



Analyse Statistique

cas des petits nombres

Jérôme Giovinazzo - CENBG

giovinaz@cenbg.in2p3.fr

Ecole Joliot-Curie – Seignosse – 2008

Analyse statistique – cas des petits nombres –

I. Introduction

II. Calcul des probabilités

concepts de la théorie des probabilités

probabilités conditionnelles

variables aléatoires

loi des grands nombres

III. Analyse statistiques

estimateurs

comportement asymptotique

intervalle de confiance

IV. Applications avec des petits nombres

cas d'une distribution discrète

cas d'une distribution continue

un peu abstrait

un peu plus concret

partie 1

Introduction

Remarques préliminaires

ce cours n'est pas donné par un statisticien...

statistique et **probabilités**: domaines des **mathématiques**

⇒ souvent maltraité par la physique !!!

des domaines avec des objectifs différents:

- **mathématiques** : rigueur formelle parfaite
- **physique** : le *nombre* n'est que l'interprétation d'une *réalité* (le monde...) :
on en attend « seulement » une signification *suffisante*
il y a une part de subjectivité intrinsèque de l'interprétation
(ce qui peut paraître paradoxal pour une science dite exacte...)

autres **cultures**, autres **langages**

physiciens et statisticiens n'ont pas le même jargon

La physique et la statistique

Probabilités : domaine des mathématiques pures

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

Statistique : domaine des mathématiques appliquées

L'outil de base de l'analyse statistique est le calcul des probabilités (mise en œuvre des lois de probabilité)

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

⇒ c'est l'objet de son utilisation en physique

(interprétation de la mesure par l'analyse statistique)

Qu'est-ce que la physique ?

description de la réalité à l'aide de théories (ou de modèles)

confrontation avec l'expérience:

ensemble de mesures pour valider ou d'invalider ces théories

⇒ nécessité de déterminer si le résultat d'une mesure est significatif (degré de confiance que l'on peut lui accorder)

Objectifs du cours (1)

subjectivité en amont: modèle **physique**

subjectivité en aval: interprétation de la mesure
 ⇒ **statistique**

en dehors des « cas d'école », cette subjectivité est donc généralement inévitable, au delà de celle inhérente aux statistiques

Nécessité d'une analyse au *cas par cas*, à partir d'outils de base

- Chaque problème est spécifique
- Utiliser les méthodes les mieux adaptées

Objectifs:

- rappeler quelques concepts fondamentaux de l'analyse statistique
- illustrations à partir de cas simples

Objectifs du cours (2)

La statistique, un vaste monde...

- calcul des **probabilités**
- théorie de l'**information**, théorie de la **décision**
- variables **aléatoires**, **estimations**, **régressions**
- **erreurs** et **incertitudes**

Calcul des probabilités

- bases
- probabilités conditionnelles
- distributions courantes

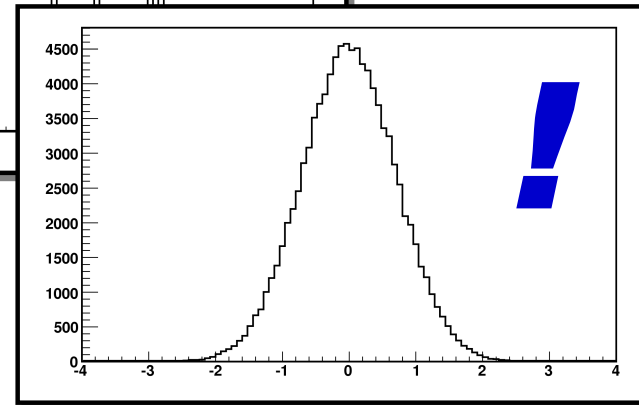
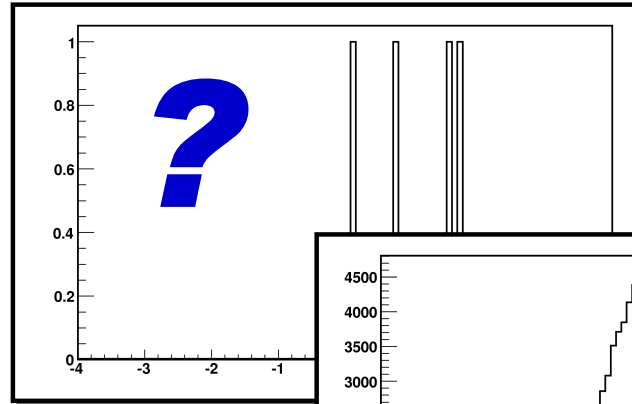
Analyse statistique

- construction d'un estimateur
- intervalle de confiance

Objectifs du cours (3)

les faibles statistiques
des exemples concrets

les grands nombres
tous devient gaussien...



Ce qui ne sera pas traité ici

⇒ **Les tests d'hypothèses**
théorie de la décision

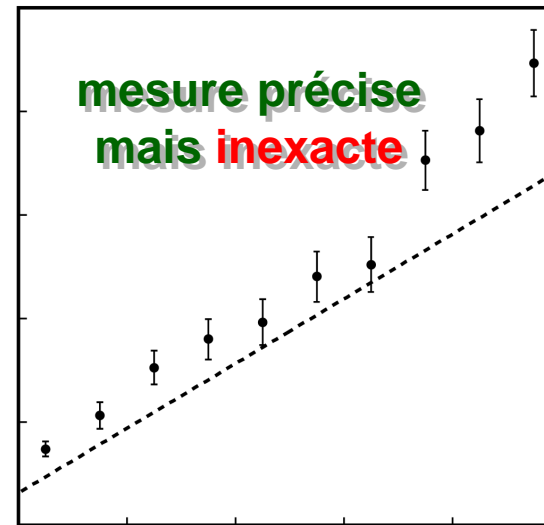
⇒ **Les régressions**
type minimisation

outils d'analyses familiers et robustes: PAW / ROOT (Minuit)

⇒ ...

La mesure et l'erreur

Précision ou exactitude ?



Erreur statistique, erreur systématique

statistiques *incertitude* liée au caractère aléatoire des grandeurs mesurées
un plus grand nombre de mesure réduit cette incertitude

systématiques il n'y a pas de théorie pour les erreurs systématiques !
elles proviennent principalement des techniques expérimentales,
susceptibles d'introduire des biais

**chaque cas est particulier, et les erreurs systématiques ne sont pas
abordées ici...**

partie 2

Calcul des probabilités

Quelques concepts de base de
la théorie des probabilités

Probabilités conditionnelles

Variables aléatoires

Loi des grands nombres

Probabilité: définition et propriétés (1)

définition en « fréquence »

- expérience (mesure) dont le résultat peut être (ou non) d'un type donné X
- on répète N fois l'expérience, et on obtient n fois le résultat X

probabilité $P(X)$ qu'un événement soit de type X :

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

événements élémentaires

soit un ensemble des résultats possibles X_i

- exhaustif : le résultat est forcément un des X_i
- exclusif : le résultat ne peut être simultanément de type X_i et X_j si $i \neq j$

alors

$$\begin{cases} P(X_i) > 0 \\ P(X_i \text{ ou } X_j) = P(X_i) + P(X_j) \\ \sum_i P(X_i) = 1 \end{cases}$$

Probabilité: définition et propriétés (2)

loi d'addition

si A et B sont 2 sous-ensembles des possibilités X_i , la probabilité qu'une observation se produise dans A ou B est :

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

probabilité conditionnelle

(loi de multiplication)

$P(A|B)$ probabilité qu'une observation se produise dans A , alors qu'elle appartient à B

la probabilité qu'une observation se produise dans A et B simultanément est alors donnée par la loi de multiplication:

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

indépendance

si la réalisation de A est indépendante de celle de B , alors $P(A|B) = P(A)$ et

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

(c'est une condition nécessaire et suffisante)

Probabilité conditionnelle : illustration géométrique

on considère un ensemble de points (x,y) équiprobables

on définit deux sous-ensembles A et B :

$P(A)$ est la proba. qu'un point soit dans A , et $P(B)$ la proba. qu'il soit dans B

$$A: \{y_{min}^A \leq y \leq y_{max}^A\}$$

$$A: \{y_{min}^A \leq y \leq y_{max}^A\}$$

$$B: \{x_{min}^B \leq x \leq x_{max}^B\}$$

$$B: \{x_{min}^B \leq x \leq x_{max}^B \text{ et } y_{min}^B \leq y \leq y_{max}^B\}$$

surfaces

$$S_A = 1 \times (y_{max}^A - y_{min}^A)$$

$$S_A = 1 \times (y_{max}^A - y_{min}^A)$$

$$S_B = (x_{max}^B - x_{min}^B) \times 1$$

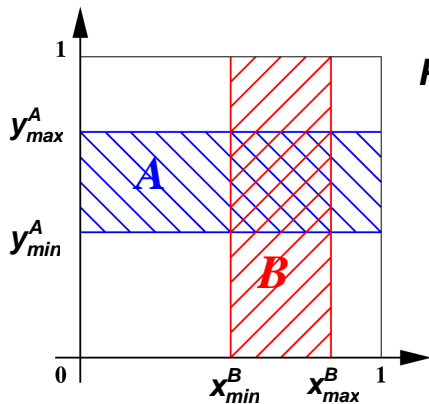
$$S_B = (x_{max}^B - x_{min}^B) \times (y_{max}^B - y_{min}^B)$$

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

$$S_{AB} = (x_{max}^B - x_{min}^B) \times (y_{max}^A - y_{min}^A)$$

probabilités

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



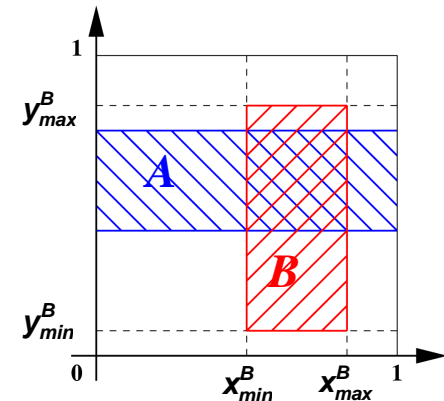
$$P(A \cap B) = S_{AB} = S_A \cdot S_B$$

$$P(A \cap B) = S_{AB}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{S_{AB}}{S_B}$$

$$P(A|B) = S_A \\ = P(A)$$

$$P(A|B) = \frac{y_{min}^A - y_{max}^A}{y_{min}^B - y_{max}^B} \\ \neq P(A)$$



cas A et B indépendants

cas A et B non indépendants

Approches classiques et modernes (1)

théorème de Bayes

d'après la loi de multiplication pour des événements discrets: $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

on peut généraliser cette relation, en considérant n sous-ensembles A_i ($i = 1, \dots, n$) d'événements, les A_i étant supposés exhaustifs et exclusifs, et un sous-ensemble quelconque B :

$$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_i P(B|A_i) \cdot P(A_i)}$$

classiques et modernes

lorsqu'à partir d'une observation (événement de type B), on veut tester une hypothèse (l'événement est-il de type A_i ?), il se pose alors un problème d'interprétation, sur lequel deux écoles s'opposent :

les « classiques » (ou *anti-bayesiens*) et les « modernes » (ou *bayesiens*)

au cœur de la question statistique : l'interprétation !

Approches classiques et modernes (2)

hypothèse des « modernes » (*bayesiens*)

on suppose une observation (mesure), donnant un résultat \mathbf{x} , que l'on cherche à interpréter dans le cadre d'une théorie :

→ l'ensemble des hypothèses θ_i possibles (ex. θ_i peut représenter les valeurs, discrètes ou continues, des paramètres de la théorie)

les « modernes » généralisent le **théorème de Bayes** au cas des hypothèses:

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

$P(\theta_i | \mathbf{x})$ probabilité que l'hypothèse soit vraie alors qu'on a observé \mathbf{x}
 ⇒ **interprétation statistique du résultat**

$P(\mathbf{x} | \theta_i)$ probabilité d'observer \mathbf{x} dans le cadre de l'hypothèse θ_i
 ⇒ **calcul de probabilité, donné par la théorie**

$P(\theta_i)$ probabilité d'une hypothèse donnée
 ⇒ **degré de confiance à priori de l'hypothèse**

$P(\mathbf{x})$ probabilité d'observer \mathbf{x} quelle que soit l'hypothèse considérée

Approches classiques et modernes (3)

opposition des *modernes* et des *classiques*

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

si l'ensemble des θ_i est exhaustif et exclusif (c'est en général le cas en physique : un paramètre de la théorie a *une valeur et une seule*) alors :

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

la controverse vient du terme $P(\theta_i)$: pour les « classiques », cette probabilité à priori ne peut être déterminée

⇒ l'interprétation du résultat X (information à posteriori que l'on en tire) n'est pas la même les 2 approches !

→ *illustration pour faibles statistiques dans la partie 3*

ce n'est pas seulement une question de valeur numérique, mais aussi une question de signification du résultat

remarques

- il est préférable d'éviter l'approche bayésienne, à cause de la subjectivité sur $P(\theta_i)$
- mais l'approche classique **ne permet pas toujours de conclure !**
- pour un grand nombre d'observations, **les 2 approches convergent** vers un même résultat

Variables aléatoires

événement aléatoire

⇒ plusieurs résultats possibles

une variable (grandeur physique), associée à un tel événement, peut prendre un ensemble de valeurs, discret ou continu

variables aléatoires discrète

valeurs x_i , avec $i = 1, \dots, n$

$$0 \leq P(x_i) \leq 1$$

si les x_i sont exhaustifs :

$$\sum_{i=1}^n P(x_i) = 1$$

variable aléatoire continue

valeur x dans domaine Ω

$f(x)dx$: probabilité d'une valeur dans $[x; x+dx[$

si le domaine Ω est exhaustif : $\int_{\Omega} f(x')dx' = 1$

$f(x)$: densité de probabilité, ou probabilité différentielle

$$f(x) = \frac{dP(x)}{dx}$$

Caractéristiques des V.A.

variable	x	discrète	continue
espérance	$E[x]$	$\mu = \sum_k x_k \cdot P(x_k)$	$\mu = \int_{\Omega_x} x \cdot f(x) \cdot dx$
variance	$V[x]$	$\sigma^2 = \sum_k (x_k - \mu)^2 \cdot P(x_k)$	$\sigma^2 = \int_{\Omega_x} (x - \mu)^2 \cdot f(x) \cdot dx$
moments	$\mu'_j(x)$	$\mu'_j = \sum_k (x_k)^j \cdot P(x_k)$	$\mu'_j = \int_{\Omega_x} x^j \cdot f(x) \cdot dx$
moments centraux	$\mu'_j(x)$	$\mu_j = \sum_k (x_k - \mu)^j \cdot P(x_k)$	$\mu_j = \int_{\Omega_x} (x - \mu)^j \cdot f(x) \cdot dx$

Fonctions de distribution

densité de probabilité pour une V.A.

elle est normée $\int_{\Omega} f(x) \cdot dx = 1$

toute fonction d'une V.A. est également une V.A.

espérance (*valeur attendue*) d'une fonction $g(x)$, où x est une V.A. de fdp $f(x)$:

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

$E[\dots]$ est un opérateur linéaire:

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

Fonction caractéristique

pour information
seulement

définition

x une var. aléatoire

$$\Phi_x(t) = E[e^{itx}]$$

variable continue : $\Phi_x(t) = \int_{-\infty}^{+\infty} e^{itx} \cdot f(x) \cdot dx$

variable discrète : $\Phi_x(t) = \sum_k p_k \cdot e^{itx_k}$

propriétés

la fonction caractéristique définit entièrement la fonction de distribution de probabilité

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

si α et β sont des constantes

$$\Phi_{\alpha x + \beta}(t) = e^{it\beta} \cdot \Phi_x(\alpha t)$$

si x et y sont 2 V.A. **indépendantes** de fonctions caractéristiques $\Phi_x(t)$ et $\Phi_y(t)$,
la fonction caractéristique de la V.A. $(x+y)$ est

$$\Phi_{x+y}(t) = \Phi_x(t) \cdot \Phi_y(t)$$

calcul des moments

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

$$\mu_r = \frac{1}{i^r} \cdot \left[\left(\frac{d}{dt} \right)^r [e^{-i\mu t} \cdot \Phi(t)] \right]_{t=0}$$

