LOIS DE PROBABILITE USUELLES

Les concepts et méthodes probabilistes ont pour but l'étude des expériences aléatoires. Ainsi, étant donné une expérience aléatoire concrète, on cherche à déterminer un modèle probabiliste qui va permettre de décrire et d'analyser le mieux possible cette expérience. Dans cette perspective, on a introduit d'une façon générale les notions de probabilité, de variable aléatoire et de distribution de probabilité.

En pratique, il y a un nombre relativement restreint d'expériences aléatoires que l'on rencontre très souvent et qui possèdent des propriétés communes, ce qui permet de réduire d'autant le nombre des modèles probabilistes nécessaires pour décrire ces expériences. On va maintenant étudier les expériences aléatoires et les modèles probabilistes correspondants que l'on rencontre le plus fréquemment en gestion, en économie, en ingénierie, en sciences sociales, etc., et particulièrement les variables aléatoires et les distributions de probabilité associées à ces modèles. La compréhension des caractéristiques générales de ces distributions nous servira par la suite à mieux comprendre les raisonnements probabilistes qui appuient des décisions en situations d'incertitude.

Dans la présentation de ces lois de probabilité dites «usuelles», on va utiliser l'expression «un processus aléatoire» pour désigner «une série d'expériences aléatoires élémentaires» en se contentant de la notion plus ou moins intuitive que l'on a d'un «processus aléatoire».

1. Loi binômiale

Le processus de Bernoulli

Un grand nombre d'expériences aléatoires que l'on rencontre en pratique apparaissent comme étant constituées d'une suite d'épreuves aléatoires élémentaires dont chacune n'a que deux résultats possibles, les deux mêmes pour chaque épreuve. Par exemple, on choisit au hasard des pièces dans un lot : chaque pièce choisie est soit bonne soit défectueuse. De même, si l'on choisit n individus au hasard dans une population et si l'on s'intéresse à leurs habitudes face à la cigarette, alors chaque individu choisi est soit fumeur, soit non-fumeur. Pour avoir une terminologie commune à toutes ces expériences, on appelle «succès» l'un des résultats et «échec» l'autre résultat.

Une épreuve de Bernoulli : c'est une expérience aléatoire élémentaire dont l'ensemble fondamental se compose de seulement deux événements simples (succès, échec).

Un processus de Bernoulli : c'est une expérience aléatoire constituée d'une *suite* d'épreuves de Bernoulli indépendantes, chaque épreuve ne pouvant conduire qu'aux deux mêmes résultats possibles (succès, échec), chacun de ces résultats ayant la même probabilité d'une épreuve à l'autre. Les probabilités de succès et d'échec sont désignées par p et q = 1-p. Ainsi, le processus de Bernoulli est un processus à un seul paramètre p que l'on appelle l'intensité du processus.

À chaque épreuve élémentaire on peut associer la variable aléatoire qui fait correspondre la valeur 1 au résultat «succès» et la valeur 0 au résultat «échec». On définit ainsi la **variable aléatoire de Bernoulli** dont la distribution de probabilité est

x_i	1	0
$f(x_i)$	p	1 - p

Remarque LO 1. ***L'espérance mathématique*** de la v.a. de Bernoulli est donnée par 1*p + 0*q = **p**. ***La variance*** de la v.a. de Bernoulli est donnée par **pq** d'après le raisonnement suivant :

$$(1-p)^2p + (0-p)^2q = (1-2p+p^2)p + p^2(1-p) = p - 2p^2 + p^3 + p^2 - p^3 = p - p^2 = p(1-p) = pq$$

Exemple : D'après un sondage dans la région de Québec, la proportion des répondants choisis au hasard qui achètent une certaine marque de jus d'orange est 0,12. L'espérance mathématique de la v.a. Bernoulli «un résident de la région achète cette marque de jus d'orange» est p = 0,12. La variance de cette v.a. est pq = 0,12*0,88 = 0,106 et l'écart type = 0,324.

Variable aléatoire binômiale

Supposons que l'on effectue n épreuves de Bernoulli et que l'on considère **la v.a. X donnant le nombre x de succès observés.** Un résultat (élémentaire) particulier d'une telle expérience aléatoire peut être écrit sous la forme (SÉSSÉ... S), ou S et É désignent «succès» et «échec» respectivement et où l'on a x fois S et (n-x) fois É. À cause de l'indépendance des épreuves successives, la probabilité d'observer ce résultat particulier est donnée par $p^x(1-p)^{n-x}$. Même si l'ordre dans lequel on observe les S et É change, la probabilité reste la même que x (le nombre de succès) soit inchangé. Donc tous les résultats (élémentaires) comprenant x succès et (n-x) échec ont la même probabilité d'être observés. Comme le nombre de résultats conduisant à une valeur X = x pour n épreuves est donné par $\binom{n}{x}$, le nombre de combinaisons de n objets pris x à la fois, la probabilité d'observer x succès est donnée par $\binom{n}{x} p^x (1-p)^{n-x}$. La v.a. binômiale est définie comme suit :

V.a binômiale. Une v.a. X qui peut prendre les valeurs entières x = 0, 1, ..., n, pour n entier positif, avec les probabilités

$$P\ (X = x) = \binom{n}{x}\ p^x\ q^{n-x}$$

où $0 < p < 1$, $q = 1 – p$ s'appelle v.a. binômiale de paramètres n et p.

Notation : on désigne par Bi (n, p) la famille de toutes les v.a. binômiales de paramètres n et p. On écrira que X est en Bi (n, p) ou X ε Bi(n,p).

En effet, il est possible de décrire le processus de Bernoulli par le modèle physique suivant : une urne contient des boules dont une proportion p sont blanches, les autres en proportion q = 1 - p étant noires; on tire n boules au hasard **avec replacement** dans cette urne et on définit la v.a. X donnant le nombre x de boules blanches obtenues au cours de ces n tirages successifs avec replacement. Cette v.a. X est une v.a. binômiale. Les tirages se faisant avec replacement, le processus est constitué d'une suite d'épreuves de Bernoulli **indépendantes.**

Exemple LO 1. Pour contrôler la qualité de la production, on tire au hasard quatre pièces d'un lot en replaçant la pièce tirée chaque fois. La probabilité de tirer une bonne pièce est 99 %. Quelle est la distribution de la variable aléatoire X = le nombre de bonnes pièces observées ?

