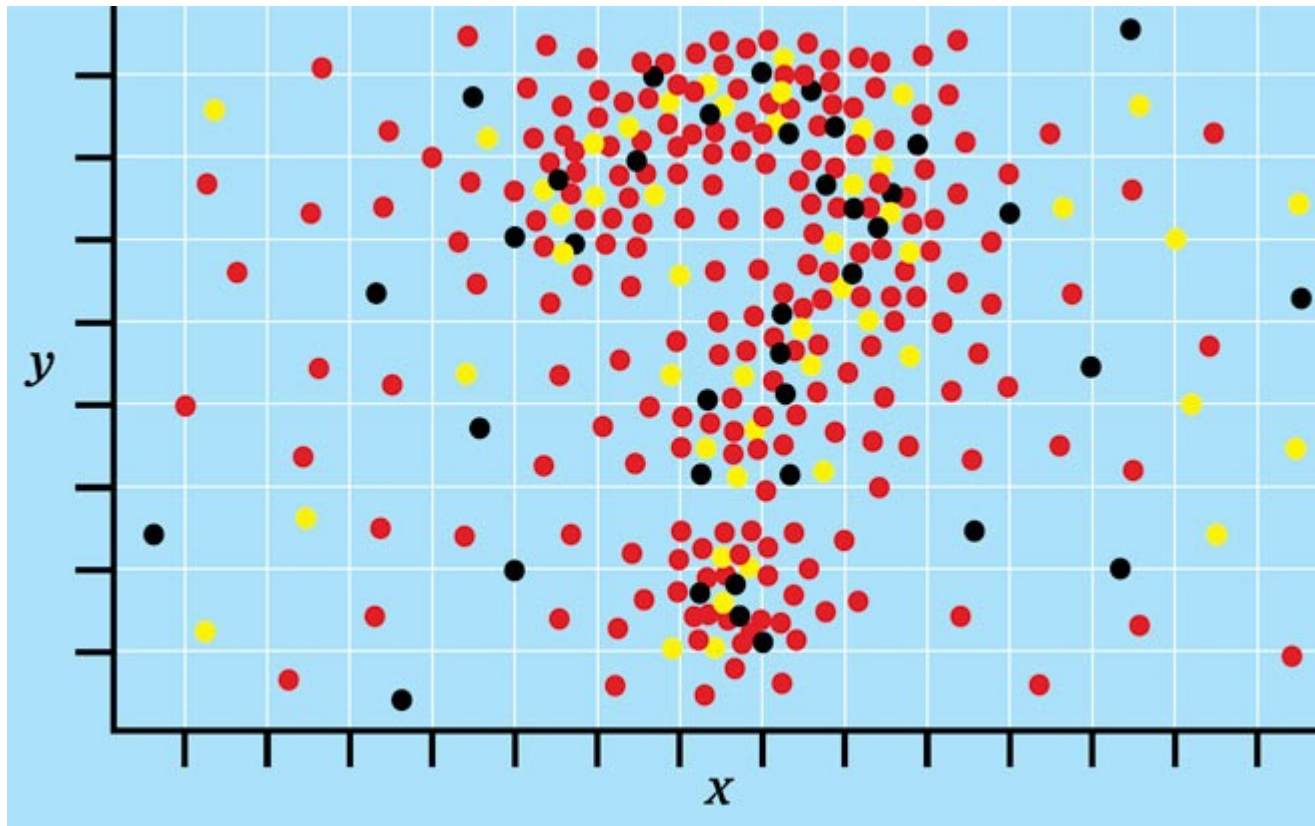


Intro à la statistique

cours TP3/TP4

Philippe Mermod, 10 décembre 2012



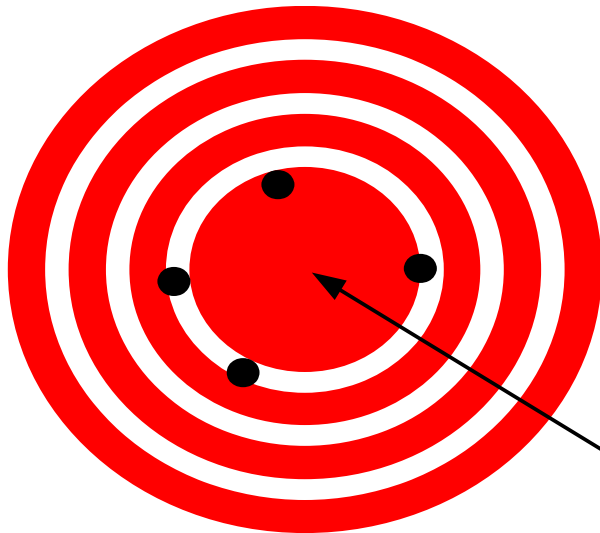
La statistique, à quoi ça sert?

- Situation: on veut **mesurer** une quantité pour la comparer avec ce que prédisent les modèles
 - Exemple: durée de vie du muon
 - En général, le résultat d'une mesure n'est pas très conclusif; donc, on la **répète** plusieurs fois (**échantillon**). Puis on pose les questions suivantes:
 - Comment extraire la quantité la plus précise possible de cette série de mesures?
 - Comment décider si cette mesure est consistante avec la théorie, ou avec d'autres mesures?
 - Pour y répondre, on définit les concepts suivants:
 - Distributions, corrélations, paramètres, erreurs...
- **THÉORIE/OUTILS STATISTIQUES**

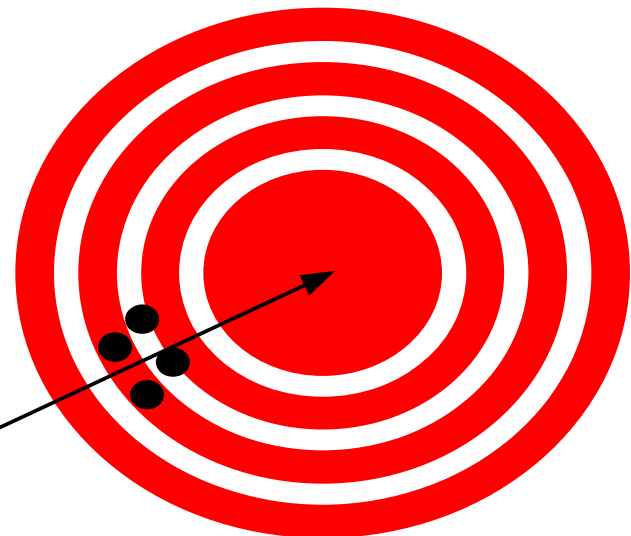
Erreurs statistiques et systématiques

Intrument bien réglé,
Statistiques insuffisantes
→ répéter la mesure un
plus grand nombre de fois;
la moyenne va converger

Intrument mal réglé,
Statistiques suffisantes
→ la mesure est biaisée
(erreur systématique)



“vraie”
valeur



Distributions, et autres définitions

- Événement: on effectue une mesure
- Variable aléatoire: on associe une valeur x_A à la variable (quantité mesurée) x pour l'événement \mathbf{A} .
- Distribution: quand on mesure x plein de fois, on remarque qu'elle tend à prendre certaines valeurs plus souvent que d'autres. On définit la distribution (ou fonction de probabilité) de x comme la probabilité qu'une mesure trouve la valeur x :
$$f(x) = P(x_A = x)$$
 - Si x est continue, on définit la probabilité qu'une mesure de x se trouve dans un intervalle:
$$P(x < x_A < x + dx) = f(x)dx$$
 - $f(x)$ est une densité de probabilité
 - On a toujours : $\int f(x)dx = 1$

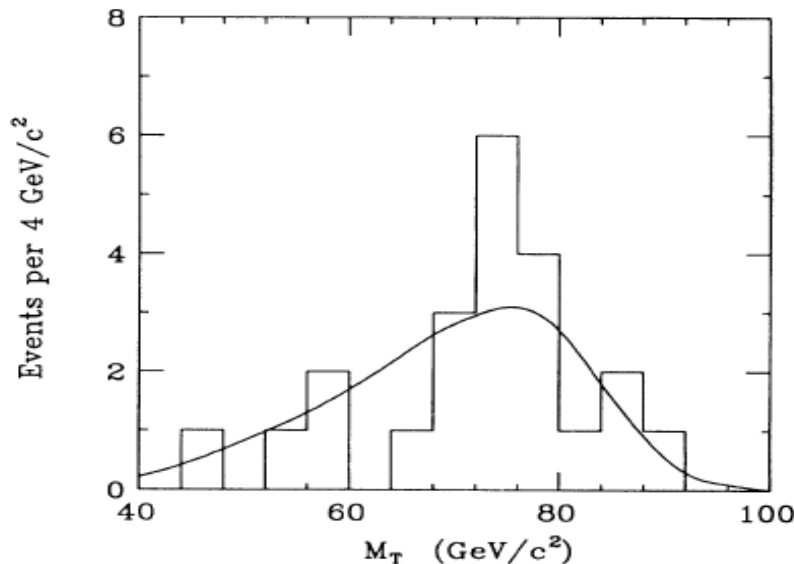
Échantillonnage: Pour représenter l'ensemble des mesures de x on utilise souvent des histogrammes

