

# PFA: documentation

---

## Table des matières

Contexte	1
Attendus	1
<b>I Intégration</b>	<b>1</b>
1 Calculs d'intégrales par primitives	2
2 Intégration numérique : formules de quadrature	3
2.1 Les premières formules de quadrature . . . . .	3
2.2 Construction d'une formule de quadrature : méthode générale . . . . .	6
2.3 Des formules de quadratures particulières . . . . .	7
2.4 Étude empirique de l'erreur . . . . .	8
<b>II Intégrales et probabilités</b>	<b>9</b>
3 Intégrales généralisées	9
3.1 Discontinuités avec limites à gauche et à droite . . . . .	10
3.2 Intégrales impropres . . . . .	11
4 Fonction de répartition et densité	12
4.1 Définition et exemples . . . . .	12
4.2 Opérations sur les variables aléatoires . . . . .	14
4.3 Espérance et variance d'une variable à densité . . . . .	15
<b>III Finance et assurance</b>	<b>15</b>
5 Finance	15
5.1 Définitions des «options» . . . . .	16
5.2 Loi du prix $S_T$ du sous-jacent à l'échéance $T$ . . . . .	16
5.3 Prix d'un call . . . . .	17
5.4 Prix d'un put . . . . .	18
6 Assurance	18
6.1 Un modèle . . . . .	19
6.2 Loi de $S$ . . . . .	20

---

## Contexte

La SAE «Probabilités, Finance et Assurance» vous demande de développer des fonctions informatiques destinées aux domaines de la finance et de l'assurance. Ces développements font appel à plusieurs notions mathématiques.

Ce document décrit, d'une part les notions mathématiques dont vous allez avoir besoin, d'autre part les modèles probabilistes que vous utiliserez pour concevoir vos fonctions informatiques.

Les domaines de la finance et de l'assurance font appel à des calculs de probabilités. Contrairement aux cas que vous avez étudiés jusqu'à présent, les variables prennent leurs valeurs dans des ensembles infinis plus vastes que celui des entiers naturels. En général, le domaine de leurs valeurs possibles couvre tout un intervalle de  $\mathbb{R}$ , voire  $\mathbb{R}$  tout entier. Dans de telles situations, les calculs de probabilités, d'espérances et de variances reviennent à des calculs d'intégrales.

On est alors souvent confronté à la situation suivante : devoir calculer une intégrale pour laquelle aucune des méthodes classiques (primitive, intégration par parties, changement de variable) ne permet d'obtenir de résultat. On construit alors une approximation numérique de la valeur de l'intégrale. Une bonne moitié du projet consistera pour vous à programmer et mettre en œuvre ces approximations, puis à analyser leurs erreurs.

Une seconde partie de ce document se consacre aux applications en probabilités. Enfin, une troisième partie se focalise sur les applications dans les domaines de la finance et de l'assurance que vous allez traiter.

## Attendus

La SAE PFA entre dans le cadre de la démarche compétences. Les deux compétences mises en jeu sont «Produire» et «Concevoir».

### 1. Produire :

- Produire des programmes informatiques de premier accès, impliquant une interaction avec l'environnement, la gestion des dépendances et la livraison d'un exécutable.
- Sélectionner des structures de données hiérarchiques de manière à répondre efficacement à un problème posé.

### 2. Concevoir :

- Collecter des données dans l'objectif de se familiariser avec un sujet, pour en déterminer les contours.
- Tester une solution existante en variant ses paramètres et en observant son comportement.

Vous avez accès sur moodle à la grille critériée de la SAE, qui formalise la façon dont vous serez évalué. Le but de ce document est de vous donner la compréhension mathématique suffisante pour écrire vos programmes, autant du point de vue théorique que des points vues finance et assurance.

## Première partie

# Intégration

Une intégrale est souvent vue, à tort, comme une primitive. En réalité, une intégrale n'est pas une primitive, ni même une fonction. C'est un nombre réel : si  $f$  désigne une fonction continue sur un intervalle  $[a, b] \subset \mathbb{R}$ , alors

$$I = \int_a^b f(x) \, dx$$

désigne l'aire comprise entre l'axe  $(O, t)$  et le graphe de la fonction  $f$ . Dans cette partie, nous examinons les différentes façons de calculer une intégrale :

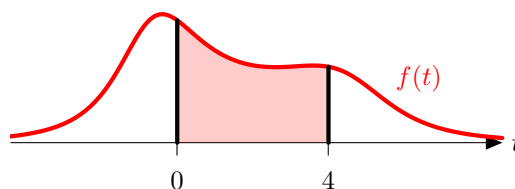


FIGURE 1 – Représentation géométrique d'une intégrale.  $\int_0^4 f(t) dt$  est l'aire comprise entre l'axe  $(O, t)$ , le graphe de la fonction et les droites verticales d'équations  $t = 0$  et  $t = 4$ , représentée en rouge sur le dessin. Cette aire est comptée positivement quand  $f(t) \geq 0$  (comme ci-dessus) et négativement quand  $f(t) \leq 0$ .

1. Intégration par la détermination d'une primitive. C'est la méthode la plus usuelle, celle qu'on vous a demandé d'appliquer. Nous incluons dans ces méthodes l'intégration par parties, qui sont en fait des méthodes perfectionnées pour trouver une primitive de la fonction intégrée  $f$ .
2. Intégration numérique : quand on ne trouve pas de primitives à la fonction  $f$ , quand ni l'intégration par parties ni aucun changement de variable ne permettent d'en trouver, on peut chercher une valeur approchée de l'intégrale. L'intégration numérique regroupe les méthodes d'approximation.

## 1 Calculs d'intégrales par primitives

Dans cette section, nous revenons sur le lien entre intégrale et primitive. Par exemple, si on se donne une fonction  $f$  continue sur  $\mathbb{R}$ , l'expression

$$I = \int_0^4 f(t) dt$$

ne désigne **pas** une primitive de  $f$ . En fait, cette expression ne définit même pas une fonction : elle définit simplement un nombre réel qui est l'aire comprise entre l'axe  $(O, t)$ , le graphe de la fonction et les droites verticales d'équations  $t = 0$  et  $t = 4$  (figure 1).

En revanche, pour toute valeur  $x \in \mathbb{R}$ , on peut définir  $\int_0^x f(t) dt$  : c'est un nombre réel qui est l'aire comprise entre l'axe  $(O, t)$ , le graphe de la fonction et les droites verticales d'équations  $t = 0$  et  $t = x$ . Cela permet de définir la fonction

$$F : \begin{cases} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \int_0^x f(t) dt \end{cases}$$

Le théorème suivant fait le lien entre intégrale et primitive.

**Théorème 1.1.** *La fonction  $F$  est dérivable en tout  $x \in \mathbb{R}$ . De plus, pour tout  $x \in \mathbb{R}$ ,  $F'(x) = f(x)$ .*

Nous ne présentons pas ici la démonstration de ce théorème, mais il faut retenir qu'elle se base sur la définition de la dérivée :

$$F'(x) = \lim_{y \rightarrow x} \frac{F(y) - F(x)}{y - x}$$

Le théorème dit donc que pour tout  $x \in \mathbb{R}$ , cette limite existe et vaut  $f(x)$ .

En d'autres termes, ce théorème dit que  $F$  est une primitive de  $f$ . Si on revient à l'intégrale initiale entre 0 et 4, on obtient

$$\begin{aligned} \int_0^4 f(t) dt &= F(4) && \text{par définition de } F \\ &= F(4) - F(0) && \text{car } F(0) = 0 \end{aligned}$$

Si maintenant on considère une autre primitive  $G$  de  $f$ , alors  $G$  est de la forme  $F + C$  où  $C$  est une constante. Ainsi,

$$G(4) - G(0) = (F(4) + C) - (F(0) + C) = F(4) - F(0) = \int_0^4 f(t) dt$$

Cela explique la méthode usuelle de calcul d'un intégrale : trouver une primitive de  $f$  et calculer la différence des valeurs de la primitive aux deux bornes de l'intégrale. C'est la méthode que vous avez toujours utilisée.

Mais pour réaliser le projet PFA il faut comprendre ceci : une intégrale ne se définit pas à partir de primitive, mais comme une aire. Chercher une primitive n'est qu'une technique pour la calculer. Et quand on ne trouve pas de primitive simple, il faut utiliser d'autres techniques. C'est ce que nous allons voir dans la suite de cette partie.

## 2 Intégration numérique : formules de quadrature

Il existe des fonctions  $f$  s'exprimant à partir des fonctions usuelles, telles qu'aucune primitive ne peut s'exprimer avec les fonctions usuelles. Par exemple, vous seriez bien embêté si vous deviez trouver une primitive de

$$f(x) = \frac{\ln(\sin(x) + 2) \times e^{\sqrt{x}}}{\cos^2(x) + e^x}$$

Et pourtant il y en a une, c'est la fonction  $F$  définie pour tout  $x \in \mathbb{R}$  par

$$F(x) = \int_0^x f(t) dt$$

Mais alors comment calculer par exemple  $F(2)$  ?