Solution : On considère une bonne pièce comme un succès. Alors, la v.a. X est une v.a. binômiale dans la famille Bi(4, 0,99). Ainsi on peut déterminer les probabilités :

$$P(X=0) = (4!/0!4!)(0,99)^0(0,01)^4 = 0,00000001;$$

$$P(X=1) = (4!/1!3!)(0,99)^1(0,01)^3 = 0,00000396;$$

$$P(X=2) = (4!/2!2!)(0,99)^2(0,01)^2 = 0,00058806;$$

$$P(X=3) = (4!/3!1!)(0,99)^3(0,01)^1 = 0,03881196;$$

$$P(X=4) = (4!/4!0!)(0,99)^4(0,01)^0 = 0,96059601.$$

Remarque LO 2. La distribution de la v.a. Y = le nombre de pièces défectueuses observées est complémentaire à la distribution de X de l'exemple LO 1. Y est dans la famille Bi(4, 0,01). On peut vérifier facilement que

P(X=0) = P(Y=4); P(X=1) = P(Y=3); P(X=2) = P(Y=2); P(X=3) = P(Y=1); P(X=4) = P(Y=0).

Dans le cas d'un processus d'échantillonnage, les tirages se font **sans replacement** et le processus est constitué d'une suite d'épreuves aléatoires **dépendantes**. Cependant, si la population échantillonnée est grande par rapport à l'échantillon, la différence entre un cas avec replacement et un cas sans replacement n'a pratiquement aucune importance.

Mesures de la distribution binômiale

La distribution binômiale est fondamentale à la définition de beaucoup d'expériences aléatoires dans tous les domaines de la gestion où on a à effectuer l'échantillonnage d'une population que ce soit des êtres ou des choses : en marketing ou en management (sondages, recensements), en production (contrôle de la qualité), en comptabilité (vérification). Cependant, il est relativement difficile et certainement fastidieux de calculer les probabilités de Bi(n,p) avec la formule affichée ci-haut. Ainsi, et assez souvent, on définit une expérience aléatoire fondée sur la distribution binômiale pour évaluer par la suite les probabilités des diverses expériences à partir d'une autre distribution connexe. Afin de faire ce genre d'évaluation, on a besoin de certaines mesures de la distribution binômiale dont principalement l'espérance mathématique et la variance :

Mesures de position et de dispersion pour X en Bi(n,p)

1) **L'espérance mathématique** : E(X) = np
2) **La variance** : Var(X) = npq

Exemple LO 2. Pour la v.a. X de l'exemple LO 1 (le nombre de bonnes pièces observées après quatre tirages au hasard), l'espérance mathématique = 4x0,99 = 3,96 et la variance = 4x0,99x0,01 = 0,0396. Ainsi l'écart type de X est $(0,0396)^{½}$ = 0,198997.

Calcul pratique des probabilités d'une v.a. binômiale

Lorsque l'on veut obtenir effectivement certaines probabilités relatives à une v.a. binômiale, le calcul de ces probabilités peut s'avérer une tâche assez ardue. Ainsi, par exemple, si X est en B(100, .40) et si l'on veut calculer P(X = 40), on a P(X = 40) = $\binom{100}{40}(0.40)^{40}(0.60)^{60}$, ce qui représente une certaine somme de calculs pour obtenir la valeur numérique correspondante. Évidemment si l'on désire P(X ≥ 40), la somme de calculs est encore plus considérable. En conséquence, pour simplifier l'utilisation de la distribution binômiale, on a regroupé les probabilités relatives à cette distribution dans des tables.

2. Loi de Poisson

Il existe plusieurs phénomènes aléatoires pour lesquels la réalisation d'un certain événement arrive peu fréquemment pendant une certaine unité de temps ou d'espace (de distance, d'aire, de volume). Il en est ainsi, par exemple, pour les expériences consistant à observer le nombre d'appels enregistrés dans un centre d'appels téléphoniques pendant 1 minute, le nombre de véhicules arrivant à un carrefour entre deux feux rouges, le nombre d'orages électriques qu'il y a dans la région pendant une année. Il en est de même pour les expériences consistant à observer le nombre de fautes de frappe dans un texte dactylographié, le nombre de défectuosités par plaque de circuits électroniques, le nombre de gâteaux mal enrobés de chocolat sur une chaîne de production alimentaire...

Ce type de processus aléatoire est appelé **processus de Poisson**; ce processus a une certaine analogie avec le processus de Bernoulli. Dans le processus de Bernoulli, on a des succès distribués de façon aléatoire au cours d'une suite d'épreuves alors que, dans le cas du processus de Poisson, on a des succès distribués d'une façon aléatoire pendant une certaine période de temps ou d'espace (divisée en unités).

Plus formellement, on peut caractériser ce processus décrivant les occurrences d'un certain événement, que l'on qualifie de succès, pendant une certaine période comme suit :

1) les occurrences de succès dans des intervalles de temps disjoints sont indépendantes entre elles;

2) pour un intervalle de temps (ou un espace) suffisamment petit :

   a) la probabilité d'un succès est proportionnelle à la longueur de l'intervalle de temps (ou la grandeur de l'espace). La constante de proportionnalité k est *l'intensité* du processus et correspond au nombre moyen de succès par unité de mesure,

   b) la probabilité de plus d'un succès est si petite par rapport à la probabilité d'un succès qu'elle peut être négligée.

Ce type de processus est fréquemment rencontré en gestion et en particulier dans l'étude des files d'attente.

La v.a. que l'on associe le plus souvent au processus de Poisson est la v.a. X qui donne le nombre x de succès obtenus dans une période de temps ou un espace donné t et que l'on appelle la v.a. de

Poisson. Si k est l'intensité du processus considéré et t la période de temps ou l'espace fixée, alors X peut prendre les valeurs entières x = 0, 1, 2, ... avec les probabilités $P(X = x) = e^{-\lambda}\lambda^x/x!$, où $\lambda = kt$ et **e = 2.718281828**... Cette dernière est une constante très souvent rencontrée en mathématiques.

La v.a. de Poisson est définie comme suit :

V.a. de Poisson. Une v.a. discrète X qui prend les valeurs entières x = 0, 1, 2, ..., avec les probabilités

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$, où λ est un nombre réel positif

s'appelle une v.a. de Poisson de paramètre λ.

Notation : On désigne par Po(λ) la famille de toutes les v.a de Poisson de paramètre λ.

Mesures de position et de dispersion de X qui est en Po(λ)

1) **E(X)** = λ ;

c'est-à-dire que le paramètre λ = kt indique le nombre moyen de succès dans la période de temps ou d'espace t considérée.

2) **Var(X)** = λ.

Calcul pratique des probabilités d'une distribution de Poisson

Étant donné l'utilisation très fréquente de la distribution de Poisson, on dispose de tables regroupant les probabilités relatives à cette distribution pour différentes valeurs du paramètre λ. Il existe des tables donnant la fonction de masse de probabilité f(x) = P(X = x) ou les probabilités cumulées du type $P(X \le x)$. À l'annexe du module ÉB (table I), on trouve une table du type $P(X \ge x)$, et cela pour diverses valeurs du paramètre λ.

