# Statistiques : aide personnalisée Terminale

On retrouve dans les programmes des classes de lycée la problématique de prise de décision, citée en exemple comme thème d'aide personnalisée en Terminale.

L'objet de ces activités est d'aborder ce sujet et de revenir sur les notions d'échantillonnage à divers niveaux d'exigence selon les capacités attendues par les programmes.

**Ce que dit le programme de Seconde** :

***dans le cadre de l'échantillonnage***

- faire réfléchir les élèves à la conception et la mise en œuvre d'une simulation ;
- sensibiliser les élèves à la fluctuation d'échantillonnage, aux notions d'intervalle de fluctuation et d'intervalle de confiance et à l'utilisation qui peut en être faite.

| CONTENUS | CAPACITES ATTENDUES | COMMENTAIRES |
|---|---|---|
| **Echantillonnage**<br><br>Notion d'échantillon.<br>Intervalle de fluctuation d'une fréquence au seuil de $95\%$*.<br><br>Réalisation d'une simulation. | . Concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice.<br><br>. Exploiter et faite une analyse critique d'un résultat d'échantillonnage. | Un échantillon de taille $n$ est constitué des résultats de $n$ répétitions indépendantes de la même expérience.<br>A l'occasion de la mise en place d'une simulation, on peut :<br>. utiliser les fonctions logiques d'un tableur ou d'une calculatrice,<br>. mettre en place des instructions conditionnelles dans un algorithme.<br><br>L'objectif est d'amener les élèves à un questionnement lors des activités suivantes :<br>. l'estimation d'une proportion inconnue à partir d'un échantillon,<br>. la prise de décision à partir d'un échantillon. |

* L'intervalle de fluctuation au seuil de $95\%$, relatif aux échantillons de taille $n$, est l'intervalle centré autour de $p$, proportion du caractère dans la population, où se situe, avec une probabilité égale à $0{,}95$, la fréquence observée dans un échantillon de taille $n$. Cet intervalle peut être obtenu, de façon approchée, par simulation.
Le professeur peut indiquer aux élèves le résultat suivant, utilisable dans la pratique pour des échantillons de taille $n \geq 25$ et des proportions comprises entre $0{,}2$ et $0{,}8$ : si $f$ désigne la fréquence du caracère dans l'échantillon, $f$ appartient à l'intervalle $\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins $0{,}95$. Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais **elle n'est pas exigible**.

**Ce que dit le programme de Première** :

***dans le cadre de l'échantillonnage***

Dans le cas particulier d'expériences identiques et indépendantes à deux issues, on introduit la loi binomiale. En s'appuyant su cette loi, on poursuit la formation des élèves dans le domaine de l'échantillonnage.

| CONTENUS | CAPACITES ATTENDUES | COMMENTAIRES |
|---|---|---|
| Echantillonnage<br>Utilisation de la loi binomiale pour une prise de décision à partir d'une fréquence. | . Exploiter l'intervalle de fluctuation à un seuil donné, déterminé à l'aide de la loi binomiale, pour rejeter ou non une hypothèse sur une proportion. | L'objectif est d'amener les élèves à expérimenter la notion de « différence significative » par rapport à une valeur attendue et à remarquer que pour une taille de l'échantillon importante, on conforte les résultats vus en classe de seconde.<br><br>L'intervalle de fluctuation peut être déterminé à l'aide d'un tableur ou d'un algorithme.<br>Le vocabulaire des tests (test d'hypothèse, hypothèse nulle, risque de première espèce) est hors programme. |

**<u>Ce que dit le programme de Terminale</u>** :

***dans le cadre de l'échantillonnage***

Afin de traiter les champs de problèmes associés aux données continues, on introduit les lois de probabilités à densité. Le programme en propose quelques exemples et, en particulier, la loi normale qui permet notamment d'initier les élèves à la statistique inférentielle par la détermination d'u intervalle de confiance d'une proportion à un niveau de confiance de $95\%$.
Cette partie se prête particulièrement à l'étude de problèmes issus d'autres disciplines (notamment des sciences économiques et sociales pour le programme de ES et L).
Le recours aux représentations graphiques et aux simulations est indispensable.

| CONTENUS | CAPACITES ATTENDUES | COMMENTAIRES |
|---|---|---|
| **Estimation**<br>Intervalle de fluctuation. | . Démontrer que si la variable $X_n$ suit la loi $\mathcal{B}(n,p)$, alors, pour tout α dans $]0;1[$, on a, $\lim\limits_{n\to+\infty} P\left(\frac{X_n}{n}\in I_n\right)=1-\alpha$, où $I_n$ désigne l'intervalle $\left[p-t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}};\, p+t_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$. | La démonstration ci-contre donne l'expression d'un intervalle de fluctuation asymptotique (*) au seuil $1-\alpha$ de la variable aléatoire $F_n=\frac{X_n}{n}$, qui à tout échantillon de taille $n$ associe la fréquence obtenue. |
| | . Connaître l'intervalle de fluctuation asymptotique (*) au seuil de $95\%$: $\left[p-1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}};\, p+1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ où $p$ désigne la proportion dans la population. | En pratique, on fait l'approximation dès que $n\geq 30$ , $np\geq 5$ et $n(1-p)\geq 5$.<br>En majorant $1{,}96\sqrt{p(1-p)}$ , on retrouve l'intervalle de fluctuation présenté en classe de seconde.<br><br>La problématique de prise de décision, déjà rencontrée, est travaillée à nouveau. |
| Intervalle de confiance (*).<br>Niveau de confiance. | . Estimer une proportion inconnue à partir d'un échantillon.<br>. Déterminer une taille d'échantillon suffisante pour obtenir une estimation d'une proportion au niveau de confiance $0{,}95$ pour une précision attendue. | Les attendus de ce paragraphe sont modestes et sont à exploiter en lien avec les autres disciplines.<br><br>. On peut démontrer que l'intervalle $\left[F_n-\frac{1}{\sqrt{n}};\, F_n+\frac{1}{\sqrt{n}}\right]$ contient la proportion à estimer avec une probabilité au moins égale à $0{,}95$.<br><br>On peut alors énoncer que $p$ est un élément de l'intervalle $\left[f-\frac{1}{\sqrt{n}};\, f+\frac{1}{\sqrt{n}}\right]$ avec un niveau de confiance de plus de $95\%$ , où $f$ désigne la fréquence observée sur un échantillon de taille $n$.<br>La simulation de sondages sur tableur permet de sensibiliser aux fourchettes de sondage.<br><br>Il est important de noter que, dans d'autres champs, on utilise l'intervalle $\left[f-1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}};\, f+1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ qu'il n'est pas possible de justifier dans le programme.<br><br>**AP** *Prise de décision lors de la comparaison de deux proportions (par exemple lors d'un essai thérapeutique).* |

(*) Avec les notations définies précédemment :
Un intervalle de fluctuation asymptotique de la variable aléatoire $F_n$ au seuil $1-\alpha$ est un intervalle déterminé à partir de $p$ et $n$ et qui contient $F_n$ avec une probabilité d'autant plus proche de $1-\alpha$ que $n$ est grand.

Un intervalle de confiance pour une proportion $p$ à un niveau de confiance $1-\alpha$ est la réalisation, à partir d'un échantillon, d'un intervalle aléatoire contenant la proportion $p$ avec une probabilité supérieure ou égale à $1-\alpha$, intervalle aléatoire déterminé à partir de la variable aléatoire $F_n$, qui à tout échantillon de taille $n$ associe la fréquence.
Les intervalles de confiance considérés ici sont centrés en $f$.

**<u>En classe de Seconde</u>** :

On considère une population où la fréquence d'un caractère est $p$ et dans laquelle on prélève, au hasard et avec remise, un échantillon de taille $n$, ($n \geq 25$).
On considère l'intervalle $\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$ de centre $p$, $p$ étant compris entre $0,2$ et $0,8$.
On réalise un grand nombre d'échantillons de taille $n$ et l'on observe à chaque fois la fréquence d'apparition $f$ du caractère étudié. On s'aperçoit que les fréquences observées pour ces échantillons de taille $n$ se situent dans cet intervalle $\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$ avec une probabilité au moins égale à $0,95$.