Cet exemple peut paraître bien inutile : on imagine mal une situation de la vie professionnelle où un ingénieur devrait intégrer cette fonction. Mais il y a d'autres exemples. Ainsi, en probabilités, nous verrons dans la partie suivante qu'il est essentiel d'intégrer la fonction  $f : t \mapsto e^{-\frac{t^2}{2}}$ . Or il n'existe pas de primitive qui s'exprime simplement à partir des fonctions usuelles.

Dans ce cas, on pose  $F(x) = \int_0^x f(t) dt$  et, si on doit calculer par exemple  $F(2)$ , on utilisera une méthode numérique pour approximer l'aire comprise entre l'axe  $(O, t)$ , le graphe de la fonction  $f$  et les droites d'équations  $t = 0$  et  $t = 2$ .

### 2.1 Les premières formules de quadrature

Dans toute cette section,  $[a, b]$  est un intervalle fermé de  $\mathbb{R}$  et  $f$  une fonction continue sur  $[a, b]$ . Les formules de quadratures sont des méthodes numériques pour approximer  $\int_a^b f(t) dt$ . Les premières d'entre elles (et vous allez programmer dans le projet PFA) sont les méthodes des rectangles et des trapèzes (figure 2). On vous les a expliquées au lycée, mais on ne vous a jamais demandé de les mettre en œuvre.

Ces méthodes consistent à :

- Partitionner  $[a, b]$  en  $N$  subdivisions  $[a_i, b_i]$  où, pour tout  $i \in \llbracket 0, N \rrbracket$ ,

$$a_i = a + i \times \frac{b-a}{N} \quad \text{et} \quad b_i = a + (i+1) \times \frac{b-a}{N}$$

Les subdivisions se suivent les uns après les autres et sont tous de mêmes longueurs :  $b_i - a_i = \frac{b-a}{N}$ .

- Sur chaque subdivision  $[a_i, b_i]$ , approcher  $\int_{a_i}^{b_i} f(t) dt$  par une «formule de quadrature»  $Q_{[a_i, b_i]}(f)$  :

$$\int_{a_i}^{b_i} f(t) dt \approx Q_{[a_i, b_i]}(f) = \int_{a_i}^{b_i} P(t) dt$$

où le polynôme  $P$  est :

Méthode 1 : le polynôme de degré 0 :  $P(t) = f(a_i)$ .

Méthode 2 : le polynôme de degré 0 :  $P(t) = f(b_i)$ .

Méthode 3 : le polynôme de degré 0 :  $P(t) = f\left(\frac{a_i+b_i}{2}\right)$ .

Méthode 4 : le polynôme  $P$  de degré 1 vérifiant  $P(a_i) = f(a_i)$  et  $P(b_i) = f(b_i)$ .

Voyons ce que donne  $Q_{[a_i, b_i]}(f) = \int_{a_i}^{b_i} P(t) dt$  avec chacune de ces quatre méthodes.

Méthode 1 :  $P$  est le polynôme constant  $P(t) = f(a_i)$ . Ainsi, l'aire comprise entre l'axe  $(Ot)$ , le graphe de  $P$  est les droites d'équations  $t = a_i$  et  $t = b_i$  est l'aire d'un rectangle de largeur  $b_i - a_i$  et de hauteur  $f(a_i)$ . Donc

$$Q_{[a_i, b_i]}(f) = \int_{a_i}^{b_i} P(t) dt = (b_i - a_i) \times f(a_i)$$

On appelle cette méthode «rectangle gauche» car la hauteur du rectangle est la valeur de  $f$  au point gauche de la subdivision  $[a_i, b_i]$ .

Méthode 2 :  $P$  est le polynôme constant  $P(t) = f(b_i)$ . Là aussi, l'aire comprise entre l'axe  $(Ot)$ , le graphe de  $P$  est les droites d'équations  $t = a_i$  et  $t = b_i$  est l'aire d'un rectangle. Celui-ci est de largeur  $b_i - a_i$  et de hauteur  $f(b_i)$ . Donc

$$Q_{[a_i, b_i]}(f) = \int_{a_i}^{b_i} P(t) dt = (b_i - a_i) \times f(b_i)$$

On appelle cette méthode «rectangle droite» car la hauteur du rectangle est la valeur de  $f$  au point droite de la subdivision  $[a_i, b_i]$ .

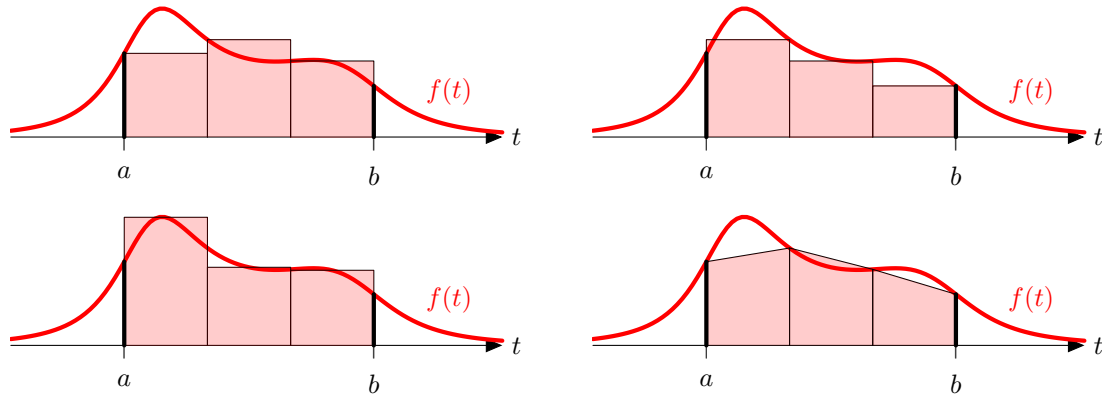


FIGURE 2 – Représentation géométrique des premières formules de quadrature. L'intervalle d'intégration  $[a, b]$  est découpé en  $N$  subdivisions  $[a_i, b_i]$  (sur les dessins,  $N = 3$ ). **En haut à gauche :** méthode «rectangle gauche». Sur chaque subdivision  $[a_i, b_i]$ , l'intégrale de  $f$  est approximée par l'intégrale de la fonction constante  $f(a_i)$ . **En haut à droite :** méthode «rectangle droite». Sur chaque subdivision  $[a_i, b_i]$ , l'intégrale de  $f$  est approximée par l'intégrale de la fonction constante  $f(b_i)$ . **En bas à gauche :** méthode «rectangle milieu». Sur chaque subdivision  $[a_i, b_i]$ , l'intégrale de  $f$  est approximée par l'intégrale de la fonction constante  $f\left(\frac{a_i+b_i}{2}\right)$ . **En bas à droite :** méthode «trapèze». Sur chaque subdivision  $[a_i, b_i]$ , l'intégrale de  $f$  est approximée par l'intégrale de la fonction polynôme de degré 1 qui coïncide avec  $f$  en  $a_i$  et en  $b_i$ .

Méthode 3 : là,  $P$  est le polynôme constant  $P(t) = f\left(\frac{a_i+b_i}{2}\right)$ . On a alors

$$Q_{[a_i,b_i]}(f) = \int_{a_i}^{b_i} P(t) dt = (b_i - a_i) \times f\left(\frac{a_i + b_i}{2}\right)$$

On appelle cette méthode «rectangle milieu» car la hauteur du rectangle est la valeur de  $f$  au point milieu de la subdivision  $[a_i, b_i]$ .

Méthode 4 :  $P$  est le polynôme de degré 1 vérifiant  $P(a_i) = f(a_i)$  et  $P(b_i) = f(b_i)$ . Pour le construire, nous allons utiliser la méthode des lagrangiens qui sera utile pour la suite. On pose :

$$L_0(t) = \frac{t - b_i}{a_i - b_i} \quad \text{et} \quad L_1(t) = \frac{t - a_i}{b_i - a_i}$$

Ce sont tous les deux des polynômes de degrés 1. De plus :

$$\begin{cases} L_0(a_i) = \frac{a_i - b_i}{a_i - b_i} = 1 & L_1(a_i) = \frac{a_i - a_i}{b_i - a_i} = 0 \\ L_0(b_i) = \frac{b_i - b_i}{a_i - b_i} = 0 & L_1(b_i) = \frac{b_i - a_i}{b_i - a_i} = 1 \end{cases} \quad \text{et}$$

Ainsi, si on pose  $P = f(a_i)L_0 + f(b_i)L_1$ , alors  $P$  est un polynôme de degré 1 qui vérifie :

$$P(a_i) = f(a_i) \underbrace{L_0(a_i)}_1 + f(b_i) \underbrace{L_1(a_i)}_0 = f(a_i) \quad \text{et} \quad P(b_i) = f(a_i) \underbrace{L_0(b_i)}_0 + f(b_i) \underbrace{L_1(b_i)}_1 = f(b_i)$$

Ce polynôme  $P$  est bien le polynôme recherché. Donc

$$Q_{[a_i,b_i]}(f) = \int_{a_i}^{b_i} P(t) dt = f(a_i) \int_{a_i}^{b_i} L_0(t) dt + f(b_i) \int_{a_i}^{b_i} L_1(t) dt$$

Il reste à calculer les intégrales des deux lagrangiens.

$$\begin{aligned} \int_{a_i}^{b_i} L_0(t) dt &= \frac{1}{a_i - b_i} \times \int_{a_i}^{b_i} (t - b_i) dt \\ &= \frac{1}{a_i - b_i} \times \left[ \frac{(t - b_i)^2}{2} \right]_{a_i}^{b_i} \\ &= \frac{1}{a_i - b_i} \times \left[ 0 - \frac{(a_i - b_i)^2}{2} \right] \\ &= \frac{b_i - a_i}{2} \end{aligned}$$

Et de même, un calcul analogue donne  $\int_{a_i}^{b_i} L_1(t) dt = \frac{b_i - a_i}{2}$ . Finalement,

$$Q_{[a_i,b_i]}(f) = (b_i - a_i) \times \left( \frac{1}{2} f(a_i) + \frac{1}{2} f(b_i) \right)$$

On appelle cette méthode «trapèze» car c'est le nom de la figure géométrique dont on calcule l'aire.

