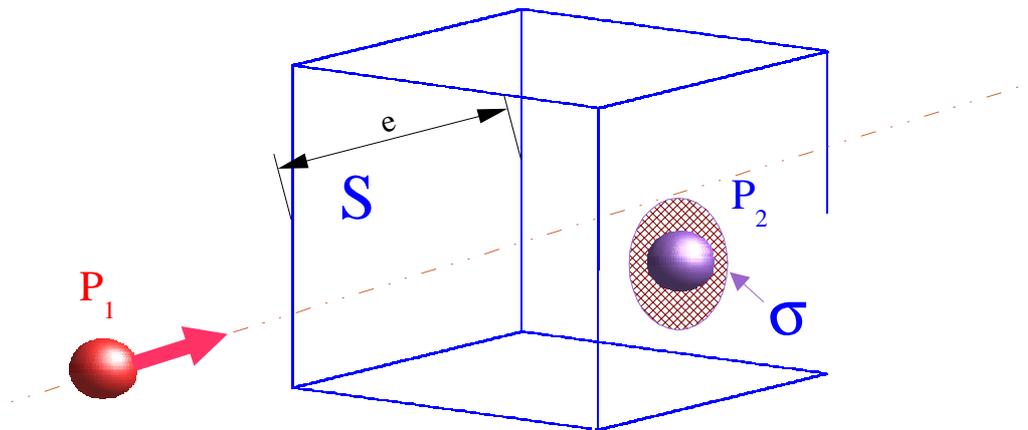


## Erreur , Probabilité et Statistique

Tous les processus étudiés en physique des hautes énergies sont, dans leurs aspects fondamentaux, régis par les lois de la mécanique quantique relativiste. Il en découle que les prévisions de mesures des quantités observables sont de nature stochastique.

Citons quelques exemples simples :

### I. Diffusion d'une particule $p_1$ sur une boîte contenant des particules $p_2$



La section efficace du processus  $p_1 p_2 \rightarrow X$  est proportionnelle à l'intégrale sur l'espace de phase ouvert du module au carré de l'élément de matrice invariant, c'est à dire :

$$\sigma(p_1 p_2 \rightarrow X) \propto \int_{\text{espace de phase}} |M|^2 d\Phi$$

$|M|^2$  est une densité de probabilité définie sur l'espace de phase ( $|M|^2 \propto d\sigma/d\Phi$ ). Si la boîte ne contient qu'une seule particule  $p_2$ , la probabilité d'occurrence de la réaction  $p_1 p_2 \rightarrow X$  s'exprime par :

$$p = \frac{\sigma(p_1 p_2 \rightarrow X)}{S},$$

où  $S$  est la surface de la face d'entrée de la boîte. Si maintenant la boîte est constituée d'un élément ordinaire et qu'on s'intéresse à la diffusion de  $p_1$  sur des noyaux  $p_2$ , la probabilité d'observer une réaction  $p_1 p_2 \rightarrow X$  devient :

$$P = p \frac{e \rho S}{M} N,$$

où :  $e$  est l'épaisseur de la boîte;

$\rho$  est la masse volumique du milieu;

$M$  est la masse atomique de l'élément;

$N$  est le nombre d'Avogadro.

## II. Décroissance d'une particule instable

Ce processus est caractérisé par une durée de vie moyenne  $\tau$  qui est l'inverse de la largeur  $\Gamma$  de la résonance :

$$\tau = \Gamma^{-1}$$

$$\text{avec, } \Gamma \propto \sum_i^{\text{branchements}} \int_{\text{espace de phase}} |M_i|^2 d\Phi_i$$

On voit de nouveau apparaître dans l'expression de  $\Gamma$  des éléments de matrices invariants qui confèrent au processus de décroissance d'une particule instable une nature stochastique. La densité de probabilité s'exprime par :

$$P(t) dt = \frac{1}{\tau} dt .$$

Soit pour  $N_0$  (à  $t=0$ ) particules instables,

$$dN(t) = -N(t) \frac{1}{\tau} dt ,$$

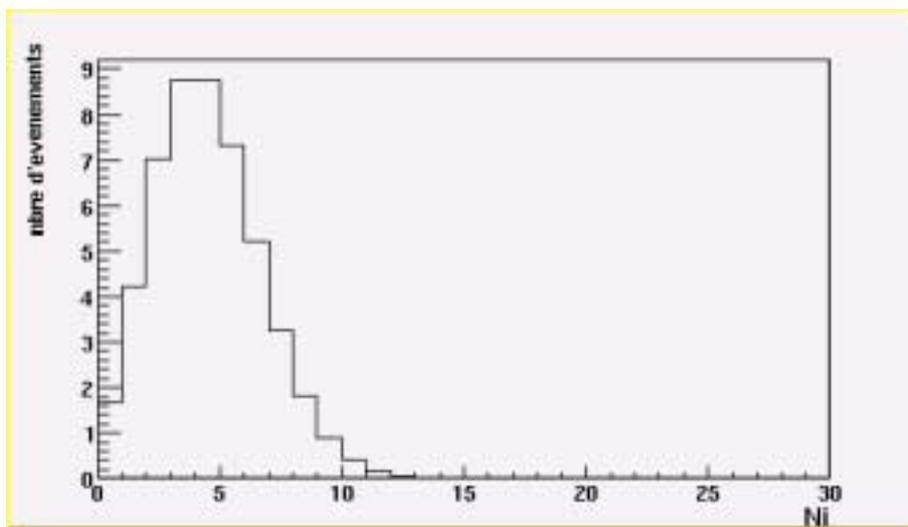
$$N(t) = N_0 e^{-t/\tau} ,$$

qui donne la population des particules restantes à l'instant  $t$ .

Si maintenant on considère  $N_0$  particules instables à  $t=0$  que l'on observe sur un intervalle de temps  $\Delta t$  petit devant  $\tau$  ( $\Delta t/\tau \ll 1$ ), le nombre de décroissances prédit sera :

$$N_p = \frac{N_0}{\tau} \Delta t .$$

Il s'agit d'une prédiction. Si on réalise l'expérience, il est fort probable qu'on obtiendra une valeur proche mais différente de  $N_p$ . Si on réalise un grand nombre de fois cette même expérience, on obtiendra une courbe similaire à la figure suivante :



qui représente le nombre d'observations ( nombre d'événements ) de  $N_i$  décroissances parmi la totalité des  $N$  expériences réalisées . L'ensemble des expériences réalisées ici est appelé un échantillon. La taille de cet échantillon est le nombre  $N$  d'expériences réalisées . La fréquence d'échantillonnage est définie comme étant :

$$f(N_i) = \frac{1}{N} n_i ,$$

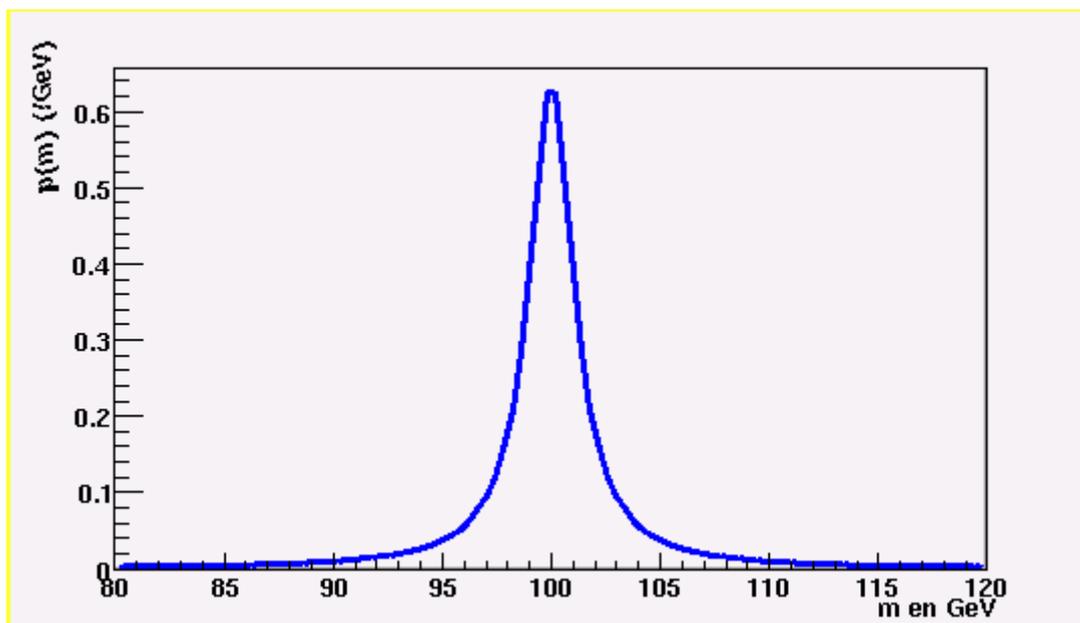
où  $n_i$  est le nombre d'observations de  $N_i$  décroissances. On a évidemment:  $\sum_i f(N_i) = 1$  .

