

## Concepts de probabilités et incertitudes

**Imad LAKTINEH**

*Institut de physique Nucléaire de Lyon, IPNL*

### 1 Introduction à la notion de Probabilité

#### Expérience aléatoire

Une expérience est qualifiée d'aléatoire si on ne peut pas prévoir par avance son résultat et si répétée, dans les mêmes conditions, elle peut donner des résultats différents.

Le résultat d'une telle expérience peut être considéré comme un élément  $\omega$  de l'ensemble des résultats possibles  $\Omega$ , appelé ensemble fondamental de l'univers des possibles (sample space).

La nature de l'ensemble  $\Omega$  n'est pas définie d'une manière unique. Elle dépend de l'usage que l'on veut faire des résultats.

*Exemple :* Le jet de deux dés  $\Omega$  peut être les différents couples possibles

$$\{(1, 1), (1, 2) \cdots \cdots (6, 6)\}$$

ou la somme possible des deux dés :

$$\{(2), (3) \cdots \cdots (12)\}$$

#### Événement

Un événement est une proposition logique relative au résultat de l'expérience.

*Exemple :* somme des deux dés  $> 7$

$$\Rightarrow \{(4, 4), (4, 5), (5, 5), (4, 6), (5, 6), (6, 6)\}$$

#### Algèbre d'événement

L'ensemble des événements constitue une classe  $C$  des parties de  $\Omega$ . On définit sur  $C$  une algèbre appelée  $\sigma$ -algèbre de Boole ou tribu :

- si  $A \in C$  alors  $\bar{A} \in C$
- pour tout ensemble dénombrable  $A_1, A_2 \cdots \in C$   
 $U_i A_i \in C$
- $\Omega \in C$

le couple  $(\Omega; C)$  est appelé espace probabilisable.

#### Espace probabilisé

Un système est appelé système complet d'événements s'il satisfait la condition suivante :

$$\{A_1, A_2 \cdots A_n\} \in \phi / \forall i \neq j \quad A_i \cap A_j = \phi, \quad U_i A_i = \Omega$$

#### Probabilité : Axiomatique de Kolmogorov

On appelle une loi de probabilité sur  $(\Omega; C)$  une application  $P$  de  $C$  dans  $[0,1]$  telle que

- $P(\Omega) = 1$

- pour tout ensemble d'événements incompatibles

$$A_1, A_2, \dots, A_n, \text{ on a } P(U_i A_i) = \sum_i P(A_i)$$

$(\Omega; \mathcal{C}; P)$  est appelé espace probabilisé

*Propriétés :*

La loi de probabilité définie auparavant possède les propriétés suivantes :

- $P(\phi) = 0$  ;  $P(\bar{A}) = 1 - P(A)$
- $P(A) \leq P(B)$  si  $A \subseteq B$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(U_i A_i) \leq \sum_i P(A_i)$
- si l'ensemble  $\{B_i\}$  constitue un système complet d'événements alors

$$\forall A : P(A) = \sum_i P(A \cap B_i)$$

### Probabilité conditionnelle

$P(A/B)$  est la probabilité conditionnelle de  $A$  sachant  $B$  et est définie par :

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Il s'agit d'une probabilité :

- $P(\Omega/B) = 1$
- $P(U_i A_i/B) = \sum_i P(A_i/B)$

### Indépendance de deux événements

$A$  et  $B$  sont indépendants si :

$$\begin{aligned} P(A/B) = P(A) &\Rightarrow \frac{P(A \cap B)}{P(B)} = P(A) \\ &\Rightarrow P(A \cap B) = P(A) P(B) \end{aligned}$$

### Formules de Bayes

La première formule de Bayes est :

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$

Elle dérive des relations :

$$P(A \cap B) = P(A/B)P(B) = P(B/A)P(A)$$

$$\begin{aligned} P(A) &= \sum_i P(A \cap B_i) \quad \{B_i\} \text{ système complet} \\ &= \sum_i P(A/B_i)P(B_i) \end{aligned}$$

D'où la deuxième formule de Bayes :

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{\sum_k P(A/B_k)P(B_k)}$$

*Exemple :* Le département de physique d'une Université a acheté 100 ordinateurs de trois marques différentes. Un taux de défaillance est connu pour chaque marque :

marque	nombre	$\epsilon$
$m_1$	30	2%
$m_2$	50	2%
$m_3$	20	3%

L'ordinateur attribué au chercheur  $\lambda$  est défectueux. Quelle est la probabilité qu'il soit d'une marque donnée ?

$$P(m_1/\text{défectueux}) = \frac{\frac{2}{100} \times \frac{30}{100} + \frac{2}{100} \times \frac{50}{100} + \frac{3}{100} \times \frac{20}{100}}{\frac{2}{100} \times \frac{30}{100} + \frac{2}{100} \times \frac{50}{100} + \frac{3}{100} \times \frac{20}{100}} = \frac{6}{22}$$

$$P(m_2/\text{défectueux}) = \frac{10}{22}$$

$$P(m_3/\text{défectueux}) = \frac{6}{22}$$

### 1.1 Notion de probabilité : approches fréquentiste et bayésienne

Il n'est pas toujours aisé d'appliquer le concept théorique de la probabilité sur des cas réels. Deux concepts ont été développés pour permettre d'utiliser la notion de probabilité dans les cas pratiques.