On tient là nos premières méthodes d'approximation numérique.

**Remarques.** 1. Ces quatre méthodes se mettent toutes sous la forme

$$Q_{[a_i,b_i]}(f) = (b_i - a_i) \times \sum_{k=0}^n w_k f(a_i + x_k(b_i - a_i))$$

Rectangle gauche :  $n = 0$ ,  $w_0 = 1$ ,  $x_0 = 0$ .

Rectangle droite :  $n = 0$ ,  $w_0 = 1$ ,  $x_0 = 1$ .

Rectangle milieu :  $n = 0$ ,  $w_0 = 1$ ,  $x_0 = \frac{1}{2}$ .

Trapèze :  $n = 1$ ,  $w_0 = w_1 = \frac{1}{2}$ ,  $x_0 = 0$ ,  $x_1 = 1$ .

2. Les autres méthodes que nous verrons ont toutes la même forme. En termes de programmation, on prévoira deux champs dans la classe `integrator` : un champs `xk` et un champs `wk`, qui seront deux vecteurs de  $n + 1$  nombres : les points  $x_0, \dots, x_n$  et les coefficients  $w_0, \dots, w_n$ .
3. Dans la suite du document, pour ne plus s'embêter avec les  $a_i$  et les  $b_i$ , on fera comme si  $a_i = 0$  et  $b_i = 1$ . En effet, si on définit  $g(x) = f(a_i + x(b_i - a_i))$ , alors le changement de variable  $t = a_i + (b_i - a_i)x$  donne :
  - $dt = (b_i - a_i) dx$
  - Les bornes pour  $x$  sont 0 (quand  $t = a_i$ ) et 1 (quand  $t = b_i$ ).
  - Donc

$$\int_{a_i}^{b_i} f(t) dt = (b_i - a_i) \int_0^1 f(a_i + x(b_i - a_i)) dx = (b_i - a_i) \int_0^1 g(x) dx$$

Or

$$Q_{[a_i, b_i]}(f) = (b_i - a_i) \times \sum_{k=0}^n w_k \underbrace{f(a_i + x_k(b_i - a_i))}_{g(x_k)} = (b_i - a_i) Q_{[0,1]}(g)$$

On notera  $Q(g)$  au lieu de  $Q_{[0,1]}(g)$ .

## 2.2 Construction d'une formule de quadrature : méthode générale

**Définition.** Soient  $\{x_0, \dots, x_n\} \subset [0, 1]$  un ensemble de  $n + 1$  points distincts dans  $[0, 1]$  et  $g$  une fonction continue sur  $[0, 1]$ . On appelle «formule de quadrature» toute formule de la forme

$$Q(g) = \sum_{k=0}^n w_k g(x_k)$$

où  $(w_0, \dots, w_n) \in \mathbb{R}^{n+1}$ . Une telle formule est destinée à approximer  $\int_0^1 g(x) dx$ .

**Remarque.** Dans la pratique, vous utiliserez une formule de quadrature pour estimer une intégrale sur  $[a_i, b_i]$  :

$$\int_{a_i}^{b_i} f(t) dt = (b_i - a_i) \times \int_0^1 g(x) dx \approx (b_i - a_i) \times Q(g)$$

où  $g$  est la fonction  $x \mapsto f(a_i + x(b_i - a_i))$ . Ainsi,

$$(b_i - a_i) \times Q(g) = (b_i - a_i) \times \sum_{k=0}^n w_k f(a_i + x_k(b_i - a_i))$$

### Construction par interpolation.

Soient  $\{x_0, \dots, x_n\} \subset [0, 1]$  un ensemble de  $n + 1$  points distincts dans  $[0, 1]$  et une fonction  $g$  continue sur  $[0, 1]$ . On se propose de déterminer les coefficient  $w_0, \dots, w_n$  intervenant dans la formule de quadrature. Voilà les étapes :

1. Il existe un unique polynôme de degré  $n$  vérifiant :

$$P(x_0) = g(x_0), \quad P(x_1) = g(x_1), \quad \dots, \quad P(x_n) = g(x_n)$$

Nous allons admettre son unicité.

2. Pour démontrer son existence, nous allons le construire à l'aide des polynômes lagrangiens :

$$\forall k \in \llbracket 0, n \rrbracket, \quad L_k(X) = \prod_{j \neq k} \frac{X - x_j}{x_k - x_j}$$

Ce sont tous des polynômes de degré  $n$ . De plus, pour tout  $(k, k') \in \llbracket 0, n \rrbracket^2$  :

$$L_k(x_{k'}) = \begin{cases} 1 & \text{si } k = k' \\ 0 & \text{si } k \neq k' \end{cases}$$

Par exemple, si  $k = 0 : L_0(X) = \frac{X - x_1}{x_0 - x_1} \times \frac{X - x_2}{x_0 - x_2} \times \cdots \times \frac{X - x_n}{x_0 - x_n}$  et

$$\left\{ \begin{array}{lll} L_0(x_0) &= \frac{x_0 - x_1}{x_0 - x_1} \times \frac{x_0 - x_2}{x_0 - x_2} \times \cdots \times \frac{x_0 - x_n}{x_0 - x_n} &= 1 \quad \text{car chaque terme du produit vaut 1} \\ L_0(x_1) &= \frac{x_1 - x_1}{x_0 - x_1} \times \frac{x_1 - x_2}{x_0 - x_2} \times \cdots \times \frac{x_1 - x_n}{x_0 - x_n} &= 0 \quad \text{car le premier terme du produit est nul} \\ L_0(x_2) &= \frac{x_2 - x_1}{x_0 - x_1} \times \frac{x_2 - x_2}{x_0 - x_2} \times \cdots \times \frac{x_2 - x_n}{x_0 - x_n} &= 0 \quad \text{car le second terme du produit est nul} \\ \vdots & & \vdots \\ L_0(x_n) &= \frac{x_n - x_1}{x_0 - x_1} \times \frac{x_n - x_2}{x_0 - x_2} \times \cdots \times \frac{x_n - x_n}{x_0 - x_n} &= 0 \quad \text{car le } n^{\text{ème}} \text{ terme du produit est nul} \end{array} \right.$$

Donc le polynôme  $P = g(x_0)L_0 + g(x_1)L_1 + \cdots + g(x_n)L_n$  convient. C'est bien un polynôme de degré  $n$  et, par exemple,

$$P(x_0) = g(x_0) \underbrace{L_0(x_0)}_1 + g(x_1) \underbrace{L_1(x_0)}_0 + \cdots + g(x_n) \underbrace{L_n(x_0)}_0 = g(x_0)$$

3. On pose alors  $Q(g) = \int_0^1 P(x) dx$ . On obtient :

$$Q(g) = g(x_0) \int_0^1 L_0(x) dx + g(x_1) \int_0^1 L_1(x) dx + \cdots + g(x_n) \int_0^1 L_n(x) dx$$

Cette formule a bien la forme  $Q(g) = \sum_{k=0}^n w_k g(x_k)$  avec :

$$\forall k \in \llbracket 0, n \rrbracket, \quad w_k = \int_0^1 L_k(x) dx$$

## 2.3 Des formules de quadratures particulières

Nous allons appliquer la méthode présentée ci-dessus dans quelques cas.

