

ANALYSE STATISTIQUE DE LA MESURE CAS DES PETITS NOMBRES

JEROME GIOVINAZZO

*Centre d'Etudes Nucléaires de Bordeaux Gradignan
CENBG / IN2P3 / CNRS
Le Haut Vigneau, BP120
F-33175 Gradignan cedex*

Résumé

Ce cours présente des éléments de base de l'analyse statistique des données, appliqués aux expériences de physiques. Il vise plus particulièrement le cas des expériences à très faible nombre d'observations (faible statistique).

Une première partie est consacrée aux concepts de base du calcul des probabilités, qui constitue le cadre formel sur lequel s'appuie l'analyse statistique. Ce chapitre présente les principales propriétés des variables aléatoires et de leurs fonctions de distributions, les lois de convergence, ainsi que quelques distributions de probabilités couramment utilisées en physique.

Le chapitre suivant est une introduction à l'analyse statistique de la mesure. Il décrit, à partir des caractéristiques d'un échantillon statistique, la construction d'estimateurs d'une variable aléatoire, dans le cadre des théorèmes de convergence, ainsi que la détermination d'intervalles de confiance pour les résultats. Les estimateurs les plus courants sont présentés.

La dernière partie présente des applications pratiques très simples illustrant les techniques d'analyse statistique. Les exemples s'appuient sur des expériences de mesures de décroissance radioactives à très faible taux de comptage.

Abstract

This lecture presents basics aspects of statistical data analysis for physics experiments. It is specially focused on the case of experiments with a very low number of observations (the case of low statistics).

A first part is devoted to the basis of the probability theory, which is the framework for statistical analysis. It presents the main properties of random variables and their distribution functions, the convergence laws, and some of the probability distribution functions that are often used in physics.

The next chapter is an introduction to statistical analysis. From the characteristics of a sample of randomly distributed values, it describes how to build an estimator for a random variable, in the framework of the convergence laws, and the corresponding confidence limits. The most usual estimators are introduced.

The last chapter shows simple real applications of the techniques presented in the previous chapter. The examples are taken from experiments measuring radioactive decay experiments with very low counting rates.

1. INTRODUCTION	3
2. CONCEPTS DE BASE DU CALCUL DES PROBABILITES	4
2.1. Aspects de la théorie des probabilités	4
2.1.1. <i>Propriétés des probabilités</i>	<i>4</i>
2.1.2. <i>Approches classiques et modernes.....</i>	<i>6</i>
2.2. Variables aléatoires.....	7
2.2.1. <i>Variables aléatoires discrètes ou continues</i>	<i>7</i>
2.2.2. <i>Caractéristiques d'une distribution de probabilité</i>	<i>7</i>
2.2.3. <i>Propriétés des fonctions de distribution et fonctions caractéristiques.....</i>	<i>7</i>
2.2.4. <i>Covariance, corrélation et indépendance de variables aléatoires</i>	<i>8</i>
2.3. Lois de convergence.....	10
2.3.1. <i>Loi des grands nombres.....</i>	<i>10</i>
2.3.2. <i>Théorème de la limite centrale.....</i>	<i>11</i>
2.4. Quelques distributions courantes de probabilités	12
2.4.1. <i>Loi binomiale</i>	<i>12</i>
2.4.2. <i>Loi de Poisson</i>	<i>13</i>
2.4.3. <i>Loi normale (gaussienne).....</i>	<i>14</i>
2.4.4. <i>Distribution du Chi².....</i>	<i>15</i>
2.4.5. <i>Relations asymptotiques.....</i>	<i>15</i>
3. ANALYSE STATISTIQUE	17
3.1. Remarques préliminaires	17
3.1.1. <i>Erreurs et incertitude</i>	<i>17</i>
3.1.2. <i>Fonction de vraisemblance et information.....</i>	<i>18</i>
3.1.3. <i>Echantillon : moyenne, variance et moments</i>	<i>18</i>
3.2. Estimateur d'une variable aléatoire	19
3.3. Construction d'un estimateur	20
3.3.1. <i>Estimateur intuitif.....</i>	<i>20</i>
3.3.2. <i>Estimateur implicite</i>	<i>21</i>
3.3.3. <i>Estimateurs par maximisation ou minimisation</i>	<i>21</i>
3.3.4. <i>Maximum de vraisemblance.....</i>	<i>22</i>
3.3.5. <i>Méthode des moindres carrés</i>	<i>22</i>
3.4. Comportement asymptotique.....	22
3.4.1. <i>Cas d'un estimateur intuitif</i>	<i>23</i>
3.4.2. <i>Cas d'un estimateur implicite</i>	<i>24</i>
3.5. Intervalle de confiance.....	24
3.5.1. <i>Approche classique : ceinture de confiance.....</i>	<i>24</i>
3.5.2. <i>Approche bayésienne</i>	<i>26</i>
3.5.3. <i>Remarques sur les intervalles de confiance</i>	<i>26</i>
4. APPLICATION AU CAS DES PETITS NOMBRES	27
4.1. Application de la loi binomiale : rapport d'embranchement	27
4.1.1. <i>Propriétés de la loi binomiale</i>	<i>27</i>
4.1.2. <i>Approche classique</i>	<i>28</i>
4.1.3. <i>Approche bayésienne</i>	<i>30</i>
4.2. Application d'une loi continue : mesure de durée de vie	31
4.2.1. <i>Loi de probabilité de la décroissance radioactive</i>	<i>32</i>
4.2.2. <i>Approche classique</i>	<i>32</i>
4.2.3. <i>Approche bayésienne</i>	<i>34</i>
4.2.4. <i>Remarques sur les résultats classique et bayésiens</i>	<i>36</i>
4.2.5. <i>Mesure de durée de vie avec bruit de fond</i>	<i>37</i>
4.3. Conclusion sur les exemples à faible statistique.....	39

1. Introduction

La physique décrit la réalité à l'aide de théories (ou de modèles), qui sont confrontés à l'expérience, c'est à dire à un ensemble de mesures, qui permettent de valider ou d'invalider ces théories.

Il est donc essentiel de pouvoir déterminer si le résultat d'une mesure est significatif, ou le degré de confiance que l'on peut lui accorder.

C'est tout l'objet de l'utilisation de la statistique en physique. L'outil de base de l'analyse statistique est le calcul des probabilités. Mais chaque problème est spécifique, et chacun doit, au regard de son cas particulier, utiliser les méthodes les mieux adaptées.

Les probabilités sont un domaine des **mathématiques pures**, alors que la statistique est un domaine des **mathématiques appliquées**.

Les probabilités permettent de déterminer le comportement de variables aléatoires à partir des lois de distribution auxquelles ces variables obéissent.

La statistique part d'une observation, c'est à dire un échantillon de variables aléatoires, et cherche à retrouver la loi de probabilité correspondante (ou, de façon plus générale, les paramètres de cette loi). La statistique donne une interprétation de l'observation : elle n'est pas exempte d'une certaine subjectivité, surtout dans le domaine des petits nombres.

Il n'est pas possible ici de prétendre être exhaustif. Ce cours se limite à rappeler les concepts de base du calcul des probabilités et quelques grandes lignes de l'analyse statistique. Ils sont illustrés par des exemples concernant des expériences à petit nombre de mesures, pour lesquelles les interprétations sont plus difficiles à donner.

Parmi les points qui ne sont pas abordés ici, les plus importants sont les tests d'hypothèses (qui découlent de la théorie de la décision) et le cas des régressions (ajustement de courbes). De plus, la théorie de l'information, dont découle la notion de statistique, est à peine effleurée. Pour ces dernières néanmoins, le physicien dispose des outils informatiques solides tels que PAW ou ROOT, basés sur les algorithmes de minimisation MINUIT, développés au CERN.

Enfin, une difficulté à laquelle ce cours est confronté est celle du langage : en effet, physiciens et statisticiens ont des jargons différents. Dans la mesure du possible, j'essaie dans ce cours d'utiliser une terminologie de physicien.

Conseils de lecture :

- W.T. Eadie, D. Drijard, F.E. Fames, M. Roos, B. Sadoulet, *Statistical Methods in Experimental Physics*, North Holland (1971)
- Philip. R. Bevington, D. Keith Robinson, *Data Reduction and Error Analysis*, 3^e édition, McGraw-Hill (2003)
- Particle Data Group, *Journal of Physics G* **33** (2006), URL: <http://pdg.lbl.gov/>

2. Concepts de base du calcul des probabilités

Ce chapitre constitue une brève introduction au calcul des probabilités. Il présente des définitions et propriétés de base des probabilités, la notion de probabilité conditionnelle et les cadres de l'interprétation d'une observation.

2.1. Aspects de la théorie des probabilités

2.1.1. Propriétés des probabilités

On considère une expérience dont le résultat peut être ou non d'un type donné X . On répète N fois l'expérience, et on obtient n fois que le résultat est de type X . La probabilité $P(X)$ qu'un événement soit de type X peut être définie par (définition en fréquence):

$$P(X) = \lim_{N \rightarrow \infty} \frac{n}{N}$$

Si on suppose l'ensemble exhaustif des résultats possibles X_i ($i = 1, \dots, n$) pour une observation (un événement), où ces résultats sont supposés exclusifs (si on obtient un résultat X_i , on n'obtient pas simultanément un résultat X_j avec $j \neq i$), alors la probabilité $P(X_i)$ d'obtenir un résultat X_i vérifie :

$$\begin{cases} P(X_i) \geq 0 \\ P(X_i \text{ ou } X_j, i \neq j) = P(X_i) + P(X_j) \\ \sum_i P(X_i) = 1 \end{cases}$$

Loi d'addition

On définit A et B comme 2 sous-ensembles des X_i .

La probabilité qu'une observation se produise dans A ou dans B est alors :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probabilité conditionnelle (loi de multiplication) et indépendance

On définit la probabilité conditionnelle $P(A|B)$ comme la probabilité qu'une observation se produise dans A , alors qu'elle appartient à B .

Alors la probabilité qu'une observation se produise dans A et B est donnée par (loi de multiplication) :

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

Si la réalisation de A est indépendante de celle de B , alors on a (c'est une condition nécessaire et suffisante) : $P(A|B) = P(A)$ et

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

Illustration géométrique (par analogie en termes de surfaces)

On considère des événements correspondant à des points dans le plan $x-y$, avec x et y compris entre 0 et 1, où tous les points sont équiprobables. On définit 2 zones A et B par leurs limites dans ce plan. On peut calculer les probabilités en terme de surfaces, rapportées à la surface totale S (qui vaut 1, donc on se ramène à des surfaces sans unité) : S_A est la surface de la zone A , S_B celle de B et S_{AB} celle de l'intersection de A et B : $P(A) = S_A$ et $P(B) = S_B$.

Cas 1 : **A** est défini par $\{y_{min}^A \leq y \leq y_{max}^A\}$ et **B** par $\{x_{min}^B \leq x \leq x_{max}^B\}$ (figure de gauche)

Surfaces : $S_A = 1 \times (y_{max}^A - y_{min}^A)$
 $S_B = (x_{max}^B - x_{min}^B) \times 1$
 $S_{AB} = (x_{max}^B - x_{min}^B) \times (y_{max}^A - y_{min}^A) = S_A \cdot S_B$

Probabilités : $P(A \cap B) = S_{AB} = S_A \cdot S_B = P(A|B) \cdot P(B)$
 on a donc $P(A|B) = \frac{S_{AB}}{P(B)} = \frac{S_{AB}}{S_B} = S_A = P(A)$

A et **B** sont donc indépendants.

Cas 2 : **A** est défini par $\{y_{min}^A \leq y \leq y_{max}^A\}$ et **B** par $\{x_{min}^B \leq x \leq x_{max}^B \text{ et } y_{min}^B \leq y \leq y_{max}^B\}$ (figure de droite)

Surfaces : $S_A = 1 \times (y_{max}^A - y_{min}^A)$
 $S_B = (x_{max}^B - x_{min}^B) \times (y_{max}^B - y_{min}^B)$
 $S_{AB} = (x_{max}^B - x_{min}^B) \times (y_{max}^A - y_{min}^A)$

Probabilités : $P(A \cap B) = S_{AB} = P(A|B) \cdot P(B)$
 donc $P(A|B) = \frac{S_{AB}}{P(B)} = \frac{S_{AB}}{S_B} = \frac{y_{min}^A - y_{max}^A}{y_{min}^B - y_{max}^B} \neq P(A)$

A et **B** ne sont donc pas indépendants.

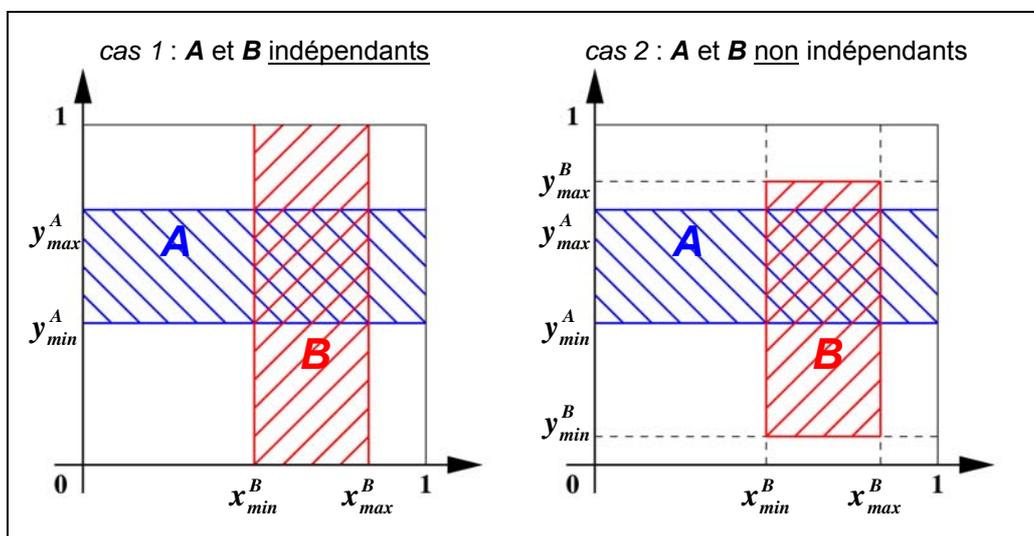


figure 2-1. Illustration géométrique de l'indépendance de deux probabilités. A gauche (cas 1), la probabilité qu'un point se trouve dans la zone A est indépendante de la probabilité qu'il se trouve dans la zone B. A droite (cas 2), ces probabilités ne sont pas indépendantes : la probabilité qu'un point se trouve dans la zone A n'est pas la même selon que ce point se trouve ou non dans la zone B.

Théorème de Bayes

La loi de multiplication donne une relation entre les probabilités conditionnelles $P(A|B)$ et $P(B|A)$, pour des événements discrets :

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

On peut généraliser cette relation en considérant les A_1, \dots, A_n des sous-ensembles d'événements, où les A_i sont exclusifs et exhaustifs (une observation appartient à un et un seul sous-ensemble A_i), et un type B quelconque d'événements. Alors la probabilité qu'une observation appartienne à A_i alors que B est réalisé est donnée par:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1, \dots, n} P(B|A_j) \cdot P(A_j)}$$

2.1.2. Approches classiques et modernes

Lorsqu'à partir d'une observation (un événement de type B), on veut tester une hypothèse (l'événement est de type A_i), il se pose un problème d'interprétation (domaine de la statistique). Deux approches s'opposent : celles des *classiques* (*anti-bayésiens*) ou des *modernes* (*bayésiens*).

Pour illustrer cette opposition (et la rendre plus concrète), on suppose faire une mesure donnant un résultat X , que l'on veut interpréter dans le cadre d'une théorie. Cette théorie correspond à l'ensemble des hypothèses θ_i possibles pour l'observation (θ_i peut par exemple représenter les valeurs, discrètes ou continues, des paramètres de la théorie).

Les *bayésiens* généralisent le cas précédent aux hypothèses, et la probabilité que l'hypothèse θ_i soit vraie si on a observé X est donc:

$$P(\theta_i|X) = \frac{P(X|\theta_i) \cdot P(\theta_i)}{P(X)}$$

avec :

- $P(\theta_i|X)$ probabilité que l'hypothèse soit vraie alors qu'on a observé X (interprétation statistique du résultat)
- $P(X|\theta_i)$ probabilité d'observer X dans le cadre de l'hypothèse θ_i (calcul de probabilité, donné par la théorie dans laquelle on interprète le résultat)
- $P(\theta_i)$ probabilité d'une hypothèse donnée (degré de confiance à *priori* de l'hypothèse)
- $P(X)$ probabilité d'observer X quelle que soit l'hypothèse considérée

Si l'ensemble des hypothèses θ_i est exhaustif et exclusif (c'est généralement le cas en physique : un paramètre de la théorie ou du modèle a une valeur et une seule), alors :

$$P(X) = \sum_j P(X|\theta_j) \cdot P(\theta_j)$$

La controverse entre *bayésiens* (ou modernes) et *anti-bayésiens* (ou classiques) vient de $P(\theta_i)$: pour les *anti-bayésiens*, cette probabilité à *priori* ne peut être déterminée.

Il en résulte que l'interprétation du résultat X d'une observation, c'est à dire l'information que l'on en tire à *posteriori*, n'est pas la même. Cela sera illustré dans les exemples à faible statistique. L'opposition entre *bayésiens* et *anti-bayésiens* est au cœur de la question de la question statistique : c'est la question de l'interprétation.