#### Concept théorique

Le concept théorique de la probabilité est basé sur l'existence d'un ensemble fini munie d'une symétrie  $\Rightarrow$  chaque événement élémentaire possède la même probabilité  $\Rightarrow$  la probabilité est une affaire de dénombrement lorsque ceci est possible.

$$P(A) = \frac{\text{nb de cas favorables}}{\text{nb de cas possibles}}$$

Or ceci n'est pas toujours le cas dans les situations rencontrées dans les expériences de physique.

#### Vision dite objective : fréquentiste

Cette approche est basée sur la loi des grands nombres on répète l'expérience un grand nombre de fois et la fréquence de répétitions définit la probabilité.

*Critiques :*

- Cette vision ne peut probabiliser les événements rares : quelle est la probabilité qu'il neige le 22/4/2104 ?
- La probabilité dite fréquentiste est basée sur la loi des grands nombres qui considère que la probabilité est déjà définie !

#### Vision dite subjective : Bayésienne

La vision objective étant limitée, la vision subjective élargit le champ d'action en faisant appel au théorème de Bayes

$$P(A/B) = \frac{Pr(B/A) P(A)}{P(B)}$$

La probabilité d'un événement est assujetti à l'évolution de l'information / mesure d'incertitude.

*Critiques :*

Cette probabilité dépend d'un choix arbitraire du priori. Il s'agit donc d'une probabilité qui varie avec l'observateur.

## 2 Variable aléatoire et probabilité

C'est une application d'un ensemble probabilisé  $(\Omega; \mathcal{C}; P)$  dans un autre ensemble  $E$ .  $E$  doit être probabilisé  $(; \tau)$ . Tout élément  $T \in \tau$  a pour image réciproque un événement.

*Exemple :*

Soit un jeu de 2 dés (application = somme des 2 dés).  $\Omega = \{(1, 1), (1, 2) \cdots (6, 6)\}$  muni de  $P : P(\omega) =$

$\frac{1}{36} \forall \omega \in \Omega \ E = \{2, 3 \dots 12\}$  somme des 2 dés  $s = 7 \in E$   
 $P(s = 7) = P(1, 6), (2, 5), (3, 4), (6, 1), (5, 2), (4, 3)) = \frac{6}{36}$

L'application permet de transporter la loi de probabilité de  $\Omega$  sur  $E \ P \rightarrow P_X$ . si  $E \equiv \mathbf{R}$  on parle d'une variable aléatoire réelle.

$$P_X(A)_{A \in \mathbf{R}} = P(\omega / X(\omega) \in A) = P(X^{-1}(A))$$

## 2.1 Fonction de densité de probabilité

Si  $X$  est une variable discrète on utilise la notion de probabilité définie auparavant. Si  $X$  est continue, la théorie de la mesure nous donne :  $P(X = x) = 0$ . Il faut donc introduire la notion de densité de probabilité :

$$P(x_{\min} < x < x_{\max}) = \int_{x_{\min}}^{x_{\max}} f(x) dx$$

## 2.2 Fonction de répartition

La fonction de répartition  $F$  est définie telle que :

$$F(x) = P(X < x)$$

$F$  est une fonction monotone continue à gauche.

si  $X$  est continue  $\Rightarrow$

$$F(x) = \int_{x_{\min}}^x f(x') dx'$$

$$F(x_{\min}) = 0 \quad ; \quad F(x_{\max}) = 1$$

$$f(x) = \frac{\partial F(x)}{\partial x}$$

*Remarque :*  $F$  est plus fondamental que  $f$ .

*Application importante :*

Soit  $X$  une variable aléatoire réelle et soit  $\phi(X)$  une fonction de  $X$  (supposée dérivable). Si  $f, F$  sont respectivement la fonction de densité de probabilité et la fonction de répartition de la variable aléatoire  $X$ , quelles sont les fonctions équivalentes de  $\phi$  que l'on notera respectivement  $(g, G)$  ?

*Réponse :*

Cas d'une fonction  $\phi$  bijective. Dans ce cas  $\phi$  est monotone.

– Cas où  $\phi$  est croissante alors  $\phi' > 0$  et

$$F(x) = G(\phi(x)) \text{ car } P(X < x) = P(\phi(X) < \phi(x))$$

$$f(x) = g(y) \phi'(x) \Rightarrow g(y) = \frac{f(x)}{\phi'(x)}; y = \phi(x)$$

– cas où  $\phi$  est décroissante alors  $\phi' < 0$  :

$$F(x) = 1 - G(\phi(x)) \text{ car } P(X < x) = P(\phi(X) > \phi(x))$$

$$f(x) = -g(y) \phi'(x) \Rightarrow g(y) = \frac{f(x)}{-\phi'(x)}; y = \phi(x)$$

donc dans les deux cas :

$$g(y) = \frac{f(x)}{|\phi'(x)|}$$

Dans le cas d'une fonction  $\phi$  quelconque, on divise le domaine de définition de la fonction en des intervalles dans lesquels la fonction est soit croissante soit décroissante. On applique ensuite à la fonction en question et dans chacun de ces intervalles la méthode utilisée dans le cas d'une fonction bijective.