- Quand on a très peu de mesures, l'histogramme ne ressemble à rien (“peu de statistiques”)
- Quand on fait un plus grand nombre de mesures, l'histogramme tend à prendre la forme de la fonction de probabilité (“hautes statistiques”)
- Pour $n \rightarrow \infty$ on obtient la “vraie” densité de probabilité
- Exemple: **production du boson W** mesurée par CDF

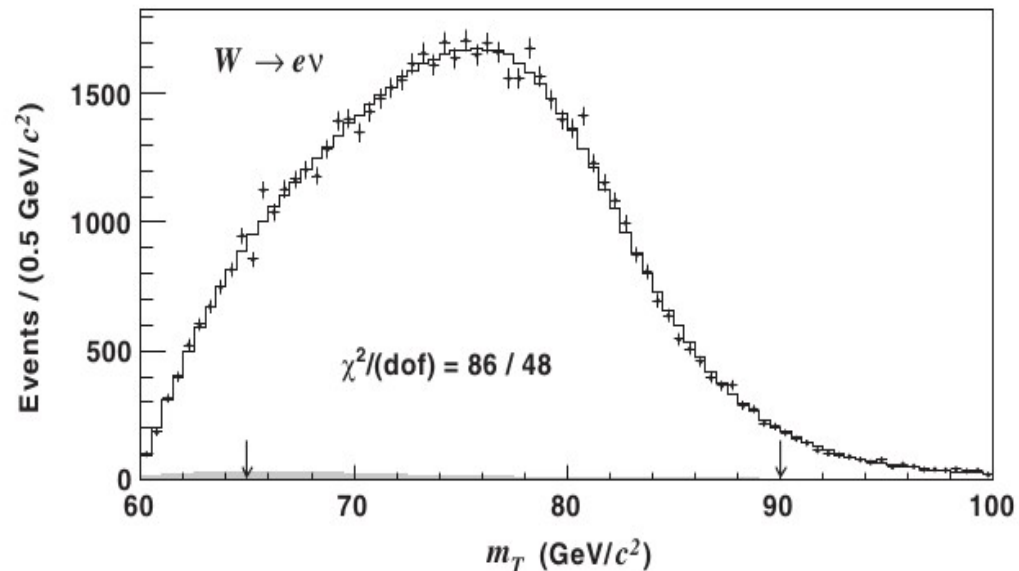
$$f(x) = \frac{N(x)}{n\Delta x}$$

$N(x)$: nombre de mesures dans le bin autour de x
 n : nombre total de mesures
 Δx : largeur du bin

En 1989



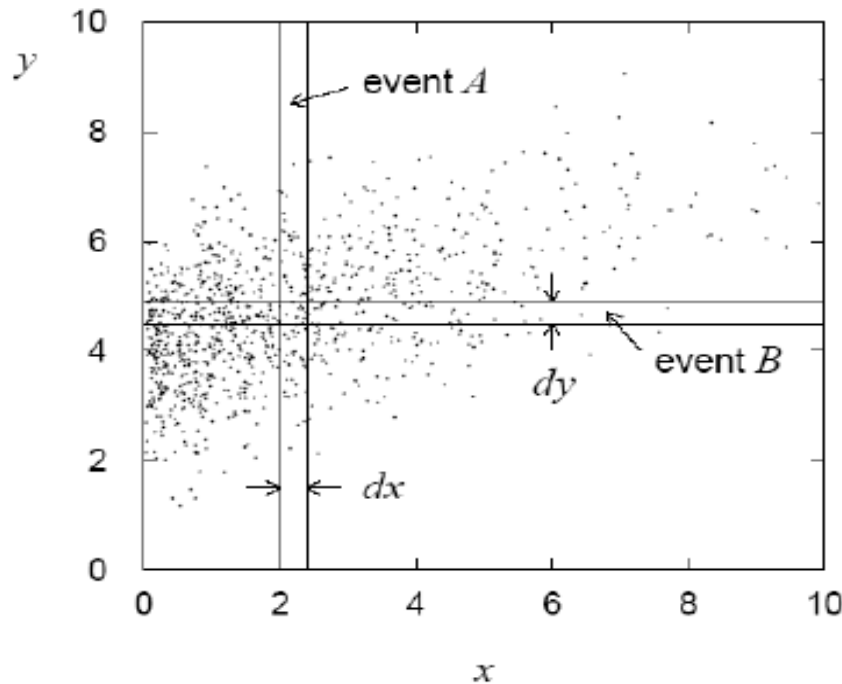
En 2007



Distributions à deux dimensions

- Considérons deux variables aléatoires \mathbf{x} et \mathbf{y} (on mesure deux quantités à la fois):

$$P(x_A < x < x_A + dx \wedge y_A < y < y_A + dy) = P(A \wedge B) = f(x, y) dx dy$$



densité de probabilité
jointe

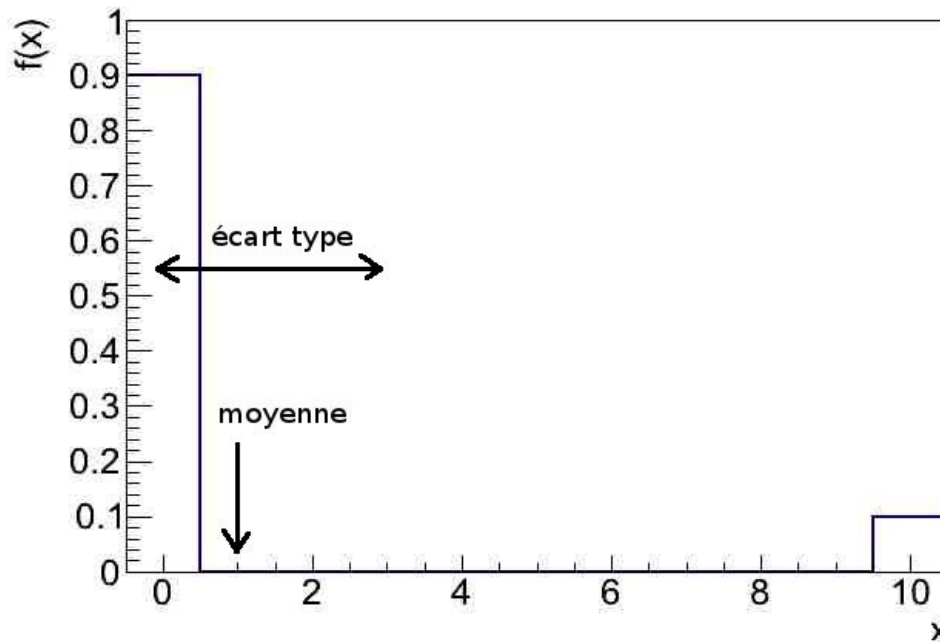
- Variables indépendantes:

$$P(A \wedge B) = P(A)P(B) = f(x)f(y) dx dy$$

- On peut généraliser à N dimensions : $f(x_1, x_2, \dots, x_N)$

- Espérance E : soit $g(x)$ une fonction de la variable aléatoire x , qui a la densité de probabilité $f(x)$
 - L'espérance de $g(x)$: $E(g(x)) = \int g(x)f(x)dx$ (on appelle ça aussi "convolution")
 - E est un opérateur linéaire: $E(ag(x) + bh(x)) = aE(g(x)) + bE(h(x))$
- Moyenne μ : c'est l'espérance de x : $\mu = E(x) = \int xf(x)dx$
- Variance V : c'est l'espérance de $(x-\mu)^2$: $V(x) = \int (x-\mu)^2 f(x)dx$
 - $V = \sigma^2$ définit l'écart type (déviation standard)

• Exemple:



$$\mu = 1 ; \sigma = 3$$

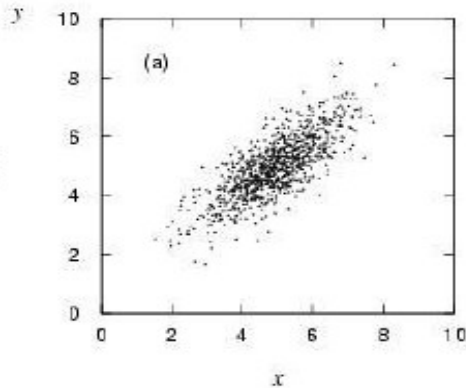
Qu'est-ce que cette distribution pourrait bien représenter?

