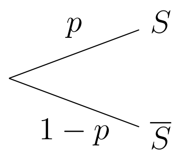
Chapitre :
Loi normale et applications

I. Rappels

1. Épreuve de Bernoulli

Définition On appelle **épreuve de Bernoulli** une expérience aléatoire présentant deux issues, l'une, notée S, appelée succès et l'autre, notée $\overline{S}$, appelée échec. On note p la probabilité du succès, puis parfois $q = 1 - p$ la probabilité de l'échec. La variable aléatoire X qui prend la valeur 1 en cas de succès et 0 en cas d'échec est appelée variable aléatoire de Bernoulli. La loi de probabilité, appelée loi de Bernoulli de paramètre p est alors donnée par :

x_i	0	1
$P(X = x_i)$	$1 - p$	p



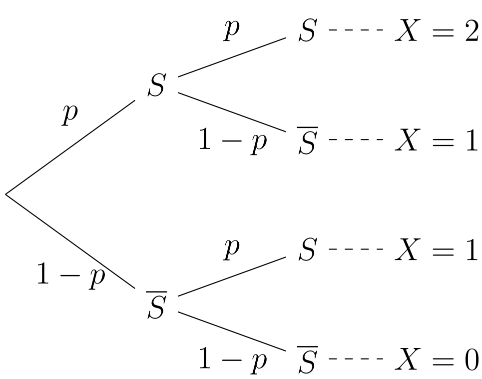
Propriété Soit X une variable aléatoire suivant une loi de Bernoulli de paramètre p. Alors :

$$E(X) = p \qquad V(X) = p(1-p) \qquad \sigma(X) = \sqrt{p(1-p)}$$

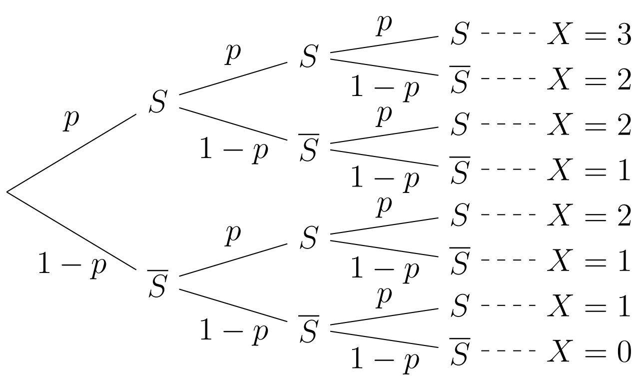
2. Schéma de Bernoulli

Définition L'expérience aléatoire consistant à répéter n fois de manière **indépendante** une épreuve de Bernoulli de paramètre p s'appelle un **schéma de Bernoulli de paramètres** n **et** p.
On considère la variable aléatoire X égale au nombre de succès obtenus au cours des n épreuves. On appelle alors **loi binomiale de paramètres** n **et** p la loi de probabilité de X. On la note $\mathcal{B}(n,p)$.

Exemple Le nombre X de succès est toujours un nombre compris entre 0 et n.



$n = 2$



$n = 3$

Propriété Soit X une variable aléatoire qui suit la loi binomiale $\mathcal{B}(n,p)$. Alors :

$$E(X) = np \qquad \text{et} \qquad \sigma(X) = \sqrt{np(1-p)}$$

3. Calcul des probabilités

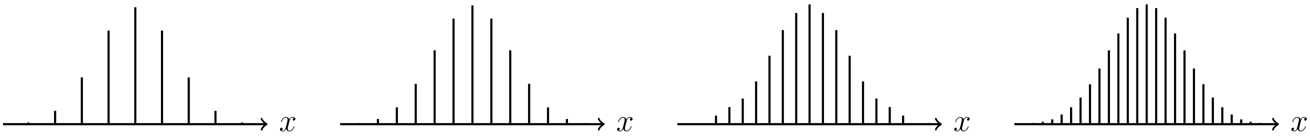
Soit X une variable aléatoire qui suit la loi binomiale $\mathcal{B}(n,p)$ de paramètres n et p. Les calculs de probabilités associés à la loi binomiale peuvent se faire avec la calculatrice. Il y a deux types de probabilités qui peuvent être obtenus (voir les pages 247 et 249 pour plus de détails) :

- $\mathbb{P}(X = k)$:
  en Casio : BinominalPD(k,n,p)
  en TI : binomFdp(n,p,k)
- $\mathbb{P}(X \leqslant k)$:
  en Casio : BinominalCD(k,n,p)
  en TI : binomFRép(n,p,k)

▶ **Exercices :** Fiche d'exercices.

