Université de Bourgogne Année 2020-2021
L3 de mathématiques Sciences, Technologies, Santé - UE Statistique

---

Mardi 18 mai 13h30-16h30

Exercice 1 : (4 pts)

Afin de mieux gérer les demandes de crédits de clients, un directeur d'agence bancaire réalise une étude relative à la durée de traitement des dossiers, supposée suivre une loi normale. Les données sont résumées ci-dessous :

Durée de traitement (min)	[0 − 10[	[10 − 20[	[20 − 30[	[30 − 40[	[40 − 50[	[50 − 60]
Nombre de dossiers	3	5	8	6	2	1

1. Calculer la moyenne et l'écart-type des durées de traitements de cet échantillon. *(on prendra les milieux des classes (5, 15, 25, 35, 45, 55) comme durée de traitement d'un dossier)*
2. Donner la variable statistique et la loi permettant de construire un intervalle de confiance à 90% pour la moyenne des durées de traitement des dossiers. **Construire de façon détaillée** cet intervalle. Puis le calculer.
3. Donner la variable statistique et la loi permettant de construire un intervalle de confiance à 90% pour l'écart-type des durées de traitement des dossiers. **Construire de façon détaillée** cet intervalle. Puis le calculer.

Exercice 2 : (3 pts)

Une machine remplit des sachets en visant un poids fixé. On pense qu'en moyenne, on a 25% de sachets en sous-poids. On souhaiterait savoir si on ne sous-estime pas le nombre de sachets en sous-poids. On choisit au hasard un **petit** échantillon de $n = 18$ sachets et on trouve 6 sachets en sous-poids.

1. Faire un test statistique au niveau 5% pour conclure en rédigeant soigneusement chaque étape, modélisation incluse.
2. Calculer la p-value. Donner l'interprétation en terme de probabilité.
3. Calculer la puissance de ce test pour 40% de sachets en sous-poids. Donner l'interprétation en terme de probabilité.

Exercice 3 : (3 pts)

Des relevés effectués pendant de nombreuses années ont permis d'établir que la hauteur des pluies dans la rivière de la Beauce en millimètres par an suit une loi normale $\mathcal{N}(600, 100^2)$. Des entrepreneurs, surnommés "faiseurs de pluie", prétendaient pouvoir augmenter la hauteur moyenne de pluie, ceci par insémination des nuages au moyen d'iodure d'argent. Leur procédé fut mis à l'essai entre 1951 et 1959 et on releva les hauteurs de pluies suivantes :

Année	1951	1952	1953	1954	1955	1956	1957	1958	1959
mm	510	614	780	512	501	534	603	788	650

1. Faire un test au niveau 0.05 pour dire si les faiseurs de pluie ont raison en rédigeant soigneusement chaque étape, modélisation incluse. *Attention ici la variance est connue, donc on utilisera la loi normale et pas la loi de Student.*
2. Calculer la p-value de ce test.
3. Calculer la puissance de ce test.

Exercice 4 : (3 pts)

Dans une entreprise le nombre moyen de jours d'absence pendant un trimestre a été de 2.05 pour un échantillon de 12 employés, avec une variance de 3.05. Dans une autre entreprise le nombre de jours d'absence pendant le même trimestre a été de 3.01 pour un échantillon de 23 employés, avec une variance de 2.14.

En faisant l'hypothèse que dans chacune des entreprises le nombre moyen de jours d'absence suit une loi normale, faire un test statistique au niveau 0.05 (en rédigeant soigneusement chaque étape, modélisation incluse) pour dire s'il y a une différence d'absentéisme entre ces deux entreprises. *On prendra garde à ce que ce test soit bien valide.*

Exercice 5 : (9 pts) *Comparaison de méthodes stochastiques d'approximation de π*

Les parties 1 et 2 sont indépendantes.

On notera
— $\|.\|$ la norme euclidienne dans $\mathbb{R}^2$ ou $\mathbb{R}^3$
— $\mathbf{1}_A$ la variable aléatoire indicatrice de l'événement A ($\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbf{1}_A(\omega) = 0$ sinon)

Partie 1 Tirage uniforme de points dans le plan et l'espace

Soient les variables aléatoires suivantes permettant de modéliser le tirage uniforme de points dans $[-1,1]^2 \subset \mathbb{R}^2$ et $[-1,1]^3 \subset \mathbb{R}^3$:

— $X_1, \ldots, X_n$ (pour les abscisses) des variables aléatoires iid de loi uniforme sur $[-1,1]$
— $Y_1, \ldots, Y_n$ (pour les ordonnées) des variables aléatoires iid de loi uniforme sur $[-1,1]$.
— $Z_1, \ldots, Z_n$ (pour les cotes) des variables aléatoires iid de loi uniforme sur $[-1,1]$.