Plusieurs variables aléatoires

densité de probabilité conjointe de 2 (ou plus) V.A. : $f(x,y)$

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

alors, pour une fonction $g(x,y)$:

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

distributions marginales

$$f_X(x) = \int_{y_{\min}}^{y_{\max}} f(x,y) \cdot dy$$

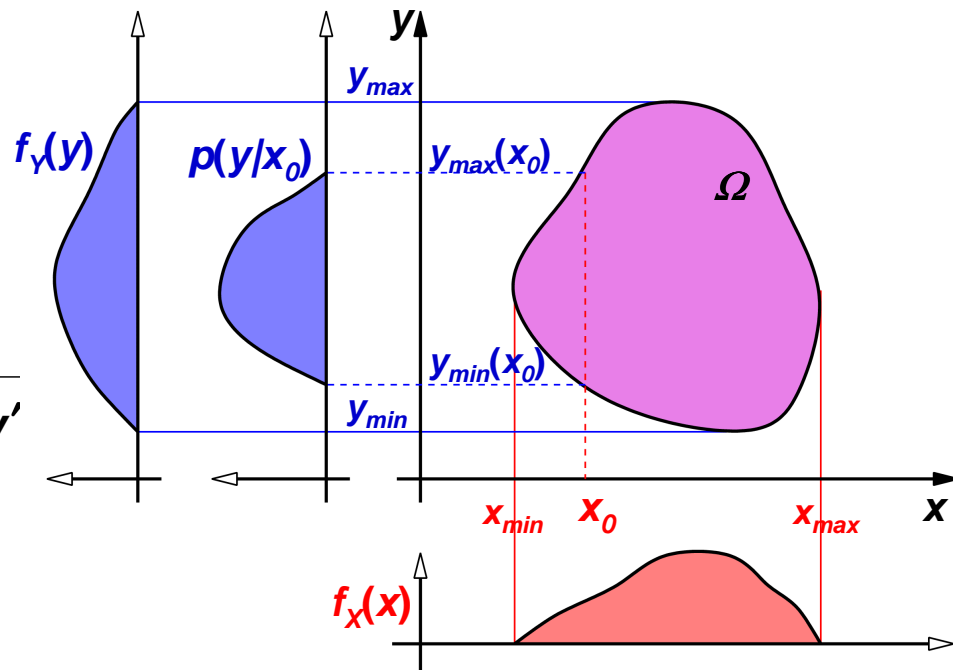
$$f_Y(y) = \int_{x_{\min}}^{x_{\max}} f(x,y) \cdot dx$$

distributions conditionnelles

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

théorème de Bayes

$$p(y|x) = \frac{p(x|y) \cdot f_Y(y)}{f_X(x)}$$



Covariance et corrélation

définitions

moyennes et **variances**, pour chaque variable :

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

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

covariance

$$\begin{aligned} \text{cov}[x, y] &= E[(x - \mu_x) \cdot (y - \mu_y)] \\ &= E[xy] - E[x] \cdot E[y] \end{aligned}$$

corrélation

$$\text{corr}[x, y] = \rho(x, y) = \frac{\text{cov}[x, y]}{\sigma_x \cdot \sigma_y}$$

on montre facilement que : $-1 \leq \rho \leq +1$

Variables aléatoires indépendantes

indépendance

les V.A. x et y sont indépendantes **si et seulement si** (condition nécessaire et suffisante)

$$f(x, y) = f_x(x) \cdot f_y(y) \quad (\text{produit des f.d.p. marginales})$$

$$\begin{aligned} \text{alors } E[xy] &= \iint_{\Omega} x \cdot y \cdot f_x(x) \cdot f_y(y) \cdot dx \cdot dy = \int x \cdot f_x(x) \cdot dx \cdot \int y \cdot f_y(y) \cdot dy \\ &= E[x] \cdot E[y] \end{aligned}$$

$$\Rightarrow \text{cov}[x, y] = 0 \text{ et } \rho(x, y) = 0$$

indépendance \Rightarrow **non corrélation** (la réciproque n'est pas nécessairement vraie)

matrice de covariance

les V.A. x_1, \dots, x_n sont indépendantes si et seulement si la f.d.p. conjointe est factorisable sous la forme:

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

les covariances et coefficients de corrélations pour (x_i, x_j) : $\sigma_{ij} = \text{cov}(x_i, x_j)$

matrice de covariance : $\bar{V} = [\sigma_{ij}]$

si la matrice de covariance n'est pas inversible, il existe au moins une relation linéaire entre les x_i

Loi des grands nombres

il existe plusieurs « **loi de convergence** »
(on n'entre pas dans les détails ici)

- 2 applications des théorèmes de convergence:
 → **loi des grands nombres**
 → **théorème de la limite centrale (TCL)**

résultats
importants
non démontrés
dans ce cours

loi des grands nombres

$\{x_1, \dots, x_N\}$ un ensemble de V.A. indépendantes de même moyenne μ , et de variances σ_i^2

moyenne de l'échantillon : $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$

(statistiques)

loi des grands nombres : $\bar{x} \xrightarrow{N \rightarrow \infty} \mu$

→ en *moyenne quadratique* si $\lim_{N \rightarrow \infty} \left(\frac{1}{N^2} \sum_{i=1}^N \sigma_i^2 \right) = 0$ (loi faible)

→ *presque certainement* si $\lim_{N \rightarrow \infty} \left(\sum_{i=1}^N \left(\frac{\sigma_i}{i} \right)^2 \right)$ est finie (loi forte)

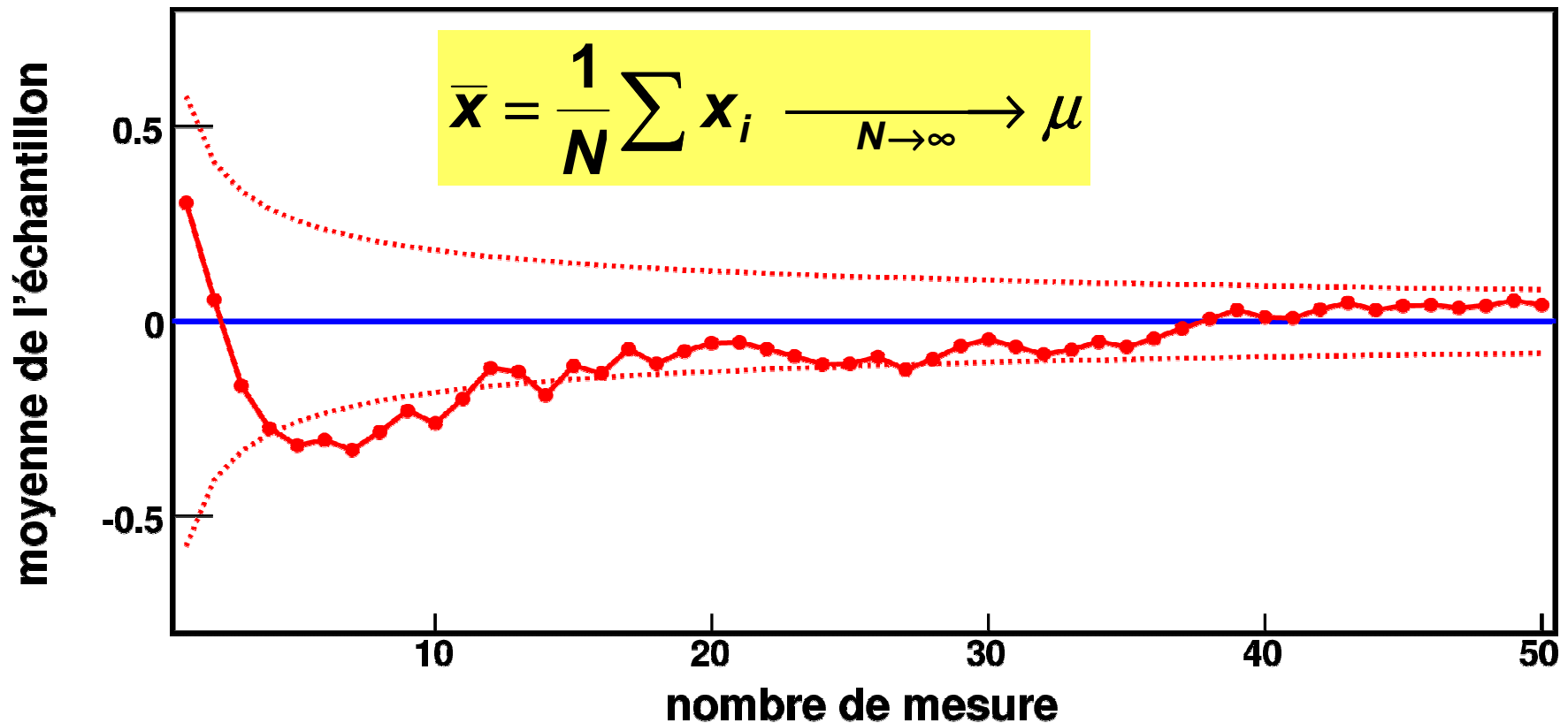
à noter : c'est le cas si il les variances σ_i^2 sont **bornées**
(et donc également si elles sont **égales**)

Loi des grands nombres

illustration : N mesures de v.a. x_i distribuées uniformément entre -1 et $+1$

loi des grands nombres :

la **moyenne de l'échantillon** tend vers la **moyenne de la fonction de densité de probabilité** des variables



Théorème de la limite centrale

théorème de la limite centrale

$\{x_1, \dots, x_N\}$ un ensemble de V.A. indépendantes de moyennes μ_i , et de variances σ_i^2
(sous des conditions – peu restrictives – limitant l'augmentation des μ_i et σ_i^2 avec i)

alors le **TLC** donne la distribution de la V.A. $S = \sum_{i=1}^N x_i$ lorsque N tend vers l'infini :

$$\left(\frac{S - \sum_i \mu_i}{\sqrt{\sum_i \sigma_i^2}} \right) \xrightarrow{N \rightarrow \infty} \mathcal{N}(0,1)$$

loi normale standard
(gaussienne $\mu = 0$ et $\sigma = 1$)

indépendamment de la distribution de départ !

cas particulier

si les x_i (indépendants) ont la même moyenne μ et la même variance σ^2 , et si on définit les V.A.:

$$y_N = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{et} \quad z_N = \frac{y_N - \mu}{\sigma / \sqrt{N}}$$

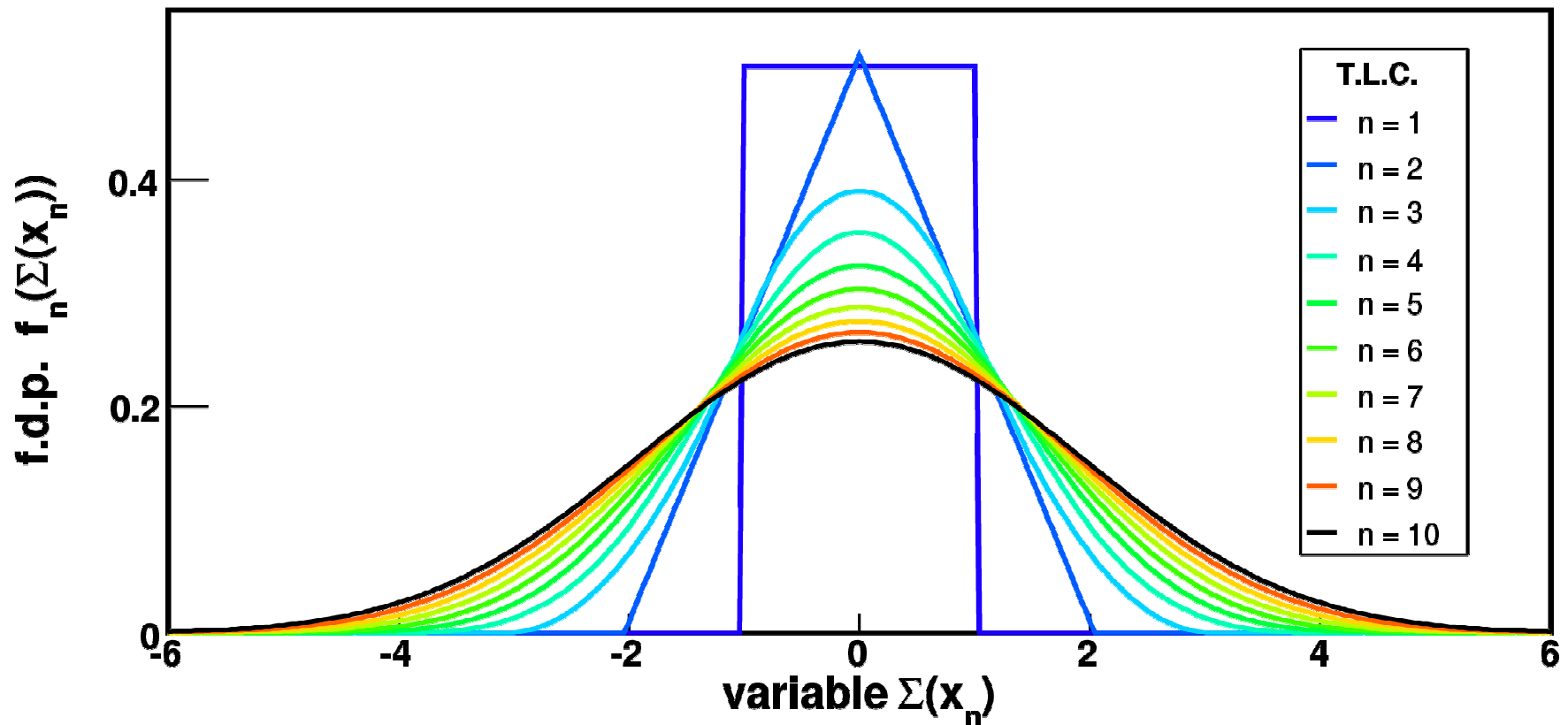
alors $E[z_N] = 0$ et $V[z_N] = 1$

Théorème de la limite centrale

T.L.C. :

la fonction de distribution de $\sum x_i$ tend vers une loi normale de centre $\sum \mu_i$ et de variance $\sum \sigma_i^2$

$$f\left(\sum_i x_i\right) \xrightarrow{N \rightarrow \infty} \mathcal{N}\left(\sum_i \mu_i, \sum_i \sigma_i^2\right)$$



Distributions courantes de probabilités

distributions discrètes

loi de Poisson

loi binomiale

distributions continues

loi « normale » (gaussienne)

distribution du Chi²

autres distributions courantes : Student, Fisher-Snedecor,...

