QUOI DE NOUVEAU EN PROBABILITE EN TERMINALE ?

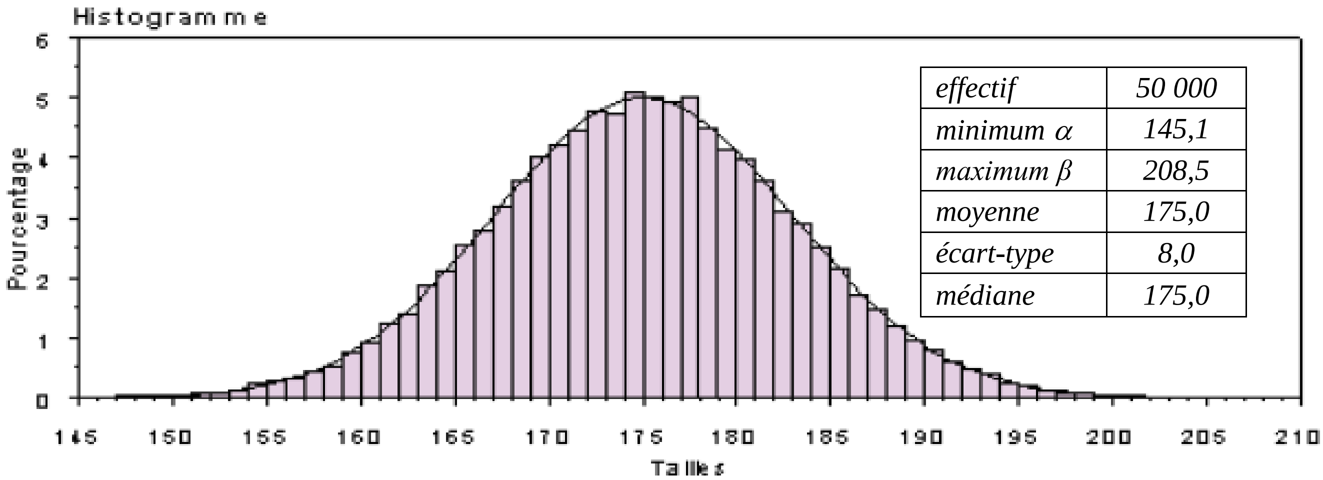
	S	**ES et L**	**STI2D et STL**
Probabilités	Conditionnement Indépendance	Conditionnement	
Lois à densité	Loi uniforme Loi exponentielle	Loi uniforme	Loi uniforme Loi exponentielle
Lois normales	Théorème de Moivre-Laplace Loi standard $\mathcal{N}(0,1)$ Loi $\mathcal{N}(\mu\ ;\ \sigma^2)$	Loi standard $\mathcal{N}(0,1)$ Loi $\mathcal{N}(\mu\ ;\ \sigma^2)$	Loi $\mathcal{N}(\mu\ ;\ \sigma^2)$ Approcher une binomiale par une normale
Fluctuation	Intervalle de fluctuation asymptotique Prise de décision	Intervalle de fluctuation asymptotique Prise de décision	Intervalle de fluctuation asymptotique Prise de décision
Estimation	Intervalle de confiance d'une proportion	Intervalle de confiance d'une proportion	Intervalle de confiance d'une proportion

1. Lois à densité

1.1. <u>Variable aléatoire continue</u>

En classe de première, la variable aléatoire, étant discrète, prend uniquement des valeurs isolées (souvent entières). Mais dans d'autres modèles, la variable aléatoire X peut prendre toute valeur de $\mathbb{R}$ ou d'un intervalle de $\mathbb{R}$. On dit alors que X est une **v.a. continue**.

Exemple : On dispose d'une population de 50 000 tailles (en cm) d'hommes adultes dont voici un résumé statistique et un histogramme des fréquences :



X est la v.a qui à tout individu pris au hasard dans la population associe sa taille.
X prend ses valeurs dans l'intervalle $[\alpha\ ;\ \beta]$ qu'on partage en classes.
L'ensemble des valeurs prises par X est l'intervalle $[\alpha\ ;\ \beta]$. X est une v.a. continue.

On construit l'histogramme des fréquences avec un pas de plus en plus petit.
Dans le cas limite, on obtient la courbe d'une **fonction continue positive *f*** qu'on appelle **densité de probabilité**.

La probabilité « totale » est alors étalée sur une infinité de valeurs possibles : $P(\alpha \le X \le \beta) = 1$, mais chaque valeur de X a une probabilité nulle de se produire :

$$\text{pour tout } x_i \in [\alpha\ ;\ \beta],\ \ P(X = x_i) = 0$$

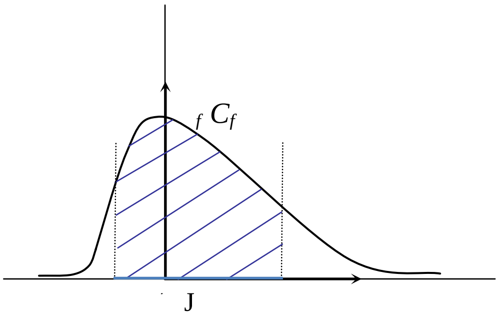
Quelle est la probabilité qu'un individu pris au hasard ait une taille comprise entre x_1 et x_2 ?
Dans un histogramme, les fréquences sont proportionnelles aux aires; par analogie : $P(x_1 \le X \le x_2)$ = aire du domaine compris entre la courbe, l'axe des abscisses et les droites d'équations $x = x_1$ et $x = x_2$:

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(t)\,\mathrm{d}t$$

f

1.2. Définition

Une v.a. continue X est définie par sa densité de probabilité : fonction f positive et définie sur un intervalle I de $\mathbb{R}$.



> La loi de probabilité de X vérifie :
> $P(X \in I) = 1$
> $P(X \in J)$ = aire du domaine $\{M(x,y) ; x \in J \text{ et } 0 \leq y \leq f(x)\}$,
> si J désigne un intervalle inclus dans I.

Remarques :

- pour tout réel x de I, $P(X = x) = 0$, donc dans les inégalités, on peut employer indifféremment $\leq$ et $<$ (ou $\geq$ et $>$)
- pour une v.a. continue, la densité f ne représente pas la probabilité de l'évènement $(X = x)$ car celle-ci est nulle. Il faut plutôt garder à l'esprit que : $P(x \leq X \leq x + \mathrm{d}x) = f(x) \times \mathrm{d}x$.