Remarques :

Dans certains cas, l'approche *anti-bayésienne* ne permet pas de conclure.

Dans les expériences avec un grand nombre d'observations, les deux hypothèses conduisent aux mêmes interprétations.

En règle générale, l'approche classique, lorsqu'elle est possible, est préférable puisqu'elle évite la subjectivité sur le terme $P(\theta_i)$.

2.2. Variables aléatoires

Un événement aléatoire est un événement pouvant donner plusieurs résultats, à chacun desquels peut être associée une probabilité.

Une variable associée à un tel événement (grandeur physique) peut donc prendre un ensemble de valeurs, qui peut être discret ou continu.

2.2.1. Variables aléatoires discrètes ou continues

Pour une variable aléatoire discrète x qui peut prendre les valeurs $x_i, i = 1, \dots, n$, on associe à chaque valeur possible une probabilité $P(x_i)$ (avec $0 \leq P(x_i) \leq 1$) d'obtenir le résultat x_i pour un événement.

Si les valeurs x_i sont exhaustives, on a $\sum_{i=1}^n P(x_i) = 1$

Pour une variable aléatoire continue x qui peut prendre les valeurs sur un domaine Ω , on associe à chaque valeur x' de Ω une probabilité $P(x')dx'$ (avec $0 \leq P(x') \leq 1$) d'obtenir un résultat compris entre x' et $x' + dx'$ pour un événement ($P(x')$ est la fonction de distribution de probabilité associée à la variable aléatoire).

Si le domaine Ω est exhaustif, on a $\int_{\Omega} P(x')dx' = 1$

2.2.2. Caractéristiques d'une distribution de probabilité

On considère une variable aléatoire discrète ou continue X , distribuée selon une loi $P(X)$.

Pour une variable discrète, on considère l'ensemble des valeurs X_k représentant toutes les valeurs possibles de X .

Pour une variable continue, on considère le domaine Ω_x sur lequel la variable X est définie.

Les principales grandeurs caractéristiques d'une variable aléatoire X (discrète ou continue) sont présentées dans le tableau ci-dessous.

Variable X	discrète	continue
Espérance $E[X]$	$\mu = \sum_k X_k \cdot P(X_k)$	$\mu = \int_{\Omega_x} x \cdot P(x) \cdot dx$
Variance $V[X]$	$\sigma^2 = \sum_k (X_k - \mu)^2 \cdot P(X_k)$	$\sigma^2 = \int_{\Omega_x} (x - \mu)^2 \cdot P(x) \cdot dx$
Moments $\mu'_j(X)$	$\mu'_j = \sum_k (X_k)^j \cdot P(X_k)$	$\mu'_j = \int_{\Omega_x} x^j \cdot P(x) \cdot dx$
Moments centraux $\mu_j(X)$	$\mu_j = \sum_k (X_k - \mu)^j \cdot P(X_k)$	$\mu'_j = \int_{\Omega_x} (x^j - \mu) \cdot P(x) \cdot dx$

Tableau récapitulatif des grandeurs caractéristiques pour des variables aléatoires discrètes ou continues.

On notera que l'espérance est le moment d'ordre 1 et que la variance est le moment central d'ordre 2. On a de plus la relation : $\sigma^2 = E[(x - \mu)^2] = E[x^2] - \mu^2$.

2.2.3. Propriétés des fonctions de distribution et fonctions caractéristiques

Soit une variable aléatoire x , de probabilité de distribution $f(x)$ sur un domaine exhaustif Ω . Alors toute fonction $g(x)$ de la variable aléatoire est elle-même une variable aléatoire. L'espérance (ou valeur attendue) de $g(x)$ est alors :

$$E[g] = \int_{\Omega} g(x) \cdot f(x) \cdot dx$$

L'espérance est un opérateur linéaire. On obtient sans difficulté :

$$\begin{aligned} E[\alpha g + \beta h] &= \alpha \cdot E[g] + \beta \cdot E[h] \\ &= \alpha \int_{\Omega} g(x) \cdot f(x) \cdot dx + \beta \int_{\Omega} h(x) \cdot f(x) \cdot dx \end{aligned}$$

On définit la fonction caractéristique de la distribution $f(x)$ par sa transformée de Fourier :

$$\begin{aligned} \Phi_x(t) = E[e^{itx}] &= \int_{\Omega} e^{itx} \cdot f(x) \cdot dx && \text{pour une variable continue} \\ &= \sum_k p_k \cdot e^{itx} && \text{pour une variable discrète} \end{aligned}$$

La fonction caractéristique définit entièrement la fonction de distribution de probabilité :

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

Elle vérifie les propriétés suivantes :

- si α et β sont des constantes, alors $\Phi_{\alpha x + \beta}(t) = e^{it\beta} \cdot \Phi_x(\alpha t)$
- si x et y sont 2 variables aléatoires indépendantes de fonctions caractéristiques $\Phi_x(t)$ et $\Phi_y(t)$ alors $\Phi_{x+y}(t) = \Phi_x(t) \cdot \Phi_y(t)$

Les moments $\mu'_j(\mathbf{X})$ et les moments centraux $\mu_j(\mathbf{X})$ de la distribution peuvent également être déterminés à partir des dérivées de la fonction caractéristique :

$$\mu'_r = \frac{1}{i^r} \cdot \left[\left(\frac{d}{dt} \right)^r \Phi(t) \right]_{t=0} \quad \text{et} \quad \mu_r = \frac{1}{i^r} \cdot \left[\left(\frac{d}{dt} \right)^r [e^{-i\mu t} \cdot \Phi(t)] \right]_{t=0}$$

2.2.4. Covariance, corrélation et indépendance de variables aléatoires

On s'intéresse ici au cas de plusieurs variables aléatoires. Dans un premier temps, on considère le cas de 2 variables aléatoires x et y , mais les propriétés présentées ici peuvent être généralisées à un nombre quelconque de variables. On leur associe une densité de probabilité conjointe $f(x, y)$.

Si Ω est un domaine exhaustif pour les valeurs possibles de x et y , alors

$$\iint_{\Omega} f(x, y) \cdot dx \cdot dy = 1$$

et toute fonction $g(x, y)$ est une variable aléatoire, donc l'espérance est donnée par :

$$E[g] = \iint_{\Omega} g(x, y) \cdot f(x, y) \cdot dx \cdot dy$$

On définit alors les distributions marginales de probabilité pour x et y (probabilité de la variable x quelle que soit la valeur de y et inversement) :

$$f_x(x) = \int_{y_{\min}}^{y_{\max}} f(x, y) \cdot dy \quad \text{et} \quad f_y(y) = \int_{x_{\min}}^{x_{\max}} f(x, y) \cdot dx$$

On définit également des probabilités conditionnelles, comme la probabilité pour la variable y alors que la variable x a une valeur donnée x_0 :

$$p(y|x_0) = \frac{f(x_0, y)}{\int_{y_{\min}(x_0)}^{y_{\max}(x_0)} f(x_0, y') \cdot dy'}$$

Les distributions marginales et probabilités conditionnelles sont illustrées sur la figure 2-2.

Le **théorème de Bayes** peut alors s'écrire, à partir des probabilités conditionnelles et des distributions marginales : $p(y|x) = \frac{p(x|y) \cdot f_y(y)}{f_x(x)}$.

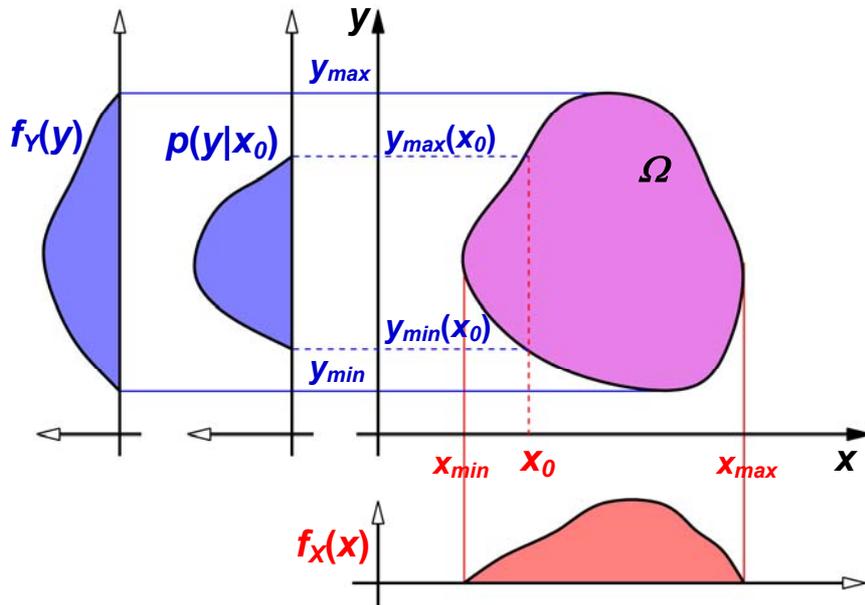


figure 2-2. Illustration des distributions marginales pour les variables x et y , et de la probabilité conditionnelle de la variable y pour une valeur donnée x_0 de la variable x .

Pour chacune des variables, on peut calculer sa moyenne et sa variance :

$$\mu_x = E[x] = \iint_{\Omega} x \cdot f(x, y) \cdot dx \cdot dy \quad \text{et} \quad \mu_y = E[y] = \iint_{\Omega} y \cdot f(x, y) \cdot dx \cdot dy$$

$$(\sigma_x)^2 = E[(x - \mu_x)^2] \quad (\sigma_y)^2 = E[(y - \mu_y)^2]$$

On peut alors définir la covariance des variables x et y , ainsi que leur corrélation, par :

- covariance : $\text{cov}[x, y] = E[(x - \mu_x) \cdot (y - \mu_y)] = E[xy] - E[x] \cdot E[y]$
- corrélation : $\text{corr}[x, y] = \rho(x, y) = \frac{\text{cov}[x, y]}{\sigma_x \cdot \sigma_y}$

On démontre facilement que la corrélation (ou coefficient de corrélation) est compris entre -1 et 1 (inclus).

Les variables aléatoires x et y sont *indépendantes* si et seulement si (condition nécessaire et suffisante) leur distribution conjointe est le produit des distributions marginales :

$$f(x, y) = f_x(x) \cdot f_y(y)$$

alors

$$E[xy] = \iint_{\Omega} x \cdot y \cdot f_x(x) \cdot f_y(y) \cdot dx \cdot dy$$

$$= \int x \cdot f_x(x) \cdot dx \times \int y \cdot f_y(y) \cdot dy$$

$$= E[x] \cdot E[y]$$

On en déduit immédiatement que dans ce cas, $\text{cov}[x, y] = 0$ et $\rho(x, y) = 0$. Si deux variables sont indépendantes, elles ne sont alors pas corrélées (la réciproque n'est pas nécessairement vraie).

On peut généraliser la définition de la covariance à un ensemble de n variables aléatoires x_1, \dots, x_n . Elles sont indépendantes si et seulement si la fonction de densité de probabilité peut être factorisée sous la forme :

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$$

Pour chaque couple de variables (x_i, x_j) , on définit la covariance σ_{ij} et le coefficient de corrélation ρ_{ij} . La matrice de covariance est alors la matrice $\overline{\mathbf{V}} = [\sigma_{ij}]$. Si la matrice de covariance n'est pas inversible, alors il existe au moins une relation linéaire entre les différentes variables x_i .

2.3. Lois de convergence

Il existe plusieurs théorèmes de convergence, mais on n'en rappelle ici que deux applications, qui sont la loi des grands nombres et le théorème de la limite centrale (TLC). Ces lois de convergence sont très importantes pour l'analyse statistique, mais la démonstration n'est pas discutée dans ce cours.

2.3.1. Loi des grands nombres

On considère ici un ensemble de variables aléatoires X_i (avec $i = 1, \dots, N$), telles que :

- elles ont la même espérance μ ;
- elles ont des variances σ_i (non nécessairement identiques).

D'un point de vue statistique, la moyenne de l'échantillon (c'est à dire d'un ensemble de valeurs des X_i) est donnée par : $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$.

Alors la loi des grands nombres stipule que, sous certaines conditions (sur les variances), la *moyenne de l'échantillon* tend vers l'*espérance des variables aléatoires* lorsque N devient très grand : $\bar{x} \xrightarrow{N \rightarrow \infty} \mu$:

- loi convergence faible (convergence en moyenne quadratique) si

$$\lim_{N \rightarrow \infty} \left(\frac{1}{N^2} \sum_{i=1}^N \sigma_i^2 \right) = 0 ;$$

- loi convergence forte (convergence presque certaine) si $\lim_{N \rightarrow \infty} \left(\sum_{i=1}^N \left(\frac{\sigma_i}{i} \right)^2 \right)$ est finie.

On note que dans les cas (fréquents en physique) où les variances sont bornées ($\sigma_i < \sigma_{\max}$), et donc également dans le cas où elles sont identiques (par exemple si tous les X_i suivent la même loi de distribution, comme illustré sur la figure 2-3), la loi des grands nombres s'applique.

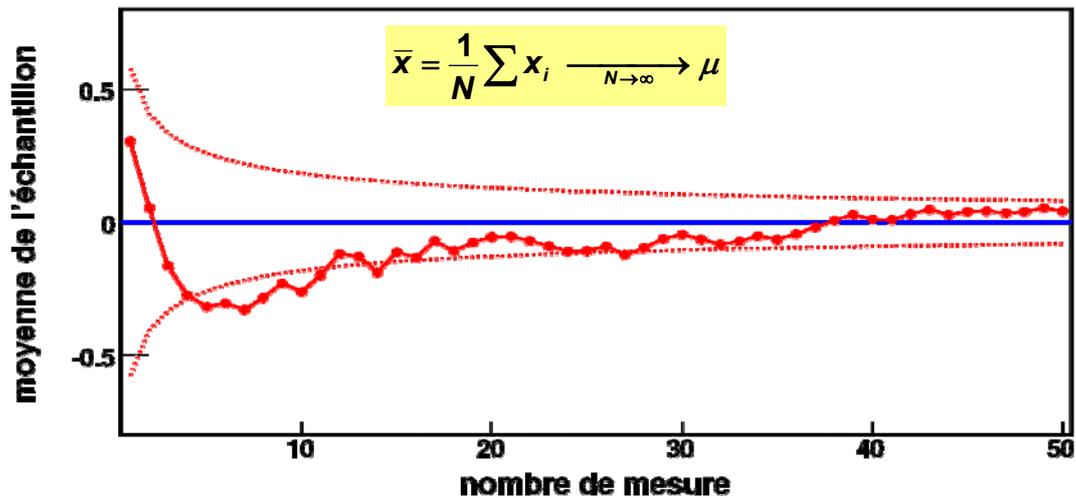


figure 2-3. Illustration de la loi des grands nombres : évolution de la moyenne d'un ensemble de N valeurs d'une variable aléatoire (de moyenne $\mu = 0$ et de variance $\sigma^2 = 1$).

2.3.2. Théorème de la limite centrale

On considère toujours un ensemble de variables aléatoires X_i (avec $i = 1, \dots, N$), d'espérances μ_i et de variances σ_i^2 . Alors, sous des conditions peu restrictives sur les μ_i et σ_i^2 (limitant leur accroissement lorsque i devient grand, ce qui est presque toujours le cas dans les problèmes courants de physique où ces grandeurs prennent des valeurs bornées), le *théorème de la limite centrale* donne le comportement de la variable aléatoire $S = \sum_{i=1}^N x_i$

lorsque N tend vers l'infini : $\left(\frac{S - \sum_{i=1}^N \mu_i}{\sqrt{\sum_{i=1}^N \sigma_i^2}} \right) \xrightarrow{N \rightarrow \infty} \mathcal{N}(0,1)$ où $\mathcal{N}(0,1)$ est la fonction de

distribution de probabilité gaussienne, de centre 0 et d'écart-type 1 (*loi normale standard*, cf. §2.4.3). Une illustration du TCL est donnée sur la figure 2-4.

Ce résultat est indépendant de la distribution a priori des X_i .

Cas particulier

Si les X_i (indépendants) ont la même espérance μ et la même variance σ^2 , la variable

$$z_N = \frac{\bar{x} - \mu}{\sigma / \sqrt{N}} \quad (\text{avec } \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i) \text{ est telle que } E[z_N] = 0 \text{ et } V[z_N] = 1.$$

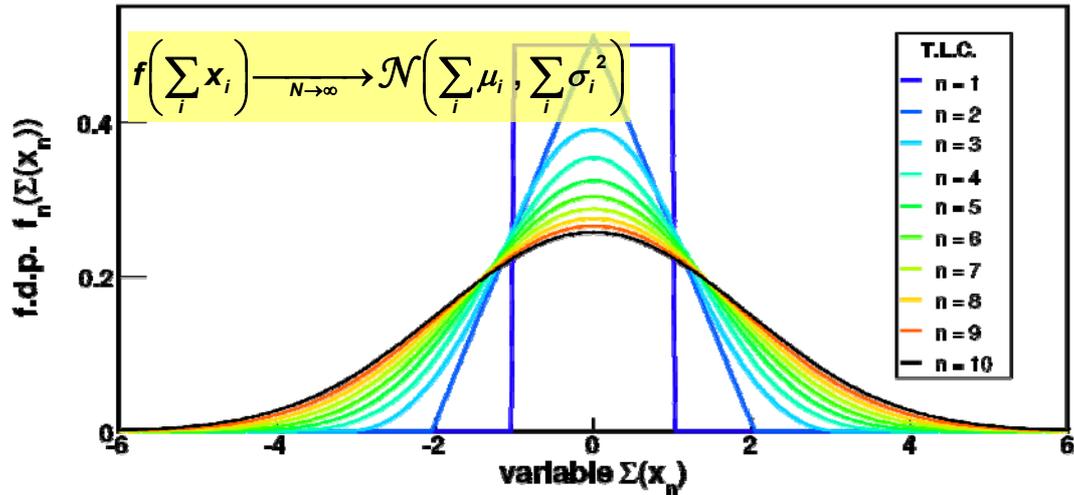


figure 2-4. Illustration du **théorème de la limite centrale** : loi de distribution de la variable aléatoire $S = \sum_{i=1}^N x_i$ (dans le cas où tous les X_i suivent la même loi de distribution). Elle tend vers une distribution gaussienne.

