

## VARIABLES ALÉATOIRES ET DISTRIBUTIONS DE PROBABILITÉ

Dans le module PR, pour étudier une expérience aléatoire, on a introduit les trois éléments :

- l'ensemble fondamental des résultats,
- une famille d'événements associée à l'ensemble fondamental et
- une fonction de probabilité  $P$ .

À une même expérience aléatoire, il est possible d'associer plusieurs groupes différents de ces trois éléments, puisqu'il n'existe pas de façon unique de définir l'ensemble des résultats. Ce n'est qu'une fois défini chacun des trois éléments dans un problème spécifique que l'on peut parler de la probabilité d'un résultat ou d'un événement relié à l'expérience aléatoire.

Cependant, pour une expérience donnée, une fois défini chacun de ces trois éléments, souvent on ne s'intéresse pas tant aux résultats eux-mêmes qu'à une ou des *caractéristiques particulières* de ces résultats, soit des sous-ensembles des résultats élémentaires. Par exemple, lorsqu'on lance une pièce de monnaie 4 fois, les résultats ont la forme (F, P, F, P), (F, P, P, F), .... Souvent, on ne s'intéresse pas tellement à tel résultat particulier (par exemple, au résultat (F, P, F, P)), mais plutôt à une caractéristique de ce résultat, tel le nombre de faces (dans ce cas, 2 faces).

De plus, pour une expérience donnée, il arrive souvent que les résultats ne sont pas définis sous forme numérique. Ainsi, dans l'exemple que l'on vient d'utiliser, un résultat a la forme de  $(\omega_1, \omega_2, \omega_3, \omega_4)$ , où  $\omega_i = P$  ou  $F$ ,  $i = 1, 2, 3, 4$ . Comme il est pratiquement impossible d'effectuer des opérations mathématiques sur des résultats de cette nature, il est utile de *faire correspondre* à chacun de ces résultats de type non numérique *un résultat de type numérique*; ainsi, à chacun des résultats  $(\omega_1, \omega_2, \omega_3, \omega_4)$  on fera correspondre, par exemple, le nombre de faces dans le résultat.

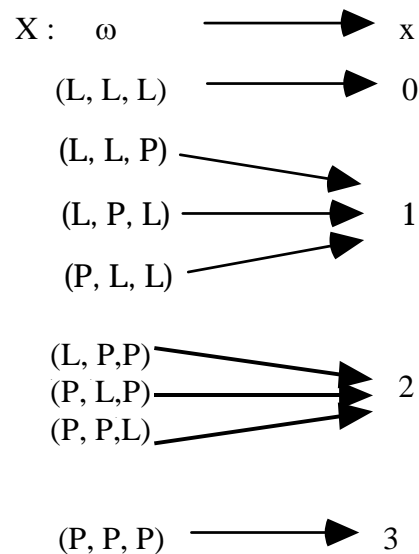
### 1. Notion de variable aléatoire

D'une façon générale, pour les calculs pratiques, il arrive très souvent que l'on fasse ainsi correspondre à chacun des résultats un nombre réel  $x$ . Cette correspondance est une fonction que l'on appellera une *variable aléatoire*. Ayant des valeurs numériques, une variable aléatoire facilite les calculs des probabilités.

**Exemple DS 1.** Dans l'espoir d'être mieux informé au sujet des intentions d'achat des marques d'auto de luxe, on décide d'effectuer un sondage-éclair. On choisit trois clients au hasard à la sortie du bureau d'une maison de placements et on les interroge sur leurs intentions d'achat. Pour simplifier le problème, on va supposer qu'il y a seulement deux marques importantes vendues au Québec : P et L. En assignant à chacun des 3 clients choisis la lettre "P" s'il déclare l'intention d'acheter la marque P ou la lettre "L" dans le cas inverse, on obtient comme ensemble fondamental

$\{(L, L, L), (L, L, P), (L, P, L), (P, L, L), (L, P, P), (P, L, P), (P, P, L), (P, P, P)\}$

Cependant, ce à quoi l'on s'intéresse, ce n'est pas tellement un résultat particulier comme, par exemple, (L, P, P) qui indique que le premier client interrogé achètera L et les deux autres P, mais plutôt, par exemple, le nombre de clients qui entendent acheter P dans l'échantillon. Dans cette perspective, il est utile de définir une fonction  $X$  qui permet d'associer à chaque résultat un nombre réel  $x$ , ce nombre  $x$  étant ici le nombre d'intentions d'acheter P dans l'échantillon de 3 clients. On aura donc



D'une façon générale, on aura :

**Variable aléatoire.** Étant donné un ensemble fondamental et une famille d'événements associée à une expérience aléatoire, on appelle variable aléatoire toute correspondance qui relie les événements aux valeurs numériques de façon à ce qu'un intervalle numérique corresponde à un événement de la famille.

**Notation.** Par convention, dans les problèmes où l'on ne traite que d'une seule variable aléatoire, on désignera cette fonction par  $X$  et ces valeurs par  $x$ . De plus, on désignera par  $\mathcal{X}$  l'ensemble

des valeurs prises par  $X$ . Dans les problèmes où l'on a affaire à plusieurs variables aléatoires, on pourra utiliser aussi parfois la lettre  $X$  indicée, c'est-à-dire  $X_1, X_2, X_3, \dots$ . Pour simplifier l'écriture, on va parfois utiliser l'abréviation "v.a." pour désigner "variable aléatoire".

### 1.1 Types de variables aléatoires

Selon la nature de l'ensemble  $X$  des valeurs prises par une variable aléatoire  $X$ , on distingue différents types de variables aléatoires.

$X$  est dite **variable aléatoire discrète** si la variable aléatoire  $X$  ne prend que des valeurs isolées.

$X$  est dite **variable aléatoire continue** si  $X$  est susceptible de prendre n'importe quelle valeur dans un certain intervalle réel.

**Exemple DS 2.** Soit l'expérience consistant à interroger trois clients choisis au hasard pour voir s'ils ont l'intention d'acheter la marque P ou L (exemple 2.1). On a alors l'ensemble fondamental  $= \{(L, L, L), (L, L, P), (L, P, L), (P, L, L), (L, P, P), (P, L, P), (P, P, L), (P, P, P)\}$ . Dans cet ensemble fondamental, on a défini (à l'exemple 2.1) la variable aléatoire  $X$  donnant le nombre de clients potentiels de P parmi les clients interrogés; ainsi, on avait, par exemple,

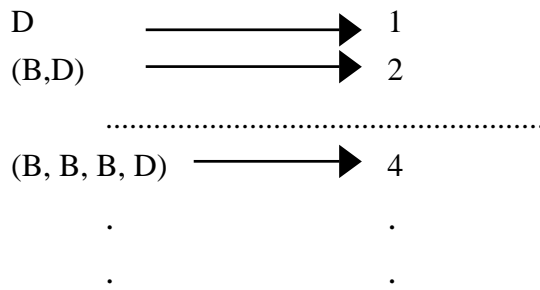
$$(L, P, P) \longrightarrow 2$$

L'ensemble des valeurs  $x$  prises par  $X$  est ici  $X = \{0, 1, 2, 3\}$ . Puisque les valeurs de  $x$  sont isolées ou distinctes,  $X$  est une variable aléatoire **discrète**.

**Exemple DS 3.** La production des transistors d'une compagnie d'appareils électroniques contient un certain pourcentage de transistors défectueux. Pour se faire une idée de ce pourcentage, on effectue l'expérience aléatoire suivante : dans la production, on choisit des transistors l'un après l'autre jusqu'à ce que l'on obtienne un premier transistor défectueux et, à ce moment-là, on arrête les tirages. Si l'on associe la lettre "B" aux bons transistors et la lettre "D" aux transistors défectueux, alors on a comme ensemble fondamental  $\{(D, (B, D), (B, B, D), (B, B, B, D), (B, B, B, B, D), \dots, (B, B, B, B, B, B, B, B, D), \dots\}$ .

On peut définir sur cet ensemble fondamental la variable aléatoire donnant le nombre de tirages nécessaires pour obtenir un premier transistor défectueux, c'est-à-dire :

$$\omega \longrightarrow x = X(\omega)$$



où  $X(\omega) = x$  n'est rien d'autre que le nombre d'éléments dans  $(\omega)$ . L'ensemble des valeurs prises par  $X$  est  $\mathcal{X} = \{1, 2, 3, 4, 5, 6, \dots\} =$  l'ensemble des entiers positifs.  $\mathcal{X}$  est un ensemble infini et dénombrable et la variable  $X$  est **discrète**.

**Exemple DS 4.** Un agent du ministère de la Consommation doit vérifier le poids des boîtes de café qu'un certain magasin présente à sa clientèle comme des boîtes de 500 grammes. À cette fin, il choisit au hasard une boîte de café sur un étalage de 20 boîtes. L'ensemble fondamental associé à cette expérience est  $\{\omega : \omega \text{ est l'une des 20 boîtes de l'étalage}\}$ . Par la suite, il définit la variable aléatoire  $X$  qui, à chaque boîte  $\omega$  choisie, fait correspondre son poids  $X(\omega) = x$ . Comme  $X$  peut prendre n'importe quelle valeur  $x$  dans un intervalle, par exemple dans l'intervalle  $[400, 600]$ ,  $X$  est une variable aléatoire **continue**.