La probabilité d'observation ( fonction de probabilité ) est égale à la fréquence d'échantillonnage lorsque la taille de l'échantillon tend vers l'infini, ou encore (c'est également ce que l'on appelle la loi des grands nombres):

$$p(N_i) = \lim_{N \rightarrow \infty} f(N_i) = \lim_{N \rightarrow \infty} \frac{n_i}{N}$$

C'est par conséquent une entité théorique qui n'est jamais rigoureusement accessible expérimentalement .  $N_i$  est appelée une variable aléatoire ( discrète dans cet exemple ).

En physique on rencontre également des variables aléatoires continues : par exemple, la masse invariante d'une particule instable.



Lorsqu'il s'agit d'une variable aléatoire continue, la fonction de probabilité est remplacée par une densité de probabilité (  $\sum \rightarrow \int$  ) .

Puisqu'il est impossible d'obtenir un échantillon d'événements ayant une taille infinie, les caractéristiques d'un processus physique aléatoire ne sont jamais exactement accessibles par l'expérience. Ainsi, toute mesure réalisée sur un échantillon fini d'événements sera entachée d'une incertitude que l'on désigne sous le nom d'erreur statistique . On peut aisément admettre que l'erreur statistique doit décroître lorsque la taille de l'échantillon croît et doit s'annuler lorsque celle-ci est infinie.

À cette incertitude (erreur) statistique, on doit en général ajouter une autre erreur que l'on appelle l'erreur systématique (ou instrumentale) qui est le reflet de l'exactitude de la méthode de mesure. Une mesure peut être très précise statistiquement mais peut s'avérer inexacte et donc nécessiter une correction qui est elle-même connue avec une précision limitée (erreur systématique).

Si on reprend l'exemple des  $N_o$  particules instables observées sur un temps  $\Delta t$ , et que l'on réalise l'expérience qui consiste à compter les décroissances avec un détecteur réel, il faudra prendre en compte son efficacité ( $\epsilon \leq 1$ ) de détection pour remonter au nombre de décroissances vraies. Une estimation de  $N_p$  sera obtenue par :  $N_p = \frac{\langle N_i \rangle}{\epsilon}$  ( $\langle \rangle$  signifie moyenne), dans laquelle, l'erreur sur  $\langle N_i \rangle$  sera de nature statistique et l'erreur sur  $\epsilon$  (mesurée ou déterminée par calcul ou simulation) sera de nature systématique ou instrumentale.

## 1 Caractéristiques d'une variable aléatoire :

Les variables aléatoires, dont il sera question ici, sont les prédictions théoriques de mesures des observables physiques: masse d'une particule instable, nombre d'événements résultant de la décroissance de  $N_o$  particules instables (à  $t=0$ ) sur  $\Delta t$  ...

D'une manière générale, une variable aléatoire  $x$  est en premier lieu caractérisée par sa densité de probabilité associée  $f(x)$ , telle que :

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad ,$$

ou pour une variable discrète ,

$$\sum_{i=1}^j f(x_i) = 1 \quad \text{avec} \quad \{x_i, i=1, j\} \quad \text{étant l'ensemble des valeurs possibles de } x \quad .$$

### 1.1 Fonction de répartition (cumulative) :

La fonction de répartition d'une variable aléatoire  $x$  est la probabilité que  $x$  soit inférieure à une valeur  $a$  :

$$F(a) = \int_{-\infty}^a f(x) dx \quad ,$$

$$\text{ou pour une variable discrète} \quad F(a) = \sum_{i=1}^n f(x_i) \quad \text{avec} \quad x_1 \leq x_2 \leq x_3 \leq \dots \leq x_n \leq a$$

### 1.2 Espérance mathématique :

L'espérance mathématique d'une fonction  $u(x)$  d'une variable aléatoire  $x$  est définie comme étant :

$$E(u(x)) = \int_{-\infty}^{+\infty} u(x) f(x) dx \quad ,$$

$$\text{ou, pour une variable discrète,} \quad E(u(x)) = \sum_{i=1}^j f(x_i) u(x_i)$$

### 1.3 Moments :

Le moment d'ordre  $n$  d'une distribution d'une variable aléatoire  $x$  s'exprime par :

$$\alpha_n = E(x^n) = \int_{-\infty}^{+\infty} x^n f(x) dx$$

### 1.4 Moyenne :

La moyenne  $\mu$  d'une variable aléatoire  $x$  n'est pas autre chose que son moment du premier ordre :

$$\mu = \alpha_1 = E(x) = \int_{-\infty}^{+\infty} x f(x) dx$$

### 1.5 Moments centrés :

Le moment centré d'ordre  $n$  d'une variable aléatoire  $x$  est défini par :

$$m_n = E((x-\mu)^n) = \int_{-\infty}^{+\infty} (x-\mu)^n f(x) dx$$

### 1.6 Variance et écart type :

La variance d'une variable aléatoire  $x$  est le carré de son écart type. C'est également le moment centré du 2<sup>ème</sup> ordre de  $x$  :

$$\begin{aligned} \text{var}(x) = \sigma^2(x) = m_2 &= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx \\ &= \alpha_2 - \mu^2 \end{aligned}$$

La moyenne d'une variable aléatoire  $x$  donne la position du centroïde de la densité de probabilité associée, alors que son écart type est une estimation de sa largeur.

### 1.7 Distributions à deux variables aléatoires :

Soit deux variables aléatoires  $x$  et  $y$  ayant une densité de probabilité  $f(x,y)$  avec  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dx dy = 1$ . La densité de probabilité marginale de  $x$  est :

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y) dy$$

( c'est à dire sans observation de  $y$  ).

Pour  $y$  on obtient :

$$f_2(y) = \int_{-\infty}^{+\infty} f(x,y) dx$$

On peut également définir les densités de probabilité conditionnelles :

$$f_3(y|x) f_1(x) = f(x,y)$$

$f_3$  : densité de probabilité d'obtenir  $y$  avec  $x$  donnée

$$\text{ou } f_4(x|y) f_2(y) = f(x,y)$$

$f_4$  : densité de probabilité d'obtenir  $x$  avec  $y$  donnée .

À l'aide de ces définitions on peut calculer les moyennes et les écarts types :

$$\mu_x = \int_{-\infty}^{+\infty} x f_1(x) dx; \sigma_x^2 = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f_1(x) dx$$

$$\mu_y = \int_{-\infty}^{+\infty} y f_2(y) dy; \sigma_y^2 = \int_{-\infty}^{+\infty} (y - \mu_y)^2 f_2(y) dy$$

La covariance de  $x$  et de  $y$  est définie par :

$$E((x - \mu_x)(y - \mu_y)) = cov(x, y)$$

C'est une mesure de la corrélation entre les variables  $x$  et  $y$ . Le coefficient de corrélation est défini par :

$$\rho_{xy} = \frac{cov(x, y)}{\sigma_x \sigma_y} \quad \text{avec} \quad -1 \leq \rho_{xy} \leq 1$$

On peut montrer que :  $E(x + y) = E(x) + E(y)$  et  $var(x + y) = var(x) + var(y) + 2 cov(x, y)$