2.4. Quelques distributions courantes de probabilités

Dans ce paragraphe, je présente quelques lois de distribution de probabilités qui sont couramment utilisées en physiques. J'ai fait le choix de deux distributions pour des variables discrètes (loi de Poisson et loi binomiale) et de deux distributions pour des variables continues (loi normale et distribution du Chi2).

2.4.1. Loi binomiale

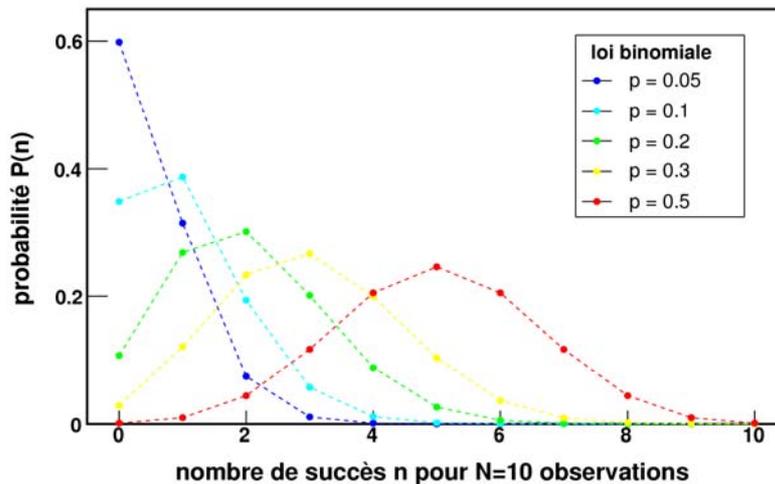


figure 2-5. Illustration de la loi binomiale pour différentes valeurs de la probabilité p et pour un nombre de tirages constant $N = 10$.

La loi binomiale est utilisée lorsqu'un tirage aléatoire a deux résultats possibles (par exemple le résultat est soit un succès, soit un échec). Si p est la probabilité, pour un tirage, d'obtenir un succès (la probabilité d'un échec est alors $1 - p$), la loi binomiale donne la probabilité d'obtenir n succès lorsqu'on procède à N tirages :

$$B(n, p|N) = \frac{N!}{n!(N-n)!} p^n \cdot (1-p)^{N-n}$$

La variable aléatoire (discrète) est le nombre de succès n ($0 \leq n \leq N$) et les paramètres de la distribution sont le nombre de tirage N (entier, $N > 0$) et la probabilité p (réelle, $0 \leq p \leq 1$).

Espérance de la loi binomiale : $E[n] = Np$

Variance de la loi binomiale : $V[n] = Np(1 - p)$

Parmi les lois de distribution dérivant de la loi binomiale, on peut citer :

- la *loi binomiale négative*, qui donne la probabilité de devoir attendre N tirages pour obtenir n succès ;
- la *loi multinomiale*, qui est la généralisation de la loi binomiale lorsque le tirage a plus de deux issues possibles.

2.4.2. Loi de Poisson

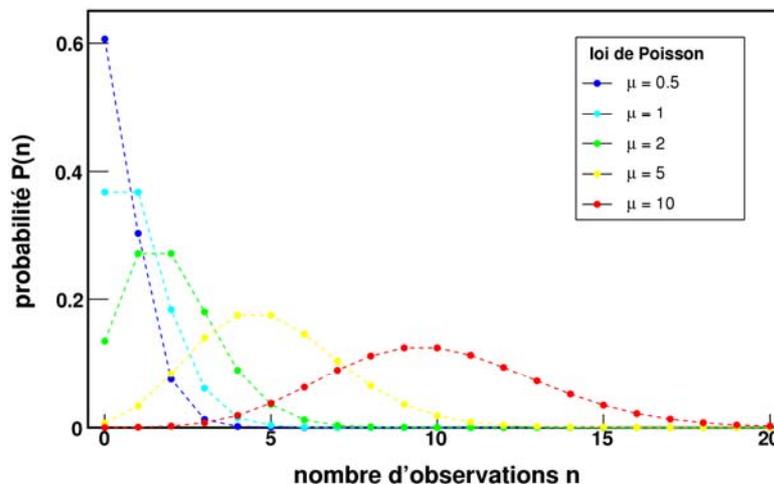


figure 2-6. Illustration de la loi de Poisson.

La loi de Poisson donne la probabilité d'observer n événements lorsqu'on en attend (en théorie) λ :

$$P(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$$

Variable (discrète) : n (≥ 0), nombre d'événements observés

Paramètre : λ (réel, $\lambda > 0$), nombre d'événements attendus

Espérance de la loi de Poisson : $E[n] = \lambda$

Variance de la loi de Poisson : $V[n] = \lambda$

Une loi (sur une variable aléatoire continue) très utile découle de la loi de Poisson : la *loi exponentielle*, qui mesure la probabilité d'avoir un intervalle de temps t entre deux événements consécutifs, lorsque les événements sont indépendants et qu'ils se produisent avec une fréquence moyenne ν :

$$P(t|\nu) = \nu \cdot e^{-\nu t}$$

$$\text{avec } E[t] = \frac{1}{\nu} \text{ et } V[t] = \frac{1}{\nu^2}$$

2.4.3. Loi normale (gaussienne)

La loi normale (distribution gaussienne) est la fonction de probabilité la plus utilisée, en raison notamment du théorème de la limite centrale (cf. §2.3.2). Cette fonction de distribution présente un ensemble de propriétés uniques qui ne sont pour la plupart pas abordées ici. L'expression de la fonction de distribution de probabilité (f.d.p.) pour une variable aléatoire continue x est la suivante :

$$P(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right]$$

Variable (continue) : x , réel

Paramètres : μ (moyenne, réel), σ (écart-type, réel)

Espérance de la loi de normale : $E[x] = \mu$

Variance de la loi de normale : $V[x] = \sigma^2$

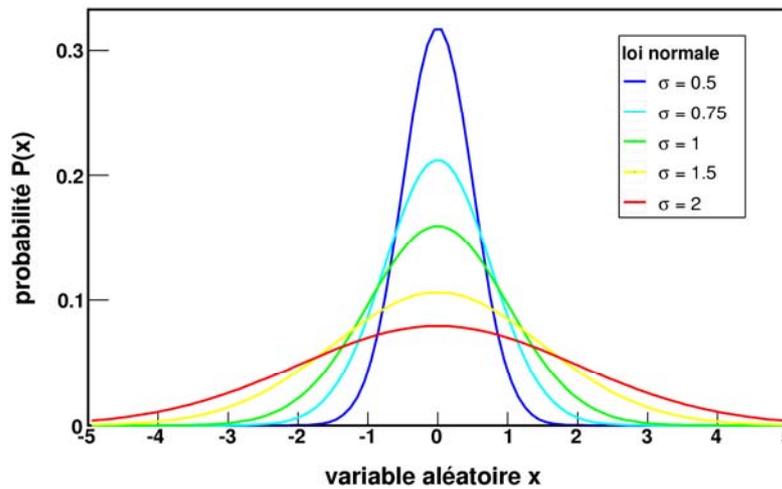


figure 2-7. Illustration de la loi normale pour une valeur moyenne $\mu = 0$ et différentes valeurs de l'écart-type.

On parle de *loi normale « standard »* dans le cas particulier où $\mu = 0$ et $\sigma = 1$.

Parmi les propriétés de la loi normale, on note :

- que toute combinaison linéaire de variables aléatoires normales est une variable aléatoire normale ;

- que les moments centraux impairs sont nuls :
$$\begin{cases} \mu_{2r-1} = 0 \\ \mu_{2r} = \frac{(2r)!}{r!} \cdot \left(\frac{\sigma^2}{2}\right)^r \end{cases}$$

On peut généraliser la loi normale à plusieurs dimensions, sous la forme (pour k dimensions) :

$$P(\vec{x}) = \frac{1}{(2\pi)^{k/2} \cdot \sqrt{|\mathbf{V}_{kk}|}} \exp\left[-\frac{1}{2} (\vec{x} - \vec{\mu})^T (\mathbf{V}_{kk})^{-1} (\vec{x} - \vec{\mu})\right]$$

où \mathbf{V}_{kk} est la matrice ($k \times k$) de covariance des composantes du vecteur de variables aléatoires \vec{x} .

2.4.4. Distribution du Chi2

Pour un ensemble de variables aléatoires indépendantes X_i (avec $i = 1, \dots, n$), distribuées selon une loi normale standard ($\mu = 0$ et $\sigma = 1$), la variable aléatoire $S = s_{(n)}^2 = \sum_{i=1}^n (x_i)^2$ est distribuée selon la loi du Chi2 (χ^2) à n degrés de libertés :

$$P(S|n) = \frac{1}{2} \cdot \frac{(s/2)^{\frac{n}{2}-1}}{\Gamma(n/2)} \cdot e^{-s/2}$$

Variable (continue) : $S = s^2$, réel

Paramètre : $n \geq 1$ (entier, nombre de degrés de liberté)

Espérance de la distribution du χ^2 : $E[S] = n$

Variance de la distribution du χ^2 : $V[S] = 2n$

Propriétés

Si on considère les variables aléatoires $s_{(n)}^2$ et $s_{(m)}^2$ indépendantes et qui suivent des distributions du χ^2 à n et m degrés de liberté, alors la variable $s_{(p)}^2 = s_{(n)}^2 + s_{(m)}^2$ suit une distribution du χ^2 à $p = n + m$ degrés de liberté.

De façon générale, la distribution du χ^2 tend vers une loi normale. Plus spécifiquement, les variables aléatoires $y_n = \frac{s_{(n)}^2 - n}{\sqrt{2n}}$ et $z_n = \sqrt{2s_{(n)}^2} - \sqrt{2n-1}$ tendent vers une loi normale standard $\mathcal{N}(0,1)$.

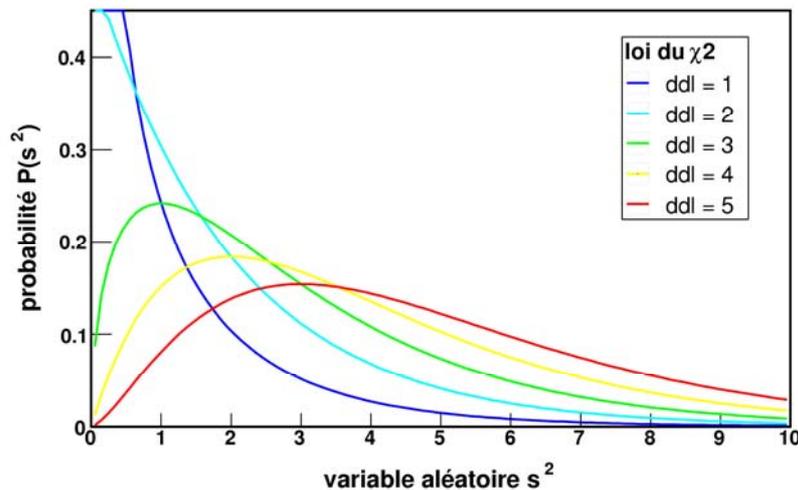


figure 2-8. Illustration de la distribution du Chi2 pour différents nombres de degrés de liberté.

2.4.5. Relations asymptotiques

La figure 2-9 présente les relations entre les lois de probabilité présentées précédemment. Ainsi, on constate qu'elles convergent toutes vers la distribution normale lorsque le nombre d'observations tend vers l'infini (conséquence du TLC).

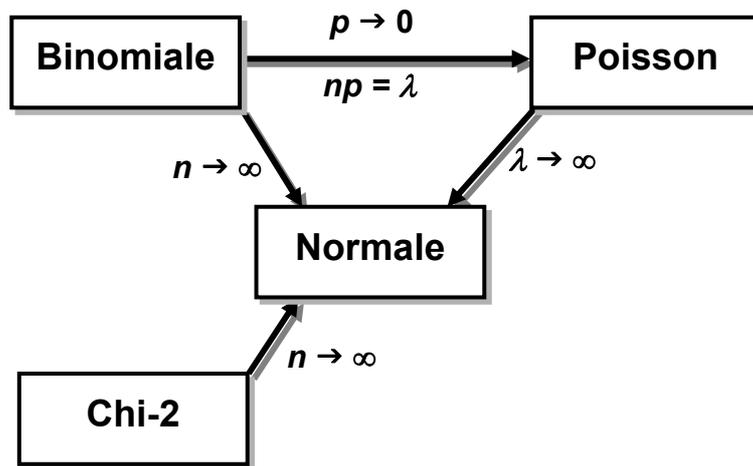


figure 2-9. Relations asymptotiques entre les lois de distributions présentées dans ce chapitre.

3. Analyse statistique

Dans le chapitre précédent, nous avons posé les bases du calcul des probabilités. En pratique (en particulier dans l'analyse des expériences de physique), on ne connaît pas le détail des lois que suivent les grandeurs mesurées. Les résultats expérimentaux (mesures) fournissent un échantillon, qui représente un ensemble de mesures des grandeurs étudiées. Ces grandeurs sont les variables aléatoires dont on cherche à déterminer les propriétés.

Après quelques remarques introduisant des notions utiles, ce chapitre s'intéresse principalement à la théorie de l'estimation (de paramètres d'une théorie). La loi des grands nombres est utilisée pour construire des estimateurs. L'incertitude associée à ces estimateurs, dans le cas des grands nombres, est déduite du théorème de la limite centrale qui donne le comportement asymptotique de la variance des estimateurs. Dans le cas des faibles statistiques, le dernier paragraphe présente les outils pour la détermination d'intervalles de confiance.

3.1. Remarques préliminaires

L'objectif de ce paragraphe est d'évoquer quelques concepts qui ne sont pas étudiés dans ce cours, mais dont il peut être utile d'avoir connaissance :

- la notion d'erreur est centrale dans l'analyse des données, mais n'est abordée dans ce cours que d'un point de vue purement statistique ;
- la théorie de l'information n'est pas détaillée, mais il sera fait référence à la fonction de vraisemblance (qui en découle) dans la construction des estimateurs (cf. §3.3.3 et §3.4)

3.1.1. Erreurs et incertitude

A la notion de mesure est intimement associée la notion d'*erreur*. Ce terme est utilisé en physique par abus de langage : il fait référence à la notion d'*incertitude*, qui reflète le degré de confiance que l'on attribue à une mesure. Il recouvre en fait plusieurs significations qu'il convient de distinguer : la précision d'une mesure d'une part, et son exactitude d'autre part (voir figure 3-1).

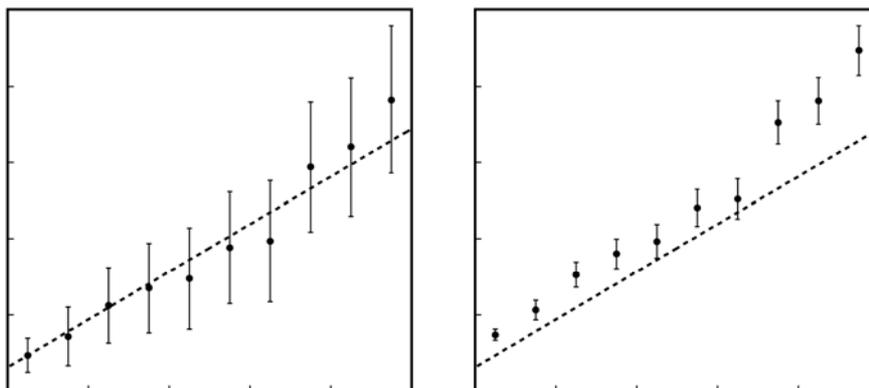


figure 3-1. Sur ces figures, on suppose que les points représentent des mesures et que la ligne pointillée représente le comportement « vrai » de la grandeur mesurée. La figure de gauche illustre le cas d'une mesure correcte mais imprécise (grande incertitude sur les points), celle de droite une mesure précise mais inexacte.

Pour l'analyse des données, on distingue généralement deux sources d'erreurs :

- l'incertitude statistique, liée à la taille de l'échantillon de données ;

- l'erreur systematique, liée à la méthode utilisée pour faire la mesure (dispositif instrumental, méthode de sélection et d'analyse des données, fiabilité de la théorie dans le cadre de laquelle les données sont interprétées,...).