- Covariance cov pour 2 variables aléatoires x et y :

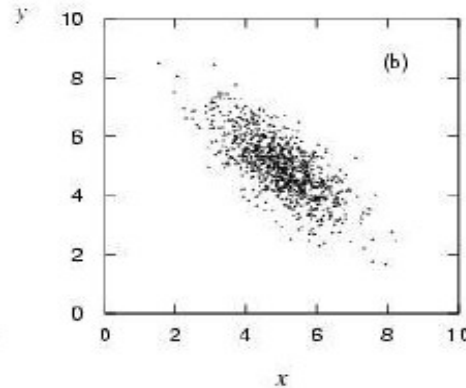
$$\text{cov}(x, y) = E((x - \mu_x)(y - \mu_y)) = \int \int (x - \mu_x)(y - \mu_y) f(x, y) dx dy$$

- Coefficient de corrélation ρ : $\rho_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad -1 \leq \rho_{xy} \leq 1$
- Si variables indépendantes $\rightarrow \rho = 0$
(l'inverse n'est pas toujours vrai)

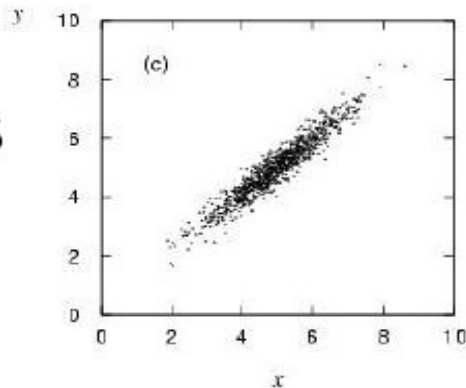
$$\rho = 0.75$$



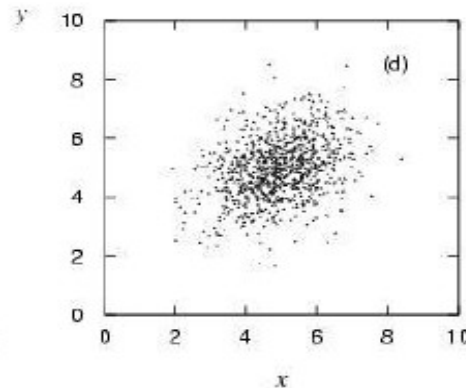
$$\rho = -0.75$$



$$\rho = 0.95$$



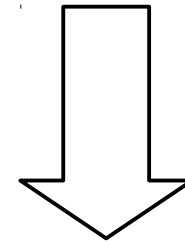
$$\rho = 0.25$$



Situation fréquente: on mesure plusieurs quantités

(x_1, x_2, \dots, x_N) (ex: température, pression, conductivité thermique,...). On connaît leurs moyennes μ_i et les covariances $V_{ij} = \text{cov}(x_i, x_j)$

- Quelle est la variance d'une variable aléatoire y qui est une fonction de tous ces x_i , c'à-d $y(x_1, x_2, \dots, x_N)$?
 - On pourrait obtenir la distribution de probabilité de y à partir de la densité de probabilité jointe $f(x_1, x_2, \dots, x_N)$, mais c'est souvent difficile
 - On peut aussi évaluer $V(y)$ en faisant un développement en série autour de μ_i pour obtenir la formule de propagation d'erreur



Formule de propagation d'erreur

Variance de la variable aléatoire combinée

$$\sigma_y^2 \approx \sum_{i,j=1}^n \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Somme sur toutes les combinaisons

Matrice de covariances

dérivées autour de la moyenne



Les corrélations sont importantes!
(contenues dans V_{ij})

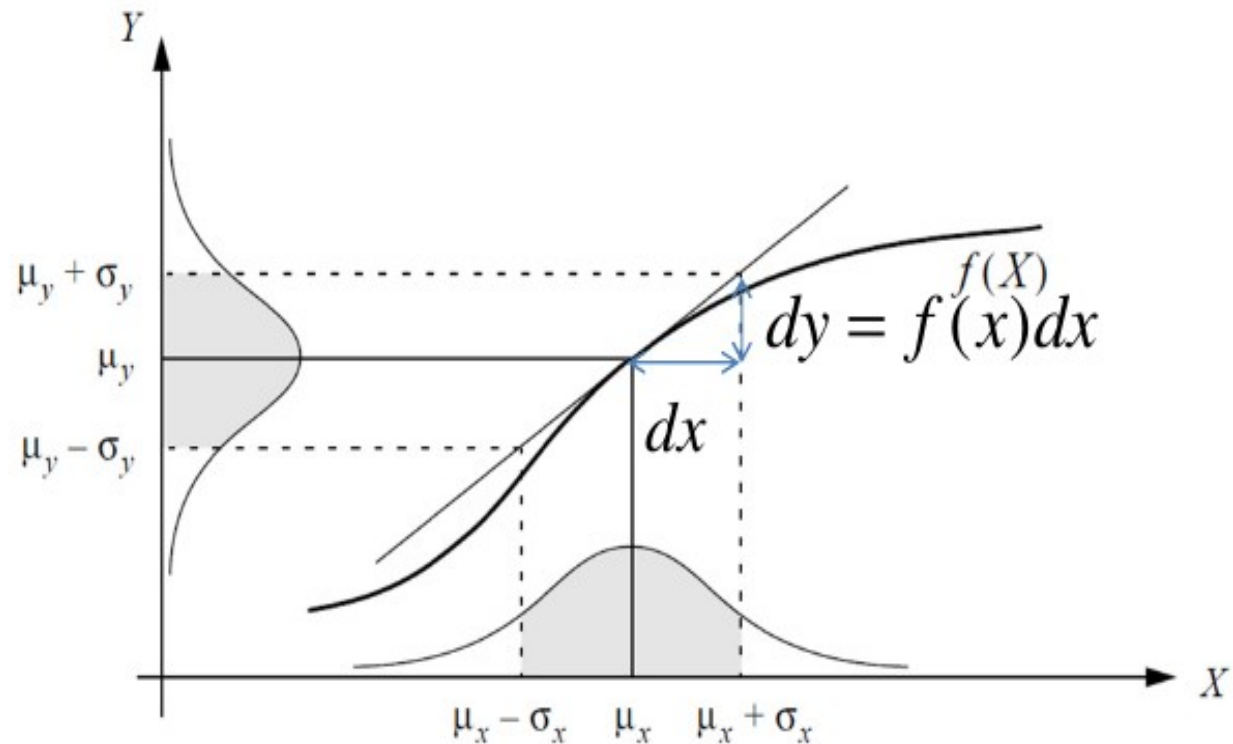
Propagation d'erreur – cas spéciaux

- Soit y une fonction des **variables indépendantes** mesurées x_i :

$$\sigma_y^2 = \left. \frac{\partial y}{\partial x_1} \right|_{\mu_1}^2 \sigma_{x_1}^2 + \left. \frac{\partial y}{\partial x_2} \right|_{\mu_2}^2 \sigma_{x_2}^2 + \dots$$

- Soit y une fonction d'**une seule variable** mesurée x :

$$\sigma_y^2 = \left. \frac{\partial y}{\partial x} \right|_{\mu_x}^2 \sigma_x^2$$



Exemples de propagation d'erreur

- $y(x) = ax + b \quad \rightarrow \quad \sigma_y = a\sigma_x$

- $y(x) = x^2 \quad \rightarrow \quad \sigma_y = 2x\sigma_x$

- $y(x_1, x_2) = x_1 + x_2 \quad \rightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\text{COV}[x_1, x_2]$

- $y(x_1, x_2) = x_1 x_2 \quad \rightarrow \quad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + 2\frac{\text{COV}[x_1, x_2]}{x_1 x_2}$

→ **Si x_1 et x_2 sont non corrélées:** $\text{cov}(x_1, x_2) = 0$, et il faut:

→ Additionner les erreurs en quadrature pour obtenir l'erreur sur la somme

→ Additionner les erreurs relatives en quadrature pour obtenir l'erreur relative sur le produit