Pour comprendre la signification de la covariance, on peut examiner le cas de deux variables aléatoires possédant la même variance :  $var(x) = var(y) = \sigma^2 = \sigma_x^2 = \sigma_y^2$

$$\text{si } \rho_{xy} = -1 \quad cov(x, y) = -\sigma^2 \Rightarrow var(x + y) = 0$$

$$\text{si } \rho_{xy} = 0 \quad cov(x, y) = 0 \Rightarrow var(x + y) = var(x) + var(y) = 2\sigma^2$$

$$\text{si } \rho_{xy} = 1 \quad cov(x, y) = \sigma^2 \Rightarrow var(x + y) = 4\sigma^2$$

Deux variables aléatoires sont indépendantes si et seulement si :

$$f(x, y) = f_1(x) f_2(y)$$

$$\text{c'est à dire : } \begin{aligned} f_3(y|x) &= f_2(y) \\ f_4(x|y) &= f_1(x) \end{aligned}$$

Si  $x$  et  $y$  sont indépendantes on a :

$$cov(x, y) = E((x - \mu_x)(y - \mu_y)) = E(x - \mu_x) E(y - \mu_y) = 0$$

$$\text{donc } \rho_{xy} = 0$$

D'une manière générale si  $u(x)$  et  $v(y)$  sont des fonctions de variables  $x$  et  $y$  indépendantes, alors :

$$E(u(x)v(y)) = E(u(x))E(v(y))$$

Pour deux variables aléatoires, la fonction de répartition est définie par :

$$F(a, b) = P(x < a, y < b) = \int_{-\infty}^a \int_{-\infty}^b f(x, y) dx dy$$

$$\text{on a évidemment : } f(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F(x, y)$$

## 1.8 Distributions à n variables aléatoires :

$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$  est un vecteur formé à partir des n variables aléatoires  $x_1 \cdots x_n$

Si la fonction de répartition est  $F(\vec{x})$ , alors :

$$f(\vec{x}) = \frac{\partial}{\partial \vec{x}} F(\vec{x}) \quad (\text{densité de probabilité})$$

L'espérance mathématique d'une fonction  $H(\vec{x})$  sera :

$$E(H(\vec{x})) = \int H(\vec{x}) f(\vec{x}) d\vec{x}$$

De plus les covariances et les variances peuvent être arrangées dans une matrice appelée matrice de covariance :

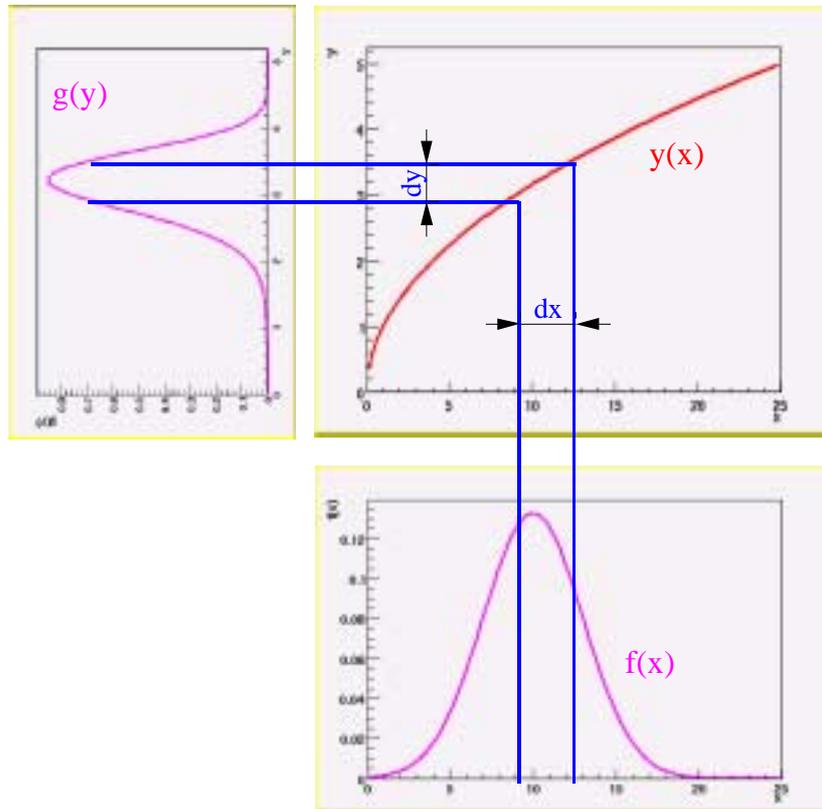
$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix} \quad \text{avec} \quad C_{ii} = \sigma^2(x_i), \quad C_{ij} = c_{ji} = \text{cov}(x_i, x_j) \quad \text{si} \quad i \neq j$$

On peut également écrire cette matrice sous la forme suivante :

$$C = E\{(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T\} \quad \text{avec} \quad \vec{x}^T = (x_1 \ x_2 \ \cdots \ x_n), \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{et} \quad \vec{\mu} = E(\vec{x})$$

## 1.9 Changement de variables :

Pour une variable :



$y=y(x)$  ( une fonction monotone de  $x$  ) est une nouvelle variable aléatoire.

Si  $f(x)$  est une densité de probabilité connue , la question est de trouver  $g(y)$  .

Au vu de la figure ci-dessus , on a :  $g(y)dy=f(x)dx$  . Puisque  $g(y),f(y)\geq 0$  , on obtient finalement :

$$g(y)=\left|\frac{dx}{dy}\right|f(x)=\left|\frac{dx}{dy}\right|f(w(y)) \quad \text{avec} \quad x=w(y) \quad \text{fonction inverse de } y .$$

Pour  $n$  variables :

$$\vec{y} \begin{pmatrix} y_1 = y_1(\vec{x}) \\ y_2 = y_2(\vec{x}) \\ \vdots \\ y_n = y_n(\vec{x}) \end{pmatrix} \quad \vec{x} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_3 \end{pmatrix} \quad f(\vec{x}) \text{ qui est une densité de probabilité connue .}$$

Alors :  $g(\vec{y}) = J\left(\frac{\vec{x}}{\vec{y}}\right) f(\vec{w}(\vec{y}))$  où  $J$  est le Jacobien de la transformation ,  $J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \dots & \frac{\partial x_n}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_n}{\partial y_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_n} & \frac{\partial x_2}{\partial y_n} & \dots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$

et  $\vec{w}(\vec{y}) = \vec{x}$  est le vecteur de fonctions inverses

## 2 Distributions de probabilité usuelles :

### 2.1 Distribution binomiale :

La distribution binomiale s'applique aux systèmes composés de  $n$  objets identiques et indépendants qui évoluent selon un processus aléatoire commun à réponse binaire (1 ou 0 ; succès ou échec ). Cette situation est mathématiquement identique à la réalisation d'une série de  $n$  expériences identiques et indépendantes, chacune représentée par une variable aléatoire binaire  $x_i$  . Soit  $x = \sum_i^n x_i$  ( $x_i = 0, 1$ ) , la variable aléatoire qui calcule le nombre de succès ( $x = 0, 1, 2, \dots, n$ ) . Si  $p$  est la probabilité de succès pour une seule expérience et  $q = 1 - p$  la probabilité d'échec , la probabilité d'obtenir la valeur  $x$  pour la série est:

$$f(x, n, p) = C_n^x p^x q^{n-x} \text{ avec } C_n^x = \frac{n!}{x!(n-x)!}$$

$$E(x) = \sum_{i=1}^n E(x_i) = n p$$

$$Var(x) = \sum_{i=1}^n var(x_i) = \sum_{i=1}^n E((x - p)^2) = n((1 - p)^2 p + (0 - p)^2 q) = n p q$$

ex : 100 décroissances de neutrons ( $n \rightarrow p \bar{\nu}_e e$ ) observées avec un détecteur dont l'efficacité est

$$\epsilon = 0,3 (p = 0,3) \quad , \quad x = \sum_{i=1}^{100} x_i \quad , \quad E(x) = 100 \times 0,3 = 30 \quad , \quad \sigma(x) = \sqrt{100 \times 0,3 \times 0,7} = 4,6$$

## 2.2 Distribution de Poisson :

Il n'est pas toujours aisé d'avoir recours à la distribution binomiale lorsque  $n$  (nombre d'objets indépendants) devient grand ou lorsqu'il est inconnu. Dans la mesure où la moyenne ( $np$ ) reste finie, on peut montrer que la distribution binomiale peut être simplifiée.

