Chapitre 13

Nul n'est censé ignorer les lois continues

Hors Sujet

Titre : « Laurence Anyways »

Auteur : Xavier Dolan

Présentation succinte : Laurence Anyways est un film franco-québécois, un mélodrame, écrit et réalisé par Xavier Dolan, sorti en 2012.

Le film a reçu le prix de meilleur film canadien au festival international du film de Toronto et le grand prix au festival du film de Cabourg, ainsi que le Prix collégial du cinéma québécois en 2013. C'est l'histoire d'un impossible amour entre un homme et une femme après que celui-ci ait décidé de changer de sexe dans les années 90.

Le jour de ses trente ans, Laurence annonce à Fred qu'il veut devenir une femme et lui demande de l'accompagner dans sa transformation. Pour Fred, c'est un coup de tonnerre, mais elle décide malgré tout de donner une chance à leur couple.

Face aux jugements et à l'incompréhension, Laurence et Fred vont tout faire pour préserver leur amour hors du commun.

Xavier Dolan, né le 20 mars 1989 à Montréal,se fait connaître en tant que scénariste et réalisateur lors de la projection de son premier long métrage : « J'ai tué ma mère » à la 41e Quinzaine des réalisateurs, au cours de la 62e édition du Festival de Cannes. Il y gagne trois prix et les trois jurys soulignent le caractère unique de sa réalisation, la vérité, la violence et la poésie de la langue, ainsi que l'acharnement du jeune cinéaste et la foi en ses projets.

Document réalisé à l'aide de LaTeX
Auteur : C. Aupérin
Site : wicky-math.fr.nf
Lycée Jules Fil (Carcassonne)

Table des matières

L'ESSENTIEL :

- $\rightsquigarrow$ Découvrir les univers infinis et les lois continues
- $\rightsquigarrow$ Connaître trois lois continues
- $\rightsquigarrow$ Approximer une loi binomiale par une loi normale
- $\rightsquigarrow$ Déterminer des probabilités de la loi normale à la calculatrice
- $\rightsquigarrow$ Raisonner par considérations géométriques pour trouver des probabilités de loi normale

« *Télécharger c'est tuer l'industrie, tuons les tous* »

THURSTON MOORE (SONIC YOUTH)

CHAPITRE 13:

NUL N'EST CENSÉ IGNORER LES LOIS CONTINUES

Résumé

Partie II

IV) La loi normale centrée réduite

IV.1. Définition

Proposition 1. (Admise)

Soit f la fonction définie sur $\mathbb{R}$ par $f(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. On a :

$$\lim_{t\to-\infty}\int_t^0 f(x)dx=\frac{1}{2} \qquad \text{et} \qquad \lim_{t\to+\infty}\int_0^t f(x)dx=\frac{1}{2}$$

Ainsi, la fonction f étant de plus clairement continue est positive sur $\mathbb{R}$, est une densité de probabilité sur $\mathbb{R}$.

Définition 1.

La loi normale centré réduite $\mathcal{N}(0,1)$ est la loi de probabilité ayant pour densité la fonction f définie sur $\mathbb{R}$ par

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

Remarques :

⇝ La loi normale modélise des phénomènes naturels très fréquents (d'où son nom), qui résultent de l'addition de plusieurs causes indépendantes, comme la taille d'individus, le taux de cholestérol, les erreurs de mesure ...

⇝ Ainsi, si une variable aléatoire X suit une loi $\mathcal{N}(0,1)$, pour tous réels a et b tels que $a < b$ on a :

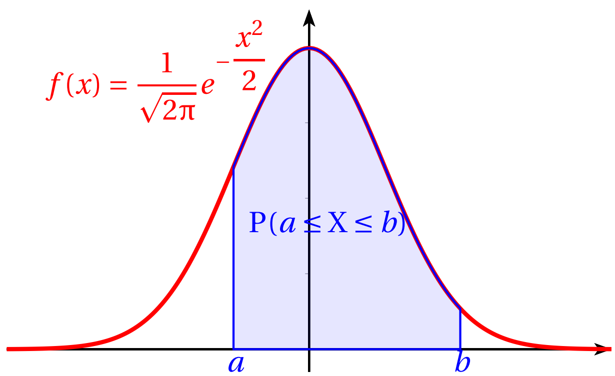
$$\mathrm{P}(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$$

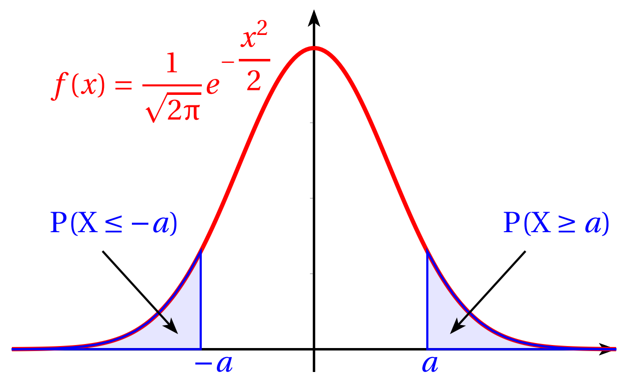
⇝ On sait donc que $\mathrm{P}(X \leq 0) = \mathrm{P}(X \geq 0) = \frac{1}{2}$

⇝ La fonction f est paire, sa courbe est donc symétrique par rapport à l'axe des ordonnées. D'où :

— Pour tout réel a on a $\mathrm{P}(X \leq -a) = \mathrm{P}(X \geq a) = 1 - \mathbb{P}(X < a) = 1 - \mathrm{P}(X \leq a)$

— Si $a \geq 0$ on a aussi $\mathrm{P}(-a \leq X \leq a) = 2\mathrm{P}(0 \leq X \leq a) = 1 - 2\mathrm{P}(X > a) = 2\mathrm{P}(X \leq a) - 1$





Propriété 1.

Soit X une variable aléatoire qui suit une loi $\mathcal{N}(0,1)$. On a :

$$\mathrm{E}(X) = \lim_{x \to -\infty} \int_x^0 t f(t)\, dt + \lim_{x \to +\infty} \int_0^x t f(t)\, dt = 0 \qquad \text{et} \qquad \mathrm{V}(X) = \mathrm{E}\left((X - \mathrm{E}(X))^2\right) = 1$$

Ainsi une variable aléatoire qui suit une loi normale de paramètre 0 et 1 est centrée réduite.

Preuve

$\int_x^0 t f(t)\, dt = \frac{1}{\sqrt{2\pi}} \int_x^0 t e^{-\frac{t^2}{2}}\, dt = \frac{1}{\sqrt{2\pi}} \left[-e^{-\frac{t^2}{2}}\right]_x^0 = \frac{1}{\sqrt{2\pi}} \left(-1 + e^{-\frac{x^2}{2}}\right)$ D'où $\lim_{t \to -\infty} \int_x^0 t f(t)\, dt = -\frac{1}{\sqrt{2\pi}}$

De même $\int_0^x t f(t)\, dt = \frac{1}{\sqrt{2\pi}} \left(e^{-\frac{x^2}{2}} - 1\right)$ D'où $\lim_{x \to +\infty} \int_0^x t f(t)\, dt = \frac{1}{\sqrt{2\pi}}$

Donc $\mathrm{E}(X) = 0$

IV.2. Approximation de la loi binomiale

Le théorème de Moivre-Laplace implique donc, si une variable X_n suit une loi $\mathscr{B}(n,p)$ et que la loi de probabilité de la variable centrée réduite associée $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$, « tend vers » la loi normale centrée réduite, lorsque n devient grand .

Dans la pratique, on estime que quand X suit une loi $\mathscr{B}(n,p)$ avec $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ on peut approximer la variable $Z = \dfrac{X - np}{np(1-p)}$ par une variable Z de loi $\mathcal{N}(0,1)$ puis se ramèner à la loi binomiale de X.

Exemple :

Si X suit la loi binomiale $\mathscr{B}(50;0.2)$ on a

$$P(15 \leq X \leq 30) = P\left(\frac{15-10}{10\times 40} \leq \frac{X-10}{10\times 40} \leq \frac{30-10}{10\times 40}\right) = P\left(0.0125 \leq Z \leq 0.05\right) \simeq \int_{0.0125}^{0.05} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \simeq 0.0150$$

Ce qui est bien plus rapide que d'ajouter chacun des $P(X = k)$ pour k variant de 15 à 30 !

Evidemment, actuellement, vous ne connaissez encore que la méthode fastidieuse, car vous ne connaissez pas de primitive à la densité de la loi normale ... Pour calculer cette intégrale, on a besoin de la calculatrice ou encore de tables de valeurs fournies par les énoncés.

La plupart des calculs liés à la loi normale seront donc des estimations.

Quelques approximations

n	15	50	100	100	100
p	0.4	0.4	0.4	0.02	0.98
$P(3 \leq X_n \leq 12)$	0.97	0.013	6×10^{-10}	0.32	2×10^{-135}
$P\left(\dfrac{3-np}{\sqrt{np(1-p)}} \leq Z \leq \dfrac{12-np}{\sqrt{np(1-p)}}\right)$	0.94	0.010	55×10^{-10}	0.23	4×10^{-10}

On constate l'écart qui peut être important lorsque les critères de validité ne sont pas remplis ...

