Exercice 2, page de droite

On en a désormais l'habitude : modéliser correctement l'énoncé à l'aide de variables aléatoires aidera à répondre à l'ensemble des questions.

On s'intéresse ici à la loi qui donne la fréquence de 7 dans une population de 121 étudiants dont chacun a tiré un nombre au hasard dans une table de nombres.

On numérote donc les étudiants de i=1 à i=121.

Une épreuve est le tirage d'un chiffre entre 0 et 9.
L'univers Ω est l'ensemble des étudiants : n = 121.

Mais on s'intéresse ici en particulier à l'épreuve qui consiste à tirer un 7.
C'est pourquoi, pour un étudiant i, on note **X_i la variable aléatoire qui vaut 1 s'il tire un 7 et qui vaut 0 dans les autres cas**.

Toutes les questions de l'énoncé portent sur la fréquence de 7.

Or cette variable vaut $\dfrac{X_1 + \ldots + X_{121}}{121} = \dfrac{S_{121}}{121}$

Appelons-la $\mathbf{F_{121}}$:

Soit $\mathbf{F_{121} = S_{121} / 121}$

Pour résoudre l'exercice, il **reste à connaître la loi de probabilité de F_{121}**.

Les 121 variables aléatoires **X_i sont indépendantes** (le chiffre tiré au hasard par un étudiant n'a aucune influence sur le chiffre tiré au hasard par un autre étudiant).
Dans une table de nombres, les chiffres de 0 à 9 sont équidistribués : chacun des dix chiffres a autant de chances d'être tiré.
La probabilité de tirer un 7 vaut donc 1/10.

Les variables **X_i** suivent donc toutes une **loi de Bernouilli de paramètre 1/10**.

Le cours nous dit que **X** = S_{121} / 121 suit une **loi pseudo-binomiale**,
de **paramètres 121 et 1/10**.

Note : la loi pseudo-binomiale est aussi appelée loi des fréquences. D'où la lettre F.

On sait calculer toutes les probabilités qu'on veut sur la loi pseudo-binomiale.
Mais pour n = 121, les calculs sont extrêmement longs.

Mais une **propriété du cours** nous dit que si n est assez grand, on peut approcher cette loi par une loi normale (c'est une conséquence du théorème central limite).

Plus précisément, **si np ≥ 10 et nq ≥ 10**,
alors **Sn / n** suit une **loi normale de paramètres p et $\sqrt{pq / n}$**

On le voit **dans le formulaire**, rubrique échantillons, troisième propriété (avec une petite erreur : lire « loi de Bernouilli de paramètre p », et pas « loi binomiale de paramètres n et p »). Cela apparaît aussi en bas du tableau de la page 3 :

> Si les variables aléatoires X_i (associées aux individus de l'échantillon) sont indépendantes et suivent une même loi de Bernouilli de paramètre p,
> alors la variable aléatoire F_n qui donne leur fréquence (qui est aussi leur moyenne $\overline{X}$) suit une loi pseudo-binomiale de paramètres n et p.
>
> Et si np >10 et nq >10,
> on peut approcher cette loi pseudo-binomiale par une loi normale de paramètres p et $\sqrt{pq / n}$

Or, dans le cas qui nous intéresse, on a np = 121/10 = 12,1 >10 et nq = 121 × 9 / 10 > 10

Donc F_{121}, la fréquence de 7, suit une loi normale de paramètres :

$\mu = 1/10$

$\sigma = \sqrt{(1/10) \times (9/10)/121} = \sqrt{(9/100)/121} = \sqrt{9} / \sqrt{100} / \sqrt{121} = 3 / 10 / 11 = 3/110$

On utilise la formule de standardisation pour transformer F_{121} en une variable centrée réduite :

Soit $F_{121}^* = \dfrac{F_{121} - 1/10}{3/110} = 110/3 \times (F_{121} - 0{,}1)$ la variable centrée réduite de F_{121}.

F_{121}^* suit une loi normale de moyenne 0 et d'écart-type 1.

On ramènera tous les calculs sur F_{121} à des calculs sur F_{121}^*.

1) $4{,}65\ \% < F_{121} < 15{,}35\ \%$

$\Leftrightarrow 0{,}0465 < F_{121} < 0{,}1535$

$\Leftrightarrow 110/3 \times (0{,}0465 - 0{,}1) < F_{121}^* < 110/3 \times (0{,}1535 - 0{,}1)$

$\Leftrightarrow -1{,}96 < F_{121}^* < 1{,}96$

> Remarque
> Il est indifférent ici de mettre les signes < ou ≤.
> Puisque la probabilité que la variable soit exactement égale à 4,65 % ou à 15,35 % est nulle.