**<u>Exemple</u>** :
*On souhaite savoir si une entreprise exerce une discrimination à l'embauche vis-à-vis du personnel féminin.*
*S'il n'y a pas de discrimination, la proportion de femmes dans cette entreprise devrait être représentative de la proportion de femmes dans la population active. On admet que la proportion de femmes dans la population active est $0,5$.*

1. *En utilisant l'intervalle de fluctuation au seuil $0,95$, déterminer si une entreprise employant $1183$ femmes sur $2540$ salariés exerce une discrimination à l'égard des femmes.*
2. *Quelle doit être le nombre minimal de femmes dans cette entreprise pour que la proportion observée $p_{obs}$ de femmes appartienne à l'intervalle de fluctuation $[0,48;0,52]$ ?*

**<u>Solution</u>** :

1. *La taille de l'échantillon est $n=2540$. On connaît la proportion théorique de femmes : $p=0,50$. Les conditions d'application de la propriété citée ci-dessus sont respectées.*
   *La proportion $p_{obs}$ de femmes dans un échantillon de taille $2540$ appartient dans plus de $95\%$ des cas à l'intervalle $\left[0,5-\frac{1}{\sqrt{2540}};0,5+\frac{1}{\sqrt{2540}}\right]$, soit l'intervalle $[0,48;0,52]$.*
   *Ici, $p_{obs}=\frac{1183}{2540}\approx 0,466$. Donc $p_{obs}\notin[0,48;0,52]$. On en conclut que cette entreprise exerce très probablement une discrimination à l'égard des femmes.*
2. *Pour que $p_{obs}$ appartienne à l'intervalle de fluctuation, on doit avoir $p_{obs}\geq 0,48$. On doit donc avoir un nombbre de femmes supérieur à $0,48\times 2540$ soit 1220. Il faudrait donc au moins $1220$ femmes dans cette entreprise pour que $p_{obs}\in[0,48;0,52]$.*

**<u>En classe de Première</u>** :
Dans une population, on suppose qu'un caractère est présent dans la proportion $p$.
Pour juger de cette hypothèse, on prélève au hasard et avec remise un échantillon de taille $n$ et on observe que la fréquence du caractère étudiée est $f$.
On cherche à savoir pour quelles valeurs de $f$ on pourra considérer qu'elles sont suffisamment éloignés de $p$ pour rejeter l'hypothèse.
La variable aléatoire $X$ compte le nombre d'individus de l'échantillon qui présentent le caractère étudié. Elle suit la loi binomiale de paramètres $n$ et $p$.
On partage l'intervalle $[0;n]$ en trois intervalles : $[0;a-1]$, $[a;b]$ et $[b+1;n]$ de façon que $X$ prenne ses valeurs (entières de $0$ à $n$) dans chacun des deux intervalles extrêmes avec une probabilité proche de $2,5\%$ sans jamais la dépasser.

**L'intervalle de fluctuation au seuil de $95\%$** de la fréquence $f$ est l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ où :

- $a$ est **le plus petit entier** tel que $P(X\leq a)>0,025$ ;
- $b$ est le plus petit entier tel que $P(X\leq b)\geq 0,975$

**Remarques** :

1. Pour $n \geq 30$ , $np \geq 5$ et $n(1-p) \geq 5$ , l'intervalle de fluctuation au seuil de $95\%$ est sensiblement le même que celui donné en classe de Seconde : $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$.
2. On déduit facilement de la définition que $P(X \leq a-1) \leq 0{,}025$ et $P(X \geq b+1) \leq 0{,}025$ .
3. $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a-1)$. Or $-P(X \leq a-1) \geq -0{,}025$ et $P(X \leq b) - P(X \leq a-1) \geq 0{,}975$ soit $P(a \leq X \leq b) \geq 0{,}0975$.

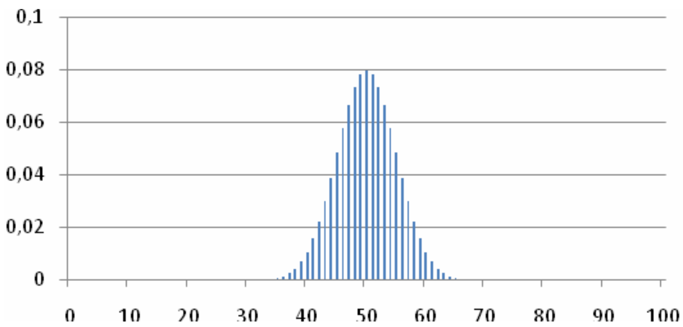
**Règle de décision** :

- Si $f \in \left[\frac{a}{n} ; \frac{b}{n}\right]$, alors **on ne peut pas rejeter l'hypothèse** selon laquelle la proportion du caractère dans la population est égale à $p$.
- Si $f \notin \left[\frac{a}{n} ; \frac{b}{n}\right]$, alors **on rejette l'hypothèse** selon laquelle la proportion du caractère dans la population est égale à $p$ au risque d'erreur de $5\%$ (la probabilité de rejeter l'hypothèse alors qu'elle est vraie est inférieure à $5\%$).

**Exemple :**

*Une entreprise affirme respecter parfaitement la parité homme-femme dans tous les secteurs. On interroge au hasard $100$ ingénieurs de cette entreprise : $36\%$ sont des femmes.*
*On fait l'hypothèse que l'entreprise respecte la parité, c'est-à-dire que la proportion de femmes ingénieurs est $p = 0{,}50$ . La variable aléatoire $X$ , égale au nombre de femmes parmi les $100$ ingénieurs interrogés, suit la loi binomiale de paramètres $100$ et $0{,}50$.*

| k | P(X≤k) | k | P(X≤k) | k | P(X≤k) |
|---|---|---|---|---|---|
| 0 | 7,88861E-31 | 39 | 0,0176001 | 59 | 0,97155603 |
| 1 | 7,9675E-29 | 40 | 0,02844397 | 60 | 0,9823999 |
| 2 | 3,98454E-27 | 41 | 0,04431304 | 61 | 0,98951063 |




*L'intervalle de fluctuation à $95\%$ pour cette loi est $[0{,}40 ; 0{,}60]$.*
*La fréquence observée lors du sondage auprès des $100$ ingénieurs est $f = 0{,}36$ .*
*Comme $f \notin [0{,}40 ; 0{,}60]$, on rejette l'hypothèse $p = 0{,}50$ avec un risque d'erreur de $5\%$. On considère donc au risque $5\%$ que l'entreprise ne respecte pas la parité homme-femme.*

**Remarque** :

En seconde, on obtiendrait avec $p = 0{,}50$ et $n = 100$ , $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right] = \left[0{,}50 - \frac{1}{\sqrt{100}} ; 0{,}50 + \frac{1}{\sqrt{100}}\right] = [0{,}40 ; 0{,}60]$.

**Exercice** :

*Un laboratoire annonce qu'un médicament sauve $40\%$ des patients atteints d'une maladie rare. Pour contrôler cette affirmation, on le teste sur $100$ patients atteints de cette maladie.*
*Soit $X$ le nombre de malades sauvés par ce médicament dans un échantillon aléatoire de malades et assimilé à un tirage avec remise de taille $100$ .*

1. *Quelle loi suit la variable aléatoire $X$ ?*
2. *Déterminer les plus petits entiers $a$ et $b$ tels que $P(X \leq a) > 0{,}025$ et $P(X \leq b) \geq 0{,}975$.*
3. *Enoncer la règle de décision permettant de rejeter ou non l'hypothèse $p = 0{,}40$ selon la valeur de la fréquence $f$ des malades sauvés dans l'échantillon.*
4. *Sur les $100$ malades auxquels on a administré ce médicament, on en a sauvé $30$. Au seuil de risque $5\%$ , que peut-on dire de l'annonce faite par le laboratoire ?*

**Solution** :

1. *L'épreuve de Bernoulli « un patient est sauvé ou non par le médicament » se répète $100$ fois de façon identique et indépendante : $X$ suit donc la loi binomiale de paramètres $100$ et $0,40$.*
2. *A l'aide de la calculatrice ou du tableur, on construit la table des valeurs $P(X \leq k)$ et en parcourant les valeurs de la table, on lit $a = 31$ et $b = 50$.*

| k | P(X≤k) | k | P(X≤k) | k | P(X≤k) |
|---|---|---|---|---|---|
| 0 | 6,53319E-23 | 29 | 0,01477532 | 48 | 0,95769858 |
| 1 | 4,42079E-21 | 30 | 0,02478282 | 49 | 0,9729008 |
| 2 | 1,48151E-19 | 31 | 0,03984788 | 50 | 0,98323831 |