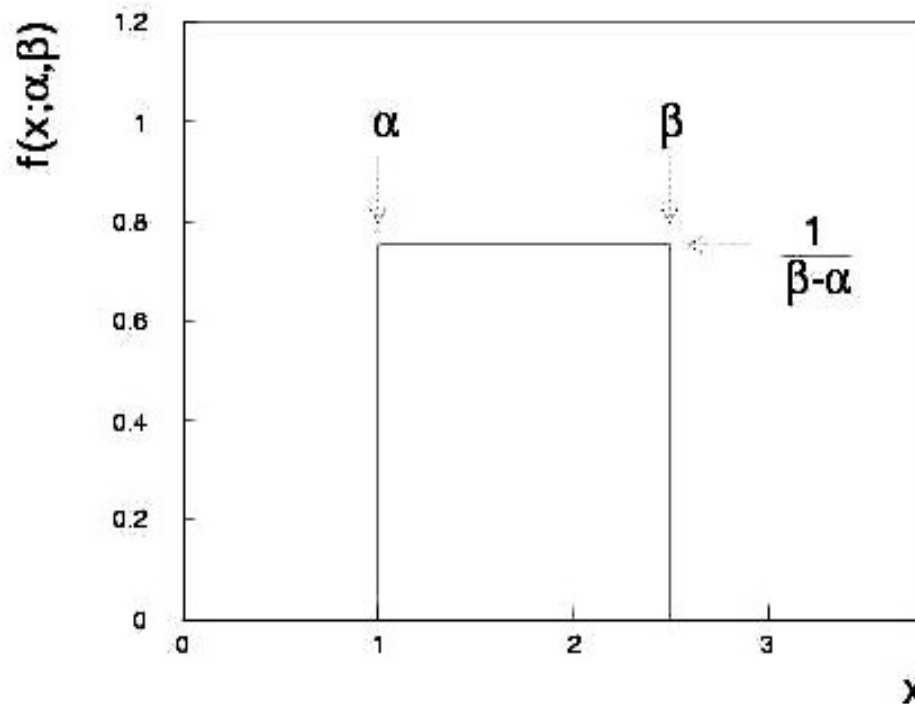
Exemples de distributions qu'on rencontre fréquemment en physique

- Uniforme
- Binomiale
- Poisson
- Exponentielle
- Normale (gaussienne)
- Chi-carré (χ^2)
- Landau, Cauchy, Breit-Wigner, beta, gamma, Student's t , etc...

Distribution uniforme

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta - \alpha} & \text{Si } \alpha < x < \beta \\ 0 & \text{Sinon} \end{cases}$$

- α et β sont des paramètres de la fonction de probabilité
- $\mu = (\alpha + \beta) / 2$
- $V = (\beta - \alpha)^2 / 12$



Distribution binomiale - 1

- Considérons un essai qui n'a que deux résultats possibles:
 - succès : probabilité p
 - échec : probabilité $1-p$
- La variable aléatoire n est le nombre de succès obtenus après N essais indépendants (deux paramètres: N et p)
 - Probabilité d'une séquence donnée avec n succès:

$$p^n (1 - p)^{N-n}$$

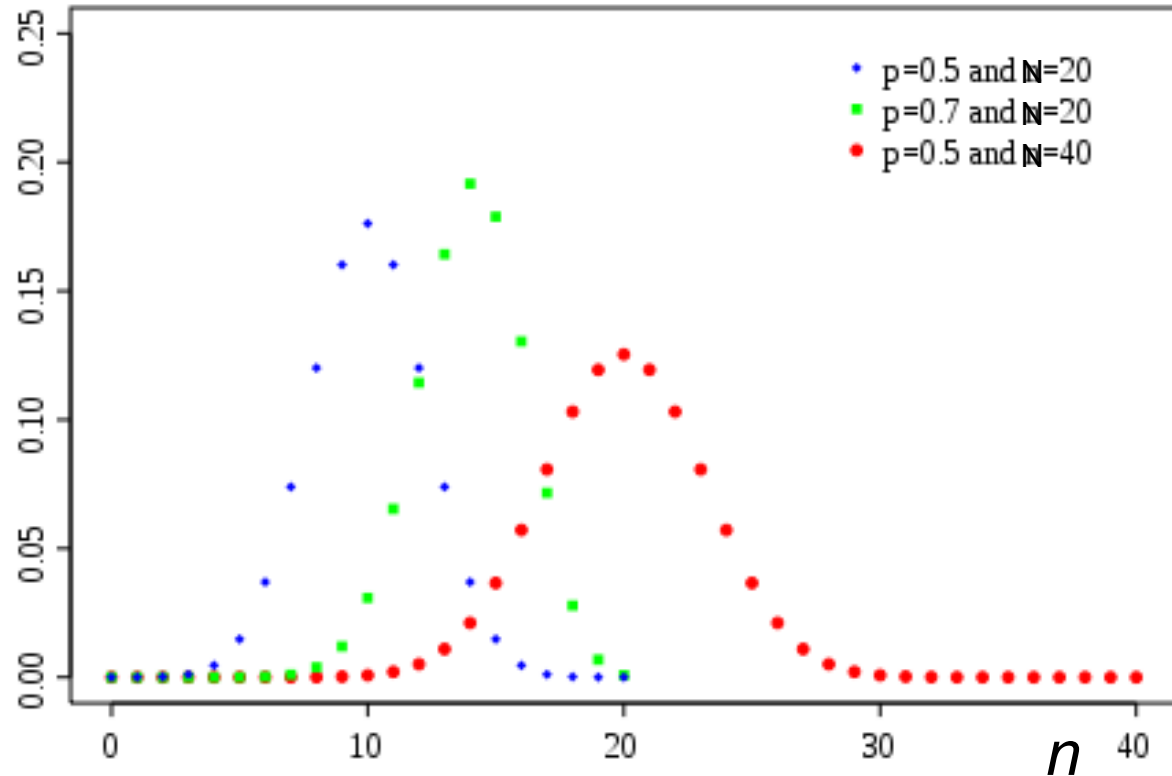
- Mais on a $\frac{N!}{n!(N-n)!}$ séquences possibles qui

donnent n succès (ex avec $N=5$: sésésé, ssésés, sééss, ...)

$$\longrightarrow f(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1 - p)^{N-n}$$

Distribution binomiale - 2

- $\mu = Np$
- $V = Np(1-p)$



- Exemple en physique des particules: parmi N désintégrations du boson W , nombre d'événements avec electron+neutrino (alors, p est le rapport de branchement)
- En fait, on trouve la binomiale un peu partout, sans vraiment s'en rendre compte...

Distribution de Poisson

On prend la distribution binomiale et on considère ce qui se passe lorsque $N \rightarrow \infty$ et $p \rightarrow 0$

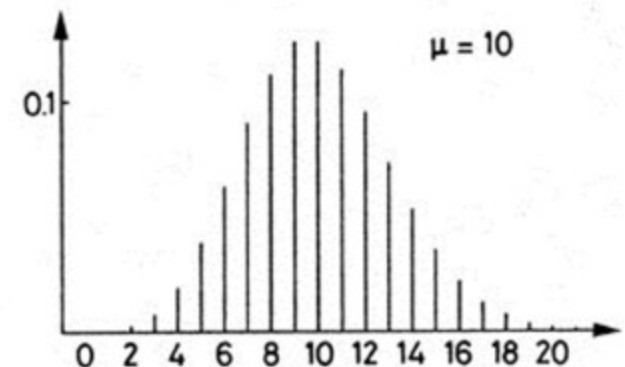
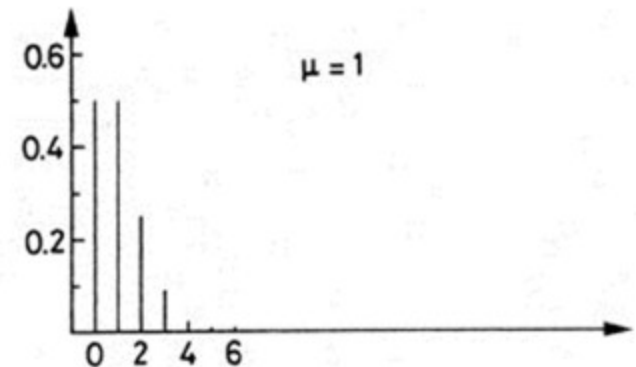
- Alors, on a $Np \rightarrow \nu$
- La probabilité d'observer n succès dans un intervalle de temps fixe suit la distribution de Poisson:

$$f(n; \nu) = \frac{\nu^n}{n!} e^{-\nu}$$

- Le taux de succès (nombre moyen de succès par intervalle de temps) est égal à la moyenne de la distribution pour cet intervalle:

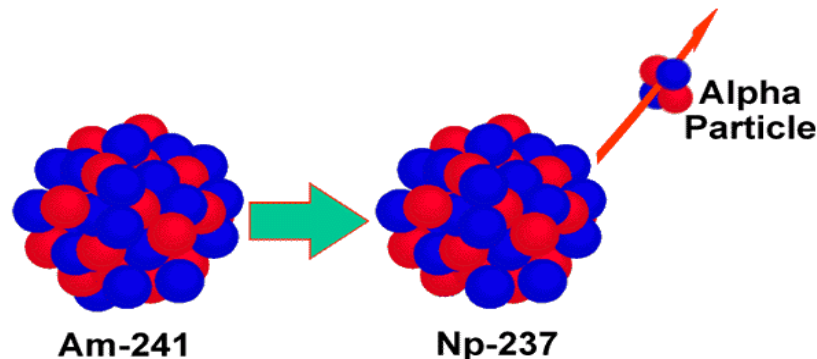
- $\mu = \nu$
- $\sigma = \sqrt{\nu}$

- Lorsque ν est $> \sim 5$, la distribution de Poisson tend vers la distribution normale (voir plus loin)