## 2. Variable aléatoire discrète

### 2.1 Fonction de probabilité

On commence l'étude des variables aléatoires par celles qui ne prennent que des valeurs isolées, que l'on a appelées variables aléatoires discrètes. Une fois définie cette transformation  $X$  des  $\omega$  en nombres entiers positifs, on cherche à définir une fonction qui à chaque valeur  $x$  prise par  $X$  va faire correspondre la probabilité que  $X$  prenne la valeur  $x$ .

**Exemple DS 5.** Dans l'exemple DS 1, on a considéré l'expérience aléatoire consistant à choisir 3 clients au hasard et à les interroger sur leurs intentions d'acheter l'une ou l'autre marque d'auto de luxe, soit P ou L. Sur l'ensemble fondamental  $\{(L, L, L), (L, L, P), (L, P, L), (P, L, L), (L, P, P), (P, L, P), (P, P, L), (P, P, P)\}$ , on a défini la variable aléatoire  $X$  donnant le nombre de clients qui se déclarent acheteurs de P parmi les 3 personnes interrogées. La v.a.  $X$  prend les valeurs 0, 1, 2, 3. On a  $X = 0$  si le résultat de l'expérience est (L, L, L). Selon l'hypothèse de l'équiprobabilité des 8 résultats (ce qui revient à supposer qu'il y a autant d'acheteurs de P que de L parmi les clients du magasin), il semble donc naturel d'écrire

$$P(X = 0) = P(\{L, L, L\}) = 1/8$$

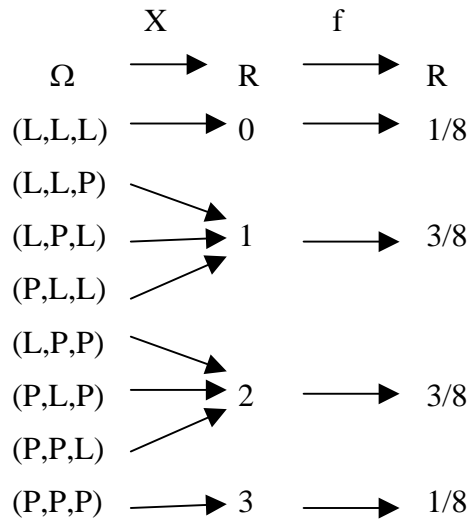
de même

$$\text{l'événement } \{\ll X = 1 \gg\} = \{(L, L, P), (L, P, L), (P, L, L)\}$$

d'où l'on a

$$P(X = 1) = P(\{(L, L, P), (L, P, L), (P, L, L)\}) = 3/8$$

On a de la même façon  $P(X = 2) = 3/8$  et  $P(X = 3) = 1/8$ . On a ainsi défini une fonction  $f$  qui, à chaque valeur  $x$  prise par  $X$ , fait correspondre la probabilité  $P(X = x)$  que la variable  $X$  prenne cette valeur  $x$ . En résumé, la variable aléatoire  $X$  et cette fonction  $f$  peuvent être représentées comme suit :



En général, on a la définition suivante :

**Fonction masse de probabilité de X.** Soit  $X$  une variable aléatoire discrète, on appelle fonction masse de probabilité de  $X$  une fonction notée  $f$  qui, à chaque valeur  $x$  prise par  $X$ , fait correspondre la probabilité que  $X$  prenne cette valeur  $x$ , c'est-à-dire

$$f(x) = P(X = x)$$

**Remarque DS 1** a) On utilise aussi les expressions plus générales **distribution de probabilité** ou encore **loi de probabilité** pour désigner la fonction de masse de probabilité. b) Pour toutes les valeurs réelles qui ne sont pas dans  $X$ , on peut considérer que la masse de probabilité de  $X$  prend la valeur zéro, c'est-à-dire  $f(x) = 0$ , pour  $x$  pas dans  $X$ . c) Il est souvent pratique de se

servir d'un tableau pour définir la masse de probabilité d'une variable aléatoire discrète : dans ce tableau, à chacune des valeurs  $x$  prises par la variable  $X$ , on associe la valeur  $f(x)$  prise par la masse de probabilité de  $X$ . Ainsi, pour définir la masse de probabilité de la variable aléatoire de l'exemple DS 5, on peut utiliser le tableau suivant :

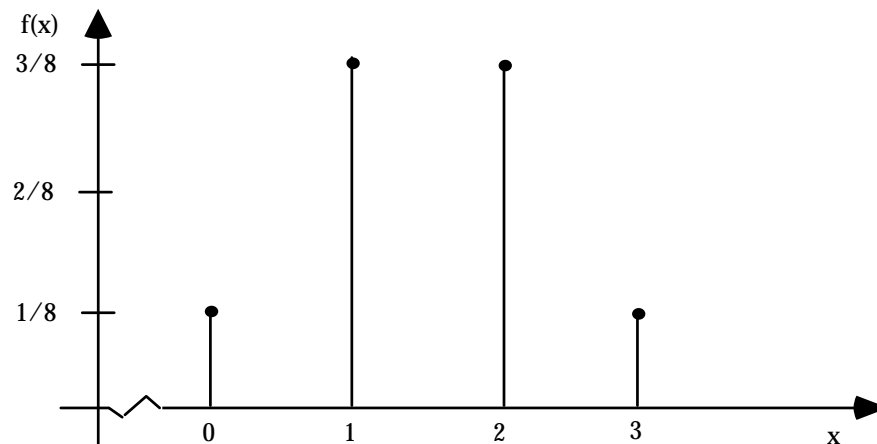
$x$	0	1	2	3	$x$ pas dans $X$
$f(x)$	1/8	3/8	3/8	1/8	0

d) On notera aussi par la suite la fonction masse de probabilité d'une v.a. discrète  $X$  par  $P(x)$ .

### Représentation graphique d'une fonction masse de probabilité

Pour représenter une fonction masse de probabilité, on utilise généralement un diagramme en bâtons et parfois l'histogramme.

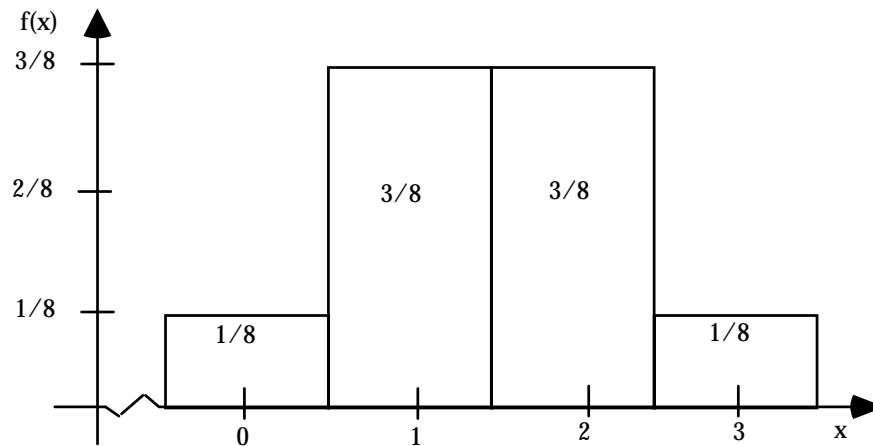
**Le diagramme en bâtons** est le mode de représentation le plus utilisé : dans le plan cartésien, on place en abscisse les valeurs  $x$  prises par la v.a.  $X$  et ensuite, sur chacune de ces valeurs, on élève un bâton (segment de droite) d'une hauteur égale à la masse de probabilité  $f(x)$ . Ainsi, pour la masse de probabilité de l'exemple DS 5 redéfinie par un tableau à la remarque DS 1 c), on obtient le diagramme en bâtons de la figure DS 1



**Figure DS 1 Diagramme en bâtons**

**L'histogramme** est parfois utilisé pour représenter une fonction masse de probabilité  $f(x)$ . Dans ce cas, on place sur l'axe horizontal les valeurs  $x$  prises par la v.a.  $X$ . Par la suite, pour chacune de ces valeurs, on élève un rectangle dont la base est centrée sur  $x$  et dont la hauteur est

proportionnelle à  $f(x)$ , la surface du rectangle étant égale à  $f(x)$ . Ainsi, la masse de probabilité de l'exemple DS 5 pourrait être représentée par l'histogramme de la figure DS 2.



**Figure DS 2 Histogramme**

Pour cet exemple particulier, il est convenable de construire des rectangles ayant une base de largeur 1, de telle sorte que la hauteur des rectangles soit égale à  $f(x)$ ; mais, en général, dans un histogramme, c'est la surface des rectangles que l'on prend comme égale à  $f(x)$ , de telle sorte que la hauteur des rectangles est alors donnée par  $f(x)/b$  où  $b$  désigne la longueur de la base des rectangles (que l'on prend en général égale pour tous les rectangles).

### Propriétés de la fonction masse de probabilité

Puisque, pour une valeur  $x$  donnée, la masse de probabilité  $f(x)$  est définie comme étant égale à la probabilité  $P(X = x)$ , il découle de la définition de la probabilité que  $f$  possède les deux propriétés suivantes :

1°  $f(x) \geq 0$ , pour tout  $x$  réel,

2°  $\sum_{x \in \mathcal{X}} f(x) = 1$ ,

cette dernière sommation pouvant être prise seulement sur les  $x$  dans  $\mathcal{X}$  puisque, pour  $x$  pas dans  $\mathcal{X}$ , on a  $f(x) = 0$ .

