

Chapitre 1.

Probabilité et statistique

1.1 Introduction

La notion de probabilité est un concept profond et complexe qui vise à quantifier, autant que faire se peut, le hasard ou la méconnaissance d'un état. Elle peut être abordée de différentes manières, depuis l'approche strictement mathématique (formulée par Kolmogorov mais difficilement applicable aux cas pratiques) à celle, intuitive, de l'expérimentateur. Elle repose sur des interprétations distinctes souvent complémentaires mais aussi parfois contradictoires.

Les deux idées majeures et fondamentales illustrant la théorie des probabilités sont d'une part, la loi des grands nombres associée à l'approche fréquentiste et objective de la probabilité et d'autre part, le conditionnement lié à l'approche bayésienne et subjective de la probabilité :

- Dans le premier cas, le concept probabiliste se traduit par la considération d'un ensemble de résultats d'une expérience $\Omega = \{E_1, E_2, \dots\}$ où l'on suppose réaliser N fois l'expérience de manière *strictement identique*. Si le résultat E_i est obtenu N_i fois, alors la probabilité du résultat E_i est donnée par la **limite** :

$$\Pr(E_i) = \lim_{N \rightarrow \infty} \frac{N_i}{N}$$

Cette définition simple, naturelle et intuitive, souffre toutefois d'exigences fortes : la probabilité d'un événement n'est pas seulement la propriété d'une expérience particulière, mais plutôt la propriété partagée de l'expérience et de l'ensemble des N expériences identiques. De plus, l'expérience doit être reproductible sous des conditions identiques conduisant à des résultats différents, exigence très forte et en pratique quasiment irréalisable. Par exemple, la question de donner une prédiction probabiliste météorologique ne fait plus de sens !

- Dans le second cas, l'approche consiste à quantifier la probabilité de réalisation de certains résultats de l'expérience en se basant sur l'information connue **a priori**, d'où l'idée afférente de subjectivité. En effet, il est fondamental de remarquer que si l'information connue change, les probabilités de réalisation changent aussi. Cette **dépendance conditionnelle** apparaît dans un théorème très simple, le théorème de Bayes qui conduit à toute une interprétation par essence différente du premier cas.

En pratique, ces deux approches convergent souvent (mais pas toujours !) vers un résultat probabiliste identique (et sont même parfois utilisées en commun pour résoudre un problème, ce qui en soi est incorrect) mais elles restent néanmoins fondamentalement différentes.

La notion de statistique est quant à elle liée à l'outil mathématique qui permettra d'estimer quantitativement le résultat d'une expérience dont les comportements aléatoires sont sujets à des variations imprévisibles ou en trop grand nombre ou trop complexes à modéliser. Cet outil ou plutôt cette boîte à outils est l'équipement de base nécessaire à tout expérimentateur

et seule sa maîtrise permet l'élaboration de résultats scientifiquement corrects. Il existe d'innombrables références générales à consulter (livres, publications, cours, etc...) relatives à la théorie de la probabilité et de la statistique définissant et expliquant les concepts fondamentaux abordés dans ce premier chapitre, aussi les références mentionnées ci-après ne sont que très subjectives et non exhaustives [1, 2, 3, 4, 5, 6, 7, 8].

1.1.1 Expérience, épreuve et événement aléatoires

Rappelons quelques définitions de base :

On appelle **expérience aléatoire** une expérience E qui, reproduite dans des conditions identiques, peut conduire à plusieurs résultats possibles et imprévisibles. L'espace de tous les résultats possibles est défini comme l'**espace d'états ou univers** (associé à l'expérience) et noté Ω .

On appelle **épreuve** la réalisation d'une expérience aléatoire.

On appelle **événement aléatoire** (associé à l'expérience E) un sous-ensemble de Ω correspondant à la propriété du système qui, une fois l'épreuve effectuée, est ou n'est pas réalisée. Un événement est donc une partie de Ω .

Ainsi par exemple, pour l'expérience du lancer d'un dé, on a $\Omega = \{1, 2, 3, 4, 5, 6\}$ et l'événement peut être $A =$ (la valeur obtenue est inférieure à 5).