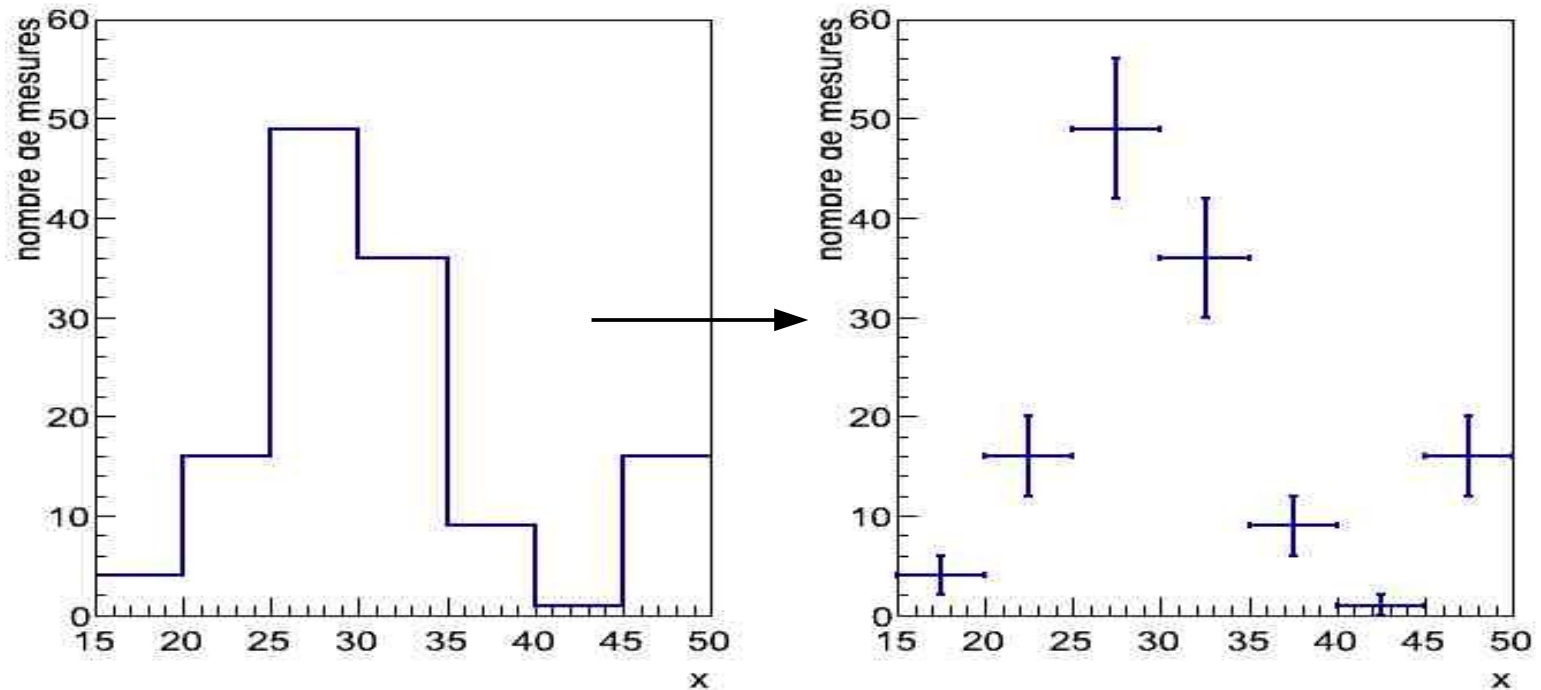
Distribution de Poisson – exemple 1

- **Désintégration radioactive**, par exemple, un matériau émet en moyenne 2 particules alpha par minute
 - Quelle est la valeur de ν si on considère le nombre de particules émises en deux minutes?
 - Quelle est la probabilité qu'une expérience unique, sur deux minutes, observe un nombre de particules alpha égal à la moyenne?
 - Dans ce cas, quelle est l'incertitude sur le nombre observé de particules?



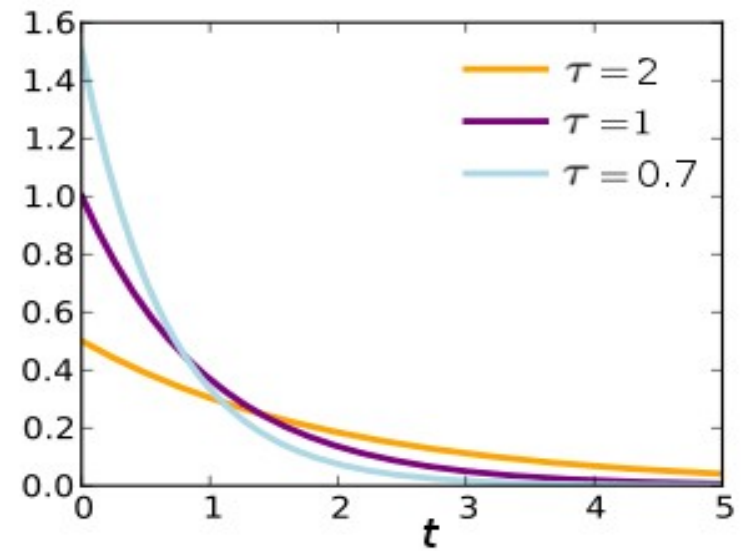
Distribution de Poisson – exemple 2

- On remplit un histogramme avec des données.
 - Le nombre d'entrées dans un bin donné suit en général une distribution de Poisson.
 - Admettons qu'on prenne le nombre d'entrées comme estimateur de la moyenne (voir plus loin). Quelle est l'erreur sur le nombre d'entrées?



Distribution exponentielle

- $t > 0$
 - $E(t) = \tau$
 - $V(t) = \tau^2$
- $$f(t; \tau) = \frac{1}{\tau} e^{-t/\tau}$$



- **Exemple:** désintégration (t =temps)
 - Le nombre de désintégrations par seconde suit une distribution de Poisson
 - Le temps de vie d'une particule (compté à partir de n'importe quel instant) suit une distribution exponentielle. La moyenne de cette distribution est le temps de vie moyen τ
(à ne pas confondre avec la demi-vie, qui est le temps moyen pour que la moitié des particules se désintègrent $t_{1/2} = \tau \ln(2)$)
 - Perte de mémoire:
(unique à l'exponentielle) $f(t - t_0 | t \geq t_0) = f(t)$

Distribution Normale (ou gaussienne) - 1

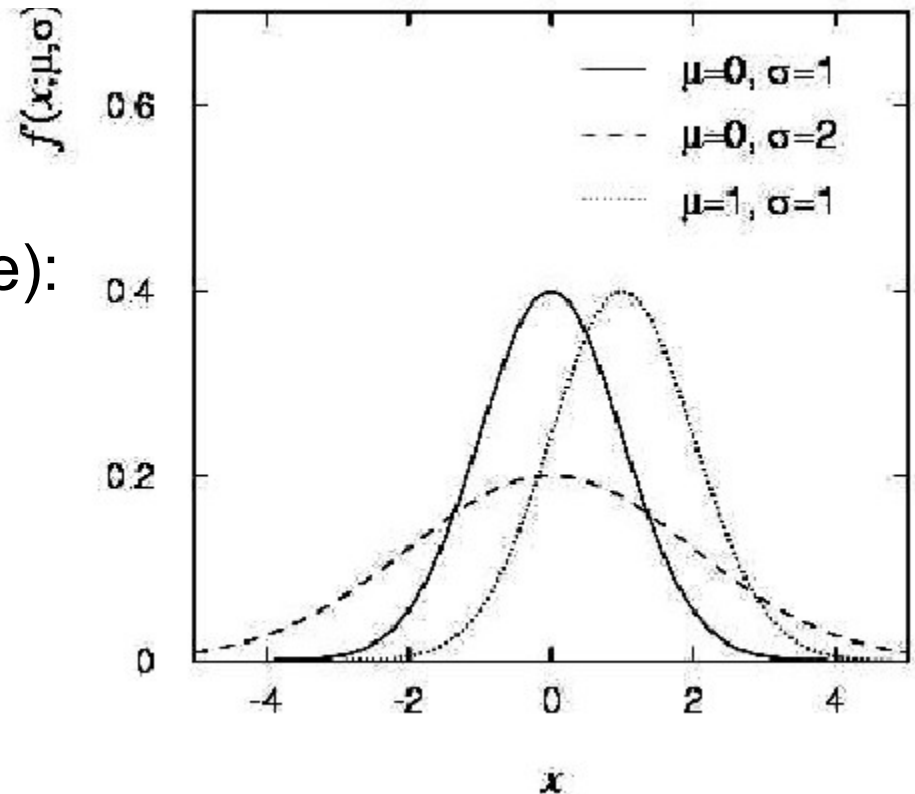
- On rencontre cette distribution un peu partout, nous verrons bientôt pourquoi (théorème central limite)

