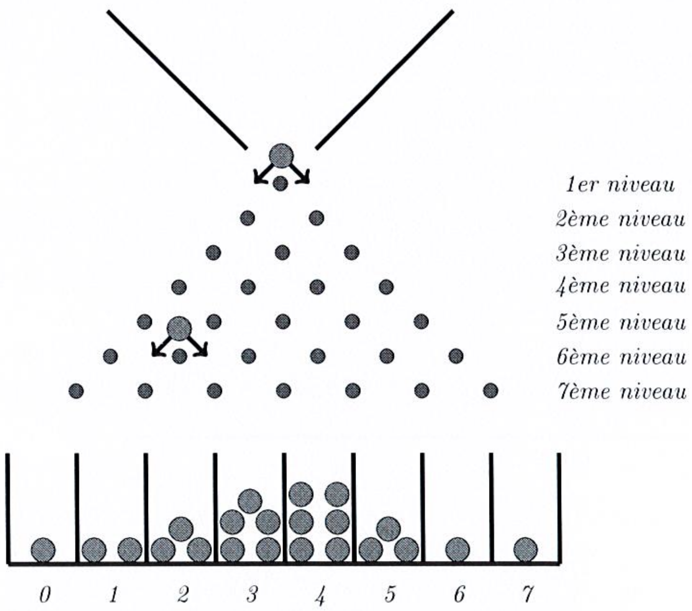
Exercice 1. *(2.5 points) On considère $(X_1, \ldots, X_n)$ un $n-$échantillon d'une loi d'espérance $\mu \in \mathbb{R}$ inconnue et de variance $\sigma^2 > 0$ inconnue modélisant les poids de léopards des mers femelles adultes choisies aléatoirement dans la population.*

1. *Rappeler la démarche mathématiques permettant de construire un intervalle de confiance asymptotique pour μ au niveau $1-\alpha$.*
2. *Pour* 150 *individus le poids moyen est* 478 kg *et la variance des poids est de* 1032 kg^2. *Calculer l'intervalle de confiance expérimental pour μ de niveau asymptotique* 0.90.

Exercice 2. *(3 points) Deux types de médicaments A et B sont administrés à des malades ayant subi une opération. Sur 554 malades ayant absorbé le médicament A, 351 ont déclaré leur douleur atténuée. Sur 687 malades ayant absorbé le médicament B, 408 ont déclaré leur douleur atténuée. Est-ce que les médicaments ont une efficacité différente ?*

1. *Faire un test statistique au niveau 5% pour conclure en rédigeant soigneusement chaque étape.*
2. *Calculer la p-valeur.*

Exercice 3. *(2.5 points) Une bricoleuse a construit une planche de Galton, schématisée ci-après, contenant sept niveaux.*



Si sa planche est parfaite, la position aléatoire d'une bille devrait suivre une loi binomiale $\mathcal{B}(7, 0.5)$. La bricoleuse a testé sa planche avec cinq milles billes et a obtenu les résultats suivants

Position	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Nombre	*39*	*274*	*848*	*1386*	*1387*	*789*	*234*	*43*

En effectuant un test statistique peut-on affirmer que la planche modélise correctement une loi binomiale $\mathcal{B}(7, 0.5)$? On prendra un risque de première espèce de 5% (le calcul de la p-valeur n'est pas demandé).

Pour information, une variable aléatoire S de loi $\mathcal{B}(7, 0.5)$ vérifie :

k	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
$\mathbb{P}(S = k)$	*0.0078*	*0.0547*	*0.1641*	*0.2734*	*0.2734*	*0.1641*	*0.0547*	*0.0078*

Exercice 4 (Comparaison de méthodes stochastiques d'approximation de $\sqrt{3}$; 8 points). *On note $\mathbf{1}_A$ la variable aléatoire indicatrice de l'événement A ($\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbf{1}_A(\omega) = 0$ sinon).*

Première méthode d'estimation de $\sqrt{3}$

Soit $U_1, \ldots, U_n$ une suite de variables aléatoires indépendantes de loi $\mathcal{U}(0, 1)$. Comme $\mathbb{P}(U_1^2 \leq 1/3) = \sqrt{3}/3$, il est naturel d'approximer $\sqrt{3}$ via la statistique suivante :