1.1.2 Axiomatique de Kolmogorov

La théorie des probabilités due principalement à Kolmogorov [9] repose sur la définition suivante de la probabilité :

Une **probabilité** sur l'espace (Ω, \mathcal{A}) , où Ω est l'espace d'états (fini ou infini) et \mathcal{A} un ensemble d'événements $\{A_i\}$, est une application de \mathcal{A} dans $[0; 1]$, notée \Pr , telle que :

- $\Pr(\Omega) = 1$,
- $0 \leq \Pr(A_i) \leq 1 \quad \forall i$,
- Pour tout ensemble d'événements incompatibles (mutuellement exclusifs) $A_i \in \Omega$,
 $\Pr(\cup_i A_i) = \sum_i \Pr(A_i)$.

Il découle de cette définition¹ que la probabilité \Pr vérifie les propriétés suivantes pour tout $A \in \mathcal{A}$:

$$\Pr(\emptyset) = 0 \tag{1.1}$$

$$\Pr(\bar{A}) = 1 - \Pr(A) \quad (\text{normalisation}) \tag{1.2}$$

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \quad (\text{règle de somme}) \tag{1.3}$$

$$\Pr(A) \leq \Pr(B) \quad \text{si } A \subset B \tag{1.4}$$

Il est possible maintenant de définir un **espace de probabilité** formé par le triplet $(\Omega, \mathcal{A}, \Pr)$. Cet espace est primordial puisqu'il établit un lien direct entre la théorie des probabilités et la théorie de la mesure. Une modélisation probabiliste, c'est-à-dire la traduction mathématique d'un phénomène physique expérimental, revient à une description de l'expérience aléatoire associée par un espace de probabilité.

1. Rapelons quelques notions et opérations élémentaires sur les ensembles :

Un ensemble est une collection d'objets appelés éléments. L'appartenance et la non-appartenance d'un élément à un ensemble sont notées respectivement \in et \notin . Un ensemble peut être composé de sous-ensembles ou parties. Un ensemble A inclus dans l'ensemble Ω est noté $A \subset \Omega$.

L'intersection des ensembles A et B , notée $A \cap B$ est l'ensemble des éléments appartenant à la fois à A et B .

La réunion des ensembles A et B , notée $A \cup B$ est l'ensemble des éléments appartenant à au moins l'un des deux ensembles. La réunion d'une famille d'ensembles (A_i) est notée $\cup_i A_i$.

L'ensemble ne contenant aucun élément est appelé ensemble vide et est noté \emptyset .

Les ensembles A et B sont appelés disjoints si $A \cap B = \emptyset$.

Si $A \subset \Omega$, l'ensemble complémentaire de A , noté \bar{A} ou encore A^c est l'ensemble des éléments n'appartenant pas à A et appartenant à Ω .

La différence de 2 sous-ensembles A et B , notée $A \setminus B$ est l'ensemble des éléments qui sont dans A mais pas dans B .

1.1.3 Variable aléatoire

En théorie des probabilités mais aussi d'un point de vue calculatoire, on préfère adopter une vision fonctionnelle plutôt qu'ensémbliste, aussi on utilise les variables aléatoires à la place des événements. Une variable aléatoire est une grandeur qui dépend du résultat de l'expérience, par exemple le nombre de 1 obtenu dans le lancer du dé. Elle crée donc une relation entre un espace probabilisé $(\Omega, \mathcal{A}, \Pr)$ (ensemble abstrait de tous les états possibles d'une expérience) et un espace mesurable (ensemble de tous les résultats quantifiables que l'on peut obtenir de l'expérience).