1. Formule d'interpolation aux nœuds  $x_0 = 0$ ,  $x_1 = \frac{1}{2}$  et  $x_2 = 1$  (formule de Simpson).

Les lagrangiens sont :

$$L_0(X) = \frac{X - \frac{1}{2}}{0 - \frac{1}{2}} \times \frac{X - 1}{0 - 1} = 2 \left( X - \frac{1}{2} \right) (X - 1), \quad L_1(X) = \frac{X - 0}{\frac{1}{2} - 0} \times \frac{X - 1}{\frac{1}{2} - 1} = -4X(X - 1)$$

$$\text{et} \quad L_2(X) = \frac{X - 0}{1 - 0} \times \frac{X - \frac{1}{2}}{1 - \frac{1}{2}} = 2X \left( X - \frac{1}{2} \right)$$

Les coefficients  $w_k$  sont donc

$$w_0 = \int_0^1 2 \left( x - \frac{1}{2} \right) (x - 1) dx \quad \text{on trouve } w_0 = \frac{1}{6}$$

$$w_1 = \int_0^1 -4x(x - 1) dx \quad \text{on trouve } w_1 = \frac{2}{3}$$

$$w_2 = \int_0^1 2x \left( x - \frac{1}{2} \right) dx \quad \text{on trouve } w_2 = \frac{1}{6}$$



Donc  $Q(g) = \frac{1}{6}g(0) + \frac{2}{3}g\left(\frac{1}{2}\right) + \frac{1}{6}g(1)$ .

Et quand on utilisera la méthode pour estimer une intégrale sur  $[a_i, b_i]$ , on aura

$$\int_{a_i}^{b_i} f(t) dt \approx (b_i - a_i)Q(g) = (b_i - a_i) \times \left( \frac{1}{6}f(a_i) + \frac{2}{3}f\left(\frac{a_i + b_i}{2}\right) + \frac{1}{6}f(b_i) \right)$$

2. Formule d'interpolation aux nœuds  $x_0 = \frac{1}{2} - \frac{1}{2\sqrt{3}}$  et  $x_1 = \frac{1}{2} + \frac{1}{2\sqrt{3}}$  (formule de Gauss-Legendre à 2 nœuds).

En se basant sur la même méthode, on trouve :

$$w_0 = w_1 = \frac{1}{2} \implies Q(g) = \frac{1}{2}g\left(\frac{1 - \frac{1}{\sqrt{3}}}{2}\right) + \frac{1}{2}g\left(\frac{1 + \frac{1}{\sqrt{3}}}{2}\right)$$

3. Formule d'interpolation aux nœuds  $x_0 = \frac{1}{2}\left(1 - \sqrt{\frac{3}{5}}\right)$ ,  $x_1 = \frac{1}{2}$  et  $x_2 = \frac{1}{2}\left(1 + \sqrt{\frac{3}{5}}\right)$  (formule de Gauss-Legendre à 3 nœuds).

On trouve ;

$$w_0 = \frac{5}{18}, w_1 = \frac{4}{9}, w_2 = \frac{5}{18} \implies Q(g) = \frac{5}{18}g\left(\frac{1}{2}\left(1 - \sqrt{\frac{3}{5}}\right)\right) + \frac{4}{9}g\left(\frac{1}{2}\right) + \frac{5}{18}g\left(\frac{1}{2}\left(1 + \sqrt{\frac{3}{5}}\right)\right)$$

## 2.4 Étude empirique de l'erreur

On considère une formule de quadrature  $Q(g) = \sum_{k=0}^n w_k g(x_k)$  pour approximer  $\int_0^1 g(x) dx$ .

Pour approcher  $\int_a^b f(t) dt$ , on va donc :

- Partitionner  $[a, b]$  en  $N$  subdivisions  $[a_i, b_i]$  :

$$\forall i \in \llbracket 0, N \rrbracket, \quad a_i = a + i \times \frac{b-a}{N} \quad \text{et} \quad b_i = a + (i+1) \times \frac{b-a}{N}$$

- Sur chaque subdivision  $[a_i, b_i]$ , faire l'approximation

$$\int_{a_i}^{b_i} f(t) dt \approx (b_i - a_i) \times \sum_{k=0}^n w_k f(a_i + x_k(b_i - a_i))$$

- Au final, on obtient l'approximation

$$\int_a^b f(t) dt \approx \sum_{i=0}^{N-1} \left( (b_i - a_i) \times \sum_{k=0}^n w_k f(a_i + x_k(b_i - a_i)) \right)$$

Une étude théorique de l'erreur d'approximation est possible. Nous ne l'aborderons pas, disons simplement que l'erreur de l'approximation tend vers 0 quand le nombre  $N$  de subdivisions tend vers  $+\infty$ , et que cette convergence est plus rapide avec certaines formules de quadratures que d'autres.

Nous pouvons néanmoins étudier l'erreur empiriquement et vous allez le faire dans le projet PFA. Pour cela, on peut se donner une fonction  $f$  dont on connaît une primitive  $F$  et un intervalle  $[a, b] \subset \mathbb{R}$ . Il est alors possible de comparer la vraie valeur

$$\int_a^b f(t) dt = F(b) - F(a)$$

avec les approximations obtenues par les différentes méthode citées plus haut.

On se propose de mettre en évidence, pour une intégrale et une formule de quadrature données, une relation de l'erreur d'estimation en fonction du nombre  $N$  de subdivisions, sous la forme :

$$|\text{Erreur}| \approx \frac{C}{N^\alpha} \quad \text{où } (\alpha, C) \in \mathbb{R}^2$$

Pour cela, on se donne plusieurs valeurs de  $N : N_1, N_2, \dots, N_p$ , et on observe comment l'erreur d'approximation  $|\text{Erreur}(N_1)|, |\text{Erreur}(N_2)|, \dots, |\text{Erreur}(N_p)|$  varie avec  $N$ .

Si cette erreur est approximativement de la forme  $\frac{C}{N^\alpha}$ , alors  $\ln(|\text{Erreur}|) \approx \ln(C) - \alpha \ln(N)$  : les observations  $(\ln(N), \ln(|\text{Erreur}|))$  seront approximativement alignées (figure 3). On peut estimer  $\ln(C)$  et  $\alpha$  comme les paramètres donnant la «meilleure droite», celle qui passe «au plus près» des observations.

En fait, on cherche à résoudre au mieux le système

$$S : \begin{cases} -\alpha \ln(N_1) + \beta &= \ln(|\text{Erreur}(N_1)|) \\ -\alpha \ln(N_2) + \beta &= \ln(|\text{Erreur}(N_2)|) \\ \vdots &\vdots \\ -\alpha \ln(N_p) + \beta &= \ln(|\text{Erreur}(N_p)|) \end{cases} \quad \text{avec } \beta = \ln(C)$$

d'inconnues  $\alpha$  et  $\beta$ . Il n'y a pas de solution exacte, mais on cherche la solution optimale, celle qui est la «moins mauvaise».

N'entrons pas dans le détail de la façon de trouver cette «meilleure droite». Il existe une formule mathématique pour le faire et nous la donnons ci-dessous sans démonstration. Ce système  $S$  s'écrit sous la forme :

$$S : \begin{cases} \alpha x_1 + \beta &= y_1 \\ \alpha x_2 + \beta &= y_2 \\ \vdots &\vdots \\ \alpha x_p + \beta &= y_p \end{cases} \quad \text{avec } \beta = \ln(C)$$

Dans le cas qui nous concerne, les  $x_i$  sont les  $-\ln(N_i)$  et les  $y_i$  sont les  $\ln(|\text{Erreur}(N_i)|)$ . Pour obtenir les valeurs optimale  $\alpha^*$  et  $\beta^*$  :

$$\left| \begin{array}{l} \text{On définit d'abord :} \\ \overline{X} = \frac{\sum_{i=1}^p x_i}{p} = \text{Moyenne des } x_i \quad \text{et} \quad \overline{Y} = \frac{\sum_{i=1}^p y_i}{p} = \text{Moyenne des } y_i \\ \text{Alors :} \\ \alpha^* = \frac{\sum_{i=1}^p (x_i - \overline{X})(y_i - \overline{Y})}{\sum_{i=1}^p (x_i - \overline{X})^2} \quad \text{et} \quad \beta^* = \overline{Y} - \alpha^* \overline{X} \end{array} \right.$$

## Deuxième partie

# Intégrales et probabilités

### 3 Intégrales généralisées

Dans tous les cas considérés jusqu'ici, quand on cherchait à calculer une intégrale  $\int_a^b f(t) dt$ , on a supposé que la fonction  $f$  était continue sur sur l'intervalle **fermé**  $[a, b]$ . Or on a parfois besoin de calculer des intégrales où ce n'est pas le cas.

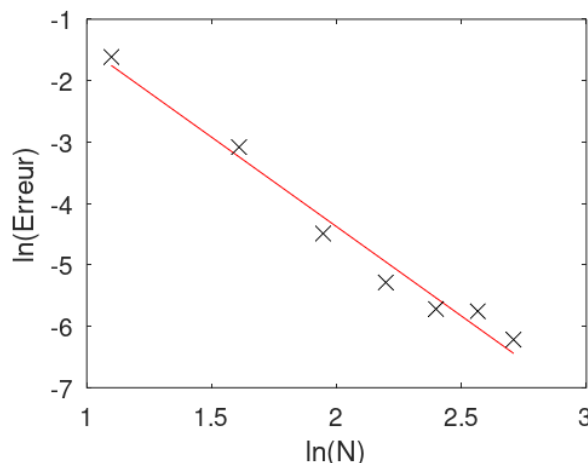


FIGURE 3 – erreur et nombre  $N$  de subdivisions en coordonnées logarithmiques. Ici,  $N$  varie de 3 à 15. On cherche à mettre en évidence une relation de la forme  $|\text{Erreur}| \approx \frac{C}{N^\alpha}$ , ainsi que les valeurs de  $C$  et  $\alpha$ . Si une telle relation est approximativement vérifiée, alors les points  $(\ln(N), \ln(|\text{Erreur}|))$  sont approximativement alignés :  $\ln(|\text{Erreur}|) \approx \ln(C) - \alpha \ln(N)$ .