Loi binomiale (dist. discrète)

cas d'utilisation

probabilité d'obtenir n succès parmi N tirages lorsque la probabilité de succès d'un tirage est p

fonction de probabilité

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

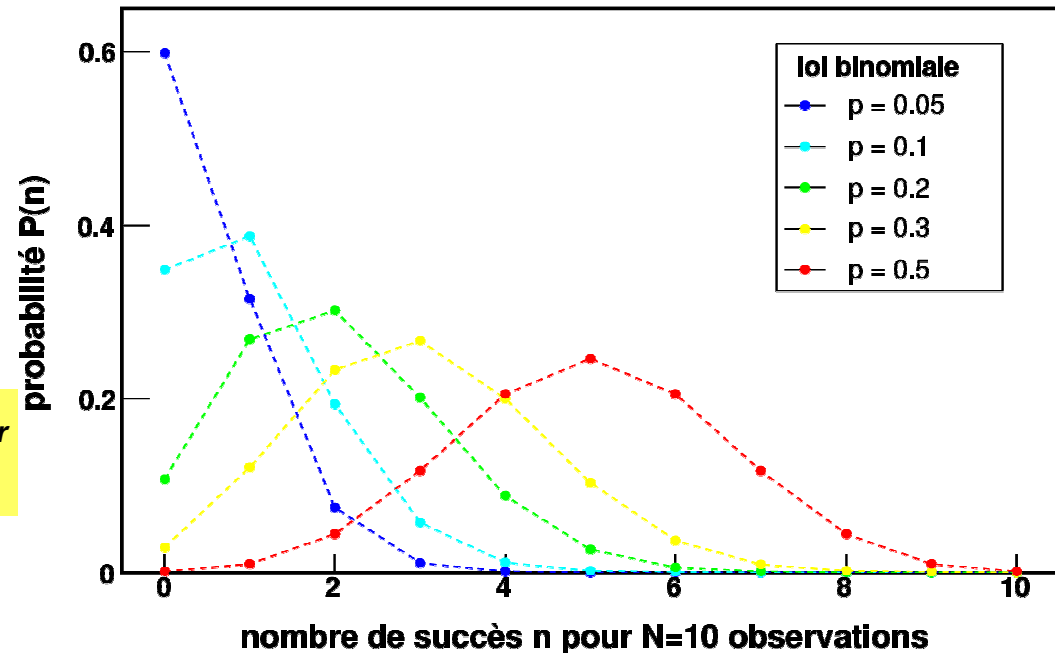
variable $0 \leq n \leq N$ (entier)

param. $N > 0$ (entier)
 $0 < p < 1$ (réel)

caractéristiques

espérance : Np

variance : $Np(1-p)$



f.d.p. liées à la binomiale

binomiale négative : probabilité de devoir attendre N tirages pour obtenir n succès

multinomiale : généralisation au cas de plus de 2 issues possibles

Loi de Poisson (dist. discrète)

cas d'utilisation

probabilité d'observer n événements lorsqu'on en attend λ

fonction de probabilité

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

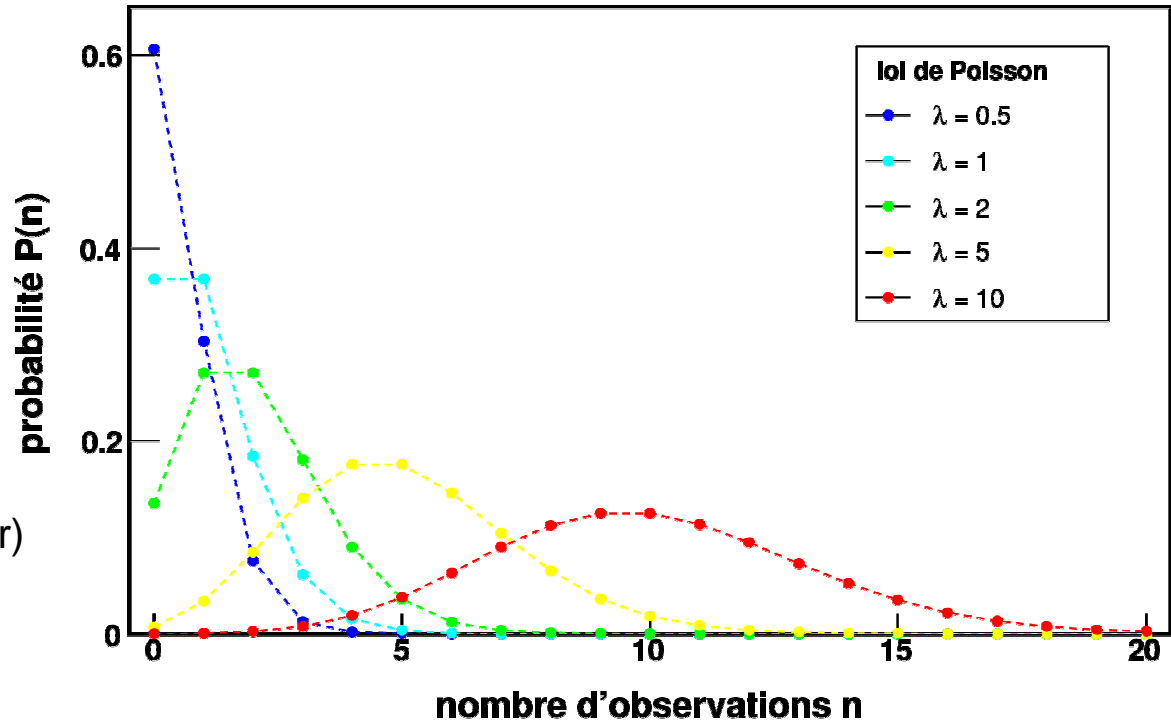
variable $n \geq 0$ (entier)

param. $\lambda > 0$ (réel)

caractéristiques

espérance : λ

variance : λ



f.d.p. dérivée de la loi de Poisson

loi exponentielle : (loi continue)

év^{ts} indépendants avec une fréquence moyenne ν

→ proba. de l'intervalle de temps t entre 2 événements consécutifs :

$$P(t|\nu) = \nu \cdot e^{-\nu t} \quad \text{moy.: } 1/\nu$$
$$\text{var.: } 1/\nu^2$$

Loi « normale » (dist. continue)

cas d'utilisation

f.d.p. la plus utilisée !...

(notamment en raison du T.L.C.)

propriétés uniques (non détaillées ici)

fonction de probabilité

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

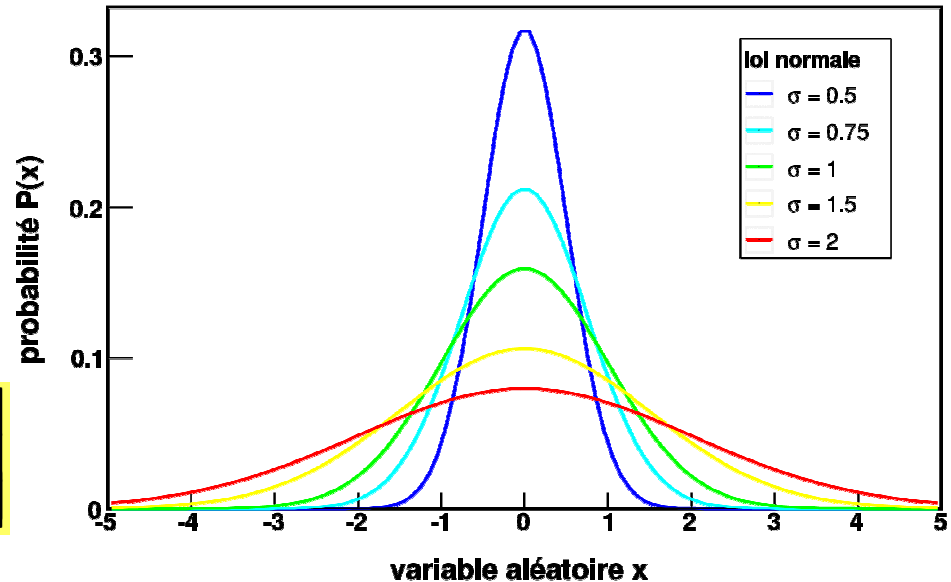
variable x (réel)

param. μ, σ (réels)

caractéristiques

espérance : μ

variance : σ



remarques

loi normale « standard » pour $\mu = 0$ et $\sigma = 1$

toute combinaison linéaire de V.A. normales est **normale**

moments (fct caractéristique) :

$$\begin{cases} \mu_{2r-1} = 0 \\ \mu_{2r} = \frac{(2r)!}{r!} \cdot \left(\frac{\sigma^2}{2} \right)^r \end{cases}$$

loi normale à plusieurs dimensions...

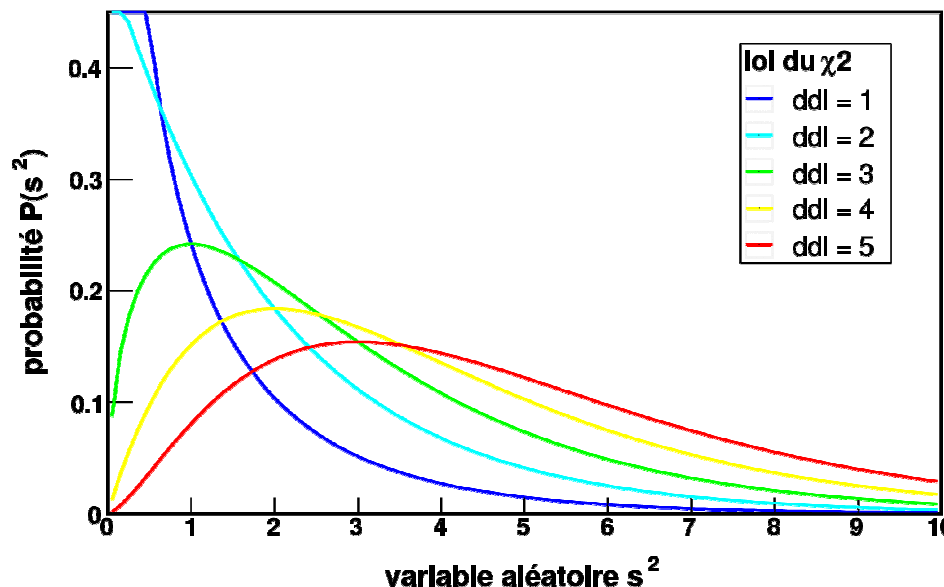
Distribution du Chi2 (dist. continue)

cas d'utilisation

si x est une v.a. distribuée selon une loi normale standard ($\mu = 0$ et $\sigma = 1$), alors la f.d.p. de

$$S = s^2 = \sum_{i=1}^n (x_i)^2$$

est une distribution du χ^2 à n degrés de liberté



fonction de probabilité

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

variable $S = s^2 \geq 0$ (réel)
 param. $n \geq 1$ (entier)

caractéristiques

espérance : n
 variance : $2n$

propriétés

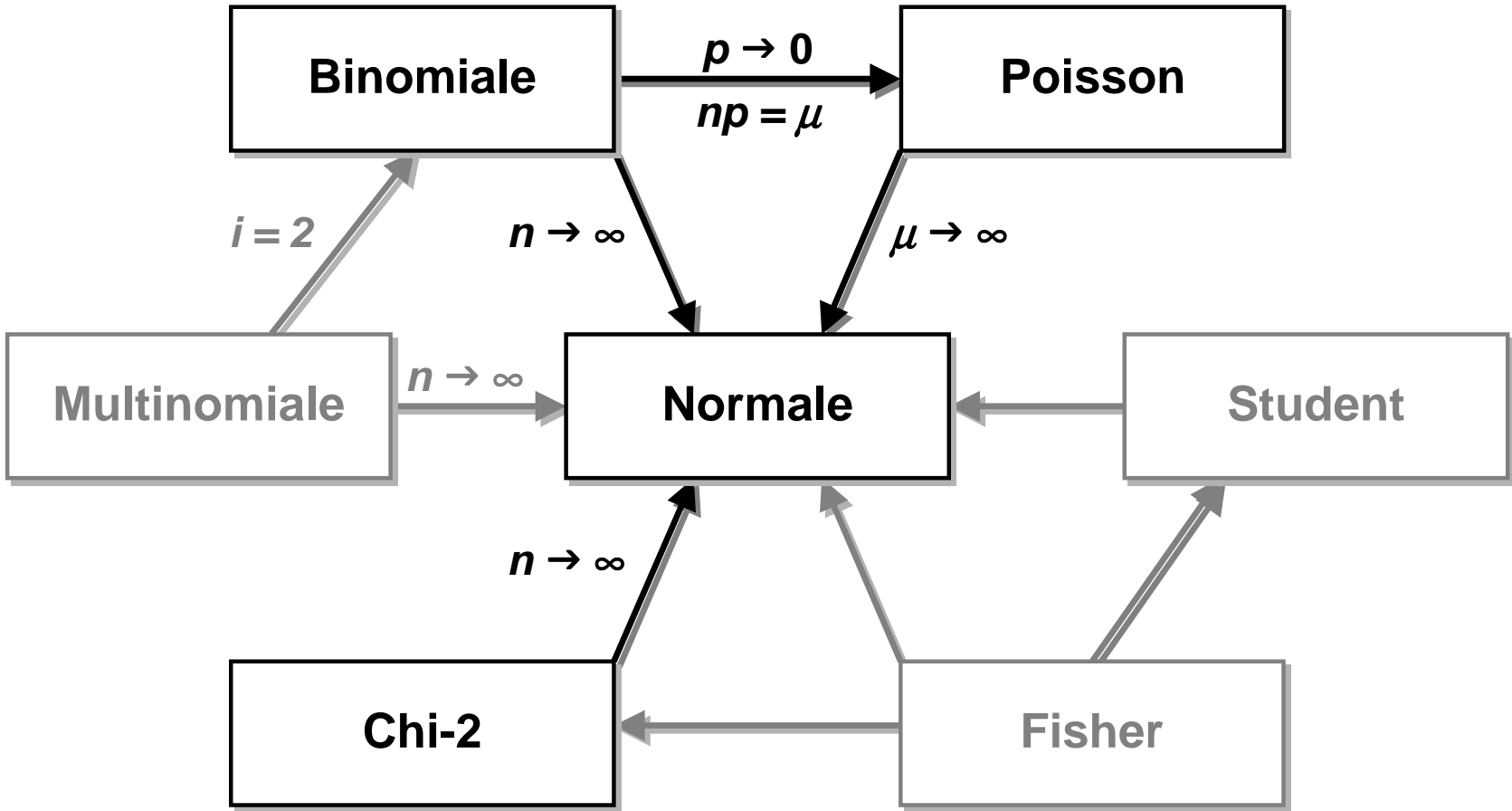
si $s_{(n)}^2$ et $s_{(m)}^2$ indépendantes suivent des dist. du χ^2 à n et m d.d.l, alors

$s_{(p)}^2 = s_{(n)}^2 + s_{(m)}^2$ suit une dist. du χ^2 à $p = n + m$ d.d.l.

les V.A. $y_n = \frac{s_{(n)}^2 - n}{\sqrt{2n}}$ et $z_n = \sqrt{2s_{(n)}^2} - \sqrt{2n-1}$

suivent une loi standard $\mathcal{N}(0,1)$

Relations asymptotiques



partie 3

Analyse statistique

Estimateurs

Comportement asymptotique

Intervalle de confiance

Échantillon

moyenne, variance, moments

soit une variable aléatoire X

on fait N mesures de cette variable : $x_1, \dots, x_N \rightarrow$ échantillon

distribution de probabilité de la V.A.

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

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

moments $\mu_j = \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N (x_i)^j \right)$

pour l'échantillon

moyenne $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$

variance $s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$

elles ne sont pas
indépendantes !

Fonction de vraisemblance

concepts de base de la théorie de l'information

fonction de vraisemblance

définition :

\mathbf{x} est une V.A. de f.d.p. $f(\mathbf{x}, \theta)$

échantillon : N observations $\mathbf{x}_1, \dots, \mathbf{x}_N$

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

⇒ « **probabilité globale** » de l'observation complète

information

f.d.p. à 1 paramètre :

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

f.d.p. à plusieurs paramètres :

$$[I_{\bar{\theta}}(\bar{\theta})]_{ij} = E \left[\frac{\partial L(\bar{\mathbf{x}}|\bar{\theta})}{\partial \theta_i} \cdot \frac{\partial L(\bar{\mathbf{x}}|\bar{\theta})}{\partial \theta_j} \right]$$

au delà des objectifs
de ce cours

Estimateur d'une variable aléatoire

qu'est-ce qu'un estimateur ?

- variable aléatoire x distribuée selon une f.d.p. $f(x|\theta_0)$
- échantillon (mesure) $X_N = \{x_i\}$ ou $i = 1, \dots, N$

Que peut-on dire de θ_0 à partir de l'échantillon ?

un estimateur $\theta(X_N)$ est une fonction de l'échantillon qui tente de donner une valeur s'approchant de la valeur « vraie » du paramètre

un estimateur est une variable aléatoire !

qualités d'un estimateur

un estimateur doit être :

- **consistant** : il converge vers la vraie valeur du paramètre
- **non biaisé** : si $E[\theta] = \theta_0$ quel que soit le nombre d'observations N
- **efficace** : s'il converge *rapidement* vers la valeur du paramètre
- **robuste** : lié à la fiabilité de l'estimateur si la f.d.p. est peu ou mal connue (sensibilité à la forme de la f.d.p.)