## 2.2 Fonction de répartition

Très souvent, lorsqu'on utilise une variable aléatoire  $X$ , on est intéressé à connaître des probabilités cumulées reliées à cette variable, c'est-à-dire des probabilités de la forme  $P(X \leq x)$ ,  $P(a \leq X \leq b)$  ou  $P(X \geq x)$ . Ainsi, pour la variable aléatoire  $X$  de l'exemple DS 1 exprimant le nombre d'acheteurs de la marque d'auto  $P$  dans un échantillon aléatoire de 3 clients, on sera intéressé à connaître, par exemple, la probabilité  $P(X \leq 2)$ , c'est-à-dire la probabilité qu'il y ait au plus 2 acheteurs de  $P$  dans l'échantillon. Une telle probabilité est déterminée **en cumulant** (soit, en additionnant) les probabilités de tous les événements correspondant à  $X \leq 2$ . Au lieu de répéter ce genre de calcul à partir de la fonction masse de probabilité à chaque fois qu'on en a besoin, on le fait une fois et on enregistre les résultats dans une fonction de distribution des **probabilités cumulées** que l'on va appeler **fonction de répartition**.

**Fonction de répartition.** On appelle fonction de répartition de  $X$  une fonction notée  $F$  à valeurs dans  $[0, 1]$  qui, à chaque valeur  $x$  prise par  $X$ , fait correspondre la probabilité que  $X$  prenne une valeur «plus petite ou égale à  $x$ », c'est-à-dire :

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i) = \sum_{x_i \leq x} f(x_i)$$

**Exemple DS 6.** La fonction de répartition  $F$  de la v.a.  $X$  exprimant le nombre d'acheteurs de  $P$  dans un échantillon aléatoire de 3 clients (exemple DS 1) est définie comme suit :

$$F(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1/8, & \text{si } 0 \leq x < 1 \\ 4/8, & \text{si } 1 \leq x < 2 \\ 7/8, & \text{si } 2 \leq x < 3 \\ 1, & \text{si } x \geq 3 \end{cases}$$

Cette fonction a comme graphe la courbe en escalier de la figure DS 3.



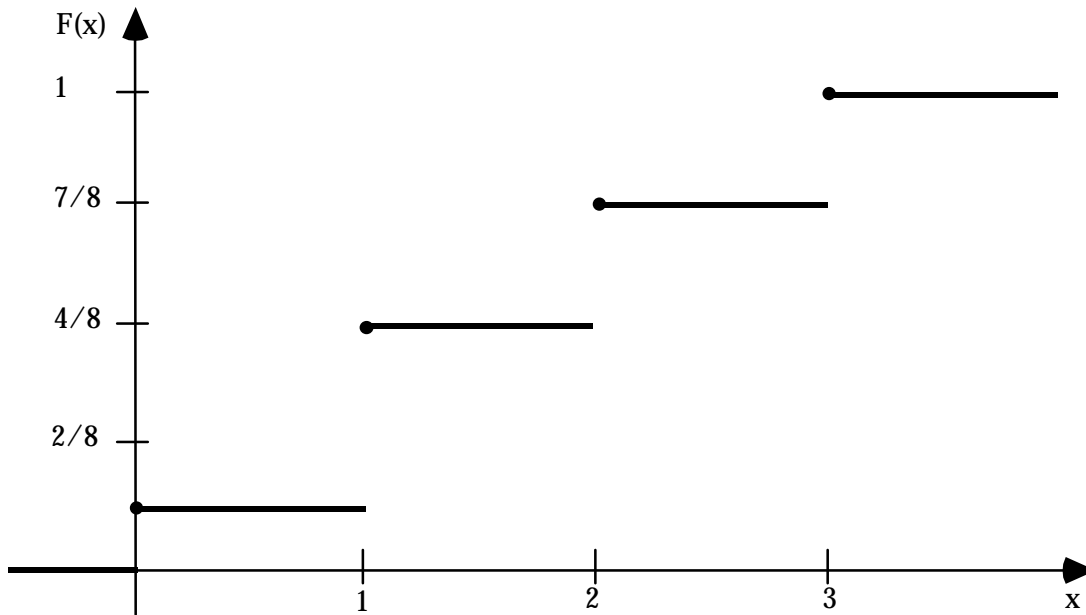


Figure DS 3 Fonction de répartition

**Propriétés de la fonction de répartition F**

- 1° F(x) est définie pour tout x réel,
- 2°  $0 \leq F(x) \leq 1$ , pour tout x réel,
- 3°  $F(-\infty) = 0, F(+\infty) = 1$ ,
- 4° si  $y \geq x, F(y) \geq F(x)$ .

Autrement dit, quelle que soit la fonction de répartition F, 1° elle est toujours définie partout sur les réels, 2° ses valeurs F(x) sont toujours entre 0 et 1, 3° elle tend vers zéro lorsque x tend vers l’infini négatif et elle tend vers 1 lorsque x tend vers l’infini positif, 4° elle ne décroît jamais.

**Calcul de la probabilité que X appartienne à un intervalle réel à l'aide de la fonction de répartition**

Soit a et b deux nombres réels; on a :

- 1°  $P(X \text{ est dans } [a, b]) = P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a^-)$  ●—————●
- 2°  $P(X \text{ est dans } [a, b[) = P(a \leq X < b) = P(X < b) - P(X < a) = F(b^-) - F(a^-)$  ●—————
- 3°  $P(X \text{ est dans } (a, b]) = P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$  —————●
- 4°  $P(X \text{ est dans } (a, b[) = P(a < X < b) = P(X < b) - P(X \leq a) = F(b^-) - F(a)$  —————

où  $F(b^-) = P(X < b)$  et  $F(a^-) = P(X < a)$ .

**Exemple DS 7.** On reprend ici les données de l'exemple DS 6 concernant la v.a.  $X$  désignant le nombre d'acheteurs de la marque P dans un échantillon aléatoire de 3 clients. À partir de la fonction de répartition définie dans l'exemple DS 6, on peut calculer :

- a)  $P(X \leq 2)$ , la probabilité qu'il y ait au plus 2 acheteurs de P dans l'échantillon : on a  $P(X \leq 2) = F(2) = 7/8$ ;
- b)  $P(X < 2)$ , la probabilité qu'il y ait moins de 2 acheteurs de P dans l'échantillon : on a  $P(X < 2) = F(2^-) = 4/8$ ;
- c)  $P(X > 2)$ , la probabilité qu'il y ait plus de 2 acheteurs de P dans l'échantillon : on a  $P(X > 2) = 1 - P(X \leq 2) = 1 - 7/8 = 1/8$ ;
- d)  $P(X \geq 2)$ , la probabilité qu'il y ait au moins 2 acheteurs de P dans l'échantillon : on a  $P(X \geq 2) = 1 - P(X < 2) = 1 - F(2^-) = 1 - 4/8 = 4/8$ ;
- e)  $P(1 < X \leq 3)$ , la probabilité qu'il y ait plus de 1 acheteur de P dans l'échantillon, mais au plus 3 : on a  $P(1 < X \leq 3) = P(X \leq 3) - P(X \leq 1) = F(3) - F(1) = 1 - 4/8 = 4/8$ ;
- f)  $P(1 < X < 3)$ , la probabilité qu'il y ait plus de 1 acheteur de P dans l'échantillon, mais moins de 3 : on a  $P(1 < X < 3) = P(X < 3) - P(X \leq 1) = F(3^-) - F(1) = 7/8 - 4/8 = 3/8$ .
- g)  $P(1 \leq X \leq 3)$ , la probabilité qu'il y a de 1 à 3 acheteurs de P dans l'échantillon : on a  $P(1 \leq X \leq 3) = P(X \leq 3) - P(X < 1) = F(3) - F(1^-) = 1 - 1/8$ .
- h)  $P(1 \leq X < 3)$ , la probabilité qu'il y a au moins un acheteur mais moins de 3 acheteurs de P dans l'échantillon : on a  $P(1 \leq X < 3) = F(3^-) - F(1^-) = 7/8 - 1/8 = 6/8$ .

### 3. Variable aléatoire continue

On a vu précédemment qu'une variable aléatoire est dite continue si  $X$  est susceptible de prendre n'importe quelle valeur d'un intervalle ou de plusieurs intervalles réels.

Comme dans les cas d'une v.a. discrète, on veut définir pour la v.a. continue  $X$  une fonction analogue à la fonction masse de probabilité. Lorsque  $X$  est *discrète*, par l'intermédiaire de la masse de probabilité, à chaque valeur  $x$  prise par  $X$ , on fait correspondre une probabilité positive  $f(x) = P(X = x)$ , c'est-à-dire la probabilité que  $X$  prenne cette valeur  $x$ . Pour une v.a. continue  $X$ , si, à chaque valeur  $x$  prise par  $X$  on fait correspondre  $f(x) = P(X = x) > 0$ , on obtient alors

$$(*) \quad \sum_{x \in X} P(X = x) = \infty$$

car  $X$  contient un nombre infini de valeurs. La relation (\*) contredit l'une des caractéristiques de la probabilité. C'est donc dire que, pour définir la notion de distribution de probabilité dans le cas d'une v.a. continue, on ne peut pas utiliser exactement la même approche que celle utilisée dans le cas d'une v.a. discrète.

Ce dont on a besoin c'est la probabilité que la variable aléatoire tombe dans un intervalle donné. On va introduire la notion de densité de probabilité d'une variable aléatoire continue par l'intermédiaire d'un exemple.