3. *L'intervalle de fluctuation à $95\%$ de la fréquence dans les échantillons de taille $100$ est $[0,31; 0,50]$.*
   *D'où la règle de décision : « si $f$ appartient à l'intervalle $[0,31; 0,50]$, on peut accepter l'hypothèse $p = 0,40$ sinon elle est rejetée au risque d'erreur de $5\%$.*
4. *La fréquence observée est $0,30$ et $0,30$ n'appartient pas à l'intervalle $[0,31; 0,50]$ donc, au risque d'erreur $5\%$, on rejette l'hypothèse selon laquelle ce médicament sauve $40\%$ des malades.*

**<u>En classe de terminale</u>** :

*Approximation de la loi binomiale $B(n, p)$ par la loi normale $\mathcal{N}(np, \sqrt{npq})$.*

La loi binomiale $B(n, p)$ est d'autant plus symétrique que $p$ est proche de $0,50$ et elle prend une forme « en cloche » quand $n$ augmente. Le Théorème Central Limite donne une justification à ce phénomène.

Une valeur de $p$ très différente de $0,50$ ($p$ petit ou $q = 1 - p$ petit) pourra être compensée par une grande valeur de $n$ et on accepte généralement de remplacer la loi binomiale $B(n, p)$ *par la loi normale* $\mathcal{N}(np, \sqrt{npq})$ lorsque $n \geq 30$ , $np \geq 5$ et $n(1 - p) \geq 5$.

**<u>Exemple</u>** :

*On estime que la probabilité qu'une graine ait perdu son pouvoir germinatif après trois ans de conservation est de $70\%$ . Sur un échantillon de $100$ graines conservées depuis trois ans, quelle est la probabilité pour que moins $25$ germent ?*

**<u>Solution</u>** :

*Notons $p$ la probabilité qu'une graine germe $(p = 0,30)$ et considérons que l'échantillon est indépendant.*

*Notons $X$ la variable aléatoire : « nombre de graines qui germent parmi les $100$ ».*

*$X$ suit la loi binomiale de paramètres $100$ et $0,30$.*

*On cherche $P(X < 25)$ qui peut s'écrire aussi $P(X \leq 24)$. La calculatrice ou le tableur donne $P(X \leq 24) \approx 0,114$.*

*En remplaçant la loi binomiale par la loi normale $(np = 30$ et $n(1 - p) = 70)$ , les paramètres de celle-ci seront :*

$m = np = 30$ *et* $\sigma = \sqrt{npq} = \sqrt{21} \approx 4,5826$.

*La variable aléatoire discrète $X$ (correspondant à la loi binomiale $B(n, p) = B(100; 0,30)$ sera alors remplacée par la variable aléatoire continue $X_c$ correspondant à la loi normale $\mathcal{N}(np, \sqrt{npq}) = \mathcal{N}(30; 4,5826)$.*

*Un problème se pose alors : faut-il calculer $P(X_c < 25)$ ou $P(X_1c \leq 24$ ? Pour une variable continue, ces valeurs ne sont pas identiques. La meilleure approximation sera obtenue en prenant la valeur intermédiaire $24,5$ . C'est ce qu'on appelle la « correction de continuité ».*

$$P(X \leq 24,5) = \Pi\left(\frac{24,5 - 30}{4,5826}\right) = \Pi(-1,20) = 1 - \Pi(1,20) = 1 - 0,885 = 0,115.$$

*On peut constater que ceci fournit une excellente approximation de la valeur trouvée à l'aide de la loi binomiale.*

**<u>Prise de décision</u>** :

Il s'agit de comparer la fréquence théorique $p$ et la fréquence observée $f$ .

On formule l'hypothèse nulle $H_0$ : l'échantillon provient d'une population de proportion $p' = p$.

et l'hypothèse alternative $H_1 : p' \neq p$ . Le seuil de risque est $\alpha = 0,05$.

On note $F$ la fréquence d'échantillon.

Sous l'hypothèse $H_0$, $\sigma(F) = \sqrt{\frac{p(1-p)}{n}}$ . On suppose que la loi de $F$ peut être approchée par une loi normale.

***Première méthode*** :

On calcule l'intervalle d'acceptation $I_\alpha$ de la fréquence observée $f$.

$$I_\alpha = [p - t_\alpha \sigma(F); p + t_\alpha \sigma(F)]$$

Si $f$ appartient à $I_\alpha$, on ne rejette pas $H_0$ au risque de $5\%$.

Si $f$ n'appartient pas à $I_\alpha$, on rejette pas $H_0$ au risque de $5\%$.

***Deuxième méthode*** :

On compare la valeur absolue de la différence centrée réduite $t = \dfrac{f-p}{\sigma(F)}$ avec la valeur limite $t_\alpha = 1{,}96$.

Si $|t| < t_\alpha$, on ne rejette pas $H_0$ au risque de $5\%$.

Si $|t| < t_\alpha$, on rejette $H_0$ au risque de $5\%$.

**Exemple** :

**Partie A**

*On effectue des contrôles d'alcoolémie d'automobilistes dans une région donnée, un jour donné, pendant une période horaire donnée. Les statistiques permettent d'établir que la probabilité pour qu'un automobiliste choisi au hasard présente un contrôle positif est* $\dfrac{1}{50}$.

*On considère un échantillon de la population constitué de* $2000$ *automobilistes dont on veut contrôler le taux d'alcoolémie dans les conditions précédentes.*

*On appelle* $X$ *la variable aléatoire prenant pour valeur le nombre de contrôles positifs parmi ces* $2000$.

1. $X$ *suit une loi binomiale. En donner l'espérance et l'écart-type (à* $10^{-2}$ *près).*
2. *Par quelle loi peut-n approcher la loi de* $X$ *? Justifier.*
3. *En utilisant la loi normale de paramètres* $40$ *et* $6{,}26$ *:*
   - **a.** *Calculer les probabilités* $P(X \geq 36)$ *et* $P(33 \leq X \leq 43)$.
   - **b.** *Déterminer une valeur approchée du réel* $\alpha$ *tel que* $P(X \leq \alpha) = 0{,}99$.

     *Quelle est la plus grande valeur de l'entier naturel* $n$ *que l'on peut en déduire tel que* $P(X \leq n) \leq 0{,}99$ *?*

**Partie B**

*Dans une ville donnée de cette région, on effectue au hasard* $200$ *contrôles de taux d'alcoolémie d'automobilistes dans les conditions précédentes.*

*Les taux d'alcoolémie mesurés en g.*$L^{-1}$ *ont donné les résultats suivants :*

| Taux | $[0;0{,}2[$ | $[0{,}2;0{,}4[$ | $[0{,}4;0{,}6[$ | $[0{,}6;0{,}8[$ | $[0{,}8;1[$ | $[1;1{,}2[$ | $[1{,}2;1{,}4[$ | $[1{,}4;1{,}6[$ |
|---|---|---|---|---|---|---|---|---|
| Effectifs | 42 | 78 | 52 | 19 | 4 | 2 | 2 | 1 |

1. *Calculer les valeurs approchées de la moyenne* $\bar{t}$ *et de l'écart-type* $s$ *de cette série statistique.*
2. *Au moment de cette enquête, un contrôle est considéré positif si le taux d'alcoolémie est supérieur à* $0{,}8$ *g.*$L^{-1}$.
   *Calculer la fréquence de contrôles positifs de cet échantillon.*
3. *Cette fréquence permet-elle de conclure que l'échantillon est représentatif au seuil de risque de* $5\%$ *de l'ensemble des contrôles de la région où l'on considère que la proportion des contrôles positifs est* $\dfrac{1}{50}$.