Exercice du Cours : Pondichéry 2013

Dans une entreprise, on s'intéresse à la probabilité qu'un salarié soit absent durant une période d'épidémie de grippe.

Cette entreprise emploie 220 salariés. On admet que la probabilité pour qu'un salarié soit malade une semaine donnée durant cette période d'épidémie est égale à $p = 0,05$.

On suppose que l'état de santé d'un salarié ne dépend pas de l'état de santé de ses collègues.

On désigne par X la variable aléatoire qui donne le nombre de salariés malades une semaine donnée.

1. Justifier que la variable aléatoire X suit une loi binomiale dont on donnera les paramètres.
   Calculer l'espérance mathématique μ et l'écart type σ de la variable aléatoire X.
2. On admet que l'on peut approcher la loi de la variable aléatoire $\dfrac{X-\mu}{\sigma}$ par la loi normale centrée réduite c'est-à-dire de paramètres 0 et 1.

On note Z une variable aléatoire suivant la loi normale centrée réduite.
Le tableau suivant donne les probabilités de l'évènement $Z < x$ pour quelques valeurs du nombre réel x.

x	$-1,55$	$-1,24$	$-0,93$	$-0,62$	$-0,31$	0,00	0,31	0,62	0,93	1,24	1,55
$P(Z < x)$	0,061	0,108	0,177	0,268	0,379	0,500	0,621	0,732	0,823	0,892	0,939

Calculer, au moyen de l'approximation proposée en question b., une valeur approchée à 10^{-2} près de la probabilité de l'évènement : « le nombre de salariés absents dans l'entreprise au cours d'une semaine donnée est supérieur ou égal à 7 et inférieur ou égal à 15 ».

IV.3. A la calculatrice

IV.3.a. TI82-83-84

Objectif	Affichage voulu	Méthode à suivre

Calculer $P(a \le X \le b)$
où $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$
	`normalFRép(-1.5,2)`
	`.9104427093`
	
	`normalFRép(-1.5,2,1,5)`
	`.2707221556`
		Appuyer sur **2nde** + **var** pour obtenir **distrib**
		Puis choisir `2:normalFRép(`
		Compléter ensuite dans l'ordre les paramètres a et b.
		Préciser éventuellement les paramètres μ et σ si la loi normale n'est pas centrée réduite.
		
		La calculatrice demande σ et non σ^2 en argument !

Déterminer α **tel que**
$P(X \le \alpha) = p$
où $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$
	`FracNormale(.2)`
	`-.8416212335`
	
	`FracNormale(.2,1,5)`
	`-3.208106167`
		Aller dans **distrib**
		Puis choisir `3:FracNormale(`
		Compléter ensuite le paramètre $p \in [0;1]$.
		Préciser éventuellement les paramètres μ et σ si la loi normale n'est pas centrée réduite.

IV.3.b. TI89

Objectif	Affichage voulu	Méthode à suivre

Calculer
$P(a \le X \le b)$ **où**
$X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$
	`■ tistat.normfdr(-1.5,2)`
	`.910443`
	`TIStat.normFdR(-1.5,2)`
	`MAIN RAD AUTO FONC 4/30`
	
	`■ tistat.normfdr(-1.5,2,1 ▸`
	`.270722`
	`…Stat.normFdR(-1.5,2,1,5)`
	`MAIN RAD AUTO FONC 5/30`
		Dans **CATALOG** ouvrir l'onglet **F3 AppFlash**
		Appuyer sur **6** pour aller à la lettre N.
		Choisir `normFdR(....TIStat`
		
		Compléter ensuite dans l'ordre les paramètres a et b.
		Préciser éventuellement les paramètres μ et σ si la loi normale n'est pas centrée réduite.
		
		La calculatrice demande σ et non σ^2 en argument !

Déterminer α **tel**
que $P(X \le \alpha) = p$,
où $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$
	`■ tistat.invnorm(.2)`
	`-.841621`
	`TIStat.invNorm(0.2)`
	`MAIN RAD AUTO FONC 6/30`
	
	`■ tistat.invnorm(.2,1,5)`
	`-3.20811`
	`TIStat.invNorm(0.2,1,5)`
	`MAIN RAD AUTO FONC 7/30`
		Dans **CATALOG** ouvrir l'onglet **F3 AppFlash**
		Appuyer sur **9** pour aller à la lettre I.
		Choisir `invNorm(....TIStat`
		
		Compléter ensuite le paramètre $p \in [0;1]$.
		Préciser éventuellement les paramètres μ et σ si la loi normale n'est pas centrée réduite.

IV.3.c. Nspire CX Cas

Objectif	Affichage voulu	Méthode à suivre
Calculer $P(a \leq X \leq b)$ **où** $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$	normCdf(-1.5,2,0,1) 0.910443 normCdf(-1.5,2,1,5) 0.270722	Dans l'onglet $2:\int\sum$ du catalogue Ouvrir la catégorie `Probabilités` . Puis la sous-catégorie `Distributions` Choisir `Normale FdR` Compléter ensuite dans l'ordre les paramètres a et b. Préciser éventuellement les paramètres μ et σ si la loi normale n'est pas centrée réduite. *La calculatrice demande σ et non σ^2 en argument !*
Déterminer α **tel que** $P(X \leq \alpha) = p$, **où** $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$	invNorm(0.2,0,1) -0.841621 invNorm(0.2,1,5) -3.20811	Dans l'onglet $2:\int\sum$ du catalogue Ouvrir la catégorie `Probabilités` . Puis la sous-catégorie `Distributions` Choisir `Inverse Normale` Compléter ensuite le paramètre $p \in [0;1]$. Préciser éventuellement les paramètres μ et σ si la loi normale n'est pas centrée réduite.

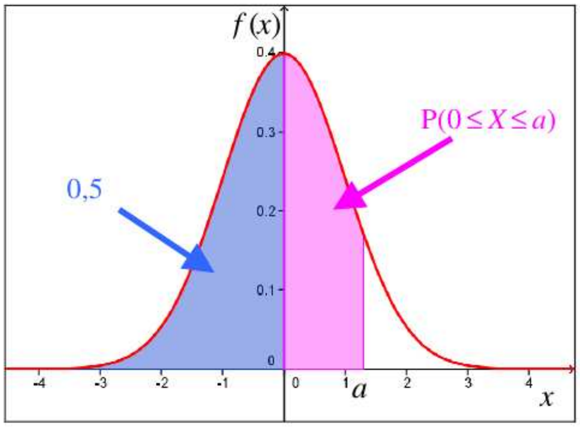
Remarques :

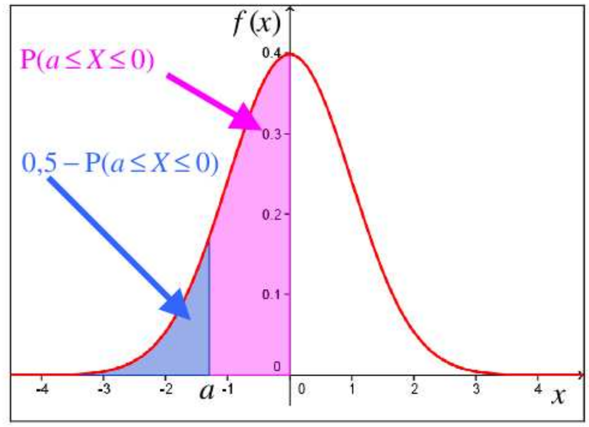
- Pour calculer $P(X \leq 1)$, il faudra penser à écrire $P(X \leq 1) = P(X \leq 0) + P(0 \leq X \leq 1) = \dfrac{1}{2} + P(0 \leq X \leq 1)$, ce qui s'obtient désormais à la calculatrice. On trouve environ 0.84.

- De même, pour avoir $P(X \leq -1)$ on écrira $P(X \leq -1) = P(X \leq 0) - P(-1 \leq X \leq 0) = \dfrac{1}{2} - P(-1 \leq X \leq 0)$. On trouve environ 0.16.

- Visualiser les aires sous la courbe aide à faire ses transformations.
  De manière générale, on peut voir que pour calculer $P(X \leq a)$:

  Si $a \geq 0$ on utilise $P(X \leq a) = 0.5 + P(0 \leq X \leq a)$ Si $a \leq 0$ on utilise $P(X \leq a) = 0.5 - P(a \leq X \leq 0)$





IV.4. Répartition de l'aire sous la courbe

Théorème 1.

Soit X une variable aléatoire suivant la loi normale centrée réduite $\mathcal{N}(0,1)$.
Pour tout réel $\alpha \in]0;1[$, il existe un unique réel positif u_α tel que

$$P(-u_\alpha \leq X \leq u_\alpha) = 1-\alpha$$

Preuve

On considère la fonction g définie sur $\mathbb{R}^+$ par $g(t) = P(-t \leq X \leq t) = \int_{-t}^{t} f(x)dx$ où f est la fonction densité de la loi normale centrée réduite.
On cherche à applique le TVI à g. Comme f est paire, on a pour tout t positif, $g(t) = 2\int_0^t f(x)dx$
De plus, f est continue et positive, donc $2f$ aussi, et g est la primitive de $2f$ qui s'annule en 0 (d'après le théorème fondamental).
Ainsi g est dérivable sur $\mathbb{R}^+$ donc continue sur $\mathbb{R}^+$.
Ensuite, il est clair que $2f$ strictement positive implique que g est strictement croissante sur $\mathbb{R}^+$.
Enfin, comme $\alpha \in]0;1[$ on a $1-\alpha \in]0,1[$ et on a $g(0) = 0$ et $\lim\limits_{t\to+\infty} g(t) = \lim\limits_{t\to+\infty} 2\int_0^t f(x)dx = 2\times\frac{1}{2} = 1$.
D'après le corollaire du TVI, il existe un uniquement réel $u_\alpha \in]0;1[$ tel que $g(u_\alpha) = 1-\alpha$.