	Variable discrète	Variable continue
Espérance mathématique de X	$E(X) = \sum x_i p_i$	$E(X) = \int_I t \times f(t)\,\mathrm{d}t$
Variance et écart-type de X	$V(X) = E\left(\left(X - E(X)\right)^2\right)$ $\sigma(X) = \sqrt{V(X)}$	$V(X) = E\left(\left(X - E(X)\right)^2\right)$ $\sigma(X) = \sqrt{V(X)}$

Espérance mathématique et variance d'une v.a. continue vérifient les mêmes propriétés que celles énoncées en première pour une v.a. discrète.

1.3. Exemples

- Olivier vient tous les matins entre 7h et 7h 45 chez Karine prendre un café.

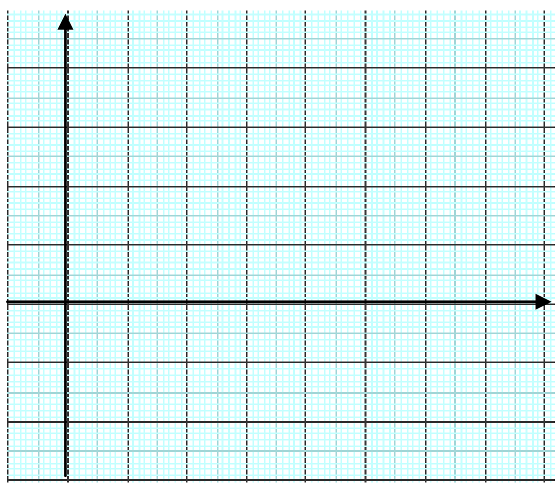
a. Sachant qu'Olivier ne vient jamais en dehors de la plage horaire indiquée et qu'il peut arriver à tout instant avec les mêmes chances, quelle densité peut-on attribuer à la variable aléatoire « heure d'arrivée d'Olivier » ?

b. Calculer la probabilité qu'Olivier sonne chez Karine après 7h 30 ? Entre 7h 20 et 7h 22 ? A 7h 30 très exactement ?

- Le temps d'accomplissement (en secondes) d'une certaine action mécanique est un phénomène aléatoire.

Soit T la variable aléatoire qui à chaque action mécanique, associe son temps d'accomplissement.
La loi physique suivie par celle-ci suggère que la fonction de densité de T est définie par :

$$f(t) = \begin{cases} 0 & si \quad t < 0 \\ \dfrac{\pi}{4}\cos\left(\dfrac{\pi}{4}t\right) & si \quad 0 \leq t \leq 2 \\ 0 & si \quad t > 2 \end{cases}$$

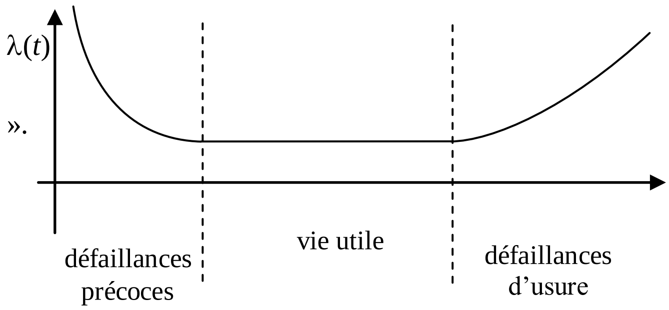


a. Donner une allure de la courbe représentative de g.

b. Vérifier que f est une densité de probabilité.

c. Quelle est la probabilité que le temps d'accomplissement dépasse 1,6 s ?

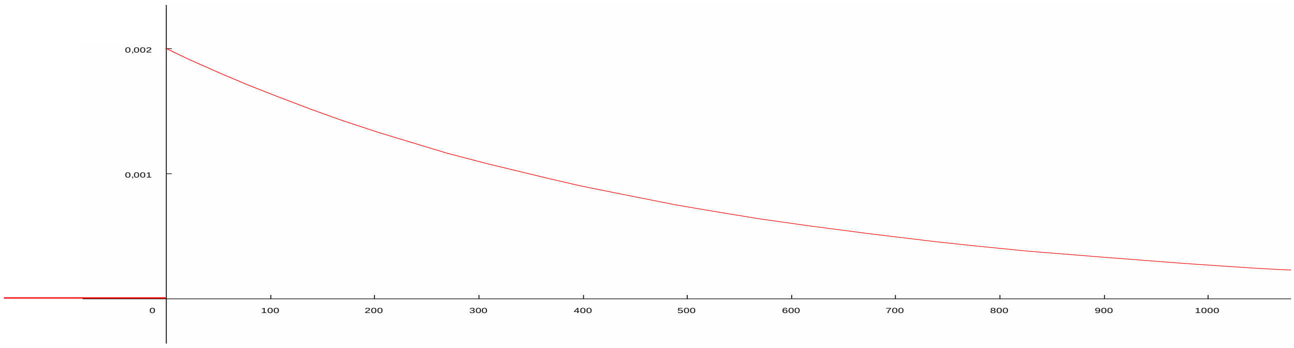
- *Evolution du taux de défaillance dans le temps*



Pour la plupart des matériels, la courbe représentative du taux d'avarie instantané a l'allure d'une « courbe en baignoire ».

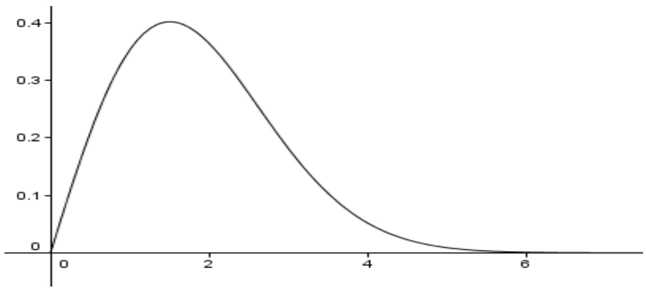
Lorsque le taux d'avarie est constant, le matériel est en période de vie utile : les défaillances sont aléatoires (mort sans vieillissement).
La variable aléatoire T, mesurant la durée de vie du matériel avant une panne, suit la loi exponentielle de paramètre λ positif.
Ce type de loi modélise les phénomènes « sans mémoire » comme les pannes de matériels électroniques, les tremblements de terre, les crues… On démontre que $P_{T \geq t}(T \geq t+h) = P(T \geq h)$.