*exercice* : Trouver la fonction de partition de la fonction  $\phi(x) = x^2$

### 2.3 Moments d'une variable aléatoire

Ce sont des valeurs réelles qui permettent de caractériser  $f(x)$  :

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

#### Espérance

Il s'agit du moment d'ordre 1 :

$$m = 1 \Rightarrow E(x) = \mu = \int_{-\infty}^{+\infty} x f(x) dx$$

#### Les moments centraux

Ce sont des moments autour de la valeur moyenne  $\mu$

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

#### Variance

Il s'agit d'un moment central d'ordre 2 :

$$m = 2 \Rightarrow V = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

En plus de l'espérance et de la variance, deux autres moments sont parfois utilisés. Il s'agit de :

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

appelé coefficient d'asymétrie (skewness). L'autre est :

$$\gamma_2 = \frac{E[(x - \mu)^4]}{\sigma^4}$$

appelé coefficient d'aplatissement (kurtosis).

### 2.4 Fonction caractéristique

Il s'agit de la transformée de Fourier inverse de la fonction de densité de probabilité :

$$\begin{aligned} \phi(t) &= E[e^{itx}] = \int_{-\infty}^{+\infty} e^{itx} f(x) dx \\ &\text{ou} \left( \sum_{k=-\infty}^{+\infty} e^{itx_k} f(x_k) \text{ cas discret} \right) \end{aligned}$$

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \phi(t) dt$$

$$TF^{-1} \quad \longleftrightarrow \quad TF$$

$$\phi(t) = E \left[ \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \right]$$

$$= \sum_{k=0}^{\infty} \frac{(it)^k}{k!} E[x^k]$$

$$E[x^k] = \frac{1}{i^k} \frac{d^k \phi(t)}{dt^k} \Big|_{t=0}$$

### Cas discret

Dans le cas où il s'agit d'une variable discrète, on parle d'une fonction génératrice  $G(z) : \phi(t) \rightarrow G(z)$

*Application*

- Trouver  $f(x)$  à partir des données expérimentales :

Réponse : on "estime" les différents moments ce qui nous conduit à déterminer une forme approchée de  $\phi(t)$  et puis en utilisant la formule précédente on espère trouver une forme approchée de  $f(x)$ .

- Trouver la fonction de densité de probabilité  $f$  d'une variable aléatoire qui est la somme de deux autres variables aléatoires  $w = x + y$

Réponse :  $\phi(t) = E[e^{iwt}] = E(e^{itx} e^{ity})$

si  $x, y$  indépendants  $\Rightarrow E(e^{iwt}) = E(e^{ixt}) E(e^{iyt})$

$\Rightarrow \phi_w(t) = \phi_x(t) \phi_y(t)$ . Une fois  $\phi_w(t)$  déterminé, on peut obtenir  $f(x+y)$  en prenant la transformée de Fourier inverse de  $\phi$ .

## 3 Cas de plusieurs variables aléatoires réelles

Si  $X, Y$  sont deux variables aléatoires définies sur le même espace probabilisé, le couple  $(X, Y)$  est une application de  $(\Omega; \mathcal{C}; \mathbb{P})$  dans  $\mathbf{R}^2$ .

### 3.1 Fonction de densité conjointe

$f(x, y)$  est définie de la manière suivante :

$$P(x < X < x + dx, y < Y < y + dy) = f(x, y) dx dy$$

### 3.2 Fonction de densité marginale

On s'intéresse à l'une ou à l'autre des 2 variables seulement

$$\begin{aligned} f_1(x) &= \int f(x, y) dy \\ f_2(y) &= \int f(x, y) dx \end{aligned}$$

### 3.3 Fonction de densité conditionnelle

$$f_c(y/x) = \frac{f(x, y)}{f_1(x)}$$

### 3.4 Notion de corrélation

L'étude des données expérimentales peuvent être effectuée à l'aide de plusieurs variables. Afin de mieux caractériser les données à travers ces variables il est naturel de se poser la question si ces variables sont corrélées et sinon quel est le degré de corrélation mutuelle. La corrélation entre deux variables est représentée par la notion de la covariance :

**Covariance**

La covariance de 2 variables  $X, Y$  est définie par :

$$\begin{aligned} COV(X, Y) &= E[(X - \mu_x)(Y - \mu_y)] \\ &= E[XY] - \mu_x \mu_y \end{aligned}$$

avec

$$\begin{aligned} \mu_x &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dx dy \\ \mu_y &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dx dy \end{aligned}$$

A partir de cette notion on introduit le coefficient de corrélation.

**Coefficient de corrélation**

On appelle coefficient de corrélation la quantité :

$$\rho_{XY} = \frac{COV(X, Y)}{\sigma_x \sigma_y} \quad ; -1 \leq \rho_{XY} \leq 1$$

**Indépendance de 2 variables aléatoires**

La notion de corrélation entre deux variables n'est pas identique à la notion de dépendance. Il s'agit là d'un point important qui nécessite d'être rappelé. En effet si deux variables sont indépendantes, elles sont forcément décorrélées. Par contre deux variables décorrélées ne sont pas forcément indépendantes.