Pour $i = 1, \ldots, n$, l'événement $\|(X_i, Y_i)\| \leq 1$ signifie que le ième point tiré dans $[-1,1]^2$ est dans le disque de rayon 1 et l'événement $\|(X_i, Y_i, Z_i)\| \leq 1$ signifie que le ième point tiré dans $[-1,1]^3$ est dans la boule de rayon 1.

Comme $\mathbb{P}(\|(X_i, Y_i)\| \leq 1) = \frac{\pi}{4}$ (aire du disque de rayon 1 divisée par l'aire de $[-1,1]^2$) et $\mathbb{P}(\|(X_i, Y_i, Z_i)\| \leq 1) = \frac{\pi}{6}$ (volume de la boule de rayon 1 divisé par le volume de $[-1,1]^3$), il est naturel d'approximer π via les statistiques suivantes :

$$T_1 = \frac{4}{n}\sum_{i=1}^{n} \mathbf{1}_{\|(X_i,Y_i)\|\le 1} \text{ et } T_2 = \frac{6}{n}\sum_{i=1}^{n} \mathbf{1}_{\|(X_i,Y_i,Z_i)\|\le 1}$$

1. Donner les lois des variables aléatoires $\mathbf{1}_{\|(X_i,Y_i)\|\le 1}$ et $\mathbf{1}_{\|(X_i,Y_i,Z_i)\|\le 1}$. Est-ce que T_1 et T_2 sont des estimateurs biaisés pour π ?
2. Calculer l'erreur quadratique moyenne $EQM(T_1)$ et $EQM(T_2)$ des estimateurs T_1 et T_2. Quel estimateur est le plus performant au sens de l'erreur quadratique moyenne ?
3. On considère le code R suivant :

```
n=1000;N=1000
T1=numeric(N)
for (i in seq(1,N))
{
  u1=runif(n);u2=runif(n)
  T1[i]=4*mean(u1^2+u2^2<=1)
}
R=mean((T1-pi)^2)
```

Que représentent les lettres n, N et R ? Que peut-on dire de R lorsque n est très grand ? Que peut-on dire de R lorsque N est très grand ? Justifiez brièvement vos réponses.

Partie 2 L'aiguille de Buffon

L'aiguille de Buffon est une expérience probabiliste qui permet également d'approximer π.

1. On considère X une variable aléatoire de loi uniforme sur $[0, 1/2]$ et θ une variable aléatoire de loi uniforme sur $[0, \pi/2]$ [1]. Montrer que

$$\mathbb{P}(2X \le \sin(\theta)) = \frac{2}{\pi}.$$

2. Soient $X_1, \ldots, X_n$ des variables aléatoires iid de loi uniforme sur $[0, 1/2]$ et $\theta_1, \ldots, \theta_n$ des variables aléatoires iid de loi uniforme sur $[0, \pi/2]$. D'après la question précédente, il est naturel d'estimer $2/\pi$ via la statistique suivante :

$$\overline{W} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{2X_i \le \sin(\theta_i)}$$

1. Le jet aléatoire d'une aiguille de longueur 1 sur un parquet dont les lattes mesurent 1 permet de générer une observation de la statistique $\mathbf{1}_{2X\le\sin(\theta)}$. En effet, en notant X la distance aléatoire entre le centre de l'aiguille et la plus proche rainure et en posant θ l'angle géométrique aléatoire entre l'aiguille et la rainure alors la probabilité que l'aiguille coupe une rainure est $\mathbb{P}(2X \le \sin(\theta))$.

On pose également $T_3 = 2/\overline{W}$. L'erreur quadratique moyenne de T_3 pour π est difficile à calculer explicitement. À la place, on va faire une étude asymptotique de l'estimateur T_3 qui sera plus facile à effectuer.

Donner la loi asymptotique de $\sqrt{n}(\overline{W} - 2/\pi)$.

3. En utilisant la "delta-méthode", donner la loi asymptotique de la statistique $\sqrt{n}(T_3 - \pi)$.

Partie 3 Comparaison des trois méthodes

1. En comparant les lois asymptotiques des statistiques : $\sqrt{n}(T_1 - \pi)$, $\sqrt{n}(T_2 - \pi)$ et $\sqrt{n}(T_3 - \pi)$, quel est d'après vous l'estimateur le plus performant pour approximer π lorsque n est grand ?
2. On considère le code R suivant :

```
1) curve(dnorm(x,0,sqrt(2.696)),col='black',lwd=2,xlim=c(-12,12),ylim=c(0,0.25))
2) curve(dnorm(x,0,sqrt(5.633)),col='red',add=TRUE,lwd=2)
3) curve(dnorm(x,0,sqrt(8.979)),col='blue',add=TRUE,lwd=2)
```

Que représentent les courbes 1), 2) et 3) ?