Soit  $\lambda = np$  avec  $\lambda$  finie ( $n \rightarrow \infty, p \rightarrow 0$ ) : c'est à dire un grand nombre d'objets indépendants et une petite efficacité de succès.

$$\begin{aligned} f(x, n, p) &= C_n^x p^x q^{n-x} = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x} \\ &= \frac{\lambda^x}{x!} \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x} \\ &= \frac{\lambda^x}{x!} \left(1-\frac{\lambda}{n}\right)^n \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^x} \end{aligned}$$

si  $n$  est grand, on a :  $\left(1-\frac{i}{n}\right) \rightarrow 1$ ,  $\left(1-\frac{\lambda}{n}\right)^x \rightarrow 1$ ,  $\lim_{n \rightarrow \infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$  on obtient finalement :

$$f(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \text{ qui est la distribution de Poisson (définie pour } \lambda > 0 \text{ strictement)}$$

On peut montrer que  $E(x) = \sigma^2(x) = \lambda$ . On peut également retenir (sans démonstration) que la variable aléatoire formée de la somme de deux variables aléatoires poissoniennes indépendantes de moyennes

$\lambda_1, \lambda_2$  est également une distribution de Poisson ayant pour moyenne  $\lambda_1 + \lambda_2$ .

Strictement parlant, la distribution de Poisson n'est utilisable que lorsque  $n \rightarrow \infty$  et  $p \rightarrow 0$ .

Cependant, en pratique elle s'avère être une très bonne approximation de la distribution binomiale plus lourde à manipuler dans les calculs numériques. Au-dessus de  $n=60$  pour  $p=0,05$  on peut considérer qu'on peut utiliser la distribution de Poisson.

ex: La décroissance d'un grand nombre de noyaux radioactifs contenus dans une source sur un temps court en comparaison avec leur période suit une distribution de Poisson. En effet  $n$  est très grand, de l'ordre de grandeur du nombre d'Avogadro, et  $p$  (probabilité de décroissance par noyau) est petite:

$$\lambda = np$$

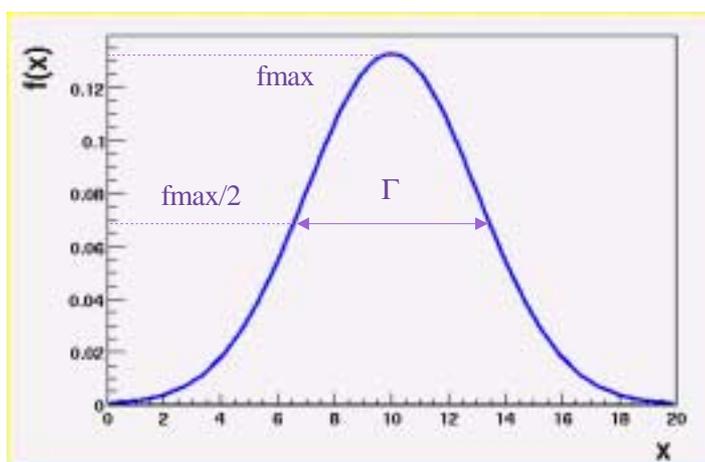
## 2.3 Gaussienne ou distribution normale :

La distribution normale est une autre approximation de la distribution binomiale dans le cas où le nombre d'objets  $n$  ( ou d'expériences ) devient grand avec une probabilité  $p$  de succès importante : c'est à dire lorsque  $np \gg 1$  .

Elle est définie par :

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

Pour une Gaussienne on obtient :  $E(x) = \mu$  ,  $var(x) = \sigma^2$



On peut caractériser une distribution ( densité de probabilité ) par sa largeur totale  $\Gamma$  à la moitié de sa hauteur (FWHM = full width at half maximum) :  $\Gamma = 2,354\sigma$  pour une Gaussienne.

On peut également rencontrer la distribution normale réduite qui a une moyenne nulle et un écart type égal à 1 :

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} ,$$

sa fonction de répartition étant :  $F_0(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \frac{1}{2} (1 + \operatorname{erf}(\frac{x}{\sqrt{2}}))$  .

Pour une distribution symétrique, on a la relation suivante:  $P(|x-\mu| \leq e) = 2F(e) - 1$  .

Ce qui donne pour une distribution normale :

$$P(|x-\mu| \leq 1\sigma) = 0,682$$

$$P(|x-\mu| \leq 2\sigma) = 0,954$$

$$P(|x-\mu| \leq 3\sigma) = 0,998$$

La distribution normale est également le modèle vers lequel une distribution de Poisson tend lorsque  $\lambda$  (valeur moyenne) devient grand.

En pratique on peut considérer qu'une Gaussienne est une bonne approximation d'une Poissonienne dès que  $\lambda$  dépasse 30.

La somme de deux variables aléatoires normales et indépendantes de moyennes  $\mu_1, \mu_2$  et d'écart type  $\sigma_1, \sigma_2$  est elle-même une variable normale avec  $\mu = \mu_1 + \mu_2$  et  $\sigma^2 = \sigma_1^2 + \sigma_2^2$ .

## 2.4 Distribution normale à n variables :

C'est une généralisation de la distribution précédente :

$$f(\vec{x}, \vec{\mu}, B) = \frac{1}{(2\pi)^{n/2} \sqrt{|B|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T B^{-1}(\vec{x}-\vec{\mu})\right) \quad \text{avec} \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \quad \text{et} \quad \mu_i = E(x_i)$$

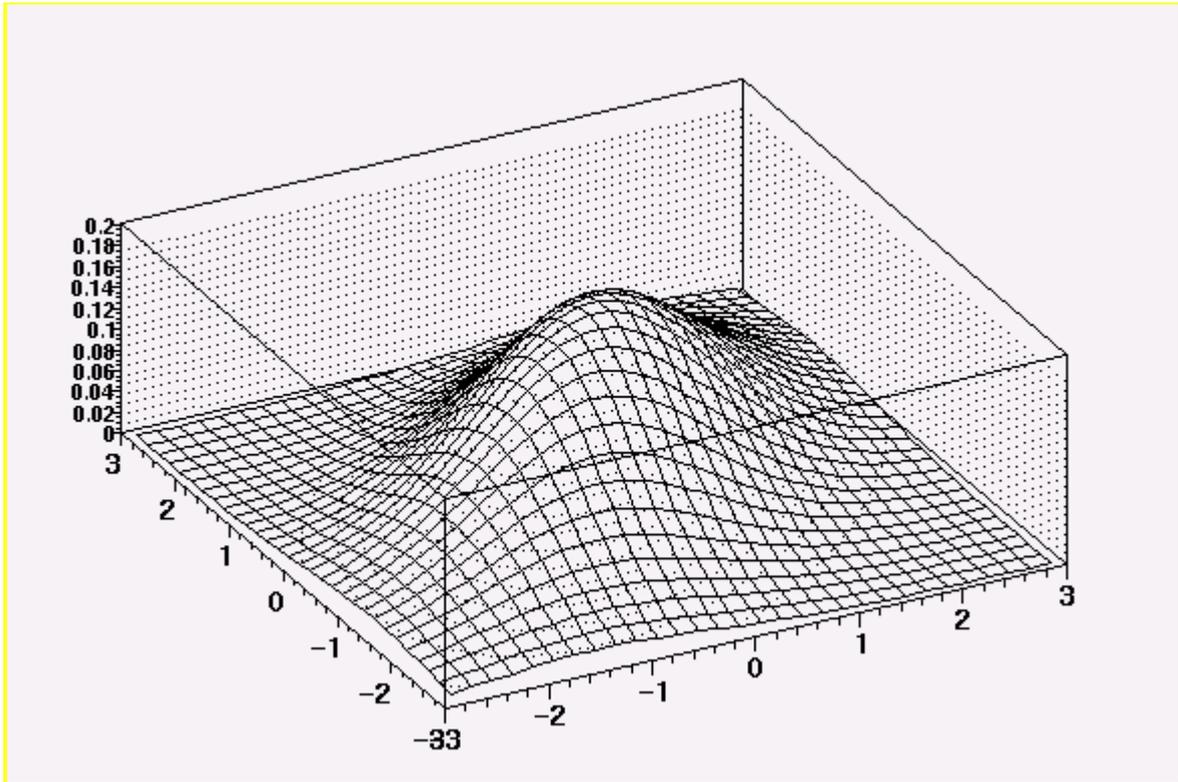
$B = C^{-1}$ , C étant la matrice de covariance.