En effet, pour deux variables indépendantes  $X, Y$  on a  $f(x, y) = f(x)f(y)$  ce qui conduit à :

$$\begin{aligned} COV(X, Y) &= \int (x - \mu_x) \int (y - \mu_y) f(x, y) dx dy \\ &= \int (x - \mu_x) f(x) \int (y - \mu_y) f(y) dy = 0 \end{aligned}$$

Mais la covariance peut s'annuler sans que  $f(x, y) = f(x)f(y)$ . Ceci a lieu lorsque la variation de l'une des deux variables ne change pas la valeur moyenne de l'autre.

*Remarque :*

Si deux variables  $X, Y$  sont corrélées ( $COV(X, Y) \neq 0$ ) on peut toujours faire un changement de variables :

$$X, Y \Rightarrow X', Y' \quad \text{telle que} \quad COV(X', Y') = 0$$

avec

$$X' = X'(X, Y), \quad Y' = Y'(X, Y) \Rightarrow g(X', Y') = f(x, y)J$$

où nous avons introduit le Jacobien  $J$

$$J = \begin{vmatrix} \frac{\partial X}{\partial X'} & \frac{\partial Y}{\partial X'} \\ \frac{\partial X}{\partial Y'} & \frac{\partial Y}{\partial Y'} \end{vmatrix}$$

**Généralisation à plusieurs variables**

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

$$F(\underline{x}) = \int_{-\infty}^{+x_1} \int_{-\infty}^{+x_2} \cdots \int_{-\infty}^{x_n} f(\underline{x}) d\underline{x} \quad ; \quad d\underline{x} = dx_1 dx_2 \cdots dx_n$$

$$f(\underline{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(\underline{x})$$

$$\mu_{\ell_1, \ell_2, \dots, \ell_n} = E(x_1^{\ell_1}, x_2^{\ell_2}, \dots, x_n^{\ell_n})$$

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

$$\underline{V} = E[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^T] = \begin{pmatrix} \sigma_{11} & \sigma_{12} \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \sigma_{2n} \\ \vdots & & \\ \sigma_{n1} & & \sigma_{nn} \end{pmatrix}$$

$$\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j$$

$V$  est une matrice appelée matrice de corrélation. Il s'agit d'une matrice symétrique et donc diagonalisable :

$$\Rightarrow \exists U \text{ tel que } g = U x \text{ avec } V(g) = \text{diag.}$$

## 4 Les distributions de probabilité

Nous venons de voir les variables aléatoires discrètes et continues. Nous sommes souvent amenés dans le cadre de l'analyse des données expérimentales à étudier la distribution des variables physiques issues des données. Les distributions que l'on peut rencontrer en physique subatomique sont nombreuses. Il en existe cependant un certain nombre dont la connaissance est indispensable à la compréhension des phénomènes physiques et leur interprétation. Nous allons citer ces quelques distributions en donnant leurs caractéristiques les plus importantes.

### 4.1 Distribution de Bernoulli

Il s'agit de la distribution discrète la plus simple. Elle donne la probabilité d'une variable  $X = k$  qui ne peut prendre comme valeurs que 1 ou 0 avec une probabilité  $p$  d'avoir la valeur 1 :

$$\begin{aligned} f(k, p) &= p^k (1-p)^{1-k} \\ E[k] &= p \\ V[k] &= p(1-p) \end{aligned}$$

### 4.2 Distribution binômiale

Il s'agit d'une distribution d'une variable discrète  $X = k$ . Cette distribution décrit la probabilité d'avoir  $k$  tirages favorables lors de  $n$  essais avec une probabilité  $p$  de succès pour chaque tirage :

$$\begin{aligned} f(n, k, p) &= B(k; n; p) = C_n^k p^k (1-p)^{n-k} \\ E[k] &= np \\ V[k] &= np(1-p) \\ \phi(t) &= (pe^{it} + 1 - p)^n \end{aligned}$$

*Exercice :*

Montrer que la somme de deux variables indépendantes  $X, Y$  suivant chacune une distribution binomiale ( $B(k_x; n_x, p)$  et  $B(k_y; n_y, p)$  respectivement) est une variable dont la distribution est à son tour binomiale et vaut  $B(k_x + k_y, n_x + n_y, p)$ .



### 4.3 Distribution multinomiale

Il s'agit d'une généralisation de la distribution binômiale avec  $m$  résultats possibles au lieu de 2 dans le cas de la distribution binômiale. La distribution multinomiale peut être schématisée par le tableau suivant :

résultats	nb	prob.
1	$k_1$	$p_1$
2	$k_2$	$p_2$
...	...	...
m	$k_m$	$p_m$

avec

$$f(k_1, k_2, \dots, k_m, p_1, p_2, \dots, p_m, n) = \frac{n!}{k_1! k_2! \dots k_m!} \times p_1^{k_1} p_2^{k_2} \dots p_m^{k_m}$$

et on a alors :

$$E[k_i] = np_i$$

$$V[k_i] = np_i(1 - p_i)$$

$$\text{CONV}(k_i, k_j) = -np_i p_j \quad i \neq j$$

$$\phi(t_2, t_3 \dots t_m) = (P_1 + P_2 e^{it_2} + \dots P_m e^{it_m})^n$$

*Application :*

- 1) La désintégration d'une particule avec plusieurs modes possibles ;
- 2) Les bins d'un histogramme.