Donc $P(4{,}65\ \% < F_{121} < 15{,}35\ \%) = P(-1{,}96 < F_{121}^* < 1{,}96) \approx 95\ \%$

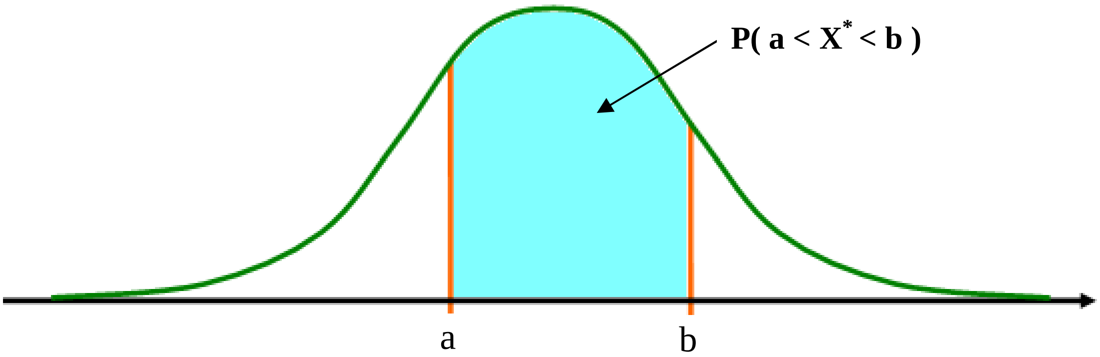
Car 1,96 est une valeur typique pour la loi normale centrée réduite.

Remarque : de même, 1,65 est une valeur typique : $P(-1{,}65 < F_{121}^* < 1{,}65) \approx 90\ \%$

Parenthèse technique

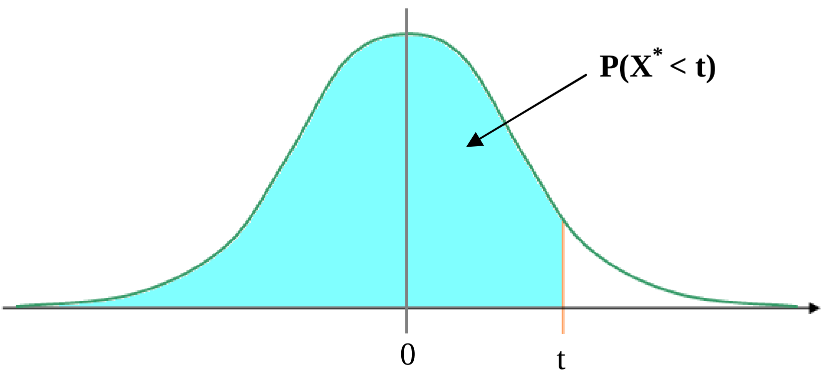
Comment déterminer une probabilité du type $P(a < X^* < b)$,
quand X^* suit une loi normale centrée réduite ?

Cette probabilité est l'aire comprise entre les droites d'abscisses a et b et sous la courbe de la fonction de densité de X^* :



Les calculatrices, en général, ne font pas ces calculs. On utilise donc une table de valeurs de la loi normale centrée réduite : la **table 1**, à la **page 5 du formulaire**.

Mais cette table ne donne que des probabilités du type $P(X^* < t)$, avec $t > 0$.



On va apprendre par étapes à calculer $P(a < X^* < b)$:

a) Calcul de $P(X^* < t)$ quand $t > 0$
b) Calcul de $P(X^* > t)$
c) Calcul de $P(X^* < t)$ et de $P(X^* > t)$ quand $t < 0$
d) Calcul de $P(a < X^* < b)$