### 3.1 Discontinuités avec limites à gauche et à droite

Un premier cas de figure est celui où, pour un nombre fini de valeurs  $t_1 < t_2 < \dots < t_K$  :

- la fonction  $f$  est discontinue en  $t_k$ .
- mais elle admet une limite finie à gauche et une limite finie à droite, qu'on notera

$$f(t_k^-) = \lim_{x \rightarrow t_k^-} f(x) \quad \text{et} \quad f(t_k^+) = \lim_{x \rightarrow t_k^+} f(x)$$

Dans ce cas, on définira l'intégrale de  $f$  sur  $[a, b]$  par :

$$\int_a^b f(t) dt = \int_a^{t_1} f(t) dt + \int_{t_1}^{t_2} f(t) dt + \dots + \int_{t_K}^b f(t) dt$$

Dans le calcul de chaque intégrale  $\int_{t_k}^{t_{k+1}} f(t) dt$ , on considère que  $f(t_k) = f(t_k^+)$  et  $f(t_{k+1}) = f(t_{k+1}^-)$ .

**Exemple.** Considérons la fonction  $f$  définie pour tout  $t \in \mathbb{R}$  par  $f(t) = \begin{cases} e^{-t} & \text{si } t > 0 \\ 0 & \text{sinon} \end{cases}$

Cette fonction n'est pas continue en 0 car

$$\lim_{x \rightarrow 0^-} f(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow 0^+} f(x) = 1$$

On peut néanmoins calculer son intégrale sur  $[-1, 2]$  :

$$\begin{aligned} \int_{-1}^2 f(t) dt &= \underbrace{\int_{-1}^0 f(t) dt}_{f(0)=0} + \underbrace{\int_0^2 f(t) dt}_{f(0)=1} \\ &= \int_{-1}^0 0 dt + \int_0^2 e^{-t} dt \\ &= 0 + [-e^{-t}]_0^2 \\ &= 1 - e^{-2} \end{aligned}$$

### 3.2 Intégrales impropres

**Définition.** Soient  $a \in \mathbb{R}$ ,  $b \in \mathbb{R} \cup \{+\infty\}$  et  $f$  une fonction définie sur  $[a, b[$ . L'intégrale  $\int_a^b f(t) dt$  est dite **impropre en  $b$**  si :

- la fonction  $f$  est continue sur  $[a, b[$ ,
- et on ne peut pas dire qu'elle est continue sur  $[a, b]$  : soit parce que  $b = +\infty$ , soit que  $b \in \mathbb{R}$  mais que  $f$  n'est pas définie (et n'a pas de limite finie à gauche) en  $b$ .

**Exemples .** 1. Les intégrales  $\int_0^{+\infty} e^{-t} dt$  et  $\int_0^{+\infty} \cos(t) dt$  sont impropres en  $+\infty$ . Les fonctions  $t \mapsto e^{-t}$  et  $t \mapsto \cos(t)$  sont continues sur  $[0, +\infty[$ .

2. L'intégrale  $\int_0^1 \frac{1}{\sqrt{1-t}} dt$  est impropre en 1. La fonction  $t \mapsto \frac{1}{\sqrt{1-t}}$  est continue sur  $[0, 1[$  mais pas en 1. De plus, elle n'admet pas de limite finie en 1.

**Définition.** Soit  $\int_a^b f(t) dt$  une intégrale impropre en  $b$ . On dit que cette intégrale **converge** si il existe une limite finie

$$\lim_{x \rightarrow b^-} \int_a^x f(t) dt$$

Sinon, on dit qu'elle **diverge**. Quand l'intégrale converge, on dit que la valeur de l'intégrale est égale à cette limite.

**Exemples .** 1. Pour tout  $x \in [0, +\infty[$ ,  $\int_0^x e^{-t} dt = [-e^{-t}]_0^x = -e^{-x} + 1 \xrightarrow{x \rightarrow +\infty} 1$ , donc  $\int_0^{+\infty} e^{-t} dt$  converge et vaut 1.

2. Pour tout  $x \in [0, +\infty[$ ,  $\int_0^x \cos(t) dt = [-\sin(t)]_0^x = -\sin(x)$  qui n'a pas de limite en  $+\infty$ . Donc  $\int_0^{+\infty} \cos(t) dt$  diverge.

3. Pour tout  $x \in [0, 1[$ ,  $\int_0^x \frac{1}{\sqrt{1-t}} dt = [-2\sqrt{1-t}]_0^x = -2\sqrt{1-x} + 2 \xrightarrow{x \rightarrow 1} 2$ . Ainsi,  $\int_0^1 \frac{1}{\sqrt{1-t}} dt$  converge et vaut 2.

On définit de façon analogue une intégrale  $\int_a^b f(t) dt$  impropre en  $a$ , et la nature (convergente ou divergente) d'une telle intégrale. Enfin, une intégrale peut être impropre à la fois en  $a$  et en  $b$ . Dans ce cas, on choisit une valeur  $c \in ]a, b[$ . L'intégrale  $\int_a^b f(t) dt$  converge si et seulement si les deux intégrales

$$\int_a^c f(t) dt \quad \text{et} \quad \int_c^b f(t) dt$$

convergent **toutes les deux**. Si une au moins de ces deux intégrales diverge, alors  $\int_a^b f(t) dt$  diverge.

**Remarque.** Mise en œuvre informatique.

Les méthode d'approximation d'intégrales que nous avons vues dans la partie précédentes se transposent mal à des intégrales impropres. Dans la pratique, quand vous devrez calculer la valeur numérique d'une intégrale impropre en  $b$ , les spécifications du projet vous demanderont toujours de calculer une intégrale entre  $a$  et  $x$  où  $x$  sera un argument de la fonction que vous pourrez supposer  $< b$  dans votre programmation.

Il faut néanmoins savoir traiter le cas de le cas suivant, qui reviendra très souvent dans les applications en probabilités.

**Exemple.** Densité de la loi normale.

On considère la fonction  $\varphi$  définie pour tout  $t \in \mathbb{R}$  par  $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ . On peut alors montrer que

$$\int_{-\infty}^{+\infty} \varphi(t) dt \text{ converge et vaut } 1$$

Alors, compte tenu de la parité de  $\varphi$ , on a :

$$\int_{-\infty}^0 \varphi(t) dt = \int_0^{+\infty} \varphi(t) dt = \frac{1}{2}$$

Il vous faudra calculer la primitive suivante de  $\varphi$  :

$$\Phi : x \mapsto \int_{-\infty}^x \varphi(t) dt$$

Cette fonction ne s'écrit pas de façon simple à partir des fonctions usuelles. Vous serez amené à en programmer une approximation numérique. Pour contourner le problème de la borne  $-\infty$ , on remarque que pour tout  $x \in \mathbb{R}$ ,

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \int_{-\infty}^0 \varphi(t) dt + \int_0^x \varphi(t) dt = \frac{1}{2} + \int_0^x \varphi(t) dt$$

Ainsi, pour tout  $x \in \mathbb{R}$ , on est amené à calculer une intégrale sur l'intervalle fermé  $[0, x]$ , qui n'est donc pas une intégrale impropre.

## 4 Fonction de répartition et densité

### 4.1 Définition et exemples

**Définition.** Fonction de répartition.

Soit une variable aléatoire  $X$ . Sa fonction de répartition est la fonction  $F$  définie pour tout  $x \in \mathbb{R}$  par

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

On a alors, pour tout intervalle  $]a, b]$  :  $P(X \in ]a, b]) = F_X(b) - F_X(a)$ .

*Notation.* Quand il n'y a pas d'ambiguïté, on note  $F$  plutôt que  $F_X$ , pour ne pas surcharger l'écriture.