D'un point de vue plus mathématique, **une variable aléatoire** X est une application de (Ω, \mathcal{A}) dans un ensemble \mathcal{F} telle que :

$$\omega \in \Omega \rightarrow X(\omega) = x \in \mathcal{F}$$

L'ensemble \mathcal{F} peut être très général (fini ou dénombrable ou \mathbb{R} ou \mathbb{R}^n ou plus complexe) :

- Une variable aléatoire X est **quantitative** si ses valeurs x sont mesurables. Ainsi elle peut être **continue** ($\mathcal{F} = \mathbb{R}$) ou **discrète** ($\mathcal{F} = \mathbb{N}$). Pour la suite du cours, on s'intéressera quasi-exclusivement à ce type de variable aléatoire.
- Une variable est **qualitative** si ses valeurs ne sont pas numériques mais sont des caractéristiques que l'on appelle **modalités**.
- Une variable qualitative est dite **ordinaire** si ses valeurs sont naturellement ordonnées (par exemple une appréciation, une classe d'âge, etc ...). Dans le cas contraire, elle est dite **nominale** (une couleur par exemple).

Remarque importante :

Malgré son nom, une variable aléatoire n'est pas une variable au sens analytique du terme mais une fonction de la variable $\omega \in \Omega$. Par souci de simplification de l'écriture, on adoptera très souvent dans la suite de manière indifférente la notation simple (mais abusive) sans capitale pour indiquer une variable aléatoire, c'est-à-dire le réel ou l'entier x qui est le résultat de l'application X . \square^1

1.1.4 Echantillon

En tant qu'expérimentateur, nous nous restreignons aux cas d'expériences où le résultat est un ensemble fini de nombres réels x_i puisque l'on n'a pas toujours accès à l'ensemble possible des valeurs (**population**). Un **échantillonnage** correspond à des tirages indépendants et équiprobables d'individus au sein de la population.

D'un point de vue probabiliste, un **échantillon** X_1, \dots, X_n de taille n est un n -uplet (X_1, \dots, X_n) de variables aléatoires X_i , indépendantes ou dépendantes, issues ou non d'une même population. Lorsque les X_i sont des variables aléatoires mutuellement indépendantes provenant de la même population alors on parle d'un échantillon **indépendant et identiquement distribué, noté i.i.d.**, autrement dit les n composantes ont la même loi de probabilité (appelée loi mère de l'échantillon). On identifie alors la réalisation (x_1, \dots, x_n) de l'échantillon i.i.d. de taille n à la série des n observations successives d'un même phénomène aléatoire, ces résultats étant indépendants les uns des autres (au sens où la i^{eme} observation n'influence pas la j^{eme} observation) et dans les mêmes conditions expérimentales.

La problématique habituellement rencontrée consiste, à partir d'un échantillon fini mesurable de données provenant d'une population répartie selon une certaine distribution appelée loi de probabilité **inconnue**, à déduire les propriétés de cette loi (principalement ses paramètres) : on parle alors d'**estimation (ponctuelle, fonctionnelle ou encore par intervalle)**. Par simplicité, nous utiliserons régulièrement le terme échantillon pour signifier à la fois l'échantillon des observations x_1, \dots, x_n et le n -uplet aléatoire X_1, \dots, X_n . Cette introduction intuitive de la répartition des données suivant une loi peut être précisée de façon mathématique comme nous allons le montrer maintenant.

1. Dans la suite de cet ouvrage, afin d'en faciliter la lecture, les Remarques se terminent par un carré blanc et les Exemples par un carré noir.

1.1.5 Densité (ou loi) de probabilité

D'un point de vue mathématique, la **loi (ou densité) de probabilité** f , notée parfois ddp en abréviation, est une mesure-image de la probabilité \Pr par l'application mesurable X . Elle décrit donc totalement le comportement fonctionnel de la variable aléatoire et peut (ou ne pas) dépendre de paramètres intrinsèques.

Dans le cas **discret**, il est possible de définir **une fonction de probabilité** (encore appelée **fonction de masse** dans ce cas là) $f(x)$ positive pour chaque valeur de x vérifiant les axiomes de probabilité :

$$\sum_{(\Omega, \mathcal{A})} f(x) = 1, \quad \Pr(A) \equiv \Pr(X \in A) = \sum_A f(x), \quad A \subset \Omega$$

Dans le cas **continu**, il n'est pas possible de définir la probabilité pour une valeur exacte, aussi on introduit la probabilité qu'un résultat X soit dans un certain intervalle de largeur dx qui est directement proportionnelle à la densité de probabilité f telle que :

$$\Pr(x \leq X \leq x + dx) = f(x)dx, \quad \int_{\Omega} f(x)dx = 1 \quad (1.5)$$

Notation :

X et par extension abusive $x \sim f$ signifie X est distribué suivant la loi de probabilité f .