Si les résultats d'un ensemble de mesures sont distribués selon une loi de probabilité donnée, l'incertitude statistique est d'autant plus faible que le nombre de mesures est grand. Il est donc possible d'augmenter la « précision statistique » en augmentant le nombre de mesures ! En principe, cette incertitude peut être calculée dans le cadre des théories de la statistique, ce qui est abordé dans ce chapitre.

Mais il n'existe pas de théorie des erreurs systématiques : comme il a été mentionné, ce type d'erreur peut avoir des origines très variées. Chaque cas est particulier et implique une très bonne compréhension du processus qui conduit de l'expérience aux données à analyser. Les erreurs systématiques ne peuvent donc pas faire l'objet de ce cours.

Il est à noter que si les erreurs systématiques sont sur-estimées, cela conduit à un résultat correct mais imprécis, alors que si elles sont sous-estimées, le résultat sera relativement précis, mais il risque d'être faux !

3.1.2. Fonction de vraisemblance et information

La théorie de l'information ne fait pas l'objet de ce cours. Néanmoins, les définitions de la fonction de vraisemblance et de l'information sont données ici, parce qu'elles interviennent dans la construction des estimateurs abordée plus loin.

On suppose que \mathbf{x} est une variable aléatoire de fonction de densité de probabilité $f(\mathbf{x}, \vec{\theta})$, où la notation vectorielle $\vec{\theta}$ indique un ensemble de paramètres θ_1 à θ_r . On considère un échantillon de N observations (ou mesures) de \mathbf{x} . Cet échantillon constitue un vecteur $\vec{\mathbf{x}}$.

La fonction de vraisemblance est alors définie par :

$$L(\vec{\mathbf{x}}|\vec{\theta}) = L(\mathbf{x}_1, \dots, \mathbf{x}_N|\vec{\theta}) = \prod_{i=1}^N f(\mathbf{x}_i, \vec{\theta})$$

Elle peut être interprétée comme la « probabilité globale » d'observer l'échantillon lorsque chaque mesure suit individuellement la distribution $f(\mathbf{x}, \vec{\theta})$.

On définit la quantité d'information contenue dans l'échantillon $\vec{\mathbf{x}}$, relativement au paramètre unique θ (dans le cas à 1 dimension) par :

$$I_{\vec{\mathbf{x}}}(\theta) = E \left[\left(\frac{dL(\vec{\mathbf{x}}|\theta)}{d\theta} \right)^2 \right] = \int_{\Omega_{\theta}} \left(\frac{dL(\vec{\mathbf{x}}|\theta)}{d\theta} \right)^2 L(\vec{\mathbf{x}}|\theta) d\vec{\mathbf{x}}$$

Cette relation se généralise dans le cas à plusieurs dimensions ($\vec{\theta}$ est alors un ensemble de paramètres). L'information contenue dans l'échantillon prend alors la forme matricielle :

$$[I_{\vec{\mathbf{x}}}(\vec{\theta})]_{ij} = E \left[\frac{dL(\vec{\mathbf{x}}|\vec{\theta})}{d\theta_i} \cdot \frac{dL(\vec{\mathbf{x}}|\vec{\theta})}{d\theta_j} \right]$$

3.1.3. Echantillon : moyenne, variance et moments

On considère \mathbf{x} comme une variable aléatoire de probabilité $P(\mathbf{x})$ (qui peut être discrète ou continue). On fait N observations de cette variable aléatoire, $\mathbf{x}_1, \dots, \mathbf{x}_N$.

La moyenne de la distribution est la limite : $\mu = \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \right)$

Pour une observation \mathbf{x}_i , la déviaton est $\mathbf{d}_i = \mathbf{x}_i - \mu$.

La moyenne de la déviation $\bar{d} = \frac{1}{N} \sum_{i=1}^N d_i$, tend vers **0** lorsque **N** tend vers l'infini (par définition de la moyenne μ).

On peut définir la déviaton moyenne α , comme une mesure de la dispersion des observations autour de la valeur moyenne μ : $\alpha = \frac{1}{N} \sum_{i=1}^N |x_i - \mu|$, mais cette définition n'est pas commode à utiliser dans les analyses.

On définit donc une autre grandeur, la déviaton standard σ (ou écart-type), et la variance σ^2 , par :

$$\sigma^2 = \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \right) = \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N x_i^2 \right) - \mu^2$$

On peut également définir la valeur médiane $\mu_{1/2}$ telle que $P(x > \mu_{1/2}) = P(x < \mu_{1/2}) = 1/2$, et le mode (ou valeur la plus probable) μ_{max} telle que quel que soit X, $P(x) \leq P(\mu_{max})$.

Les moments de la distribution $P(x)$ sont (μ_j désigne le moment d'ordre j) :
 $\mu_j = \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N (x_i)^j \right)$. On voit notamment que $\mu = \mu_1$ et $\sigma^2 = \mu_2 - (\mu_1)^2$.

Remarque

Pour l'échantillon correspondant à **N** observations x_1 à x_N , la moyenne de l'échantillon est $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ et sa variance $s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$. Le terme **N - 1** plutôt que **N** au dénominateur vient du fait que la moyenne et la variance sont estimées avec les mêmes données, et ne sont donc pas indépendantes.

3.2. Estimateur d'une variable aléatoire

On considère la variable aléatoire x de fonction de densité de probabilité $f(x, \theta_0)$, ou le vecteur θ_0 représente un paramètre de la distribution (cela peut être généralisé au cas de plusieurs paramètres), et un ensemble de **N** mesures (échantillon) $\bar{x}_N = \{x_i, i = 1, \dots, N\}$. Que peut-on dire de θ_0 à partir de l'échantillon mesuré ?

Un estimateur $\hat{\theta}(\bar{x}_N)$ est une fonction de l'échantillon qui tente de donner une valeur s'approchant de la « vraie » valeur du paramètre θ_0 . Un estimateur est donc lui-même une variable aléatoire !

Un estimateur doit être :

- **Consistant** : il doit converger vers la vraie valeur du paramètre ;
- **Non biaisé** : l'estimateur, pour **N** observations est une variable aléatoire ; on définit le biais par $b_N(\hat{\theta}) = E(\hat{\theta}) - \theta_0 = E(\hat{\theta} - \theta_0)$: l'estimateur est non biaisé si $b_N(\hat{\theta}) = 0$ quel que soit le nombre d'observations **N** ;
- **Efficace** : s'il converge *rapidement* vers la valeur du paramètre ;
- **Robuste** : cette caractéristique est liée à la fiabilité de l'estimateur si la f.d.p. est peu ou mal connue (c'est la sensibilité à la forme de la f.d.p.).

3.3. Construction d'un estimateur

Ce paragraphe présente les principes généraux qui permettent de construire un estimateur pour un paramètre d'une distribution aléatoire, à partir de la loi des grands nombres.

Pour une variable aléatoire \mathbf{X} distribuée selon une loi de probabilité $f(\mathbf{x}, \theta_0)$, où θ_0 représente un paramètre (vrai) de la distribution, une fonction quelconque $\mathbf{a}(\mathbf{X})$ est également une variable aléatoire, et si sa variance est finie, la loi des grands nombres implique alors que, pour un ensemble de mesures (échantillon) \mathbf{x}_i , avec $i = 1$ à N :

$$\lim_{N \rightarrow \infty} \left[\frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i) \right] = E[\mathbf{a}(\mathbf{X})|\theta_0] = \int_{\Omega_x} \mathbf{a}(\mathbf{x})f(\mathbf{x}, \theta_0)d\mathbf{x}$$

(La notation $E[\mathbf{a}(\mathbf{X})|\theta_0]$ indique l'espérance de la variable (fonction) aléatoire $\mathbf{a}(\mathbf{X})$, sachant que \mathbf{X} est une variable aléatoire distribuée selon $f(\mathbf{X}, \theta_0)$).

En d'autres termes, la grandeur expérimentale $\xi = \frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i)$ converge vers $E[\mathbf{a}(\mathbf{X})|\theta_0]$.

3.3.1. Estimateur intuitif

Si on peut trouver une fonction $\mathbf{a}(\mathbf{X})$ telle que $E[\mathbf{a}(\mathbf{X})|\theta] = \mathbf{h}(\theta)$ ou $\mathbf{h}(\theta)$ est une fonction connue, alors la valeur « vraie » θ_0 du paramètre θ est (si \mathbf{h} est inversible à la valeur θ_0), de façon triviale : $\theta_0 = \mathbf{h}^{-1}(E[\mathbf{a}(\mathbf{X})|\theta_0])$.

Tout le problème réside dans le fait que ne connaissant pas à priori la valeur « vraie » de θ_0 , on ne connaît pas $E[\mathbf{a}(\mathbf{X})|\theta_0]$. On peut alors considérer que la grandeur ξ , obtenue à partir de l'échantillon (les mesures), en donne une valeur *approchée*. Par conséquent, un estimateur intuitif (par la loi des grands nombre, il est *consistant*) $\hat{\theta}$ de θ_0 est donc :

$$\hat{\theta} = \mathbf{h}^{-1}\left(\frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i)\right) = \mathbf{h}^{-1}(\xi)$$

Dans le cas de plusieurs paramètres, $\vec{\theta} = (\theta_j, j = 1 \text{ à } r)$, il est alors nécessaire de définir r fonctions $\mathbf{a}_j(\mathbf{x})$, telles que $E[\mathbf{a}_j(\mathbf{X})|\vec{\theta}] = \mathbf{h}_j(\vec{\theta})$, où les $\mathbf{h}_j(\vec{\theta})$ sont des fonctions connues.

Alors un estimateur $\hat{\theta}$ la valeur vraie des paramètres $\vec{\theta}_0$ est telle que, quel que soit j :

$$\xi_j = \frac{1}{N} \sum_{i=1}^N \mathbf{a}_j(\mathbf{x}_i) = \mathbf{h}_j(\hat{\theta}_1, \dots, \hat{\theta}_r) \text{ pour } j = 1, r$$

Il s'agit donc de résoudre un système de r équations à r inconnues (les θ_j). Il s'agit donc d'inverser le système sous la forme :

$$\begin{cases} \xi_1 = \mathbf{h}_1(\hat{\theta}_1, \dots, \hat{\theta}_r) \\ \vdots \\ \xi_r = \mathbf{h}_r(\hat{\theta}_1, \dots, \hat{\theta}_r) \end{cases} \Leftrightarrow \begin{cases} \hat{\theta}_1 = (\mathbf{h}^{-1})_1(\xi_1, \dots, \xi_r) \\ \vdots \\ \hat{\theta}_r = (\mathbf{h}^{-1})_r(\xi_1, \dots, \xi_r) \end{cases}$$

Il est à noter que dans cette écriture $(\mathbf{h}^{-1})_j$ n'est (en général) pas l'inverse de la fonction \mathbf{h}_j . Par ailleurs, sauf dans quelques cas simples, la fonction-vecteur \mathbf{h} est difficilement inversible analytiquement.

Méthode des moments

Elle correspond au choix : $\mathbf{a}_j(\mathbf{X}) = \mathbf{X}^j$ et le système d'équations à résoudre est donc :

$$\xi_j = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i)^j = \int_{\Omega_x} \mathbf{x}^j \cdot f(\mathbf{x}, \hat{\theta}) \cdot d\mathbf{x} = h_j(\hat{\theta})$$

Dans le cas où la variable est distribuée selon une loi normale de centre μ et de variance σ^2 , on a de façon très simple un estimateur des paramètres $\vec{\theta} = (\mu, \sigma)$ par la méthode des moments :

$$\begin{cases} \xi_1 = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i & \xrightarrow{N \rightarrow \infty} \mu = h_1(\mu, \sigma) = \int_{\Omega_x} \mathbf{x} \cdot \exp\left(-\frac{(\mathbf{x} - \mu)^2}{2\sigma^2}\right) \cdot d\mathbf{x} \\ \xi_2 = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i)^2 & \xrightarrow{N \rightarrow \infty} \sigma^2 + \mu^2 = h_2(\mu, \sigma) = \int_{\Omega_x} \mathbf{x}^2 \cdot \exp\left(-\frac{(\mathbf{x} - \mu)^2}{2\sigma^2}\right) \cdot d\mathbf{x} \end{cases}$$

$$\text{soit : } \begin{cases} \hat{\mu} = (\mathbf{h}^{-1})_1(\xi_1, \xi_2) = \xi_1 \\ \hat{\sigma} = (\mathbf{h}^{-1})_2(\xi_1, \xi_2) = \sqrt{\xi_2 - (\xi_1)^2} \end{cases}$$

3.3.2. Estimateur implicite

De façon plus générale, on peut choisir, à la place de $\mathbf{a}(\mathbf{X})$, une fonction de \mathbf{X} et de θ , de telle sorte que $\mathbf{h}(\theta|\theta_0)$ soit nulle pour $\theta = \theta_0$:

$$\mathbf{h}(\theta|\theta_0) = \mathbf{E}[\mathbf{a}(\mathbf{X}, \theta)|\theta_0] = \int_{\Omega_x} \mathbf{a}(\mathbf{x}, \theta) f(\mathbf{x}, \theta_0) d\mathbf{x} \text{ avec } \mathbf{h}(\theta_0|\theta_0) = \mathbf{0}$$

La loi des grands nombres implique donc que

$$\left[\frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i, \theta = \theta_0) \right] \xrightarrow{N \rightarrow \infty} \mathbf{E}[\mathbf{a}(\mathbf{X}, \theta_0)] = \mathbf{0}$$

On obtient alors un estimateur implicite $\hat{\theta}$ de θ_0 en résolvant ($\xi(\theta)$ est une fonction de l'observation):

$$\xi(\theta) = \frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i, \theta) = \mathbf{0}$$

L'estimateur est consistant sous certaines conditions sur la fonction $\xi(\theta)$ (dérivabilité, dérivée non nulle en θ_0).

Des fonctions du type : $\mathbf{a}(\mathbf{X}, \theta) = \mathbf{a}(\mathbf{X}) - \mathbf{E}[\mathbf{a}(\mathbf{X})|\theta]$ satisfont naturellement la condition $\mathbf{h}(\theta_0) = \mathbf{0}$.

3.3.3. Estimateurs par maximisation ou minimisation

En pratique, on peut remplacer la résolution de $\xi(\theta) = \mathbf{0}$ par une minimisation ou une maximisation de fonction, en considérant une fonction $\mathbf{g}(\mathbf{X}, \theta)$ telle que :

$$\mathbf{a}(\mathbf{X}, \theta) = \frac{d}{d\theta} \mathbf{g}(\mathbf{X}, \theta)$$

On obtient alors un estimateur $\hat{\theta}$ de θ_0 comme solution de $\xi(\theta) = \frac{1}{N} \frac{d}{d\theta} \sum_{i=1}^N \mathbf{g}(\mathbf{x}_i, \theta) = \mathbf{0}$.

C'est notamment le cas du maximum de vraisemblance ou de la méthode des moindres carrés.

Dans le cas de plusieurs paramètres, $\vec{\theta} = (\theta_j, j = 1 \text{ à } r)$, on définit alors les r fonctions

$\mathbf{a}_j(\mathbf{X}, \theta) = \frac{d}{d\theta_j} \mathbf{g}(\mathbf{X}, \theta)$ et le système de r équations à résoudre est donc :

$$\xi_j(\bar{\theta}) = \frac{1}{N} \frac{d}{d\theta_j} \sum_{i=1}^N g(x_i, \bar{\theta}) = 0 \quad \text{pour } j = 1 \text{ à } r$$

3.3.4. Maximum de vraisemblance

Il s'agit d'un cas particulier (très fréquemment utilisé) d'estimateur par maximisation. Il correspond au choix suivant pour la fonction g : $g(\mathbf{x}, \theta) = \ln f(\mathbf{x}, \theta)$.

L'estimation du paramètre est obtenue en résolvant l'équation (ou un système d'équation dans le cas de plusieurs paramètres) :

$$\xi(\theta) = \frac{1}{N} \frac{d}{d\theta} \ln \left(\prod_{i=1}^N f(x_i, \theta) \right) = \frac{1}{N} \frac{d}{d\theta} \ln(L(\bar{\mathbf{x}}, \theta)) = 0$$

On peut montrer que l'estimateur est consistant sous certaines conditions de dérivabilité (ce qui est généralement le cas en physique) et de commutation des opérateurs $\frac{d}{d\theta}$ et $\int d\mathbf{x}$.

3.3.5. Méthode des moindres carrés

La méthode des moindres carrés est un cas d'estimateur implicite par minimisation. Dans le cas de paramètres multiples $\bar{\theta}$, la fonction g prend la forme quadratique suivante :

$$g(\bar{\mathbf{x}}, \bar{\theta}) = [\bar{\mathbf{x}} - \bar{\mathbf{M}}(\bar{\theta})]_{(N)}^T \cdot \bar{\mathbf{W}}_{(N \times N)} \cdot [\bar{\mathbf{x}} - \bar{\mathbf{M}}(\bar{\theta})]_{(N)}$$

où $\bar{\mathbf{M}}(\bar{\theta}) = \mathbf{E}[\bar{\mathbf{x}} | \bar{\theta}]$ est le vecteur (de dimension N) des valeurs attendues pour les mesures,

et $\bar{\mathbf{W}}$ est une matrice de poids (de dimension $N \times N$).