**Cas n=2 :**

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$f(x_1, x_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) =$$

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right)$$



Les courbes d'équiprobabilité sont des ellipses :

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = (1 - \rho^2) n^2 ,$$

où n est le nombre d'écart type ( pour le cas n=1 , il s'agit de l'ellipse de covariance).

La probabilité de trouver  $x_1, x_2$  à l'intérieur d'une ellipse à n écarts type est donnée par :

$$P(n\sigma) = 1 - e^{-\frac{1}{2}n^2}$$

### 2.5 Loi de khi-deux :

Si  $x_1, x_2, \dots, x_n$  sont des variables aléatoires indépendantes et normales , la somme :

$$\chi_n^2 = \sum_{i=1}^n \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2 \text{ qui est elle-même une variable aléatoire,}$$

suit une loi dite de Khi-deux à n degrés de liberté. La densité de probabilité de cette loi, qui ne dépend que de n, est :

$$f(\chi_n^2, n) = \frac{(\chi_n^2)^{\frac{n}{2}-1} e^{-\frac{\chi_n^2}{2}}}{2^{n/2} \Gamma(\frac{n}{2})} \text{ avec } \Gamma(\frac{n}{2}) = \int_0^{+\infty} t^{\frac{n}{2}-1} e^{-t} dt \text{ étant la fonction d'Euler } (\Gamma(x+1) = x\Gamma(x))$$

On peut montrer que :  $E(\chi_n^2) = n$  ,  $var(\chi_n^2) = 2n$  .

On trouve également que la somme  $\sum_i \chi_{n_i}^2$  suit une loi de khi-deux à  $n = \sum_i n_i$  degrés de liberté .

On peut également rencontrer la variable de  $\chi^2$  réduite qui est  $\frac{\chi^2}{n}$  dont l'espérance mathématique vaut 1.

Pour un échantillon de  $n$  événements  $x_i$  provenant d'une même variable aléatoire normale de moyenne  $\mu$  et d'écart type  $\sigma$  connus , la somme :  $\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2$  , suivra une loi de khi-deux à  $n$  degrés de liberté.

Le résultat de cette expérience devrait fournir une valeur de  $\chi_n^2$  proche de  $n$  du fait que la moyenne de la loi de chi-deux est précisément  $n$  .

### 3 Théorème de la valeur centrale et modèle de l'erreur expérimentale

Ce théorème établit que la somme de  $n$  variables aléatoires indépendantes ayant chacune la même moyenne ( $\mu$ ) et la même variance finie ( $\sigma$ ) tend vers une Gaussienne lorsque  $n$  tend vers l'infini:

$$x = \sum_{i=1}^n x_i \rightarrow \text{Gaussienne de moyenne } n\mu \text{ et de variance } n\sigma^2 \text{ lorsque } n \rightarrow \infty.$$

Lorsque les  $x_i$  ont des distributions différentes, ce théorème reste valable, dans le sens où à grand  $n$ , la somme des  $x_i$  tend vers une Gaussienne. Dans ce cas,  $E(x) = \sum_{i=1}^n \mu_i$  et  $var(x) = \sum_{i=1}^n \sigma_i^2$  .

À partir de ce théorème, en 1783 Laplace a proposé un modèle de l'erreur expérimentale. On peut considérer qu'une mesure s'écrit comme :

$$x = x_{vrai} + e_1 + e_2 + \dots + e_n = x_{vrai} + \sum_{i=1}^n e_i \text{ avec } E(e_i) = 0 \quad \sigma(x_{vrai}) = 0$$

La somme dans le deuxième membre représente toutes les sources d'erreurs (souvent inconnues) qui, si elles sont nombreuses, produisent une distribution de probabilité d'erreur totale proche d'une distribution normale en vertu de l'application du théorème de valeur centrale.

Ce modèle permet de justifier l'utilisation fréquente d'une Gaussienne pour la densité de probabilité des erreurs de mesure.

Lorsqu'une mesure est effectuée dans des conditions qui satisfont le théorème de la valeur centrale, les résultats sont communiqués de la façon suivante :

$$m \pm \sigma \quad \text{où } m \text{ est le résultat de la mesure et } \sigma \text{ son erreur .}$$

Cet intervalle est à 68,3% de confiance: c'est à dire 68,3% de probabilité que la vraie valeur de la quantité mesurée soit contenue dans l'intervalle.

Dans la plupart des expériences, l'applicabilité du théorème de la valeur centrale n'est jamais réellement démontrée et il faut se méfier du niveau de confiance des intervalles communiqués.

## 4 Échantillonnage et estimation :

Une densité de probabilité constitue un modèle théorique qui permet de prédire la fréquence avec laquelle la valeur d'une variable aléatoire sera observée. Les paramètres (masse d'une particule, durée de vie, moyenne, variance,...) de cette densité de probabilité sont généralement inconnus. Ce sont les mesures qui peuvent nous permettre de les déterminer. C'est du reste l'objet, le but des mesures dans la plupart des cas.

Une série de mesures d'une variable aléatoire constitue un échantillon. L'ensemble de tous les résultats de mesures possibles (toutes les expériences possibles) d'une variable aléatoire est la population. Un échantillon est un sous-ensemble d'une population. Un échantillon de taille  $n$  est représenté par une série :

$$x_1, x_2, \dots, x_n \text{ .}$$

Un échantillon est lui-même une variable aléatoire à  $n$  dimensions (l'expérience qui mesure un échantillon peut être répétée) . On appelle statistique une fonction qui dépend d'un échantillon:

$$f(x_1, x_2, \dots, x_n) \text{ .}$$

Un estimateur est une statistique destinée à produire une valeur estimative de l'un des paramètres de la densité de probabilité de la variable aléatoire mesurée. Un estimateur est dit sans biais si quelle que soit la taille d'un échantillon, sa moyenne est égale à la valeur du paramètre recherché.

### 4.1 Estimation d'une moyenne :

$n$  événements indépendants d'une variable aléatoire  $x$  sont mesurés :  $\bar{x} = \sum_{i=1}^n x_i$  est un estimateur sans biais de sa moyenne.

$$E(\bar{x}) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} n E(x) = E(x) \quad \forall n$$

$$\begin{aligned} \text{var}(\bar{x}) &= E((\bar{x} - E(x))^2) = E\left(\left(\frac{x_1 + x_2 + \dots + x_n}{n} - E(x)\right)^2\right) = \frac{1}{n^2} E((x_1 - E(x))^2 + (x_2 - E(x))^2 + \dots + (x_n - E(x))^2) \\ &= \frac{\sigma^2}{n} \end{aligned}$$

$$\text{ou encore} \quad \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

### 4.2 Estimation d'une variance :

$n$  événements indépendants d'une variable aléatoire  $x$  (de moyenne inconnue) sont mesurés :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \bar{x}^2 \right) \text{ est un estimateur sans biais de la variance .}$$

Démonstration :

$$\begin{aligned}
 E(S^2) &= \frac{1}{n-1} E\left( \sum_{i=1}^n (x_i - \bar{x})^2 \right) = \frac{1}{n-1} E\left( \sum_{i=1}^n (x_i - E(x) + E(x) - \bar{x})^2 \right) \\
 &= \frac{1}{n-1} E\left( \sum_{i=1}^n (x_i - E(x))^2 + \sum_{i=1}^n (E(x) - \bar{x})^2 + 2 \sum_{i=1}^n (x_i - E(x))(E(x) - \bar{x}) \right) \\
 &= \frac{1}{n-1} E\left( \sum_{i=1}^n (x_i - E(x))^2 + \sum_{i=1}^n (E(x) - \bar{x})^2 - 2n(E(x) - \bar{x})^2 \right) \\
 &= \frac{1}{n-1} E\left( \sum_{i=1}^n (x_i - E(x))^2 - \sum_{i=1}^n (E(x) - \bar{x})^2 \right) \\
 &= \frac{1}{n-1} \left( E\left( \sum_{i=1}^n (x_i - E(x))^2 \right) - E\left( \sum_{i=1}^n (\bar{x} - E(x))^2 \right) \right) \\
 &= \frac{1}{n-1} \left( \sum_{i=1}^n E((x_i - E(x))^2) - \sum_{i=1}^n E((\bar{x} - E(x))^2) \right) \\
 &= \frac{1}{n-1} (n\sigma^2 - n\frac{\sigma^2}{n}) = \sigma^2 \quad \forall n > 1
 \end{aligned}$$