Estimation ponctuelle (1)

estimateur intuitif

\mathbf{x} est une V.A. avec une f.d.p. $f(\mathbf{x}|\theta_0)$ \Rightarrow toute fonction $\mathbf{a}(\mathbf{x})$ est une V.A.
échantillon $\mathbf{X} = \{\mathbf{x}_i; i = 1, \dots, N\}$

loi des grands nombres :

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

si $\mathbf{a}(\mathbf{x})$ est telle que $E[\mathbf{a}(\mathbf{x})|\theta] = h(\theta)$

alors $\theta_0 = h^{-1}(E[\mathbf{a}(\mathbf{x})|\theta_0]) = h^{-1}\left(\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i)\right)$

estimateur de θ_0 :

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

grandeur expérimentale $\xi = \frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i)$

Estimation ponctuelle (2)

cas de plusieurs paramètres

si la f.d.p. dépend de plusieurs paramètres $\{\theta_j; j = 1, \dots, r\}$

→ r fonctions $\mathbf{a}_j(\mathbf{x})$

→ r fonctions $h_j(\theta_1, \dots, \theta_r)$

→ r grandeurs expérimentales $\xi_j = \frac{1}{N} \sum_{i=1}^N \mathbf{a}_j(\mathbf{x}_i) = h_j(\hat{\theta}_1, \dots, \hat{\theta}_r)$

système d'équations:

$$\begin{cases} \xi_1 = h_1(\hat{\theta}_1, \dots, \hat{\theta}_r) \\ \vdots \\ \xi_r = h_r(\hat{\theta}_1, \dots, \hat{\theta}_r) \end{cases} \Leftrightarrow \begin{cases} \hat{\theta}_1 = (h^{-1})_1(\xi_1, \dots, \xi_r) \\ \vdots \\ \hat{\theta}_r = (h^{-1})_r(\xi_1, \dots, \xi_r) \end{cases}$$

méthode des moments

elle correspond au choix $\mathbf{a}_j(\mathbf{x}) = \mathbf{x}^j$

ex.: \mathbf{X} distribuée selon une f.d.p. de moyenne μ et de variance σ^2 (param. $\theta_1 = \mu$ et $\theta_2 = \sigma$)

$$\begin{cases} \xi_1 = \frac{1}{N} \sum_{i=1}^N x_i & \xrightarrow{N \rightarrow \infty} h_1(\mu, \sigma) = \int_{\Omega_x} x \cdot f(x|\mu, \sigma) \cdot dx = \mu \\ \xi_2 = \frac{1}{N} \sum_{i=1}^N (x_i)^2 & \xrightarrow{N \rightarrow \infty} h_2(\mu, \sigma) = \int_{\Omega_x} x^2 \cdot f(x|\mu, \sigma) \cdot dx = \sigma^2 + \mu^2 \end{cases}$$

soit :

$$\hat{\mu} = (h^{-1})_1(\xi_1, \xi_2) = \xi_1$$

$$\hat{\sigma} = (h^{-1})_2(\xi_1, \xi_2) = \sqrt{\xi_2 - (\xi_1)^2}$$

Estimation ponctuelle (3)

estimateur implicite

on peut généraliser au cas d'une fonction de \mathbf{x} et de θ : $\mathbf{a}(\mathbf{x}, \theta)$
en choisissant \mathbf{a} de sorte que $\mathbf{h}(\theta_0) = \mathbf{0}$

$$\mathbf{h}(\theta|\theta_0) = E[\mathbf{a}(\mathbf{x}, \theta)|\theta_0] = \int_{\Omega_{\mathbf{x}}} \mathbf{a}(\mathbf{x}, \theta) f(\mathbf{x}, \theta_0) d\mathbf{x}$$

avec $\mathbf{h}(\theta_0|\theta_0) = \mathbf{0}$

la grandeur expérimentale devient une *fonction du paramètre* : $\xi(\theta) = \frac{1}{N} \sum_{i=1}^N \mathbf{a}(\mathbf{x}_i, \theta)$

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

un **estimateur implicite** vérifie donc :

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

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

méthode des moments

$\mathbf{a}_1(\mathbf{x}) = \mathbf{x} - \mu$ et $\mathbf{a}_2(\mathbf{x}) = \mathbf{x}^2 - (\mu^2 + \sigma^2)$

on retrouve le résultat de l'estimateur intuitif

$$\begin{cases} \xi_1(\hat{\mu}, \hat{\sigma}) = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \hat{\mu}) \\ \xi_2(\hat{\mu}, \hat{\sigma}) = \frac{1}{N} \sum_{i=1}^N ((\mathbf{x}_i)^2 - (\hat{\mu}^2 + \hat{\sigma}^2)) \end{cases}$$

Estimation ponctuelle (4)

estimateur par maximisation / minimisation

estimateur implicite : résoudre $\xi(\theta) = 0$

si on considère une fonction $g(\mathbf{x}, \theta)$ telle que : $\mathbf{a}(\mathbf{x}, \theta) = \frac{d}{d\theta} g(\mathbf{x}, \theta)$

on obtient alors de façon triviale
un estimateur de θ_0 en résolvant :

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

maximum de vraisemblance

il correspond au choix de $g(\mathbf{x}_i, \theta) = \ln(f(\mathbf{x}_i | \theta))$

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

estimateur consistant :
conditions de dérivabilité
et de commutation

moindres carrés

$g(\mathbf{x}, \theta)$ est une forme quadratique du type (cas de θ à r dimensions) :

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

\mathbf{W} : matrice de poids ; $\mathbf{M}(\theta) = E[\mathbf{X} | \theta]$ (espérance)

soit à résoudre un système de r équations :

$$\bar{\xi}(\bar{\theta})_{(r)} = \bar{\nabla}_{\theta} (g(\bar{\mathbf{X}}, \bar{\theta}))_{(r)} = -\bar{\nabla}_{\theta} [\bar{\mathbf{M}}(\bar{\theta})]^T \cdot \bar{\mathbf{W}} \cdot [\bar{\mathbf{X}} - \bar{\mathbf{M}}(\bar{\theta})] = 0$$

Cas pathologique

récréatif...

le **maximum de vraisemblance** est un des estimateurs les plus utilisés
mais il n'est pas toujours le mieux adapté...

un cas d'école

x une variable aléatoire uniformément distribuée entre 0 et θ_0 : $f(x) = 1 / \theta_0$

on fait N mesures : $L(x_1, \dots, x_N | \theta) = 1 / \theta^N$

→ maximum pour $\theta = x_+$, où x_+ est la plus grande des valeurs observées (θ ne peut pas être plus petit...)

f.d.p. de l'estimateur $\theta = x_+$:

$$N = 1 \quad \theta = x_1 \quad f_1(\theta) = 1 / \theta_0$$

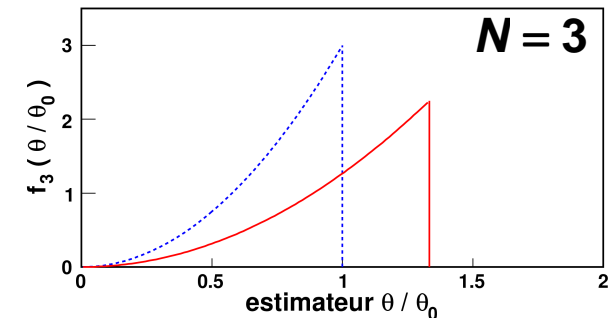
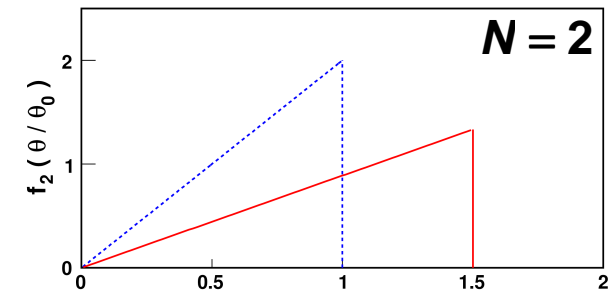
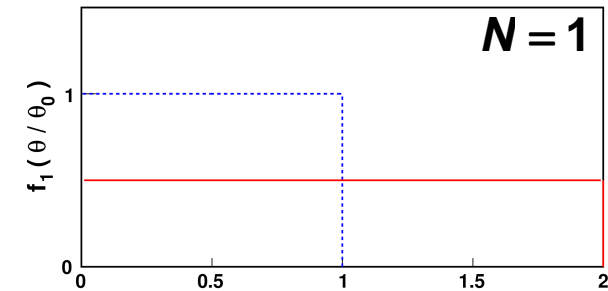
$$N = 2 \quad f_2(\theta) = P(x_1 < \theta) \cdot f(x_2 = \theta) + P(x_2 < \theta) \cdot f(x_1 = \theta) = 2 \theta / \theta_0^2$$

$$N \text{ qcq} \quad f_N(\theta) = N \cdot \theta^{N-1} / (\theta_0)^N$$

biaisé quel que soit le nombre N de mesures
 $E[\theta] < \theta_0$

estimateur non biaisé : $\theta = x_+ + (x_+ / N)$

$$f_N(\theta) = N \cdot \left(\frac{N}{N+1} \right)^N \cdot \frac{\theta^{N-1}}{\theta_0^N}$$



Comportement asymptotique (1)

quelle incertitude sur l'estimation ?

loi des grands nombres → estimateur
théorème de la limite centrale → incertitude

cas d'un estimateur intuitif

TCL : $\xi = \frac{1}{N} \sum_{i=1}^N \mathbf{a}(x_i)$ *asymptotiquement* ($N \rightarrow \infty$) distribué selon une loi normale de moyenne $E[\mathbf{a}(x)]$ et de variance $N^{-1} V[\mathbf{a}(x)]$

développement au premier ordre autour de $\xi_0 = E[\mathbf{a}(x)]$

$$\hat{\theta} = h^{-1}\left(\frac{1}{N} \sum_{i=1}^N \mathbf{a}(x_i)\right) = h^{-1}(\xi) \quad \hat{\theta} \approx h^{-1}(\xi_0) + \frac{\partial h^{-1}(\xi_0)}{\partial \xi} \times \{\xi - \xi_0\} + \dots$$

$$\hat{\theta} \approx h^{-1}(E[\mathbf{a}]) + \frac{\partial h^{-1}(E[\mathbf{a}])}{\partial \xi} \times \left\{ \frac{1}{N} \sum_{i=1}^N \mathbf{a}(x_i) - E[\mathbf{a}] \right\} + \dots$$

tous les termes sont constants, sauf ξ \Rightarrow **normalement distribués**, et

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

\Rightarrow **estimation** de l'incertitude : à partir de la moyenne et de la variance de l'échantillon $\mathbf{a}(x_i)$

Comportement asymptotique (2)

généralisation

calcul similaire pour un estimateur implicite

$$\begin{aligned} \xi &\rightarrow \xi(\theta) \\ \frac{\partial h^{-1}(E[a])}{\partial \xi} &\rightarrow \frac{1}{E\left[\frac{\partial \xi(\theta)}{\partial \theta}\right]_{\theta=\theta_0}} \end{aligned}$$

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

maximum de vraisemblance

$$a(x_i, \theta) = \frac{\partial}{\partial \theta} \ln(f(x_i | \theta))$$

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

où I_N est l' « information » vue précédemment

cas de plusieurs paramètres

→ écriture matricielle

cas de paramètres contraints

→ changements de variables

non détaillé ici

Intervalle de confiance (1)

grands nombres et limite centrale

lorsque $N \rightarrow \infty$

loi des grands nombres → **estimateur**

théorème de la limite centrale → **incertitude**

qu'en est-il lorsque le nombre d'observation est faible ?

on ne peut plus s'appuyer sur le comportement asymptotique

Intervalle de confiance (2)

Approche classique : ceinture de confiance

échantillon X , distribué selon $f(X | \theta)$, où θ est inconnu

⇒ estimateur de θ : $t(X)$

$t(X)$ est une variable aléatoire (une *statistique* de l'échantillon) avec une f.d.p. $f_t(t, \theta)$ qui peut être déterminée en fonction de θ

on peut définir $t_1(\theta)$ et $t_2(\theta)$ tel que $P(t < t_1 | \theta) = \alpha_1$

$$P(t > t_2 | \theta) = \alpha_2$$

alors $P(t_1 < t < t_2 | \theta) = \beta = 1 - (\alpha_1 + \alpha_2)$

$$\int_{t_1}^{t_2} f_t(t | \theta) \cdot dt = \beta$$

θ_0 , la valeur « vraie », est inconnue...

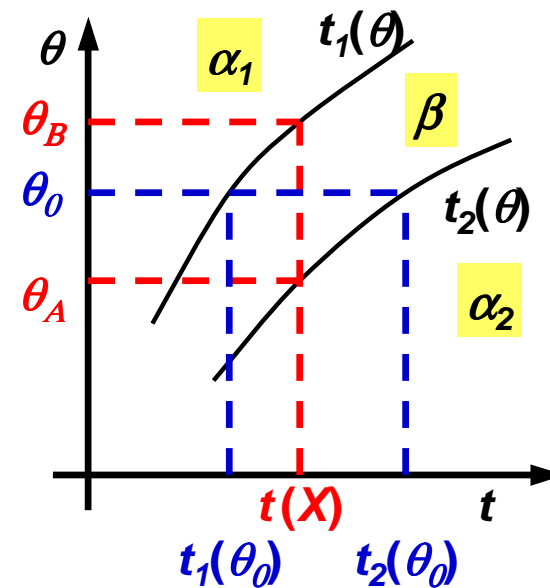
à partir de l'estimateur $t(X)$, on définit un intervalle de confiance $[\theta_A, \theta_B]$:

$$\theta_A \rightarrow t_2(\theta_A) = t(X)$$

$$\theta_B \rightarrow t_1(\theta_B) = t(X)$$

comme $P(t_1 < t < t_2 | \theta_0) = \beta$, alors la probabilité que la vraie valeur θ_0 soit dans $[\theta_A, \theta_B]$ est aussi β

(tous les intervalles $[\theta_A, \theta_B]$ obtenus pour $t_1 < t < t_2$ contiennent θ_0)



Intervalle de confiance (3)

limites supérieure ou inférieure

l'intervalle de confiance (avec un degré de confiance β) n'est pas unique
en règle général, on fait le choix symétrique :

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

il n'y a pas à priori, d'un point de vue statistique, de choix préférable

dans certain cas cependant, il peut être préférable de choisir α_1 et α_2 de façon asymétrique

les cas extrêmes sont obtenus pour $\alpha_1 = 0$ ou $\alpha_2 = 0$

$$\text{lim. inf.} \quad \alpha_1 = 0 \quad \rightarrow \quad t_1 = -\infty \rightarrow P(\theta > \theta_A) = \beta$$

$$\text{lim. sup.} \quad \alpha_2 = 0 \quad \rightarrow \quad t_2 = +\infty \rightarrow P(\theta < \theta_B) = \beta$$

ex.: mesure d'un nombre d'événements d'un type donné, lorsque **aucun** événement de ce type n'est observé : intervalle $[0; \theta_{sup}]$

Intervalle de confiance (4)

approche bayésienne

la probabilité *à posteriori* est déterminée à partir de la probabilité *à priori*

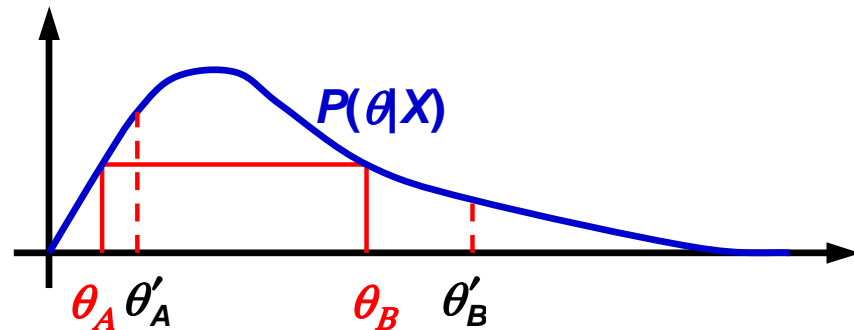
$$P(\theta|X) = \frac{P(X|\theta) \cdot P(\theta)}{\int_{\Omega_\theta} P(X|\theta') \cdot P(\theta') \cdot d\theta'}$$

un intervalle de confiance $[\theta_A, \theta_B]$ avec un niveau de confiance β vérifie la condition :

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

le choix de l'intervalle n'est pas unique
si on impose également que $P(\theta_A|X) = P(\theta_B|X)$ 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 à l'extérieur