Exercice du Cours : Extrait de Nouvelle Calédonie mars 2014

Restitution organisée des connaissances

L'objectif de cette partie est de démontrer le théorème suivant :

> Si X est une variable aléatoire suivant la loi normale centrée réduite, alors pour tout réel α appartenant à l'intervalle]0 ; 1[, il existe un unique réel strictement positif χ_α tel que $P(-\chi_\alpha \leqslant X \leqslant \chi_\alpha) = 1-\alpha$.

Soit f la fonction définie sur l'ensemble des nombres réels $\mathbb{R}$ par

$$f(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}}.$$

Soit H la fonction définie et dérivable sur $[0\,;\,+\infty[$ par

$$\mathrm{H}(x) = \mathrm{P}(-x \leqslant \mathrm{X} \leqslant x) = \int_{-x}^{x} f(t)\,\mathrm{d}t.$$

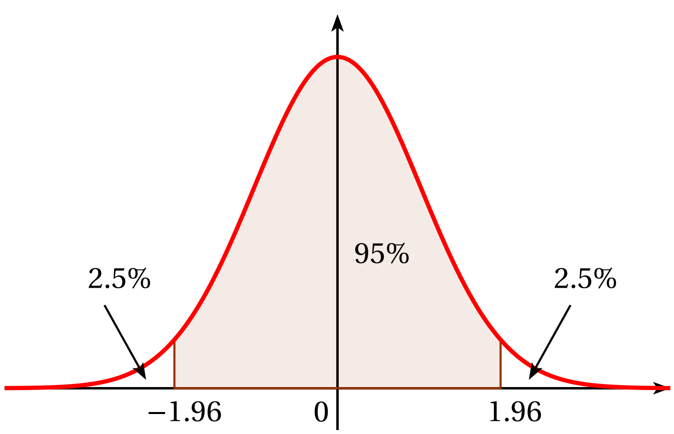
1. Que représente la fonction f pour la loi normale centrée réduite ?
2. Préciser $\mathrm{H}(0)$ et la limite de $\mathrm{H}(x)$ quand x tend vers $+\infty$.
3. À l'aide de considérations graphiques, montrer que pour tout nombre réel positif x, $\mathrm{H}(x) = 2\displaystyle\int_0^x f(t)\,\mathrm{d}t$.
4. En déduire que la dérivée H' de la fonction H sur $[0\,;\,+\infty[$ est la fonction $2f$ et dresser le tableau de variations de H sur $[0\,;\,+\infty[$.
5. Démontrer alors le théorème énoncé.

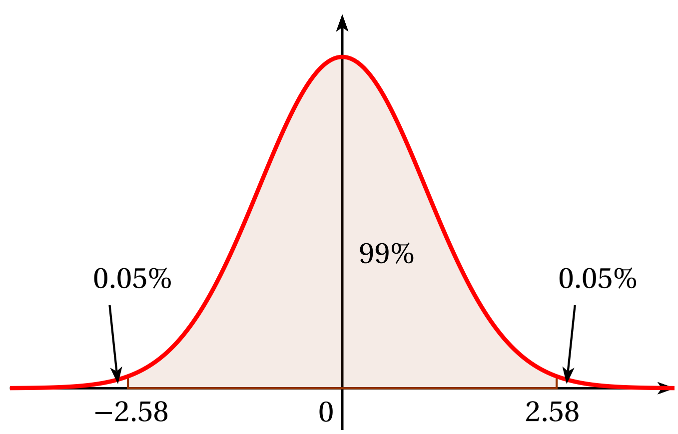
Corollaire 1.

Avec les notations précédentes, on retiendra deux valeurs : $u_{0.05} \simeq 1.96$ et $u_{0.01} \simeq 2.58$.
Ainsi, $\boxed{\mathrm{P}(-1.96 \leq \mathrm{X} \leq 1.96) \simeq 0.95}$ et $\boxed{\mathrm{P}(-2.58 \leq \mathrm{X} \leq 2.58) \simeq 0.99}$

Remarque : Cela donne une idée de la répartition :

- Environ 95% des réalisations de X sont entre -1.96 et 1.96.
- Environ 99% des réalisations de X sont entre -2.58 et 2.58.





V) Fluctuation et Estimation

Des statistiques économiques jusqu'aux extraterrestres ...

Vous maitrisez assez bien les statistiques descriptives depuis le collège (moyenne, quartiles, etc).
Mais les statistiques inférentielles sont tout autre. Vous les avez effleurés du doigt en Seconde pour la première fois, vous les avez retrouvés en première très succinctement, mais le domaine est vaste.

C'est en 1746 que les statistiques endossent leur premier grand rôle dans la vie moderne : celui d'un outil de prévision et non de description.
Le mathématicien Deparcieux établit dans un essai le « profil » de la mortalité de populations à partir de données statistiques venant d'échantillons pris dans les registres et nécrologies. En se servant des méthodes d'échantillonnage, de calcul de moyennes et d'écarts-types, Deparcieux crée les premières « tables de mortalité », permettant d'évaluer le risque moyen de mort d'un individu en fonction de son profil (âge, sexe, profession, ...). Ce risque est alors directement transformé en pécule, par exemple dans le calcul du montant de rentes viagères (rente versée à quelqu'un durant toute sa vie en échange de l'acquisition de son bien à sa mort). Avec Deparcieux, la statistique fait son entrée dans l'économie.

Mais ce n'est qu'au XIX^e siècle que la statistique finit de prendre la place qu'est la sienne aujourd'hui : celle d'une science mathématique mais aussi humaine, omniprésente dans le débat publique. C'est Adolphe Quételet, astronome belge, qui intègre dans un ouvrage toutes les lois de probabilités développées depuis Pascal et Fermat : mesure des erreurs, méthode des moindres carrés, loi binomiale, etc.
Quételet croit à la possibilité d'établir une science mathématique humaine capable de relier les phénomènes sociaux de masse à des lois.
Ainsi, il tente d'établir des lois statistiques des suicides et des crimes en fonction de paramètres comme l'origine sociale, l'âge, le sexe, le climat, le niveau d'études, le revenu, etc. Mais Quételet se voit reprocher de faire de l'homme un être dont le comportement est prédéterminé par des lois mathématiques...

De plus, avec les statistiques sociales, une question se pose : sont-ce les statistiques qui déterminent nos comportements ou l'inverse ?

Par exemple, si le taux de meurtriers dans la population est de 5% par an, cela signifie-t-il qu'il existe une sorte de loi qui nous dépasse et qui « oblige » 5% de personnes à se transformer en meutriers ?
Ce type de questionnement hantera tout le XX^e siècle et conduira à de tragiques dérapages où l'on voudra « neutraliser » dès le berceau tout homme né avec les « paramètres statistiques du crime »...

Dans tous les cas, la statistique inférentielle, armée des lois de probabilité, envahit le XX^e siècle avec une force qui déborde la planète. Ainsi, depuis les années 1960, les astronomes se servent des statistiques pour pister les extraterrestres !

V.1. Echantillonnage : rappels de Seconde et de Première

Travail de l'élève 1 : En Syldavie, Fabrice fait des études statistiques sur le daltonisme des éléphants syldaves.

Il sait que 46% des éléphants syldaves sont des mâles et il pense que 18% des éléphants ont plus de 60 ans.

Il a prélevé un échantillon de 400 éléphants syldaves parmi lesquels 195 sont des mâles et 313 ont moins de 60 ans.

L'échantillon a été réalisé par un tirage au hasard et il peut être assimilé à un tirage avec remise.

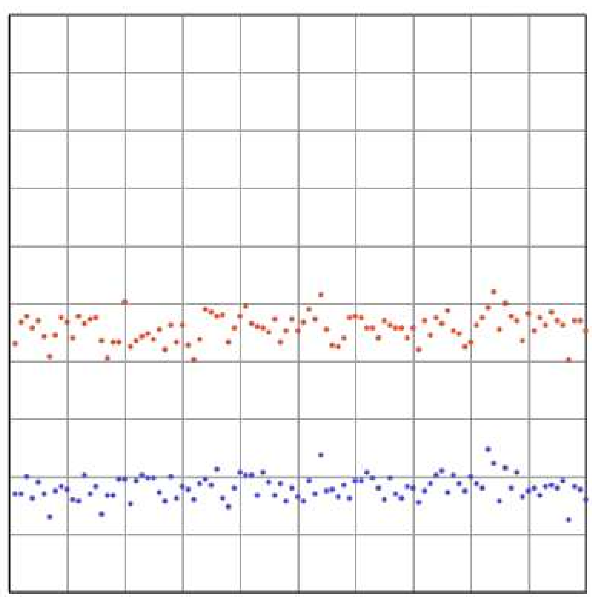
1. L'échantillon de Fabrice est-il représentatif pour le sexe des éléphants ?
2. Son hypothèse sur le nombre d'éléphants âgés de plus de 60 ans est-elle crédible au risque de 5% ?
3. Fabrice trouve que dans cet échantillon, 29% des éléphants sont daltoniens.
   Estimer la proportion d'éléphants daltoniens dans la population d'éléphants syldaves.