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

- Moyenne μ , variance σ^2
- Cas spécial: $\mu=0$, $\sigma=1$
(distribution normale réduite):

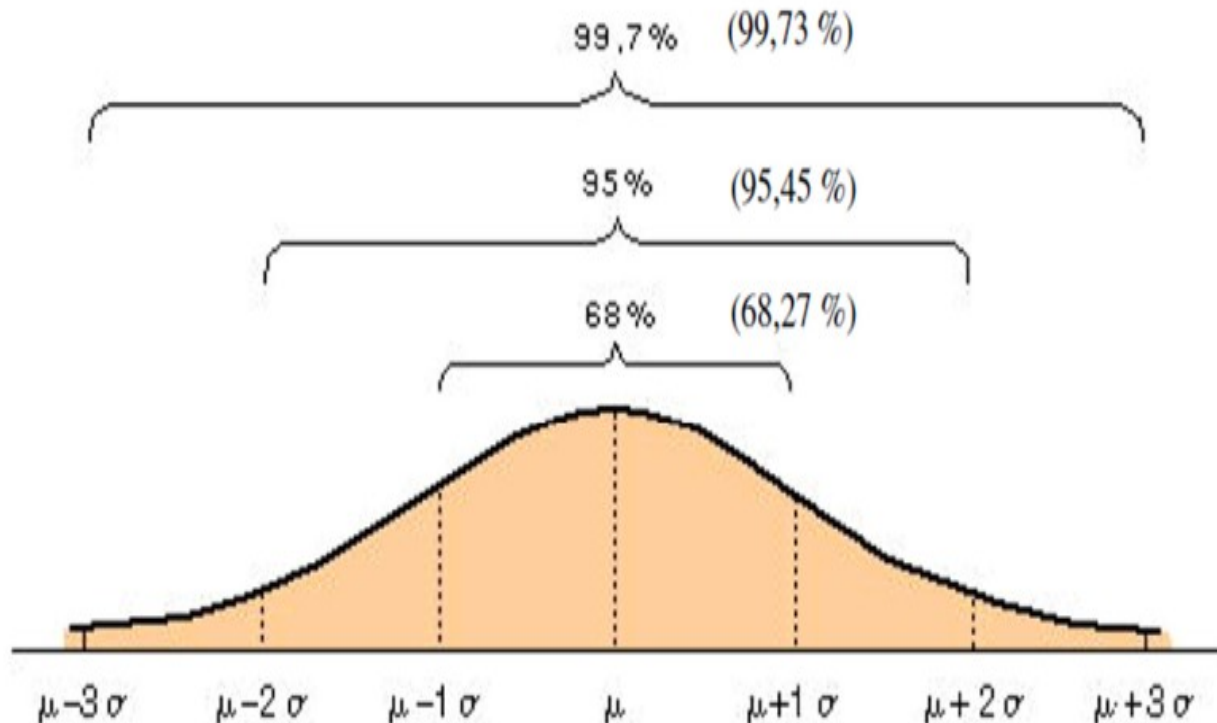
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

- $P(-1 < x < 1) = 0.6827$
- $P(-2 < x < 2) = 0.9545$
- $P(-3 < x < 3) = 0.9973$



Distribution Normale (ou gaussienne) – 2

- On fait une **mesure unique** de la quantité x .
- **Convention:** l'erreur sur la mesure, Δx , est définie telle que la probabilité que la vraie valeur de x soit dans l'intervalle $x \pm \Delta x$ est 68.27% (on parle d'un “écart standard”)
- Cela correspond à **$\Delta x = \sigma$** si on admet que la distribution de probabilité de x est approximativement normale (nous verrons bientôt pourquoi).



Distribution Normale (ou gaussienne) – 3

Théorème central limite: Une somme d'un grand nombre de variables aléatoires indépendantes (quelles qu'elles soient) suit une distribution normale

- Cela découle du fait que la moyenne de la somme est la somme des moyennes et la variance de la somme est la somme des variances

$$y = \sum_{i=1}^n x_i \quad \begin{array}{l} x_i \text{ indépendantes; dans la limite } n \rightarrow \infty, \\ y \text{ est une distribution normale avec} \end{array}$$

$$E[y] = \sum_{i=1}^n \mu_i \quad V[y] = \sum_{i=1}^n \sigma_i^2$$

- Le résultat précis d'une mesure unique dépend souvent de la somme de plein de petites contributions expérimentales
 - Une mesure individuelle suit souvent (mais pas toujours!) une distribution approximativement normale
 - erreur sur la mesure

Distribution χ^2

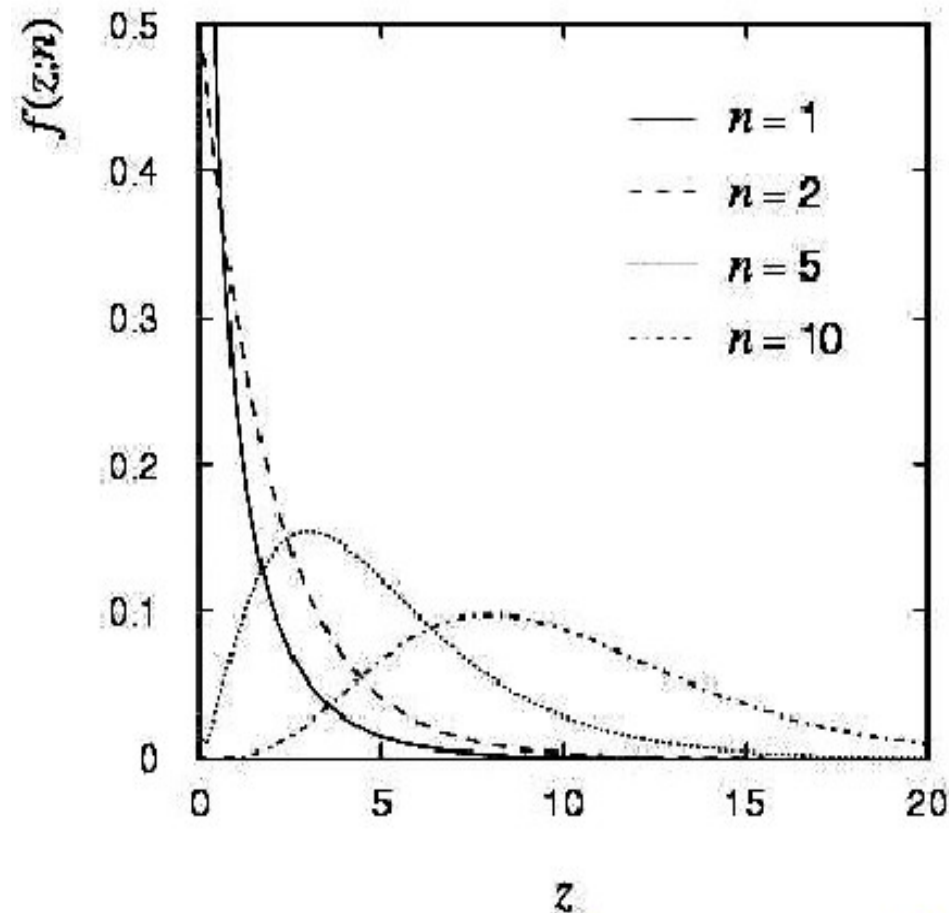
$$f(z; n) = \frac{1}{2^{n/2} \Gamma(n/2)} z^{n/2-1} e^{-z/2}$$

- $n=1,2,\dots$ est le nombre de degrés de liberté (aussi dénoté *Ndof*)
- $E(z) = n$
- $V(z) = 2n$
- Pour des variables x_1, \dots, x_n

qui ont des moyennes μ_i
et variances σ_i , la variable

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

suit une distribution χ^2 avec
 n degrés de liberté.



Estimation de paramètres

Situation fréquente: on mesure une quantité. On connaît la forme de sa distribution, mais pas la valeur de ses paramètres.

