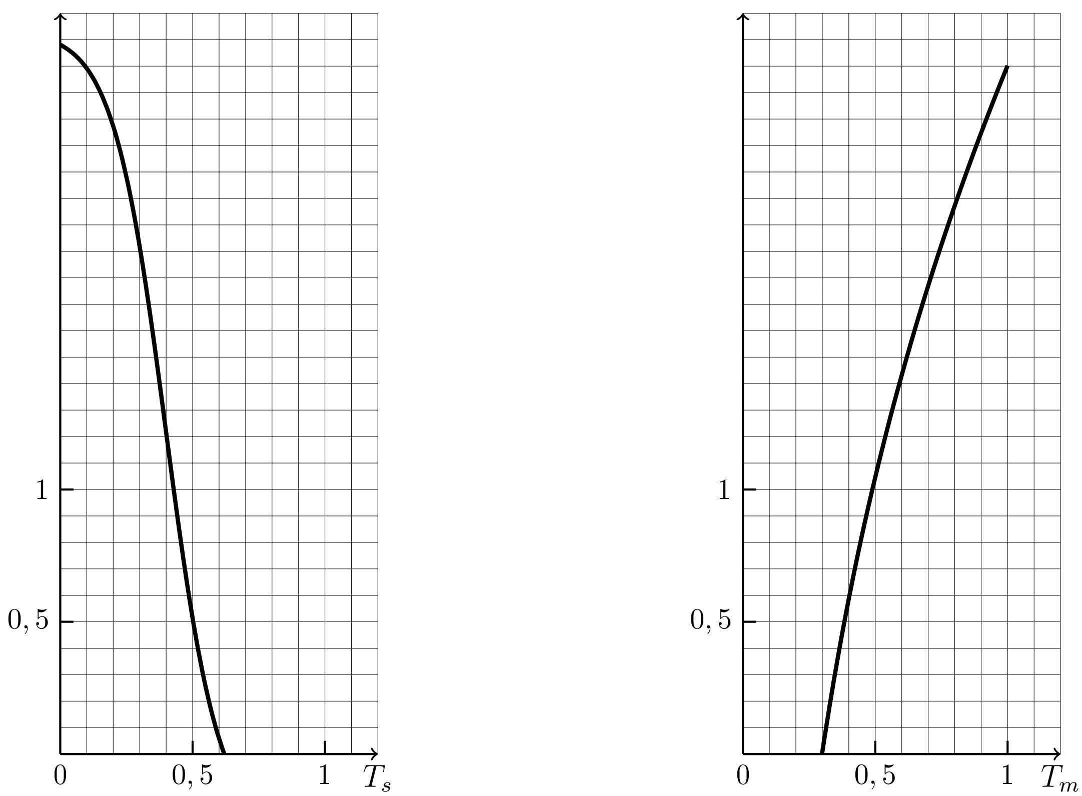
# Devoir

*Documents et calculatrice sont autorisés.*
*Le barème est donné à titre indicatif.*

## Exercice 1 : test médical (5 pts)

Afin de détecter une certaine maladie, on utilise un test médical dont le résultat est une valeur comprise entre 0 et 1.

Des expériences ont permis d'établir la répartition statistique des résultats du test chez les individus sains et chez les individus malades. On note $T_s$ le résultat du test chez un individu sain et $T_m$ celui d'un individu malade. Les graphes ci-dessous représentent les densités de $T_s$ et $T_m$ (l'aire d'un carreau est $0,01$).




1. Estimer grossièrement les espérances $\mathbf{E}(T_s)$ et $\mathbf{E}(T_m)$.
2. Proposer une valeur $c$ telle que $\mathbf{P}(T_m > c) \approx 0,97$.
3. Estimer alors $\mathbf{P}(T_s > c)$.

   On utilise $c$ comme valeur seuil : lorsque le test d'un individu fournit une valeur supérieure à $c$, il est déclaré positif, sinon il est déclaré négatif.

4. Sachant que cette maladie touche un individu sur cent, estimer les deux risques d'erreur : la probabilité d'être malade lorsque le test est négatif et la probabilité d'être sain lorsque le test est positif.