Démarche pour mener l'activité : On laisse les élèves chercher et comprendre eux-mêmes l'expression « échantillon représentatif », ainsi que la nécessité d'un critère de décision.

On attend notamment les mots « fréquence » et « fluctuation d'échantillonnage ».

On peut parallèlement proposer aux élèves le programme suivant sur algobox qui simule la réalisation de 100 échantillons de tailles 400, et renvoit les nuages de points associés aux fréquences d'apparition des caractères mâles et plus de 60 ans.

```
VARIABLES
  a EST_DU_TYPE NOMBRE
  Males EST_DU_TYPE NOMBRE
  Vieux EST_DU_TYPE NOMBRE
  compteur EST_DU_TYPE NOMBRE
  compteur2 EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  POUR compteur ALLANT_DE 1 A 100
    DEBUT_POUR
    Males PREND_LA_VALEUR 0
    Vieux PREND_LA_VALEUR 0
    POUR compteur2 ALLANT_DE 1 A 400
      DEBUT_POUR
      a PREND_LA_VALEUR random()
      SI (a<=0.46) ALORS
        DEBUT_SI
        Males PREND_LA_VALEUR Males+1
        FIN_SI
      SI (a<=0.18) ALORS
        DEBUT_SI
        Vieux PREND_LA_VALEUR Vieux+1
        FIN_SI
      FIN_POUR
    TRACER_POINT (compteur,Males/400)
    TRACER_POINT (compteur,Vieux/400)
    FIN_POUR
FIN_ALGORITHME
```



Xmin: 0 ; Xmax: 100 ; Ymin: 0 ; Ymax: 1 ; GradX: 10 ; GradY: 0.1

La notion d'échantillon représentatif est une question délicate, en particulier lorsqu'elle concerne des personnes dans le cadre d'un sondage. Elle l'est clairement moins lorsqu'il s'agit d'un échantillon de pièces dans une chaine de fabrication. Cette notion d'échantillon représentatif est évoquée ici afin de contextualiser un peu l'activité mais ne constitue en aucun cas un objectif du programme.
Il convient également de souligner que, dans les sondages, les tirages sont pour la plupart effectués sans remise mais peuvent s'apparenter à des tirages avec remise dès que la taille de l'échantillon est petite devant la taille de la population totale, ce qui est le cas dans les sondages classiques.
On peut d'ailleurs observer que dans le cas contraire, l'intérêt de ne questionner qu'un échantillon diminue.

Critère de décision

Pour une étude statistique sur un caractère connu de la population (âge, sexe, taille, etc), on considère en général :

- ⇝ qu'un échantillon de taille n est **représentatif**,
- ⇝ ou encore qu'une hypothèse sur la propabilité d'apparition p de ce caractère dans la population totale est crédible

si la fréquence f observée de ce caractère dans l'échantillon est dans l'**intervalle de fluctuation au seuil de** 95%.

Remarques :

- ⇝ Il s'agit d'un intervalle dans lequel se situent 95% des échantillons établis dans ces conditions.

- ⇝ Dans la pratique, cet intervalle de fluctuation sert à :
  - — vérifier qu'un échantillon est représentatif quand on connait p
  - — vérifier la crédibilité d'une hypothèse sur p

  Ainsi, si la fréquence observée f n'appartient pas à l'intervalle de fluctuation I au seuil de 95%, alors :
  - — on considérera que l'échantillon n'est pas représentatif
  - — on rejettera l'hypothèse sur p

  I ne contenant que 95% des échantillons établis dans les mêmes conditions, on risque de rejeter 5% d'hypothèses correctes sur p « en trop ».

  Par contre, si $f \in \mathrm{I}$, alors :
  - — on considérera que l'échantillon est représentatif
  - — on concluera que l'hypothèse sur p est **crédible**

  Cette fois, on ne connait pas le risque d'erreur, ie le nombre d'hypothèses fausses sur p acceptées « en trop ».

  Dans tous les cas, on est sûr de rien ! Donc on évitera le vocabulaire « vrai » ou « faux ».

$\rightsquigarrow$ Tout est une question d'équilibre :
Si l'on veut diminuer l'erreur de rejet, par exemple en prenant un intervalle de fluctuation au seuil de 100%, on ne rejettera donc aucune hypothèse « en trop », par contre, on les acceptera toutes, donc évidemment, même les fausses. Toute hypothèse sur p semblera crédible, ce qui n'a aucun intérêt.
Ainsi, dans la pratique, on utilise surtout les seuils de 95% et de 99%.

On en vient aux rappels de seconde et de première.

En seconde

Pour un caractère donné, on note p sa probabilité d'apparition et f sa fréquence d'apparition dans un échantillon.
La fréquence f se situe dans au moins 95% des cas dans l'intervalle $\left[p-\dfrac{1}{\sqrt{n}}, p+\dfrac{1}{\sqrt{n}}\right]$.

- **Avantages** : rapide à déterminer
- **Inconvénients** : valable uniquement si $n \geq 25$ ET $0.2 \leq p \leq 0.8$
  Sans parler du fait que cet intervalle sort de nulle part et n'a jamais été justifié par votre enseignant ...

On peut appliquer ce résultat pour le sexe des éléphants, mais pas pour l'âge.
On trouve $\mathrm{I} = [0.41; 0.51]$ et $f = \dfrac{195}{400} = 0.4875 \in \mathrm{I}$.
Donc l'échantillon est représentatif pour le caractère sexe.

En Première

On appelle X la variable aléatoire qui compte le nombre de personnes d'un échantillon de taille n, possédant un certain caratère, de probabilité d'apparition p. On a donc $\mathrm{X} \hookrightarrow \mathrm{B}(n, p)$
On note $f = \dfrac{\mathrm{X}}{n}$ la fréquence d'apparition de caractère.
La fréquence f se situe dans au moins 95% des cas dans l'intervalle $\left[\dfrac{a}{n}, \dfrac{b}{n}\right]$, où :

$\rightsquigarrow$ a désigne le plus petit entier k à partir duquel $\mathrm{P}(\mathrm{X} \leq k)$ dépasse 0.025.

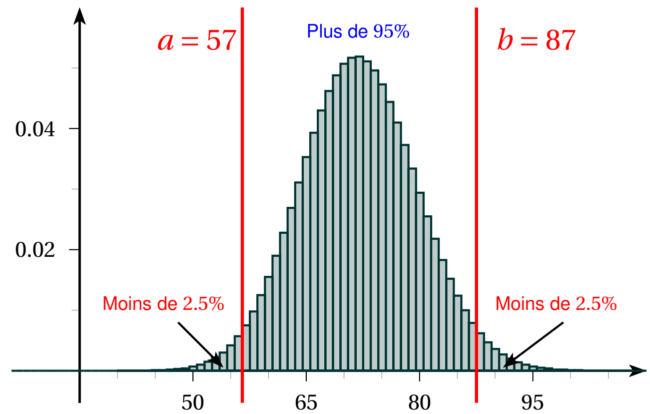
$\rightsquigarrow$ b désigne le plus petit entier k à partir duquel $\mathrm{P}(\mathrm{X} \leq k)$ dépasse 0.975.

- **Avantages** : valable pour toutes les valeurs de n et p
  Sans parler du fait que cet intervalle est justifié par vos connaissances sur la loi binomiale ...
- **Inconvénients** : Très fastidieux si n est grand.

On peut appliquer ce résultat pour l'âge des éléphants.

k	$P(X \leq k)$
...	...
54	0.00953
55	0.01375
56	0.01944
57	**0.02699**
58	0.03678
59	0.04924
...	...

k	$P(X \leq k)$
...	...
83	0.94587
84	0.94587
85	0.95825
86	0.96821
87	**0.97609**
88	0.98225
...	...



Pour l'âge, on trouve $I' = [0.1425; 0.2175]$ et $f = \dfrac{400-313}{400} = 0.2175 \in I'$.
Donc l'hypothèse $p = 0.18$ est crédible.
Notons que pour le sexe, on trouve $I' = [0.4125; 0.51]$ et $I \subset I'$. De plus, $f = \dfrac{195}{400} \in I'$.
Donc l'échantillon est bien représentatif pour le caractère sexe.

Pour la question suivante, on ignore la proportion p d'éléphants daltoniens dans la population.
On cherche cette fois un *intervalle de confiance*.

En Seconde

Pour un caractère donné, on note f sa fréquence d'apparition dans un échantillon de taille n.
Sa probabilité p d'apparition dans la population totale se situe dans au moins 95% des cas dans l'intervalle

$$\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$$

- **Avantages** : rapide à déterminer
- **Inconvénients** : valable uniquement si $n \geq 30$ ET $nf \geq 5$ ET $n(1-f) \geq 5$
  Sans parler du fait que cet intervalle sort de nulle part et n'a jamais été justifié par votre enseignant ...

Ainsi, on estime que la proportion de daltoniens est dans l'intervalle $[0.24; 0.34]$.