- Soit la v.a. T mesurant la durée de bon fonctionnement (en jours) d'un équipement électronique fabriqué en grande série. T est une variable continue de densité de probabilité f définie sur]0 ; + ∞[par : $f(t) = 0{,}002\mathrm{e}^{-0{,}002t}$



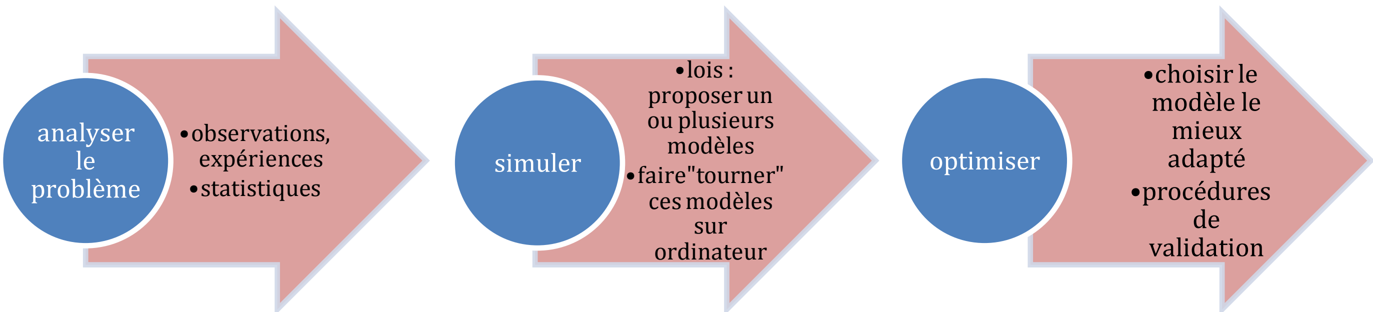
a. Vérifier que f est une densité de probabilité.
b. Quelle est la probabilité que la durée de bon fonctionnement d'un équipement soit inférieure à 1000 jours ?
c. Quelle est la probabilité que la durée de bon fonctionnement d'un équipement soit supérieure à 400 jours ?

- Soit X la v.a. qui, à une année prise au hasard, fait correspondre la hauteur maximale annuelle de l'Oise au pont d'Auvers (en mètres). X suit une loi de Rayleigh, de densité f définie sur]0 ; + ∞[par : $f(t) = 0{,}44\ t\ \mathrm{e}^{-0{,}22t^2}$



a. Vérifier que f est une densité de probabilité.
b. Quelle est la probabilité que la hauteur de cette rivière dépasse 3,80 m (ce qui correspond à une crue exceptionnelle) ?

1.4. En guise de conclusion



2. Variable centrée réduite

2.1 Définition

Une v.a. est dite **centrée** et **réduite** si son espérance est nulle et si son écart type est égal à 1.

Soit X une v.a. discrète d'espérance $\mathrm{E}(X) = m$, de variance $\mathrm{V}(X)$ et d'écart type $\sigma = \sqrt{\mathrm{V}(X)}$ (non nul).

On pose $Z = \dfrac{X-m}{\sigma}$. Alors on montre que Z est une v.a. centrée réduite. Ses paramètres (espérance et variance) étant égaux à 0 et à 1, ils ne dépendent plus de ceux de X.

Pour la démonstration, on rappelle que si X est une v.a., alors pour tous réels a et b : $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$

$\mathrm{V}(aX + b) = a^2\mathrm{V}(X)$

$\sigma(aX + b) = |a|\sigma(X)$

2.2 Application à la loi binomiale

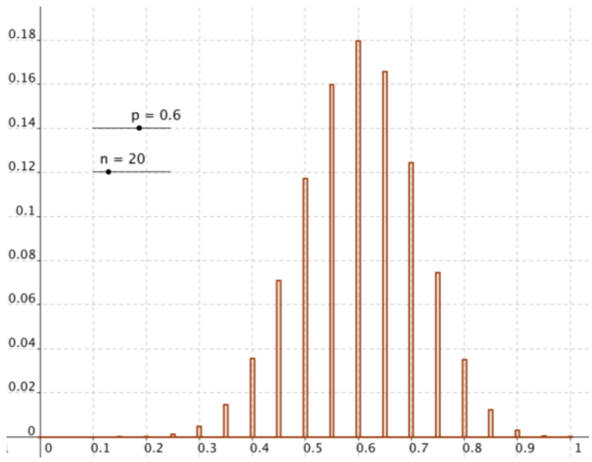
On considère X_n une v.a. qui suit **la loi binomiale $\mathcal{B}(n\,;p)$** de paramètres n et p.

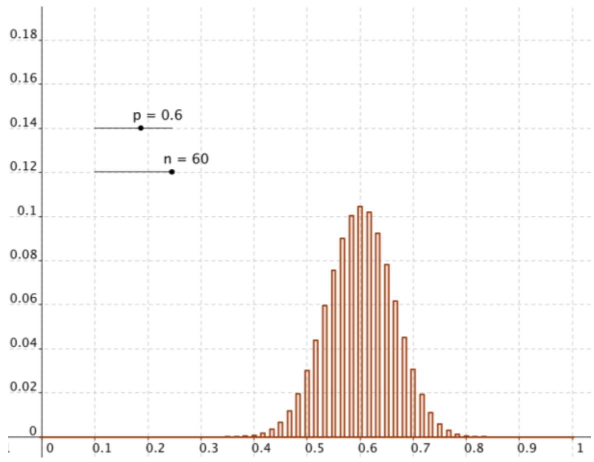
Alors : $\mathrm{E}(X_n) = np$; $\mathrm{V}(X_n) = np(1-p) = npq$; $\sigma(X_n) = \sqrt{npq}$.

- La v.a. $F_n = \dfrac{X_n}{n}$ correspond à la proportion de succès.

  $\mathrm{E}(F_n) = p$; elle ne dépend pas de n (mais dépend toujours de p).