**Eléments de correction** :

**Partie A**

1. $X$ *suit une loi binomiale de paramètres* $n = 2000$ *et* $p = \dfrac{1}{50}$ *donc* $E(X) = np = 40$ *et*
   $$\sigma(X) = \sqrt{2000 \times \frac{1}{50} \times \frac{49}{50}} = \sqrt{39{,}2} \approx 6{,}26.$$
2. *Puisque* $n \geq 30$, $np \geq 5$ *et* $n(1-p) \geq 5$, *on peut approcher la loi de* $X$ *par la loi normale de paramètres* $40$ *et* $6{,}26$.
3. *Soit* $Y$ *la variable aléatoire suivant cette loi normale ; la variable centrée réduite* $T = \dfrac{Y-40}{6{,}26}$ *suit la loi normale* $\mathcal{N}(0;1)$ *de fonction de répartition* $\Pi$.
   - **a.** $P(X \geq 36) = P(Y \geq 35{,}5) = P(T \geq -0{,}72) = 1 - P(T < 0{,}72) = 1 - \Pi(0{,}72) = 0{,}2358.$

$P(33 \leq X \leq 43) = \leq P(32{,}5 \leq Y \leq 43{,}5) = P(-1{,}20 \leq T \leq 0{,}56) = \Pi(0{,}56) - \Pi(-1{,}20) = \Pi(0{,}56) - (1 - \Pi(1{,}20))$

*b.* $P(X \leq a) = P(Y \leq a + 0{,}5) = P\left(T \leq \frac{a - 39{,}5}{6{,}26}\right) = \Pi\left(\frac{a - 39{,}5}{6{,}26}\right) = 0{,}99$ *d'où* $\frac{a - 39{,}5}{6{,}26} = 2{,}33$ *et*

$a = 54{,}09.$

*On en déduit que* $n = 54.$

**Partie B**

1. *On obtient* $\bar{t} = 0{,}39$ *g.*$L^{-1}$ *et* $s = 0{,}25$ *g.*$L^{-1}$
2. *Il y a* $9$ *contrôles positifs soit* $4{,}5\%$
3. *On doit comparer la fréquence théorique* $p = 0{,}02$ *et la fréquence observée* $f = 0{,}045$ *et l'on fait que l'hypothèse que l'échantillon provient d'une population de proportion* $0{,}02.$

   *Soit* $F$ *la fréquence d'échantillon. En supposant que la loi de* $F$ *peut être approchée par une loi normale, o calcule l'intervalle d'acceptation de la fréquence observée :* $[p - t_\alpha \sigma(F); p + t_\alpha \sigma(F)] = [0{,}0006; 0{,}0394]$ *avec*

   $$\sigma(F) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}02 \times 0{,}98}{200}} = 0{,}0099$$

   $f$ *n'appartient pas à cet intervalle donc au rejette l'hypothèse au risque de* $5\%.$

# Savoir-faire

__En classe de seconde__ :

- savoir **simuler une expérience** et **déterminer un intervalle de fluctuation**

*On lance un dé équilibré dont les faces sont numérotées de* $1$ *à* $6$ *.*

*On appelle succès l'obtention de la face* $5$ *ou de la face* $6$ *et échec l'obtention d'une autre face.*

1. *Simuler à l'aide d'un tableur cette épreuve sur un échantillon de taille* $50$ *. On pourra noter* $1$ *pour le succès et* $0$ *pour l'échec.*
2. *Faire figurer dans la feuille de calcul le nombre et la fréquence de succès de l'échantillon.*

*Au jeu de la boule, on a quatre chances sur neuf de gagner lorsque l'on mise sur la couleur rouge.*

1. *Simuler* $50$ *tirages (algorithme, calculatrice, tableur)*
2. *A quel intervalle de fluctuation appartient la fréquence de gain pour une simulation de* $50$ *jeux ?*
3. *En déduire l'intervalle auquel appartient, dans plus de* $95\%$ *des cas, le nombre de fois où l'on gagne avec la simulation de* $50$ *jeux ci-dessus.*
4. *Le résultat de votre simulation est-il en accord avec le résultat précédent ?*

- savoir **représenter graphiquement les distributions de fréquences** et **tenir compte de la taille des échantillons**

*Dans un secteur d'activité où il y a autant de femmes que d'hommes, une petite entreprise* $A$ *de* $35$ *personnes emploie* $40\%$ *de femmes alors qu'une importante société* $B$ *compte* $46\%$ *de femmes parmi ses* $3500$ *salariés.*

1. *Simuler* $100$ *échantillons de taille* $35$ *et* $100$ *échantillons de taille* $3500$ *de lancers d'une pièce supposée équilibrée. Les points obtenus correspondent à la fréquence de « pile » sur chaque échantillon (« pile » représentant « femme »).*
2. *Calculer les bornes de l'intervalle* $I_1 = \left[0{,}50 - \frac{1}{\sqrt{35}}; 0{,}50 + \frac{1}{\sqrt{35}}\right]$ *et le visualiser sur le graphique. Combien de fréquences se trouvent en dehors de l'intervalle* $I_1$ *? Peut-on dire que pour* $95\%$ *au moins de ces échantillons de taille* $35$ *, la fréquence de « pile » appartient à cet intervalle ?*
3. *Mêmes questions avec l'intervalle* $I_2 = \left[0{,}50 - \frac{1}{\sqrt{3500}}; 0{,}50 + \frac{1}{\sqrt{3500}}\right]$ *pour les échantillons de taille* $3500.$
4. *Que peut-on en conclure sur les proportions de femmes dans ces deux entreprises ?*

<u>**En classe de première**</u> :

- savoir **construire une table des valeurs** $P(X \leq k)$ et **représenter graphiquement la loi binomiale par un diagramme en bâtons**

*Soit* $X$ *une variable aléatoire suivant la loi binomiale de paramètres* $n = 50$ *et* $p = 3.$

1. *Construire un tableau des valeurs* $P(X \leq k)$ *à la calculatrice ou avec un tableur.*
2. *Déterminer le plus petit entier* $a$ *tel que* $P(X \leq a) > 0{,}025.$
3. *Déterminer le plus petit entier* $b$ *tel que* $P(X \leq b) \geq 0{,}975.$
4. *Représenter graphiquement la loi de probabilité de* $X$ *par un diagramme en bâtons.*

- savoir **déterminer un intervalle de fluctuation à** $95\%$ **à l'aide de la loi binomiale**

*On s'intéresse à la proportion de faces marquées « 1 » obtenues quand on lance un dé tétraédrique bien équilibré dont les faces sont numérotées* $1, 2, 3$ *et* $4.$

1. *Soit* $X$ *la variable aléatoire égale au nombre de faces marquées «* $1$ *» obtenues quand on lance* $100$ *fois ce dé. Quelle loi suit* $X$ *?*
2. *Déterminer les entiers* $a$ *et* $b$ *tels que* $a$ *soit le plus petit entier tel que* $P(X \leq a) > 0{,}025$ *et* $b$ *le plus petit entier tel que* $P(X \leq b) \geq 0{,}975.$
3. *En déduire l'intervalle de fluctuation à* $95\%$ *de la fréquence des faces marquées «* $1$ *» dans les échantillons de taille* $100.$
4. *Comparer avec l'intervalle obtenu en classe de Seconde.*

- savoir **prendre une décision** (lors d'un essai thérapeutique)

*Pour obtenir l'autorisation de mise sur le marché, les médicaments doivent passer avec succès différents types d'essais. Pour les essais pré-cliniques, un médicament est testé sur des souris. On constate la guérison de* $90\%$ *des souris malades.*
*Pour les essais cliniques, ce même médicament est administré à* $200$ *patients atteints par la maladie. On constate la guérison de* $162$ *malades.*
*On se propose de déterminer si l'efficacité du médicament constatée chez les souris se confirme chez les humains.*
*On suppose que la probabilité de guérison d'un patient est* $p = 0{,}90.$
$X$ *est la variable aléatoire qui compte le nombre de patients guéri dans l'échantillon de* $200$ *patients.*

1. *Expliquer pourquoi* $X$ *suit une loi binomiale et préciser ses paramètres.*
2. *Etablir une feuille de calcul.*
3. *Déterminer les entiers* $a$ *et* $b$ *tels que* $a$ *soit le plus petit entier tel que* $P(X \leq a) > 0{,}025$ *et* $b$ *le plus petit entier tel que* $P(X \leq b) \geq 0{,}975.$
4. *En déduire l'intervalle de fluctuation à* $95\%$ $\left[\frac{a}{n}; \frac{b}{n}\right]$ *puis comparer cet intervalle à l'intervalle de fluctuation* $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$ *au seuil de* $95\%$ *utilisé en classe de Seconde.*