**Exemple.** Considérons une variable aléatoire  $X$  suivant une loi de Bernoulli de paramètre  $p = \frac{1}{3}$  :

$$X(\Omega) = \{0, 1\} \quad \text{avec} \quad P(X=1) = \frac{1}{3} \text{ et } P(X=0) = \frac{2}{3}$$

Sa fonction de répartition est alors une fonction en escalier : pour tout  $x \in \mathbb{R}$ ,

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{2}{3} & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

**Proposition.** Pour toute variable aléatoire  $X$ , sa fonction de répartition  $F$  vérifie :

1.  $F$  est croissante (pas strictement)
2.  $F(x) \xrightarrow{x \rightarrow -\infty} 0$  et  $F(x) \xrightarrow{x \rightarrow +\infty} 1$

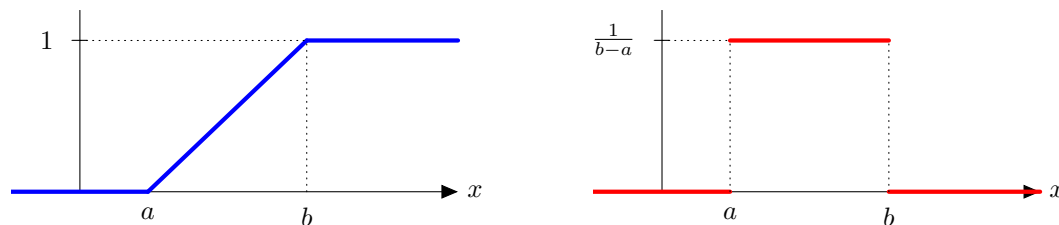


FIGURE 4 – Loi uniforme sur un intervalle  $[a, b]$ . **Courbe de gauche :** la fonction de répartition de  $X$ . Elle est dérivable sur  $\mathbb{R}$ , sauf en  $a$  et en  $b$ . Néanmoins, en ces deux points, elle admet une dérivée à gauche et une dérivée à droite. **Courbe de droite :** densité de  $X$ . Elle est égale à la dérivée  $F'$ , sauf en  $a$  et  $b$ . En ces deux points, elle peut prendre n'importe quelle valeur positive, c'est pourquoi on parle d'une densité plutôt que de la densité. Dans cet exemple, il est bien sûr naturel de lui donner en  $a$  et en  $b$  une des valeurs 0 ou  $\frac{1}{b-a}$ .

**Définition.** Densité.

Soient  $X$  une variable aléatoire et  $F$  sa fonction de répartition. S'il existe une fonction  $f$  telle que pour tout  $x \in \mathbb{R}$ ,

$$F(x) = \int_{-\infty}^x f(t) dt$$

alors  $f$  est une densité de  $X$ .

**Exemple.** Loi uniforme sur  $[a, b]$  (figure 4).

Soit  $(a, b) \in \mathbb{R}^2$  tel que  $a < b$ . Considérons une variable aléatoire  $X$  admettant la fonction de

$$\text{répartition définie pour tout } x \in \mathbb{R} \text{ par } F(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } b \leq x \end{cases}$$

Alors une densité de  $X$  est la fonction  $f$  définie par  $f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$

En effet, pour tout  $x \in \mathbb{R}$  :

- Si  $x < a$ , alors

$$\underbrace{\int_{-\infty}^x f(t) dt}_{f(t)=0} = \int_{-\infty}^x 0 dt = 0 = F(x)$$

- Si  $x \in [a, b]$ ,

$$\begin{aligned} \int_{-\infty}^x f(t) dt &= \underbrace{\int_{-\infty}^a f(t) dt}_{f(t)=0} + \underbrace{\int_a^x f(t) dt}_{f(t)=\frac{1}{b-a}} \\ &= 0 + \left[ \frac{t}{b-a} \right]_a^x \\ &= \frac{x-a}{b-a} \\ &= F(x) \end{aligned}$$

- Et si  $x > b$ ,

$$\begin{aligned} \int_{-\infty}^x f(t) dt &= \underbrace{\int_{-\infty}^a f(t) dt}_{f(t)=0} + \underbrace{\int_a^b f(t) dt}_{f(t)=\frac{1}{b-a}} + \underbrace{\int_b^x f(t) dt}_{f(t)=0} \\ &= 0 + \left[ \frac{t}{b-a} \right]_a^b + 0 \\ &= 1 \\ &= F(x) \end{aligned}$$

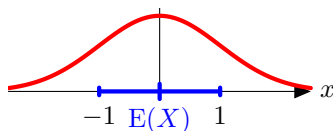


FIGURE 5 – Densité  $\varphi$  d'une variable normale centrée réduite.

De plus, pour tout  $x \notin \{a, b\}$ ,  $F$  est dérivable en  $x$  et  $F'(x) = f(x)$ .

**Proposition.** Soit une variable aléatoire  $X$  admettant une densité  $f$ . Alors :

1. La densité  $f$  est positive.
2. L'intégrale  $\int_{-\infty}^{+\infty} f(t) dt$  converge et vaut 1.
3. La fonction de répartition  $F$  est continue sur  $\mathbb{R}$ . De plus, si  $f$  est continue (ou si elle a un ensemble fini de points de discontinuité), alors  $F'(x) = f(x)$  (sauf aux éventuels points de discontinuité de  $f$ ).
4. Pour tout  $(a, b) \in (\mathbb{R} \cup \{\pm\infty\})^2$  tel que  $a \leq b$ ,

$$P(a \leq X \leq b) = \int_a^b f(t) dt \quad \text{et en particulier,} \quad P(X=a) = \int_a^a f(t) dt = 0$$

**Exemple.** Loi normale.

Soit  $(m, \sigma) \in \mathbb{R}^2$  tel que  $\sigma > 0$ . Une variable aléatoire  $X$  suit une loi normale de paramètres  $m$  et  $\sigma^2$  si elle admet pour densité la fonction

$$f(t) = \frac{1}{\sigma} \varphi\left(\frac{t-m}{\sigma}\right)$$

où la fonction  $\varphi$  a été définie dans l'exemple de la page 12 :  $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ . On note :  $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$ .

En particulier, si  $m = 0$  et  $\sigma = 1$ , alors  $f = \varphi$  et on dit que  $X$  suit une loi normale centrée réduite. Dans ce cas, sa fonction de répartition est  $\Phi(x) = \int_{-\infty}^x \varphi(t) dt$ .

Dans le cadre du projet PFA, il faut pouvoir calculer cette fonction de répartition  $\Phi$ . On a vu dans l'exemple page 12 que

$$\Phi(x) = \frac{1}{2} + \int_0^x \varphi(t) dt$$

## 4.2 Opérations sur les variables aléatoires

**Théorème 4.1.** Densité de  $aX + b$ .

Soit  $X$  une variable aléatoire de densité  $f$ . Alors pour tout  $(a, b) \in \mathbb{R}^2$  tel que  $a \neq 0$ ,  $aX + b$  admet pour densité la fonction  $g$  définie pour tout  $x \in \mathbb{R}$  par

$$g(x) = \frac{1}{|a|} f\left(\frac{x-b}{a}\right)$$

**Explication.** Notons  $F$  et  $G$  les fonctions de répartition de  $X$  et de  $aX + b$ . Supposons par exemple que  $a > 0$ . Alors pour tout  $x \in \mathbb{R}$ ,

$$\begin{aligned} G(x) &= P(aX + b \leq x) \quad \text{par définition de } G \\ &= P\left(X \leq \frac{x-b}{a}\right) \\ &= F\left(\frac{x-b}{a}\right) \end{aligned}$$

Par dérivation, on obtient  $g(x) = G'(x) = \frac{1}{a} f\left(\frac{x-b}{a}\right)$ .

Par ce théorème, on montre que toute variable  $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$  peut s'écrire  $X = m + \sigma Z$  où  $Z \rightsquigarrow \mathcal{N}(0, 1)$ .

**Théorème 4.2.** *Somme de variables aléatoires indépendantes.*

Considérons deux variables aléatoires indépendantes  $X$  et  $Y$  admettant des densités  $f$  et  $g$ . Alors  $X + Y$  admet pour densité la fonction  $h$  définie pour tout  $x \in \mathbb{R}$  par

$$h(x) = \int_{-\infty}^{+\infty} f(t)g(x-t) dt$$

Par ce théorème, on montre que si  $X \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2)$  et  $Y \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2)$  sont deux variables aléatoires indépendantes, alors  $X + Y \rightsquigarrow \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ .

### 4.3 Espérance et variance d'une variable à densité

**Définition.** Soit  $X$  une variable aléatoire admettant une densité  $f$ .

1. L'espérance de  $X$  est définie par

$$E(X) = \int_{-\infty}^{+\infty} t f(t) dt$$

quand l'intégrale converge. Si l'intégrale ne converge pas absolument, alors  $X$  n'admet pas d'espérance.

2. Si  $X$  admet une espérance, alors sa variance est définie par

$$\text{Var}(X) = E\left((X - E(X))^2\right) = \int_{-\infty}^{+\infty} (t - E(X))^2 f(t) dt$$

quand l'intégrale converge. Si l'intégrale diverge, alors  $X$  n'admet pas de variance.

**Remarque.** Quand  $X$  admet une variance, alors

$$\text{Var}(X) = E(X^2) - E^2(X) = \int_{-\infty}^{+\infty} t^2 f(t) dt - E^2(X)$$

**Théorème 4.3.** *Cas de la loi normale.*

Soit une variable aléatoire  $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$ . Alors

$$E(X) = m \quad \text{et} \quad \text{Var}(X) = \sigma^2$$

## Troisième partie

# Finance et assurance

## 5 Finance

Dans le cadre du projet, vous devez calculer un prix à une «option». Le but de cette partie est de définir une «option» et de voir comment on lui affecter un prix.



## 5.1 Définitions des «options»

On considère un actif financier, par exemple une action, et on va considérer une option d'achat («call») ou de vente («put») conclue à la date  $t = 0$  sur cet actif «sous-jacent», à échéance  $T$ . Pour tout instant  $t \in [0, T]$ , le prix de l'actif est noté  $S_t$ .