En général, passer du cas continu au cas discret revient à transformer les intégrales en somme.

1.1.6 Distribution de probabilité cumulative (fonction de répartition)

La fonction appelée distribution de probabilité cumulative (notée dpc) est la probabilité que la variable aléatoire soit inférieure ou égale à une valeur définie. Elle est désignée par la lettre majuscule F indicée avec la lettre minuscule de la densité de probabilité correspondante :

$$F_f(x) = \Pr(X \leq x) = \int_{-\infty}^x f(x')dx' \quad (1.6)$$

Lorsque le contexte est clair, l'indice est souvent omis.

Remarque importante :

Cette fonction est aussi appelée **fonction de distribution cumulée** mais elle est principalement utilisée sous la dénomination de **fonction de répartition** de la variable aléatoire X . \square

Les propriétés de la fonction de répartition sont :

- $F(-\infty) = 0, F(+\infty) = 1, 0 \leq F(x) \leq 1,$
- $F(x)$ est monotone croissante et $f(x) = \frac{\partial F(x)}{\partial x}$

Le contenu probabiliste est donné par la relation fondamentale suivante :

$$\Pr(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a) \quad (1.7)$$

La fonction de répartition joue un rôle très important en statistique comme on le verra par la suite même si on utilise aussi fréquemment la densité associée.

1.1.7 Inverse généralisée d'une fonction de répartition

Soit $F(x) = \Pr(X \leq x)$ une fonction de répartition (version continue à droite). On définit la **fonction de répartition inverse généralisée** F^{-1} de F comme la fonction croissante et continue à gauche :

$$F^{-1}(t) = \min\{x : F(x) \geq t\}, \quad 0 < t < 1 \quad (1.8)$$

A partir de cette définition, on en déduit l'importante proposition suivante appelée

Transformation de probabilité inverse : soit F une fonction de répartition quelconque,

- si U est une variable aléatoire de distribution uniforme entre 0 et 1 (notée $\mathcal{U}(0, 1)$ comme il sera vu ultérieurement avec l'équation 1.93), alors la loi de la variable aléatoire $F^{-1}(U)$ admet F pour fonction de répartition ;
- soit X une variable aléatoire dont la loi admet F pour fonction de répartition. Si F est continue, alors la variable aléatoire $F(X)$ est distribuée uniformément :

$$F(X) \sim \mathcal{U}(0, 1) \quad (1.9)$$

Ce résultat est intéressant car il permet en particulier de simuler des variables aléatoires de fonction de répartition F à partir de variables aléatoires uniformes sur $[0; 1]$. Il est également à la base de procédures de tests d'hypothèses (adéquation) comme nous le verrons dans un chapitre ultérieur.

1.1.8 Moments, moyenne, variance et quantile

Le **moment** d'ordre m , où $m \in \mathbb{N}^*$, d'une variable aléatoire continue x pour laquelle la loi de probabilité est déterminée par $f(x)$ s'écrit en fonction de l'**espérance mathématique** notée E :

$$E[x^m] = \int_{-\infty}^{+\infty} x^m f(x) dx \quad (1.10)$$

Pour une variable aléatoire discrète $x = \{x_1, \dots, x_n\}$ dont la probabilité associée à x_i est p_i , le moment d'ordre m devient la somme :

$$E[x^m] = \sum_{i=1}^n x_i^m p_i \quad (1.11)$$

Le moment existe si sa valeur est finie (c'est-à-dire si l'intégrale ou la somme est définie). Il est à noter que $E[x^m]$ n'existe pas toujours suivant m et $f(x)$.