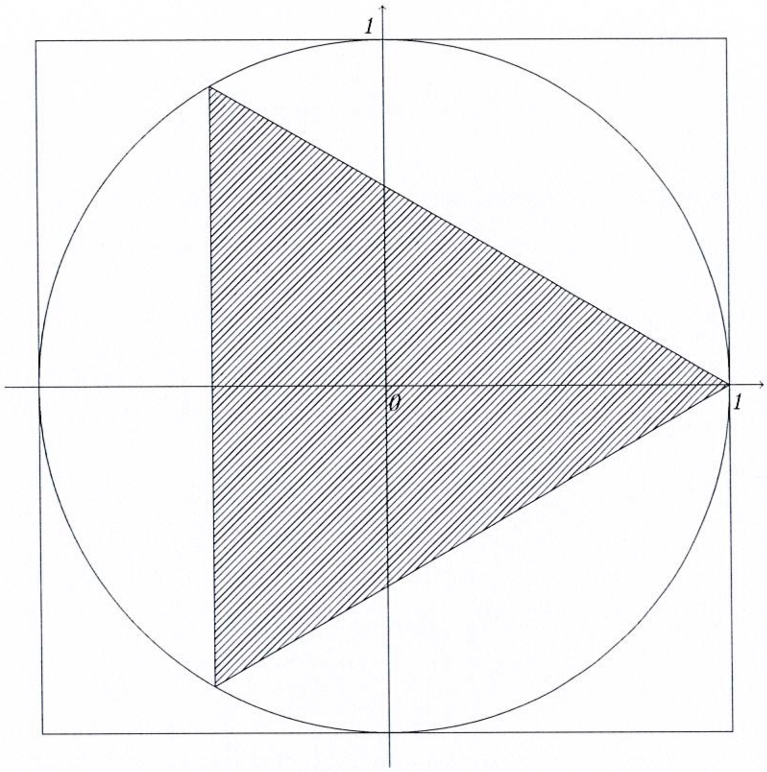
$$T_1 = \frac{3}{n} \sum_{i=1}^{n} \mathbf{1}_{\{U_i^2 \leq 1/3\}}.$$

1. *Quelle est la loi de la variable aléatoire $\mathbf{1}_{\{U_i^2 \leq 1/3\}}$?*
2. *Calculer l'erreur quadratique moyenne de l'estimateur T_1.*
3. *Que cherche à illustrer le code R suivant ?*

```
NB=10000
N=1000
C=numeric(NB)
for(i in (1:NB))
{
A=runif(N)
B=3*mean(A^2<1/3)
C[i]=(B-sqrt(3))/sqrt((sqrt(3)*(3-sqrt(3))/N))
}
plot(density(C))
curve(dnorm(x),add=TRUE,col="red")
```

Deuxième méthode d'estimation de $\sqrt{3}$

On considère la figure suivante illustrant un triangle équilatéral inscrit dans un cercle centré en l'origine et de rayon et 1 et un carré exinscrit au cercle.



1. *Déterminer l'aire du triangle équilatéral hachuré.*

Soit X (représentant l'axe des abscisses) et Y (représentant l'axe des ordonnées) deux variables aléatoires indépendantes de loi uniforme sur l'intervalle $[-1,1]$. On considère la variable aléatoire W définie ci-dessous :

$$W = \mathbf{1}_{\{(X,Y) \text{ appartient au triangle hachuré}\}}.$$

On admet que la variable aléatoire W suit une loi $\mathcal{B}(3\sqrt{3}/16)$ (la probabilité du succès est le rapport entre l'aire du triangle équilatéral et l'aire du carré). On considère $W_1, \dots, W_n$ une suite de variables aléatoires indépendantes et de même loi $\mathcal{B}(3\sqrt{3}/16)$. Il est pertinent d'estimer $\sqrt{3}$ via l'estimateur suivant :

$$T_2 = \frac{16}{3n} \sum_{i=1}^{n} W_i.$$

2. *Le code R suivant génère une observation de la variable W*

```
A=runif(1,-1,1)
B=runif(1,-1,1)
C=1*(A>-0.5)*1*(9*B^2<3*(A-1)^2)
```

   En vous aidant du code, donner une formulation explicite de la variable

$$W = \mathbf{1}_{\{(X,Y) \text{ appartient au triangle hachuré}\}}.$$

3. *Calculer l'erreur quadratique moyenne de l'estimateur T_2.*
4. *Quel estimateur T_1 ou T_2 est le plus performant pour estimer $\sqrt{3}$? Justifier votre réponse.*