Exemple LO 3. L'arrivée des clients à un supermarché est considérée comme un processus de Poisson. On sait que le nombre moyen de clients arrivant par minute à ce supermarché est 2. Calculer la probabilité que, pendant une période particulière de 5 minutes, il arrive 12 clients.

Solution : Le processus de Poisson considéré est un processus d'intensité k = 2 par minute. La v.a. X donnant le nombre x de clients arrivant au supermarché pendant le temps t = 5 minutes est une v.a. de Poisson de paramètre $\lambda = kt = 2 \times 5 = 10$, c'est-à-dire X est en Po(10). On cherche la probabilité que X égale 12. On a

$$P(X = 12) = \frac{e^{-10}\, 10^{12}}{12!}$$

Si l'on recourt à une table donnant les probabilités $P(X \geq x)$ (Table I module ÉB), on procède ainsi :

> La table donne les probabilités pour les valeurs (entières) de X groupées sous diverses valeurs de λ. Elle est repartie entre des ensembles de 10 valeurs de λ qui sont inscrites aux entêtes des colonnes de chaque groupe de valeurs de X. Les valeurs de X sont inscrites dans une colonne à gauche dans chaque groupe et elles commencent toujours avec x=1 (au lieu de 0) parce que $P(X \geq 0)=1$. Pour obtenir $P(X = 12)$, il faut chercher $P(X \geq 12) - P(X \geq 13)$. Or, dans la partie de la table où se trouve l'entête de colonne $\lambda = 10$, on cherche les deux probabilités à la ligne x = 12 et x = 13 respectivement, soit 0.303 et 0.208. Enfin la différence donne $P(X=12) = 0.095$.

Exemple LO 4. Sachant que la probabilité pour un individu d'avoir une mauvaise réaction à la suite de l'injection d'un sérum est de 0.001, déterminer la probabilité que, sur un contingent de 2 000 personnes vaccinées, il y en ait **au plus trois** qui aient une mauvaise réaction ?

Solution : Comme les vaccins sont administrés à des personnes provenant d'une très grande population, on peut considérer que les personnes sont choisies avec replacement. Alors, la v.a. X donnant le nombre de personnes ayant une mauvaise réaction parmi les 2 000 est une v.a. binômiale, et plus précisément X est en Bi (2 000, .001). Puisque n est assez grand et p petit, on peut considérer que la v.a. X suit approximativement une distribution de Poisson de paramètre $\lambda = 2\,000 \times .001 = 2$. Si l'on recourt à une table de Poisson, on obtient,

$$P(X \leq 3) = 1 - P(X > 3) = 1 - P(X \geq 4) = 1 - 0.143 = 0.857$$

Exemple LO 5. Un manufacturier reçoit des lots de 10 000 pièces de différents fournisseurs. Il sait très bien qu'inévitablement il y aura un certain pourcentage de pièces défectueuses. Il cherche une règle de contrôle de la qualité qui le conduirait à refuser les lots contenant 5 % ou plus de pièces défectueuses, à accepter ceux contenant 2 % ou moins de pièces défectueuses, et il est indifférent pour les lots dont le pourcentage de pièces défectueuses se situe entre 2 % et 5 %. Il fixe son choix sur la règle qui consiste à tirer de chaque lot un échantillon (A.S.S.R.) de 100 pièces et à accepter seulement les lots pour lesquels il observe moins de 4 pièces défectueuses dans l'échantillon. Construire la **courbe d'efficacité** de cette règle de décision et indiquer jusqu'à quel point elle permet au manufacturier de satisfaire ses désirs.

Solution : La courbe d'efficacité représente les probabilités que le lot soit accepté, selon la règle choisie, pour différentes valeurs de p (le pourcentage de pièces défectueuses dans le lot). On désigne par X la v.a. correspondant au nombre de pièces défectueuses dans l'échantillon. Alors X suit une distribution apparentée à la binômiale (qui s'appelle l'hypergéométrique).

On veut savoir la probabilité d'avoir *moins de 4* défectueuses dans un lot où la probabilité de défectueuses est p et la taille du lot est 100, soit $P(X<4)$. Cependant, on ne connaît pas la vraie probabilité de défectueuses dans un lot donné. Même un bon fournisseur peut générer des lots de qualité plus ou moins variable. Ainsi, pour se protéger contre cette variabilité (et encore plus pour contrer la variabilité entre les divers fournisseurs du cas), on veut savoir cette probabilité $P(X < 4)$ pour diverses valeurs de p.

Or, la taille du lot est fixe (n=100). Donc, on a un cas analogue à la distribution du nombre d'erreurs de frappe dans un document donné ou la distribution du nombre de véhicules qui arrivent à un feu de circulation dans un intervalle de temps fixe. C'est un cas qui se prête à la distribution de Poisson où l'intensité de la distribution est $\lambda = n*p = 100*p$ pour diverses valeurs de p. Ainsi, on calculera $P(X < 4)$, où X est Poisson, pour différentes valeurs de $\lambda = 100*p$. À la Figure LO 1 on a inscrit les résultats de ces calculs de probabilité en se servant de la table de la distribution de Poisson en annexe du module ÉB. Il s'agit de chercher dans cette table les probabilités $P(X \geq 4 \mid \lambda = 100*p)$ et d'effectuer par la suite l'opération $P(X < 4) = 1 - P(X \geq 4)$.

p	$P(X < 4 \mid \lambda = 100*p) =$ $1 - P(X \geq 4 \mid \lambda = 100*p)$
.01	.981
.02	.857
.03	.647
.04	.433
.05	.265
.06	.151
.	.
.10	.010
.	.
.15	0

Courbe d'efficacité

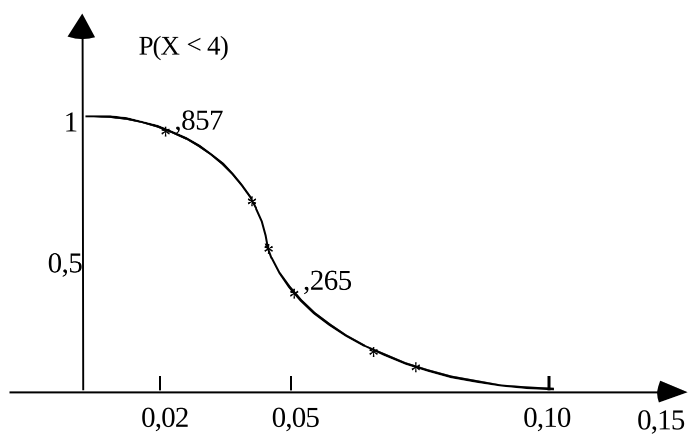


Figure LO 1

On obtient donc, pour la règle choisie, que le manufacturier a une probabilité d'au moins 0.857 d'accepter les lots contenant 2 % ou moins de pièces défectueuses et une probabilité d'au moins 1 – 0.265 = 0.735 de refuser les lots contenant 5 % ou plus de pièces défectueuses.