- Comment estimer les paramètres à partir d'un échantillon (histogramme à nombre d'entrées limité)?
- Comment évaluer l'erreur sur l'estimation du paramètre?
- **Exemple:** boson **Z** au LEP

Distribution Breit-Wigner

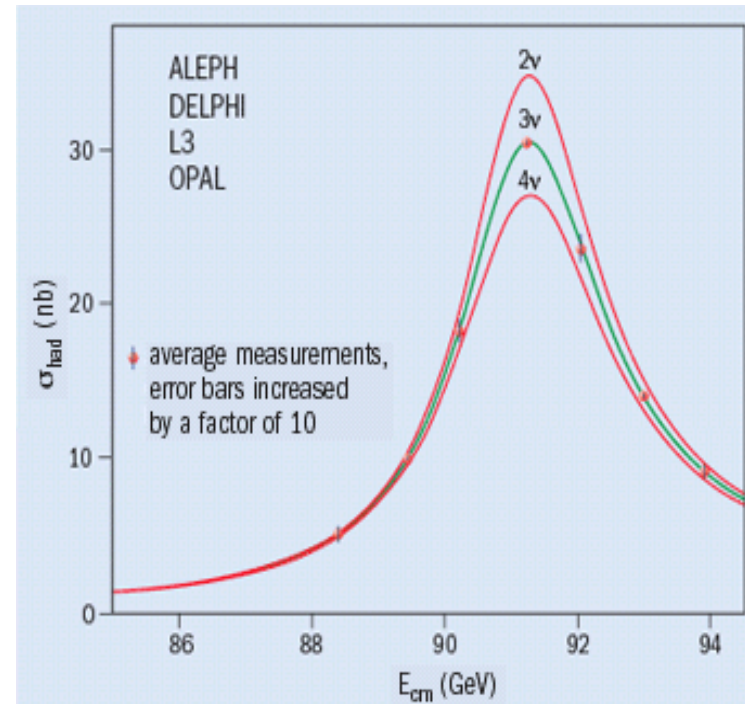
$$f(x; x_0, \Gamma) = \frac{1}{\pi} \frac{\Gamma/2}{(x - x_0)^2 + \Gamma^2/4}$$

On peut estimer:

$x_0 \rightarrow$ masse du Z

$\Gamma \rightarrow$ nombre d'espèces de neutrinos

$1/\Gamma \rightarrow$ durée de vie du Z

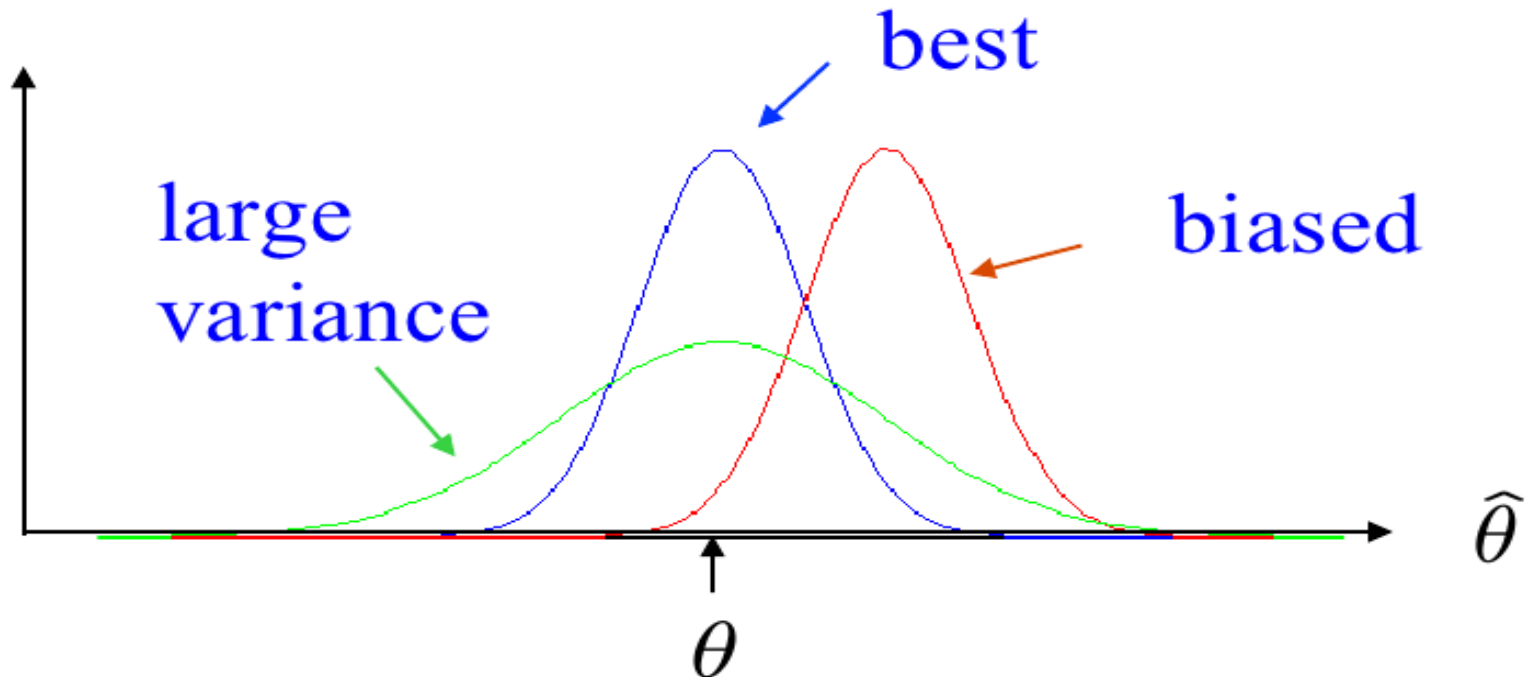


Estimation de paramètres: définitions – 1

- Paramètre θ : faire varier θ fait varier la fonction de probabilité:
$$f(x; \theta)$$
- Échantillon (x_1, \dots, x_n) : ensemble des mesures de la variable aléatoire x
- Estimateur $\hat{\theta}(x_1, \dots, x_n)$: fonction des données de l'échantillon, qui estime le paramètre θ
 - L'estimateur est lui-même une variable aléatoire, dont la moyenne, idéalement, est égale à θ
- Convergence: quand on augmente n , $\hat{\theta}$ se rapproche de θ
- Le théorème central limite assure que si n est grand ($> \sim 30$), $\hat{\theta}$ suit une distribution normale

Estimation de paramètres: définitions – 2

- Estimateur biaisé: converge vers une “fausse” valeur
→ (erreur systématique)
- Estimateur efficace: faible variance
- Estimateur robuste: efficace indépendamment de la distribution



Estimateurs pour la moyenne et la variance

- On suppose que chaque mesure d'échantillon (x_1, \dots, x_n) a la même erreur (inconnue) $\sigma_i = \sigma$
- Paramètre à estimer: $\theta = \mu = E(x)$

– Estimateur:
(moyenne d'échantillon)

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

– On trouve: $E(\hat{\mu}) - \mu = 0$ (pas de bias)

– On trouve: $V(\hat{\mu}) = \frac{\sigma^2}{n}$ $\left(\sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}} \right)$

- Paramètre à estimer: $\theta = \sigma^2 = V(x)$

– Estimateur:
(variance d'échantillon)

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Estimateur pour la moyenne: moyenne pondérée

- Cette fois-ci, chaque mesure d'échantillon (x_1, \dots, x_n) a une erreur connue, mais différente σ_i

- Paramètre à estimer: $\theta = \mu = E(x)$

- Estimateur:
$$\hat{\mu} = \frac{1}{w} \sum_{i=1}^N w_i x_i$$

(moyenne pondérée d'échantillon)

- w_i est le "poids": $w_i = 1/\sigma_i^2$ et $w = \sum_i w_i$

- Cet estimateur est non biaisé et a une meilleure efficacité que la moyenne simple:

$$\sigma_{\hat{\mu}} = 1/\sqrt{w}.$$

Méthode du maximum de vraisemblance

On a une mesure d'échantillon (x_1, \dots, x_N) . La distribution de x dépend de plusieurs paramètres $(\theta_1, \dots, \theta_n)$. Chaque mesure indépendante x_i suit la même densité de probabilité $f(x; \theta_1, \dots, \theta_n)$. La densité de probabilité jointe pour obtenir le résultat (x_1, \dots, x_N)

est: $f(x_1, \dots, x_N; \theta_1, \dots, \theta_n) = f(x_1; \theta_1, \dots, \theta_n) \cdot \dots \cdot f(x_i; \theta_1, \dots, \theta_n) \cdot \dots \cdot f(x_N; \theta_1, \dots, \theta_n)$.