5. *A l'aide d'un tableur, représenter par un diagramme en bâtons la loi de probabilité de* $X$ *puis localiser sur le graphique l'intervalle de fluctuation à* $95\%$.
6. *Déterminer la fréquence de guérison* $f$ *observée chez les humains de cet échantillon.*
7. *Peut-on affirmer, avec un risque d'erreur de* $5\%$ *, que ce médicament est aussi efficace sur les patients que sur les souris ?*

- savoir **trouver un intervalle de fluctuation à l'aide d'un algorithme**

*L'algorithme ci-contre permet de déterminer la plus petite valeur de* $k$ *telle que* $P(X \leq k) > S$ *avec* $X$ *une variable aléatoire qui suit une Loi binomiale de paramètres* $(n, p)$ *et* $S$ *un réel.*

1. *Programmer cet algorithme.*
2. *Déterminer, au seuil de* $95\%$ *, les intervalles de fluctuation obtenus à l'aide du programme précédent d'une fréquence correspondant à la réalisation, sur un échantillon aléatoire de taille* $n$ *, d'une variable aléatoire* $X$ *de loi binomiale de paramètres* $n = 200$ *et* $p = 0{,}30$ *, puis de paramètres* $n = 400$ *et* $p = 0{,}02$.
3. *Comparer les résultats précédents avec l'intervalle de fluctuation vu en classe de Seconde.*

Variables : $S, P, R$ trois nombres réels ; $N, J$ deux entiers naturels
Début
Saisir $N, P, S$
$0 \rightarrow J$
$(1-P)^N \rightarrow R$
**Répète**
$J + 1 \rightarrow J$

Combinaison $(N, J) \times P^J (1-P)^{N-J} + R \rightarrow R$
**Tant que** $R < S$

**Eléments de réponses** :

*Pour* $n = 200$ *et* $p = 0{,}30$ *, on trouve* $a = 48$ *et* $b = 73$ *soit l'intervalle* $[0{,}24; 0{,}365]$.

*Pour* $n = 400$ *et* $p = 0{,}02$ *, on trouve* $a = 3$ *et* $b = 14$ *soit l'intervalle* $[0{,}075; 0{,}035]$.

*Pour* $n = 200$ *et* $p = 0{,}30$ *, l'intervalle utilisé en Seconde est* $\left[0{,}30 - \frac{1}{\sqrt{200}}; 0{,}30 + \frac{1}{\sqrt{200}}\right]$*, soit environ* $[0{,}229; 0{,}371]$.

*Pour* $n = 400$ *et* $p = 0{,}02$ *, les conditions ne sont pas remplies.*

**En classe de Terminale** :

- savoir **se ramener à une loi normale centrée réduite**

*Centrer et réduire une variable, c'est raisonner en nombre d'écarts-types par rapport à la moyenne. Par exemple, si la masse d'un foie gras est une variable de moyenne* $550$ *g et d'écart-type* $100$ *g, on dira d'un foie gras de masse* $x = 650$ *g que, en donnée centrée réduite, il pèse* $t = 1$ *sous-entendu : un écart-type de plus que la moyenne) alors que la masse centrée réduite d'un foie gras de* $x = 500$ *g sera de* $t = -0{,}5$ *(un demi écart-type en dessous de la moyenne).*

*Tous les événements relatifs à la variable aléatoire* $X$ *peuvent donc être exprimés en fonction de la variable aléatoire* $T = \dfrac{X - 550}{100}$. *Plus généralement :* $T = \dfrac{X - m}{\sigma}$ *où,* $m$ *est la moyenne et* $\sigma$ *l'écart-type.*

*Ainsi, la masse d'un foie gras comprise entre* $500$ *g et* $650$ *g* $(500 \leq X \leq 650)$ *pourra être remplacée par la masse centrée réduite comprise entre* $-0{,}5$ *et* $1$ $(-0{,}5 \leq T \leq 1)$.

- savoir **faire des calculs avec la loi normale**

*Si l'on note* $\Pi$ *la fonction de répartition de la loi normale centrée réduite* $\mathcal{N}(0;1)$*, on a donc* $\Pi(t) = P(T \leq t)$.
*On obtient par exemple :*

| t | Π(t) | t | Π(t) | T | Π(t) | t | Π(t) | t | Π(t) | t | Π(t) | t | Π(t) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -1,97 | 0,0244192 | -1,01 | 0,1562476 | -0,51 | 0,3050257 | -0,01 | 0,4960106 | 0,49 | 0,6879331 | 0,99 | 0,8389129 | 1,95 | 0,9744119 |
| -1,96 | 0,0249979 | -1 | 0,1586553 | -0,5 | 0,3085375 | 0 | 0,5 | 0,5 | 0,6914625 | 1 | 0,8413447 | 1,96 | 0,9750021 |
| -1,95 | 0,0255881 | -0,99 | 0,1610871 | -0,49 | 0,3120669 | 0,01 | 0,5039894 | 0,51 | 0,6949743 | 1,01 | 0,8437524 | 1,97 | 0,9755808 |

$\Pi(0) = 0{,}5$
$\Pi(1) = 0{,}8413$
$\Pi(0{,}5) = 0{,}6915$

$\Pi(1{,}96) = 0{,}975$

*Par symétrie, on obtient :*

$\Pi(-1) = 1 - \Pi(1) = 1 - 0{,}8413 = 0{,}1587$

$\Pi(-0{,}5) = 1 - \Pi(0{,}5) = 1 - 0{,}6915 = 0{,}3085$

$\Pi(-1{,}96) = 1 - \Pi(1{,}96) = 1 - 0{,}975 = 0{,}025$

*Si l'on reprend l'exemple du foie gras qui suit une loi normale de paramètres* $550$ *et* $100$ *et que l'on veuille calculer la probabilité que ce foie gras pèse moins de* $650$ *g, plus de* $746$ *g, moins de* $500$ *g ou encore entre* $500$ *et* $600$ *g, on calcule :*

$$P(X < 650) = P\left(T < \frac{650 - 550}{100}\right) = P(T < 1) = \Pi(1) = 0{,}8413$$

$$P(X > 746) = P\left(T > \frac{746 - 550}{100}\right) = P(T > 1{,}96) = 1 - P(T \le 1{,}96) = 1 - \Pi(1{,}96) = 1 - 0{,}975 = 0{,}025$$

$$P(X < 500) = P\left(T < \frac{500 - 550}{100}\right) = P(T < -0{,}5) = \Pi(-0{,}5) = 1 - \Pi(0{,}5) = 1 - 0{,}6915 = 0{,}3085$$

$$P(550 < X < 600) = P\left(\frac{550 - 550}{100} T < \frac{600 - 550}{100}\right) = P(0 < T < 0{,}5) = \Pi(0{,}5) - \Pi(0) = 0{,}6915 - 0{,}5 = 0{,}1915$$

*Rappelons que, pour une variable continue, il n'y a pas de différence entre* $P(X < k)$ *et* $P(X \le k)$*, la probabilité de* $(X = k)$ *étant nulle.*

- savoir **approcher une loi binomiale par une loi normale**

*On considère une variable aléatoire* $X$ *suivant la loi binomiale* $\mathcal{B}(400; 0{,}50)$*. On peut approcher cette loi par la loi normale* $\mathcal{N}(200; 10)$*. En effet* $n$ *et* $p$ *vérifient les conditions requises et* $np = 400 \times 0{,}50 = 200$ *;*

$\sqrt{npq} = \sqrt{400 \times 0{,}50 \times 0{,}50} = \sqrt{100} = 10.$

*Ainsi, si l'on considère la variable* $Y$ *suivant cette loi normale, on approchera* $P(X = 190)$ *par* $P(189{,}5 \le Y \le 190{,}5)$ *en effectuant le changement de variable adéquat.*

*On a* $P(189{,}5 \le Y \le 190{,}5) = P(-1{,}05 \le T \le -0{,}95) = 0{,}0242$ *très proche du calcul de* $P(X = 190)$ *à l'aide de la loi binomiale.*

*De même,* $P(X \le 210)$ *sera approchée par* $P(Y \le 210{,}5) = P(T \le 1{,}05) = 0{,}8531.$

*Si l'on veut approcher* $P(X < 205) = P(X \le 204)$*, il faudra prendre* $P(Y \le 204{,}5)$.