II. Loi normale

Voici une illustration de l'apparence des diagrammes en bâton de la loi binomiale lorsque n augmente :

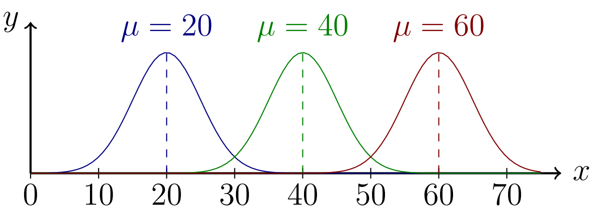


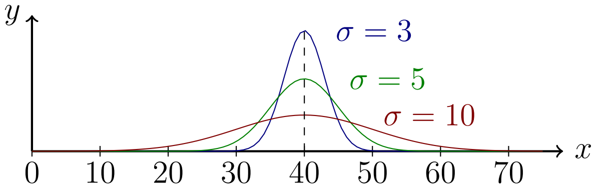
$n = 8$ $n = 16$ $n = 32$ $n = 64$

Propriété **(et définition)** Le diagramme en bâtons d'une loi binomiale $\mathcal{B}(n,p)$ de paramètres n et p, lorsque n est très grand et que p n'est pas « voisin » de 0 ou de 1, peut être approché par une courbe « en cloche ». Cette courbe est celle d'une fonction qui détermine une nouvelle loi de probabilité, appelée **loi normale** qui possède deux paramètres :

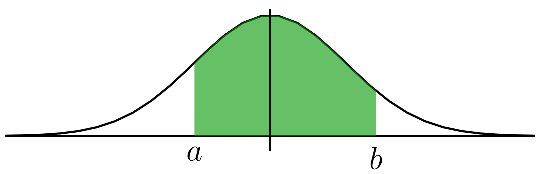
- son espérance μ, qui correspond à celle de la loi binomiale qu'elle approche, soit np ;
- son écart-type σ, qui correspond aussi à celui de la loi binomiale, soit $\sqrt{np(1-p)}$.

Voici une illustration des conséquences de ces paramètres sur les courbes :





Propriété Soit X une variable aléatoire qui suit la loi normale et a et b deux réels tels que $a \leqslant b$. Alors la probabilité $\mathbb{P}(a \leqslant X \leqslant b)$ est l'aire du domaine délimité par la courbe de la loi normale, l'axe des abscisses et les droites d'équations respectives $x = a$ et $x = b$.



On peut aussi chercher à calculer des probabilités comme $\mathbb{P}(X \leqslant a)$ ou $\mathbb{P}(X \geqslant b)$.
Voir le livre page 180 pour l'utilisation de la calculatrice.

La syntaxe est la suivante :

- En TI : $\mathbb{P}(a \leqslant X \leqslant b) = \text{normalFRép}(a,b,\mu,\sigma)$
- En Casio : $\mathbb{P}(a \leqslant X \leqslant b) = \text{NormCD}(a,b,\sigma,\mu)$

Remarque

- La courbe est symétrique autour de l'espérance μ et l'aire totale sous la courbe vaut 1.
  Par conséquent, on a toujours $\mathbb{P}(X \geqslant \mu) = 0{,}5$.
- Une variable aléatoire X qui suit une loi normale peut prendre n'importe quelle valeur de l'intervalle $]-\infty;+\infty[$.
  Par suite, quelque soit $x \in \mathbb{R}$, on a $\mathbb{P}(X = x) = 0$.
  Et ainsi par exemple $\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(a < X < b)$: cela n'a pas de conséquence sur les probabilités d'inclure ou non les bornes.
- La courbe s'approche très rapidement de l'axe des abscisses lorsque x s'éloigne de μ.
  Par conséquent, on peut par exemple donner une bonne approximation de $\mathbb{P}(X < b)$ en calculant $\mathbb{P}(-100000 < X < b)$ (remplacer -100000 par n'importe quel « grand nombre négatif »).