3. Loi exponentielle

Si l'on considère le processus de Poisson sous un angle un peu différent en ne fixant pas l'intervalle de temps ou l'espace, mais en définissant une v.a. X correspondante au temps ou à l'espace qu'il faut pour obtenir un premier succès, ou encore le temps ou l'espace x *entre deux succès consécutifs*. Cette v.a. X est appelée une v.a. exponentielle.

V.a. exponentielle. Une v.a. continue X pouvant prendre toutes les valeurs dans l'intervalle $[0,\infty)$, pour k réel positif, dont la fonction de densité est

$$f(x)=\begin{cases}0, & \text{si } x<0\\ k\,e^{-kx}, & \text{si } x\geq 0\end{cases}$$

s'appelle une v.a. exponentielle de paramètre k.

Le paramètre k, l'intensité du processus de Poisson sous-jacent, est le nombre moyen de succès *par unité de temps ou d'espace*.

Mesures de position et de dispersion de X en Exp(k)

1) **E(X) = 1/ k**

2) **Var(X) = 1/ k^2**

La fonction de répartition (fonction de distribution de probabilité cumulée) de la v.a. exponentielle X de paramètre k est donnée par

$$F(x)=\begin{cases}0, & \text{si } x<0\\ 1-e^{-kx}, & \text{si } x\geq 0\end{cases}$$

Dans l'étude des problèmes de file d'attente, on fait souvent l'hypothèse que la durée de service se distribue selon une v.a. exponentielle.

Exemple LO 6. À l'exemple LO 3, on a défini la v.a. X donnant le nombre x de clients arrivant à un supermarché pendant une certaine période de temps. On savait que l'arrivée des clients à ce supermarché pouvait être considérée comme un processus de Poisson d'intensité k = 2, c'est-à-dire que le nombre moyen de clients arrivant par minute était 2. Dans le contexte de ce processus, au lieu de considérer la v.a. X, on s'intéresse maintenant à la v.a. Y donnant le temps y entre 2 arrivées consécutives. Alors, sachant qu'un client vient tout juste d'arriver au supermarché, on cherche la probabilité qu'il faille attendre plus d'une minute avant qu'un autre client arrive.

Solution : Les v.a. Y donnant le temps d'attente entre 2 arrivées consécutives est une v.a. exponentielle de paramètre k = 2. La densité de Y est :

$$f(y) = \begin{cases} 0, & \text{si } y < 0 \\ 2e^{-2y}, & \text{si } y \geq 0 \end{cases},$$

d'où on trouve $F(y) = 1 - e^{-2y}$ si $y \geq 0$. On cherche $P(Y \geq 1)$. En se servant du fait que $P(Y\geq1) + P(Y\leq1) = 1$, on obtient alors,

$P(Y\geq1) = 1 – P(Y\leq1) = 1 – F(1) = 1 – (1 – e^{-2*1}) = e^{-2} = 0{,}135$.

4. Loi normale

La loi normale (appelée aussi distribution de Laplace-Gauss) est de loin la plus connue et la plus importante de toutes les distributions de probabilités utilisées en probabilités et en statistique. La v.a. normale est une v.a. qui peut être associée à une multitude d'expériences aléatoires qui se présentent en gestion, en économie, en ingénierie, en sciences sociales, etc. De plus, cette distribution s'avère fondamentale en théorie de l'échantillonnage et en inférence statistique où elle va se servir de distribution approximative pour plusieurs statistiques calculées à partir d'un échantillon aléatoire. Enfin, cette distribution présente des propriétés mathématiques exceptionnelles et joue le rôle de distribution limite pour plusieurs autres distributions, en particulier pour la loi binômiale et pour la loi de Poisson.

Idée intuitive. Il existe un grand nombre d'expériences aléatoires dans le cadre desquelles la variation d'une certaine caractéristique se présente comme la résultante d'un grand nombre de causes indépendantes à effets additifs, chacune de ces causes ayant un effet négligeable sur l'ensemble de l'expérience : cette variation est souvent alors une v.a. normale.

Par exemple, en production maraîchère, les poids des têtes de laitue dépendent des variations du temps, de la température, du micro-climat des champs, de l'habilité des cueilleurs et des cueilleuses. Si l'on pèse les têtes à la réception de l'entrepôt et si on range les poids en ordre croissant, on constate que les poids se concentrent autour d'une certaine valeur. En plus, il y aura une certaine variabilité où plus on s'éloigne ou vers les poids lourds, ou vers les poids légers, on a moins d'observations.

On peut résumer les observations du poids X de la façon suivante : X prend des valeurs x concentrées autour d'une certaine valeur centrale μ avec des probabilités fortes, et plus ces valeurs x sont éloignées de cette valeur centrale, plus les probabilités avec lesquelles X prend ces valeurs sont faibles. Ainsi, par exemple, on considère souvent des caractéristiques telles que la taille et le quotient intellectuel des individus dans une population comme des variables normales.

La v.a. normale est définie comme suit :

V.a. normale. Une v.a. continue X pouvant prendre toutes les valeurs réelles x, pour μ réel positif ou négatif, σ réel positif, dont la fonction de densité est :

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

s'appelle une v.a. normale (ou v.a. de Laplace-Gauss) de paramètre μ et σ^2.

Notation : On désigne par N (μ, σ^2) la famille de toutes les v.a. normales de paramètres μ et σ^2.

Comme la distribution normale est d'une très grande importance, on va attacher une attention particulière à ses propriétés. Il s'agit bien d'une distribution de probabilité parce que

1º f(x) > 0, pour tout x réel, et que

2º l'accumulation de probabilité dans tout l'intervalle (de moins infini jusqu'à plus infini) = 1

On pourrait aussi montrer que la fonction f(x) possède les caractéristiques suivantes :

1) f(x) a **un maximum à x = μ,**

2) f(x) a des points d'inflexion à x = μ ± σ (où la pente croissante devient décroissante et vice versa),

3) f(x) est **asymptotique à l'axe 0X** pour x allant vers ± ∞ , c'est-à-dire que f(x) tend vers zéro au fur et à mesure que x approche +∞ ou −∞ .

4) f(x) est **symétrique** par rapport à x = μ.