Le moment le plus communément utilisé est l'espérance de la variable aléatoire elle-même, appelée **moyenne théorique** ou encore **moyenne exacte** :

$$\mu \equiv E[x] = \int_{-\infty}^{+\infty} x f(x) dx \quad (1.12)$$

La moyenne est souvent une bonne mesure de position et caractérise la tendance centrale de la distribution. Elle indique souvent (mais pas toujours!) où se situe la région la plus probable. En statistique, nous verrons que **la moyenne d'un échantillon** ou **moyenne empirique**, \bar{x} , représentant la moyenne des résultats d'une expérience, peut être utilisée pour estimer la moyenne théorique (μ) de la loi correspondante. Si l'on a une fonction $u(x)$, l'espérance de $u(x)$ pour la loi $f(x)$ de la variable aléatoire X est naturellement définie par :

$$E_X[u(x)] = \int_{-\infty}^{+\infty} u(x) f(x) dx \quad (1.13)$$

Usuellement, lorsque le contexte est clair, on note plus simplement $E[u(x)]$.

Les propriétés de l'espérance sont :

- Si k et m sont des constantes réelles et $u(x)$ une fonction,

$$E[ku(x) + m] = kE[u(x)] + m \quad (1.14)$$

donc si k_1 et k_2 sont des constantes réelles et $u_1(x)$, $u_2(x)$ des fonctions,

$$E[k_1 u_1(x) + k_2 u_2(x)] = k_1 E[u_1(x)] + k_2 E[u_2(x)].$$

- En général, pour deux variables aléatoires x et y : $E[xy] \neq E[x]E[y]$. L'égalité est **vraie** si les deux variables sont **indépendantes**.

Remarque :

Certains ouvrages utilisent des notations différentes ($\langle u(x) \rangle$ ou encore $\overline{u(x)}$). \square

Le **moment central d'ordre** m d'une variable aléatoire x pour laquelle la densité de probabilité est définie par $f(x)$, s'écrit :

$$E[(x - \mu)^m] = \int_{-\infty}^{+\infty} (x - \mu)^m f(x) dx \quad (1.15)$$

On s'intéresse surtout au moment central second appelé **variance** et noté $V[x]$ ou encore plus simplement σ^2 . Elle caractérise la dispersion de la distribution autour de la valeur moyenne théorique μ :

$$V[x] = E[(x - \mu)^2] = E[x^2] - \mu^2 = E[x^2] - (E[x])^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \quad (1.16)$$

Si l'on a une variable aléatoire x de loi $f(x)$, par extension, la variance d'une fonction de x , $u(x)$ s'écrit :

$$V[u(x)] = E[(u(x) - E[u(x)])^2] = \int_{-\infty}^{+\infty} (u(x) - E[u(x)])^2 f(x) dx \quad (1.17)$$

Etant une loi quadratique, la dimension de la variance n'est pas celle de la moyenne. C'est pourquoi on utilise souvent l'**écart-type** ou **écart standard** noté σ (racine carrée de la variance).

La variance peut traduire la notion d'incertitude statistique (ou stochastique) : Plus la variance est faible et moins le résultat de l'expérience aléatoire est dispersif et donc incertain. Comme nous le verrons plus tard, la moyenne et la variance jouent chacune un rôle fondamental en statistique.

On appelle **variable aléatoire centrée réduite**, une variable aléatoire y construite à partir de la variable aléatoire x selon la définition :

$$y = \frac{x - E[x]}{\sqrt{V[x]}}$$

La propriété fondamentale de la variance (qui se démontre aisément d'après la définition et les propriétés de l'espérance ci-dessus) est, pour k, m réels et u une fonction de x :

$$V[ku(x) + m] = k^2 V[u(x)] \quad (1.18)$$

Après avoir caractérisé la position et la dispersion de la densité de probabilité avec la moyenne et la variance, il est possible également de caractériser la forme de la densité. Du fait que toutes les distributions symétriques ont des moments centraux impairs nuls, le troisième moment central fournit une mesure de l'asymétrie de la distribution. On définit ainsi la variable sans dimension appelée **coefficient d'asymétrie** (skewness en anglais) par :

$$\gamma_1 \equiv \frac{E[(x - \mu)^3]}{\sigma^3} \quad (1.19)$$

Ainsi $\gamma_1 = 0$ indique une loi symétrique autour de sa moyenne. Si une queue de distribution se profile pour les valeurs supérieures (inférieures) à la moyenne alors $\gamma_1 > 0$ (< 0 respectivement).