Remarques sur les intervalles de confiance

présentation du résultat

l'estimateur donne **une valeur** (la plus probable) : $\hat{\theta}_0$

ex.: *maximum de vraisemblance*

intervalle de confiance à 68,3 % : $[\theta_A, \theta_B]$

incertitude à « 1σ » (par abus de langage)

par analogie avec la gaussienne

$$\Rightarrow \theta = \hat{\theta}_0 + \begin{cases} \theta_B - \hat{\theta}_0 \\ \hat{\theta}_0 - \theta_A \end{cases}$$

comparaison entre les approches classiques et bayésiennes

l'hypothèse sur $P(\theta)$ dans l'approche bayésienne apporte de l'information

⇒ **l'intervalle de confiance bayésien est toujours plus petit que l'intervalle de confiance classique**

partie 4

**Applications dans le cas
de faibles statistiques**

Loi discrète
mesure d'un rapport
d'embranchement

Loi continue
mesure de durée de vie

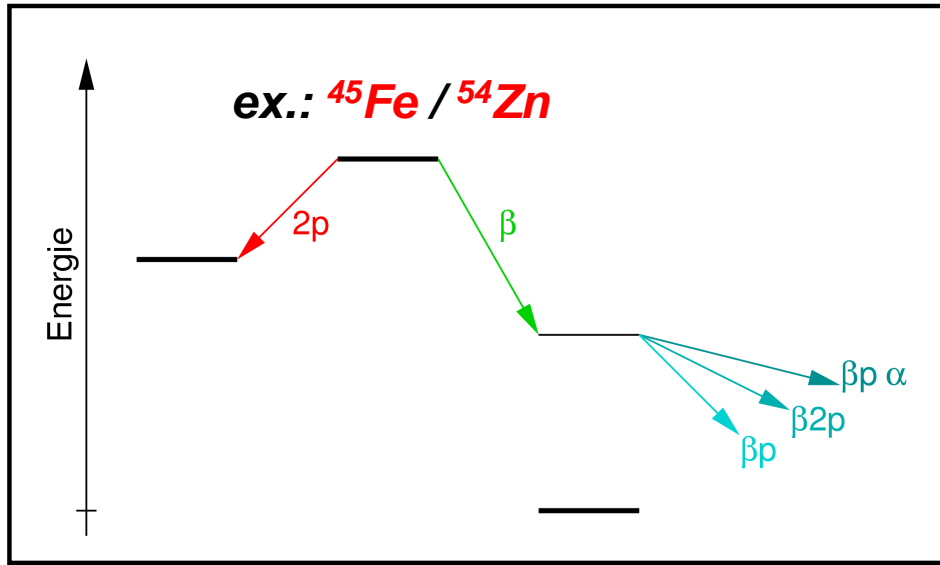
**les données numériques
des exemples ne sont pas
les résultats expérimentaux
véritables**

Application 1 : rapport d'embranchement

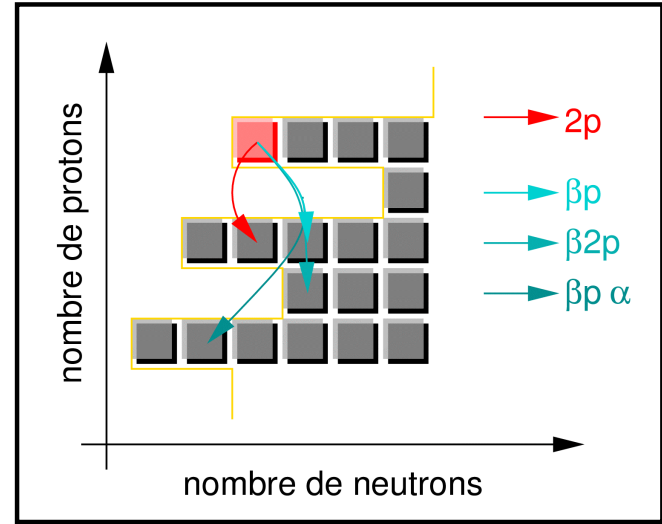
exemple de la radioactivité 2-protons

succès : radioactivité **2p**

échec : décroissance **β** (+p, +2p, +p α , ...)



expérience : on produit $N_{obs} = 10$ noyaux ^{54}Zn
 on observe $n_{2p} = 7$ décroissances **2p** \Rightarrow



loi binomiale

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

que peut-on dire pour le rapport d'embranchement **2p** ?

(probabilité $p = R_{2p}$)

Loi binomiale

un événement à 2 issues possibles :
 - « succès » avec une probabilité p
 - « échec » avec une probabilité $(1-p)$

on considère un échantillon de N événements
 ⇒ quelle est la probabilité d'avoir n succès ?

$$N = 1 \quad P(0) = 1-p$$

$$P(1) = p$$

$$N = 2 \quad P(0) = P\{e,e\} = (1-p)^2$$

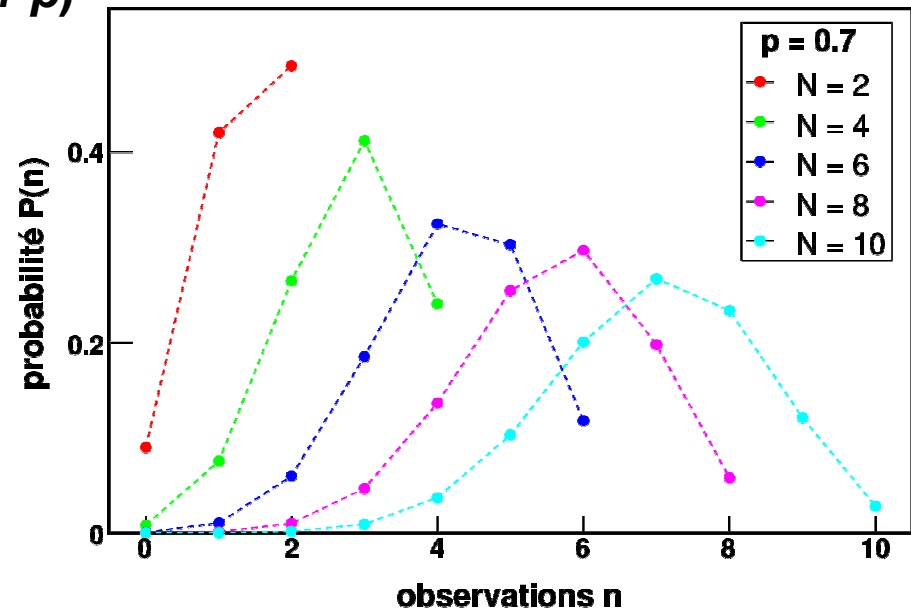
$$P(1) = P\{s,e\} + P\{e,s\} = 2p(1-p)$$

$$P(2) = P\{s,s\} = p^2$$

...

loi binomiale

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



Ceinture de confiance de la loi binomiale

estimateur : M.V.

$$L(n|p, N) = \prod_{i=1}^n P_i = p^n \cdot (1-p)^{N-n}$$

$$\frac{L(n|\hat{p}, N)}{dp} = 0 \Leftrightarrow \hat{p} = \frac{n}{N} \propto n$$

intervalle de confiance :

pour $\hat{p} = \frac{n_{\text{exp}}}{N}$

p_B tel que $P(n \leq n_{\text{exp}} | p_B) < \alpha_2$

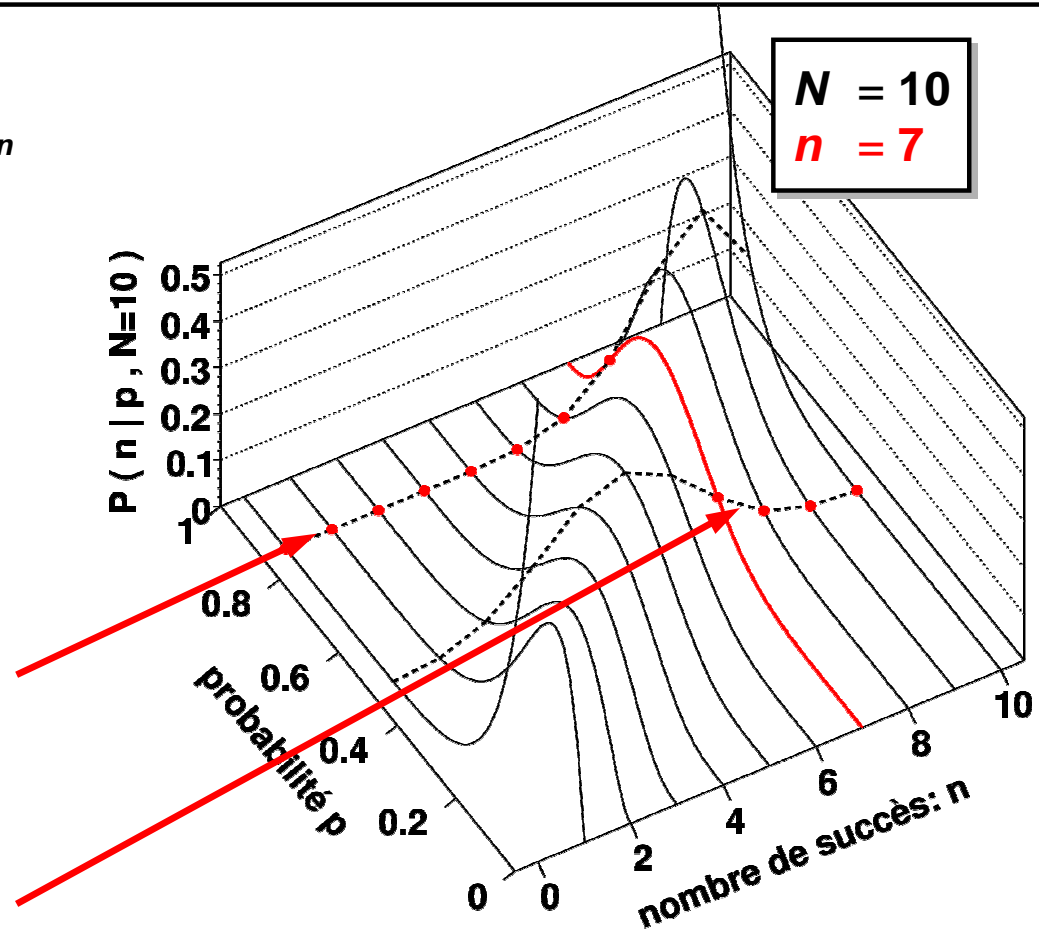
solution de $\sum_{n=n_{\text{exp}}}^N P(n|p_A, N) = \alpha_2$

p_A tel que $P(n \geq n_{\text{exp}} | p_A) < \alpha_1$

solution de $\sum_{n=0}^{n_{\text{exp}}} P(n|p_B, N) = \alpha_1$

intervalle avec un degré de confiance

$$\beta = 1 - (\alpha_1 + \alpha_2)$$



Ceinture de confiance de la loi binomiale

estimateur : M.V.

$$L(n|p, N) = \prod_{i=1}^n P_i = p^n \cdot (1-p)^{N-n}$$

$$\frac{L(n|\hat{p}, N)}{dp} = 0 \Leftrightarrow \hat{p} = \frac{n}{N} \propto n$$

intervalle de confiance :

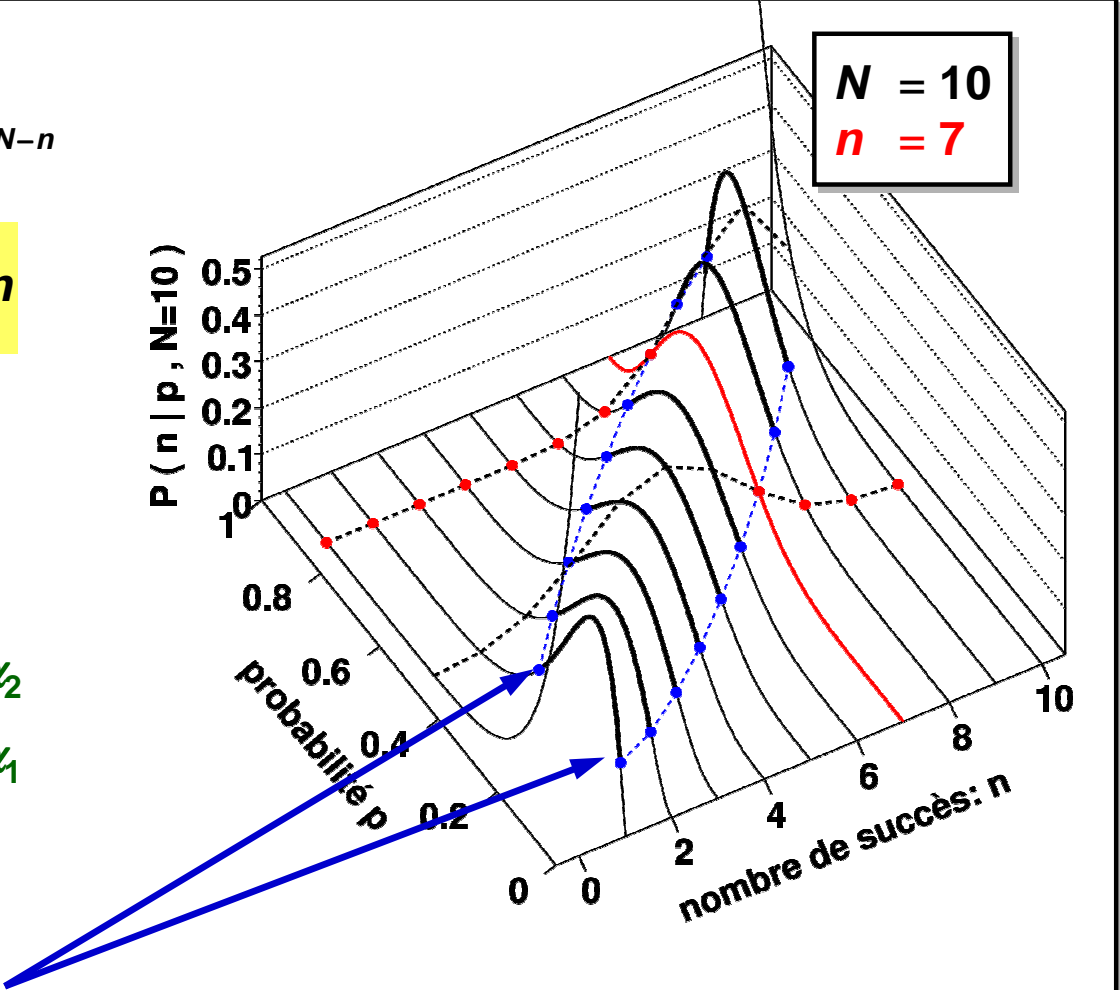
pour $\hat{p} = \frac{n_{\text{exp}}}{N}$

p_B tel que $P(n \leq n_{\text{exp}} | p_B) < \alpha_2$

p_A tel que $P(n \geq n_{\text{exp}} | p_A) < \alpha_1$

ceinture de confiance :

pour l'ensemble des nombres d'observations possibles

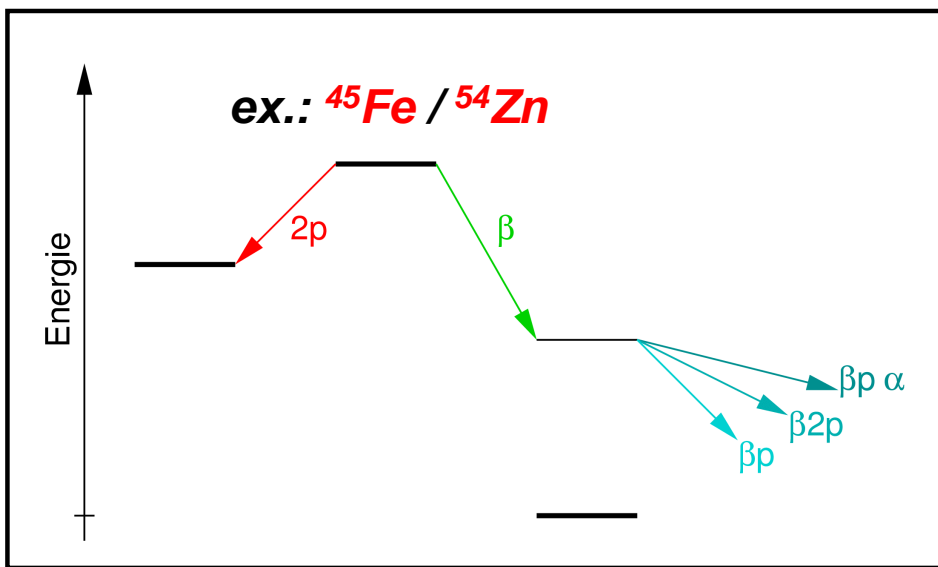


Résultat : approche classique

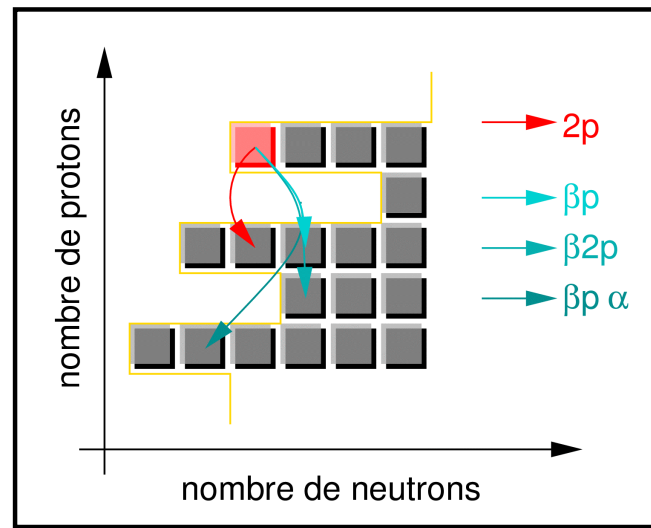
exemple de la radioactivité 2-protons

succès : radioactivité **2p**

échec : décroissance β (+p, +2p, +p α , ...)