- savoir **prendre une décision**

**1. Exemple de comparaison d'un pourcentage à une valeur de référence (test de la première loi de Mendel)**

*Soit un couple* $[A, a]$ *de gènes où* $A$ *est le dominant. Le croisement d'hétérozygotes fournit des génotypes* $Aa$ *,* $aA$ *,* $aa$ *et* $AA$*. Les génotypes* $Aa$ *,* $aA$ *et* $AA$ *présentent le phénotype* $[A]$ *tandis que le génotype* $aa$ *présente le phénotype* $[a]$*.*

*Lorsque la première loi de Mendel s'applique, les proportions attendues de* $[A]$ *et de* $[a]$ *dans la descendance sont respectivement* $\frac{3}{4}$ *et* $\frac{1}{4}$*.*

*Dans une descendance de* $40$ *individus, on observe* $27$ *individus* $[A]$*.*

*On veut tester si ces données permettent de rejeter la première loi de Mendel avec un risque de première espèce égal à* $5\%$*.*

__Eléments de correction__ *:*

$H_0$ *:* $p = \frac{3}{4}$ *et* $H_1$ *:* $p \neq \frac{3}{4}$

$$\varepsilon_0 = \frac{\left|\frac{27}{40} - \frac{3}{4}\right|}{\sqrt{\frac{3}{4} \times \frac{1}{4} \times \frac{1}{40}}} = 1{,}095$$

*Dans la table de* $\mathcal{N}(0, 1)$*, on lit pour* $\alpha = 0{,}05$ *:* *seuil* $= 1{,}96$

$\varepsilon_0 < 1{,}96$ *donc on ne rejette pas* $H_0$ *.*

__Conclusion__ **: on ne rejette pas la première loi de Mendel au risque** $5\%$**.**

## 2. Exemple de comparaison de deux proportions

*Deux médicaments $A$ et $B$ en concurrence ont été administrés à deux groupes distincts de $50$ personnes atteintes de la même maladie. Les patients sont suivis de façon absolument identique. Après $15$ jours, on remarque que $44$ personnes parmi les $50$ traitées par $A$ ont guéri alors qu'il y en a $40$ sur les $50$ traitées par $B$.*
*On veut tester au risque de $5\%$ si les deux médicaments sont significativement différents.*

<u>**Eléments de correction**</u> :

$H_0 \ : P_1 = P_2$ *et* $H_1 \ : P_1 \neq P_2$

$$P_1 = \frac{44}{50} = 0{,}88 \quad \text{et} \quad P_2 = \frac{40}{50} = 0{,}80$$

$$P = \frac{50 \times 0{,}88 + 50 \times 0{,}80}{100} = 0{,}84$$

$$\varepsilon_0 = \frac{|0{,}88 - 0{,}80|}{\sqrt{0{,}84 \times (1 - 0{,}84) \times \left(\frac{1}{50} + \frac{1}{50}\right)}}$$

*Soit* $\varepsilon_0 \cong 1{,}09$

*Dans la table de* $\mathcal{N}(0,1)$, *on lit pour* $\alpha = 0{,}05$ : *seuil* $= 1{,}96$

$\varepsilon_0 < 1{,}96$ *donc on ne rejette pas* $H_0$ *au risque* $5\%$.

<u>**Conclusion**</u> **: les deux médicaments ne sont pas significativement différents au risque $5\%$.**

## 3. Exemple de comparaison d'une proportion à une valeur de référence

*Sachant que $20\%$ des personnes vaccinées par la formule $F$ d'un vaccin présentent des allergies ou troubles secondaires, un laboratoire pharmaceutique propose une formule améliorée de ce vaccin $F_\alpha$ et espère diminuer le taux d'allergies et de troubles secondaires. Sur un échantillon de $400$ personnes prises au hasard et ayant opté pour la formule $F_\alpha$, on observe $60$ cas d'allergies ou troubles secondaires.*
*On veut tester au risque $\alpha = 0{,}05$ puis au risque $\alpha = 0{,}01$ si la formule améliorée apporte réellement quelque chose de positif.*

<u>**Eléments de correction**</u> :

$H_0 \ : P = 0{,}20$ *et* $H_1 \ : P < 0{,}20$

$$\varepsilon_0 = \frac{\frac{60}{400} - 0{,}20}{\sqrt{\frac{0{,}20 \times (1 - 0{,}20)}{400}}} = -2{,}5$$

*Pour* $\alpha = 0{,}05$ *, on lit (et on calcule) dans la table :* $-1{,}645$

*Pour* $\alpha = 0{,}01$ *, on lit (et on calcule) dans la table :* $-2{,}33$

$\varepsilon_0$ *est inférieur à ces deux valeurs donc dans les deux cas on rejette* $H_0$.

***<u>Conclusion</u> : le taux d'allergie ou troubles secondaires est diminué par la formule $F_\alpha$ du vaccin aux risques de $5\%$ et de $1\%$.***

# Exercices

**Exercice 1** :

*Soit* $X$ *une variable aléatoire suivant la loi binomiale* $\mathcal{B}(100; 0,20)$*. En utilisant une approximation de cette loi par une loi normale dont on précisera les paramètres, calculer une valeur approchée de* $P(X = 20)$, $P(X \leq 22)$, $P(18 \leq X \leq 22)$ *et* $P(X > 18)$*.*

**Exercice 2** :

*Un revendeur de matériel photographique désire s'implanter dans une galerie marchande.*

*Il estime qu'il pourra vendre* $40$ *appareils photographiques par jour et les ventes sont deux à deux indépendantes.*

*Une étude lui a montré que, parmi les différentes marques disponibles, la marque* $A$ *réalise* $38,6\%$ *du marché.*

1. *On note* $X$ *la variable aléatoire qui, un jour donné, associe le nombre d'appareils de marque* $A$ *vendus.*
   - **a.** *Expliquer pourquoi* $X$ *suit une loi binomiale et préciser les paramètres de cette loi.*
   - **b.** *Calculer la probabilité que, sur* $40$ *appareils vendus ce jour,* $20$ *soient de la marque* $A$*. En donner une valeur approchée arrondie à* $0,01$ *près.*
   - **c.** *Calculer l'espérance de* $X$ *et son écart-type arrondi à l'unité.*
2. *On décide d'approcher cette loi par la loi normale de paramètres* $m$ *et* $\sigma$
   - **a.** *Expliquer pourquoi* $m = 15,44$ *et* $\sigma = 3$*.*
   - **b.** *On note* $Y$ *la variable aléatoire suivant la loi normale* $\mathcal{N}(15,44; 3)$*. Donner une approximation à* $0,01$ *près de la probabilité de l'événement : « un jour pris au hasard, il y a exactement* $20$ *appareils de marque* $A$ *vendus », ce qui revient à calculer* $P(19,5 \leq Y \leq 20,5)$*.*
   - **c.** *Déterminer une valeur approchée à* $0,01$ *près de la probabilité de l'événement : « un jour pris au hasard, le nombre d'appareils de marque* $A$ *vendus est compris entre* $15$ *et* $25$ *», ce qui revient à calculer* $P(14,5 \leq Y \leq 25,5)$*.*
   - **d.** *Déterminer une valeur approchée à* $0,01$ *près de la probabilité de l'événement : « un jour pris au hasard,* $20$ *au moins des appareils vendus sont de marque* $A$ *», ce qui revient à calculer* $P(Y \geq 19,5)$*.*

**Exercice 3** :

*Dans une revue, on peut lire : « On estime à* $60,5\%$ *le pourcentage de Français partant au moins une fois en vacances au cours de l'année ».*

*On considère* $100$ *personnes prises au hasard dans la population française (on assimile à un tirage sans remise).*