Méthodes

Soit p la probabilité d'apparition d'un caractère dans une population totale.

⇝ **Lorsque l'on connait** p, ou lorsque l'on fait une hypothèse sur la valeur de p, on utilise un **intervalle de fluctuation à** 95% pour estimer la représentativité d'un échantillon de taille n ou la crédibilité de l'hypothèse sur p.

⇝ **Lorsque l'on ne connait pas** p, on utilise un **intervalle de confiance à** 95% pour estimer sa valeur (dans la population totale) à partir de la fréquence f observée dans un échantillon de taille n.

V.2. Intervalle de fluctuation asymptotique

Travail de l'élève 2 : On reprend le contexte de l'activité précédente.

Fabrice veut vérifier la crédibilité de son résultat sur la proportion d'éléphants daltoniens dans la population d'éléphants syldaves. Il fait donc l'hypothèse $p = 0.29$ et prend un échantillon de 1200 éléphants.

1. **a.** Pourquoi aimeront-on trouver une autre méthode que celles employées précédemment pour savoir si l'échantillon de Fabrice est représentatif ?

   b. Quel nouvel outil de Terminale pourrait-on utiliser ?

   c. Proposer un nouvel intervalle de fluctuation au seuil de 95%.
   Quelle différence a-t-il avec les précédents ? Quels sont ses avantages ? ses inconvénients ?

2. L'étude de Fabrice montre que dans cet échantillon, 32% des éléphants sont daltoniens.
   Son hypothèse $p = 0.29$ est-elle crédible ?

Théorème 2. (Définition)

Soit X_n une variable aléatoire suivant une loi binomiale $\mathscr{B}(n, p)$ et α un réel tel que $0 < \alpha < 1$.
Soit Z une variable aléatoire suivant la loi normale $\mathscr{N}(0,1)$ et u_α l'unique réel tel que :

$$\mathrm{P}\left(-u_\alpha \leq \mathrm{Z} \leq u_\alpha\right) = 1 - \alpha$$

On note I_n l'intervalle :

$$\mathrm{I}_n = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

Alors $\displaystyle\lim_{n \to +\infty} \mathrm{P}\left(\frac{\mathrm{X}_n}{n} \in \mathrm{I}_n\right) = 1 - \alpha$

Ainsi, l'intervalle I_n contient la fréquence $\mathrm{F}_n = \dfrac{\mathrm{X}_n}{n}$ avec une probabilité qui se rapproche de $1 - \alpha$ lorsque n augmente : on dit que

- α est le risque (5%, 1% ...)
- I_n est un **intervalle de fluctuation asymptotique** de F_n au seuil $1 - \alpha$.

***Preuve* ROC**

On pose $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ et on applique le théorème de Moivre-Laplace.
Si X suit la loi normale $\mathcal{N}(0,1)$:

$$\lim_{n\to+\infty} P(-u_\alpha \leq Z_n \leq u_\alpha) = P(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$$

Or

$$\begin{aligned}
-u_\alpha \leq Z_n \leq u_\alpha &\iff -u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha \\
&\iff np - u_\alpha\sqrt{np(1-p)} \leq X_n \leq np + u_\alpha\sqrt{np(1-p)} \\
&\iff p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + u_\alpha \frac{\sqrt{np(1-p)}}{\sqrt{n}} \\
&\iff F_n \in I_n
\end{aligned}$$

Donc $\lim\limits_{n\to+\infty} P\left(\dfrac{X_n}{n} \in I_n\right) = \lim\limits_{n\to+\infty} P(-u_\alpha \leq Z_n \leq u_\alpha) = 1-\alpha$

Remarques :

- I_n 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.
- Quand on sait qu'une suite (u_n) converge vers une limite L, on peut considérer que pour n assez grand le terme u_n constitue une approximation de L.
  Ici, on inverse les rôles. On connaît la limite, mais pas les valeurs des termes de la suite. On admet donc que, sous certaines conditions, on peut approcher le terme de rang n de la suite $P\left(\dfrac{X_n}{n} \in I_n\right)$ par sa limite $1-\alpha$.
- On considère que l'approximation est satisfaisante dès que $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$

Corollaire 2.

L'intervalle de fluctuation asymptotique au seuil de 95% pour une variable aléatoire X_n suivant une loi binomiale $\mathscr{B}(n,p)$ est l'intervalle :

$$I_n = \left[p - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

Preuve

On a vu précédemment que $u_{0.05} \simeq 1.96$ et $u_{0.01} \simeq 2.58$

Remarques :

- Ainsi, environ 95% des fréquences observées se situent dans l'intervalle ci-dessus.
- Quel est l'intervalle de fluctuation asymptotique au seuil de 99% ?
- L'intervalle de fluctuation vu en seconde est une approximation de l'intervalle I_n au seuil de 95%.

  En effet, $p+1.96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq p+2\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}$

  De plus, en étudiant la fonction f définie sur $[0,1]$ par $f(x)=x(1-x)$, on constate que son maximum est $\dfrac{1}{4}$.

  Ainsi, pour tout $p \in [0,1]$, on a $p(1-p) \leq \dfrac{1}{4} \iff \sqrt{p(1-p)} \leq \dfrac{1}{2}$.

  Par conséquent, $p+1.96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq p+2\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq p+\dfrac{1}{\sqrt{n}}$

  De même on a : $p-1.96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \geq p-2\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \geq p-\dfrac{1}{\sqrt{n}}$

  Ainsi $\left[p-1.96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}};p+1.96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \subset \left[p-\dfrac{1}{\sqrt{n}};p+\dfrac{1}{\sqrt{n}}\right]$.

  Ceci prouve que l'intervalle de fluctuation vu en seconde est un intervalle de fluctuation asymptotique à un seuil au moins égal à celui de Terminale (proche de 0.95).

  Compte tenu du caractère asymptotique, il reste inexact d'affirmer que la probabilité que la variable $\dfrac{\mathrm{X}_n}{n}$ prenne ses valeurs dans cet intervalle est supérieur à 0.95 pour toute valeur de n, d'où les conditions $n \geq 25$ et $0.2 \leq p \leq 0.8$.

  En réalité, on a encore simplifié les conditions, car par exemple, pour $p=0.35$, on faudrait prendre $n \geq 31$, pour $p=0.5$, il faudrait $n \geq 529$ ...

Exemple :

On admet que dans la population d'enfants de 11 à 14 ans d'un département français le pourcentage d'enfants ayant déjà eu une crise d'asthme dans leur vie est de 13%.

Un médecin d'une ville de ce département est surpris du nombre important d'enfants le consultant ayant des crises d'asthme et en informe les services sanitaires. Ceux–ci décident d'entreprendre une étude et d'évaluer la proportion d'enfants de 11 à 14 ans ayant déjà eu des crises d'asthme.

Ils sélectionnent de manière aléatoire 100 jeunes de 11 à 14 ans de la ville.

La règle de décision prise est la suivante : si la proportion observée est supérieure à la borne supérieure de l'intervalle de fluctuation asymptotique au seuil de 95% alors une investigation plus complète sera mise en place afin de rechercher les facteurs de risque pouvant expliquer cette proportion élevée.

1. Déterminer l'intervalle de fluctuation asymptotique au seuil de 95% de la proportion de jeunes de 11 à 14 ans ayant eu une crise d'asthme dans un échantillon de taille 100.
2. L'étude réalisée auprès des 100 personnes a dénombré 19 jeunes ayant déjà eu des crises d'asthme. Que pouvez-vous conclure ?
3. Le médecin n'est pas convaincu par cette conclusion et déclare que le nombre de personnes interrogées était insuffisant pour mettre en évidence qu'il y avait plus de jeunes ayant eu des crises d'asthme que dans le reste du département.
   Combien faudrait-il prendre de sujets pour qu'une proportion observée de 19% soit en dehors de l'intervalle de fluctuation asymptotique ?

Solution **:**

1. $[0,06;0,20]$
2. La valeur 0,19 est à l'intérieur de l'intervalle de fluctuation asymptotique au seuil de 95%.
   On en conclut que la règle de décision choisie ne prévoit pas de réaliser une enquête supplémentaire.
3. Il faut et il suffit que la borne supérieure de l'intervalle asymptotique de fluctuation soit inférieur à 0.19, ce qui équivaut à $0.13+1.96\dfrac{\sqrt{0.13\times0.87}}{\sqrt{n}}<0.19$.
   On trouve $n>120$.
   La taille doit donc être de 121 sujets au minimum si on souhaite mettre en évidence une proportion anormalement élevée dans la ville étudiée.

V.3. Intervalle de confiance et estimation

Il est souvent difficile pour des raisons à la fois financières et logistiques de pouvoir recueillir des données sur la population toute entière. Le plus souvent, on se contente de travailler sur un échantillon, c'est à dire une fraction ou sous-ensemble de cette population. Ceci présente bien sûr des avantages en termes de faisabilité et de coût, mais impose des contraintes pour que l'information recueillie au niveau de l'échantillon (estimation) soit la plus proche possible de celle de la population entière. La démarche pratique est donc la suivante :

- $\rightsquigarrow$ On sélectionne un échantillon de la population que l'on étudie, on appelle cela l'échantillonnage.
- $\rightsquigarrow$ On vérifie, selon les cas, à partir d'intervalles de fluctuation que l'échantillon ainsi obtenu est « **représentatif** » de la population pour des critères qui sont connus dans la population et en lien avec le critère étudié.
- $\rightsquigarrow$ On généralise les informations recueillies sur notre échantillon à la population entière.