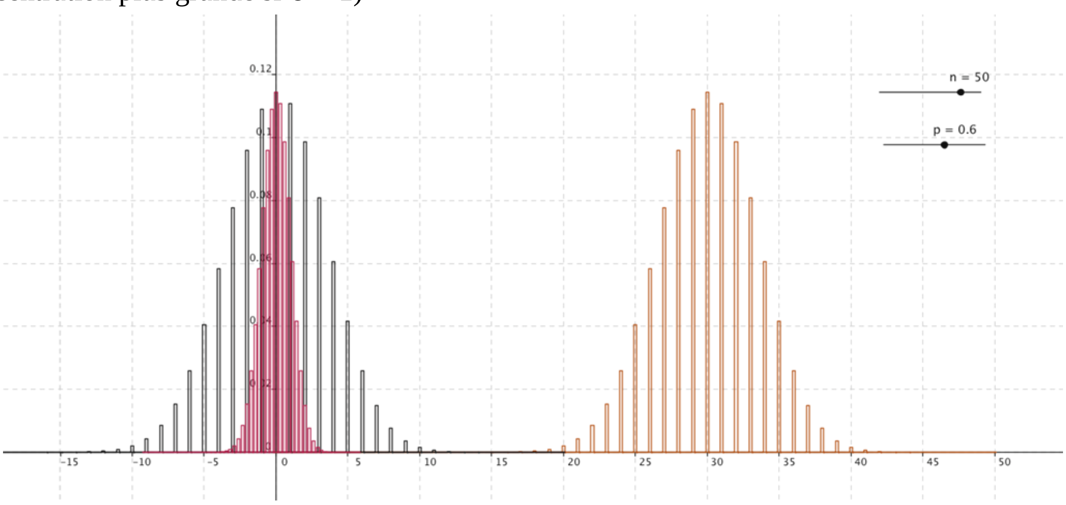
  $\mathrm{V}(F_n) = \dfrac{p(1-p)}{n}$; elle diminue si n augmente (autrement dit plus n augmente et plus les valeurs de F_n se resserrent autour de p comme l'illustrent les schémas ci-dessous).





- Pour centrer et réduire X_n, on pose $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$. Ses paramètres ne dépendent plus de n et de p.

X_n prend ses valeurs entre 0 et n, Z_n les prend entre $-\dfrac{m}{\sigma}$ et $\dfrac{n-m}{\sigma}$ avec les mêmes probabilités (mais avec une concentration plus grande si $\sigma > 1$)

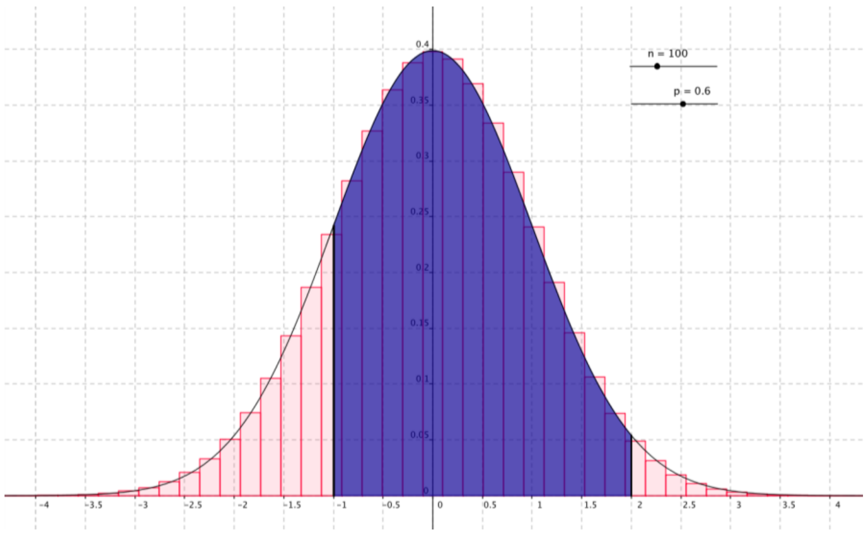


2.3 Théorème de Moivre-Laplace

On considère X_n une v.a. suivant la loi binomiale $\mathcal{B}(n\,;\,p)$ et $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}} = \dfrac{X_n - m}{\sigma}$, la v.a. centrée réduite associée.

On représente Z_n non plus par un diagramme en bâtons mais par un histogramme : à chaque valeur k prise par X_n on fait correspondre un rectangle dont l'aire est égale à $P(X_n = k)$ et dont la base est de longueur $\dfrac{1}{\sigma}$ (qui est l'écart entre deux valeurs prises par Z_n) et la hauteur : $\sigma \times P(X_n = k)$.

On va s'intéresser par exemple à $P(-1 \leq Z_n \leq 2)$. Cette probabilité est donc représentée par la réunion de tous les rectangles dont la base est le segment $\left[-1-\dfrac{1}{2\sigma}\ ;\ 2+\dfrac{1}{2\sigma}\right]$.



Il apparaît alors une courbe régulière et symétrique délimitant une aire qui est voisine de celle de la réunion des rectangles.

Abraham de Moivre a découvert que cette courbe est la courbe représentative de la fonction $x \mapsto \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

On obtient alors : $P(-1 \leq Z_n \leq 2) \approx \displaystyle\int_{-1}^{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$.

A p fixé, l'approximation sera meilleure si n est grand (car la largeur des rectangles diminue).

Théorème de Moivre-Laplace :

> Soit X_n une v.a. suivant la loi binomiale $\mathcal{B}(n\,;\,p)$ et $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$, la v.a. centrée réduite associée.
>
> Alors pour tous réels a et b tels que $a < b$, on a :
>
> $$\lim_{n\to+\infty} P(a \leq Z_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$$

3. Loi normale ou loi de Laplace - Gauss

3.1. Historique

Cette loi est un modèle essentiel dans l'étude de la variabilité et a un usage très important en statistique.
Elle est apparue au début du XIX$^{\text{ème}}$ et s'est introduite par deux voies :

Loi des erreurs (moindres carrés) en astronomie et géodésie avec Legendre puis Gauss.

Théorèmes limites avec Bernoulli puis Laplace ;
Approximation d'une binomiale par une normale avec Moivre puis Laplace.

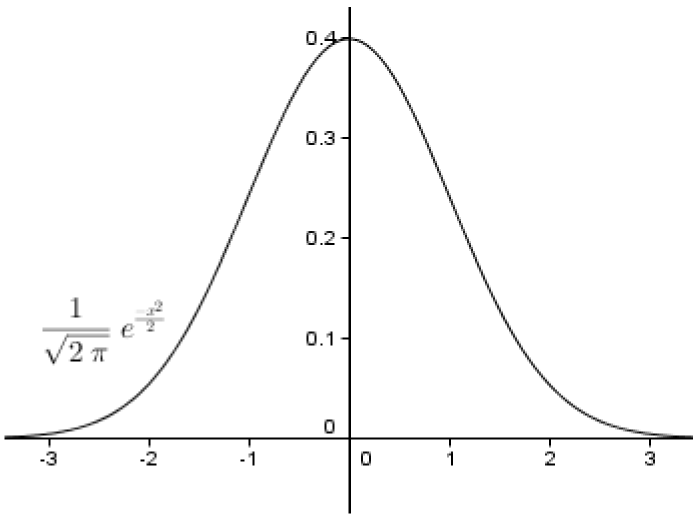
3.2. Loi normale centrée réduite

Une variable aléatoire X suit **la loi normale $\mathcal{N}$(0 ; 1)** de paramètres 0 et 1 lorsque sa densité de probabilité est définie sur $\mathbb{R}$ par : $\boxed{f(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}}$

$f'(x)=\frac{-x}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$ est du signe de $(-x)$,

d'où le tableau de variation ci-contre et la courbe « en cloche » ci-dessous :

x	$-\infty$		0		$+\infty$
$f'(x)$		$+$	0	$-$	
$f(x)$	0	$\nearrow$	$\frac{1}{\sqrt{2\pi}}$	$\searrow$	0



Propriétés de cette courbe :

- elle est symétrique par rapport à la droite d'équation $x = 0$.
- ses points d'inflexion ont pour abscisses $\pm$ 1.
- l'aire totale du domaine compris entre la courbe et l'axe des abscisses est égale à 1.