### 4.4 Distribution de Poisson

C'est une des plus intéressantes avec la distribution normale. Il s'agit d'une distribution discrète qui se réalise dans les conditions suivantes :

- le nombre de succès est connu mais pas le nombre d'essais,
- le nombre de succès (événements) dans un intervalle donné ne dépend que de la longueur de cet intervalle,
- 2 événements ne peuvent pas avoir lieu en même temps (assez rare pour exclure toute coïncidence).

Soit  $\Delta x$  un intervalle dans lequel on peut, au plus, avoir un événement.

Soit  $\lambda$  la probabilité d'avoir 1 événement dans  $n\Delta x \Rightarrow$  la probabilité d'avoir 1 événement dans  $\Delta x$  sera  $p = \frac{\lambda}{n}$ .

Calculons la probabilité d'avoir  $K$  événements dans les  $n \gg k$  intervalles ( $n\Delta x$ ). Il est aisé, tenant compte des conditions citées auparavant, de voir que cette probabilité suit une distribution binômiale :

$$P(k, p) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

or

$$k \ll n \quad (\text{év. rare}) \quad \Rightarrow \quad \frac{n!}{(n-k)!} \simeq n^k$$

$$p = \frac{\lambda}{n} \Rightarrow \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}$$

$$\rightarrow f(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$\begin{aligned} E[k] &= \lambda \\ V[k] &= \lambda \\ \gamma_1 &= \lambda^{-\frac{1}{2}} \\ \gamma_2 &= \lambda^{-1} + 3 \\ \phi(t) &= \exp(\lambda(e^{it} - 1)) \end{aligned}$$

*Exercice :*

Montrer que la somme de deux variables  $X, Y$  de Poisson est une variable de Poisson selon le schéma suivant :

$$\begin{aligned} X &\rightarrow \frac{\lambda_x^{K_x}}{K_x!} e^{-\lambda_x} \\ Y &\rightarrow \frac{\lambda_y^{K_y}}{K_y!} e^{-\lambda_y} \\ X + Y &\rightarrow \frac{(\lambda_x + \lambda_y)^K}{K!} e^{-(\lambda_x + \lambda_y)} \end{aligned}$$

*Application :*

- 1) La radioactivité;
- 2) L'interaction d'un faisceau à faible flux avec une cible épaisse..

## 4.5 La distribution uniforme

Il s'agit de la loi continue la plus simple :

$$\begin{aligned} f(x; a, b) &= \frac{1}{b-a} & \text{si } x \in [a, b] \\ f(x; a, b) &= 0 & \text{sinon} \end{aligned}$$

$$\begin{aligned} E[x] &= \frac{b+a}{2} \\ V[x] &= \frac{(b-a)^2}{12} \end{aligned}$$

*Application :*

- 1) Les arrondis d'un calcul;
- 2) L'abscisse du point d'impact d'un photon sur un pixel.

## 4.6 La distribution normale (De Moivre, Laplace, Gauss)

C'est la loi de distribution la plus répandue. Nous allons voir plus tard pourquoi cette loi joue un rôle essentiel dans les études statistiques.

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

$$\begin{aligned} E[x] &= \mu \\ V[x] &= \sigma^2 \end{aligned}$$

En introduisant le changement de variable suivant :

$$x \rightarrow z = \frac{x - \mu}{\sigma}$$

On aboutit à la distribution dite normale standard centrée à zéro et de largeur égale à l'unité :  $N(x; \mu, \sigma) \rightarrow N(z; 0, 1)$ , dont la fonction caractéristique est donnée par :

$$\phi(t) = e^{-\frac{t^2}{2}}$$

#### 4.7 Distribution normale de $n$ variables

$$x \rightarrow \underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \mu \rightarrow \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

$$-\frac{(x-\mu)^2}{2\sigma^2} \rightarrow -\frac{1}{2}(\underline{x}-\underline{\mu})^T A(\underline{x}-\underline{\mu})$$

on utilise  $E[\underline{x} - \underline{\mu}] = 0$

$$\int (\underline{x} - \underline{\mu}) e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T A(\underline{x}-\underline{\mu})} d\underline{x} = 0$$

On différencie par rapport à  $\underline{\mu}$

$$\Rightarrow E[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^T] A = \mathbf{1}$$

$$\Rightarrow V A = \mathbf{1} \Rightarrow A = V^{-1}$$

$$N(\underline{x}; \underline{\mu}; V) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} \exp \left[ -\frac{1}{2}(\underline{x} - \underline{\mu})^T V^{-1}(\underline{x} - \underline{\mu}) \right]$$

où  $|V|$  est le déterminant de la matrice  $V$

**Application : Distribution normale à 2 dimensions  $(X, Y)$**

$$\underline{V} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \quad \text{COV}(x, y) = \rho\sigma_x\sigma_y$$

$$\underline{V}^{-1} = \frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)} = \begin{pmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix}$$

$$\Rightarrow f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2}G}$$

$$G = \frac{1}{1-\rho^2} \left[ \frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right]$$

*Remarque :* Les contours de densité de probabilité à valeur donnée dans le pln  $(X, Y)$  sont obtenus en fixant  $G$ . Il s'agit évidemment d'ellipses.