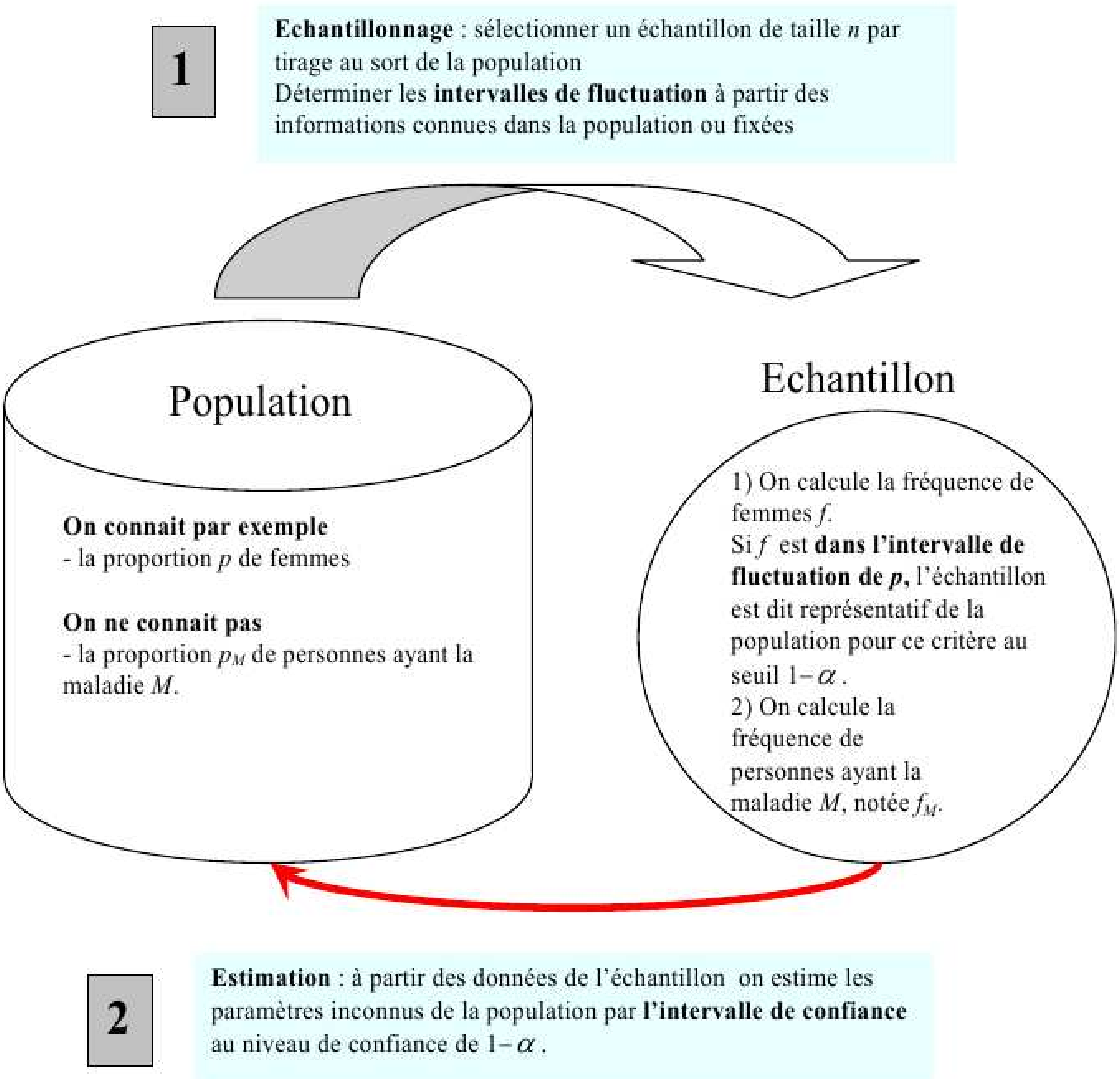
Attention !

La problématique s'inverse : à partir d'une **fréquence observée** f d'un caractère dans un échantillon, on **estime** une **proportion** p dans une population.

Cependant, on sait que notre estimation va varier d'un échantillon à l'autre, de par la fluctuation d'échantillonnage, autour de p.

Comment faire pour limiter la marge d'erreur ?

Il est donc nécessaire d'apprécier l'incertitude en fournissant une estimation par intervalle, appelé **intervalle de confiance** de p. Cet intervalle est obtenu en fonction d'un coefficient lié au niveau de confiance que l'on accorde à cette estimation.



On a déjà rappelé ce que l'on faisait en Seconde. Et bien c'est simple, on n'a pas mieux en Terminale ! Donc on reste avec le même **intervalle de confiance** au seuil de 95%, à savoir :

$$\mathrm{J} = \left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$$

Preuve

On sait que pour n suffisamment grand

$$\begin{aligned}
& \mathrm{P}\left(p-\frac{1}{\sqrt{n}} \leq \mathrm{F}_n \leq p+\frac{1}{\sqrt{n}}\right) \geq 0.95 \\
\iff \quad & \mathrm{P}\left(-\frac{1}{\sqrt{n}} \leq \mathrm{F}_n - p \leq \frac{1}{\sqrt{n}}\right) \geq 0.95 \\
\iff \quad & \mathrm{P}\left(-\frac{1}{\sqrt{n}} - \mathrm{F}_n \leq -p \leq \frac{1}{\sqrt{n}} - \mathrm{F}_n\right) \geq 0.95 \\
\iff \quad & \mathrm{P}\left(\frac{1}{\sqrt{n}} + \mathrm{F}_n \geq p \leq -\frac{1}{\sqrt{n}} + \mathrm{F}_n\right) \geq 0.95 \\
\iff \quad & \mathrm{P}\left(\mathrm{F}_n - \frac{1}{\sqrt{n}} \leq -p \leq \mathrm{F}_n \frac{1}{\sqrt{n}}\right) \geq 0.95
\end{aligned}$$

Ce qui peut se traduire ainsi : l'intervalle $\left[\mathrm{F}_n - \frac{1}{\sqrt{n}} ; \mathrm{F}_n + \frac{1}{\sqrt{n}}\right]$ a une probabilité au moins égale à 0.95 de contenir p.

A partir de cet intervalle aléatoire, on obtient, en effectuant un échantillon, une réalisation de cet intervalle numérique de la forme $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$.

Si l'on fait un très grand nombre de tirages par échantillon, on sait que théoriquement on devrait [a] avoir pour au plus 5% des échantillons, des intervalles ne contenant pas la proportion inconnue p.

a. Il s'agit toujours d'un nombre fini de réalisations et il peut y avoir plus de 5% d'entre elles qui ne contiennent pas p.

Remarques :

- On se place dans le cas $n \geq 30$, $nf \geq 5$ et $n(1-f) \geq 5$ pour considérer notre estimation convenable.
- On pourrait avoir de meilleures approximations, mais elles sont hors-programme
- Un intervalle de confiance étant un intervalle numérique, il est incorrect de conclure la détermination d'un intervalle de confiance par une phrase du type « p a une probabilité de 0,95 d'être entre $f - \frac{1}{\sqrt{n}}$ et $f + \frac{1}{\sqrt{n}}$ » car il n'y a plus d'aléatoire à ce stade. Il est en revanche convenable d'écrire : « $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ est un intervalle de confiance de la proportion inconnue p au niveau de confiance 0,95 ».

Résumé

	Intervalle de **fluctuation** Au seuil 0.95 (p connue)	Intervalle de **confiance** Au seuil 0.95 (p inconnue)
SECONDE	$n \geq 25$, $0.2 \leq p \leq 0.8$ $\mathrm{I} = \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$	Sensibilisation
PREMIERE	Avec la loi binomiale $\mathrm{I} = \left[\frac{a}{n}; \frac{b}{n}\right]$	
TERMINALE	$n \geq 30$, $np \leq 5$ et $n(1-p) \leq 5$ Asymptotique $\mathrm{I}_n = \left[p - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$	$\mathrm{J} = \left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$

Exemple :

Une entreprise désire contrôler la qualité de sa production.

Un échantillon de 370 produits a été testé. Il ressort de cette étude que 26 produits sur 370 ne sont pas conformes au cahier des charges.

On considère que la production de l'entreprise est suffisamment importante pour qu'on puisse assimiler le tirage de l'échantillon à un tirage avec remise.

1. A partir de cette étude, estimer la probabilité qu'un produit ne soit pas conforme au cahier des charges à l'aide d'un intervalle de confiance au niveau asymptotique 95%.
2. Combien faudrait-il prélever de produit pour avoir un intervalle de confiance au niveau 95% de longueur inférieure à 0.05 ?