**Exemple DS 8.** On fait une étude sur la durée de vie d'un certain type d'ampoule électrique produite par une compagnie. Cette durée de vie varie d'une ampoule à l'autre. On tire au hasard une ampoule de ce type dans un lot produit par cette compagnie et l'on observe sa durée de vie. L'ensemble fondamental associé à cette expérience est  $\{\omega : \omega \text{ est une ampoule dans le lot}\}$ . La durée de vie de l'ampoule choisie peut être considérée comme une variable aléatoire, c'est-à-dire comme une fonction

$$X : \longrightarrow \mathbf{R}$$

$$\omega : \longrightarrow x = X(\omega)$$

où  $x = X(\omega)$  = la durée de vie de l'ampoule choisie (en années). D'après les données du passé, on sait qu'une ampoule de ce type ne dure jamais plus de 2 ans.  $X$  peut donc prendre a priori n'importe quelle valeur réelle entre 0 et 2 ans. D'après les données historiques, on sait aussi que :

- la durée de vie la plus probable se situe autour de 1 an ;
- il est très improbable que la durée de vie de l'ampoule choisie se situe près de 0 ou près de 2 ans ;
- une durée de vie d'environ 9 mois (0.75 année) est beaucoup plus probable qu'une durée de vie d'environ 3 mois (0.25 année), et une durée de vie d'environ 15 mois (1.25 année) est beaucoup plus probable qu'une durée de vie d'environ 21 mois (1.75 année).

La v.a.  $X$  prend toute valeur entre 0 et 2. Dans un premier temps, au lieu de parler de la probabilité que  $X$  prenne une valeur  $x$ , on va parler de la probabilité que  $X$  prenne une «valeur voisine de  $x$ », soit une valeur dans « un intervalle autour de  $x$  ». Par exemple, la probabilité que  $X$  prenne une «valeur voisine de 3 mois» (0.25 année) pourrait être prise égale à la probabilité que  $X$  prenne une valeur entre 0 et 6 mois. Ainsi, on diviserait l'intervalle  $[0,2]$  en 4 sous-intervalles :  $[0, .5]$ ,  $(.5, 1]$ ,  $(1, 1.5]$ ,  $(1.5, 2]$ . Par la suite, à partir de données historiques, on estimerait la probabilité que la durée de vie de l'ampoule choisie se situe dans l'un de ces 4 sous-

intervalles. On obtiendrait ainsi une distribution de probabilité  $f$  qui pourrait, par exemple, être celle que l'on a représentée par l'histogramme tracé à la figure DS 4.

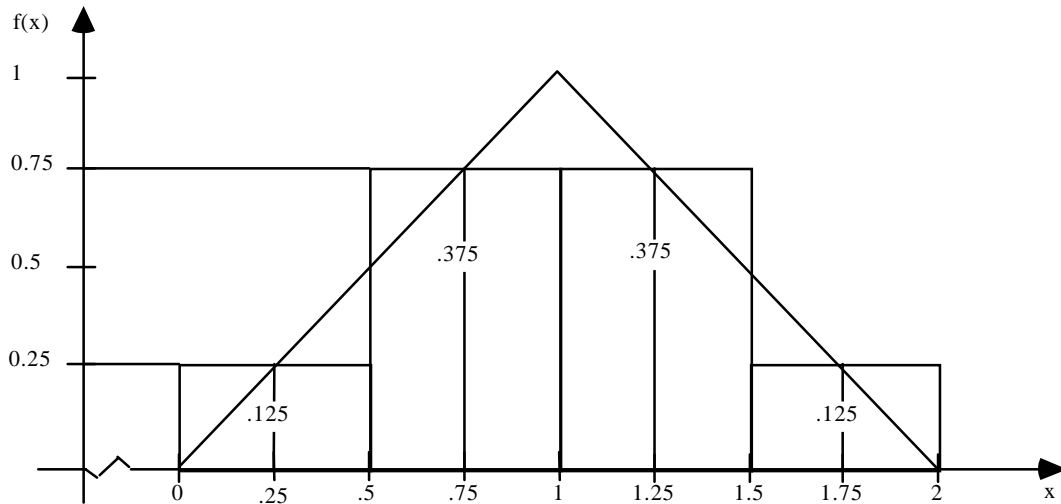


Figure DS 4

On aurait une idée plus précise de la probabilité que  $X$  prenne une «valeur voisine de  $x$ » en divisant l'intervalle  $[0,2]$  en sous-intervalles plus courts que ceux choisis dans la figure DS 4. Ainsi, en divisant l'intervalle  $[0,2]$  en 8 sous-intervalles ( $[0, .25]$ ,  $(.25, .5]$ ,  $(.5, .75]$ ,  $(.75, 1]$ ,  $(1, 1.25]$ ,  $(1.25, 1.5]$ ,  $(1.5, 1.75]$ ,  $(1.75, 2]$ ), la distribution de probabilité de la v.a.  $X$  pourrait être, par exemple celle représentée par l'histogramme de la figure DS 5.

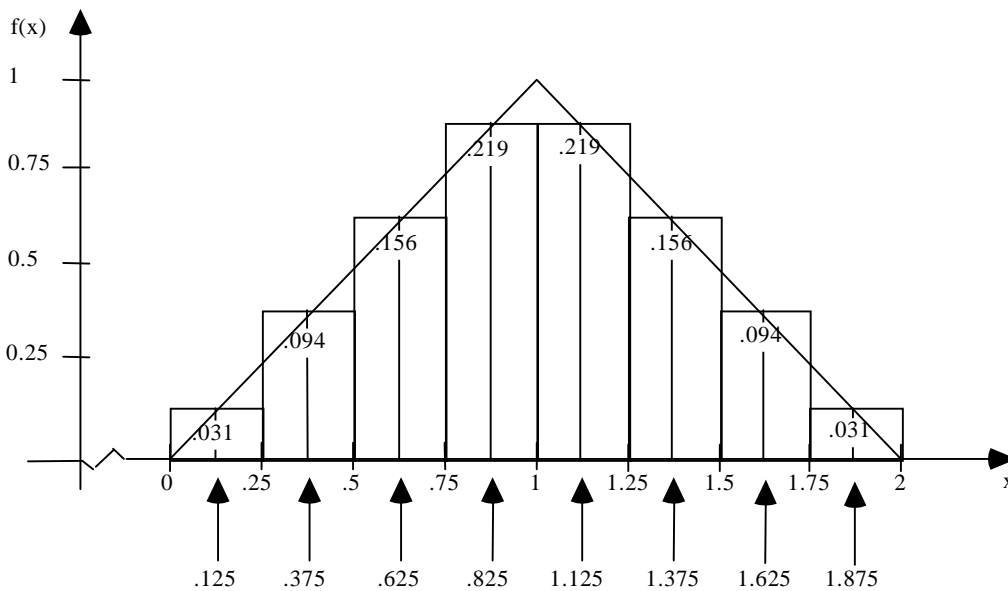
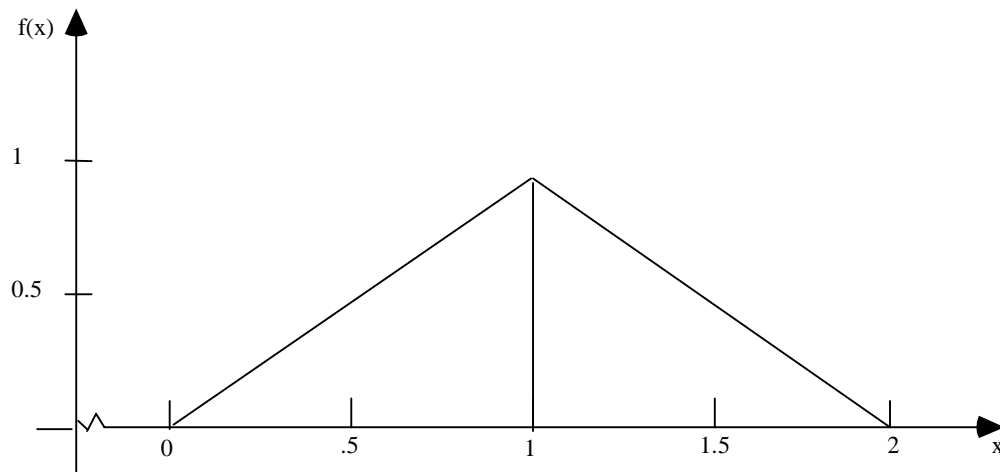


Figure DS 5

On pourrait continuer à subdiviser l'intervalle  $[0,2]$  en sous-intervalles de plus en plus courts de façon à avoir une idée de plus en plus précise de la probabilité que  $X$  prenne une « valeur voisine de  $x$  ». À la limite, lorsque la longueur des sous-intervalles tendrait vers zéro, on représenterait la distribution de probabilité de  $X$  non plus par un histogramme mais par la courbe continue tracée à la figure DS 6.



**Figure DS 6**

La fonction  $f$  représentée par la courbe tracée à la figure DS 6 est définie mathématiquement comme suit :

$$f(x) = \begin{cases} 0, & \text{si } x < 0 \\ x, & \text{si } 0 \leq x < 1 \\ 2 - x, & \text{si } 1 \leq x < 2 \\ 0, & \text{si } x \geq 2 \end{cases}$$

Cette fonction  $f$  sera appelée **densité de probabilité** de la v.a. continue  $X$ . Le terme “densité” est utilisé pour mettre l'accent sur le fait que cette fonction permet de distribuer la probabilité d'une façon dense sur tous les points d'un intervalle réel. La courbe de la fonction  $f$  représentée dans la figure DS 6 peut être interprétée intuitivement comme la forme limite de l'histogramme.