En conséquence, on a comme graphe de la fonction de densité f(x) la célèbre **courbe normale** en forme de cloche. (figure LO 2)

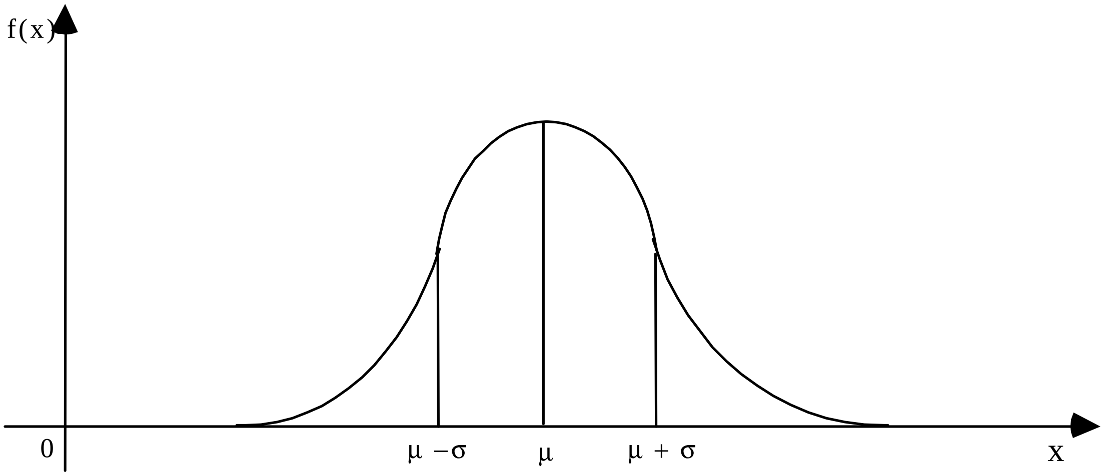


Figure LO 2

Pour la courbe normale, des variations des constantes μ et σ ont des conséquences bien spécifiques; en effet, une variation de μ (σ demeurant fixe) entraîne une translation de la courbe suivant l'axe des x, c'est-à-dire un déplacement de la courbe pour que le maximum soit toujours à μ. D'autre part, une augmentation de σ entraîne un aplatissement de la courbe, une dispersion plus large de la probabilité alors qu'une diminution de σ entraîne une contraction de la courbe, une concentration de la probabilité.

Mesures de position et de dispersion pour une v.a. X qui est en N (μ, s^2)

1º **E(X) = μ**; également Mé(X) = μ et Mode(X) = μ;

2º **Var (X) = s^2**;

Distribution de probabilité d'une fonction d'une ou de plusieurs v.a. normales

Distribution de probabilité d'une fonction linéaire de X qui est en N(μ, s^2)

Soit X en N(μ, σ^2) et Y = aX ± b où a est réel et non-nul, et b est réel, alors

Y est en N(aμ ± b, $a^2\sigma^2$).

Autrement dit, la loi de probabilité de la v.a. Y = aX ± b est une loi normale de moyenne E(Y) = aμ ± b et de variance Var(Y) = $a^2\sigma^2$.

Exemple LO 7 : Si X en N (6,3) est le temps de fabrication d'un lot de moteurs électriques, et Y = 10*X + 25 est le coût de fabrication (en dollars), alors Y est une v.a. en N(10*6+25,10^2*3) ou Y est en N(85,300).

Propriété de reproduction d'une fonction linéaire de plusieurs v.a. normales

Cette propriété présente le fait que toute somme pondérée de v.a. normales et indépendantes est elle-même une variable normale avec une espérance mathématique et une variance définies d'une façon naturelle.

Soit X_1, X_2, ..., X_n des v.a. indépendantes et telles que $X_i \in N(m_i, s_i^2)$, i = 1, ..., n, et c_i, i=1, ..., n, des constantes réelles non toutes nulles, **alors** la v.a. $Y = \sum_{i=1}^{n} c_i X_i$ a comme distribution de probabilité la distribution normale $N\left(\sum_{i=1}^{n} c_i m_i, \sum_{i=1}^{n} c_i^2 s_i^2\right)$.

Cas particulier : Si X_1, ..., X_n sont des v.a. indépendantes et telles que X_i, ε N(μ, σ^2), i = 1, ..., n, et si c_i = 1/n pour i = 1, ..., n , alors la moyenne arithmétique des X_i est

$$\overline{X} = \sum_{i=1}^{n} X_i / n \text{ est en } N(m, s^2/n) \cdot$$

Ainsi, si on fait des observations dans une population d'une caractéristique qui suit une distribution normale, la moyenne arithmétique des observations suit elle-même une distribution normale **et la variance est réduite par le nombre d'observations.**

Calcul pratique de probabilités d'une v.a. normale

Lorsque X est en $N(\mu, \sigma^2)$, on sait que μ peut prendre n'importe quelle valeur réelle et que σ^2 peut prendre n'importe quelle valeur réelle positive. La plupart du temps, lorsqu'on utilise une v.a. normale, on est intéressé à obtenir des probabilités cumulées de la forme $P(X \leq c)$, $P(X \geq c)$ ou $P(c_1 \leq x \leq c_2)$, où c, c_1 et c_2 sont des constantes. Si, par exemple, on veut calculer $P(X \leq c)$, on a en principe à calculer une intégrale qui n'est pas facile à évaluer. Pour faciliter cette tâche on peut utiliser des tables. Cependant, comme μ et σ^2 peuvent prendre une infinité de valeurs différentes, il aurait fallu construire une infinité de tables. Pour contourner cette difficulté, on peut utiliser les propriétés 1º et 2º décrites ci-haut. Ainsi, à partir d'une table donnant les probabilités relatives à la distribution N(0, 1), on peut obtenir toutes les probabilités relatives à une v.a. X en $N(\mu, \sigma^2)$. Il suffit de déterminer une nouvelle variable Z telle que

$$Z = (X - \mu)/\sigma .$$

La variable Z est en N(0,1) et est appelée la variable aléatoire **normale centrée réduite**.

Il existe des tables de plusieurs formes pour la distribution N(0,1). La table II de l'annexe du module ÉB donne des probabilités $P(Z \geq z) = 1 - P(Z \leq z) = 1 - F(z)$. Puisque la courbe de la densité Z en N(0, 1) est symétrique par rapport à la droite z = 0, la table des probabilités cumulées ne donne $P(Z \geq z)$ que pour $z > 0$, les valeurs de ces probabilités pour $z < 0$ étant obtenues par **symétrie**. Ainsi, **P(Z £ -z) = P(Z ≥ z)** parce que les valeurs de Z correspondant à chacune des probabilités sont également éloignées de la valeur 0, l'une vers le bas ou « à gauche » et l'autre vers le le haut ou « à droite ». On remarque qu'une v.a. normale est une variable continue et alors, $P(Z < z) = P(Z \leq z)$ et $P(Z > z) = P(Z \geq z)$.