- Fonction de vraisemblance: $L(\theta_1, \dots, \theta_n) = \prod_{i=1}^N f(x_i; \theta_1, \dots, \theta_n)$
- Le meilleur estimateur de $(\theta_1, \dots, \theta_n)$ est celui pour lequel L est maximum; cela revient à minimiser son logarithme $\ln(L)$:

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \quad i = 1, \dots, n$$

- Souvent on doit trouver une solution numérique
- Méthode approximativement non biaisée, efficace pour de grands échantillons, et robuste

Méthode du maximum de vraisemblance – exemple

Considérons une distribution exponentielle:

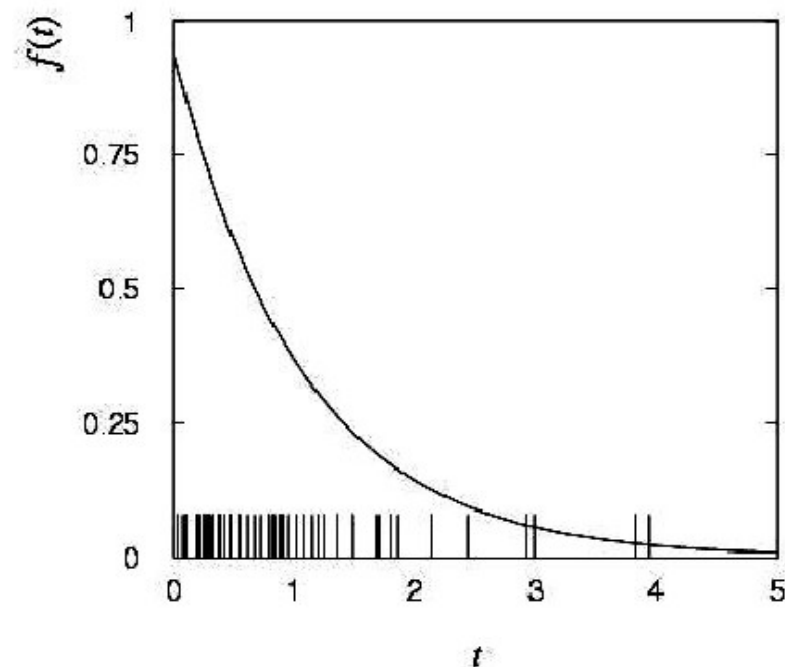
$$f(t; \tau) = \frac{1}{\tau} e^{-t/\tau}$$

- Par exemple on mesure n temps de désintégration t_1, \dots, t_n d'une particule et on veut estimer sa durée de vie τ

$$\ln L(\tau) = \sum_{i=1}^n \ln f(t_i; \tau) = \sum_{i=1}^n \left(\ln \frac{1}{\tau} - \frac{t_i}{\tau} \right)$$

$$\frac{\partial \ln L(\tau)}{\partial \tau} = 0 \rightarrow \hat{\tau} = \frac{1}{n} \sum_{i=1}^n t_i$$

→ Le meilleur estimateur du temps de vie est la moyenne d'échantillon



Méthode des moindres carrés – 1

- On mesure N quantités $y_1(x_1), \dots, y_N(x_N)$ assumant qu'elles sont indépendantes et suivent chacune une distribution de probabilité normale avec:

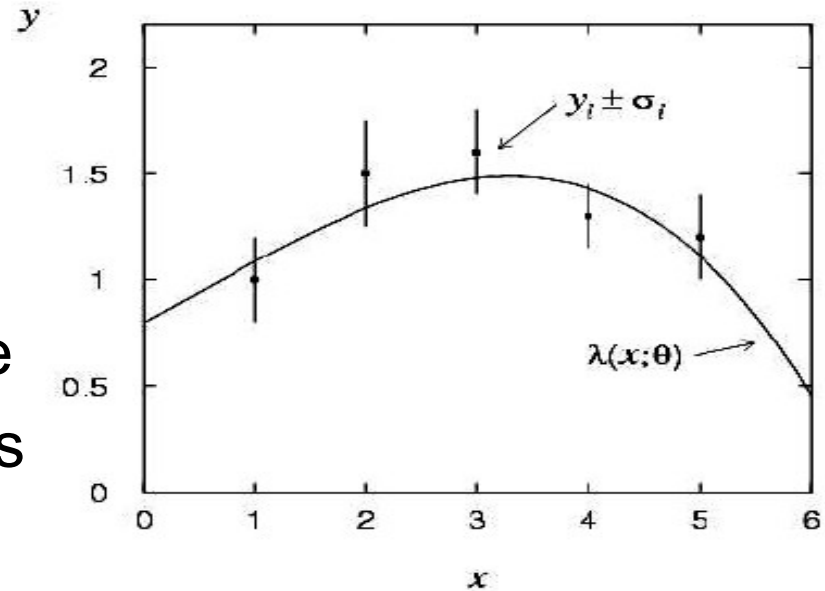
- $E(y_i) = \lambda(x_i, \theta)$
- $V(y_i) = \sigma_i^2$
- x_i et σ_i connues

- On veut estimer θ , le paramètre de la courbe que suivent les points

- Fonction de vraisemblance:

$$L(\theta) = \prod_{i=1}^N f(y_i; \theta) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[-\frac{(y_i - \lambda(x_i; \theta))^2}{2\sigma_i^2} \right]$$

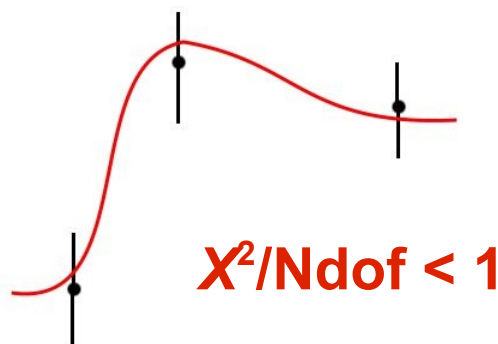
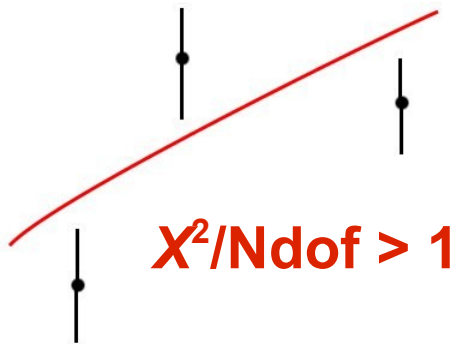
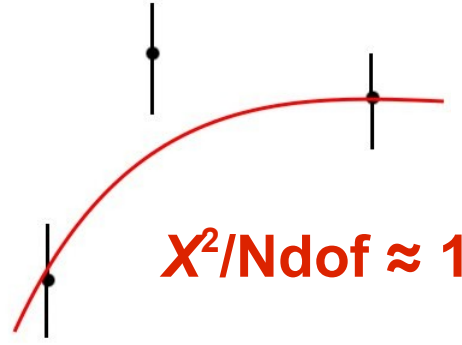
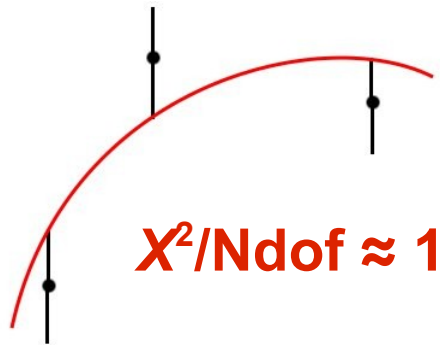
$$\ln L(\theta) = -\frac{1}{2} \sum_{i=1}^N \frac{(y_i - \lambda(x_i; \theta))^2}{\sigma_i^2} + \text{terms not depending on } \theta$$



Méthode des moindres carrés – 2

- On trouve que maximiser la vraisemblance dans le cas d'erreurs normales est équivalent à minimiser la fonction:

$$\chi^2(\theta) = \sum_{i=1}^N \frac{(y_i - \lambda(x_i; \theta))^2}{\sigma_i^2}$$



χ^2 réduit: $\chi^2/Ndof$

- $Ndof$ est le nombre de degrés de liberté: nombre de points – nombre de paramètres libres
- Idéalement, $\chi^2 \approx ndof$
 - χ^2 trop grand → mauvaise fonction
 - χ^2 trop petit → erreurs trop larges, ou trop de paramètres

Résumé

- On a défini le concept abstrait de **distribution**, ou **densité de probabilité**
- La **variance** d'une distribution d'une quantité mesurée représente l'erreur commise lors d'une mesure unique
 - On a vu comment l'erreur se propage lorsqu'on calcule une nouvelle quantité à partir de plusieurs quantités mesurées → attention aux corrélations!
- On a vu différentes distributions importantes en physique, en particulier, **Poisson**, **Normale**
- On a vu comment **estimer des paramètres** de distributions à partir d'un échantillon de mesures
 - Estimateurs simples pour la moyenne et la variance
 - Méthode du maximum de vraisemblance
 - Méthode des moindres carrés

Références utiles

- **G. Cowan**, Statistical Data Analysis, Clarendon, Oxford, 1998
- **R.J. Barlow**, Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences, Wiley, 1989
- **L. Lyons**, Statistics for Nuclear and Particle Physics, CUP, 1986
- **F. James**, Statistical and Computational Methods in Experimental Physics, 2nd ed., World Scientific, 2006
- **S. Brandt**, Statistical and Computational Methods in Data Analysis, Springer, New York, 1998
- **C. Amsler et al. (Particle Data Group)**, Review of Particle Physics: <http://pdg.lbl.gov/2011/reviews/rpp2011-rev-statistics.pdf>