a) Calcul de $P(X^* < t)$ quand $t > 0$

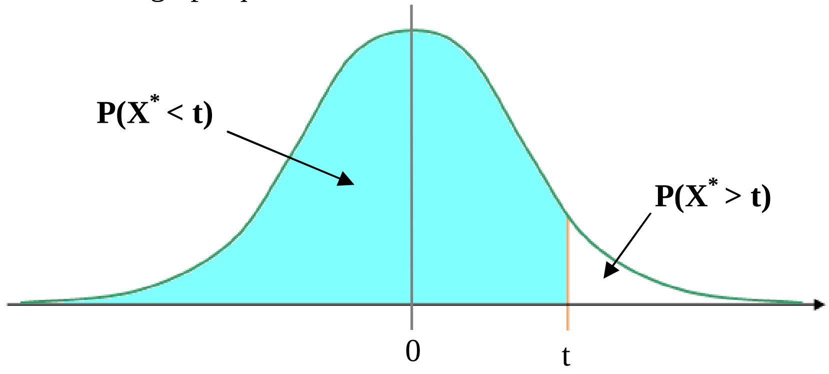
Pour calculer $P(X^* < t)$ avec la table, on cherche t (deux décimales) dans la table.
La valeur de la probabilité recherchée est sur la case correspondant à la première décimale (en ligne) et à la deuxième décimale (en colonne).

Par exemple, $P(X^* < 1,96) = 0,9750 = 97,50$ %

b) Calcul de $P(X^* > t)$

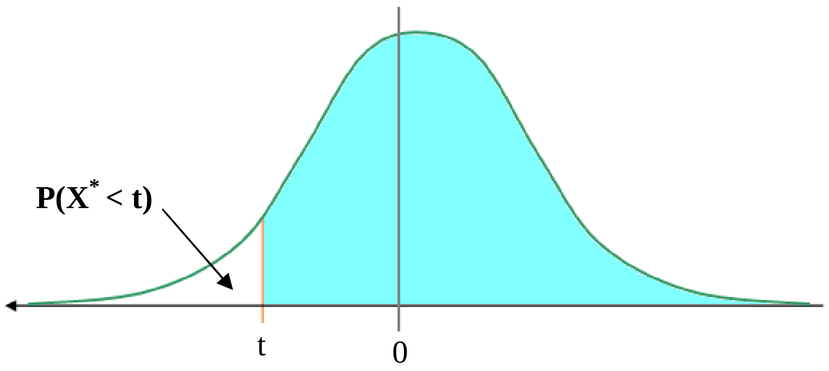
Le calcul est simple : $P(X^* > t) = 1 - P(X^* < t)$

On le voit sur le graphique suivant :



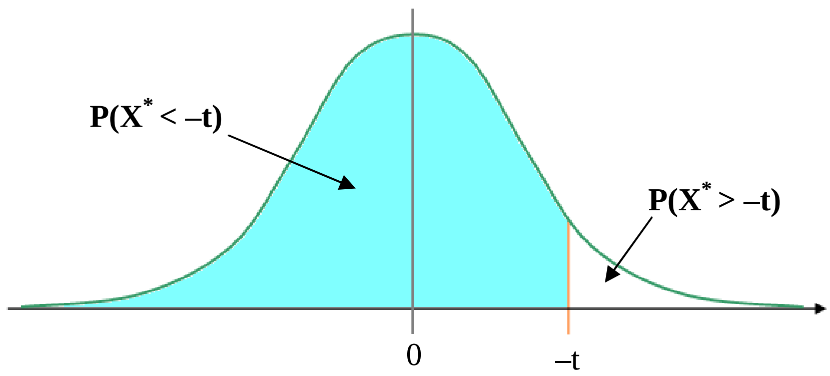
c) Calcul de $P(X^* < t)$ et de $P(X^* > t)$ quand $t < 0$

Dans les cas où $t < 0$, la table ne nous donne pas la valeur de $P(X^* < t)$.



Mais on peut calculer cette valeur par un argument de symétrie.
En effet, le nombre $-t$ est positif, et comme la courbe est symétrique par rapport à l'axe des ordonnées, alors : $\mathbf{P(X^* < t) = P(X^* > -t)}$

On le voit en comparant les deux portions blanches sous la courbe, dans le graphique ci-dessus et dans le graphique ci-dessous.