expérience : on produit $N = 10$ noyaux ^{54}Zn
on observe $n = 7$ décroissances **2p** \Rightarrow



résultat

estimateur (MV)	$R_{2p} = 0,7$
int. conf. ($\beta = 68,3\%$)	$R_{min} = 0,492$
	$R_{max} = 0,858$

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

Approche bayésienne

théorème de Bayes

échantillon X : N observations,
 → n succès,
 → $(N - n)$ échecs

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

hypothèse : toutes les valeurs du rapport d'embranchements $p = R_{2P}$ sont possibles, sans préférence « à priori » → $P(p)$ uniforme

maximum de vraisemblance

fonction de vraisemblance

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

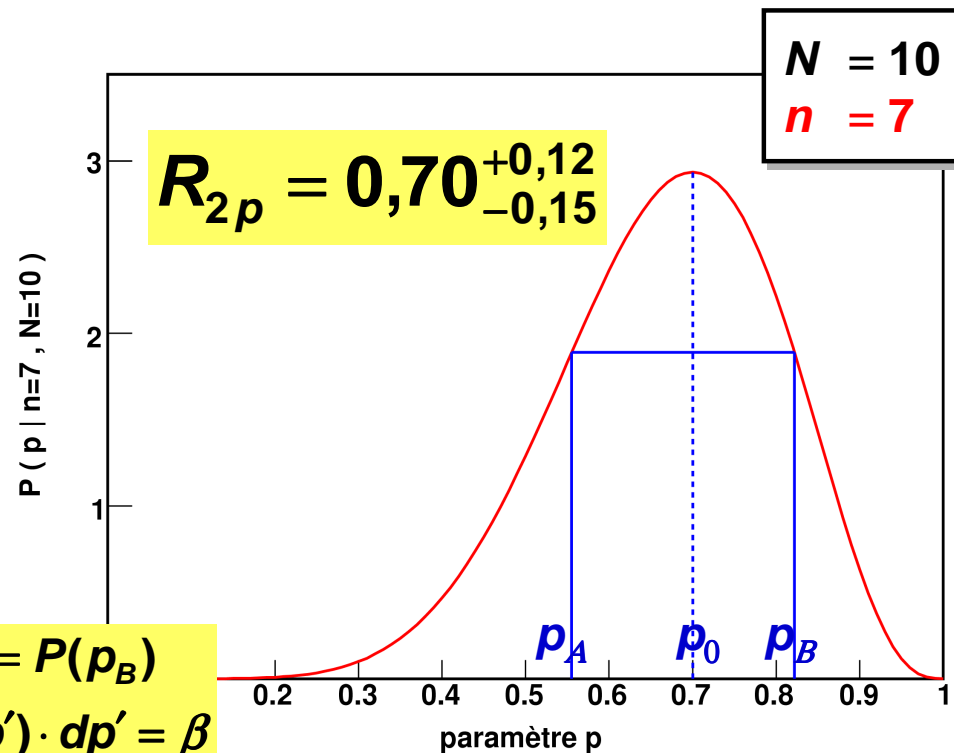
normalisation $\int_0^1 L(n|p, N) = \frac{N+1}{\binom{N}{n}}$

résolution

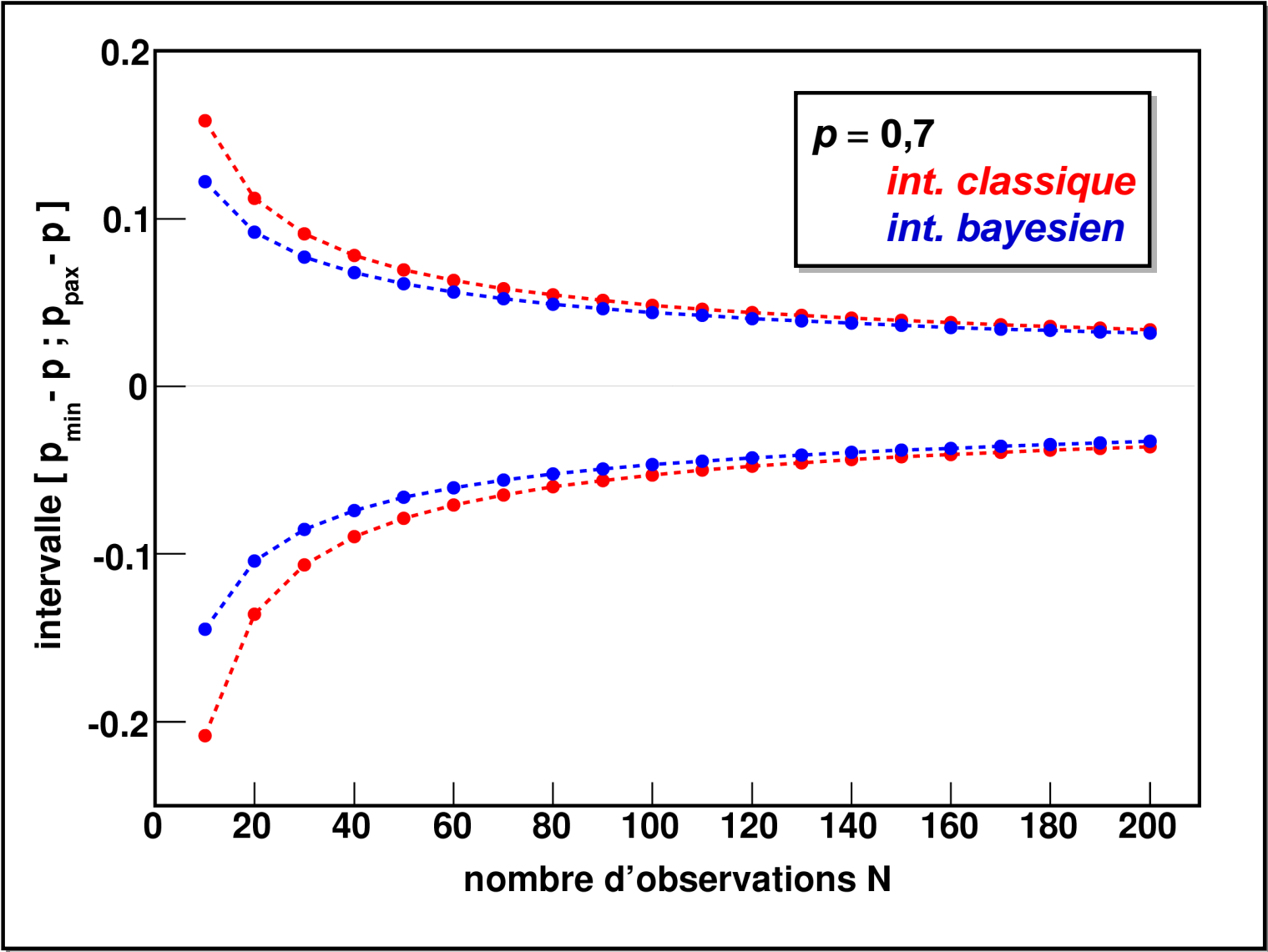
$$P(p|X) = \frac{1}{N+1} \cdot \binom{N}{n} \cdot p^n \cdot (1-p)^{N-n}$$

les bornes de l'intervalle de confiance sont les solutions de :
 (résolution numérique)

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



Comparaison classique – bayésien



Application 2 : durée de vie

(suite de l'exemple : la radioactivité 2-protons)

on a produit N noyaux *très exotiques* ($\rightarrow N$ est « petit ») et on mesure pour chacun à quel instant (après sa création) a lieu sa décroissance radioactive

loi de décroissance

probabilité de décroissance par unité de temps : λ (constante)

proba. d'avoir la décr. entre t et $t + dt$ (si elle n'a pas déjà eu lieu) : λdt

proba. de ne pas avoir cette décroissance : $1 - \lambda dt$

si $P(t)$ est la probabilité de ne pas avoir la décr. entre 0 et t , alors

$$P(t + dt) = P(t) \cdot (1 - \lambda dt) \Leftrightarrow \frac{P(t + dt) - P(t)}{P(t)} = -\lambda dt \Leftrightarrow P(t) \propto \exp(-\lambda t)$$

\Rightarrow loi de probabilité des décroissances radioactives : $f(t|\lambda) = \lambda \cdot e^{-\lambda t}$

durée de vie (temps pour lequel la probabilité que la décroissance ait eu lieu est de 50%) :

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

Durée de vie : la mesure

observation (mesure)

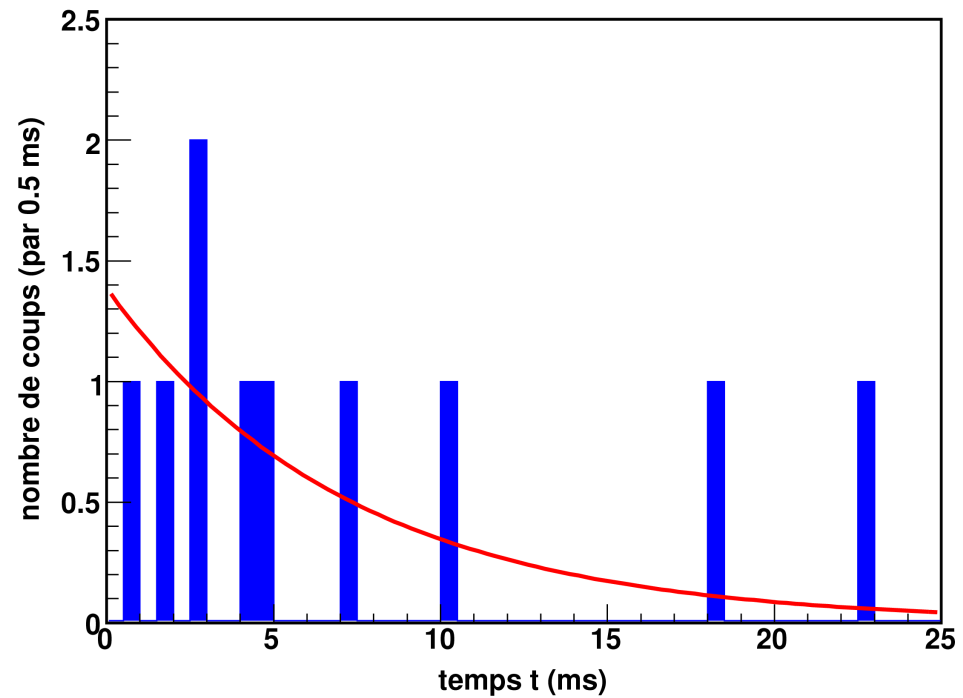
on a observé $N = 10$ noyaux (ceux de l'exemple précédent)

⇒ l'échantillon est l'ensemble des temps décroissance mesurés :

$$\{t_1, t_2, \dots, t_{N=10}\}$$

exemple :

i	t_i (ms)
1	0.905
2	4.550
3	10.278
4	4.456
5	18.310
6	22.741
7	2.856
8	1.590
9	2.940
10	7.331



⇒ que peut-on dire de $T_{1/2}$ (ou de λ) ?

Durée de vie : estimateur

maximum de vraisemblance

fonction de
vraisemblance

$$\begin{aligned} L(\vec{t} = \{t_1, \dots, t_N\} | \lambda) &= \prod_{i=1}^N P_i = \prod_{i=1}^N \lambda \cdot e^{-\lambda t_i} \\ &= \lambda^N \cdot \exp\left[\lambda \cdot \sum_{i=1}^N t_i\right] = \lambda^N \cdot \exp[\lambda t_\Sigma] \end{aligned} \quad t_\Sigma = \sum_{i=1}^N t_i$$

la fonction de vraisemblance ne dépend que de la somme des t_i , mais pas des valeurs individuelles

maximum

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

$$\ln L(t_\Sigma | \lambda) = N \cdot \ln(\lambda) - \lambda t_\Sigma \quad \frac{d(\ln L(t_\Sigma | \lambda))}{d\lambda} = \frac{N}{\lambda} - t_\Sigma$$

estimateur

$$\hat{\lambda} = \frac{N}{t_\Sigma}$$

Durée de vie : f.d.p. de l'estimateur

probabilité de l'estimateur : $f_N(\lambda | \lambda_0)$

probabilité d'observer t_Σ avec on a un échantillon de N mesures

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

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

on retrouve la forme d'une **loi de Poisson** pour une variable $x = \lambda_0 t_\Sigma$
le facteur λ_0 vient du fait que $f(x)dx = f(t)dt$

changement de variable $\lambda = \frac{N}{t_\Sigma} \quad \frac{\partial t_\Sigma}{\partial \lambda} = -\frac{N}{\lambda^2} = -\frac{t_\Sigma^2}{N}$