(Explication du facteur  $(n-1)$ : pour  $n=1$ , on a évidemment  $\bar{x}=x_1$  et  $var(\bar{x}) = \frac{1}{n-1} E(x_1^2 - x_1^2) = \frac{0}{0}$ , c'est à dire indéterminée)

Si la moyenne théorique ( $E(x)$ ) d'une variable aléatoire échantillonnée est connue, un meilleur estimateur de sa variance est obtenue par :

$$S'^2 = \frac{1}{n} \sum_{i=1}^n (x_i - E(x))^2 = \frac{1}{n} \left( \sum_{i=1}^n x_i^2 + nE(x)^2 - 2n\bar{x}E(x) \right)$$

### 4.3 Estimation de comptage :

Très fréquemment au cours d'une expérience, on est simplement amené à faire un comptage des particules qui satisfont certaines conditions (limites en énergie, limites angulaires ...). D'un point de vue formel, cela revient à prélever un échantillon d'une population binomiale: pour chaque particule on teste si elle satisfait ou non les conditions d'acceptation.

Soit  $n$  la taille de cet échantillon (ou encore le nombre d'événements enregistrés) et  $k$  le nombre de ces événements qui satisfont les conditions de sélection, on a :  $\bar{x} = \sum_{i=1}^n x_i = \frac{k}{n}$ . C'est le meilleur estimateur pour le paramètre  $p$  (probabilité de succès) de la distribution binomiale. Puisque  $\sigma^2(\bar{x}) = \frac{\sigma^2}{n}$  et que la variance d'une distribution binomiale est  $p(1-p)$ , on a :  $\sigma^2(\bar{x}) = \frac{p(1-p)}{n}$ . Si on remplace  $p$  par son estimation, on obtient :

$$\sigma^2(\bar{x}) = \frac{k}{n^2} \left(1 - \frac{k}{n}\right), \text{ ou encore } \sigma^2(n\bar{x}) = k \left(1 - \frac{k}{n}\right), \text{ c'est à dire } \sigma^2(k) = k \left(1 - \frac{k}{n}\right).$$

$\sigma(k) = \sqrt{k \left(1 - \frac{k}{n}\right)}$  est appelée l'erreur statistique de comptage. Comme en général  $n$  est très grand, cette expression est plus souvent connue sous la forme de :  $\sigma(k) = \sqrt{k}$ . Il convient de rester vigilant vis à vis de l'applicabilité de cette expression simplifiée. Si l'efficacité de comptage est voisine de 1, on doit utiliser l'expression complète pour obtenir un résultat correct.

**K est grand :**

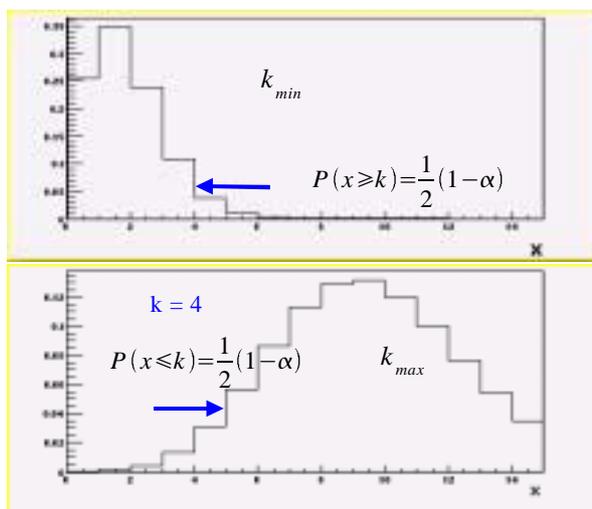
Si k est grand, la distribution binomiale tend vers une loi normale de moyenne k et d'écart type

$\sigma(k) = \sqrt{k(1 - \frac{k}{n})}$  (  $\sqrt{k}$  si l'efficacité de comptage est faible ). Le résultat est donné par un intervalle:  $k \pm \sigma(k)$  , qui est à 68,2% de confiance ( 95,4% pour  $k \pm 2\sigma(k)$  ).

**k est petit et p << 1 :**

Si k est petit et que p est également petit , on est dans le cadre d'application de la loi de Poisson qui est asymétrique, ce qui conduit à un intervalle de confiance également asymétrique autour de la moyenne k . Notez que l'écart type reste  $\sqrt{k}$  ce qui est conforme aux caractéristiques de la loi de Poisson. Pour obtenir un encadrement du comptage k à un certain niveau de confiance  $\alpha$  ( $0 < \alpha < 1$ ) , la méthode consiste à déterminer à l'aide de lois de Poisson , les valeurs  $k_{min}$  et  $k_{max}$  telles que :

$$P(x \geq k, k_{min}) \leq \frac{1}{2}(1 - \alpha) \quad \text{et} \quad P(x \leq k, k_{max}) \leq \frac{1}{2}(1 - \alpha)$$



La table ci-dessous donne les valeurs de  $k_{min}$  et  $k_{max}$  pour les valeurs usuelles de k et de  $\alpha$  :

$\alpha$ pour $k_{min} < k_{vrai}$	$\alpha$ pour $k_{min} < k_{vrai} < k_{max}$	$\alpha$ pour $k_{vrai} < k_{max}$	$k_{min}$	$k$	$k_{max}$
		84%	0	0	1,84
84%	68%	84%	0,17	1	3,3
84%	68%	84%	0,71	2	4,64
84%	68%	84%	1,37	3	5,92
84%	68%	84%	2,09	4	7,16
84%	68%	84%	3,62	6	9,58
84%	68%	84%	6,89	10	14,26
84%	68%	84%	15,57	20	25,54

$\alpha$ pour $k_{min} < k_{vrai}$	$\alpha$ pour $k_{min} < k_{vrai} < k_{max}$	$\alpha$ pour $k_{vrai} < k_{max}$	$k_{min}$	$k$	$k_{max}$
84%	68%	84%	42,95	50	58,11
		95%	0	0	3
95%	90%	95%	0,05	1	4,74
95%	90%	95%	0,35	2	6,3
95%	90%	95%	0,82	3	7,75
95%	90%	95%	1,37	4	9,15
95%	90%	95%	2,61	6	11,84
95%	90%	95%	5,43	10	16,96
95%	90%	95%	13,26	20	29,06
95%	90%	95%	38,96	50	63,28

#### 4.4 Estimation par la méthode du maximum de vraisemblance :

Cette méthode permet d'estimer un paramètre quelconque d'une densité de probabilité à partir d'un échantillon de taille n de mesures indépendantes.

Soit  $x_1, x_2, \dots, x_n$  un échantillon mesuré à partir de la variable aléatoire  $x$ . La fonction de vraisemblance est définie par :

$$L(\vec{\alpha}) = \prod_{i=1}^n f(x_i, \vec{\alpha}) \quad \text{où } f(x, \vec{\alpha}) \text{ est la densité de probabilité supposée de forme connue de la variable échantillonnée, } \vec{\alpha} \text{ est le vecteur des paramètres à estimer.}$$

Pour briser le produit  $\prod_i$  en une somme, on a souvent recours à la fonction logarithmique de vraisemblance :

$$\ln(L(\vec{\alpha})) = \sum_{i=1}^n \ln(f(x_i, \vec{\alpha})) .$$

Dans cette méthode, l'estimation des paramètres est fournie en recherchant le vecteur  $\vec{\alpha}_e$  qui maximise la fonction logarithmique de vraisemblance, ce qui consiste à résoudre :

$$\frac{\partial \{ \ln(L(\vec{\alpha})) \}}{\partial \vec{\alpha}} = 0 .$$

##### Exemple 1:

Données échantillonnées à partir d'une population de Poisson :  $f(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$

$$\frac{\partial}{\partial \lambda} \left( \sum_{i=1}^n \ln \left( \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) \right) = 0$$

$$\frac{\partial}{\partial \lambda} \left( \sum_{i=1}^n (\ln(\lambda^{x_i}) - \ln(x_i!) - \lambda) \right) = 0$$

$$\sum_{i=1}^n \left( \frac{x_i}{\lambda} - 1 \right) = 0$$

$$\lambda = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

**Exemple 2:**

Mesures indépendantes et de précisions différentes . Nous prendrons comme hypothèse que les erreurs sont distribuées de façon normale . La densité de probabilité est :  $f(x, \sigma ; \lambda) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\lambda}{\sigma}\right)^2}$  . Les données sont  $x_i, \sigma_i$  . Donc  $f$  ne dépend que d'un seul paramètre :  $\lambda$  .