Pour une masse de probabilité, c'est la notion de point qui est fondamentale, alors que, dans le cas d'une densité de probabilité, c'est la notion **d'intervalle** qui est à la base de la définition de  $f$ , car les probabilités sont maintenant assignées à des intervalles de valeurs de  $X$  plutôt qu'à des valeurs particulières. L'aire totale sous la courbe  $f$  doit être égale à 1 tout comme l'aire totale de

l'ensemble des rectangles d'un histogramme, et l'aire sous la courbe  $f$  entre deux valeurs  $a$  et  $b$  représente la probabilité que  $X$  prenne une valeur dans l'intervalle  $[a, b]$ .

À partir de cet exemple particulier, on peut dégager les principales caractéristiques de la densité (ou distribution) de probabilité et de la fonction de répartition d'une variable aléatoire continue  $X$ . Supposons que l'ensemble des valeurs prises par  $X$  soit  $X = [a, b]$ ,  $a, b$  sont des réels. On veut définir une fonction  $f$  qui, aux valeurs  $x$  dans  $[a, b]$ , fera correspondre les probabilités que  $X$  prenne ces valeurs, de telle sorte qu'à toute valeur  $x$  dans  $[a, b]$  corresponde une valeur  $f(x)$ .

Mais étant donné que  $X$  prend une infinité non dénombrable de valeurs sur  $[a, b]$  et qu'il est impossible d'assigner une probabilité positive à chacune de ces valeurs sans aller à l'encontre des caractéristiques de base des probabilités, on fait les hypothèses suivantes :

- 1° la probabilité assignée à une *valeur particulière*  $x$  est zéro, c'est-à-dire  $P(X = x) = 0$ , pour tout  $X$  dans  $[a, b]$ ;
- 2° la probabilité assignée à **un intervalle** de la forme  $(x, x + d]$  où  $d$  est un accroissement positif de  $X$  est définie positive, c'est-à-dire  $P(x < X \leq x + d) > 0$ , pour tout  $x, x+d$  dans  $[a, b]$ .

Comme  $P(x < X \leq x + d)$  dépend de  $x$  et de  $d$ , on pose :

$$P(x < X \leq x + d) = f(x) * d$$

où  $f(x)$  est une certaine fonction de  $x$  et  $d$  est un accroissement positif de  $X$ . Il s'agit en effet d'une sorte d'interpolation linéaire où  $P(x < X \leq x + d)$  est l'accumulation de la probabilité entre  $x$  et  $x+d$  alors que  $f(x)$  est un taux moyen d'accumulation dans cet intervalle de largeur  $d$ . Ainsi, au taux  $f(x)$  sur un intervalle de largeur  $d$ , on accumule  $f(x)*d$ . Vu que l'intervalle  $d$  est un intervalle de la variable  $x$ , on peut dire que  $f(x)$  représente la densité (de la probabilité qui est accumulée) et donc,  $f(x)*d$  est le «poids» de la probabilité contenue dans l'intervalle.

Puisqu'il y a une probabilité positive définie sur chacun des sous-intervalles  $(x, x + d]$  de  $[a, b]$ , la définition suivante a du sens :

**Fonction de répartition.** On appelle fonction de répartition de la v.a. aléatoire continue  $X$  et l'on note  $F$  la fonction à valeurs dans  $[0, 1]$ , par la relation

$$F(x) = P(X \leq x)$$

**Densité de probabilité.** On appelle densité de probabilité de la v.a. continue  $X$  une fonction notée  $f$  à valeurs dans les réels dont l'accumulation de probabilité entre  $X = x$  et  $X = x+d$ , où  $x < x+d$  est donnée par  $F(x+d) - F(x) = f(x)*d$ .

On désigne aussi la densité de probabilité par les expressions plus générales **distribution de probabilité** ou **loi de probabilité**.

Puisque l'aire sous la courbe de densité *jusqu'à* la valeur  $X=x$  est l'accumulation de probabilité jusqu'à  $x$ ,  $F(x)$ , la valeur de la fonction de répartition au point  $x$  est égale à l'**aire sous la courbe de la densité** de  $-\infty$  *jusqu'à*  $x$ .

**Exemple DS 9.** À l'exemple DS 8, on s'est intéressé à la durée de vie  $X$  des ampoules électriques d'un certain type produites par une compagnie. On avait obtenu comme densité de  $X$  la fonction suivante :

$$f(x) = \begin{cases} 0, & \text{si } x < 0 \\ x, & \text{si } 0 \leq x < 1 \\ 2 - x, & \text{si } 1 \leq x < 2 \\ 0, & \text{si } x \geq 2 \end{cases}$$

On peut maintenant calculer les valeurs  $F(x)$  de la fonction de répartition qui correspond à cette densité  $f$ . Si la méthode pour calculer une telle fonction est le calcul intégral, dans ce cas on peut utiliser une méthode géométrique.

Si  $x$  est plus petit que zéro, on n'a aucune valeur de  $X < 0$  et ainsi, aucune densité, aucune accumulation, soit  $F(x) = P(X \leq x) = 0$ .

Si l'on prend maintenant  $x$  tel que  $0 \leq x < 1$ , on a l'accumulation de 0 jusqu'à  $x$ . Il y a deux cas à considérer :  $0 \leq x \leq 1$  et  $1 \leq x \leq 2$ . Dans les deux cas, il s'agit de trouver l'aire sous la courbe triangulaire telle qu'affichée à la Figure DS 6.

Cas :  $0 \leq x \leq 1$

La région sous la courbe jusqu'à  $x$  est un triangle dont l'aire est : largeur\*hauteur/2. Or la largeur est égale à  $x$ . La hauteur est déterminée par la pente de la courbe qui est ici une ligne droite. Si la pente de la courbe est  $m$ , ça veut dire que la courbe croît par  $m$  unités pour chaque unité de  $x$ . À la Figure DS 6, on constate que la courbe monte à 1 lorsque  $x$  croît de 0 à 1. Ainsi, la pente dans ce cas est  $1/1 = 1$ . Donc, la hauteur de la courbe est  $m*x = 1*x = x$ . D'où on tire la conclusion que l'aire sous la courbe jusqu'à  $x$  est  $x*x/2$  ou, autrement,  $x^2/2$ . On remarque que

l'aire sous la courbe de densité de probabilité jusqu'à  $x = 1$  est égale à 0,5, soit la moitié de la probabilité totale sous la courbe.

Cas :  $1 \leq x \leq 2$

La région sous la courbe dans ce cas est le triangle à gauche (de  $x = 0$  jusqu'à  $x = 1$ ) et la partie de  $x = 1$  jusqu'à la valeur de  $x < 2$ . Par symétrie, l'aire sous la courbe de  $x = 1$  jusqu'à  $x = 2$  est égal à 0,5. Or, la partie sous la courbe de  $x = 1$  jusqu'à un  $x < 2$  est tout simplement la partie restant du triangle à droite, c'est-à-dire, son aire est égale à  $0,5 -$  (l'aire sous la courbe de  $x$  jusqu'à 2). Or, la largeur de ce dernier triangle à droite est  $(2-x)$ . Sa hauteur est également  $(2-x)$  par un raisonnement analogue à celui de gauche. Alors, dans ce cas pour un  $x > 1$  :

l'aire sous la courbe de  $x$  jusqu'à 2 =  $(2-x) \cdot (2-x) / 2 = (4 - 4x + x^2) / 2 = 2 - 2x + x^2 / 2$ .

Enfin, l'aire sous la courbe de 0 jusqu'à  $x$  dans ce cas est :

$$0,5 + (0,5 - (2 - 2x + x^2 / 2)) = -1 + 2x - x^2 / 2.$$

On remarque que si  $x = 2$ , l'aire sous la courbe jusqu'à  $x$  est  $-1 + 2 \cdot 2 - 2^2 / 2 = 1$  comme il se doit pour une probabilité.

En combinant les résultats des deux cas, on détermine que la fonction de répartition de  $X$  est définie comme suit :

$$F(x) = \begin{cases} 0, & \text{si } x < 0 \\ x^2 / 2, & \text{si } 0 \leq x < 1 \\ -x^2 / 2 + 2x - 1, & \text{si } 1 \leq x < 2 \\ 1, & \text{si } x \geq 2 \end{cases}$$

La densité de probabilité et la fonction de répartition d'une v.a. continue possèdent des propriétés analogues à celles que l'on a mentionnées pour la masse de probabilité et la répartition d'une v.a. discrète.

### Calcul de la probabilité que $X$ appartienne à un intervalle

Puisque, pour une v.a. continue  $X$ , on a  $P(X = x) = 0$  pour tout  $x$ , alors il s'ensuit que, pour tous nombres réels  $a$  et  $b$  tels que  $a < b$ , on a

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = F(b) - F(a).$$



**Exemple DS 10.** Les boîtes de détergent en poudre remplies par une machine pèsent en principe 2 kilogrammes; cependant, en pratique, leur poids est plutôt une variable aléatoire. Dans le passé, on n'a jamais trouvé de boîte de ce détergent pesant moins de 1.75 kg ou pesant plus de 2.25 kg. De plus, on va supposer (même si ce n'est probablement pas réaliste) que tous les poids entre 1.75 et 2.25 kg sont également probables. Déterminer a) la densité de X, b) sa répartition, c) la probabilité qu'une de ces boîtes de détergent pèse moins de 2 kilogrammes et la probabilité qu'une de ces boîtes pèse entre 1.95 et 2.05 kilogrammes.