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

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

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

avec $x = \lambda_0 \cdot t_\Sigma = N \cdot \frac{\lambda_0}{\lambda}$

Ceinture de confiance

intervalle de confiance

λ_B tel que $P(\lambda \leq \lambda_{\text{exp}} | \lambda_B) < \alpha_2$

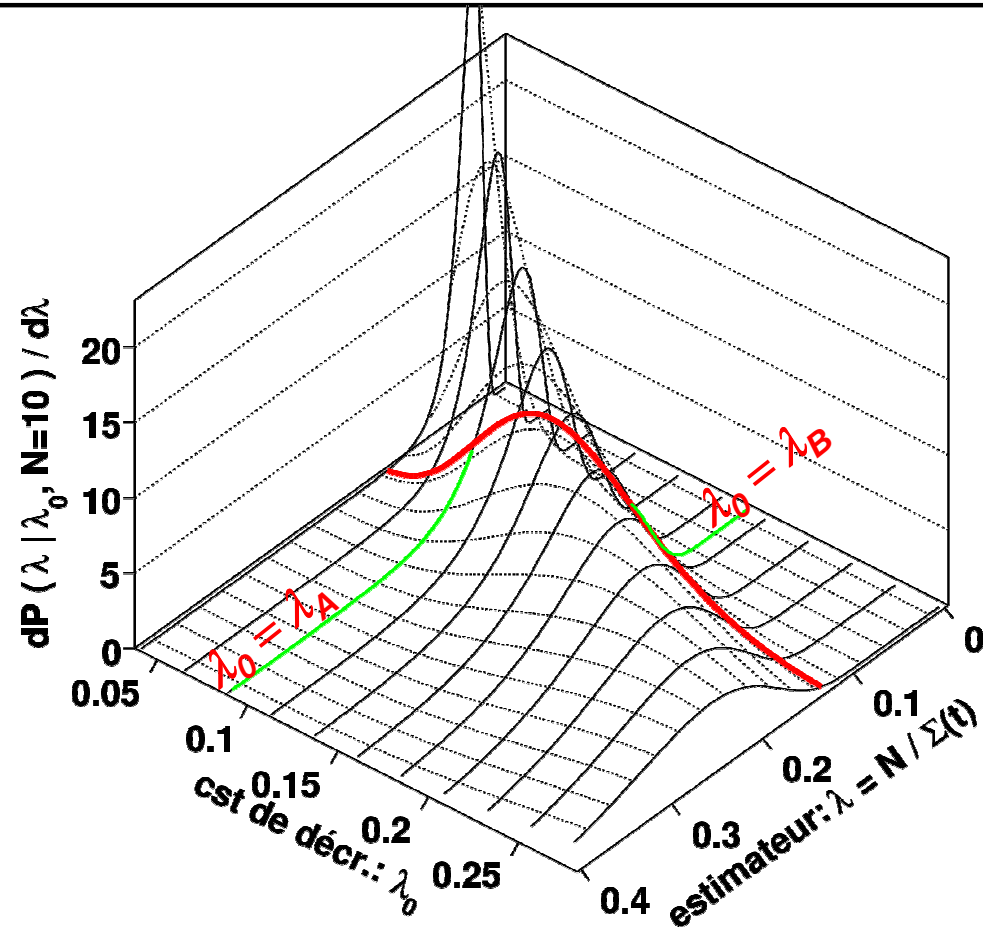
solution de

$$\int_0^{\lambda_{\text{exp}}} f_N(\lambda | \lambda_0) \cdot d\lambda = \alpha_2$$

λ_A tel que $P(\lambda \geq \lambda_{\text{exp}} | \lambda_B) < \alpha_1$

solution de

$$\int_{\lambda_{\text{exp}}}^{\infty} f_N(\lambda | \lambda_0) \cdot d\lambda = \alpha_1$$



Ceinture de confiance

intervalle de confiance

λ_B tel que $P(\lambda \leq \lambda_{\text{exp}} | \lambda_B) < \alpha_2$

solution de

$$\int_0^{\lambda_{\text{exp}}} f_N(\lambda | \lambda_0) \cdot d\lambda = \alpha_2$$

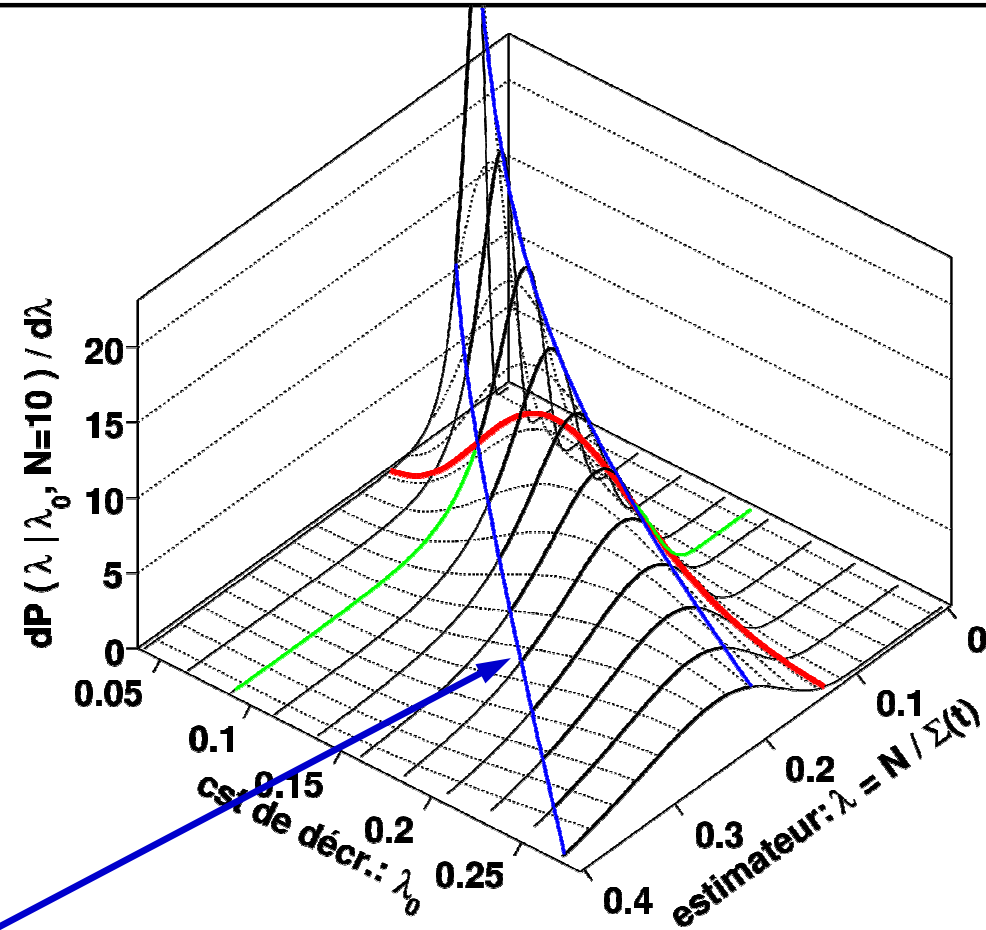
λ_A tel que $P(\lambda \geq \lambda_{\text{exp}} | \lambda_B) < \alpha_1$

solution de

$$\int_{\lambda_{\text{exp}}}^{\infty} f_N(\lambda | \lambda_0) \cdot d\lambda = \alpha_1$$

ceinture de confiance

intervalle de confiance pour l'ensemble des valeurs de l'estimateur



Résultat « classique »

observation (mesure)

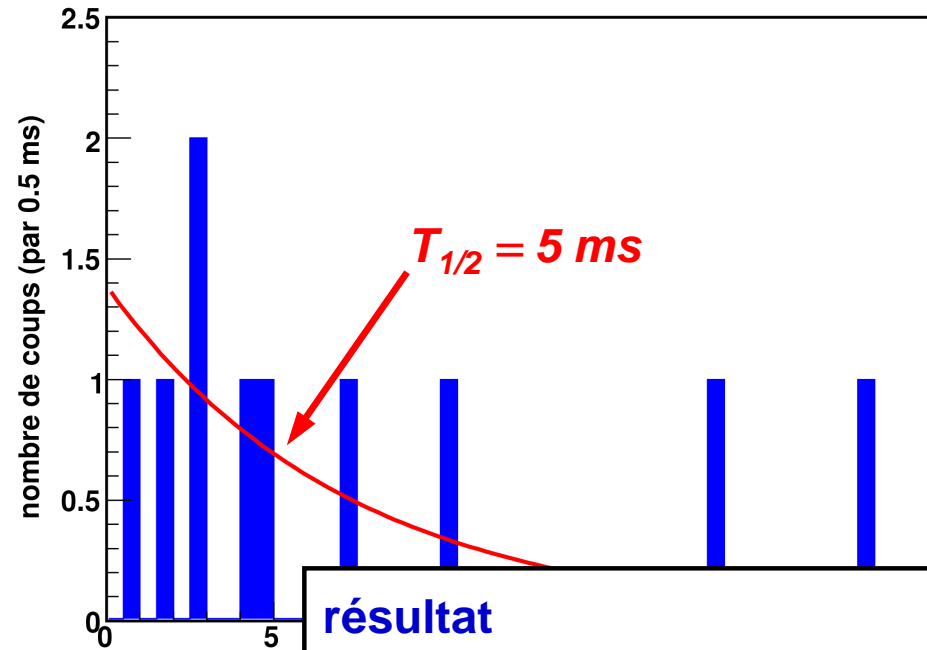
on a observé $N = 10$ noyaux (ceux de l'exemple précédent)

⇒ l'échantillon est l'ensemble des temps décroissance mesurés :

$$\{t_1, t_2, \dots, t_{N=10}\}$$

exemple :

i	t_i (ms)
1	0.905
2	4.550
3	10.278
4	4.456
5	18.310
6	22.741
7	2.856
8	1.590
9	2.940
10	7.331



⇒ que peut-on dire de $T_{1/2}$ (ou de λ) ?

résultat

estimateur $\lambda = 0,132$ $T_{1/2} = 5,268$

int. conf. ($\beta=68,3\%$)

$\lambda_B = 0,091$ $T_{1/2} = 7,647$

$\lambda_A = 0,173$ $T_{1/2} = 4,018$

$T_{1/2} = 5,3^{+2,4}_{-1,3}$

Approche bayesienne

f.d.p. dans l'approche bayesienne

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

hypothèse : toutes les valeurs du paramètre λ sont possibles, sans préférence « *à priori* » → $P(\lambda)$ uniforme

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

résolution

les bornes de l'intervalle de confiance sont les solutions de (résolution numérique) :

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

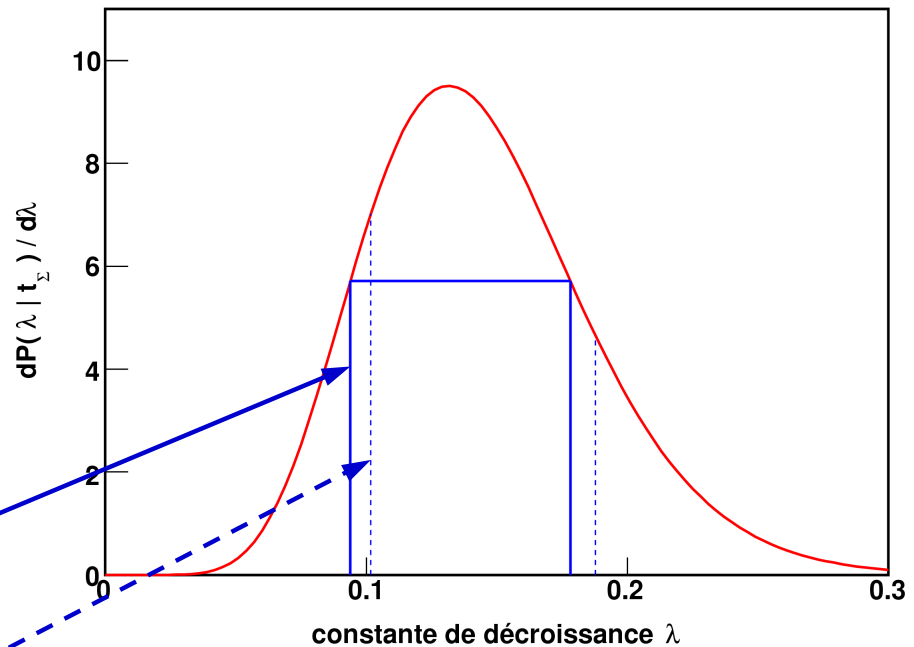
$\lambda = 0,132^{+0,047}_{-0,038} \rightarrow T_{1/2} = 5,3^{+2,1}_{-1,4}$

intervalle tel que

$P(\lambda < \lambda_A) = P(\lambda > \lambda_B) = \alpha \quad T_{1/2} = 5,3^{+1,5}_{-1,6}$

classique :

$T_{1/2} = 5,3^{+2,4}_{-1,3}$



int. bayésien le plus petit pour λ ,
mais pas pour $T_{1/2}$!

Autre choix du paramètre

on peut faire le choix d'un autre paramètre :
(temps caractéristique)

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

loi de décroissance : $f(t|\tau) = \frac{1}{\tau} \cdot e^{-(t/\tau)}$

f.d.p. pour le paramètre τ dans l'approche bayésienne

hypothèse : $P(\tau)$ uniforme

$\neq P(\lambda)$ uniforme !!!

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

résolution

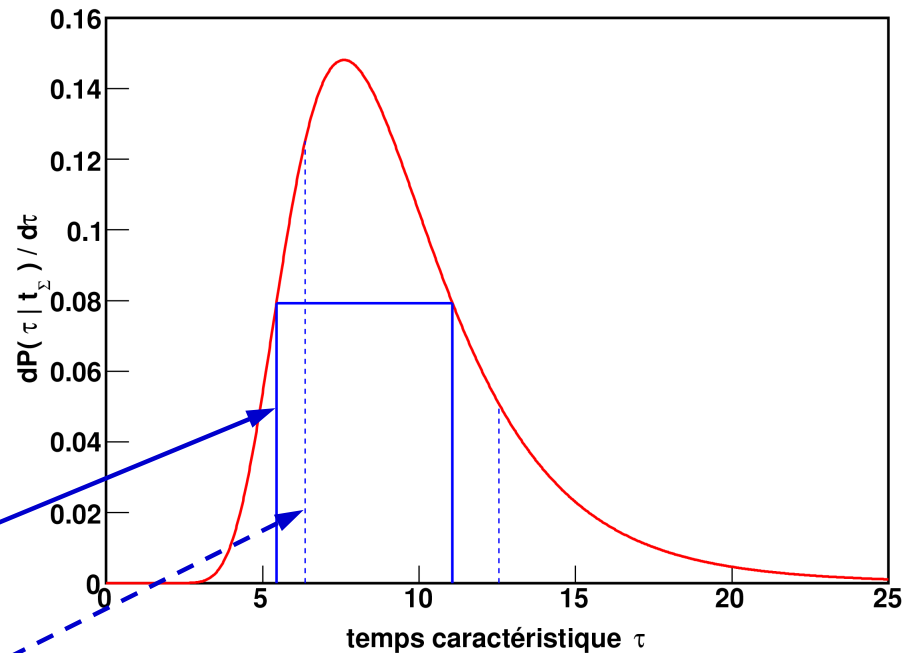
$\tau = 0,132^{+0,047}_{-0,038} \rightarrow T_{1/2} = 5,3^{+2,4}_{-1,5}$

intervalle tel que

$P(\tau < \tau_A) = P(\tau > \tau_B) = \alpha \quad T_{1/2} = 5,3^{+3,4}_{-0,9}$

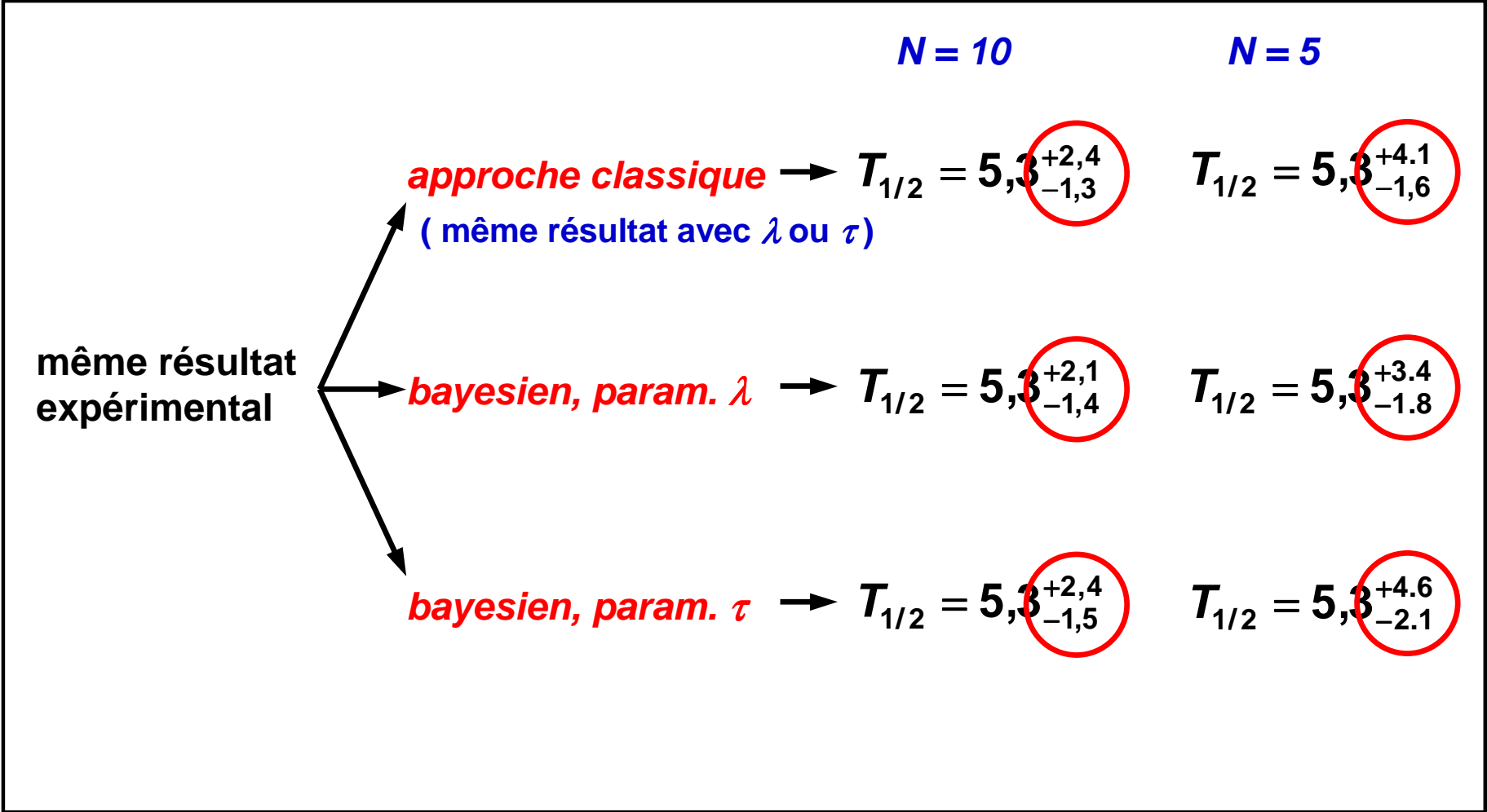
classique :

$T_{1/2} = 5,3^{+2,4}_{-1,3}$



résultat différent selon le choix du paramètre !