1. *On désigne par* $X$ *la variable aléatoire mesurant, parmi ces* $100$ *personnes, le nombre de celles qui ne partent pas en vacances au cours de l'année.*
   - **a.** *Expliquer pourquoi* $X$ *suit une loi binomiale et préciser les paramètres de cette loi.*
   - **b.** *Calculer l'espérance et l'écart-type de cette loi (résultats arrondis à* $0,01$ *près).*
   - **c.** *Calculer* $P(X = 45)$*.*
2. *On décide d'approcher cette loi par la loi normale* $\mathcal{N}(39,5; 4,89)$*. Soit* $Y$ *la variable aléatoire suivant cette loi.*
   - **a.** *Calculer une valeur approchée de l'événement : «* $45$ *personnes parmi les* $100$ *ne partent pas en vacances au cours de l'année », ce qui revient à calculer* $P(44,5 \leq Y \leq 45,5)$*.*
   - **b.** *Calculer une valeur approchée de l'événement : « au plus* $30$ *personnes parmi les* $100$ *ne partent pas en vacances au cours de l'année », ce qui revient à calculer* $P(Y \leq 30,5)$*.*

**Exercice 4** :

1. *Au cours d'une répartition de pénicilline en flacons sur une machine automatique, on prélève à intervalles plus ou moins réguliers, un flacon dont on pèse le contenu, au dixième de milligramme. On prélève ainsi un échantillon de* $250$ *flacons dont les masses exprimées en mg se répartissent en* $20$ *classes comme suit :*

| Classes | $[116,5; 117,5[$ | $[117,5; 118,5[$ | $[118,5; 119,5[$ | $[119,5; 120,5[$ | $[120,5; 121,5[$ | $[121,5; 122,5[$ | $[122,5; 123,5[$ | $[123,5; 124,5[$ |
|---|---|---|---|---|---|---|---|---|
| Effectifs | $2$ | $3$ | $4$ | $10$ | $10$ | $16$ | $20$ | $23$ |
| Classes | $[124,5; 125,5[$ | $[125,5; 126,5[$ | $[126,5; 127,5[$ | $[127,5; 128,5[$ | $[128,5; 129,5[$ | $[129,5; 130,5[$ | $[130,5; 131,5[$ | $[131,5; 132,5[$ |
| Effectif | $21$ | $30$ | $25$ | $22$ | $14$ | $15$ | $10$ | $7$ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| s | | | | | | | | | |
| Classes | $[132{,}5;133{,}5[$ | $[133{,}5;134{,}5[$ | $[134{,}5;135{,}5[$ | $[135{,}5;136{,}5[$ | | | | | |
| Effectifs | $5$ | $5$ | $4$ | $4$ | | | | | |

*On appelle $X$ la variable aléatoire qui, à tout flacon, associe la masse de son contenu, exprimée en milligrammes.*

**a.** *Calculer la moyenne $\bar{x}$, la variance $V(X)$ et l'écart-type $s(X)$ de cette série statistique (arrondis à $10^{-3}$).*

**b.** *Quel est le pourcentage de flacons ayant un contenu dont la masse est de moins de $125{,}5$ g ?*

**2.** *Cet échantillon a permis de prendre comme estimation des paramètres de la population correspondante une moyenne $\mu = 126$ et un écart type $\sigma = 4$.*

*On suppose que $X$ suit une loi normale.*

*Calculer $P(X < 125)$, $P(118 < X < 134)$ et $P(X > 129{,}2)$.*

**3.** *On a utilisé un sel de pénicilline titrant $1600$ unités au milligramme. L'activité moyenne par flacon est $201600$ unités. Pour obéir aux prescriptions des normes en vigueur, l'activité trouvée par flacon doit être au moins de $95\%$ de celle annoncée sur l'étiquette, soit $200000$ unités.*

**a.** *Quelle est la masse $m$ du contenu du flacon correspondant à cette activité minimum ?*

**b.** *Quelle est la proportion moyenne de flacons dans la population ayant une masse inférieure à $m$ ?*

<u>**Eléments de correction**</u> :

**1.** $\bar{x} = 126{,}172$ ; $V(X) = 15{,}886$ *et* $s(X) = 3{,}985$

$109$ *flacons ont un contenu de masse inférieure à* $125{,}5$ *g, soit* $43{,}6\%$

**2.** *Soit* $T = \dfrac{X-126}{4}$ *la variable centrée réduite associée à* $X$ *et* $\Pi$ *la fonction de répartition de la loi normale*

$P(X < 125) = P(T < -0{,}25) = \Pi(-0{,}25) = 1 - \Pi(0{,}25) = 1 - 0{,}5987 = 0{,}4013$

$P(118 < X < 134) = P(-2 < T < 2) = \Pi(2) - \Pi(-2) = 2 \times \Pi(2) - 1 = 0{,}9544$

$P(X > 129{,}2) = P(T > 0{,}8) = 1 - P(T \leq 0{,}8) = 1 - \Pi(0{,}8) = 1 - 0{,}7881 = 0{,}2119$

**3.** $m = \dfrac{0{,}95 \times 200000}{1600} = 118{,}75$ *mg*

$P(X < 118{,}75) = P(T < -1{,}8125) = \Pi(-1{,}8125) = 1 - \Pi(1{,}8125) = 0{,}035$

$3{,}5\%$ *des flacons ont une masse inférieure à* $118{,}75$ *mg*

<u>**Exercice 5**</u> :

*Les individus d'une population donnée peuvent être atteints de deux maladies $A$ et $B$.*

*La probabilité qu'un individu pris au hasard dans la population soit atteint de la maladie $A$ est $0{,}03$.*

*La probabilité qu'un individu pris au hasard dans la population soit atteint de la maladie $B$ est $0{,}05$.*

**1.** *On sait que la probabilité qu'un individu pris au hasard dans la population soit atteint de la maladie $B$ sachant qu'il est atteint de la maladie $A$ est $0{,}6$. On choisit au hasard un individu dans cette population.*

**a.** *Calculer la probabilité que cet individu soit atteint à la fois des maladies $A$ et $B$.*

**b.** *Calculer la probabilité que cet individu soit atteint de la maladie $A$ sachant qu'il est atteint de la maladie $B$.*

**c.** *Calculer la probabilité que cet individu soit atteint de la maladie $A$ ou de la maladie $B$.*

**2.** *Dans cette question, on considère un échantillon de la population constitué de $10000$ personnes. Soit $X$ la variable aléatoire ayant pour valeur le nombre de personnes atteintes de la maladie $A$ dans cette population.*

**a.** *Déterminer la loi de probabilité de $X$ et préciser ses paramètres.*

**b.** *Par quelle loi normale peut-on approcher cette loi de probabilité ?*

**c.** *En déduire $P(X < 250)$.*

<u>**Eléments de correction**</u> **:**

**1.** $P(A) = 0{,}03$ , $P(B) = 0{,}05$ *et* $P_A(B) = 0{,}6$

**a.** $P(A \cap B) = P_A(B) \times P(A) = 0{,}018$

**b.** $P_B(A) = \dfrac{P(A \cap B)}{P(B)} = 0{,}36$

**c.** $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0{,}062$

**2.** *On suppose que les personnes de l'échantillon sont choisies de manière indépendante*

a. $X$ suit la loi binomiale de paramètres $n = 10000$ et $p = 0,03$

b. On peut approcher la loi de $X$ par la loi normale de paramètres $\mu = E(X) = np = 300$ et $\sigma = \sigma(X) = \sqrt{np(1-p)} = \sqrt{300 \times 0,97} \approx 17$

c. Soit $Y$ l'approximation normale de $X$ et $T = \dfrac{Y - 300}{17}$ la variable centrée réduite associée à $Y$.
$P(X < 250) = P(Y \leq 249,5) = P(T \leq -2,97) = \Pi(-2,97) = 1 - \Pi(2,97) = 0,0015$

**Exercice 6** :

*Dans un pays d'Afrique, $15\%$ de la population est atteinte du virus du sida.*

**Partie A**

*La stratégie de dépistage met en place un test biologique qui doit être négatif si le sujet est sain et qui doit être positif si le sujet est contaminé.*

*La probabilité qu'un test soit positif sachant que le sujet est sain est $0,004$.*

*La probabilité qu'un test soit négatif sachant que le sujet est contaminé est $0,024$.*

*On choisit au hasard un individu de ce pays.*

1. *Calculer la probabilité que le test soit positif et l'individu sain.*
2. *Calculer la probabilité que le test soit négatif et l'individu contaminé.*
3. *En déduire la probabilité que le résultat soit erroné.*