**Solution :** a) La v.a. X (poids) est continue et sa densité aura la forme suivante :

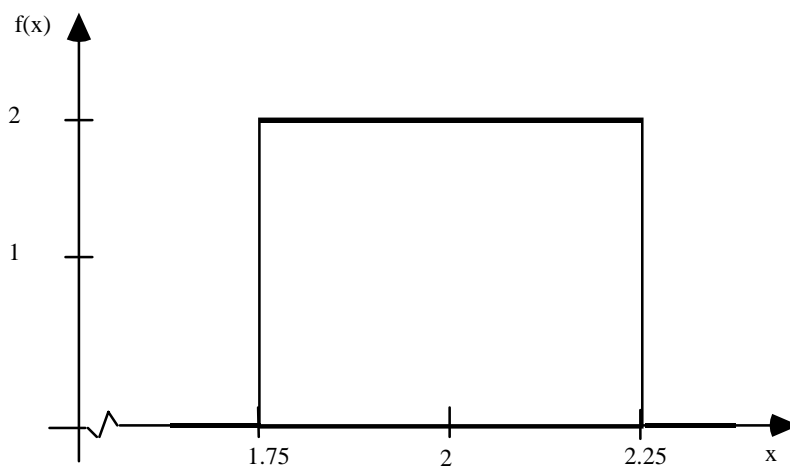
$$f(x) = \begin{cases} 0, & \text{si } x < 1.75 \\ k, & \text{si } 1.75 \leq x \leq 2.25 \\ 0, & \text{si } x > 2.25 \end{cases}$$

où k est une constante réelle qu'il faut déterminer. En utilisant l'idée que k est le taux d'accumulation de la probabilité dans l'intervalle de 1.75 à 2.25, on déduit que l'accumulation totale est k fois la largeur de l'intervalle. Ainsi,  $k \cdot (2.25 - 1.75) = k \cdot (0.5) = 1$ , d'où  $k = 2$ .

La densité de X est donc

$$f(x) = \begin{cases} 2, & \text{si } 1.75 \leq x \leq 2.25 \\ 0, & \text{ailleurs} \end{cases}$$

et elle est représentée par le graphe de la figure DS 7. Cette distribution est appelée la **distribution uniforme**.

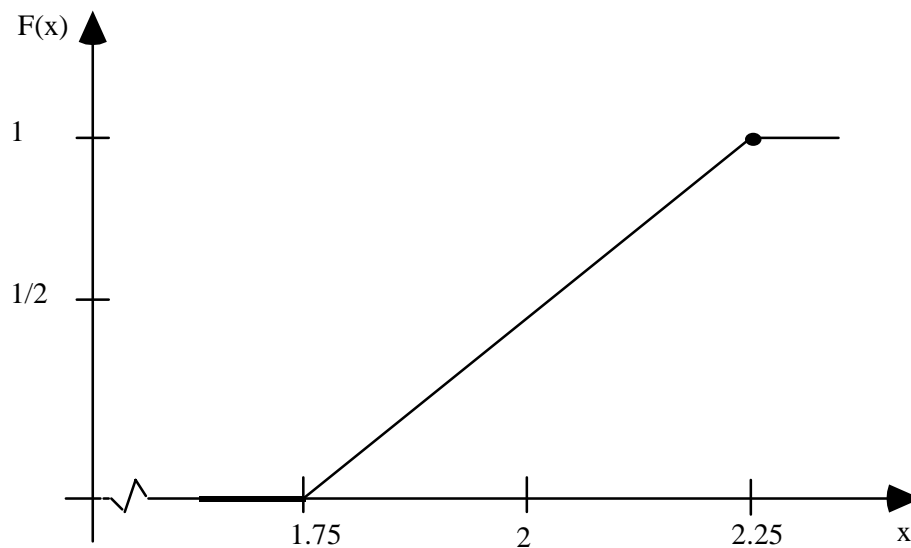


**Figure DS 7**

b) À partir de la densité de  $X$ , on peut déterminer la **fonction de répartition**  $F(x)$ . On sait que l'accumulation de probabilité dans ce cas se fait à partir de  $x = 1,75$  et au rythme de  $k = 2$ . Ainsi, la probabilité accumulée jusqu'à  $x$  est  $k$ \*la largeur de l'intervalle entre  $1,75$  et  $x$ , soit  $2*(x - 1,75)$  ou  $2x - 3,5$ . Ainsi, la fonction de répartition est :

$$F(x) = \begin{cases} 0, & \text{si } x < 1,75, \\ 2x - 3,5, & \text{si } 1,75 \leq x < 2,25, \\ 1, & \text{si } x \geq 2,25, \end{cases}$$

fonction qui est représentée par le graphe de la figure DS 8. À la figure DS 8, on voit clairement l'accumulation de la probabilité au rythme constant de  $k = 2$

**Figure DS 8**

À partir de la densité ou de la répartition, on peut calculer diverses probabilités reliées à  $X$  :

1° la probabilité qu'une boîte pèse moins de 2 kilogrammes

$$= P(X < 2) = P(X \leq 2) = F(2) = 2 \times 2 - 3,5 = 0,5$$

= surface sous la densité entre 1.75 et 2 (figure DS 9).

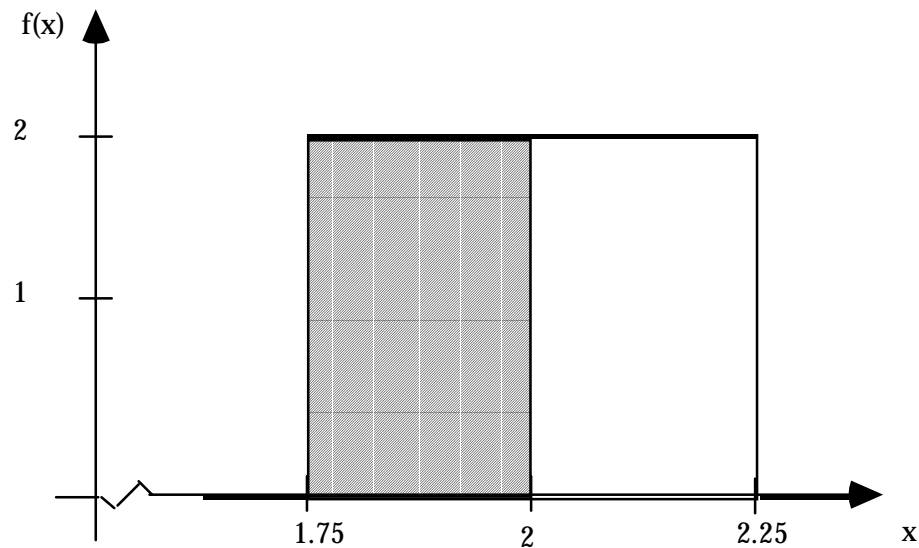


Figure DS 9

2° la probabilité qu'une boîte pèse entre 1.95 et 2.05 kilos =

$$P(1.95 \leq X \leq 2.05) = P(X \leq 2.05) - P(X < 1.95) = P(X \leq 2.05) - P(X \leq 1.95) = F(2.05) - F(1.95) = \text{surface sous la densité entre 1.95 et 2.05} = 0.2.$$

## 2.4 Caractéristiques d'une variable aléatoire

Une variable aléatoire est complètement décrite par l'ensemble de ses valeurs et par sa distribution de probabilité. Cependant, il arrive souvent que l'on n'a pas besoin de toute l'information que fournit la distribution; on peut alors se contenter d'une description plus sommaire de la v.a. exprimée par l'intermédiaire de certaines caractéristiques qui résument les propriétés de cette v.a. Ces caractéristiques sont des valeurs numériques que l'on obtient à partir de la distribution de probabilité à l'aide de certaines règles. On distingue plusieurs catégories de caractéristiques pour une v.a.; les deux principales sont les caractéristiques de position et les caractéristiques de dispersion.

### Caractéristiques de position

Les caractéristiques de position indiquent globalement où se situent les valeurs prises par une v.a. et en particulier autour de quel point se regroupent principalement ces valeurs; les principales caractéristiques de position sont:

- (a) l'espérance mathématique;
- (b) les fractiles (dont la médiane);
- (c) le mode.

**(a) L'espérance mathématique**

L'espérance mathématique est la caractéristique de position qui est de loin la plus utilisée. Historiquement, la notion d'espérance mathématique a toujours été associée plus ou moins étroitement aux jeux de hasard. Cependant, la notion d'espérance mathématique ne s'applique pas seulement pour les variables aléatoires associées aux jeux de hasard.

**Exemple DS 11.** Un vendeur d'appareils ménagers a compilé, au cours des 100 derniers jours de vente, le nombre de cuisinières vendues chaque jour. En se basant sur ces données, il se croit justifié d'assigner à la variable aléatoire  $X$  donnant «le nombre  $x$  de cuisinières vendues au cours d'une journée» la distribution de probabilité suivante :

$x$	1	2	3	4	5	6
$P(x) = f(x)$	.04	.06	.25	.35	.19	.11

Dans l'espoir de pouvoir calculer ultérieurement le profit qu'il peut espérer réaliser grâce à la vente de cuisinières, ce vendeur cherche d'abord à connaître le nombre moyen de cuisinières qu'il peut espérer vendre quotidiennement sur une longue période.

**Solution :** Il y a 4 % des jours où il vend 1 cuisinière, 6 % des jours où il vend 2 cuisinières, 25 % des jours où il vend 3 cuisinières, etc. En conséquence, sur une longue période, le nombre moyen de cuisinières qu'il peut espérer vendre quotidiennement et que l'on appellera l'espérance de  $X$  est donné par  $1 (.04) + 2 (.06) + 3 (.25) + 4 (.35) + 5 (.19) + 6 (.11)$ , ce qui est égal à 3.92 cuisinières. Ce nombre 3.92 n'est pas une valeur actuelle : il n'y aura pas de journée où ce vendeur vendra effectivement 3.92 cuisinières. Cependant, si le vendeur envisage ses ventes de cuisinières sur une très longue période, il peut utiliser 3.92 comme représentant ses ventes quotidiennes de cuisinières. Par exemple sur 100 jours, les ventes moyennes seraient de  $100 \cdot 3,92 = 392$ .