#### 4.8 Distribution de Cauchy

$$C(x, \mu, \alpha) = \frac{1}{\pi\alpha} \frac{1}{1 + (x - \mu)^2/\alpha^2}$$

Cette distribution est connue en physique sous le nom de Breit-Wigner. Elle est donnée sous la forme :

$$f(m; M, \Gamma) = \frac{1}{2\pi} \frac{\Gamma}{(m - M)^2 + \frac{\Gamma^2}{4}}$$

*Remarque :* tous les moments de  $C$  tels que  $E[x], \dots$  sont divergents. En pratique, on tronque la distribution en se limitant entre  $-L$  et  $+L$  en choisissant  $L \gg \alpha$ .

## 4.9 La distribution $\chi^2$

Soient  $x_1, x_2, x_3 \dots x_n$ ,  $n$  variables indépendantes et distribuées normalement, alors on définit

$$\chi^2(n) = \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

où les  $\mu_i, \sigma_i^2$  sont les espérances et les variances associées à  $x_i$ .

Cas de  $n=1$  :  $\chi^2(1)$

$$\chi^2 = \left( \frac{x - \mu}{\sigma} \right)^2 = z^2$$

$$Q = \chi^2 = z^2$$

$$Q = z^2$$

La fonction de distribution de  $z$  (distribuée normalement) est donnée par :

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Quelle est la fonction de distribution  $g(Q)$  associée à  $Q$  ?

$$P(Q < q) = P(-\sqrt{q} < z < \sqrt{q})$$

$$G(Q) = F(\sqrt{q}) - F(-\sqrt{q})$$

$$g(Q) = \frac{f(\sqrt{q})}{2\sqrt{q}} + \frac{f(-\sqrt{q})}{2\sqrt{q}}$$

$$\chi^2(1) \rightarrow g(Q) = \frac{1}{\sqrt{2\pi q}} e^{-\frac{q}{2}} = \frac{1}{\sqrt{2\pi \chi^2}} e^{-\frac{\chi^2}{2}}$$

Ce résultat se généralise au cas de  $n$  quelconque :

$$\chi^2(n) \rightarrow g(\chi^2, n) = \frac{(\chi^2)^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2}) 2^{n/2}} e^{-\frac{\chi^2}{2}}$$

$$E[\chi^2(n)] = n$$

$$V[\chi^2(n)] = 2n$$

$$\phi(t) = (1 - 2it)^{-n/2}$$

*Exercice :*

Montrer que la somme d'une distribution  $\chi^2(n_1)$  et d'une distribution  $\chi^2(n_2)$  est une distribution  $\chi^2(n_1 + n_2)$ .

## 4.10 La distribution student

La variable  $t = \frac{z}{\sqrt{\frac{\chi^2(n)}{n}}}$  suit la distribution student si la variable  $z$  est distribuée normalement :

$$f(t, n) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{(1 + \frac{t^2}{n})^{\frac{n+1}{2}}}$$

Il s'agit d'une distribution particulière qui s'avère très utile pour l'estimation de la variance d'une distribution normale dont la valeur centrale est inconnue.

## 5 Théorème central-limite

Soient  $X_1, X_2 \dots X_n, n$  variables indépendantes aléatoires d'espérance  $m_i$  et de variance  $\sigma_i^2$  suivant les lois  $f_i$  (pas forcément identiques).

Soit  $S_n^2 = \sum_{i=1}^n \sigma_i^2$  et  $F_i$  la fonction de répartition de  $(X_i - m_i)$  alors

$$\sum_{i=1}^n \frac{(X_i - m_i)}{S_n} \text{ tend vers } U = N(X; 0, 1)$$

si la condition de Lindeberg-Cramer est satisfaite :

$$\lim_{n \rightarrow \infty} \left[ \frac{1}{S_n^2} \sum_{i=1}^n \int_{|x| > \epsilon S_n} x^2 dF_i(x) \right] = 0$$

Cette condition peut se traduire par le fait qu'aucune variable ne domine les autres et que toutes les variables soient "uniformément petites".

### Démonstration dans un cas spécial :

Si les  $X_i$  sont identiques avec comme espérance  $m$  et variance  $\sigma^2$  alors

$$\frac{1}{\sqrt{n}} \left( \frac{X_1 + X_2 + \dots + X_n - nm}{\sigma} \right) = \frac{1}{\sqrt{n}} \sum_i \frac{(X_i - m)}{\sigma} = \sum_i \frac{X_i - m}{\sqrt{n}\sigma}$$

est la somme de  $n$  variables centrées (espérance nulle) et de variance  $= \frac{1}{n}$ . Chacune possède une fonction caractéristique  $\phi_{X_i}(t) = E(e^{itX_i}) = 1 + 0 - \frac{t^2}{2n} + o\left(\frac{1}{n^2}\right)$ . La fonction caractéristique de la somme de ces variables indépendantes est  $\phi_X(t) = \prod_{i=1}^n \phi_i = \left(1 - \frac{t^2}{2n}\right)^n$ . Si  $n \rightarrow \infty$ , on a alors :  $\phi_X(t) \rightarrow \exp\left(-\frac{t^2}{2}\right)$  qui est rien d'autre que la fonction caractéristique de  $N(X, 0, 1)$ .