Durée de vie : récapitulatif



Durée de vie avec bruit de fond

dans la réalité le problème n'est jamais aussi simple :

- pour chaque noyau produit, on mesure la décroissance pendant un **temps fini**
- il y a généralement du « **bruit de fond** », c'est à dire des événements indésirables (qu'on ne sait pas distinguer du signal...)
- il y a du **temps mort** (un événement peut en cacher un autre...)
- ...

cas d'une mesure de bruit de durée de vie avec bruit de fond

pour chacun des N noyaux produits, on étudie la décroissance pendant un intervalle de temps fixe : T

pour une « mesure » sur un intervalle T on peut avoir :

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

problème

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

2 approches (traitement bayésien)

- analyse événement par événement
- analyse par histogramme

pour simplifier, on suppose la fréquence du fond connue ($\nu_b = 0,01 \text{ ms}^{-1}$)

Analyse par événements (1)

décomposition probabiliste

- N mesures (nombre de noyaux) sur un intervalle de temps T
- pour chaque mesure i , on observe n_i événements (signal vrai et/ou fond)

prob. d'observer la décr. : $P_d = (1 - e^{-\lambda T})$ $Q_d = 1 - P_d$

prob. d'observer n_b év^{ts} de fond : $B[n_b] = P(n_b | \nu_b)$ (loi de Poisson)

prob., pour une mesure, d'observer n événements

$$P[n] = Q_d \times B[n] + P_d \times B[n-1]$$

$$P[0] = Q_d \times B[0]$$

- probabilité d'observer les n_i événements à des temps $[t_1 ; t_1 + dt_1]$, $[t_2 ; t_2 + dt_2]$, ...

$$f_n(t_1, t_2, \dots, t_n) \times dt_1 \times dt_2 \dots \times dt_n$$

fonction de vraisemblance

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

Analyse par événements (2)

fonction de vraisemblance

exemple pour une mesure avec $n = 3$ (probabilité conditionnée...) : t_1, t_2, t_3

2 cas possibles (2^{ème} condition)

→ soit on n'a pas la décroissance (fond uniforme, temps équiprobables)

$$f_{0/3}(t_1, t_2, t_3) = \frac{1}{T^3}$$

→ soit on a la décroissance $f_T(t) = \frac{\lambda}{1 - e^{-\lambda T}} \cdot e^{-\lambda t}$ (normalisé sur $[0, T]$)

au temps t_1 : $f_{1/3}(t_1, t_2, t_3) = \frac{1}{T^2} \cdot f_T(t_1)$

au temps t_2 : $f_{2/3}(t_1, t_2, t_3) = \frac{1}{T^2} \cdot f_T(t_2)$

au temps t_3 : $f_{3/3}(t_1, t_2, t_3) = \frac{1}{T^2} \cdot f_T(t_3)$

donc la probabilité différentielle est

$$f_3(t_1, t_2, t_3) = \left(\frac{Q_d \times B[3]}{P[3]} \right) \frac{1}{T^3} + \left(\frac{P_d \times B[2]}{P[3]} \right) \frac{1}{T^2} \cdot [f_T(t_1) + f_T(t_2) + f_T(t_3)]$$

cond. 2

généralisation

$$f_n(t_1, \dots, t_n) = \left(\frac{Q_d \times B[n]}{P[n]} \right) \cdot \frac{1}{T^n} + \left(\frac{P_d \times B[n-1]}{P[n]} \right) \frac{1}{T^{n-1}} \cdot \left[\sum_{k=1}^n f_T(t_k) \right]$$

Analyse par événements (3)

maximum de vraisemblance

la fonction de vraisemblance se simplifie :

cond. 1 : nombre d'observations

cond. 2 : avec / sans décroissance

$$L(\text{obs}|\lambda) = \prod_{i=1}^N P[n_i] \times \left\{ \left(\frac{Q_d \times B[n_i]}{P[n_i]} \right) \cdot \frac{1}{T^3} + \left(\frac{P_d \times B[n_i - 1]}{P[n_i]} \right) \frac{1}{T^2} \cdot \frac{\lambda}{1 - e^{-\lambda T}} \left[\sum_{k=1}^{n_i} \exp(-\lambda t_k) \right] \right\}$$

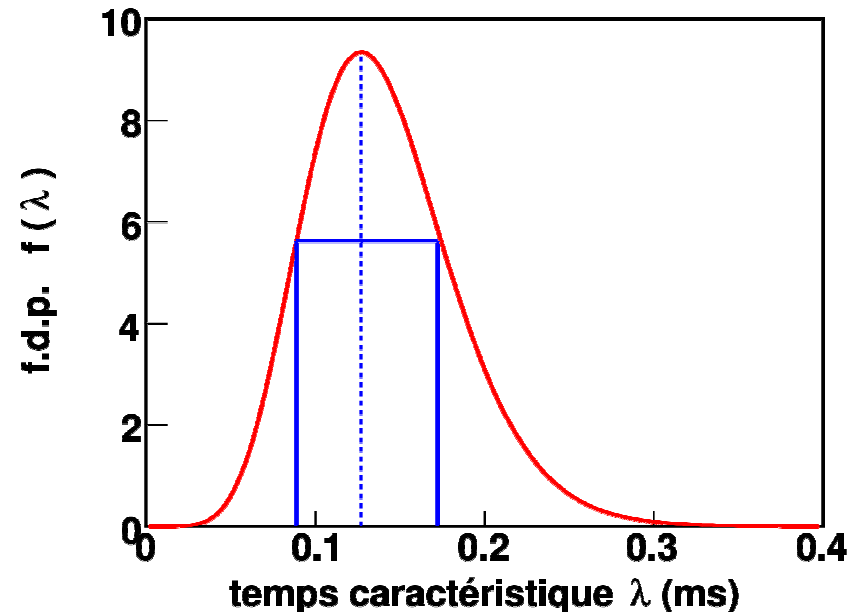
$$= \prod_{i=1}^N \left\{ (Q_d \times B[n_i]) \cdot \frac{1}{T^3} + B[n_i - 1] \cdot \frac{\lambda}{T^2} \cdot \left[\sum_{k=1}^{n_i} \exp(-\lambda t_k) \right] \right\}$$

l'**approche classique** ne permet pas de traiter le problème : il faudrait envisager toutes les possibilités conduisant au même estimateur (infinité) → inextricable !

solution bayésienne

(λ « à priori » uniforme) $T_{1/2} = 5,4^{+2,4}_{-1,5}$

(à partir des mêmes données de décr. que précédemment en ajoutant du fond)



Analyse d'histogramme (1)

l'ensemble des données est placé dans un histogramme

- discrétisation du problème
- perte d'information liée au « binning »

forme de l'histogramme

(nombre de coups attendus par *bin*)

pour une mesure

- décroissance : $\frac{dn_{dec}(t)}{dt} = \lambda \cdot e^{-\lambda t}$

- fond constant $\frac{dn_{fond}(t)}{dt} = \nu_b$

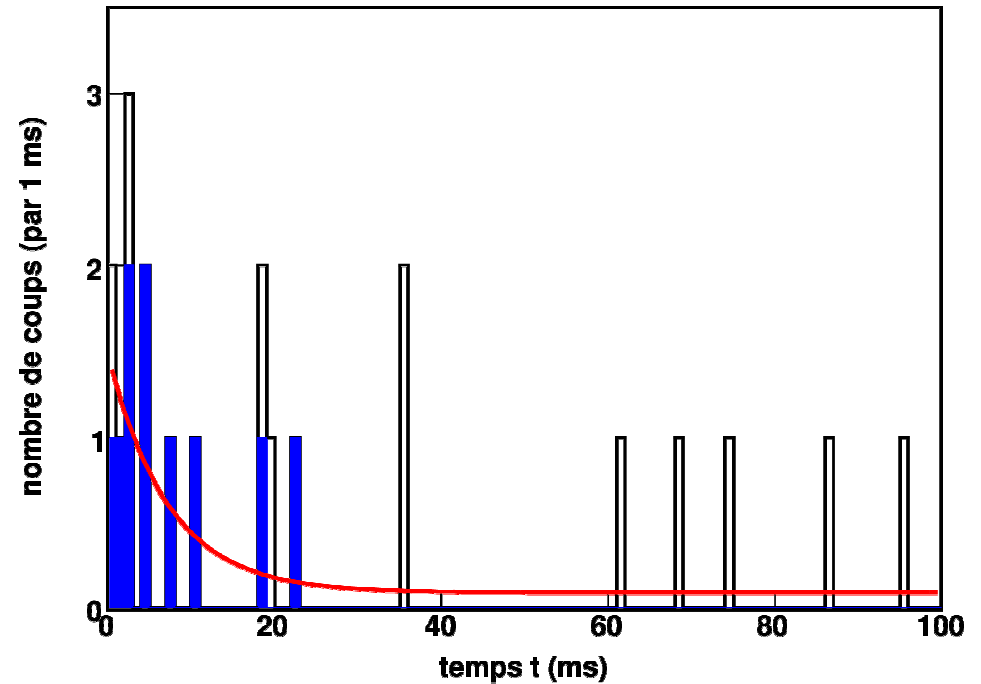
pour *N* mesures

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

« binning » Δt

nombre de coups attendus pour un *bin* $i = [t_i; t_i + \Delta t]$, avec $t_i = i \times \Delta t$

$$\mu_i = \int_{t_i}^{t_i + \Delta t} \frac{dn(t')}{dt'} dt' \approx \frac{dn(t_i)}{dt} \cdot \Delta t$$



Analyse d'histogramme (2)

maximum de vraisemblance

pour chaque *bin*, on détermine la probabilité d'observer n_i coups lorsqu'on en attend μ_i
 → loi de Poisson

fonction de vraisemblance
$$L(\text{obs}|\lambda) = \prod_i \frac{(\mu_i(\lambda))^{n_i}}{n_i!} \cdot e^{-\mu_i(\lambda)}$$

approche bayésienne

(en supposant les valeurs de λ équiprobables)

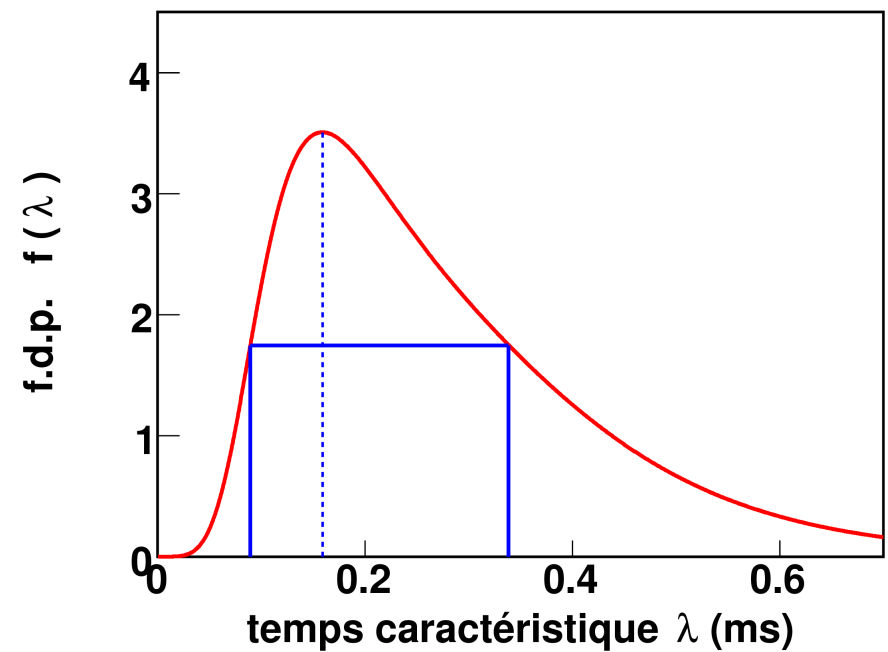
$$T_{1/2} = 4,4^{+3,4}_{-2,3}$$

même calcul sur τ
$$T_{1/2} = 4,4^{+3,1}_{-2,2}$$

traitement év. par év.
$$T_{1/2} = 5,4^{+2,4}_{-1,5}$$

→ le traitement par ensemble meilleur : 5 ms
 effet du « **bin** » plus petit ???

estimateur M.V. non consistant



Remarques

- un cas au départ très simple peut vite devenir plus complexe
- traitement bayésien incontournable
- en général, pas de solution analytique
(résolution numérique des problèmes, pas forcément très ardu)

illustration dans le cas de la radioactivité 2-protons
pas encore complètement réaliste

- rapport d'embranchement : à partir des événements dans un pic du spectre en énergie : « bruit de fond » dans le pic ?
- durée de vie : prise en compte du temps mort

« principe de réalité »

que gagne-t-on vraiment à rendre le problème trop complexe ?

- les physiciens continueront encore un peu à maltraiter les statistiques...

Conclusion

- propriétés de base du **calcul des probabilités** et des **variables aléatoires**
- construction d'un **estimateur** à partir d'un échantillon (mesures expérimentales)
- détermination de l'**incertitude** sur l'estimation
 - **comportement asymptotique** dans le cas des grands nombres
 - **intervalle de confiance**

illustré à partir d'exemples très simples, qui montrent que dans le cas des petits nombres d'observations

- que le choix de la méthode n'est pas neutre
- que les mêmes données peuvent conduire à des résultats différents

pour les faibles statistiques, un résultat se comprend en connaissance de la méthode utilisée

merci de votre attention

quelques références :

statistical methods in experimental physics, W.T. Eadie *et al.*

data reduction and error analysis for the physical sciences, P.R. Bevington

Particle Data Group (CERN) : <http://www.pdg.lbl.gov>

il n'y a plus qu'à...

Intégrales utiles

$$J_n(t) = \int_0^{\infty} \lambda^n \cdot \exp(-\lambda t) \cdot dt = \frac{n!}{t^{n+1}}$$

$$\int_0^{\infty} \frac{t}{n!} \cdot (\lambda t)^n \cdot \exp(-\lambda t) \cdot dt = 1$$

$$K_n(t) = \int_0^{\infty} \frac{1}{\tau^n} \cdot \exp\left(-\frac{t}{\tau}\right) \cdot d\tau = \frac{(n-2)!}{t^{n-1}}$$

$$\int_0^{\infty} \frac{1}{t \cdot (n-2)!} \cdot \left(\frac{t}{\tau}\right)^n \cdot \exp\left(-\frac{t}{\tau}\right) \cdot d\tau = 1$$