# Devoir

## Exercice 1

*Remarque préliminaire : soient $X$ et $Y$ deux variables aléatoires. La covariance permet de mesurer la dépendance entre ces deux variables. Elle est définie par*

$$\mathrm{Cov}(X,Y) = \mathbf{E}((X-\mathbf{E}(X))(Y-\mathbf{E}(Y))).$$

*Si $X$ et $Y$ sont indépendantes, leur covariance est nulle.*
*Si $X-\mathbf{E}(X)$ et $Y-\mathbf{E}(Y)$ ont (en moyenne) tendance à être de même signe, leur covariance est positive. Sinon elle est négative.*
*Enfin, la covariance apparaît dans le calcul de variances :*

$$\sigma^2(X-Y) = \sigma^2(X) + \sigma^2(Y) - 2\mathrm{Cov}(X,Y).$$

*On rappelle par ailleurs $\sigma^2(\lambda X) = \lambda^2\sigma^2(X)$.*

On souhaite étudier l'érosion d'une certaine pièce mécanique au cours du temps. On pèse pour cela des pièces neuves et des pièces ayant servi un mois.

Notons $X$ la variable aléatoire modélisant le poids d'une pièce neuve et $Y$ celle modélisant le poids d'une pièce usagée. On suppose que $X$ suit la loi normale $\mathcal{N}(m_1, \sigma_1^2)$ et $Y$ la loi normale $\mathcal{N}(m_2, \sigma_2^2)$. Notre objectif est d'obtenir une estimation de $d = m_1 - m_2$.

1. On choisit aléatoirement et de façon indépendantes $n$ pièces neuves et $n$ pièces usagées. On note $X_1, \dots, X_n$ les poids des premières et $Y_1, \dots, Y_n$ les poids des secondes. On propose d'estimer $d = m_1 - m_2$ à l'aide de l'estimateur

$$\bar{d} = \frac{1}{n}(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i) = \bar{X} - \bar{Y} = \overline{X-Y}.$$

   (a) Montrer que $\bar{d}$ est un estimateur sans biais.

   (b) Donner la variance de $\bar{d}$ et en déduire que l'estimateur est convergent.

   (c) Le tableau ci-dessous donne les poids (en g) de 10 pièces neuves et de 10 pièces usagées.

| $X_i$ | 1001 | 997 | 1000 | 1004 | 999 | 1000 | 998 | 1002 | 997 | 1002 |
|---|---|---|---|---|---|---|---|---|---|---|
| $Y_i$ | 990 | 995 | 990 | 989 | 988 | 993 | 989 | 988 | 985 | 993 |

   Déterminer un intervalle de confiance pour la valeur de $d$ au risque de 5%.

2. On considère maintenant une seconde méthode. On choisit aléatoirement et de façon indépendante $n$ pièces neuves que l'on pèse. Puis on pèse de nouveau ces mêmes pièces un mois plus tard (c'est le principe du panel). On note $X_1, \ldots, X_n$ leurs poids initiaux et $Y_1, \ldots, Y_n$ leurs poids finaux. On considère encore le même estimateur de $d$ : $\bar{d}' = \bar{X} - \bar{Y}$. Malgré le fait que les tirages ne sont plus indépendants, on fait l'hypothèse que les variables $X_i - Y_i$ suivent une loi normale.
   (a) Donner les paramètres de la loi des $X_i - Y_i$ (espérance et variance).
   (b) En déduire l'espérance et la variance de $\bar{d}'$.
   (c) L'estimateur $\bar{d}'$ est-il meilleur que $\bar{d}$ ?
   (d) On reprend les données collectées à la question 1-c mais on suppose cette fois qu'elles proviennent d'un panel. Et on suppose également que la covariance des variables $X_i$ et $Y_i$ a été estimée : $\mathrm{Cov}(X_i, Y_i) \approx 0,7\sigma_1\sigma_2$.
   Déterminer un intervalle de confiance pour la valeur de $d$ au risque de 5%.
3. On souhaite estimer la taille moyenne de la population. on propose deux méthodes :
   - On choisit aléatoirement et de façon indépendantes 10 hommes et 10 femmes puis on calcule la moyenne de leurs tailles.
   - On choisit aléatoirement et de façon indépendantes 20 personnes puis on calcule la moyenne de leurs tailles.

   (La première méthode est analogue à ce qu'on a fait dans la première partie mais pour une estimation de $\frac{m_1+m_2}{2}$.)

   Laquelle de ces méthodes permet d'obtenir l'estimation la plus précise ? On pourra répondre sans effectuer de calcul.

## Exercice 2 : bruit blanc

On émet un signal via un canal. Au cours de la transmission, le signal est perturbé par un bruit. L'objectif de cet exercice est de proposer une manière de débruiter le signal.