## 6 Techniques de simulation - Monte Carlo

Il est de plus en plus fréquent de recourir à des méthodes de simulation pour étudier la validité d'une nouvelle expérience avant sa réalisation. L'optimisation d'un tel ou tel outil peut également se faire grâce à des méthodes de simulation. Dans tous les cas cela permet de limiter les coûts en éliminant les projets dont la réalisation ne permet pas d'obtenir les résultats escomptés. Il ne faut cependant pas que l'utilisation massive des méthodes de simulation cache la limitation de telles méthodes. Il devient ainsi important de connaître et même de maîtriser les techniques utilisées dans ces méthodes afin de mieux les exploiter et d'éviter en même temps une utilisation inappropriée.

### 6.1 Nombres aléatoires

La génération de nombres aléatoires est une opération très courante lorsque l'on veut simuler des phénomènes dans lesquels, des variables aléatoires sont présentes. Citons la génération de points d'interaction d'un faisceau de particules dans une cible mince. Chacune des coordonnées de chacun de ces points est un nombre aléatoire. Il est possible de générer les nombres aléatoires à l'aide des phénomènes physiques aléatoires.

*Exemple :*

Afin de générer des nombres entiers aléatoires, on mesure pendant un temps fixe  $t$ , le nombre de cosmiques dans un détecteur. Le temps doit être assez large pour en avoir une quantité importante :

Si le nombre est pair  $\Rightarrow bit = 1$

Si le nombre est impair  $\Rightarrow bit = 0$

On rajoute les bits les uns à coté des autres jusqu'à l'obtention du nombre recherché de chiffres (ex. 11010011110).

L'opération peut être répétée autant de fois que nécessaire. Cette procédure peut être biaisée si pour une raison quelconque le nombre de désintégrations paires et celui du nombre de désintégrations impaires sont différents. On élimine ce biais en prenant deux mesures successives : si les deux donnent des bits identiques la solution est rejetée, sinon on prend le 2ème bit.

### Nombres pseudo-aléatoires

Les méthodes basées sur les phénomènes naturelles ont l'inconvénient d'exiger un temps de mesure très long. Ceci rend leur utilisation exclue lorsqu'un grand nombre de nombres aléatoires est nécessaire. Des nombres dits pseudo-aléatoires remplacent dans la pratique les nombres aléatoires. Leur génération fait appel à des méthodes basées sur des suites récurrentes et donc malheureusement périodiques. Parmi ces méthodes celle de Lehmer est la plus utilisée. Elle est résumée par la relation récurrente suivante :

$$r_{i+1} = a r_i \quad \text{modulo } m$$

On fixe  $a$  et  $m$  en choisissant  $m$  le plus large possible (période de la suite  $\simeq \frac{m}{4}$  pour un choix approprié de  $a$ ).

On commence par  $r_0 \rightarrow r_1 \rightarrow r_2 \dots$ . A partir de cette suite de nombres on obtient les nombres réels  $\frac{r_i}{m-1}$  "aléatoire"  $\in [0, 1]$ .

*Remarque :* Il est nécessaire de penser à changer la valeur de  $r_0$  pour chaque tirage lorsque l'on produit des nombres aléatoires car dans le cas contraire on risque de reproduire les mêmes nombres et donc les mêmes configurations lors de deux tirages différents sans se rendre compte.

## 6.2 Tirage d'un échantillon d'une variable $X$

### Inversion de la fonction de répartition

Cette méthode s'applique lorsque la fonction de répartition est inversible et  $F^{-1}$  possède une forme analytique simple. Dans ce cas la variable  $Y = F(X)$  est uniformément distribuée dans  $[0, 1]$  car la fonction de densité associée à cette nouvelle variable est donnée par :

$$g(y) = \frac{f(F^{-1}(y))}{F'(F^{-1}(y))} = 1$$

Ainsi, en choisissant  $n$  nombre aléatoires  $r_1, r_2 \dots r_n$  dans l'intervalle  $[0, 1]$  nous pouvons obtenir un échantillon de  $n$  nombre aléatoires de  $X$  en prenant  $F^{-1}(r_i)$ .