Le système d'équations à résoudre pour obtenir une estimation des paramètres est alors :

$$\xi_j(\bar{\theta}) = \frac{\partial}{\partial \theta_j} (g(\bar{\mathbf{x}}, \bar{\theta})) = \frac{\partial}{\partial \theta_j} [\bar{\mathbf{M}}(\bar{\theta})] \cdot \bar{\mathbf{W}} \cdot [\bar{\mathbf{x}} - \bar{\mathbf{M}}(\bar{\theta})] = 0$$

Cette méthode est principalement utilisée dans le cas des régressions (ajustements de courbes à un ensemble de points expérimentaux), en remplaçant formellement les valeurs \mathbf{x}_i par des points expérimentaux (x_i, y_i) et les termes $\mathbf{M}_i(\bar{\theta})$ par les valeurs attendues (théoriques) $\mathbf{y}^{\text{théo}}(\mathbf{x}_i, \bar{\theta})$. La matrice de poids est alors en général la matrice des incertitudes expérimentales sur les grandeurs y_i (matrice de covariance).

3.4. Comportement asymptotique

Un estimateur $\hat{\theta}$ est une variable aléatoire, avec par conséquent une espérance (valeur attendue) et une variance. Pour un estimateur non biaisé pour toute valeur de N (nombre d'observations), $\mathbf{E}[\hat{\theta}] = \theta_0$ et la variance $\mathbf{V}[\hat{\theta}]$ donne une mesure de l'incertitude sur l'estimateur.

Le détail du formalisme pour le calcul du comportement asymptotique de la variance des estimateurs n'est pas dans les objectifs de ce cours, dans la mesure où cela s'applique pour les grands nombres d'observations. Ce paragraphe a essentiellement pour but d'en présenter les bases (dans le cas d'un estimateur intuitif), afin d'éclairer sur les calculs mis en œuvre dans les outils standards (dans nos disciplines) tels que PAW ou ROOT.

Le comportement asymptotique (lorsque N tend vers l'infini) de la variance et du biais peuvent être déduit du *théorème de la limite centrale* qui implique un comportement asymptotiquement gaussien (normal) de la variable aléatoire ξ (ou de la fonction $\xi(\theta)$ dans le cas d'un estimateur implicite).

On utilise un développement de Taylor autour de la valeur « vraie » θ_0 du paramètre. La variance est obtenue à partir d'un développement au premier ordre, et le formalisme général

est présenté ci-dessous. Le biais résulte de la prise en compte des ordres supérieurs, et le cas général n'est pas détaillé ici.

3.4.1. Cas d'un estimateur intuitif

D'après le TLC, la « grandeur expérimentale » $\xi = \frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i)$ est asymptotiquement (lorsque N tend vers l'infini) distribuée selon une loi normale, de moyenne $\mathbf{E}[\mathbf{a}(\mathbf{X})|\theta_0]$ et de variance $N^{-1} \cdot \mathbf{V}[\mathbf{a}(\mathbf{X})|\theta_0]$. On fait un développement limité de l'estimateur $\hat{\theta} = \mathbf{h}^{-1}(\xi)$ autour de la valeur $\xi_0 = \mathbf{E}[\mathbf{a}(\mathbf{X})|\theta_0]$ (notée $\mathbf{E}[\mathbf{a}]$ pour simplifier l'écriture):

$$\begin{aligned} \hat{\theta} &\approx \mathbf{h}^{-1}(\xi_0) + \frac{\partial \mathbf{h}^{-1}(\xi_0)}{\partial \xi} \cdot \{\xi - \xi_0\} + \dots \\ &\approx \mathbf{h}^{-1}(\mathbf{E}[\mathbf{a}]) + \frac{\partial \mathbf{h}^{-1}(\mathbf{E}[\mathbf{a}])}{\partial \xi} \cdot \left\{ \left(\frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i) \right) - \mathbf{E}[\mathbf{a}] \right\} + \dots \end{aligned}$$

Dans cette relation, tous les termes sont constants à l'exception de $\xi = \frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i)$. La variance de l'estimateur est alors (en rappelant que $\theta_0 = \mathbf{h}^{-1}(\xi_0) = \mathbf{h}^{-1}(\mathbf{E}[\mathbf{a}])$):

$$\mathbf{V}[\hat{\theta}] = \mathbf{E}[(\hat{\theta} - \theta_0)^2] = \left[\frac{\partial \mathbf{h}^{-1}(\mathbf{E}[\mathbf{a}])}{\partial \xi} \right]^2 \cdot \frac{\mathbf{V}[\mathbf{a}]}{N}$$

Bien entendu, $\mathbf{E}[\mathbf{a}]$ et $\mathbf{V}[\mathbf{a}]$ ne sont en général pas connus, puisque θ_0 est inconnu ! On détermine par conséquent une *estimation* de la variance (et donc de l'incertitude statistique) à partir de la moyenne et de la variance de l'échantillon.

Dans le cas de plusieurs paramètres, il faut passer à une écriture vectorielle :

$$\begin{aligned} \mathbf{a}(\mathbf{X}) &\rightarrow \bar{\mathbf{a}}(\mathbf{X}) = \{\mathbf{a}_j(\mathbf{X}), j = 1, \dots, r\} \\ \theta_0 &\rightarrow \bar{\theta}_0 \\ \mathbf{h}(\theta) &\rightarrow \bar{\mathbf{h}}(\bar{\theta}) = \{\mathbf{h}_j(\bar{\theta}) = \mathbf{E}[\mathbf{a}_j(\mathbf{X})|\bar{\theta}]\} \\ \xi &\rightarrow \bar{\xi} = \left\{ \xi_j = \frac{1}{N} \sum_{i=1}^N \mathbf{a}_j(\mathbf{x}_i) \right\} \end{aligned}$$

Si on appelle $\bar{\mathbf{k}}(\bar{\alpha})$ la fonction inverse de $\bar{\mathbf{h}}(\bar{\theta})$ (c'est à dire $\bar{\mathbf{k}}(\bar{\mathbf{h}}(\bar{\theta})) = \bar{\theta}$) et $\bar{\alpha}_0 = \bar{\mathbf{h}}(\bar{\theta}_0)$, on a alors, pour un développement au 1^{er} ordre autour de $\bar{\alpha}_0$:

$$\hat{\theta} = \bar{\mathbf{k}}(\bar{\xi}) = \bar{\mathbf{k}}(\bar{\alpha}_0) + \bar{\nabla}_{\alpha} \bar{\mathbf{k}}(\bar{\alpha}_0) \cdot (\bar{\xi} - \bar{\alpha}_0) + \dots$$

La variance (matrice de covariance) de $\hat{\theta}$ (l'estimateur de $\bar{\theta}_0$) est donc :

$$\begin{aligned} \mathbf{V}[\hat{\theta}] &= \mathbf{E} \left[(\hat{\theta} - \bar{\theta}_0) \cdot (\hat{\theta} - \bar{\theta}_0)^T \right] \\ &= \frac{1}{N} [\bar{\nabla}_{\alpha} \bar{\mathbf{k}}(\bar{\alpha}_0)] \cdot \bar{\mathbf{V}}[\bar{\mathbf{a}}|\bar{\theta}_0] \cdot [\bar{\nabla}_{\alpha} \bar{\mathbf{k}}(\bar{\alpha}_0)]^T \end{aligned}$$

De la même manière que précédemment, $\bar{\alpha}_0$ et $\bar{\mathbf{V}}[\bar{\mathbf{a}}|\bar{\theta}_0]$ ne sont pas connus, et la variance de l'estimateur doit être estimée à partir des moyennes et de la matrice de covariance calculées pour l'échantillon.

3.4.2. Cas d'un estimateur implicite

Le même type de raisonnement peut être mené dans le cas d'un estimateur implicite : à une dimension, la grandeur ξ devient une fonction du paramètre : $\xi(\theta)$, et le terme $\frac{\partial h^{-1}(E[\mathbf{a}])}{\partial \xi}$

est remplacé par $\frac{1}{E\left[\frac{d\xi(\theta)}{d\theta}\right]_{\theta=\theta_0}}$ (l'inversion des dérivées n'est pas aussi immédiate dans le

cas de plusieurs paramètres). La variance de l'estimateur est alors obtenue par :

$$V[\hat{\theta}] = \frac{V[a(X, \theta_0)]}{N \cdot \left(E\left[\frac{d\xi(\theta)}{d\theta}\right]_{\theta=\theta_0}\right)^2}$$

Dans le cas du maximum de vraisemblance, cette expression prend la forme suivante :

$$V[\hat{\theta}] = \frac{1}{N \cdot I_N(\theta_0)}$$

où $I_N(\theta_0)$ est l'information telle qu'elle a été définie précédemment (cf. §3.1.2).

3.5. Intervalle de confiance

On a vu dans les paragraphes précédents que lorsque le nombre de mesures est très grand ($N \rightarrow \infty$), la loi des grands nombres permet de déduire un estimateur consistant, et le théorème de la limite centrale donne une estimation de l'incertitude.

Qu'en est-il lorsque le nombre d'observation est petit ? On ne peut alors plus s'appuyer sur les comportements asymptotiques. Dans une telle situation, on ne cherche plus la valeur d'un estimateur et sa variance, mais un intervalle dans lequel le paramètre recherché a une certaine probabilité de se trouver.

On se limite ici à la recherche d'intervalles de confiance pour un seul paramètre (problème à 1 dimension).

3.5.1. Approche classique : ceinture de confiance

On considère la variable aléatoire \mathbf{X} distribuée selon une loi de probabilité $f(\mathbf{x}, \theta_0)$, où θ_0 représente un paramètre (vrai) de la distribution. On dispose d'un ensemble de mesures (échantillon) $\vec{\mathbf{x}} = \{\mathbf{x}_i, i = 1, \dots, N\}$.

On considère de plus un estimateur $\mathbf{t} = \mathbf{t}(\vec{\mathbf{x}})$ du paramètre θ_0 , pour l'échantillon (cf. §3.3).

En tant que fonction des \mathbf{x}_i , $\mathbf{t}(\vec{\mathbf{x}})$ est une variable aléatoire, avec une densité de probabilité $f_t(\mathbf{t}|\theta)$ qui peut être déterminée en fonction de θ .

On peut alors définir 2 fonctions $\mathbf{t}_1(\theta)$ et $\mathbf{t}_2(\theta)$ telles que :

- la probabilité d'obtenir un estimateur \mathbf{t} inférieur à $\mathbf{t}_1(\theta)$ si le paramètre vrai vaut θ est α_1 : $P(\mathbf{t} < \mathbf{t}_1(\theta)|\theta) = \int_{-\infty}^{\mathbf{t}_1(\theta)} f_t(\mathbf{t}|\theta) \cdot d\mathbf{t} = \alpha_1$
- la probabilité d'obtenir un estimateur \mathbf{t} supérieur à $\mathbf{t}_2(\theta)$ si le paramètre vrai vaut θ est α_2 : $P(\mathbf{t} > \mathbf{t}_2(\theta)|\theta) = \int_{\mathbf{t}_2(\theta)}^{+\infty} f_t(\mathbf{t}|\theta) \cdot d\mathbf{t} = \alpha_2$

Alors la probabilité d'obtenir un estimateur compris entre $\mathbf{t}_1(\theta)$ et $\mathbf{t}_2(\theta)$ est :

$$\beta = 1 - (\alpha_1 + \alpha_2) \text{ soit } \int_{t_1(\theta)}^{t_2(\theta)} f_t(t|\theta) \cdot dt = \beta$$

La vraie valeur du paramètre, θ_0 est bien entendue inconnue. Mais à partir de la valeur obtenue pour l'estimateur $t(\bar{x})$, on peut définir un intervalle $[\theta_A, \theta_B]$ tel que :

$$\begin{aligned} t_1(\theta_B) &= t(\bar{x}) \\ t_2(\theta_A) &= t(\bar{x}) \end{aligned} \quad (\text{voir figure 3-2})$$

figure 3-2)

Par construction des fonctions t_1 et t_2 , la probabilité d'avoir un estimateur t pour l'échantillon tel que $t_1(\theta_0) < t(\bar{x}) < t_2(\theta_0)$ est β . Or pour toute valeur de t comprise entre $t_1(\theta_0)$ et $t_2(\theta_0)$, l'intervalle $[\theta_A, \theta_B]$ obtenu contient la valeur « vraie » du paramètre (voir figure 3-2). La probabilité de construire un intervalle $[\theta_A, \theta_B]$ qui contienne la vraie valeur θ_0 est donc β . Dit autrement, la probabilité que θ_0 soit dans l'intervalle $[\theta_A, \theta_B]$ est β .

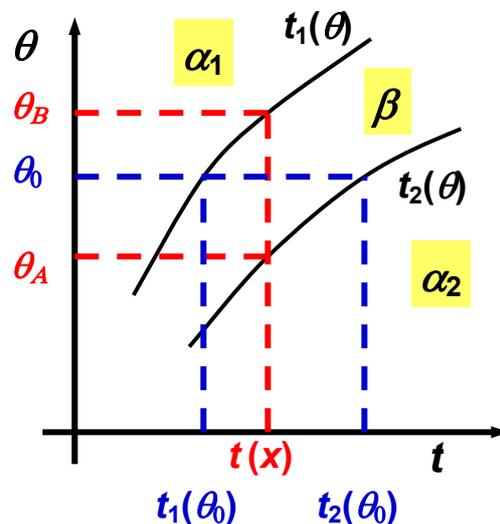


figure 3-2. Construction d'un intervalle de confiance à partir de l'estimateur $t(x)$: l'intervalle $[\theta_A, \theta_B]$ est construit verticalement. Horizontalement, la probabilité que l'estimateur obtenu soit compris entre $t_1(\theta_0)$ et $t_2(\theta_0)$ est β , donc la probabilité de construire un intervalle qui contienne la vraie valeur θ_0 est également β .

Les courbes $t_1(\theta)$ et $t_2(\theta)$ définissent une *ceinture de confiance* de probabilité β : c'est l'ensemble des intervalles de confiance pour les valeurs possibles de l'estimateur.

L'intervalle de confiance $[\theta_A, \theta_B]$ (avec un degré de confiance β) obtenu pour la valeur de l'estimateur $t(\bar{x})$ n'est pas unique. En règle générale, on fait le choix symétrique, c'est à dire :

$$\alpha_1 = \alpha_2 = \frac{1 - \beta}{2}$$

D'un point de vue statistique, il n'y pas à priori de choix préférable.

Concernant les choix asymétriques, les cas extrêmes permettent d'obtenir des limites inférieures ou supérieures du paramètre :

- limite inférieure θ_A : elle correspond au choix $\alpha_1 = 0$, soit $t_1 = -\infty$ et $P(\theta_0 > \theta_A) = \beta$
- limite supérieure θ_B : elle correspond au choix $\alpha_2 = 0$, soit $t_2 = +\infty$ et $P(\theta_0 < \theta_B) = \beta$

3.5.2. Approche bayésienne

Dans l'approche bayésienne, la probabilité à *posteriori* $P(\theta|\bar{x})$ que le paramètre « vrai » soit θ lorsqu'on a fait l'observation \bar{x} est déterminée à partir de la probabilité à *priori* $P(\bar{x}|\theta)$ de faire l'observation \bar{x} si le paramètre vaut θ , en utilisant le théorème de Bayes (cf. §2.1.2) :

$$P(\theta|\bar{x}) = \frac{P(\bar{x}|\theta) \cdot P(\theta)}{\int_{\Omega_\theta} P(\bar{x}|\theta') \cdot P(\theta') \cdot d\theta'}$$

L'intervalle $[\theta_A, \theta_B]$ définit un intervalle de confiance avec un degré de confiance β s'il vérifie (voir figure 3-3) :

$$\int_{\theta_A}^{\theta_B} P(\theta|\bar{x}) \cdot d\theta = \beta$$

Tout comme dans l'approche classique, le choix de l'intervalle de confiance n'est pas unique. Si on impose de plus que $P(\theta_A|\bar{x}) = P(\theta_B|\bar{x})$, alors :

- l'intervalle est unique ;
- c'est le plus petit intervalle pour le niveau de confiance β ;
- chaque valeur à l'intérieur de l'intervalle est plus probable que n'importe quelle valeur en dehors de l'intervalle.

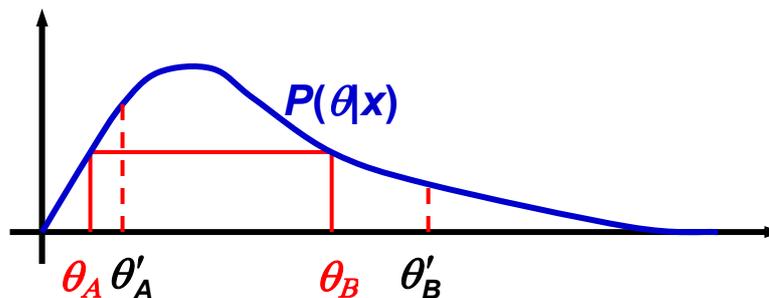


figure 3-3. Détermination de l'intervalle de confiance dans l'approche bayésienne.

Remarque.

Il est cependant à noter que dans les cas complexes où la probabilité $P(\theta|\bar{x})$ présente plusieurs maxima, alors il se peut que l'intervalle de confiance doive être remplacé par la réunion de plusieurs intervalles disjoints.