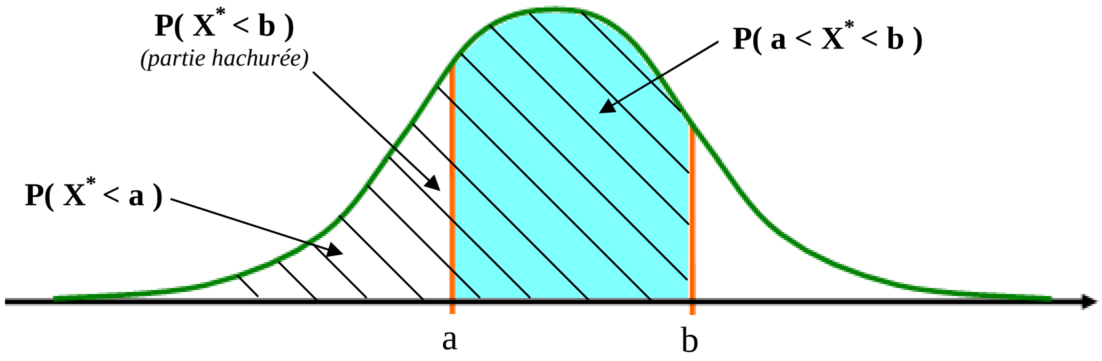
Et on sait calculer $P(X^* > -t)$: c'est égal à $1 - P(X^* < -t)$

Donc finalement, si $t < 0$, $P(X^* < t) = 1 - P(X^* < -t)$,
et la table permet de calculer $P(X^* < -t)$ car $-t > 0$.
Inutile de retenir cette formule : il faut savoir faire le raisonnement en traçant des graphiques au brouillon.

d) Calcul de $P(a < X^* < b)$

Un raisonnement sur les aires nous montre que :

$P(a < X^* < b) = P(X^* < b) - P(X^* < a)$



En effet, on obtient la portion en couleur en enlevant la portion blanche hachurée à l'ensemble de la portion hachurée.

Il ne reste plus qu'à calculer $P(X^* < b)$ et $P(X^* < a)$ avec les techniques expliquées ci-dessus.

Après cette parenthèse technique, retour à l'exercice.

Pour la question 1), on calcule donc :

$P(X^* < 1{,}96) \approx 0{,}9750$ d'après la table 1.

$P(X^* < -1{,}96) = P(X^* > 1{,}96)$ par symétrie
$= 1 - P(X^* < 1{,}96)$ par passage au complémentaire
$\approx 1 - 0{,}9750$ avec la table
$\approx 0{,}0250$

Donc $P(-1{,}96 < X^* < 1{,}96) \approx 0{,}9750 - 0{,}0250 \approx 0{,}095 \approx 95$ %

Il y a donc 95 % de chances pour qu'on obtienne entre 4,65 % et 15,35 % de 7.

2) $X > 14\%$ ⇔ $X^* > 110/3 \times (0{,}14 - 0{,}1)$

⇔ $X^* > 1{,}47$

Donc $P(X > 14\%) = P(X^* > 1{,}47)$

Or, $P(X^* > 1{,}47) = 1 - P(X^* < 1{,}47)$

Et la table 1 nous donne : $P(X^* < 1{,}47) \approx 0{,}9292$

Donc $P(X > 14\%) \approx 1 - 0{,}9292 \approx 0{,}0708$

Il y a donc 7,1 % de chances pour qu'on obtienne plus de 14 % de 7.

3) $X < 8\%$ ⇔ $X^* < 110/3 \times (0{,}08 - 0{,}1)$

⇔ $X^* < -0{,}73$

Or, $P(X^* < -0{,}73) = P(X^* > 0{,}73)$ par symétrie
$= 1 - P(X^* < 0{,}73)$ par passage au complémentaire

Et la table nous donne $P(X^* < 0{,}73) \approx 0{,}7673$

Donc $P(X < 8\%) \approx 1 - 0{,}7673 \approx 0{,}2327$

Il y a donc 23 % de chances pour qu'on obtienne moins de 8 % de 7.

Dernière question

On a demandé à 121 étudiants de donner un chiffre au hasard entre 0 et 9, sans utiliser de table de nombres.

On a obtenu 22,9 % de 7, soit plus de 22 %.

Calculons les chances que ce soit le cas, dans l'hypothèse que les chiffres sont vraiment choisis au hasard :

$P(X > 22\%) = P(X^* > 110/3 \times (0{,}22 - 0{,}1))$
$= P(X^* > 4{,}4)$
$= 1 - P(X^* < 4{,}4) \approx 0$ d'après la table 1 (en fait, moins de 1/10 000).

En admettant que les chiffres sont tirés au hasard, il n'y a quasiment aucune chance d'obtenir plus de 22 % de 7.

Conclusion :
Le présupposé selon lequel les chiffres sont tirés au hasard n'est donc pas valable : la probabilité réelle d'obtenir un 7 est supérieure à 10 %, les étudiants privilégient le chiffre 7 par rapport à un tirage aléatoire.