Espérance mathématique :

$$\mathrm{E}(X) = \lim_{x\to-\infty}\int_x^0 t\,f(t)dt + \lim_{y\to+\infty}\int_0^y t\,f(t)dt = 0$$

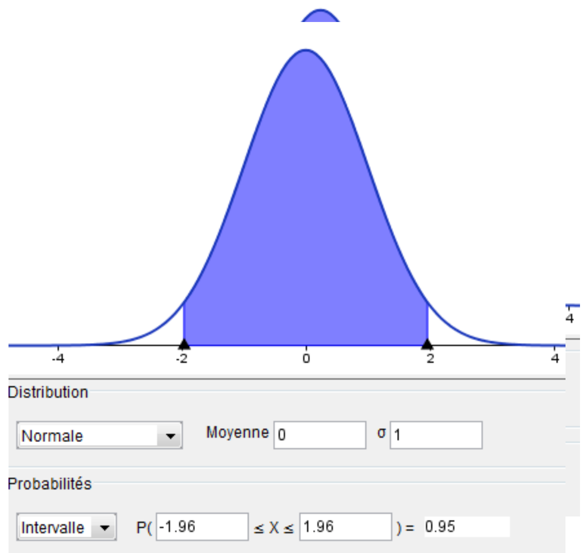
Variance : $\mathrm{V}(X) = \mathrm{E}\left(\left(X-\mathrm{E}(X)\right)^2\right) = 1$

Ecart type : $\sigma(X) = 1$

Propriété *(démontrée en S) :*

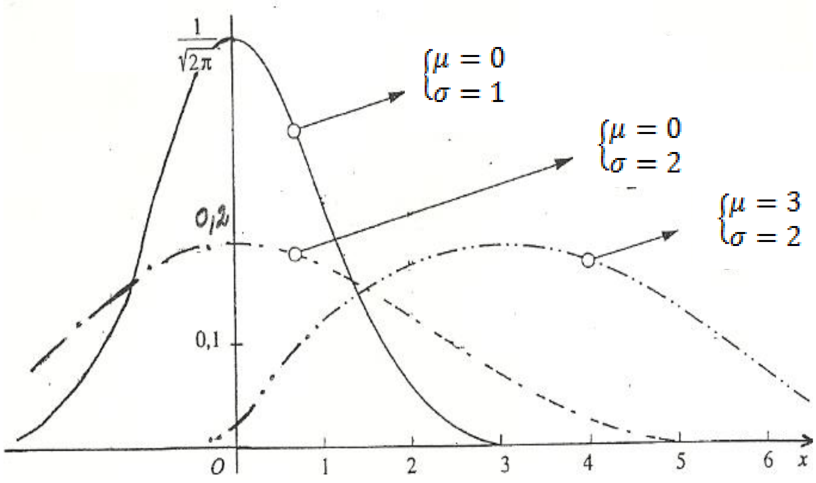
> Pour α dans $]0\,;1[$, il existe un unique réel positif u_α tel que $\mathrm{P}(-u_\alpha \le X \le u_\alpha) = 1-\alpha$ si X suit la loi $\mathcal{N}(0\,;1)$

Par exemple, on obtient $u_{0,05} \approx 1{,}96$ et $u_{0,01} \approx 2{,}58$, ce qui peut s'illustrer par le graphique ci-contre :



3.3. Loi normale $\mathcal{N}(\mu ; \sigma^2)$

Une variable aléatoire X suit la loi normale $\mathcal{N}(\mu ; \sigma^2)$ lorsque sa densité de probabilité est définie sur $\mathbb{R}$ par :



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

On obtient des courbes « en cloche » dont la forme dépend des paramètres μ et σ.

Propriété :

si X suit la loi normale $\mathcal{N}(\mu ; \sigma^2)$, alors la variable aléatoire Z définie par $Z = \dfrac{X-\mu}{\sigma}$ suit la loi normale centrée réduite $\mathcal{N}(0 ; 1)$.

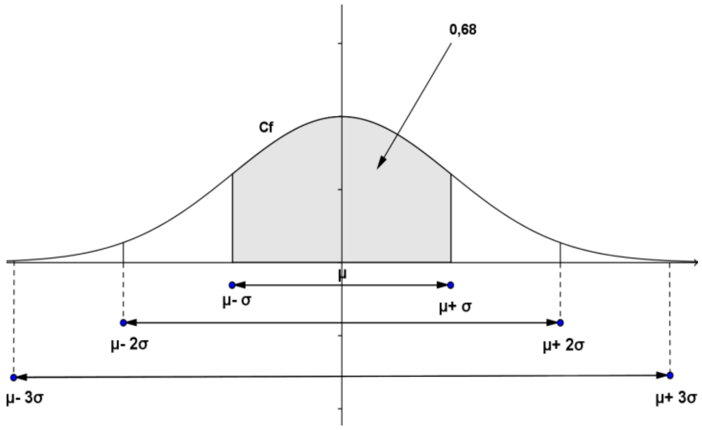
$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq Z \leq 1) \approx 0{,}68$

$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(-2 \leq Z \leq 2) \approx 0{,}95$

$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(-3 \leq Z \leq 3) \approx 0{,}997$.