3.5.3. Remarques sur les intervalles de confiance

Une façon courante, en physique, de présenter le résultat, est d'indiquer la valeur la plus probable $\hat{\theta}_0$ donnée par un estimateur (par exemple le maximum de vraisemblance), et de donner l'intervalle $[\theta_A, \theta_B]$ à 68,3 % de confiance. Par analogie avec la loi normale (gaussienne), on parle alors de l'incertitude à « 1σ » (abus de langage), et on note le résultat sous la forme :

$$\theta = \hat{\theta}_{0-(\hat{\theta}_0-\theta_A)}^{+(\theta_B-\hat{\theta}_0)}$$

Il est à noter que si en physique, c'est généralement un intervalle à 68,3 % qui est donné, ce n'est pas systématiquement le cas, et dans d'autres domaines, les usages sont parfois différents (un cas fréquent est ce donner un intervalle à 95 ou 99 % de confiance). Il est donc préférable de le mentionner explicitement lorsqu'on donne un résultat.

4. Application au cas des petits nombres

Le chapitre précédent présentait le formalisme permettant de déterminer les paramètres d'une théorie à partir d'un ensemble d'observation, et d'estimer les incertitudes sur les résultats. L'objectif du présent chapitre est de mettre ces résultats en application, dans des cas très simples, inspirés d'expériences réelles (mais les données numériques des exemples proposés ne sont pas réelles et ont été générées à des fins illustratives).

Le contexte des exemples proposés est celui d'expériences au cours desquelles on cherche à observer un phénomène très rare, et pour lequel le nombre d'observations est donc petit. C'est le cas des expériences de physique nucléaire menées sur la recherche et l'étude de la radioactivité « 2-protons », un nouveau type de radioactivité observé très récemment pour des noyaux atomiques situés aux limites d'existence, tels que ^{45}Fe ou ^{54}Ni (voir figure 4-1). Ce mode de décroissance est en compétition avec la désintégration β du noyau.

Dans les exemples proposés, on suppose qu'on a pu produire N_{obs} noyaux précurseurs, et que pour chacun de ces noyaux, on a observé sa décroissance, c'est à dire qu'on a identifié le type de décroissance, et l'instant auquel elle a lieu après la production du noyau. Les applications sont faites avec $N_{\text{obs}} = 10$ et un nombre d'événements « 2-protons » $n_{2p} = 7$, et des instants de décroissance simulés avec une durée de vie de 5 ms.

La détermination du rapport d'embranchement (intensité relative) de la radioactivité « 2-protons » ($2P$) illustre l'analyse statistique dans le cas d'une loi discrète, et la mesure de la durée de vie du noyau illustre le cas d'une loi de probabilité continue.

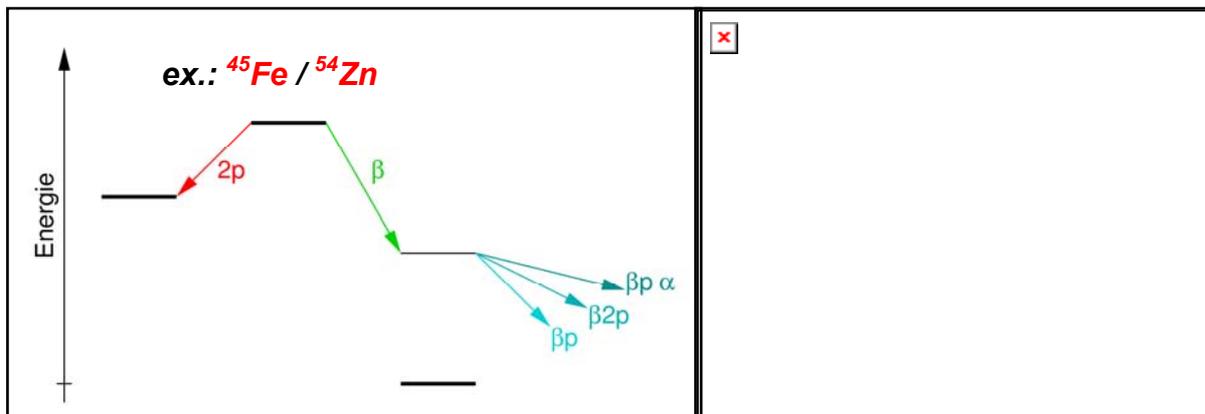


figure 4-1. Contexte de radioactivité « 2-protons ». La figure de gauche montre le schéma de décroissance simplifié : l'état fondamental du noyau précurseur très riche en protons (^{45}Fe ou ^{54}Ni) peut soit décroître par une émission directe et simultanée de 2 protons, soit par une radioactivité β (β^+ ou capture électronique). La figure de droite montre un zoom sur la table des isotopes à la drip-line proton : les flèches indiquant les différents modes de décroissance partent du noyau précurseur ayant un nombre pair de protons : du fait de l'appariement des nucléons (ici les protons), l'émission d'un seul proton n'est pas possible énergétiquement, et dans le cas d'une décroissance β , le noyau produit (avec un proton de moins et un neutron de plus) n'est pas lié, et la radioactivité β est suivie de l'émission de une ou plusieurs particules (proton(s) et/ou alpha).

4.1. Application de la loi binomiale : rapport d'embranchement

4.1.1. Propriétés de la loi binomiale

Pour la détermination du rapport d'embranchement R_{2p} , on considère que chaque décroissance a 2 issues possibles : la radioactivité $2P$ correspond à un succès, et la décroissance β à un échec. Si p est la probabilité de succès pour un événement, alors $p = R_{2p}$.

La loi binomiale (cf. §2.4.1) donne la probabilité d'obtenir n succès lorsqu'on a N événements. Si on note s un succès et e un échec, $P_N\{x_1, \dots, x_N\}$ la probabilité d'un ensemble de résultats (où chaque x_i est soit un succès s soit un échec e), alors $P_N(n)$ la probabilité d'avoir n succès vaut :

- pour $N = 1$: $P_1(0) = P_1\{e\} = 1 - p$
 $P_1(1) = P_1\{s\} = p$
- pour $N = 2$: $P_2(0) = P_2\{e, e\} = P_1\{e\} \times P_1\{e\} = (1 - p)^2$
 $P_2(1) = P_2\{s, e\} + P_2\{e, s\} = 2 \times P_1\{e\} \times P_1\{s\} = 2p(1 - p)$
 $P_2(2) = P_2\{s, s\} = P_1\{s\} \times P_1\{s\} = p^2$
- ...

Si on généralise, on retrouve l'expression de la loi binomiale (voir figure 2-5 pour différentes valeurs de p à nombre d'observations N constant, et figure 4-2 pour une probabilité p constante et un nombre variable d'observations) :

$$P_N(n|p) = \frac{N!}{n!(N-n)!} p^n \cdot (1-p)^{N-n}$$

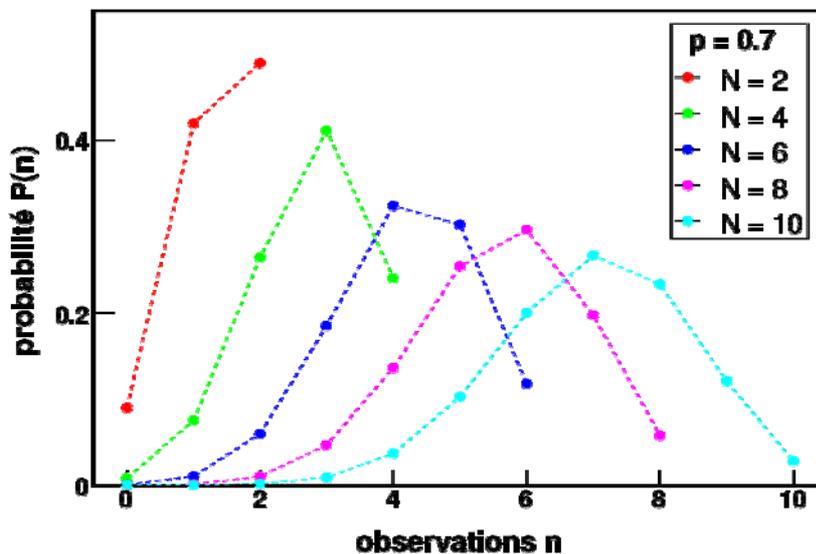


figure 4-2. Loi binomiale pour une probabilité individuelle de succès p fixée, et différents nombres d'observations N .

4.1.2. Approche classique

Le paramètre recherché est la probabilité p de succès, c'est à dire le rapport d'embranchement R_{2P} , que l'on cherche à estimer à partir des observations expérimentales. On peut utiliser le maximum de vraisemblance (M.V.) comme estimateur. Si on a observé n_{2P} succès (radioactivité $2P$) et $N_{obs} - n_{2P}$ échecs (décroissance β) alors la fonction de vraisemblance s'écrit, pour une probabilité p :

$$L(n_{2P}|p = R_{2P}, N_{obs}) = \prod_{i=1}^N P_i = p^{n_{2P}} \cdot (1-p)^{N_{obs}-n_{2P}}$$

L'estimateur est la valeur qui maximise cette fonction, et qui s'écrit alors très simplement :

$$\frac{L(n_{2P}|\hat{p}, N_{obs})}{dp} = 0 \Leftrightarrow \hat{p} = \frac{n_{2P}}{N_{obs}}$$

On détermine la ceinture de confiance de la loi binomiale comme il a été décrit au paragraphe 3.5.1 (voir figure 4-3). La probabilité d'obtenir un estimateur \hat{p} , pour un paramètre vrai p est simplement (puisque $\hat{p} \propto n_{2P}$) : $P(\hat{p}|p) = P(n_{2P}|p, N_{obs})$.

Pour un nombre de succès n_{2P} , (on note p_{2P} la valeur correspondante de l'estimateur), on détermine l'intervalle $[p_A, p_B]$ tel que (voir le principe de la figure 3-2) :

- p_A est telle que : $P\left(\hat{p} \geq p_{2P} = \frac{n_{2P}}{N_{obs}} \mid p_A\right) = \sum_{n=n_{2P}}^{N_{obs}} P(n|p_A, N_{obs}) = \alpha_2$
- p_B est telle que : $P\left(\hat{p} \leq p_{2P} = \frac{n_{2P}}{N_{obs}} \mid p_B\right) = \sum_{n=0}^{n_{2P}} P(n|p_B, N_{obs}) = \alpha_1$

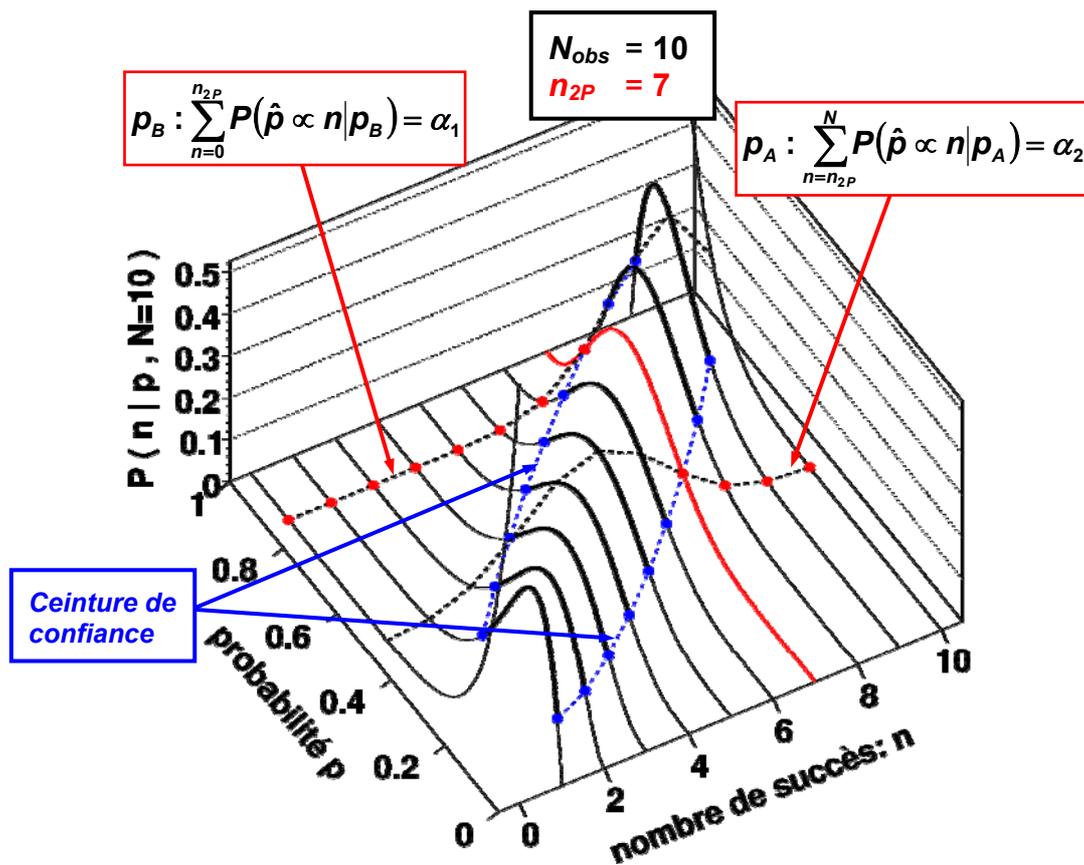


figure 4-3. Ceinture de confiance pour la loi binomiale. Le tracé présente la probabilité $P(n|p, N)$. On détermine pour chaque valeur possible du nombre de succès n les bornes de l'intervalle de confiance $[p_A, p_B]$. Dans cet exemple, pour $n_{2P} = 7$, p_A est tel que $P(n \geq n_{2P} | p_A) = \alpha_2$ et p_B est tel que $P(n \leq n_{2P} | p_B) = \alpha_1$.

Le résultat est obtenu en résolvant numériquement les équations pour déterminer p_A et p_B . Dans notre exemple, pour 10 noyaux produits, dont 7 qui décroissent par radioactivité 2-protons, l'intervalle de confiance symétrique ($\alpha_1 = \alpha_2 = (1 - \beta) / 2$) pour un degré de confiance à « 1σ » ($\beta = 68,3\%$), on a alors :

- estimateur : $R_{2P} = 0,7$
- intervalle de confiance : $[R_A, R_B] = [0,492 ; 0,858]$

Soit finalement :

$$R_{2P} = 0,70^{+0,16}_{-0,21}$$

4.1.3. Approche bayésienne

La probabilité du paramètre vrai p , sachant qu'on a observé n succès parmi N observations est donnée par la relation :

$$P(p|n, N) = \frac{P(n|p, N) \cdot P(p)}{\int_0^1 P(n|p', N) \cdot P(p') \cdot dp'}$$

Il est nécessaire de faire une hypothèse sur la probabilité à priori de $P(p)$. Comme toutes les valeurs du rapport d'embranchement sont possibles, et sans préférence « à priori », on fait ici le choix d'une distribution uniforme de $P(p) \propto 1$.

La probabilité de l'observation est donnée par la fonction de vraisemblance :

$$P(n|p, N) \propto L(n|p, N) = p^n \cdot (1-p)^{N-n}$$

Comme $\int_0^1 L(n|p, N) = \frac{N+1}{\binom{N}{n}}$, où $\binom{N}{n} = \frac{N!}{n!(N-n)!}$, la probabilité du paramètre p est

alors (elle est représentée sur la figure 4-4) :

$$P(p|n, N) = \frac{1}{N+1} \cdot \binom{N}{n} \cdot p^n \cdot (1-p)^{N-n}$$

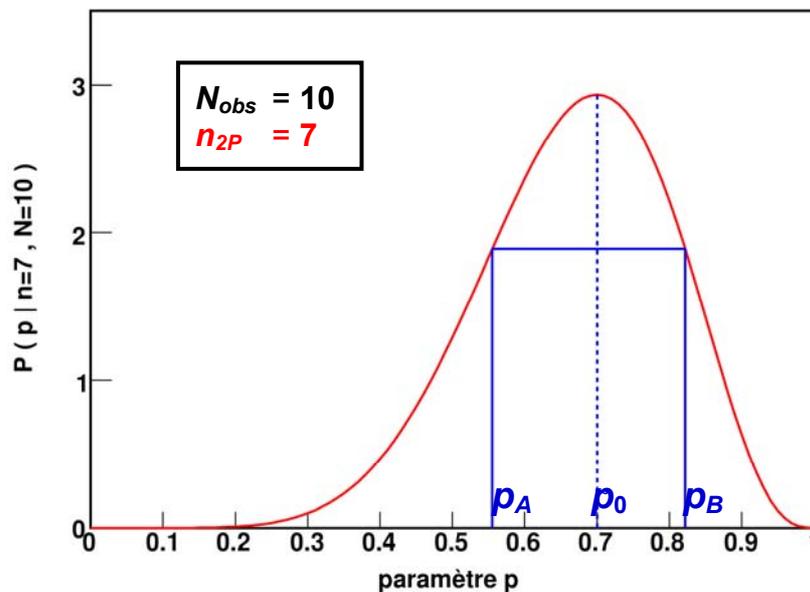


figure 4-4. Probabilité à posteriori du rapport d'embranchement dans le cadre de l'approche bayésienne, pour un nombre de radioactivités 2-protons $n_{2P} = 7$ sur un nombre d'observations de $N_{obs} = 10$. La ligne pointillée représente la valeur pour laquelle cette probabilité est maximum p_0 , et les lignes verticales pleines indiquent l'intervalle de confiance $[p_A, p_B]$ à 68,3 % (on a $P(p_A) = P(p_B)$).