L'utilisation de la **Table II au module ÉB** se fait de deux façons :

- pour trouver une ou des probabilités correspondant à une ou des valeurs de la v.a.
- pour trouver une ou des valeurs de la v.a. correspondant à une ou des probabilités.

Toujours, avant de chercher des valeurs dans la table, on transforme la v.a. en une v.a. centrée (en soustrayant la valeur espérée, soit $(X - \mu)$), réduite (en divisant par l'écart type), soit $(X - \mu)/\sigma$.

La table comprend des valeurs de Z inscrites à la tête des colonnes et à la gauche des lignes. Les valeurs sur les lignes changent par intervalle de 0,1 et dans les colonnes par intervalle de 0,01. Ainsi, à la croisée d'une ligne où on trouve 1,2, p.ex., et une colonne où on trouve 0,08, on a la probabilité $P(Z \geq 1{,}2+0{,}08=1{,}28)=0{,}10$. D'une façon inverse, en cherchant la valeur de Z, appelons-la z, telle que $P(Z \geq z)=0{,}10$, on trouve la valeur de 0,10 dans le corps de la table et on fait la somme de la valeur à gauche sur la ligne et la valeur à la tête de la colonne où réside la valeur de la probabilité de 0,10 soit 1,2+0,08=1,28.

Remarque LO 3. La valeur de $P(-3 \leq Z \leq 3)$ est très voisine de l'unité, c'est-à-dire que la probabilité de trouver des valeurs d'une v.a. normale plus loin que 3 fois l'écart type est très faible.

Exemple LO 8. Dans une certaine usine, le salaire moyen est de 8,00 $ l'heure avec une déviation standard de 0,50 $. On suppose que la v.a. X représentant le salaire est une v.a. normale.

1) Si on choisit un travailleur au hasard dans cette usine, quelle est la probabilité que ce travailleur

   a) gagne au moins 8,50 $ l'heure ?

   b) gagne entre 7,50 $ et 8,50 $ l'heure ?

   c) gagne moins de 7,00 $ l'heure ?

2) Quelle est le salaire maximum du 10 % des travailleurs les moins bien rémunérés ?

3) Le 5 % des travailleurs les mieux rémunérés reçoivent un salaire d'au moins quel montant l'heure ?

Solution : 1a) On a $X \in N(8{,}00\ \$, (0{,}50\ \$)^2)$, que l'on écrit plus simplement $N(8,(0.5)^2)$. On cherche $P(X \geq 8{,}5)$.

$$P(X \geq 8{,}5) = P\left(\frac{X - m}{s} \geq \frac{8{,}5-8{,}0}{0{,}5}\right) = P(Z \geq 1) = 0{,}159\,,$$

selon la table des probabilités cumulées de la loi normale centrée réduite (table II, module ÉB).

1b) On cherche $P(7{,}5 < X < 8{,}5)$. En se ramenant à une v.a. normale centrée réduite, on obtient

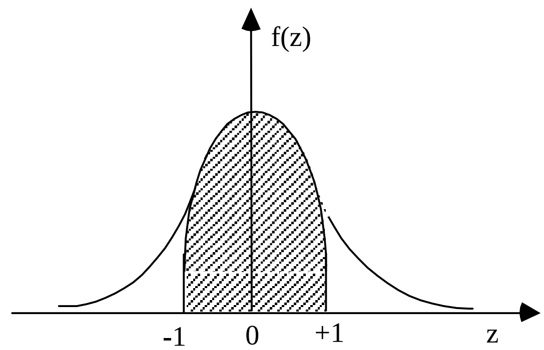
$$P\left(\frac{7{,}5-8{,}0}{0{,}5} < \frac{X - m}{s} < \frac{8{,}5-8{,}0}{0{,}5}\right) = P(-1 < Z < 1)$$

On se rappelle que $P(-1 < Z < 1) = P(-1 \leq Z \leq 1)$ parce que Z est une variable aléatoire continue.

On a P(0 < Z <1)

= 0.5 - P(Z ≥ 1)

= 0.5 - 0.159 = 0.341



Par symétrie P(-1 < Z < 0) = P(0 < Z < 1) = 0,341 (voir le diagramme précédent).

Donc P(7.5 < X < 8.5) = P(-1 < Z < 1) = P(-1 <Z < 0) + P(0< Z < 1)

= 2 x 0,341 = 0,682.

c) On cherche P(X < 7.0). On a

$$P(X < 7,0) = P\left(\frac{X - m}{s} < \frac{7,0 - 8,0}{0,5}\right) = P(Z < -2,0)$$

mais par symétrie P(Z < -2,0) = P(Z > 2,0) = 0,023

Donc P(X < 7,0) = 0,023.

2) On cherche le salaire $x_{0,1}$ où $P(X<x_{0,1}) = 0,10$. En termes de Z, on a $P(Z< z_{0,1}) = 0,10$. Or, par symétrie, on a $P(Z< z_{0,1}) = P(Z> - z_{0,1}) = 0,10$. Ainsi, de la Table II on a $- z_{0,1} = 1,28$, soit $z_{0,1} = -1,28$ et $x_{0,1} = 8,0 - 1,28*0,5 = 7,36$ \$ l'heure. Dix pour cent des travailleurs gagnent 7,36 $ l'heure ou moins.

3) Les salaires des 5 % des travailleurs les mieux rémunérés sont plus grands ou égaux à un salaire limite $x_{0,05}$, p.ex. où $P(X \geq x_{0,05}) = P(Z \geq z_{0,05}) = 0,05$. Or, $z_{0,05} = 1,645$ (Table II) et donc $x_{0,05} = 8,0 + 1,645*0,5 = 8,82$ \$ l'heure.

Exemple LO 9. Une compagnie se demande si elle devrait acheter une certaine machine très automatisée, laquelle permettrait de sauver plusieurs heures de travail. Cette machine coûterait 15 000 $, serait utilisable pendant un an et aurait une valeur de rachat de 3 000 $.

Le responsable de la production estime qu'il pourrait sauver 2 500 heures de travail durant l'année avec cette machine et que chaque heure sauvée aurait une valeur de 5 $.

Après réflexion, il nuance un peu son énoncé initial quant au nombre d'heures sauvées en disant qu'il y a autant de chances que le nombre d'heures sauvées soit au-dessus de 2 500 qu'en dessous de 2 500 et qu'il y a 50 % de chances que le nombre d'heures sauvées soit entre 2 000 et 3 000.

a) Si l'on fait l'hypothèse que le nombre d'heures sauvées suit une distribution normale, déterminer les paramètres de cette normale sur la base des énoncés du responsable de la production.

b) La compagnie devrait-elle se procurer cette machine?