D'une façon générale, on a

**Espérance mathématique.** On appelle espérance mathématique de la variable aléatoire  $X$  de distribution  $f(x)$  et l'on désigne par  $E(X)$  ou  $(\mu)$  le nombre réel défini par la relation

$$E(X) = \sum_i x_i f(x_i), \text{ si } X \text{ est discrète.}$$

Souvent on utilise l'expression «moyenne de X» pour désigner l'espérance mathématique; cependant, la notion d'espérance mathématique diffère de celle de moyenne arithmétique. Ainsi, pour une v.a. discrète qui prend n valeurs, la moyenne arithmétique est simplement la somme de ces valeurs divisée par n. L'espérance mathématique de cette v.a. est la somme de ces n valeurs, chacune de ces valeurs étant pondérée par la probabilité qui lui est attachée. Pour une v.a. discrète qui prend n valeurs distinctes avec une probabilité  $1/n$ , l'espérance mathématique et la moyenne arithmétique coïncident.

Dans le cas d'une v.a. continue, la sommation qu'implique le concept d'espérance mathématique est naturellement remplacée par une opération mathématique où on effectue la même sorte d'accumulation des valeurs de la v.a. pondérées par la densité de probabilité correspondante. (C'est l'intégration telle que définie par le calcul intégral.) Cependant, en pratique, dans le cas d'une v.a. continue on a recours à des méthodes de calcul numérique approximatives. Ainsi, des chiffriers comprennent des fonctions analogues à la fonction de répartition de la distribution triangulaire de l'exemple DS 8.

Étant donné l'importance que revêt la notion d'espérance mathématique en probabilités et statistique, on va maintenant en souligner certaines propriétés particulières.

Soit a et b des constantes et X une variable aléatoire; alors, on a

$$1^{\circ} \quad E(a) = a,$$

$$2^{\circ} \quad E(a X) = a E(X),$$

$$3^{\circ} \quad E(a X \pm b) = a E(X) \pm b.$$

**Exemple DS 12 :** L'espérance mathématique des ventes de cuisinières de l'exemple DS 11 est 3,92. Si le prix d'une cuisinière est 600 \$, alors l'espérance mathématique des revenus est  $600 \times 3,92 = 2\,352$  \$.

**b) Les fractiles**

Intuitivement, un fractile d'une v.a. est une valeur numérique en bas de laquelle repose une fraction donnée de la probabilité définie par la distribution de cette variable.

**Fractile d'ordre  $\alpha$ .** On appelle fractile d'ordre  $\alpha$ ,  $0 \leq \alpha \leq 1$ , de la v.a.  $X$  le nombre réel noté  $x_\alpha$  qui satisfait à la fois aux deux inégalités

$$P(X \leq x_\alpha) \geq \alpha \text{ et } P(X \geq x_\alpha) \geq 1 - \alpha$$

**Exemple DS 13.** Soit la v.a. continue de  $X$  exprimant le poids des boîtes de détergent remplies par une machine (exemple DS 10); sa densité est

$$f(x) = \begin{cases} 2, & \text{si } 1.75 \leq x \leq 2.25 \\ 0, & \text{ailleurs} \end{cases}$$

On pourrait, par exemple, être intéressé à connaître le poids  $x$  tel que la probabilité qu'une boîte pèse au plus ce poids  $x$  serait au moins égale à 0.8. Ce poids qui sera noté  $x_{0.8}$  est le fractile d'ordre  $\alpha = 0.8$  de la v.a.  $X$  et il doit satisfaire aux 2 inégalités

$$P(X \leq x_{0.8}) \geq 0.8 \text{ et } P(X \geq x_{0.8}) \geq 1 - 0.8 = 0.2.$$

La fonction de répartition de  $X$  est

$$F(x) = \begin{cases} 0, & \text{si } x < 1.75; \\ 2x - 3.5, & \text{si } 1.75 \leq x < 2.25; \\ 1, & \text{si } x \geq 2.25. \end{cases}$$

Comme  $X$  est une v.a. continue, ces deux inégalités se réduisent alors à l'égalité  $P(X \leq x_{0.8}) = 0.8$ , c'est-à-dire  $F(x_{0.8}) = 0.8$ . Par l'intermédiaire de la définition de  $F(x)$ , on obtient alors  $2x_{0.8} - 3.5 = 0.8$ , d'où le fractile d'ordre 0.8 de  $X$  est  $x_{0.8} = 2.15$ . Voir la représentation de  $x_{0.8}$  dans la figure DS 10.

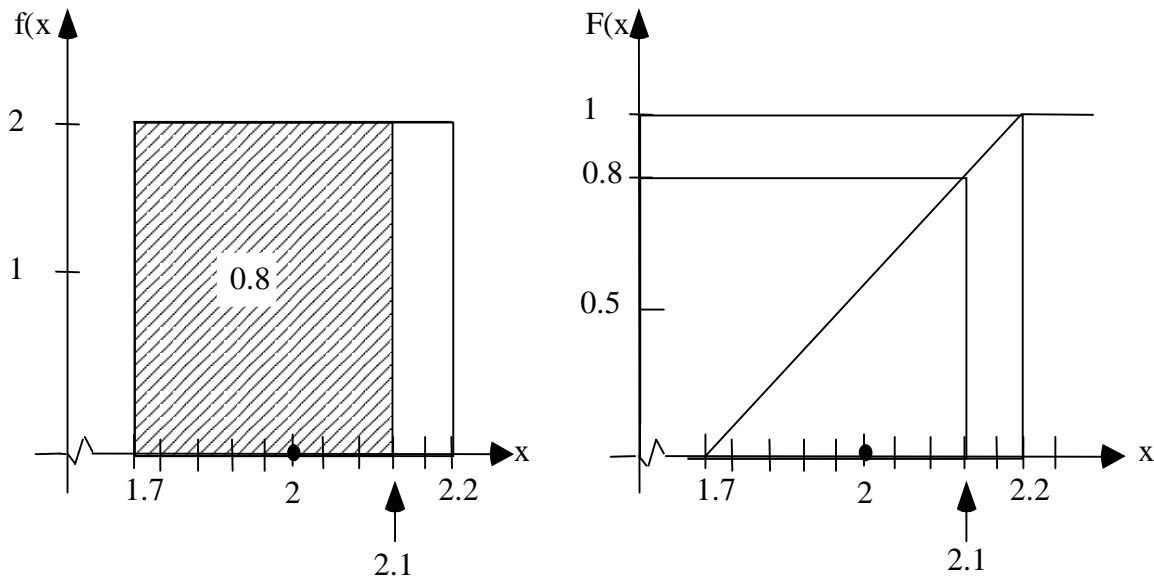


Figure DS 10

**La médiane.** Le fractile d'ordre  $\alpha = 0,5$  est connu sous le nom de **médiane**. Pour la v.a.  $X$  de l'exemple **DS 13**, la médiane est  $x_{0,5} = 2$ . Pour la v.a. discrète  $X$  de l'exemple **DS 6** tout nombre  $x$  de l'intervalle  $[1, 2)$  peut être pris comme médiane (et donc, dans ce cas, la médiane n'est pas unique).

**Exemple DS 14 :** Dans l'exemple DS 11, la fonction de répartition des ventes de cuisinières est donnée dans le tableau suivant :

$x$	1	2	3	4	5	6
$F(x) = P(X \leq x)$	0,04	0,1	0,35	0,7	0,89	1,0

La médiane  $x_{0,5}$  est telle que  $P(X \leq x_{0,5}) \geq 0,5$  ce qui donne  $x = 4$ . De plus, la médiane doit satisfaire la condition  $P(X \geq x_{0,5}) \geq 1 - 0,5 = 0,5$ . Alors, c'est  $x = 4$  qui répond à cette condition. Donc, la médiane est 4.

**Remarque :** Un fractile comme celui d'ordre  $\alpha=0,10$  dans cet exemple est un cas spécial parce que la v.a. est discrète. La fonction de répartition  $F(x)=0,1$  pour  $2 \leq x < 3$ . Ainsi, toute valeur dans l'intervalle  $[2, 3)$  répond aux deux conditions  $P(X \leq x_{0,1}) \geq 0,1$  et  $P(X \geq x_{0,1}) \geq 1 - 0,1 = 0,9$ . Ceci est un exemple général correspondant au cas de la médiane de l'Exemple **DS 6** cité ci-haut.

**(c) Le mode**

Le mode d'une v.a.  $X$  de distribution  $f(x)$  est la valeur de  $X$  pour laquelle  $f(x)$  atteint son maximum, c'est-à-dire la valeur prise par  $X$  qui est la plus probable. Si  $f(x)$  a plusieurs maxima, chacun de ces maxima peut être considéré comme le mode de  $X$ .

**Exemple DS 15.** Dans l'exemple DS 11, on a considéré la v.a.  $X$  donnant le nombre de cuisinières vendues quotidiennement par un détaillant d'appareils ménagers. La distribution de probabilité de  $X$  était

$x$	1	2	3	4	5	6
$P(X = x) = f(x)$	0,04	0,06	0,25	0,35	0,19	0,11

Dans ce cas, le mode de  $X$  (qui représente le nombre de ventes le plus probable) est 4.