VI) La loi normale $\mathcal{N}(\mu;\sigma^2)$

VI.1. Définition

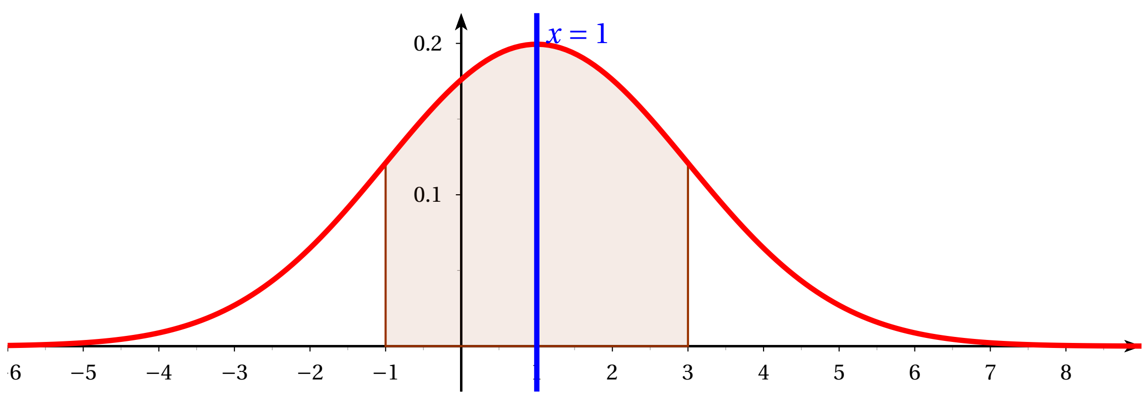
Définition 2.

Soient un réel μ et un réel strictement positif σ.
On dit qu'une variable aléatoire X suit la loi normale $\mathcal{N}(\mu,\sigma^2)$ si la variable aléatoire $\dfrac{X-\mu}{\sigma}$ suit la loi normale centrée réduite $\mathcal{N}(0,1)$.

Remarque : Dans la pratique, grâce au théorème de Moivre-Laplace, on approxime une loi binomiale $\mathcal{B}(n,p)$ par une loi $\mathcal{N}(\mu,\sigma^2)$ avec $\mu = np$ et $\sigma^2 = np(1-p)$.
On considère que l'approximation est satisfaisante dès que $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

Exemple :

Loi normale $\mathcal{N}(1,4)$: $\mu = 1$ et $\sigma = 2$



Propriété 2.

Soit une variable aléatoire X qui suit la loi normale $\mathcal{N}(\mu,\sigma^2)$. On a :

$$E(X) = \mu \quad , \quad V(X) = \sigma^2 \quad \text{et} \quad \sigma(X) = \sigma$$

Preuve

Par définition, $\dfrac{X-\mu}{\sigma}$ suit la loi $\mathcal{N}(0,1)$. On a donc

$$E\left(\frac{X-\mu}{\sigma}\right) = 0 \iff \frac{1}{\sigma}\big(E(X)-\mu\big) = 0 \iff E(X)-\mu = 0 \iff E(X) = \mu$$

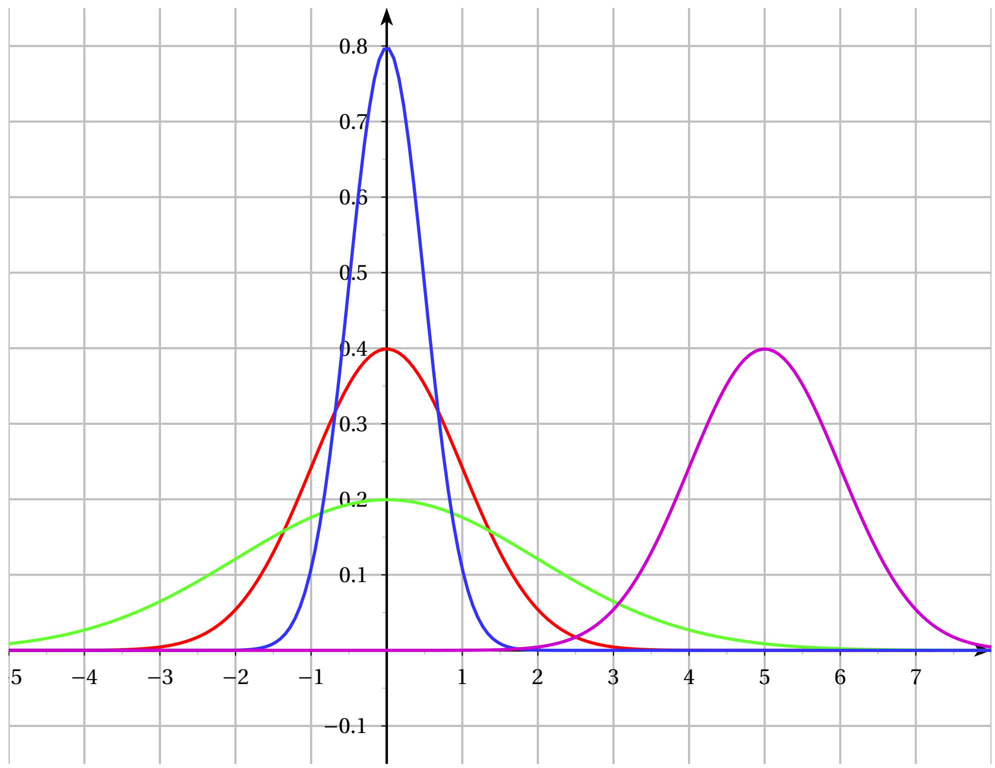
De même,

$$V\left(\frac{X-\mu}{\sigma}\right) = 1 \iff \frac{1}{\sigma^2}V(X) = 1 \iff V(X) = \sigma^2 \iff \sigma(X) = |\sigma| = \sigma$$

Remarques :

- μ est un paramètre de position et σ un paramètre de dispersion.
- La densité de la loi normale $\mathcal{N}(\mu,\sigma)$ est représentée par une courbe en cloche, dont l'axe de symétrie vertical est $x=\mu$.
  Plus σ est petit, et moins la dispersion est grande, donc plus la cloche est resserrée autour de son axe de symétrie.

Reconnaître les lois $\mathcal{N}(0,1)$, $\mathcal{N}(0,1/4)$, $\mathcal{N}(0,2)$ et $\mathcal{N}(5,1)$ parmi les quatre courbes représentées ci-dessous.



VI.2. Méthodes de calcul

$X \hookrightarrow \mathcal{N}(\mu,\sigma^2)$. Par exemple $X \hookrightarrow \mathcal{N}(1,4)$

1. Calculer $P(a \leq X \leq b)$. Par exemple $P(-1 \leq X \leq 3)$.

- On se ramène à une loi normale centrée réduite en écrivant :
  $P(-1 \leq X \leq 3) = P(-2 \leq X-1 \leq 2) = P\left(-1 \leq \dfrac{X-1}{2} \leq 1\right) = P(-1 \leq Z \leq 1)$ où $Z \hookrightarrow \mathcal{N}(0,1)$.
  A la calculatrice, on trouve $P(-1 \leq X \leq 3) \simeq 0.683$.
- Ou alors on calcule directement cette valeur sur la calculatrice en rajoutant les paramètres μ et σ aux commandes déjà connues.

2. Déterminer α tel que $P(X \leq \alpha) = p$, avec p donné appartenant à $[0;1]$.
Par exemple, on cherche α tel que $P(X \leq \alpha) = 0.9$.

- On se ramène à une loi normale centrée réduite en écrivant :
  $P(X \leq \alpha) = P(\leq X-1 \leq \alpha-1) = P\left(\frac{X-1}{2} \leq \frac{\alpha-1}{2}\right) = P(Z \leq \beta)$ où $Z \hookrightarrow \mathcal{N}(0,1)$ et $\beta = \frac{\alpha-1}{2}$.
  A la calculatrice, on trouve $\beta \simeq 1.28$ donc $\alpha \simeq 3.56$.
- Ou alors on trouve directement α sur la calculatrice en rajoutant les paramètres μ et σ aux commandes déjà connues.

Attention !
Les calculatrices demandent μ et σ en arguments, et non σ^2.
Si on ne les précise pas, elle considère que $\mu = 0$ et $\sigma = 1$ (donc elle fait les calculs pour une loi normale centrée réduite).

Exemple :
La masse en kg des nouveaux nés à la naissance est une variable aléatoire qui peut-être modélisée par une loi normale [a] de moyenne $\mu = 3.3$ et d'écart-type $\sigma = 0.5$.