Solution : a) On fait l'hypothèse que la v.a. X donnant le nombre x d'heures sauvées si la compagnie achète la nouvelle machine est une v.a. normale, c'est-à-dire que $X \varepsilon N(\mu, \sigma^2)$. Sur la base des énoncés du responsable de la production, on peut conclure que μ = 2 500 heures et que P(X > 3 000) = 0,25. En se ramenant à la v.a. normale centrée réduite, on a

$$P(X \geq 3\,000) = P\left(Z \geq \frac{3\,000 - 2\,500}{s}\right) = 0{,}25$$

Selon la Table II au module ÉB, on obtient $\left(\frac{3\,000 - 2\,500}{s}\right) = 0{,}67$, donc $s = \frac{500}{0{,}67} = 746.28$ heures.

b) En se référant au modèle du point mort (probabilisé), il est facile de déterminer une valeur x^o telle que, si $x < x^o$, la compagnie n'a pas intérêt à acheter la nouvelle machine alors que, si $x \geq x^o$, elle a intérêt à le faire. Son bénéfice, si elle se procure la nouvelle machine, est donné par $\pi = 5x - 12\,000$ $. Donc, pour qu'il soit avantageux d'acheter la machine ($\pi \geq 0$, ce bénéfice est nul si l'on ne se procure pas la nouvelle machine), il faut que $x \geq x^o = \frac{12\,000}{5} =$ 2 400 heures. Il est facile de vérifier que selon le critère de la valeur espérée, l'action optimale sera d'acheter la machine si $E(X) \geq x^o$. Puisque $\mu = E(X) = 2\,500 > x^o = 2\,400$, la compagnie devrait acheter la nouvelle machine.

4.1 Théorème central limite

L'une des propriétés les plus remarquables de la loi normale est le fait que plusieurs v.a. ont des distributions de probabilité ***qui convergent justement vers la loi normale***. L'expression «théorèmes limites» fait reférence à des résultats théoriques qui démontrent la convergence de ces diverses v.a. vers la v.a. normale. Le plus célèbre et le plus important de ces théorèmes est le **théorème central limite** qui prouve, en bref, que plus on a des observations indépendantes d'une v.a., plus ***la distribution de la somme des observations ressemble à une distribution normale***.

Plus précisément, le théorème s'énonce comme suit : soit X_1, ... X_n, des v.a. *indépendantes de même distribution de probabilité*, d'espérances μ_i et de variances s_i^2, $i = 1, \ldots, n$, alors, lorsque n tend vers l'infini, la distribution de probabilité de la v.a. $Y = \sum_{i=1}^{n} X_i$ converge vers la distribution normale d'espérance $m_y = \sum_{i=1}^{n} m_i$ et de variance $s_y^2 = \sum_{i=1}^{n} s_i^2$. Autrement dit, lorsque n croît sans cesse, la loi de probabilité de la v.a. centrée réduite $\left(Y - \sum m_i\right)/\sqrt{\sum s_i^2}$ est approximativement **une distribution normale centrée réduite**.

Remarque LO 4. Ce qu'il y a de particulièrement étonnant dans ce théorème, c'est que cette convergence de la distribution de Y vers la distribution normale a lieu ***quelle que soit la nature de la distribution des v.a.*** X_i, i = 1, ..., n.

Exemple LO 10. On considère l'expérience consistant à lancer un dé un certain nombre de fois. On le lance d'abord une première fois et on associe à cette première expérience la v.a. X_1 donnant le nombre x_1 de points obtenus sur la face supérieure du dé. On sait que $f(x_1) = 1/6$, pour $x_1 = 1, 2, 3, 4, 5, 6$, masse de probabilité qui est représentée par un diagramme en bâtons dans la figure LO 3 (cas n = 1). On peut considérer que $X_1 = Y_1$. Si on lance le dé une deuxième fois, on associe à ce deuxième lancer la v.a. X_2 donnant le nombre de x_2 de points obtenus sur le dé. Par la suite, on définit la v.a. $Y_2 = X_1 + X_2$ dans la figure LO 3 (cas n = 2). De la même façon, on lance le dé une troisième fois, une quatrième fois, etc., associant à chaque fois au n$^{\text{ième}}$ lancer la v.a. X_n et par la suite la v.a. $Y_n = \sum_{k=1}^{n} X_k$, où Y_n donne la somme des points obtenus lorsqu'on lance le dé n fois. La figure LO 3 donne le diagramme en bâtons de la distribution de Y_n, pour n = 4. La v.a. Y_n est une somme des n v.a. indépendantes X_i, i = 1, ..., n, pouvant chacune prendre les valeurs 1, 2, 3, 4, 5 et 6 et ayant même distribution de probabilité. Alors, le théorème central limite s'applique et, lorsque n tend vers l'infini, la distribution Y_n converge vers la distribution normale d'espérance $E(Y_n)$ et de variance Var (Y_n). On peut voir dans les figures LO 3 comment se modifie la forme du diagramme en bâtons représentant la distribution de Y_n à mesure que n augmente. Déjà, pour n = 4, la forme de la courbe normale se dessine clairement.