**Caractéristiques de dispersion**

Souvent les caractéristiques de position ne suffisent pas pour faire ressortir toutes les propriétés essentielles d'une variable aléatoire. Considérons l'exemple suivant :

**Exemple DS 16.** Un joueur a le choix de participer à l'un ou l'autre des deux jeux suivants : un premier jeu où son gain  $X$  peut prendre les valeurs  $-2, -1, 1, 2$ , avec probabilité  $1/4$  pour chacune de ces valeurs, et un deuxième jeu où son gain  $Y$  peut prendre les valeurs  $-10, -5, 5, 10$  aussi avec probabilité  $1/4$  pour chacune de ces valeurs. Quel jeu doit-il choisir ? Normalement, le joueur choisira le jeu qui lui permettra de maximiser son gain espéré. Il calcule donc le gain espéré de chaque jeu, ce qui lui donne  $E(X) = E(Y) = 0$ . Ainsi, l'espérance mathématique ne suffit pas à caractériser suffisamment bien les deux v.a.  $X$  et  $Y$ . S'il veut faire un choix, le joueur peut en plus chercher à minimiser son risque. Dans cette perspective, le deuxième jeu semble présenter beaucoup plus de risque que le premier car le joueur peut perdre dans ce deuxième jeu jusqu'à 10 (dollars). Ce risque traduit en quelque sorte ici une plus grande dispersion pour la v.a.  $Y$  que pour la v.a.  $X$ .

Ainsi, pour une v.a., en plus des caractéristiques de position, on introduit des caractéristiques de dispersion qui indiquent globalement dans quelle mesure les valeurs prises par la v.a. ont tendance à être plus ou moins dispersées autour d'une certaine valeur et en particulier autour de



l'espérance mathématique. On ne considère ici comme caractéristique de dispersion que la variance.

### La variance

La variance est l'espérance mathématique de  $(X - \mu)^2$  soit  $E[(X - \mu)^2]$ . C'est, parmi les mesures de dispersion, celle qui est de loin la plus utilisée; c'est pourquoi on introduit la définition suivante :

**Variance.** On appelle variance de la v.a.  $X$  de distribution  $f(x)$  que l'on désigne par  $\sigma^2(X)$ , le moment centré d'ordre 2 de  $X$  défini par

$$\sigma^2(X) = E[(X - \mu)^2] = \sum (x_i - \mu)^2 f(x_i) \text{ pour une v.a. discrète.}$$

En plus d'utiliser  $\sigma^2(X)$  (ou simplement  $\sigma^2$ ) pour désigner la variance, on utilise aussi  $\text{Var}(X)$  ou parfois  $\mu_2(X)$ .

De plus, on appelle **écart type** (ou **déviat ion standard**) de la v.a.  $X$  le nombre réel non négatif noté  $\sigma(X)$  (ou simplement  $\sigma$ ) défini comme la racine carrée positive de la variance de  $X$ , c'est-à-dire

$$\sigma(X) = \sqrt{\text{VAR}(X)}$$

L'écart type de la v.a.  $X$  représente en quelque sorte la déviation moyenne de  $X$  autour de sa moyenne  $\mu$ .

**Exemple DS 17.** Dans l'exemple DS 16, les v.a.  $X$  et  $Y$  représentant le gain du joueur pour deux jeux différents avaient toutes deux une espérance mathématique égale à zéro. Calculons maintenant les variances de  $X$  et de  $Y$ .

$$\begin{aligned} \sigma^2(X) &= \text{Var}(X) = E[(X - \mu)^2] = E[(X - 0)^2] \\ &= \sum_i (x_i - 0)^2 f(x_i) \\ &= (-2)^2 1/4 + (-1)^2 1/4 + (1)^2 1/4 + (2)^2 1/4 = 2.5 \end{aligned}$$

De même,

$$\begin{aligned}\sigma^2(Y) &= \text{Var}(Y) = E[(Y - \mu)^2] = E[(Y - 0)^2] \\ &= \sum_i (y_i - 0)^2 f(y_i) \\ &= (-10)^2 1/4 + (-5)^2 1/4 + (5)^2 1/4 + (10)^2 1/4 = 62.5\end{aligned}$$

Ainsi, même si  $E(X) = E(Y)$ , on a  $\text{Var}(X) = 2,5$  et  $\text{Var}(Y) = 62,5$  et donc  $\sigma(X) \approx 1,58$  et  $\sigma(Y) \approx 7,91$ . La grande différence qu'il y a entre  $\text{Var}(X)$  et  $\text{Var}(Y)$  confirme le fait que la v.a.  $Y$  est en moyenne beaucoup plus dispersée autour de  $E(Y)$  que la v.a.  $X$  n'est dispersée autour de sa propre moyenne  $E(X)$ . Si cette plus ou moins grande dispersion de la variable gain dans un jeu traduit un risque plus ou moins grand, il est évident que le deuxième jeu (de gain  $Y$ ) présente un risque de loin supérieur à celui du premier jeu (de gain  $X$ ).

Le calcul de la variance d'une v.a.  $X$  est souvent simplifié si l'on utilise la relation suivante :

$$\text{Var}(X) = E(X^2) - [E(X)]^2 \text{ où } E(X^2) = \sum_i x_i^2 f(x_i)$$

**Propriétés de la variance.** Soit  $a$  et  $b$  des constantes et  $X$  une v.a. quelconque, alors

- 1°  $\text{Var}(a) = 0$  ;
- 2°  $\text{Var}(aX) = a^2 \cdot \text{Var}(X)$  ;
- 3°  $\text{Var}(aX \pm b) = a^2 \cdot \text{Var}(X)$ .

**Exemple DS 18.** Dans l'exemple DS 11, on a considéré la v.a.  $X$  donnant le nombre de cuisinières vendues quotidiennement par un détaillant d'appareils ménagers. La distribution de probabilité de  $X$  était

$x$	1	2	3	4	5	6
$P(X = x) = f(x)$	0,04	0,06	0,25	0,35	0,19	0,11

Sachant que ce détaillant réalise un profit de 100 \$ par cuisinière vendue, déterminer le profit qu'il peut espérer réaliser quotidiennement grâce à la vente de cuisinières ainsi que la variance et l'écart type de ce profit quotidien.

**Solution :** Dans l'exemple DS 11, on a calculé le nombre de cuisinières qu'il peut espérer vendre quotidiennement : ce nombre est donné par

$$E(X) = \mu = \sum_i x_i f(x_i) = 3.92$$

Le profit Y quotidien réalisé grâce à la vente de cuisinières est une variable aléatoire définie par  $Y = 100 X$ . En conséquence, en utilisant les propriétés de l'espérance mathématique E, le profit que ce détaillant peut espérer réaliser quotidiennement grâce à la vente de cuisinières est donné par

$$\begin{aligned} E(Y) &= E(100 X) = 100 E(X) = 100 (3.92) \\ &= 392,00 \$ \end{aligned}$$

Le détaillant sera sans doute intéressé à calculer la variabilité moyenne de ce profit quotidien tiré de la vente de cuisinières. Cette variabilité est donnée par la variance (ou par l'écart type) de la variable Y. Pour calculer la variance de Y, on calcule d'abord la variance de X. Si l'on utilise la définition de la variance, on obtient

$$\begin{aligned} \text{Var}(X) &= \sigma^2(X) = \sum_{i=1}^6 (x_i - \mu)^2 f(x_i) \\ &= (1 - 3.92)^2 (.04) + (2 - 3.92)^2 (.06) + (3 - 3.92)^2 (.25) \\ &\quad + (4 - 3.92)^2 (.35) + (5 - 3.92)^2 (.19) + (6 - 3.92)^2 (.11) \\ &\approx 1.47 \end{aligned}$$

Plus simplement, on peut utiliser la formule  $\text{Var}(X) = E(X^2) - [E(X)]^2$ . Le calcul de  $E(X^2)$  donne

$$E(X^2) = \sum_{i=1}^6 x_i^2 f(x_i) = 1^2 (.04) + 2^2 (.06) + 3^2 (.25) + 4^2 (.35) + 5^2 (.19) + 6^2 (.11) = 16.84$$

d'où on a de nouveau  $\text{Var}(X) = 16.84 - 15.37 = 1.47$

Par la suite, l'écart type est donné par

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{1.47} = 1.21$$

Comme l'indique l'écart type de 1.21, le nombre de cuisinières vendues est en moyenne assez près de 3.92. En utilisant l'une des propriétés de la variance, on peut alors calculer la variance et l'écart type du profit quotidien Y réalisé par ce détaillant grâce à la vente de cuisinière; on a

$$\begin{aligned}\text{Var}(Y) &= \text{Var}(100 X) = (100)^2 \text{Var}(X) \\ &= 10\,000 (1.47) = 14\,700\end{aligned}$$

d'où l'écart type de Y est donné par

$$\sigma(Y) = \sqrt{14\,700} = 121,24 \$$$

En conclusion, grâce à la vente de cuisinières, ce détaillant peut donc espérer réaliser un profit quotidien moyen de 392 \$ et il peut espérer que ce profit quotidien variera en “moyenne” de 121,24 \$ autour de 392 \$.

Pour compléter votre apprentissage de la notion de variables aléatoires de distributions de probabilité, nous vous invitons à essayer de résoudre les exercices du module DS qui se trouvent en annexe. En cas de difficulté, vous pouvez consulter les solutions de ces exercices qui se trouvent en annexe ou communiquer avec le responsable du cours ou son adjoint pour plus d'explications.