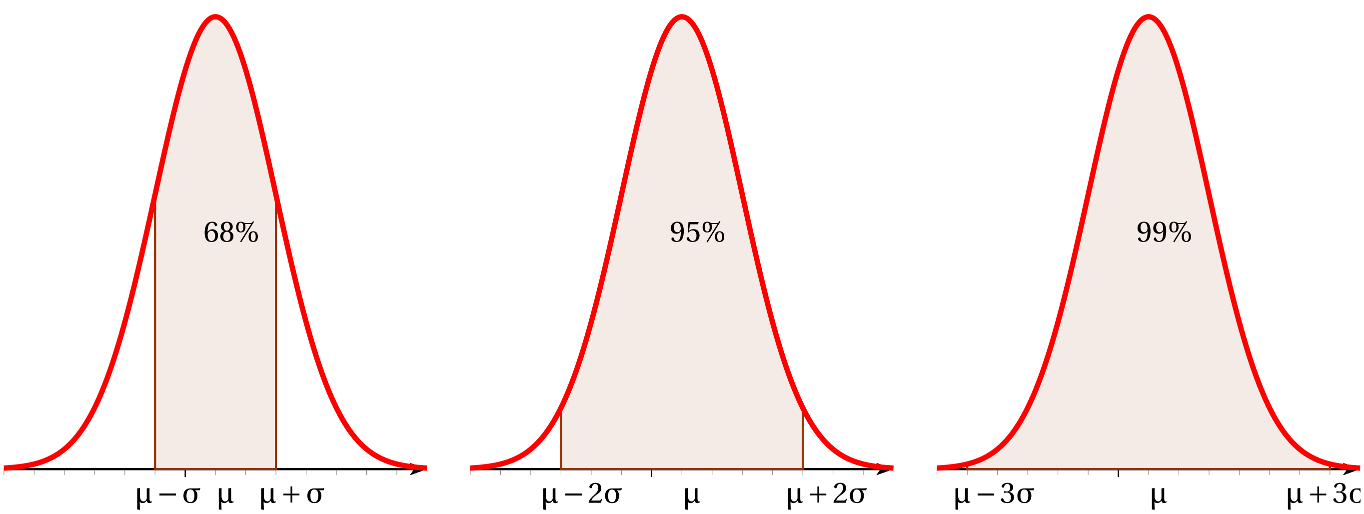
1. Déterminer $P(X < 2.5)$ en vous ramenant à une loi normale centrée réduite.
   Retrouver $P(X < 2.5)$ en utlisant la symétrie de la courbe représentative de la densité de la loi $\mathcal{N}(3.3, 0.5)$.

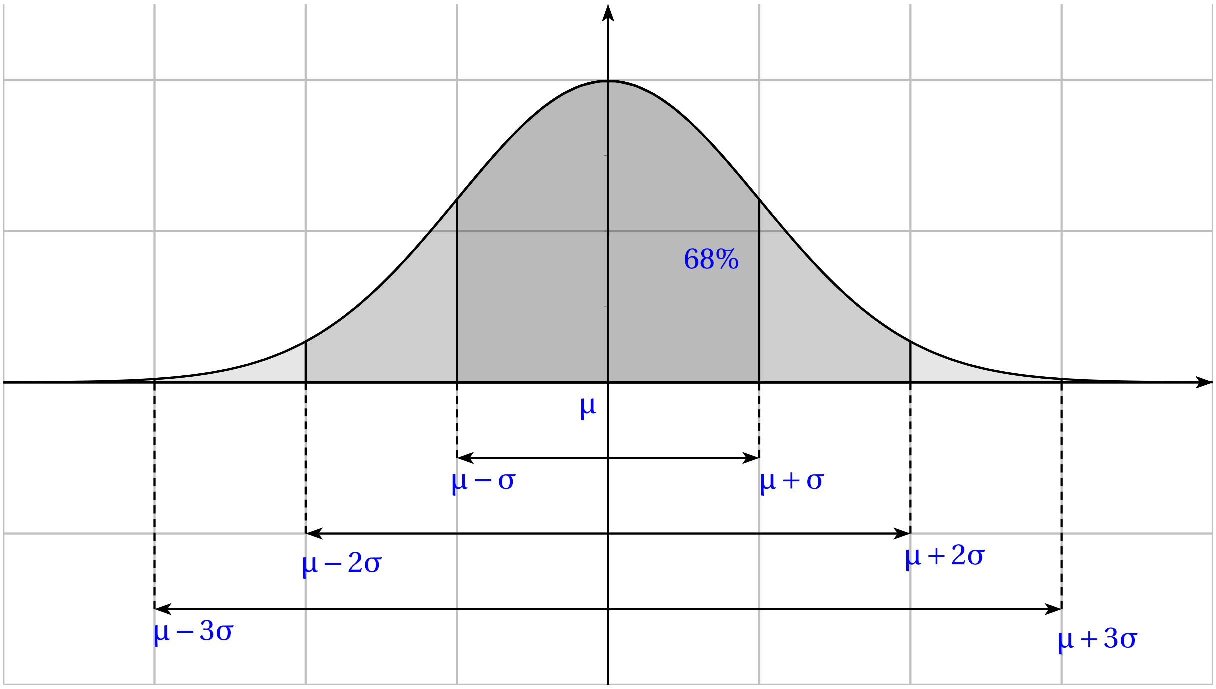
a. Le poids d'un nouveau né ne prend pas de valeurs négatives ni trop grandes, mais on peut vérifier que $P(X < 0)$ et $P(X > 5)$ sont négligeable.

VI.3. Répartition de l'aire sous la courbe

Propriété 3.

Soit une variable aléatoire X qui suit la loi normale $\mathcal{N}(\mu, \sigma^2)$. On a :
- $P(\mu-\sigma \leq X \leq \mu+\sigma) \simeq 0.683$
- $P(\mu-2\sigma \leq X \leq \mu+2\sigma) \simeq 0.954$
- $P(\mu-3\sigma \leq X \leq \mu+3\sigma) \simeq 0.997$





Preuve

$$P(\mu-\sigma \leq X \leq \mu+\sigma) = P(-\sigma \leq X-\mu \leq \sigma) = P\left(-1 \leq \frac{X-\mu}{\sigma} \leq 1\right) \simeq 0.683$$

car $\dfrac{X-\mu}{\sigma}$ suit la loi $\mathcal{N}(0,1)$.
De même pour les autres approximations.

Exercice du Cours : Une compagnie d'empaquetage de jus possède un équipement versant une quantité X de jus dans chacune des bouteilles de format 1000 ml. Cette quantité peut varier un peu d'une bouteille à l'autre.

On considère alors que X suit une loi normale de paramètre μ et $\sigma^2 = 25$ ml^2 et on appelle Z la variable aléatoire $Z = \dfrac{X-\mu}{\sigma}$.

1. Dans un premier temps, on considère $\mu = 1000$.
   a. Déterminer la probabilité qu'une bouteille contiennent moins de 1000 ml.
   b. Déterminer l'intervalle dans lequel se situe 99% de la production.
2. Reprendre les deux questions précédentes avec $\mu = 1005$ ml.
3. La législation impose qu'il y ait moins de 1% des bouteilles contiennent moins de 1000 ml.
   a. Déterminer μ pour que l'embouteilleur respecte la législation et interpréter.
   b. La contenance des bouteilles étant de 1020 ml, quelle est la probabilité qu'une bouteille déborde lors du remplissage ?
   c. Un inspecteur choisit un échantillon de 20 bouteilles, au hasard et avec remise, parmi la production de la journée.
   L'embouteilleur recevra une amende si l'inspecteur trouve au moins une bouteille contenant moins de 1000 ml.

On appelle Y la variable aléatoire qui compte le nombre de bouteilles dans l'échantillon qui contiennent moins de 1000 ml. Quelle est la probabilité que l'embouteilleur reçoive une amende ?

Solution :

1. **a.** Par symétrie de la courbe, on sait que $P(X \leq 1000) = \dfrac{1}{2}$

   b. On sait que

   c. $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \simeq 0.99$ donc $X \in [985; 1015]$

2. **a.** On a $P(X \leq 1000) = 0.5 - P(1000 \leq X \leq 1005) = 0.5 - \dfrac{1}{2}P(1000 \leq X \leq 1010) \simeq 0.5 - \dfrac{1}{2}0.683 = 0.1585$

   Ou encore, à la calculatrice, en calculant $P(1000 \leq X \leq 1005)$, on trouve environ 0.1587, donc presque pareil.

   b. On sait que $X \in [1005 - 15; 1005 + 15]$ ie $X \in [990; 1020]$ dans 99% des cas.

3. **a.** On sait que Z suit une loi normale $\mathcal{N}(0; 1)$. On cherche alors μ tel que

$$P(X \leq 1000) = 0.01 \iff P\left(\frac{X-\mu}{\sigma} \leq \frac{1000-\mu}{\sigma}\right) = 0.01 \iff P\left(Z \leq \frac{1000-\mu}{5}\right) = 0.01$$

   On cherche donc α tel que $P(Z \leq \alpha) = 0.01$.
   Grâce à la calculatrice, on trouve $\alpha \simeq -2.326$. Ainsi

$$\frac{1000-\mu}{5} \simeq -2.326 \iff \mu \simeq 1011.632$$

   L'embouteilleur doit verser en moyenne 1012 ml par bouteille pour que environ 1% des bouteilles contiennent moins de 1000 ml.
   On peut vérifier en cherchant à la calculatrice α *tel que* $P(X \leq \alpha) = 0.01$ *pour la loi de* X.
   On doit évidemment trouver une valeur proche de 1000.

   b. A la calculatrice, avec $\mu \simeq 1011.6$, on trouve $P(X > 1020) = 0.5 - P(1011.6 < X < 1020) \simeq 0.046$.

   c. On répète 20 fois de manière identique et indépendante, l'épreuve de Bernoulli qui consiste à choisir une bouteille, le succès étant « la bouteille contient moins de 1000 ml », de probabilité 0.01 d'après ce qui précède.
   Y compte le nombre de succès, donc Y suit une loi binomiale de paramètres 20 et 0.01.
   On cherche $P(Y \leq 1) = 1 - P(Y = 0) = 1 - 0.99^2 0 \simeq 0.18$
   La probabilité que l'embouteilleur reçoive une amende est de 0.18.