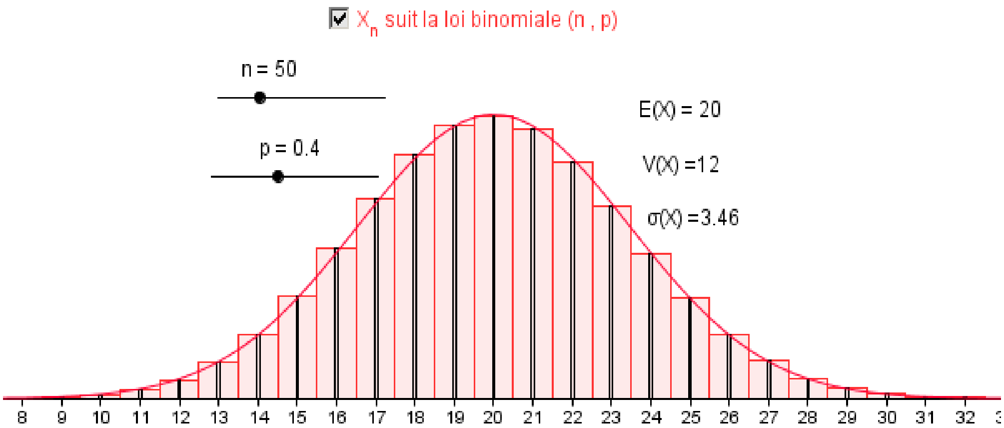
On peut justifier que :

$E(X) = \mu$ et $V(X) = \sigma^2$



3.4. Approximation d'une loi binomiale par une loi normale

Soit X une variable aléatoire qui suit la loi binomiale $\mathcal{B}(n ; p)$.
D'après une conséquence du théorème de Moivre-Laplace, si n est assez grand (en général, lorsque $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$),
alors X suit approximativement la loi normale $\mathcal{N}(\mu ; \sigma^2)$ avec $\mu = np$ et $\sigma = \sqrt{np(1-p)}$: l'espérance et l'écart type sont conservés.



La planche de Galton est une illustration classique de cette approximation.

Francis Galton éprouva le besoin d'expérimenter pour comprendre les propriétés de la loi normale.

4. Intervalle de fluctuation

4.1 Intervalle de fluctuation asymptotique

Soit X_n une v.a. suivant la loi binomiale $\mathcal{B}(n\ ;\ p)$ et α un réel dans $]0\ ;\ 1[$.
Un intervalle de fluctuation de X_n au seuil $1-\alpha$ est un intervalle $[a\ ;\ b]$ tel que $P(X_n \in [a\ ;\ b]) \geq 1-\alpha$.

- Dans le programme de seconde, on a défini un intervalle de fluctuation approché au seuil 0,95, de la variable fréquence $F_n = \dfrac{X_n}{n}$, par $\left[p-\dfrac{1}{\sqrt{n}}\ ;\ p+\dfrac{1}{\sqrt{n}}\right]$, valable sous certaines conditions de n et de p.
- En première, on a déterminé l'intervalle de fluctuation $\left[\dfrac{a}{n};\dfrac{b}{n}\right]$, calculé à partir de la loi binomiale, a étant le plus petit entier tel que $P(X_n \geq a) > 0{,}025$ et b le plus petit entier tel que $P(X_n \leq b) \geq 0{,}975$.
  Cet intervalle convient pour toutes les valeurs de n et de p.

- En terminale, le théorème de Moivre-Laplace va permettre de donner un intervalle de fluctuation plus facilement calculable qu'en première, sous réserve que n soit assez grand (mais valable pour toute valeur de p). Comme il est obtenu grâce à une convergence, on le qualifie d'intervalle de fluctuation asymptotique.

Notons Z une v.a. suivant la loi normale $\mathcal{N}(0\ ;\ 1)$. Il existe un unique réel u_α tel que $P(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$.

En posant $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$, d'après le théorème de Moivre-Laplace on a : $\lim\limits_{n\to+\infty} P(-u_\alpha \leq Z_n \leq u_\alpha) = P(-u_\alpha \leq Z \leq u_\alpha)$

$$\text{Or } P(-u_\alpha \leq Z_n \leq u_\alpha) = P\left(np - u_\alpha\sqrt{np(1-p)} \leq X_n \leq np + u_\alpha\sqrt{np(1-p)}\right)$$

$$= P\left(p - u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

$$\text{donc : } \lim_{n\to+\infty} P\left(p - u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1-\alpha.$$

Théorème : si X_n suit la loi binomiale $\mathcal{B}(n\ ;\ p)$, alors pour tout α dans $]0\ ;\ 1[$, on a :
$\lim\limits_{n\to+\infty} P\left(\dfrac{X_n}{n} \in I_n\right) = 1-\alpha$ avec $I_n = \left[p - u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.

$\left[p - u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + u_\alpha\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est appelé **intervalle de fluctuation asymptotique au seuil $1-\alpha$**.

Cet intervalle contient $F_n = \dfrac{X_n}{n}$ avec une probabilité d'autant plus proche de $1-\alpha$ que n est grand.

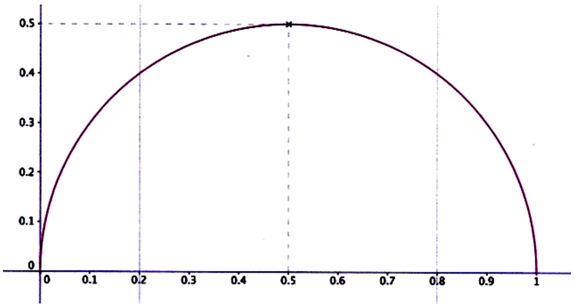
On convient d'utiliser cette approximation si $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$.

4.2 Cas particulier

Si $\alpha = 5\%$, on a $u_\alpha = 1{,}96$.

L'intervalle de fluctuation asymptotique au seuil de 95% est donc : $\left[p - 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$

Considérons la représentation graphique de la fonction f définie par $f(x) = \sqrt{x(1-x)}$



Pour tout x de $[0 ; 1]$, $\sqrt{x(1-x)} \leq 0,5$ et $u_\alpha = 1,96 \leq 2$

On a alors : $u_\alpha \sqrt{p(1-p)} \leq 1$, et l'on obtient l'intervalle de fluctuation au seuil de 95 % : $\left[p-\frac{1}{\sqrt{n}}\ ;\ p+\frac{1}{\sqrt{n}}\right]$.

On retrouve l'intervalle de fluctuation vu en seconde. Cette majoration est valable, d'un point de vue statistique, pour $0,2 \leq p \leq 0,8$ sinon on augmenterait trop l'amplitude de l'intervalle de fluctuation.

4.3 Exemple de prise de décision : la parité, c'est quoi ?

Deux entreprises A et B recrutent dans un bassin d'emploi où il y a autant de femmes que d'hommes, avec la contrainte du respect de la parité. Dans l'entreprise A, il y a 100 employés dont 43 femmes; dans l'entreprise B, il y a 2500 employés dont 1150 femmes (soit 46%).
Ces entreprises respectent-elles la parité ?