$$\frac{\partial}{\partial \lambda} \left( \sum_{i=1}^n \ln \left( \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i-\lambda)^2}{2\sigma_i^2}} \right) \right) = 0$$

$$\frac{\partial}{\partial \lambda} \left( \sum_{i=1}^n \ln \left( \frac{1}{\sqrt{2\pi}\sigma_i} \right) - \frac{1}{2} \frac{(x_i-\lambda)^2}{\sigma_i^2} \right) = 0$$

$$\sum_{i=1}^n \frac{x_i - \lambda}{\sigma_i^2} = 0 \quad \text{qui a pour solution} \quad \lambda = \frac{\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} .$$

**4.5 Estimation par la méthode des moindres carrés :**

Cette méthode est celle que l'on rencontre le plus souvent dans les problèmes d'estimation de paramètres inconnus à partir de données mesurées .

**Cas de mesures directes :**

Les mesures sont directement reliées à la quantité inconnue  $x$  . On dispose d'un échantillon de  $n$  mesures indépendantes  $y_i$  de la même quantité recherchée  $x$  ; c'est à dire  $y_i = x + \epsilon_i$  où  $\epsilon_i$  est normalement distribuée autour de zéro avec  $E(\epsilon_i^2) = \sigma_i^2$  . Au passage , nous avons fait l'hypothèse que les erreurs de mesure sont normalement distribuées (application du théorème de la valeur centrale). La fonction logarithmique de vraisemblance est:

$$\ln(L) = -\frac{1}{2} \sum_{i=1}^n \left( \frac{y_i - x}{\sigma_i} \right)^2 + \text{const}$$

Imposer que cette fonction soit maximale revient à demander que :  $M = \sum_{i=1}^n \left( \frac{y_i - x}{\sigma_i} \right)^2$  soit minimale .

C'est ce que l'on appelle la méthode des moindres carrés.

En pratique , on obtient :

$$\frac{\partial M}{\partial x} = 0 \Rightarrow \bar{x} = \frac{\sum_{i=1}^n \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} .$$

Les  $n$  résidus sont définis comme étant:  $\epsilon_i = y_i - \bar{x}$  . Ils sont normalement distribués et possèdent des variances  $\sigma_i^2$  . Par conséquent ,  $M = \sum_{i=1}^n \left(\frac{\epsilon_i}{\sigma_i}\right)^2$  suit une loi de khi-deux à  $(n-1)$  degrés de liberté. Le nombre de degrés de liberté (1 de moins que ce que l'on pourrait attendre) provient du fait que les résidus ne sont pas indépendants. En effet, la condition de minimisation s'écrit  $\sum_{i=1}^n \frac{(y_i - \bar{x})}{\sigma_i^2} = 0$  soit encore :  $\sum_{i=1}^n \frac{\epsilon_i}{\sigma_i^2} = 0$  .

Le niveau de confiance du résultat  $\bar{x}$  peut être estimé par un test de chi-deux. On se donne le niveau de confiance  $\alpha$  auquel on souhaiterait parvenir pour le résultat: attention la définition de ce niveau de confiance est inverse à celle utilisée pour les intervalles de mesures. Ensuite, on cherche la valeur de  $\chi_{max}^2$  pour laquelle:

$$\int_0^{\chi_{max}^2} f(\chi^2, n-1) d\chi^2 = (1-\alpha) \quad \text{où} \quad f(\chi^2, n-1) \quad \text{est une distribution de chi-deux à } n-1 \text{ degrés de liberté.}$$

Si  $M$  dépasse  $\chi_{max}^2$  , on est en droit de se douter du résultat au niveau de confiance choisi. On remarquera que plus  $\alpha$  est grand, plus  $\chi_{max}^2$  est petit ; ce qui est inverse à la logique utilisée lors de l'estimation du niveau de confiance d'un intervalle de comptage (4.3) qui faisait qu'à un niveau de confiance plus grand était associé un intervalle de comptage plus large .

Les fractiles  $f_q$  d'une densité de probabilité sont définies par :  $F(f_q) = q$  où  $F$  est la fonction de répartition de la densité de probabilité considérée et  $q$  est une probabilité comprise entre 0 et 1 . Dans le test du khi-deux exposé dans le paragraphe précédent , la valeur de  $\chi_{max}^2$  est en fait la fractile  $1-\alpha$  de la loi de khi-deux à  $(n-1)$  degrés de liberté.

La table ci-contre donne les fractiles de la loi de khi-deux pour quelques cas courants :

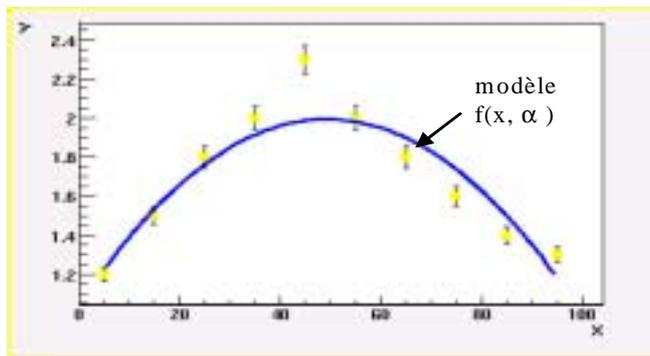
$n \setminus q$	0,01	0,05	0,1	0,5	0,9	0,99
1	–	–	0,02	0,46	2,71	6,63
2	0,02	0,1	0,21	1,39	4,61	9,21
3	0,12	0,35	0,58	2,37	6,25	11,3
4	0,3	0,71	1,06	3,36	7,78	13,3
5	0,55	1,15	1,61	4,35	9,24	15,1
6	0,87	1,64	2,2	5,35	10,6	16,8
7	1,24	2,17	2,83	6,35	12	18,5
8	1,65	2,73	3,49	7,34	13,4	20,1
9	2,09	3,33	4,17	8,34	14,7	21,7
10	2,56	3,94	4,87	9,34	16	23,2
11	3,05	4,57	5,58	10,3	17,3	24,7
12	3,57	5,23	6,3	11,3	18,5	26,2
13	4,11	5,89	7,04	12,3	19,8	27,7
14	4,66	6,57	7,79	13,3	21,1	29,1
15	5,23	7,26	8,55	14,3	22,3	30,6
20	8,26	10,9	12,4	19,3	28,4	37,6
25	11,5	14,6	16,5	24,3	34,4	44,3
30	15	18,5	20,6	29,3	40,3	50,9

### Cas de mesures indirectes :

$n$  mesures indépendantes sont effectuées aux points  $x_i$ . Les données  $y_i$  (indépendantes) doivent maintenant être comparées aux valeurs d'une fonction,  $f(x_i, \vec{\alpha})$  où le vecteur  $\vec{\alpha}$  représente les paramètres à estimer. On forme la somme suivante :

$$M = \sum_{i=1}^n \frac{(y_i - f(x_i, \vec{\alpha}))^2}{\sigma_i^2} .$$

Ce problème d'estimation est celui que l'on rencontre le plus fréquemment en physique.



Le vecteur  $\vec{\alpha}_s$  solution du système d'équations de minimisation  $\frac{\partial M}{\partial \vec{\alpha}}=0$  constitue l'estimation des paramètres recherchés .

Si les erreurs (  $\sigma_i$  ) sont distribuées de façon normale, on peut montrer que :

$$M(\vec{\alpha}_s) = \sum_{i=1}^n \left( \frac{y_i - f(x_i, \vec{\alpha}_s)}{\sigma_i} \right)^2 ,$$

suit une loi de khi-deux à  $n-k$  degrés de liberté ;  $k$  étant le nombre de paramètres ou la dimension de  $\vec{\alpha}$  .