▶ **Exercices :** 1,3,4,7,8,9p183

▶ **Exercices :** 20,23,24,26p184 et 34p185 (calcul de $\mathbb{P}(X \geqslant b)$ sans tricher)

Propriété Soit X une variable aléatoire qui suit une loi normale de moyenne μ et d'écart-type σ.
Alors :

$$\mathbb{P}(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) \simeq 0{,}95$$

▶ **Exercices :** 11,13,16p183

▶ **Exercices :** (DM) 41p186 et 46p187 (avec loi binomiale) et 61p188 (2σ)

III. Echantillonnage

1. Intervalles de fluctuation

⊛ **Activité :** 1p202 (intervalle de fluctuation des fréquences)

Soit X une variable aléatoire suivant une loi binomiale de paramètres n et p. On définit F la variable aléatoire représentant la **fréquence de succès**, autrement dit $F = \dfrac{X}{n}$.
Dans certains cas on sait que l'on peut approcher une loi binomiale par une loi normale. Ainsi, on peut donner un intervalle contenant la valeur de F dans au moins 95% des cas :

Définition Soit p la proportion (connue) d'un caractère dans une population.
L'intervalle $\left[p - \dfrac{1}{\sqrt{n}}; p + \dfrac{1}{\sqrt{n}}\right]$ est un **intervalle de fluctuation** à au moins 95% de la fréquence de ce caractère dans un échantillon de taille n issu de la population.
Les conditions pour pouvoir utiliser cet intervalle sont : $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.

Exemple Supposons $p = 0{,}3$ et $n = 400$. Alors $n = 400 \geqslant 30$, $np = 120 \geqslant 5$ et $n(1-p) = 280 \geqslant 5$.
Par conséquent on peut donner pour intervalle de fluctuation à au moins 95% l'intervalle suivant :

$$\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right] = \left[0{,}3 - \frac{1}{\sqrt{400}}; 0{,}3 + \frac{1}{\sqrt{400}}\right] = [0{,}25; 0{,}35]$$

Cela signifie que pour 95% au moins des échantillons de taille 400, la fréquence obtenue appartient à cet intervalle.

Rappel Cet intervalle de fluctuation $\left[p - \dfrac{1}{\sqrt{n}}; p + \dfrac{1}{\sqrt{n}}\right]$ a déjà été vu en seconde, mais les conditions données étaient différentes : $n \geqslant 25$ et $0{,}2 \leqslant p \leqslant 0{,}8$.
En première, on voit un autre intervalle : $\left[\dfrac{a}{n}; \dfrac{b}{n}\right]$, avec ($X$ suit une loi $\mathcal{B}(n,p)$) :

- a le plus petit entier tel que $\mathbb{P}(X \leqslant a) > 0{,}025$;
- b le plus petit entier entier tel que $\mathbb{P}(X \leqslant b) \geqslant 0{,}975$.

Il est plus précis, mais moins facile à déterminer.

▶ **Exercices :** 1,3,6,11,13,14p206

▶ **Exercices :** 24,25p208 puis 27p208 (algorithme) et 31p209 (recherche de n)

Méthode (Prise de décision) On fait l'**hypothèse** que, dans une population donnée, la proportion d'un caractère donné est p. On **observe** une fréquence f dans un échantillon de taille n de la population. Soit I l'intervalle de fluctuation à au moins 95% pour un échantillon de taille n. Alors la règle de décision est la suivante (à savoir énoncer) :

- Si $f \in I$, on **accepte l'hypothèse au seuil de confiance 95%** (car l'hypothèse n'est pas remise en cause) ;
- Sinon on **rejette l'hypothèse**.

▶ **Exercices :** 15p206, 17p207

▶ **Exercices :** 35,36,37p209

▶ **Exercice :** (DM) 39p210 (intervalle de première)

2. Intervalles de confiance

⊛ **Activité :** 2p202 (salle informatique) estimation d'une proportion
Dans le cas de l'intervalle de fluctuation, on connaît la proportion p dans la population totale.
Ici on s'intéresse au problème « inverse » : on ne connaît pas la valeur de p et on veut donner un intervalle qui peut la contenir avec un seuil de confiance 95%, à partir d'une fréquence f observée.

Définition Soit f la fréquence observée d'un caractère dans un échantillon de taille n extrait dans une population dans laquelle la proportion du caractère, inconnu, est p.
Alors l'intervalle $\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}\right]$ est un **intervalle de confiance de la proportion p au seuil de confiance 95**%.

On peut utiliser cet intervalle dès que n est suffisamment grand.

Propriété Au moins 95% des intervalles de la forme $\left[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}\right]$ contient la proportion p.

▶ **Exercices :** 18,19,21,22,23p207

▶ **Exercices :** 43,44,46p210 et 50p211 (comparaison)