La parité signifie que l'identité sexuelle n'intervient pas au niveau du recrutement, c'est-à-dire qu'au niveau du caractère homme ou femme, les résultats observés pourraient être obtenus par choix au hasard des individus dans la population.
Les entreprises peuvent donc être assimilées à des échantillons de taille n prélevés dans une population où la fréquence étudiée p est égale à 0,5.
Les intervalles de fluctuation au niveau 0,95 sont :

$$\left[p-1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p+1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \text{, soit } \left[0,5-\frac{0,98}{\sqrt{n}}\ ;\ 0,5+\frac{0,98}{\sqrt{n}}\right]$$

L'entreprise A est un échantillon de taille 100 dont l'intervalle de fluctuation est $[0,402 ; 0,598]$.

L'entreprise B est un échantillon de taille 2500 dont l'intervalle de fluctuation est $[0,48 ; 0,52]$.

Conclusion : **la valeur 43% est dans l'intervalle de fluctuation pour un échantillon de taille 100 alors que 46% n'est pas dans l'intervalle de fluctuation pour un échantillon de taille 2500 !**

Autrement dit : **pour l'entreprise B**, la proportion de 46% s'observe dans moins de 5% des échantillons de taille 2500 prélevés au hasard dans une population où il y a autant d'hommes que de femmes. On peut alors **rejeter l'hypothèse** que cette entreprise respecte la parité. Mais en prenant cette décision, on prend le risque de se tromper, ce risque étant égal à 5%.

Pour l'entreprise A, on considère que le résultat observé est compatible avec le modèle (l'écart entre f et p est probable, au sens où il est contenu dans la fourchette que le hasard produirait avec 95% des échantillons envisageables). De ce fait, on peut **accepter l'hypothèse** que cette entreprise respecte la parité. Mais là aussi en prenant cette décision, on a un risque de se tromper, risque inconnu dans ce cas.

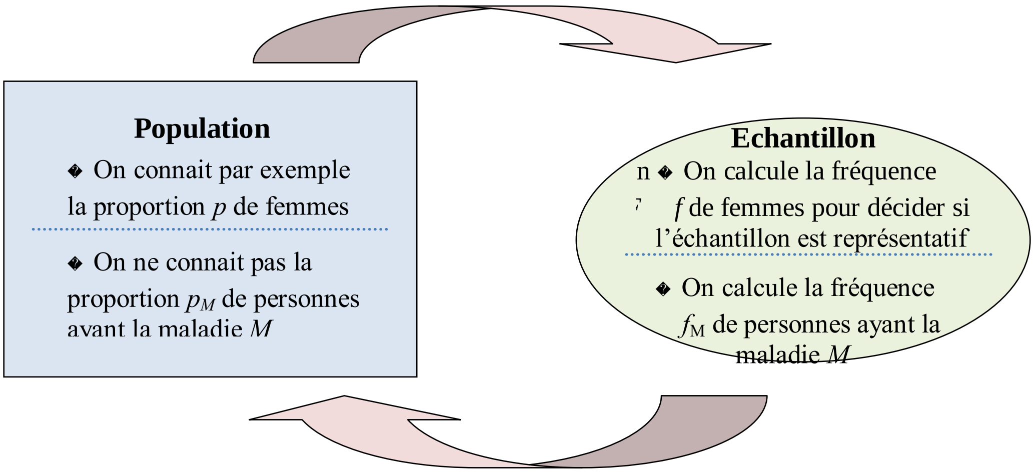
5. Intervalle de confiance

5.1 Introduction

Il est souvent difficile de pouvoir recueillir des données sur la population toute entière. Le plus souvent, on travaille à partir d'un échantillon prélevé de manière aléatoire pour estimer le paramètre inconnu de la population (par exemple une proportion p). Or cette estimation va varier d'un échantillon à l'autre du fait de la fluctuation d'échantillonnage. Il est donc nécessaire d'apprécier l'incertitude en fournissant une estimation par intervalle, appelé intervalle de confiance de p au niveau qu'on s'est fixé.

s **ECHANTILLONNAGE** : à partir des informations connues dans la population, on détermine si l'échantillon prélevé est représentatif ou non.

INTERVALLE DE FLUCTUATION



s **ESTIMATION** : à partir des données de l'échantillon, on estime les paramètres inconnus de la population.

INTERVALLE DE CONFIANCE

5.2 Intervalle de confiance

Lorsque $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$, on a déterminé l'intervalle de fluctuation de $F_n = \dfrac{X_n}{n}$ au niveau 0,95 par la formule $\left[p - \dfrac{1}{\sqrt{n}}\ ;\ p + \dfrac{1}{\sqrt{n}}\right]$.

Si p est inconnu, on peut approximer p par la fréquence f obtenue avec l'échantillon mais cette estimation ponctuelle n'est pas satisfaisante car elle ne tient pas compte de n. Or l'estimation est forcément meilleure sur un échantillon de grande taille.

On sait que pour n suffisamment grand on a : $\quad \mathrm{P}\left(\left[p - \dfrac{1}{\sqrt{n}} \leq \dfrac{X_n}{n} \leq p + \dfrac{1}{\sqrt{n}}\right]\right) \geq 0{,}95$

Comme $p - \dfrac{1}{\sqrt{n}} \leq F_n \leq p + \dfrac{1}{\sqrt{n}} \Leftrightarrow F_n - \dfrac{1}{\sqrt{n}} \leq p \leq F_n + \dfrac{1}{\sqrt{n}}$, alors : $\mathrm{P}\left(\left[F_n - \dfrac{1}{\sqrt{n}} \leq p \leq F_n + \dfrac{1}{\sqrt{n}}\right]\right) \geq 0{,}95$

L'intervalle aléatoire $\left[F_n - \dfrac{1}{\sqrt{n}}\ ;\ F_n + \dfrac{1}{\sqrt{n}}\right]$ a donc une probabilité au moins égale à 0,95 de contenir p. Il est appelé **intervalle de confiance** (théorique) **de p au niveau 95%.**

Pour une observation donnée d'un échantillon, on obtient une réalisation de cet intervalle avec la fourchette $\left[f - \dfrac{1}{\sqrt{n}}\ ;\ f + \dfrac{1}{\sqrt{n}}\right]$ qui représente un **intervalle de confiance de p au niveau 95% d'après l'échantillon observé.**

On a 5% de risque que cette construction aboutisse à un intervalle ne contenant pas p. Autrement dit, si l'on prélève un très grand nombre d'échantillons, environ 95% des intervalles de confiance obtenus contiennent p.