L'intervalle de confiance (le plus petit) pour un degré de confiance β est obtenu en résolvant (numériquement) les équations :

$$\begin{cases} P(p_A|n, N) = P(p_B|n, N) \\ \int_{p_A}^{p_B} P(p'|n, N) \cdot dp' = \beta \end{cases}$$

Le résultat est alors, dans les mêmes conditions que précédemment : $R_{2P} = 0,70^{+0,12}_{-0,15}$.

On constate que l'intervalle de confiance obtenu est plus petit que celui de l'approche classique (du fait de l'apport d'information par la probabilité à priori). La figure 4-5 montre la comparaison des intervalles de confiance obtenus dans les 2 approches, pour différents nombres d'observations.

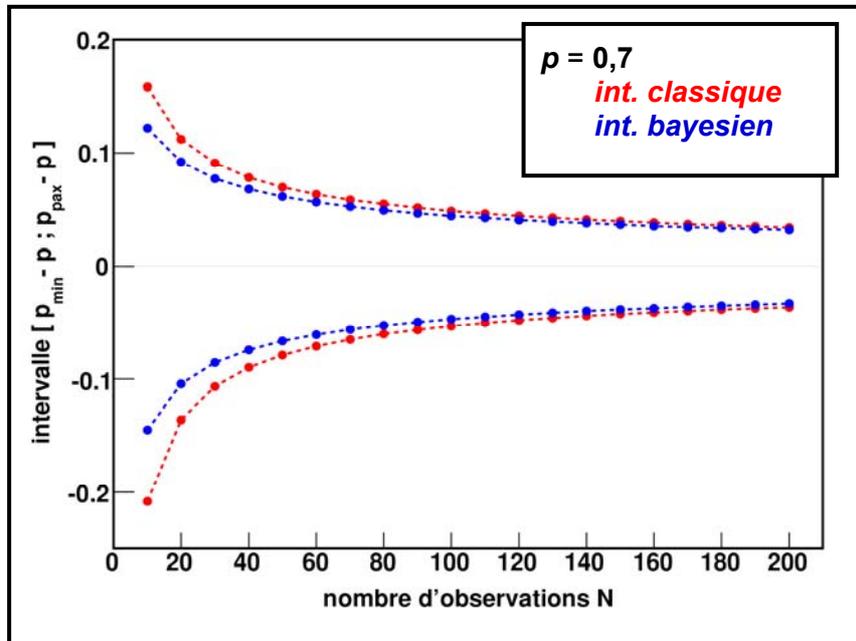


figure 4-5. Comparaison des intervalles de confiance (avec $\beta = 68,3 \%$) obtenus pour le rapport d'embranchement 2P (loi binomiale) dans les approches classique et bayésiennes, pour différents nombre d'observations (en abscisse), en gardant constant l'estimateur $n_{2P} / N_{obs} = 0,7$.

4.2. Application d'une loi continue : mesure de durée de vie

On cherche maintenant à déterminer la durée de vie du noyau produit, quelle que soit son mode de décroissance (on peut déduire la durée de vie pour chaque mode de décroissance à partir de la durée de vie du noyau et des rapports d'embranchement de chaque mode).

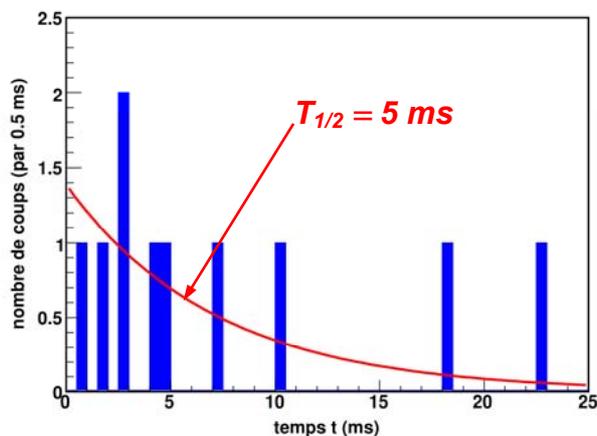


figure 4-6. Histogramme des instants de décroissance pour l'ensemble des noyaux produits. La courbe représente la loi de décroissance pour la valeur théorique de la durée de vie (5 ms).

Pour les $N_{obs} = 10$ noyaux produits, on observe les temps de décroissance suivant (en ms) pour chaque noyau (voir représentation figure 4-6) :

$$\begin{aligned} \vec{t} &= \{t_1, \dots, t_{N_{obs}}\} \\ &= \left\{ \begin{array}{ccccc} 0,905 & 4,550 & 10,278 & 4,456 & 18,310 \\ 22,741 & 2,856 & 1,590 & 2,940 & 7,331 \end{array} \right\} \end{aligned}$$

4.2.1. Loi de probabilité de la décroissance radioactive

La probabilité de décroissance par unité de temps est une constante : λ . Alors, si la décroissance n'a pas encore eu lieu à un instant t , on a :

- probabilité que la décroissance ait lieu entre t et $t + dt$: λt ;
- probabilité que la décroissance n'ait pas lieu entre t et $t + dt$: $1 - \lambda t$.

La probabilité que la décroissance ait lieu entre t et $t + dt$ est $f(t|\lambda) \cdot dt$. C'est le produit de la probabilité que la décroissance n'ait pas encore eu lieu $\bar{P}(t|\lambda)$ et de la probabilité pour l'intervalle de temps dt : soit $f(t|\lambda) \cdot dt = \bar{P}(t|\lambda) \cdot \lambda \cdot dt$.

La probabilité que la décroissance n'ait pas eu lieu à un instant $t + dt$ est de la même manière : $\bar{P}(t + dt|\lambda) = \bar{P}(t|\lambda) \cdot (1 - \lambda dt)$. Ce qui donne une équation différentielle dont le résultat est : $\bar{P}(t|\lambda) \propto e^{-\lambda t}$.

La loi de probabilité (f.d.p.) pour la décroissance radioactive est donc :

$$f(t|\lambda) = \lambda \cdot e^{-\lambda t}$$

La durée de vie du noyau, $T_{1/2}$, est donnée par le temps au bout duquel la probabilité que la décroissance ait eu lieu est de 50 %. Elle est reliée à la constante de décroissance par :

$$T_{1/2} = \frac{\ln 2}{\lambda}$$

4.2.2. Approche classique

On utilise ici encore le maximum de vraisemblance comme estimateur du paramètre λ . La fonction de vraisemblance s'écrit :

$$\begin{aligned} L(\vec{t}|\lambda) &= \prod_{i=1}^N P_i = \prod_{i=1}^N \lambda \cdot e^{-\lambda t_i} \\ &= \lambda^N \cdot \exp\left[-\lambda \cdot \sum_{i=1}^N t_i\right] = \lambda^N \cdot \exp[-\lambda t_\Sigma] \end{aligned} \quad \text{avec } t_\Sigma = \sum_{i=1}^N t_i$$

On constate que la fonction de vraisemblance ne dépend que de la somme des instants de décroissance, et non des valeurs individuelles. Son maximum est tel que :

$$\frac{dL(t_\Sigma|\lambda)}{d\lambda} = 0 \quad \Leftrightarrow \quad \frac{d(\ln L(t_\Sigma|\lambda))}{d\lambda} = 0$$

Et comme $\ln L(t_\Sigma|\lambda) = N \cdot \ln(\lambda) - \lambda t_\Sigma$, on a donc $\frac{d(\ln L(t_\Sigma|\lambda))}{d\lambda} = \frac{N}{\lambda} - t_\Sigma$. Et l'estimateur du

maximum de vraisemblance est alors : $\hat{\lambda} = \frac{N}{t_\Sigma}$.

Pour déterminer la ceinture de confiance pour cet estimateur, on détermine la fonction de probabilité de l'estimateur $f_N(\hat{\lambda}|\lambda_0)$, pour N observations et en supposant un paramètre « vrai » λ_0 . On calcule cette fonction à partir de la probabilité $f_N^{(t)}(t_\Sigma|\lambda_0)$ d'obtenir t_Σ . Pour N mesure, cette probabilité s'écrit :

$$f_N^{(t)}(t_\Sigma|\lambda_0) = \int_{t_1=0}^{t_\Sigma} f(t_1|\lambda_0) \times \int_{t_2=0}^{t_\Sigma-t_1} f(t_2|\lambda_0) \dots \times \int_{t_{n-1}=0}^{t_\Sigma-t_1-\dots-t_{n-2}} f(t_{n-1}|\lambda_0) \times f\left(t_N = t_\Sigma - \sum_{j=1}^{N-1} t_j \middle| \lambda_0\right) \cdot dt_1 \cdot dt_2 \dots \cdot dt_{N-1}$$

Un calcul par récurrence permet d'obtenir :

$$f_N^{(t)}(t_\Sigma|\lambda_0) = \lambda_0 \frac{(\lambda_0 \cdot t_\Sigma)^{N-1}}{(N-1)!} \cdot e^{-\lambda_0 \cdot t_\Sigma}$$

On retrouve la forme d'une loi de Poisson (cf. §2.4.2) pour une variable $x = \lambda_0 \cdot t_\Sigma$, qui donne la probabilité d'observer $N - 1$ événements (sur l'intervalle de temps t_Σ) alors qu'on en attend en moyenne x . Le facteur λ_0 vient du changement de variable de x à t ($f(x) \cdot dx = f(t) \cdot dt$).

Pour obtenir la f.d.p. de l'estimateur, on procède au changement de variable :

$$\lambda = \frac{N}{t_\Sigma} \text{ et } \frac{\partial t_\Sigma}{\partial \lambda} = -\frac{N}{\lambda^2} = -\frac{t_\Sigma^2}{N}$$

On a alors :

$$f_N(\lambda|\lambda_0) \cdot d\lambda = f_N^{(t)}(t_\Sigma|\lambda_0) \cdot dt_\Sigma$$

et finalement :

$$f_N(\lambda|\lambda_0) = \frac{N}{\lambda} \cdot \frac{\left(\frac{N \lambda_0}{\lambda}\right)^N}{N!} \cdot e^{-\left(\frac{N \lambda_0}{\lambda}\right)}$$

Si on note $x = \lambda_0 \cdot t_\Sigma = N \cdot \frac{\lambda_0}{\lambda}$, alors la f.d.p. pour l'estimateur s'écrit de façon simplifiée :

$$f_N(\lambda|\lambda_0) = \frac{N}{\lambda} \cdot \frac{x^N}{N!} \cdot e^{-x}$$

On peut alors calculer l'intervalle de confiance $[\lambda_A, \lambda_B]$ comme vu précédemment (cf. §3.5.1). Si λ_{exp} est la valeur de l'estimateur pour l'observation expérimentale, les bornes de l'intervalle de confiance sont données par (voir figure 4-7) :

$$P(\lambda \leq \lambda_{exp} | \lambda_B) = \int_0^{\lambda_{exp}} f_N(\lambda|\lambda_B) \cdot d\lambda = \alpha_1$$

et

$$P(\lambda \geq \lambda_{exp} | \lambda_A) = \int_{\lambda_{exp}}^{+\infty} f_N(\lambda|\lambda_A) \cdot d\lambda = \alpha_2$$

Pour l'échantillon mesuré (qui donne la valeur λ_{exp}), on calcule la valeur de l'estimateur et on résout numériquement les équations pour l'intervalle de confiance, ce qui donne :

- estimateur : $\hat{\lambda} = 0,132$ soit $T_{1/2} = 5,268$
- intervalle de confiance : $[\lambda_A, \lambda_B] = [0,091; 0,173]$ soit $T_A = 7,647$ et $T_B = 4,018$

Soit finalement, pour l'approche classique : $T_{1/2} = 5,3_{-1,3}^{+2,4} \text{ ms}$.

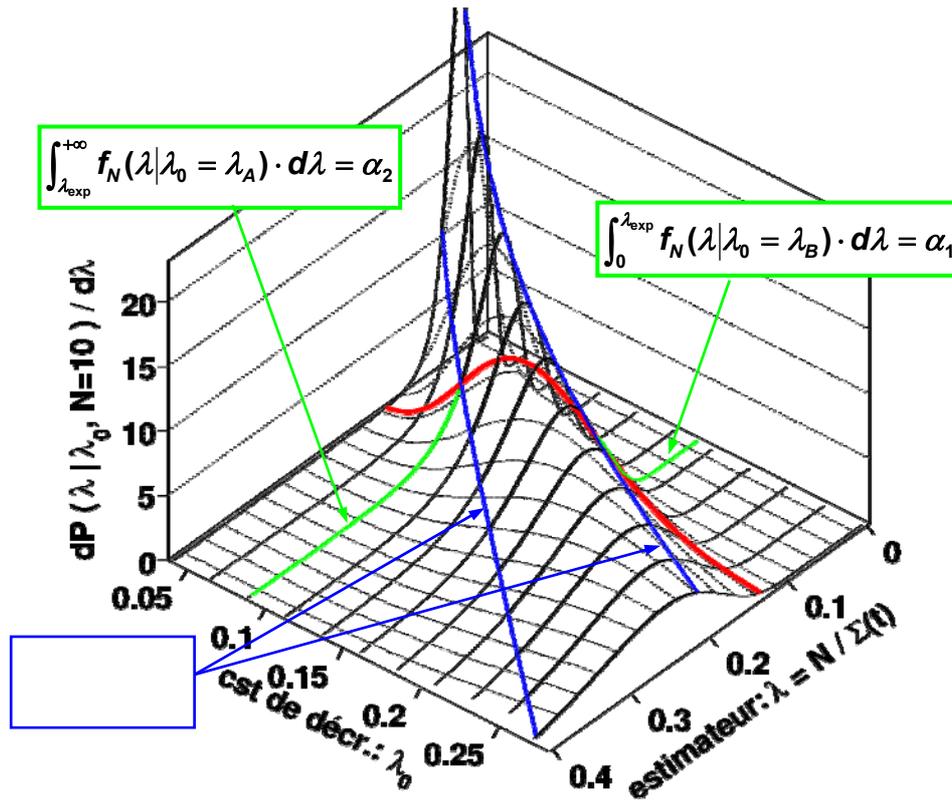


figure 4-7. Détermination de la ceinture de confiance pour la constante de décroissance radioactive. Pour une valeur observée de t_{Σ} , on détermine la valeur λ_B telle que $P(\lambda \geq \lambda_{exp} | \lambda_B) = \alpha_1$ et celle de λ_A telle que $P(\lambda \leq \lambda_{exp} | \lambda_A) = \alpha_2$.

4.2.3. Approche bayésienne

La f.d.p. du paramètre λ , dans l'approche bayésienne, est donnée par :

$$f(\lambda | t_{\Sigma}) = \frac{f(t_{\Sigma} | \lambda) \cdot P(\lambda)}{\int_0^{\infty} f(t_{\Sigma} | \lambda') \cdot P(\lambda') \cdot d\lambda'}$$

Si toutes les valeurs (positives) de λ sont possibles, et sans préférence à priori, alors on fait le choix (arbitraire) de $P(\lambda)$ uniforme. La f.d.p. du paramètre est alors :

$$f(\lambda | t_{\Sigma}) = t_{\Sigma} \cdot \frac{(\lambda \cdot t_{\Sigma})^{N-1}}{(N-1)!} \cdot e^{-\lambda \cdot t_{\Sigma}}$$

Cette fonction est représentée sur la figure 4-8.

Les bornes de l'intervalle de confiance ($\beta = 68,3\%$) peuvent être obtenues en résolvant le système d'équations (voir figure 4-8):

$$\begin{cases} f(\lambda_A | t_{\Sigma}) = f(\lambda_B | t_{\Sigma}) \\ \int_{\lambda_A}^{\lambda_B} f(\lambda' | t_{\Sigma}) \cdot d\lambda' = \beta \end{cases}$$

La résolution numérique donne alors le résultat $\lambda = 0,132_{-0,038}^{+0,047}$, soit $T_{1/2} = 5,3_{-1,4}^{+2,1} \text{ ms}$.

Comme vu au paragraphe 3.5.2, on peut faire un autre choix pour l'intervalle. Par exemple, faire le choix que $P(\lambda \leq \lambda_A) = P(\lambda \geq \lambda_B) = \alpha = \frac{1-\beta}{2}$. Dans ce cas, on obtient le résultat

suivant : $T_{1/2} = 5,3_{-1,6}^{+1,5} \text{ ms}$. On constate alors que l'intervalle précédent n'est pas le plus petit, contrairement à ce qui a été dit au paragraphe 3.5.2. En fait, c'est dû au changement de variable de λ à $T_{1/2}$. En effet, l'intervalle obtenu dans le premier cas est bien le plus petit pour le paramètre λ (voir figure 4-8).