## Exercice 2 : simulation de la loi normale

On souhaite simuler une variable aléatoire suivant la loi $\mathcal{N}(0,1)$. Nous proposons trois méthodes reposant toutes sur des simulations de la loi uniforme $\mathcal{U}([0,1])$. On note $f_{\mathcal{N}}$ et $F_{\mathcal{N}}$ les fonctions de densité et de répartition de la loi $\mathcal{N}(0,1)$ et $\mathbf{P}_{\mathcal{N}}$ la mesure de probabilité associée.

### 1. Première méthode (2 pts)

Notons $F_{\mathcal{N}}^{-1}$ la bijection réciproque de $F_{\mathcal{N}}$ définie de $]0,1[$ vers $\mathbf{R}$. On considère une variable $X$ de loi $\mathcal{U}([0,1])$ et on pose $N_1 = F_{\mathcal{N}}^{-1}(X)$.

a. Montrer que $N_1$ suit la loi normale $\mathcal{N}(0,1)$.

b. Pourquoi est-il en pratique difficile de simuler la loi normale de cette façon là ?

### 2. Deuxième méthode (5 pts)

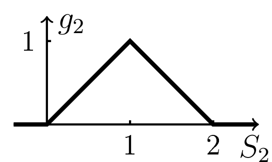
Soit $n \geqslant 2$. On note $X_1, \ldots, X_n$ des variables indépendantes de même loi $\mathcal{U}([0,1])$. On pose

$$S_n = \sum_{i=1}^{n} X_i \quad \text{et} \quad N_n = \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}.$$

a. En quoi la variable $N_n$ permet-elle de simuler la loi $\mathcal{N}(0,1)$ ?

On aimerait préciser la valeur de $n$ à partir de laquelle la variable $N_n$ permet de calculer des probabilités pour la loi $\mathcal{N}(0,1)$ avec une précision de $10^{-2}$.

La loi de la variable $S_n$ est connue : il s'agit de la loi d'Irwin-Hall. Sa densité $g_n$ permet de déterminer explicitement la densité $f_n$ de $N_n$. Pour $n = 2$, la densité $g_2$ de $S_2$, calculée dans le cours, est donnée ci-contre.




b. Calculer à l'aide de cette densité la probabilité $\mathbf{P}(-\sqrt{3/2} \leqslant N_2 \leqslant \sqrt{3/2})$.
La comparer avec la probabilité $\mathbf{P}_{\mathcal{N}}([-\sqrt{3/2}, \sqrt{3/2}])$.

c. On note $\|f\|_1 = \int_{-\infty}^{+\infty} |f(x)| \mathrm{d}x$. *On reconnaîtra une norme bien connue de $L^1(\mathbf{R})$.*
Justifier que si $\|f_n - f_{\mathcal{N}}\|_1 \leqslant 0{,}01$ , on peut en déduire que pour toute partie $A$ de $\mathbf{R}$,

$$|\mathbf{P}(N_n \in A) - \mathbf{P}_{\mathcal{N}}(A)| \leqslant 0{,}01.$$

*On admet que cette condition est satisfaite à partir de $n = 15$.*

### 3. Troisième méthode (dite de Box-Muller) (8 pts)

a. Soient $X$ et $Y$ des variables indépendantes de loi normale $\mathcal{N}(0,1)$. Soit $B$ une partie du plan. Exprimer sous forme intégrale la probabilité $\mathbf{P}((X,Y) \in B)$.

b. Effectuer pour cette intégrale un changement de variables en coordonnées polaires :

$$x = r\cos(\theta), \quad y = r\sin(\theta) \quad \text{et} \quad \mathrm{d}x\mathrm{d}y = r\mathrm{d}r\mathrm{d}\theta.$$

c. Un exemple : soit $B$ le disque de centre $(0,0)$ et de rayon 1. Que vaut $\mathbf{P}((X,Y) \in B)$ ?

Les variables $r$ et $\theta$ ci-dessus vont nous permettre de simuler des variables gaussiennes. Considérons maintenant deux variables $U_1$ et $U_2$ indépendantes de même loi uniforme $\mathcal{U}([0,1])$.

d. Posons $R = \sqrt{-2\ln(U_1)}$. Déterminer la fonction de répartition de $R$ puis sa densité.

e. Posons $\Theta = 2\pi U_2$. Quelle est la loi de $\Theta$ ?

f. En déduire que $X = R\cos(\Theta)$ et $Y = R\sin(\Theta)$ suivent la loi $\mathcal{N}(0,1)$.

g. En quoi cette méthode est-elle meilleure que la deuxième méthode ?