L'aplatissement d'une loi de probabilité est mesurée au moyen de son **coefficient d'aplatissement** (kurtosis en anglais) défini par :

$$\gamma_2 \equiv \frac{E[(x - \mu)^4]}{\sigma^4} - 3 \quad (1.20)$$

La quantité -3 est introduite afin que ce coefficient soit nul pour une loi gaussienne (la plus connue des lois de probabilité que l'on verra ultérieurement). Une valeur $\gamma_2 > 0 (< 0)$ implique que la densité est plus (moins) piquée qu'une gaussienne. Ces deux coefficients de forme sont illustrés par le **figure 1.1**.

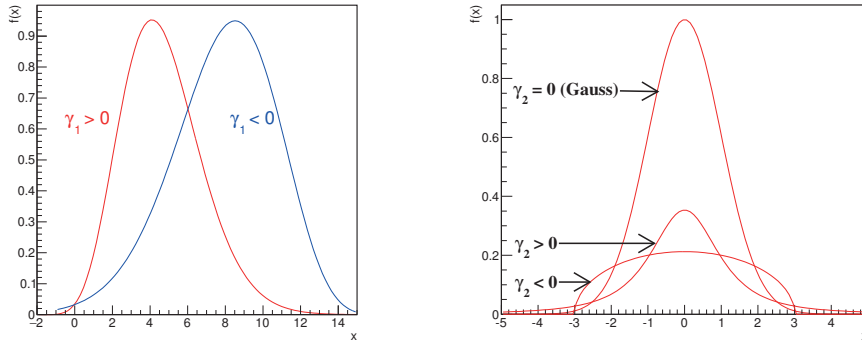


FIGURE 1.1 – Effet du coefficient d'asymétrie γ_1 non nul et du coefficient d'aplatissement γ_2 sur une loi de probabilité.

Remarques :

Les coefficients d'asymétrie et d'aplatissement aussi appelés coefficients de Fisher sont parfois définis différemment. Ces indicateurs de forme ne sont utilisables que dans le cas d'une distribution possédant un seul maximum (unimodale). \square

On peut montrer que si tous les moments centraux d'une distribution existent alors celle-ci est entièrement caractérisée par ses moments. En statistique, il est possible d'estimer chaque moment de la distribution par le moment correspondant de l'échantillon et ainsi, en principe, de complètement reconstruire la distribution.

D'autres caractéristiques d'une densité de probabilité sont aussi :

- Le **mode** qui correspond à la position du maximum de la loi. Une loi de probabilité peut être multimodale.
- La **médiane** correspond à la valeur de x pour laquelle $F(x) = 1/2$. La médiane n'est pas toujours définie de manière non ambiguë car plusieurs valeurs de x peuvent vérifier la définition.
- Le **quantile** $Q_{f,\alpha}$ d'une distribution de probabilité f est défini de la façon suivante : Soit $\alpha \in]0; 1[$ et F la fonction de répartition de f . Le quantile d'ordre α , $Q_{f,\alpha}$, de la distribution f est donné par :

$$F(Q_{f,\alpha}) = \alpha \Rightarrow Q_{f,\alpha} = F^{-1}(\alpha) \quad (1.21)$$

Le **quantile d'ordre α** est la valeur de la fonction de répartition inverse au point α . Il représente la valeur associée à la variable aléatoire décrite par la distribution f telle que la surface sous f comprise entre la valeur minimale et la valeur quantile de la variable aléatoire représente une fraction α de la surface totale de la distribution. Lorsque le contexte est clair, on écrit aussi parfois plus simplement Q_α .