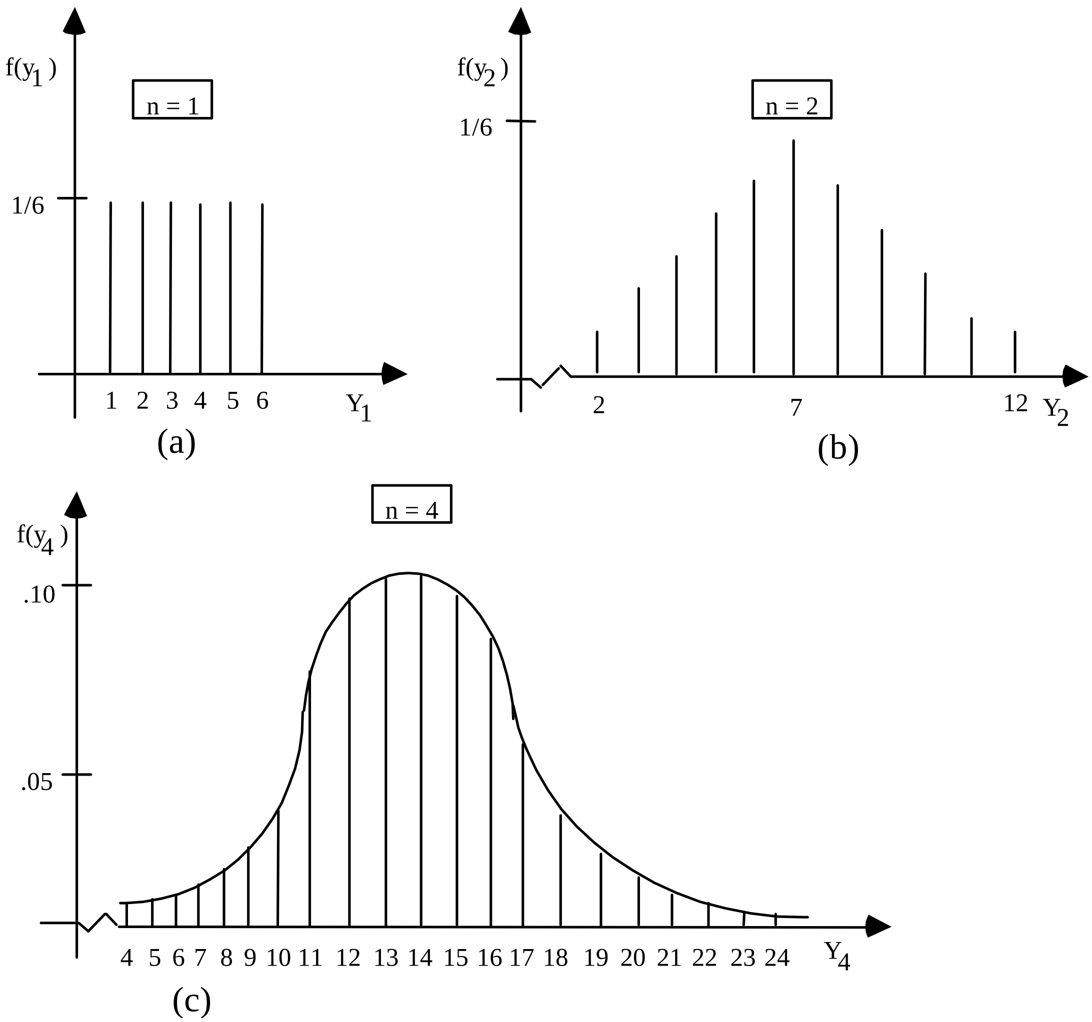


Figure LO 3

4.2 Approximation de la loi binômiale par la loi normale

Si X est une v.a. binômiale de paramètres n et p, c'est-à-dire X est en Bi(n, p), alors lorsque n tend vers l'infini et p demeure fixe, la distribution de probabilité de la v.a. X converge vers la distribution normale d'espérance np et de variance npq, ce que l'on exprime plus simplement par

$$\mathrm{Bi}(n, p) \xrightarrow[n \to \infty]{} N(np, npq).$$

Autrement dit, lorsque n croît vers l'infini, la distribution de la v.a. $(X - np)/\sqrt{npq}$ converge vers la distribution normale centrée réduite.

Ce résultat est une application directe du théorème central limite. En effet, toute v.a X_i en Bi(n, p) peut être considérée comme la somme de n v.a. Bernoulli $X_1, X_2, ..., X_n$ où X est en Bi(1, p), i = 1, ..., n, indépendantes, c'est-à-dire $X = \sum X_i$. On sait de plus que $E(X) = E\left(\sum X_i\right) = np$ et que $\mathrm{Var}(X) = \mathrm{Var}\left(\sum X_i\right) = npq$. Alors, par le théorème central limite, on conclut que, lorsque n tend vers l'infini, la distribution de $X = \sum X_i$ converge vers la distribution normale d'espérance np et de variance npq.

Remarque LO 5. En général, on considère que l'approximation de la binômiale par la normale est bonne lorsque n > 30 et p voisin de 1/2. On convient généralement que l'approximation est acceptable lorsque n > 30, np > 5 **et** n (1-p) > 5.

Exemple LO 11. Les contribuables d'une ville vont être appelés à se prononcer par référendum au sujet d'un projet d'emprunt pour la municipalité. Pour se faire un idée de l'opinion des gens, on effectue un sondage auprès de 100 contribuables. Si en réalité la proportion de ceux qui favorisent le projet d'emprunt est égale à la proportion de ceux qui s'y opposent dans toute la population, quelle est la probabilité que, dans l'échantillon de 100 contribuables, au moins 60 contribuables soient en faveur du projet ?

Solution : On définit d'abord la v.a. X qui donne le nombre x de contribuables qui sont en faveur du projet d'emprunt parmi les 100 personnes interrogées. On peut supposer ici que la ville possède une population nombreuse et que, même si l'échantillonnage s'est fait sans remplacement, on peut pour cette raison considérer que l'on a affaire à un échantillon obtenu avec replacement. Alors, X est une v.a. binômiale de paramètres n = 100 et p = 1/2. c.est-à-dire X est en Bi(100, 1/2). On cherche $P(X \geq 60)$, ce qui donne

$$P(X \geq 60) = \sum_{x=60}^{100} \binom{100}{x} (1/2)^x (1/2)^{100-x} = 0.028$$

cette dernière valeur pourrait être lue dans une table de la loi binômiale avec n = 100 et p = 1/2.

Cependant, comme n (=100) est passablement grand et comme p est égal à 1/2, on sait que la distribution binômiale Bi(100, 1/2) va converger vers la distribution normale N(100 x 1/2, 100 x 1/2 x 1/2), soit N(50, 25).

On peut donc considérer que la distribution de X est approximativement une distribution normale d'espérance 50 et de variance 25. En conséquence,

$$\begin{aligned} P(X \geq 60) &\cong P\left(\frac{X - m}{s} \geq \frac{60 - 50}{5}\right) \\ &= P(Z \geq 2) = 0.023, \end{aligned}$$

d'où l'on voit que l'approximation de la binômiale par la normale est passablement précise dans ce cas particulier.

Remarque LO 6. Dans l'exemple précédent, si l'on veut calculer la probabilité qu'il y ait exactement 60 contribuables qui soient en faveur du projet parmi les 100 interrogés, on peut utiliser l'approximation normale en supposant que la probabilité $P(X = 60)$ lorsque X est considérée comme une binômiale est approximativement égale à $P(59.5 < X < 60.5)$ lorsque X est considérée comme une normale. Cette méthode correspond à une interpolation linéaire de la fonction de répartition de la distribution normale.

Pour compléter votre apprentissage des lois de probabilité usuelles nous vous recommandons d'essayer de résoudre les exercices du module LO qui se trouve en annexe. En cas de difficultés, vous pouvez consulter les solutions de ces exercices qui se trouvent également en annexe, ou communiquer avec le responsable du cours ou son adjoint pour plus d'explications.