Le niveau de confiance de la solution  $\vec{\alpha}_s$  peut être testé conformément à la méthode décrite dans le paragraphe précédent .

### Exemple :

Paramétrisation ou ajustement (fit en anglais) d'une droite sur une série de points . On dispose de  $n$  points  $(x_i, (y_i, \sigma_i))$  . Le modèle est  $f(x; a, b) = ax + b$  . La somme des carrés est donnée par :

$$M = \sum_{i=1}^n \left( \frac{y_i - ax_i - b}{\sigma_i} \right)^2 .$$

La minimisation conduit à :

$$\frac{\partial M}{\partial a} \Leftrightarrow \sum_{i=1}^n \frac{1}{\sigma_i^2} (y_i - ax_i - b) x_i = 0$$

$$\frac{\partial M}{\partial b} \Leftrightarrow \sum_{i=1}^n \frac{1}{\sigma_i^2} (y_i - ax_i - b) = 0$$

$$\sum_{i=1}^n \frac{1}{\sigma_i^2} x_i y_i = a \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} + b \sum_{i=1}^n \frac{x_i}{\sigma_i^2}$$

$$\sum_{i=1}^n \frac{1}{\sigma_i^2} y_i = a \sum_{i=1}^n \frac{x_i}{\sigma_i^2} + b \sum_{i=1}^n \frac{1}{\sigma_i^2}$$

C'est un système d'équations à deux inconnues :  $a$  et  $b$  .

## 5 Loi de propagation des erreurs :

Soit  $n$  variables aléatoires  $x_i$  ( $i=1, \dots, n$ ) et leur matrice de covariance associée :

$$C_x = E\{(\vec{x} - \vec{\mu}_x)(\vec{x} - \vec{\mu}_x)^T\} \text{ avec } \vec{\mu}_x = E(\vec{x}) .$$

On construit par transformations linéaires de ces  $x_i$  ,  $r$  nouvelles variables aléatoires  $y_j$  :

$$y_1 = a_1 + t_{11} x_1 + t_{12} x_2 + \dots + t_{1n} x_n$$

$$y_2 = a_2 + t_{21} x_1 + t_{22} x_2 + \dots + t_{2n} x_n$$

$$\vdots$$

$$y_r = a_r + t_{r1} x_1 + t_{r2} x_2 + \dots + t_{rn} x_n$$

ou encore  $\vec{y} = T\vec{x} + \vec{a}$  , on obtient alors :  $\vec{\mu}_y = T\vec{\mu}_x + \vec{a}$  .

Et pour ce qui concerne la nouvelle matrice de covariance :

$$\begin{aligned} C_y &= E\{(\vec{y}-\vec{\mu}_y)(\vec{y}-\vec{\mu}_y)^T\} = E\{(T\vec{x}+\vec{a}-T\vec{\mu}_x-\vec{a})(T\vec{x}+\vec{a}-T\vec{\mu}_x-\vec{a})^T\} \\ &= E\{T(\vec{x}-\vec{\mu}_x)(\vec{x}-\vec{\mu}_x)^T T^T\} \\ &= T E\{(\vec{x}-\vec{\mu}_x)(\vec{x}-\vec{\mu}_x)^T\} T^T \end{aligned}$$

finalement , on obtient :  $C_y = T C_x T^T$

Supposons maintenant que nous ayons mesuré un échantillon à  $n$  variables aléatoires  $x_i$  et que nous ayons ainsi estimé leurs moyennes ( $\vec{\mu}_x$ ) et leur matrice de covariance associée  $C_x$ . Nous formons à partir de l'échantillon des  $x_i$  un nouvel échantillon à  $r$  variables aléatoires  $y_j$  obtenues par des fonctions du type :  $y_j(\vec{x})$ . Si les erreurs de mesure des  $x_i$  sont assez petites en comparaison à leurs moyennes, la densité de probabilité des erreurs ne sera grande qu'au voisinage des moyennes. On peut alors développer les  $y_j$  au voisinage des moyennes  $\vec{\mu}_x$  :

$$y_j(\vec{x}) = y_j(\vec{\mu}_x) + \left(\frac{\partial y_j}{\partial x_1}\right)_{\vec{x}=\vec{\mu}_x} (x_1 - \mu_{x1}) + \dots + \left(\frac{\partial y_j}{\partial x_n}\right)_{\vec{x}=\vec{\mu}_x} (x_n - \mu_{xn}) \quad , \quad (j=1, r),$$

ou encore :

$$\vec{y}(\vec{x}) = \vec{y}(\vec{\mu}_x) + T(\vec{x} - \vec{\mu}_x) \quad \text{avec} \quad T = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_r}{\partial x_1} & \frac{\partial y_r}{\partial x_2} & \dots & \frac{\partial y_r}{\partial x_n} \end{pmatrix} \quad \vec{x} = \vec{\mu}_x$$

En appliquant la formule obtenue précédemment, on a :

$$C_y = T C_x T^T .$$

Dans le cas général, il convient de tenir compte des covariances et des variances dans la matrice  $C_x$  pour obtenir les termes diagonaux de  $C_y$ . Si par contre toutes les covariances dans la matrice  $C_x$  sont nulles, c'est à dire si les variables  $x_i$  sont indépendantes, alors les termes diagonaux de  $C_y$  sont :

$$\sigma^2(y_j) = \sum_{i=1}^n \left(\frac{\partial y_j}{\partial x_i}\right)_{\vec{x}=\vec{\mu}_x}^2 \sigma^2(x_i) .$$

C'est ce qui est communément appelé la loi de propagation des erreurs. Il faut garder à l'esprit qu'elle n'est applicable que dans le cas où les variables  $x_i$  sont indépendantes.

**Exemples :**

$$\sigma(\ln(y)) = \frac{\sigma_y}{y}$$

$$y = \frac{a}{b} \Rightarrow \left(\frac{\sigma_y}{y}\right)^2 = \left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2$$

$$y = a + b \Rightarrow \sigma_y^2 = \sigma_a^2 + \sigma_b^2$$

$$y = a - b \Rightarrow \sigma_y^2 = \sigma_a^2 + \sigma_b^2$$

$$y = a^2 + \sqrt{b} \Rightarrow \sigma_y^2 = 4a^2\sigma_a^2 + \frac{1}{4}\frac{\sigma_b^2}{b}$$

*a* et *b* sont supposées indépendantes

$y = f(x_1, x_2)$  où  $x_1$  et  $x_2$  ne sont pas indépendantes.

$$\sigma_y^2 = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{pmatrix} \begin{pmatrix} \sigma_{x_1}^2 & cov \\ cov & \sigma_{x_2}^2 \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + 2\left(\frac{\partial f}{\partial x_1}\right)\left(\frac{\partial f}{\partial x_2}\right) cov(x_1, x_2)$$

## 6 Équivalents français de termes anglo-saxons :

<i>Anglais</i>	<i>Français</i>
Acceptance	Acceptation
Bias	Biais
Bivariate normal distribution	Loi normale à deux variables
Confidence level	Niveau de confiance
Consistent estimator	Estimateur convergent
Distribution function	Fonction de répartition
Estimate	Estimation
Expectation	Espérance mathématique
Range	Étendue
Sample	Échantillon
Significance level	Seuil de signification
Standard deviation	Écart-type

<i>Anglais</i>	<i>Français</i>
Unbiased estimator	Estimateur sans biais
Fractile	Fractile

## 7 Pour en savoir plus :

- Data Analysis . Statistical and computational methods for Scientists and Engineers , Siegmund Brandt , Springer 1999
- Review of Particle Properties , Physical Review D
- Aide-Mémoire Statistique , CISIA – CERESTA éditeur
- Statistics for nuclear and particle physicists , Louis Lyons, Cambridge University Press
- Data Reduction and error analysis for the physical sciences, Ph. Bevington , D. Keith Robinson , Mc. Graw-Hill , inc.
- Statistical methods in experimental physics : W.T. Eadie, D. Drijard, F.E. James, M. Roos, B. Sadoulet , North-Holland