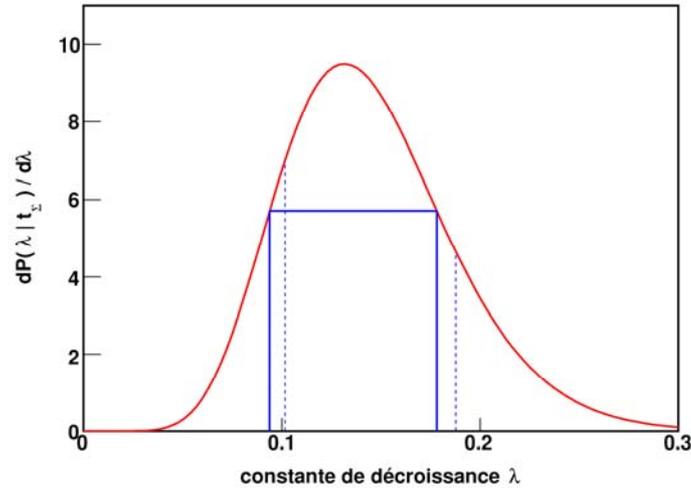


figure 4-8. Calcul bayésien de l'intervalle de confiance pour la constante de décroissance λ . L'intervalle en trait plein correspond au plus petit intervalle $[\lambda_A, \lambda_B]$ tel que $f(\lambda_A) = f(\lambda_B)$, alors que l'intervalle en pointillés correspond au choix « symétrique », tel que $P(\lambda \leq \lambda_A) = P(\lambda \geq \lambda_B) = \alpha$.

La remarque précédente suggère que l'on peut faire le choix d'un autre paramètre. Si on définit le temps caractéristique τ tel que :

$$\tau = \frac{1}{\lambda} = \frac{T_{1/2}}{\ln 2}$$

alors la f.d.p. de l'instant de décroissance devient : $f(t|\tau) = \frac{1}{\tau} \cdot e^{-(t/\tau)}$

et la fonction de vraisemblance : $L(\vec{t}|\tau) = \frac{1}{\tau^N} \cdot \exp\left[-\frac{t_\Sigma}{\tau}\right]$

De la même manière que précédemment (avec le paramètre λ), on détermine la f.d.p. de l'estimateur du paramètre τ dans l'approche bayésienne. Il faut donc faire une hypothèse sur la probabilité à priori de τ . On peut bien entendu faire le choix d'une distribution uniforme. Il est à noter que ce choix est différent du cas précédent, puisqu'une distribution uniforme de τ implique une distribution qui n'est pas uniforme pour $\lambda = 1/\tau$. Les deux hypothèses (probabilité uniforme de τ ou de λ) sont donc différentes !

Avec le choix du paramètre τ , la f.d.p. est alors, à partir du théorème de Bayes (en supposant $P(\tau)$ uniforme) :

$$f(\tau|t_\Sigma) = t_\Sigma \cdot \frac{\left(\frac{t_\Sigma}{\tau}\right)^{N-2}}{(N-2)!} \cdot e^{-(t_\Sigma/\tau)}$$

Cette probabilité est représentée sur la figure 2-1.

Pour un intervalle à 68,3 % de confiance, on obtient $\tau = 7,601_{-2,158}^{+3,466}$, soit $T_{1/2} = 5,3_{-1,5}^{+2,4}$ ms, en choisissant les bornes de l'intervalle telles que $f(\tau_A|t_\Sigma) = f(\tau_B|t_\Sigma)$.

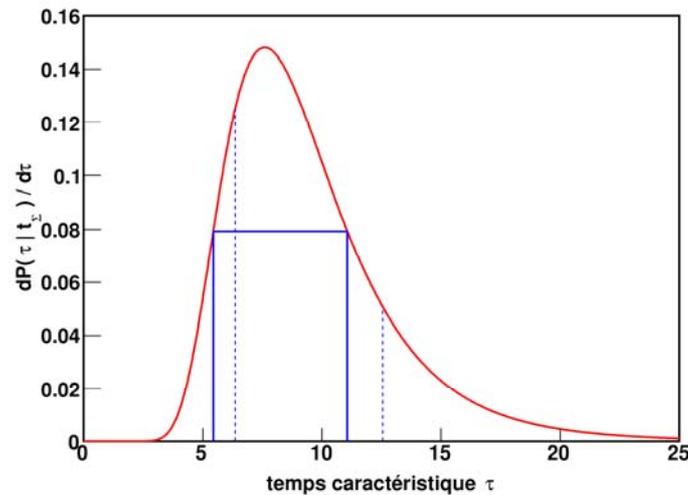


figure 4-9. Calcul bayésien de l'intervalle de confiance pour le temps caractéristique τ . L'intervalle en trait plein correspond au plus petit intervalle $[\tau_A, \tau_B]$ tel que $f(\tau_A) = f(\tau_B)$, alors que l'intervalle en pointillés correspond au choix « symétrique », tel que $P(\tau \leq \tau_A) = P(\tau \geq \tau_B) = \alpha$.

4.2.4. Remarques sur les résultats classiques et bayésiens

Cet exemple d'estimation de durée de vie illustre les différentes approches. Dans le cas classique, quel que soit le paramètre considéré (constante de décroissance λ ou temps caractéristique τ), le même résultat est obtenu (aucune hypothèse n'est faite a priori). Dans le cas bayésien, on constate que le résultat n'est pas exactement le même selon le paramètre choisi.

Dans les cas présentés, qui correspondent à 10 mesures, les résultats ne diffèrent pas beaucoup, mais plus la statistique est faible, plus ces différences peuvent être importantes. Le tableau ci-dessous montre les résultats des différentes approches pour différents nombres d'observations.

Nombre d'observations	Durée de vie selon le type d'analyse		
	classique	bayésienne (plus petit intervalle pour le paramètre)	
		paramètre λ	paramètre τ
N = 5	$T_{1/2} = 5,34_{-1,61}^{+4,06}$ ms	$T_{1/2} = 5,34_{-1,84}^{+3,39}$ ms	$T_{1/2} = 5,34_{-2,14}^{+4,48}$ ms
N = 10	$T_{1/2} = 5,27_{-1,25}^{+2,38}$ ms	$T_{1/2} = 5,27_{-1,38}^{+2,11}$ ms	$T_{1/2} = 5,27_{-1,50}^{+2,40}$ ms
N = 20	$T_{1/2} = 4,92_{-0,89}^{+1,40}$ ms	$T_{1/2} = 4,92_{-0,96}^{+1,29}$ ms	$T_{1/2} = 4,92_{-1,00}^{+1,37}$ ms
N = 50	$T_{1/2} = 4,70_{-0,58}^{+0,77}$ ms	$T_{1/2} = 4,70_{-0,61}^{+0,73}$ ms	$T_{1/2} = 4,70_{-0,62}^{+0,75}$ ms

Tableau des durées de vies obtenues avec des données simulées avec une valeur effective $T_{1/2} = 5$ ms. Pour les différents nombres N d'observations, les données sont les mêmes, en se limitant aux N premiers résultats de la simulation.

4.2.5. Mesure de durée de vie avec bruit de fond

L'exemple précédent (mesure de durée de vie) est très simplifié par rapport aux situations généralement rencontrées dans la réalité. Parmi les facteurs qui sont susceptibles de rendre l'analyse plus délicate, on peut noter (par exemple) :

- que pour chaque noyau produit, la décroissance est mesurée pendant un temps fini ;
- qu'il y a généralement du bruit de fond, c'est à dire des événements indésirables qu'on ne sait pas distinguer du signal de décroissance ;
- qu'il peut y avoir du temps mort, c'est à dire qu'on peut manquer des événements ;
- ...

On se propose ici de reprendre le cas précédent d'une mesure de durée de vie, en présence de bruit de fond. Pour chacun des N noyaux produits, on mesure la décroissance pendant un intervalle de temps fixe T .

Pour une « mesure » (observations des signaux de décroissance pour 1 noyau, pendant un temps T), on peut avoir :

- la décroissance du noyau (**0** ou **1** événement) ;
- du bruit de fond : événements aléatoires de fréquence ν_b supposée constante.

L'information contenue dans l'échantillon (N mesures avec un nombre variable d'événements) comporte à la fois des aspects discrets (nombre d'événements) et des aspects continus (instant des événements).

Pour simplifier le problème, on suppose la fréquence du bruit de fond connue (si par exemple elle peut être déterminée par des mesures complémentaires indépendantes) : $\nu_b = 0,01 \text{ ms}^{-1}$. Il s'agit alors d'estimer un seul paramètre : la durée de vie.

Analyse par événements

On peut envisager une analyse événement par événement, en déterminant pour chaque mesure i (c'est à dire pour chaque noyau produit), la probabilité d'observer n_i événements (décroissance vraie et/ou bruit de fond) à des temps $t_1^{(i)}, \dots, t_{n_i}^{(i)}$.

On peut écrire la fonction de vraisemblance sous la forme :

$$L(\text{obs}|\lambda) = \prod_{i=1}^N P(n_i) \times f_{n_i}(t_1^{(i)}, \dots, t_{n_i}^{(i)})$$

où $P(n_i)$ est la probabilité d'observer n_i événements pour la mesure i et $f_{n_i}(t_1^{(i)}, \dots, t_{n_i}^{(i)})$ est la probabilité (conditionnelle) lorsqu'on a observé n_i événements que ces événements aient lieu aux temps $t_1^{(i)}, \dots, t_{n_i}^{(i)}$. On n'entre pas ici dans les détails du calcul (qui sont dans la présentation orale), mais la fonction de vraisemblance prend la forme :

$$L(\text{obs}|\lambda) = \prod_{i=1}^N \left\{ e^{-\lambda T} \cdot B(n_i) \cdot \frac{1}{T^{n_i}} + B(n_i - 1) \cdot \frac{\lambda}{T} \cdot \left[\sum_{k=1}^{n_i} \exp(-\lambda t_k^{(i)}) \right] \right\}$$

où $B(n) = \frac{(\nu_b \cdot T)^n}{n!} \cdot e^{-\nu_b \cdot T}$ est la probabilité pour une mesure d'observer n événements de bruit de fond (il s'agit d'une loi de poisson donnant la probabilité d'observer n événements lorsqu'on en attend en moyenne $\nu_b \cdot T$).

L'approche classique ne permet pas ici de traiter le problème : il faudrait envisager l'ensemble des possibilités conduisant au même estimateur : il y en a une infinité. Seule l'approche bayésienne est possible.

En utilisant l'estimateur du maximum de vraisemblance (avec par exemple l'hypothèse que $P(\lambda)$ est uniforme), on peut calculer un intervalle de confiance pour le paramètre.

Néanmoins, si le calcul semble relativement simple, il n'est pas exact : dans notre exemple, le maximum de vraisemblance ne satisfait pas les conditions (commutation des opérateurs de dérivation et d'intégration, voir §3.3.4) qui garantissent que l'estimateur est consistant.

Analyse par histogramme

La solution la plus simple dans ce type de problème est de créer un histogramme des mesures (voir figure 4-10). Cela implique une discrétisation du problème, et une perte d'information liée au « binning » (discrétisation des instants d'observation des événements). Le problème consiste alors à analyser la forme de l'histogramme en fonction du nombre de coups attendus dans chaque intervalle de temps Δt (bin).

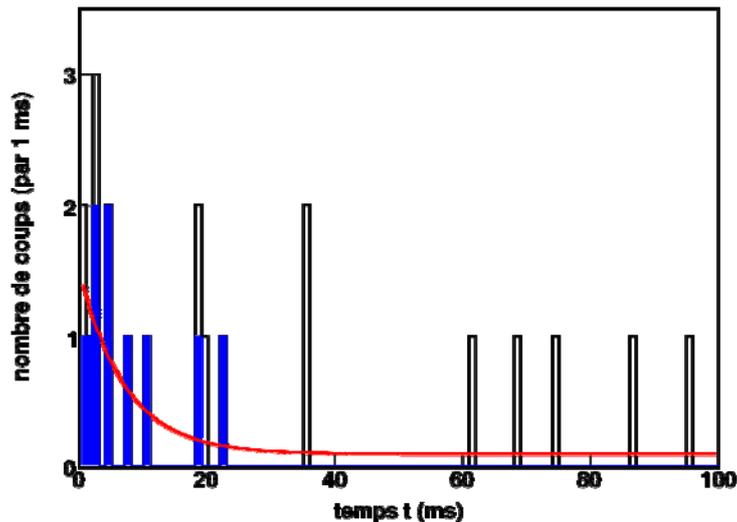


figure 4-10. Histogramme des mesures (issues d'une simulation) obtenu avec un échantillonnage toute les millisecondes ($\Delta t = 1$ ms) de la durée de chaque mesure ($T = 100$ ms) : l'histogramme non coloré correspond à l'ensemble des événements (décroissance et bruit de fond), l'histogramme coloré correspond aux seuls événements de décroissance (identiques aux événements analysé dans la mesure de durée de vie sans bruit de fond), et la courbe correspond à la distribution moyenne pour les paramètres d'entrée de la simulation : $T_{1/2} = 5$ ms et $\nu_b = 0,01$ ms^{-1} .

Pour une mesure (1 noyau produit), le nombre d'événements par unité de temps est donné par :

- pour la décroissance : $\frac{dn_{dec}(t)}{dt} = \lambda \cdot e^{-\lambda t}$
- pour le fond (constant) : $\frac{dn_{fond}(t)}{dt} = \nu_b$

Par conséquent, pour l'ensemble de l'histogramme, comprenant les décroissances et le fond de chacune des N mesures, le taux d'événements est donné par :

$$\frac{dn(t)}{dt} = N \cdot [\lambda \cdot e^{-\lambda t} + \nu_b]$$

Le nombre moyen μ_i de coups attendus dans un intervalle i compris entre t_i et $t_i + \Delta t$ (avec $t_i = i \times \Delta t$) est alors :

$$\begin{aligned} \mu_i &= \int_{t_i}^{t_i+\Delta t} \frac{dn(t')}{dt} dt' \approx \frac{dn(t_i)}{dt} \cdot \Delta t \\ &\approx N \cdot \Delta t \cdot [\lambda \cdot e^{-\lambda \cdot i \cdot \Delta t} + \nu_b] \end{aligned}$$

La probabilité d'observer n_j événements dans l'intervalle j lorsqu'on en attend μ_j est donnée par la loi de Poisson (voir §2.4.2), et la fonction de vraisemblance est alors :

$$L(\text{obs}|\lambda) = \prod_j \frac{(\mu_j(\lambda))^{n_j}}{n_j!} \cdot e^{-\mu_j(\lambda)}$$

Ici encore, l'approche classique n'est pas possible, car pour calculer la probabilité d'un estimateur $\hat{\lambda}$ étant donné un paramètre vrai λ_0 il faut envisager tous les cas possibles. On peut cependant faire le calcul dans le cas bayésien, en utilisant le maximum de vraisemblance comme estimateur de λ .

En supposant une probabilité à priori uniforme de λ , on obtient le résultat suivant (voir figure 4-11) : $T_{1/2} = 4,4^{+3,4}_{-2,3}$

L'analyse en choisissant le temps caractéristique τ comme paramètre donne : $T_{1/2} = 4,4^{+3,1}_{-2,2}$

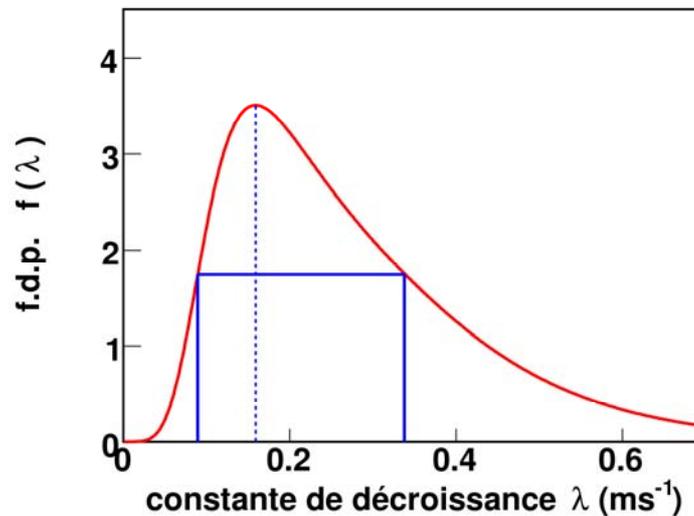


figure 4-11. Intervalle de confiance bayésien dans la détermination de la durée de vie avec bruit de fond, en utilisant la constante de décroissance λ comme paramètre.

4.3. Conclusion sur les exemples à faible statistique

Les exemples présentés sont au départ très simples. Dans le cas de la mesure de durée de vie, on a pu constater qu'en ajoutant seulement le bruit de fond, le problème devient plus complexe. De plus, en dehors de « cas d'école », l'approche classique de l'analyse statistique est souvent impossible, et le traitement bayésien du problème devient incontournable. Enfin, en règle générale (et c'est déjà le cas pour les cas très simples présentés ici), il n'y a pas de solution analytique au problème, et les intervalles de confiance doivent être déterminés par des méthodes numériques de résolution des équations.

En ce qui concerne l'illustration du traitement des faibles statistiques dans le cas de la radioactivité 2-protons, on est encore loin d'une situation réaliste. Pour la détermination du rapport d'embranchement, il faudrait introduire la probabilité d'erreur dans la détermination du type d'événement (émission 2-protons ou radioactivité β). Pour la durée de vie, outre le bruit de fond, le temps mort devrait également être pris en compte.

Comme un problème simple peut très vite devenir complexe, il convient de bien étudier chaque problème dans le détail, pour déterminer les aspects réellement significatifs. Il n'est guère utile de rendre un problème inextricable pour des détails qui en fin de compte ont un effet négligeable.