1. **Analyse du bruit**
   Le bruit blanc est modélisé par une suite de variables $X_k$ indépendantes de loi normale $\mathcal{N}(0, \sigma^2)$. Afin d'estimer l'ampleur du bruit (modélisée par $\sigma$), on commence par émettre un signal constant égal à $m$. On note $Y_k = m + X_k$ le message reçu à l'instant $k$. On a capté 100 valeurs et on a obtenu le résultat suivant :
   $$\sum_{k=1}^{100} (Y_k - m)^2 = 400.$$
   En déduire une majoration de $\sigma^2$ au risque de 1% (c'est-à-dire un intervalle de confiance de la forme $[0, M]$).

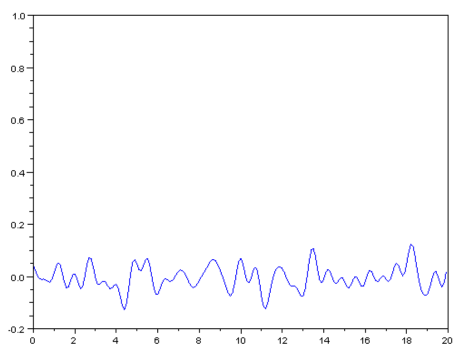
2. **Débruitage**

On émet désormais un signal $S$ représenté par une suite de valeurs successives $S_k$. Le message est bruité et ce que l'on capte est la suite de valeurs $B_k = S_k + X_k$. On souhaite estimer la valeur moyenne du signal $S$. Elle est égale à $m = \frac{1}{N}\sum_{k=1}^{N} S_k$. On propose de l'estimer par $\bar{m} = \frac{1}{N}\sum_{k=1}^{N} B_k$.

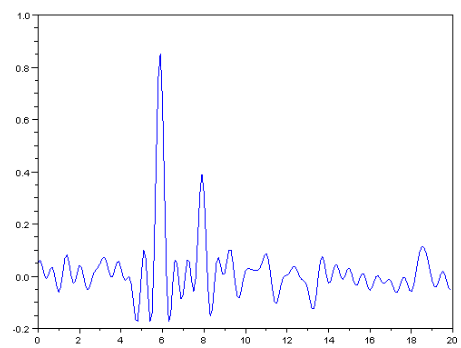
(a) Montrer avec le théorème central limite qu'au risque de 1%, $m$ est dans l'intervalle $[\bar{m} - \frac{\alpha}{\sqrt{N}}, \bar{m} + \frac{\alpha}{\sqrt{N}}]$ où $\alpha$ est une constante à déterminer.

(b) Pour quelle valeur de $N$ obtient-on une estimation de $m$ avec une précision de $0,2$ ?

Plus généralement, on peut pour cette valeur de $N$ mesurer avec la même précision la présence de chaque fréquence dans le signal $S$ et ainsi recomposer le signal $S$ entièrement. Le signal $S$ peut se décomposer sous la forme d'une série de Fourier : $S(t) = m + \sum_{\lambda \in F} a_\lambda \cos(\lambda t) + b_\lambda \sin(\lambda t)$, où $F$ est l'ensemble des fréquences présentes dans le signal. De manière analogue à ce qu'on a fait pour estimer $m$, on peut estimer les coefficients $a_\lambda$ et $b_\lambda$ avec la même précision. On présente ci-dessous les graphes des coefficients $a_\lambda$ et $b_\lambda$ que l'on a estimés.



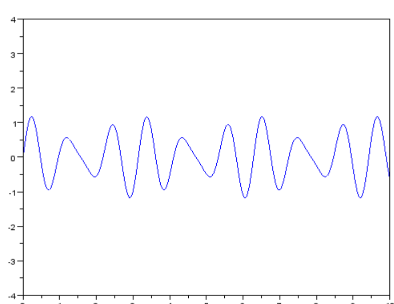
Valeurs de $a_\lambda$

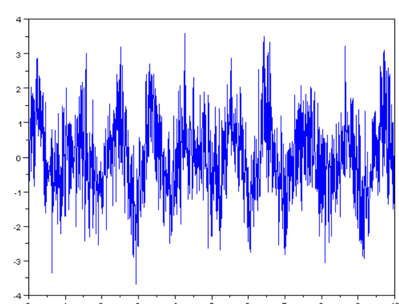


Valeurs de $b_\lambda$

(c) Compte tenu de la précision avec laquelle on a estimé ces coefficients, proposer une façon de débruiter le signal et donner l'expression du signal émis correspondant.

On donne ci-dessous les graphes du signal émis et du signal bruité.





(d) Expliquer pourquoi un traitement direct du signal bruité n'aurait pas permis de débruiter aussi précisément.