Les 1^{er} et 3^{eme} **quartiles** (un quartile étant un quantile multiple du quart) sont définis par $Q_1 = Q_{0.25} = F^{-1}(0.25)$ et $Q_3 = Q_{0.75} = F^{-1}(0.75)$ respectivement. L'**intervalle interquartile** $[Q_1; Q_3]$ contient donc 50% des données. Par définition, le second quartile correspond à la médiane.

- L'**étendue** ou intervalle de variation est la différence entre les 2 valeurs extrêmes, si elles existent : $x_{max} - x_{min}$.

1.2 Un peu plus sur les probabilités

1.2.1 Fonction caractéristique d'une densité de probabilité

La fonction caractéristique $\phi(t)$ d'une densité $f(x)$ est définie comme la transformée de Fourier de $f(x)$:

$$\phi(t) \equiv \int_{-\infty}^{+\infty} e^{itx} f(x) dx = E \left[e^{itx} \right] \quad (1.22)$$

Elle caractérise complètement la loi de probabilité et simplifie souvent les calculs des moments ainsi que la convolution de deux distributions. En effet, on a :

$$\left. \frac{d^m \phi(t)}{dt^m} \right|_0 = \int_{-\infty}^{+\infty} (ix)^m e^0 f(x) dx = i^m E[x^m]$$

On peut ainsi par exemple considérer le cas d'une distribution de Poisson (que l'on verra ultérieurement) $\mathcal{P}(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$. La fonction caractéristique s'écrit donc :

$$\phi(t) = e^{-\lambda} \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} = e^{-\lambda} \exp(e^{it} \lambda) = \exp(\lambda(e^{it} - 1))$$

A partir de là, il vient :

$$E[k] = \mu = \frac{1}{i} \left. \frac{d\phi}{dt} \right|_0 = i\lambda \frac{1}{i} = \lambda \quad \text{et} :$$

$$E[k^2] = \frac{1}{i^2} \left. \frac{d^2\phi}{dt^2} \right|_0 = -\exp(\lambda(e^{i0} - 1))((\lambda i e^{i0})^2 - \lambda e^{i0}) = (\lambda^2 + \lambda) \Rightarrow V[k] = E[k^2] - \mu^2 = \lambda$$

1.2.2 Densité de probabilité de plusieurs variables aléatoires

Jusqu'à présent, nous avons considéré une densité dépendant d'une variable aléatoire unique. L'extension n -dimensionnelle est une variable aléatoire notée \underline{x} qui est définie par une collection de n variables aléatoires x_1, x_2, \dots, x_n telle que :

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \underline{x}^T = (x_1 \quad x_2 \quad \dots \quad x_n)$$

La variable aléatoire \underline{x} est distribuée selon **une densité de probabilité multidimensionnelle** $f(\underline{x}) = f(x_1, x_2, \dots, x_n)$ positive $\forall \underline{x}$ et normalisée :

$$\int_{\Omega} f(\underline{x}) d\underline{x} = 1 \quad , \quad d\underline{x} = dx_1 dx_2 \dots dx_n$$

Pour le cas où $n = 2$, on parlera souvent de **densité de probabilité conjointe**.

La moyenne de la variable aléatoire x_i (composante i) est simplement :

$$\mu_i \equiv E[x_i] = \int_{\Omega} x_i f(\underline{x}) d\underline{x}$$

Ces moyennes peuvent être réécrites sous forme matricielle :

$$\underline{\mu} \equiv E[\underline{x}] = \int_{\Omega} \underline{x} f(\underline{x}) d\underline{x} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

L'équation 1.16 définissant la variance doit être adaptée au cas multidimensionnel en introduisant une **matrice de covariance** \underline{V} :

$$\underline{V} \equiv V[\underline{x}] = E \left[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^T \right] \quad (1.23)$$

dont les éléments de matrice sont donnés par :

$$V_{ij} = \sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)] = \int_{\Omega} (x_i - \mu_i)(x_j - \mu_j) f(\underline{x}) d\underline{x} \quad (1.24)$$