**Partie B**

*Une campagne de dépistage est mise en place sur un échantillon de $500$ personnes prises au hasard dans la population. On suppose que l'effectif de la population est très grand. On suppose que les risques d'erreur du test sont négligeables et on admet que la probabilité qu'un test réalisé sur une personne prise au hasard dans la population soit positif est $0,15$. on appelle $X$ la variable aléatoire qui prend pour valeur le nombre de tests positifs sur les $500$ tests effectués.*

1. *Quelle est la loi de probabilité de la variable aléatoire $X$.*
   *Calculer l'espérance mathématique et l'écart-type de $X$.*
2. *Par quelle loi normale peut-on approcher la loi définie ci-dessus ?*
3. *En utilisant cette loi normale, calculer la probabilité pour que plus de $80$ tests se révèlent positifs.*

**Eléments de correction** :

**Partie A**

1. $P(\overline{S}) = 0,15$ , $P_S(T) = 0,004$ et $P_{\overline{S}}(\overline{T}) = 0,024$
   $P(T \cap S) = P_S(T) \times P(S) = 0,85 \times 0,004 = 0,003$
2. $P(\overline{T} \cap \overline{S}) = P_{\overline{S}}(\overline{T}) \times P(\overline{S}) = 0,024 \times 0,15 = 0,004$
3. Le résultat est erroné si le test est positif alors que l'individu est sain ou que le test est négatif alors que l'individu est contaminé : c'est donc l'événement $(T \cap S) \cup (\overline{T} \cap \overline{S})$
   Les deux événements sont disjoints donc $P\big((T \cap S) \cup (\overline{T} \cap \overline{S})\big) = P(T \cap S) + P(\overline{T} \cap \overline{S}) = 0,007$

**Partie B**

1. $X$ suit la loi binomiale de paramètres $n = 500$ et $p = 0,15$
   $E(X) = np = 75$ et $\sigma(X) = \sqrt{np(1-p)} = \sqrt{63,75} \approx 7,98$
2. Cette loi binomiale peut être approchée par la loi normale de paramètres $\mu = 75$ et $\sigma = 7,98$
3. Soit $Y$ l'approximation normale de $X$ et $T = \dfrac{Y - 75}{7,98}$ la variable aléatoire centrée réduite associée à $Y$ .
   $T$ suit la loi normale centrée réduite de fonction de répartition $\Pi$.
   $P(X > 80) = P(Y \geq 80,5) = P(T \geq 0,69) = 1 - P(T < 0,69) = 1 - \Pi(0,69) = 0,245$

**Exercice 7** : **Evaluation d'un procédé de fabrication**

*Un laboratoire veut fabriquer des pilules se composant de deux substances $A$ et $B$. Pour chaque pilule on considère les masses $a$ et $b$ respectives des substances $A$ et $B$ qui la constituent.*

*On désigne par $X$ la variable aléatoire qui associe à chaque pilule la masse $a$ de la substance $A$ et par $Y$ la variable aléatoire qui associe à chaque pilule la masse $b$ de la substance $B$.*

*On suppose que ces deux variables aléatoires sont indépendantes et suivent des lois normales de moyennes respectives* $m_X = 8{,}55$ *mg et* $m_Y = 5{,}20$ *mg et d'écarts-types respectifs* $\sigma_X = 0{,}05$ *mg et* $\sigma_Y = 0{,}05$ *mg.*

1. *Déterminer la probabilité* $P(8{,}45 \leq X \leq 8{,}70)$.
2. *Déterminer la probabilité* $P(5{,}07 \leq Y \leq 5{,}33)$.
3. *On impose les normes de fabrication suivantes :* $8{,}45 \leq a \leq 8{,}70$ *et* $5{,}07 \leq b \leq 5{,}33$.
   *Calculer le pourcentage de pilules qui seront hors normes à la sortie de la chaîne de fabrication.*
   *Peut-on alors retenir ce procédé de fabrication, sachant que le pourcentage de pilules défectueuses ne peut dépasser* $1\%$ *?*

**Eléments de correction** :

1. *On pose* $T = \dfrac{X - 8{,}55}{0{,}05}$. $T$ *suit la loi normale* $\mathcal{N}(0;1)$ *de fonction de répartition* $\Pi$.
   $P(8{,}45 \leq X \leq 8{,}70) = P(-2 \leq T \leq 3) = \Pi(3) - \Pi(-2) = 0{,}99865 - 0{,}02275 = 0{,}9759$
2. *On pose* $T' = \dfrac{Y - 5{,}20}{0{,}05}$. $T'$ *suit la loi normale* $\mathcal{N}(0;1)$ *de fonction de répartition* $\Pi$.

$P(5{,}07 \leq Y \leq 5{,}33) = P(-2{,}6 \leq T' \leq 2{,}6) = \Pi(2{,}6) - \Pi(-2{,}6) = 2\Pi(2{,}6) - 1 = 2 \times 0{,}99534 - 1 = \quad 0{,}99068$

3. *Les variables* $X$ *et* $Y$ *sont indépendantes donc*
   $P((8{,}45 \leq X \leq 8{,}70) \cap (5{,}07 \leq Y \leq 5{,}33)) = (8{,}45 \leq X \leq 8{,}70) \times P(5{,}07 \leq Y \leq 5{,}33) = 0{,}9759 \times 0{,}99068 = 0$

   *la proportion de pilules hors normes est donc* $1 - 0{,}9668 = 0{,}0332 = 3{,}32\%$
   *On rejette ce procédé car le pourcentage de pilules défectueuses dépasse* $1\%$

**Exercice 8** *:* **Effet d'une maladie sur une population**
*Une étude statistique réalisée sur une population de* $5$ *millions d'individus atteints d'une maladie a permis de montrer que* $2000$ *d'entre eux sont morts de celle –ci soit une proportion de* $0{,}0004$.
*On appelle* $X$ *la variable aléatoire donnant le nombre de décès dus à cette maladie sur un échantillon de taille* $n$.
*On suppose qu'il y a indépendance des décès constatés et que* $n$ *est inférieur à 100000.*

1. *Déterminer la loi de probabilité de* $X$. *Préciser ses paramètres.*
2. *On considère un échantillon de* $100000$ *individus de cette population.*
   a. *Par quelle loi normale peut-on approcher la loi de probabilité de* $X$ *?*
   b. *On note* $m$ *l'espérance mathématique de* $X$.
      *Déterminer le nombre* $a$ *tel que* $P(m - a \leq X \leq m + a) = 0{,}95$.
      *En déduire le plus petit entier* $n_0$ *tel que* $P(m - n_0 \leq X \leq m + n_0) \geq 0{,}95$.

**Eléments de correction** :

1. *On répète* $n$ *fois la même épreuve de Bernoulli de paramètre* $p = 0{,}0004$ *(probabilité de succès) de manière indépendante ; la variable aléatoire suit la loi binomiale* $\mathcal{B}(n; 0{,}0004)$
2. $n = 10000$ *0*
   a. $E(X) = np = 40$ *et* $\sigma(X) = \sqrt{np(1-p)} = \sqrt{39{,}984} \approx 6{,}32$
      *On peut approcher la loi de* $X$ *par la loi normale* $\mathcal{N}(40; 6{,}32)$
   b. *Soit* $Y$ *l'approximation normale de* $X$ *et* $T = \dfrac{Y - m}{\sigma}$ *la variable aléatoire centrée réduite associée à* $Y$. $T$ *suit la loi normale centrée réduite de fonction de répartition* $\Pi$.
      $P(m - a \leq X \leq m + a) = P\left(\dfrac{-a - 0{,}5}{\sigma} \leq T \leq \dfrac{a + 0{,}5}{\sigma}\right) = 2 \times \Pi\left(\dfrac{a + 0{,}5}{\sigma}\right) - 1 = 0{,}95$
      *On en tire :* $\Pi\left(\dfrac{a + 0{,}5}{\sigma}\right) = 0{,}975 \approx \Pi(1{,}96)$ *d'où* $\dfrac{a + 0{,}5}{\sigma} \approx 1{,}96$ *ce qui donne*
      $a \approx 1{,}96\sigma - 0{,}5 \approx 11{,}89$
      $n_0 = 12$