5.3 Remarque

Si on part de l'intervalle de fluctuation au seuil de 5% vu en terminale, pour n suffisamment grand on a :

$$\mathrm{P}\left(\left[p-1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p+1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) \geq 0{,}95,$$

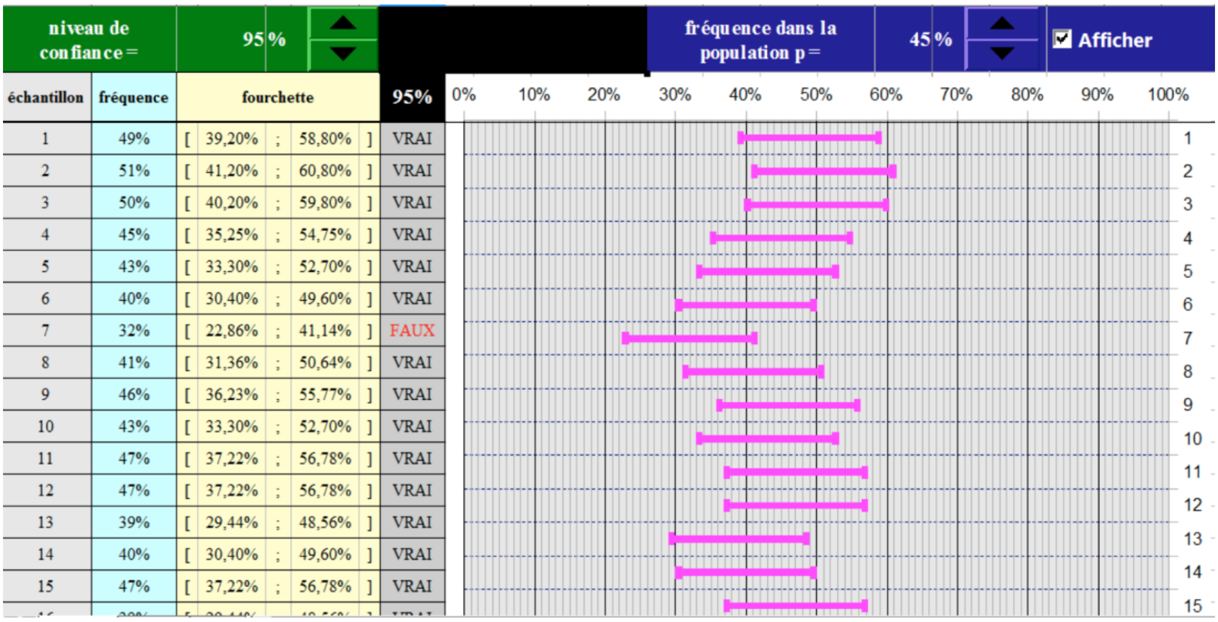
ce qui est équivalent à : $\mathrm{P}\left(\left[F_n-1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq p \leq F_n+1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) \geq 0{,}95$

On obtient l'intervalle aléatoire $\left[F_n-1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ F_n+1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ qui dépend du paramètre p inconnu.

On remplace p par son estimation ponctuelle f (car on peut démontrer que dans l'écart type, f est un bon estimateur de p) et on obtient alors **l'intervalle de confiance de p au niveau 0,95 d'après l'échantillon observé :** $\left[f-1{,}96\frac{\sqrt{f(1-f)}}{\sqrt{n}}\ ;\ f+1{,}96\frac{\sqrt{f(1-f)}}{\sqrt{n}}\right]$

5.4 Simulation

Sur tableur, pour une proportion qu'on se fixe, on peut simuler un grand nombre d'intervalles de confiance au niveau 0,95 et observer que tous ne contiennent pas nécessairement p.



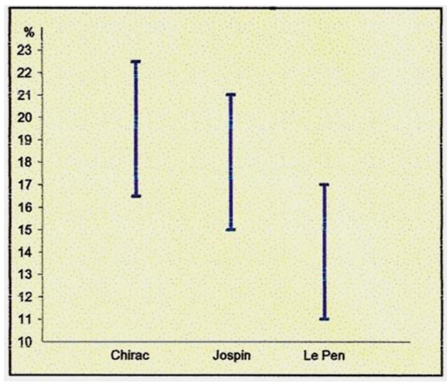
échantillon	fréquence	fourchette					95%
1	49%	[	39,20%	;	58,80%	]	VRAI
2	51%	[	41,20%	;	60,80%	]	VRAI
3	50%	[	40,20%	;	59,80%	]	VRAI
4	45%	[	35,25%	;	54,75%	]	VRAI
5	43%	[	33,30%	;	52,70%	]	VRAI
6	40%	[	30,40%	;	49,60%	]	VRAI
7	32%	[	22,86%	;	41,14%	]	FAUX
8	41%	[	31,36%	;	50,64%	]	VRAI
9	46%	[	36,23%	;	55,77%	]	VRAI
10	43%	[	33,30%	;	52,70%	]	VRAI
11	47%	[	37,22%	;	56,78%	]	VRAI
12	47%	[	37,22%	;	56,78%	]	VRAI
13	39%	[	29,44%	;	48,56%	]	VRAI
14	40%	[	30,40%	;	49,60%	]	VRAI
15	47%	[	37,22%	;	56,78%	]	VRAI

5.5 Exemple : marge d'erreur de 3% du sondage par quotas

La plupart des sondages en France sont effectués auprès de 1000 personnes. On obtient alors l'intervalle de confiance : $[f - 0{,}03\ ;\ f + 0{,}03]$.

Le tableau ci-dessous présente les résultats du dernier sondage effectué les 17/18 Avril 2002 et les résultats du premier tour de l'élection présidentielle du 21 Avril 2002 :

	Sondage	Election
Chirac	19,5 %	19,7 %
Jospin	18,0 %	16,1 %
Le Pen	14,0 %	16,9 %



Si les résultats du dernier sondage avaient été annoncés avec une marge d'erreur de 3%, voici le graphique qui aurait été donné :

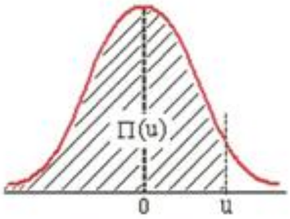
Il apparaît que toutes les configurations de l'ordre des trois premiers candidats étaient possibles.

On peut se demander si les sondeurs et les journalistes annonçaient que leur marge d'erreur est de 3%, est-ce que nous serions inondés de sondages comme c'est le cas aujourd'hui ?

Table de Loi Normale

Fonction de répartition Π de la loi normale centrée réduite.

Probabilité de trouver une valeur inférieure à u .



$$\Pi(-u) = 1 - \Pi(u)$$

u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Annexe :