### Méthode de rejet

Cette méthode s'applique lorsque la densité de probabilité de  $X$  est à support borné et reste finie. On suppose que  $0 \leq X \leq 1$  pour simplifier.

soit  $M$  un nombre réel tel que  $\forall X \in [0, 1] \quad f(X) \leq M$

- 1) On tire au hasard un nombre  $U$  entre 0 et 1 (d'une façon uniforme) ;
- 2) On tire un nombre aléatoire  $V$  entre 0 et  $M$  ;
- 3) On compare  $f(U)$  et  $V$ , si  $f(U) \geq V$  on garde  $U$  sinon on le rejette

*Applications :*

- 1) Génération d'une distribution gaussienne  $N(X, \mu, \sigma)$  : On va utiliser le théorème central-limite : la somme des variables aléatoires uniformément distribuées sur l'intervalle  $[0, 1]$  est une variable gaussienne si  $n \nearrow (n \geq 12)$

$$X = \mu + \sigma \left[ \sum_{i=1}^{12} r_i - 6 \right]$$

en fait

$$\sum_{i=1}^{12} r_i \longrightarrow N(r, 12 \times \frac{1}{2}, 12 \times \frac{1}{12}) = N(r, 6, 1)$$

2) Génération d'une distribution de Poisson : On engendre des variables aléatoires uniformes :

$$r_i \in [0, 1] \Rightarrow \Gamma_i = -\ln(r_i)$$

La réalisation  $n$  de la loi de Poisson est le plus grand nombre  $n$  tel que  $\sum_{i=1}^n \Gamma_i < \lambda$

### 6.3 Méthodes d'intégration

Les techniques de simulation peuvent être mises à contribution dans l'estimation des intégrales dont la forme ne permet pas une estimation analytique. Nous connaissons des méthodes numériques d'intégration dont la plus élémentaire est représentée par :

$$I = \int_a^b f(x)dx \Rightarrow \frac{b-a}{n} \sum_i f(x_i)$$
$$x_i = a + (i-1) \frac{b-a}{n} \quad i = 1, \dots, n$$

Une méthode équivalente à cette intégration numérique est donnée par :

$$\frac{b-a}{n} \sum_i f(x_i)$$
$$x_i = a + r_i(b-a)$$
$$r_i \in [0, 1]$$

On peut s'interroger sur l'avantage de cette méthode par rapport aux méthodes numériques.

Il est clair que cette méthode ne représente pas d'avantage s'il s'agit d'une intégration à une dimension. En effet, la résolution obtenue dans le cas d'une intégration numérique à  $d$  dimension en utilisant  $n$  points est proportionnelle à  $n^{-2/d}$  tandis que celle obtenue avec les méthodes utilisant une sélection aléatoires est toujours proportionnelle à  $n^{-1/2}$ .

L'intérêt de la méthode précédente se manifeste dans le cas d'une intégration à  $d$  dimensions avec  $d > 4$ . De plus si les bornes d'intégration d'une variable dépendent d'une autre variable, l'intégration numérique devient très compliquée contrairement aux méthodes aléatoires.

### 6.4 Méthodes de réduction de la variance

Il existe des techniques qui permettent d'améliorer la résolution obtenue avec les méthodes d'intégration basées sur la simulation. Nous donnons ici quelques techniques parmi les plus utilisées.

**Stratification :** Il s'agit de diviser l'intervalle d'intégration en 2 et de choisir la moitié des points générés dans un et l'autre dans l'autre moitié ce qui a pour but de mieux répartir les points.

**Echantillonnage non-uniforme :** Dans cette technique on choisit de générer plus de points là où l'intégrand varie rapidement ce qui implique une stratification avec nombre de points différent dans chaque moitié et plus généralement dans les différents intervalles choisis.

**Echantillonnage proportionnel "Importance Sampling" :** La variance dépend de l'intégrand, réduire la fluctuation de l'intégrand conduit à réduire la variance de l'intégrale.

$$\int f(x)dx = \int \frac{f(x)}{g(x)} g(x) dx = \int h(y) dy$$

On choisit  $g$  telle que  $f(x)/g(x)$  varie le moins possible. Il faut pour cela trouver une fonction analytique qui ressemble au mieux à la fonction  $f$ . En effet dans le cas limite où  $g \equiv f$ , l'intégrand vaut 1 et la variance est nulle.

**Symétrisation :** On intègre une fonction quelconque  $f$  sur un intervalle  $[a, b]$ . A chaque point généré  $x_i \in [a, b]$  on associe un autre point  $x'_i$  qui est son symétrique par rapport à  $\frac{b-a}{2}$ , cela permet de supprimer la partie impaire par rapport à  $\frac{b-a}{2}$ .

*Remarque :*

Pour les intégrations à plusieurs dimensions pour lesquelles les bornes d'une variable dépendent des autres, il faut faire attention à bien répartir les points générés.

*Exemple*  $I = \int_0^1 dy \int_0^y dx f(x, y)$

*1ère méthode :* générer  $x$  uniformément dans  $[0, 1]$  et générer  $y$  uniformément dans  $[0, x]$ .

*2ème méthode :* générer  $x \in [0, 1]$  et  $y \in [0, 1]$  et rejeter ensuite les points tels que  $y_i > x_i$ .

Dans la deuxième méthode le nombre de points utilisés pour estimer l'intégrale est inférieur à celui de la première méthode. Cela ne veut pas dire qu'une meilleure précision est obtenue avec la première méthode car la densité des points dans cette méthode n'est pas la même partout. Cette densité est plus élevée pour les valeurs de  $x$  proches de zéro que pour celles proches de 1.

## Références

- [1] Probabilités, Analyse des données et Statistique, G.Saporata, Edition Technip.
- [2] Statistics for nuclear and particle physics, L.Lyons, Cambridge University Press
- [3] Statistical Methods in Data Analysis, W.J. Metzger