**Définition.** Option de type «Call».

Un «call» d'échéance  $T$  et de «prix d'exercice»  $K$  est un contrat passé à la date  $t = 0$  entre deux agents, qu'on va nommer  $A$  et  $B$ .

- À la date  $t = 0$ ,  $A$  paye un certain prix à  $B$ , c'est le prix d'achat du call.
- À la date  $t = T$ ,  $A$  pourra choisir d'acheter ou non à  $B$  l'actif financier au prix  $K$  fixé initialement. L'agent  $B$  sera alors tenu de vendre l'actif à ce prix  $K$ , si  $A$  le demande.

Concrètement, ce qui va se passer à la date d'échéance  $T$  :

- Si l'actif s'échange sur le marché à un prix  $S_T < K$ ,  $A$  ne va pas exercer l'option. En effet, il préférera alors acheter l'actif sur le marché, moins cher que le prix  $K$  convenu dans l'option.
- Si en revanche l'actif s'échange sur le marché à un prix  $S_T > K$ , alors  $A$  pourra l'acheter à  $B$  au prix  $K$ , prix avantageux puisque inférieur à la valeur de marché.

Ainsi, à échéance  $T$ , c'est  $A$  qui aura la main et  $B$  sera tenu de suivre. Le contrat donne à  $A$  le droit d'acheter ou non au prix  $K$  et  $B$  est tenu d'accepter. Il est donc normal que, en échange de ce droit,  $A$  paye initialement un certain prix à  $B$  : c'est ce prix qui fait que  $B$  pourra trouver un intérêt à se mettre dans cette situation.

Vous devrez écrire une fonction `price_call` donnant le prix du call, la somme que  $A$  doit payer à  $B$  à la date  $t = 0$  pour acquérir ce droit.

L'autre type d'options que l'on va considérer est le «put». C'est une option basée sur le même principe mais, cette fois, l'acheteur de l'option acquiert le droit de **vendre** l'actif (au lieu du droit de l'acheter).

**Définition.** Option de type «Put»

Un «put» d'échéance  $T$  et de prix d'exercice  $K$  est un contrat passé à la date  $t = 0$  entre deux agents,  $A$  et  $B$ .

- À la date  $t = 0$ ,  $A$  paye un certain prix à  $B$ , c'est le prix d'achat du put.
- À la date  $t = T$ ,  $A$  pourra choisir de vendre ou non à  $B$  l'actif financier au prix  $K$  fixé initialement. L'agent  $B$  sera alors tenu d'acheter l'actif à ce prix  $K$ , si  $A$  le lui demande.

Cette fois, voilà ce qui va se passer à la date d'échéance  $T$  :

- Si l'actif s'échange sur le marché à un prix  $S_T > K$ ,  $A$  ne va pas exercer l'option. En effet, il pourra alors réussir à vendre l'actif sur le marché, plus cher que le prix  $K$  convenu dans l'option.
- Si en revanche l'actif s'échange sur le marché à un prix  $S_T < K$ , alors  $A$  choisira de le vendre à  $B$  au prix  $K$ , prix avantageux puisque supérieur à la valeur de marché.

Là aussi, à échéance  $T$ ,  $A$  aura la main et  $B$  sera tenu de suivre. Cette fois, le contrat donne à  $A$  le droit de **vendre** ou non au prix  $K$  et  $B$  est tenu d'accepter. Il est donc tout aussi normal que, en échange de ce droit,  $A$  paye un certain prix à  $B$  : c'est ce prix qui fait que  $B$  pourra trouver un intérêt à se mettre dans cette situation.

Vous devrez écrire une seconde fonction `price_put` donnant le prix du put, la somme que  $A$  doit payer à  $B$  à la date  $t = 0$  pour acquérir ce droit.

## 5.2 Loi du prix $S_T$ du sous-jacent à l'échéance $T$

Les prix du call et du put dépendent de la loi du prix  $S_T$  de l'actif sous-jacent à la date d'échéance  $T$ .

Un modèle communément admis parmi les financiers est le modèle de Black et Scholes : on suppose qu'il existe  $(\mu, \sigma) \in \mathbb{R}^2$  avec  $\sigma > 0$  tel que

$$S_T = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z} \quad \text{où} \quad Z \rightsquigarrow \mathcal{N}(0, 1)$$

Cette expression de  $S_T$  est suffisante pour la suite du projet et nous ne la démontrerons pas. Pour être plus précis, elle se base sur l'hypothèse que, dans chaque petit intervalle de temps  $[t, t + \delta t]$ , la variation de  $S$  a une partie déterministe (paramétrée par  $\mu$ ) et une partie aléatoire qui suit une loi normale (paramétrée par  $\sigma$ ).

### 5.3 Prix d'un call

On considère un call de prix d'exercice  $K$  à échéance  $T$ , sur un actif sous-jacent de prix  $S_0$ . Le prix du sous-jacent à la date  $T$  suit la loi proposée ci-dessus.

Le calcul mathématique de son prix est un calcul lourd. Il n'est pas nécessaire de le comprendre pour écrire le code de la fonction qui le mettra en œuvre dans le projet PFA.

Ce calcul repose sur les deux étapes suivantes :

1. Exprimer le gain à échéance pour l'acheteur. C'est une fonction  $G(S_T, K)$  qui dépend du prix d'exercice  $K$  du call (valeur connue certaine) et de la valeur  $S_T$  qu'aura l'actif échéance (valeur inconnue, aléatoire). Ce gain est donc une variable aléatoire, il représente aussi la perte du vendeur à l'échéance  $T$ .
2. Le prix du call à la date  $t = 0$  est alors  $E(G(S_T, K))$ .

On rappelle que  $\varphi$  désigne la densité de la loi normale centrée réduite et  $\Phi$  sa fonction de répartition :

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{et} \quad \Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{2} + \int_0^x \varphi(t) dt$$

1. Le gain final pour l'acheteur est une variable aléatoire qui dépend de  $S_T$ . Il vaut :

$$G(S_T, K) = \begin{cases} S_T - K & \text{si } S_T - K > 0 \quad (\text{il achète au prix } K \text{ un titre qu'il peut revendre au prix } S_T) \\ 0 & \text{si } S_T - K \leq 0 \quad (\text{il n'exerce pas l'option}) \end{cases}$$

Et pour le vendeur, cette valeur  $G(S_T, K)$  est sa perte d'argent à l'échéance.

2. Le prix du call est l'espérance de ce gain. On a vu que

$$S_T = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z} \quad \text{où} \quad Z \rightsquigarrow \mathcal{N}(0, 1)$$

Ainsi,  $S_T$  et donc  $G(S_T, K)$  ne sont aléatoire que par la variable  $Z \rightsquigarrow \mathcal{N}(0, 1)$ . D'où :

$$C = E(G(S_T, K)) = \int_{-\infty}^{+\infty} G\left(S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}z}, K\right) \varphi(z) dz$$

Il faut calculer cette intégrale :  $G\left(S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}z}, K\right)$  vaut 0 quand  $S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}z} < K$ , c'est à dire quand

$$z < z_0 = \frac{\ln\left(\frac{K}{S_0}\right) - \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

Ainsi,

$$\begin{aligned} C &= \int_{z_0}^{+\infty} \left( S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}z} - K \right) \varphi(z) dz \\ &= \int_{z_0}^{+\infty} S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}z} \varphi(z) dz - \underbrace{\int_{z_0}^{+\infty} K \varphi(z) dz}_{K(1-\Phi(z_0))} \end{aligned}$$

Pour l'intégrale qu'il reste à calculer, on écrit explicitement  $\varphi(z)$  :

$$\begin{aligned}
 \int_{z_0}^{+\infty} S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}z} \varphi(z) dz &= S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T} \int_{z_0}^{+\infty} \frac{e^{\sigma\sqrt{T}z - \frac{z^2}{2}}}{\sqrt{2\pi}} dz \\
 &= S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T} \int_{z_0}^{+\infty} \frac{e^{-\frac{(z - \sigma\sqrt{T})^2}{2} + \frac{\sigma^2 T}{2}}}{\sqrt{2\pi}} dz \\
 &= S_0 e^{\mu T} \int_{z_0}^{+\infty} \frac{e^{-\frac{(z - \sigma\sqrt{T})^2}{2}}}{\sqrt{2\pi}} dz \\
 &= S_0 e^{\mu T} \int_{z_0 - \sigma\sqrt{T}}^{+\infty} \varphi(u) du \quad \text{en posant } u = z - \sigma\sqrt{T} \\
 &= S_0 e^{\mu T} \left(1 - \Phi(z_0 - \sigma\sqrt{T})\right)
 \end{aligned}$$

Finalement, le prix du call est

$$C = S_0 e^{\mu T} \left(1 - \Phi(z_0 - \sigma\sqrt{T})\right) - K(1 - \Phi(z_0))$$

qu'on peut réécrire (car  $1 - \Phi(z) = \Phi(-z)$ ) :

$$C = S_0 e^{\mu T} \Phi(\sigma\sqrt{T} - z_0) - K\Phi(-z_0)$$

Au final, le prix du call est donnée par :

$$C = S_0 e^{\mu T} \Phi(\sigma\sqrt{T} - z_0) - K\Phi(-z_0) \quad \text{où} \quad z_0 = \frac{\ln\left(\frac{K}{S_0}\right) - \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

## 5.4 Prix d'un put

On considère un put de prix d'exercice  $K$  à échéance  $T$ , sur un actif sous-jacent de prix  $S_0$ . Le prix du sous-jacent à la date  $T$  suit la loi proposée dans la section 5.2. Le calcul du prix du put se fait de la même façon que celui du call.

1. Dans ce cas, le gain pour l'acheteur est

$$G(S_T, K) = \begin{cases} K - S_T & \text{si } K - S_T > 0 \quad (\text{il vend au prix } K \text{ un titre qu'il peut racheter au prix } S_T) \\ 0 & \text{si } K - S_T \leq 0 \quad (\text{il n'exerce pas l'option}) \end{cases}$$

2. Le prix du put est alors  $P = E(G(S_T, K))$ . Par un calcul analogue à celui pour le call, on trouve :

$$P = K\Phi(z_0) - S_0 e^{\mu T} \Phi(z_0 - \sigma\sqrt{T}) \quad \text{où} \quad z_0 = \frac{\ln\left(\frac{K}{S_0}\right) - \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

## 6 Assurance

On considère un client individuel d'une compagnie d'assurance, qui a un contrat de protection contre un risque donné, disons par exemple contre le risque de dégât des eaux. On considère la variable aléatoire

$S$  = «Somme des remboursements que la compagnie devra verser à ce client, cette année»

On se donne pour but de proposer une loi de  $S$ .

## 6.1 Un modèle

La variable  $S$  dépend de deux types de variables aléatoires :

- Le nombre  $N$  de sinistres déclarés par le client. Cette variable prend ses valeurs dans  $\mathbb{N}$ .
- Le coût que la compagnie doit rembourser pour chaque sinistre. Appelons  $X_1, X_2, \dots$  les coûts correspondant à chacun des sinistres. Leurs nombres est  $N$ , ce sont des variables qui prennent leurs valeurs dans  $\mathbb{R}_+^*$ . Nous allons supposer que ces variables sont indépendantes les unes des autres.

Nous proposons ici un modèle simplifié :

- Nous allons nous limiter au cas où  $N$  prend ses valeurs dans  $\{0, 1, 2\}$ . Ainsi, la loi de  $N$  est donnée par un triplet  $(p_0, p_1, p_2) \in (\mathbb{R}_+)^3$  tel que

$$P(N=0) = p_0, \quad P(N=1) = p_1 \quad \text{et} \quad P(N=2) = p_2$$

On a bien sûr  $p_0 + p_1 + p_2 = 1$ .

- Pour tout  $i \in \llbracket 1, N \rrbracket$ , la variable aléatoire  $X_i$  correspondant au remboursement du dommage  $i$  suit une loi «log-normale» de paramètres  $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$ .

**À propos de ce modèle :**

- La plus forte approximation est de considérer que le nombre  $N$  de sinistres a ses valeurs possibles dans  $\{0, 1, 2\}$ .
- En revanche, l'hypothèse que le montant de chaque sinistre suit une loi log-normale est fréquemment utilisée par les assureurs.
- Dès que  $N \geq 2$ , la variable  $S$  est une somme de variables aléatoires, dont il va falloir calculer la densité. Ce calcul est faisable quand on n'ajoute que deux variables, mais plus on en ajoute plus cela devient compliqué.  
Les assureurs considèrent des modèles où  $N$  peut prendre des valeurs plus grandes que 2 mais, en contrepartie, font des approximations sur la loi de la somme des montants  $X_i$ .

**Définition.** Loi log-normale.

On dit qu'une variable aléatoire  $X$  suit une loi log-normale de paramètres  $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$  si

$$X(\Omega) = \mathbb{R}_+^* \quad \text{et} \quad \ln(X) \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$$

On note :  $X \rightsquigarrow \text{Log-}\mathcal{N}(\mu, \sigma^2)$

**Proposition.** Propriétés de la loi log-normale.

Soient  $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$ . Considérons une variable aléatoire  $X \rightsquigarrow \text{Log-}\mathcal{N}(\mu, \sigma^2)$ .

1. Alors  $X$  se met sous la forme  $X = e^{\mu + \sigma Z}$  où  $Z \rightsquigarrow \mathcal{N}(0, 1)$ .

En effet, par définition,  $\ln(X) \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$ . Mais alors :

$$\ln(X) \rightsquigarrow \mathcal{N}(\mu, \sigma^2) \iff \ln(X) \text{ s'écrit } \mu + \sigma Z \text{ où } Z \rightsquigarrow \mathcal{N}(0, 1)$$

On obtient  $X = e^{\mu + \sigma Z}$  et  $Z \rightsquigarrow \mathcal{N}(0, 1)$ .

2. Fonction de répartition  $F_X$  de  $X$  : si  $x \leq 0$ , alors  $F_X(x) = P(X \leq x) = 0$ . En revanche, si  $x > 0$ ,

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(e^{\mu + \sigma Z} \leq x) \\ &= P\left(Z \leq \frac{\ln(x) - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) \end{aligned}$$

3. Densité de  $X$  : par dérivation de  $F_X$ , on obtient : si  $x \leq 0$ ,  $f_X(x) = 0$  et si  $x > 0$ ,

$$f_X(x) = F'_X(x) = \frac{1}{\sigma x} \times \varphi\left(\frac{\ln(x) - \mu}{\sigma}\right)$$

De plus,  $\lim_{x \rightarrow 0^+} f_X(x) = 0$ . Ainsi,  $f_X$  est continue en 0 et  $f_X(0) = 0$ .

## 6.2 Loi de $S$

On se donne donc  $(p_0, p_1, p_2 \in [0, 1]^3$  tels que  $p_0 + p_1 + p_2 = 1$  et  $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$ . Le nombre de sinistres  $N$  prend ses valeurs dans  $\{0, 1, 2\}$  avec les probabilités  $p_0, p_1$  et  $p_2$ . De plus, pour chaque sinistre  $i \leq N$ , le remboursement à effectuer est  $X_i \rightsquigarrow \text{Log-}\mathcal{N}(\mu, \sigma^2)$ .

On considère la variable aléatoire  $S$  égale au total des montants remboursés par l'assureur à son client. Sa fonction de répartition est alors la fonction  $G$  définie par :

- Pour tout  $x < 0$ ,  $G(X) = P(S \leq x) = 0$ .
- Pour  $x = 0$ ,

$$\begin{aligned} G(0) &= P(S \leq 0) \\ &= \underbrace{P(S \leq 0 \mid N=0)}_1 P(N=0) + \underbrace{P(S \leq 0 \mid N=1)}_0 P(N=1) + \underbrace{P(S \leq 0 \mid N=2)}_0 P(N=2) \\ &= p_0 \end{aligned}$$

- Et pour  $x > 0$ ,

$$\begin{aligned} G(x) &= P(S \leq x) \\ &= \underbrace{P(S \leq x \mid N=0)}_1 P(N=0) + \underbrace{P(S \leq x \mid N=1)}_{F_{X_1}(x)} P(N=1) + \underbrace{P(S \leq x \mid N=2)}_{F_{X_1+X_2}(x)} P(N=2) \\ &= p_0 + p_1 F_{X_1}(x) + p_2 F_{X_1+X_2}(x) \end{aligned}$$

La fonction  $G$  n'est pas continue en 0, la variable  $S$  n'admet donc pas de densité.

Dans cette expression de  $G$ , il reste à préciser les fonctions de répartition  $F_{X_1}$  et  $F_{X_1+X_2}$ . La première est la fonction de répartition d'une variable log-normale :

$$F_{X_1}(x) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right)$$

Et pour calculer  $F_{X_1+X_2}$  :

1. Il faut d'abord déterminer la densité de  $X_1 + X_2$  : les densités de  $X_1$  et de  $X_2$  sont identiques :

$$f(t) = \frac{1}{t\sigma} \times \varphi\left(\frac{\ln(t) - \mu}{\sigma}\right) \quad \text{si } t > 0 \text{ et } 0 \text{ sinon}$$

En vertu du théorème 4.2, la densité de  $X_1 + X_2$  est donnée pour tout  $x > 0$  par :

$$f_{X_1+X_2}(x) = \int_{-\infty}^{+\infty} f(t)f(x-t) dt = \int_0^x f(t)f(x-t) dt$$

et vaut 0 si  $x \leq 0$ .

Il faut faire une intégration numérique pour avoir une approximation de  $f_{X_1+X_2}(x)$  (ne pas oublier de remplacer  $f(0)$  par 0, en  $t = 0$  et en  $t = x$ ).

2. On a alors

$$F_{X_1+X_2}(x) = \int_0^x f_{X_1+X_2}(t) dt$$

Ce calcul demande une seconde intégration numérique.