III - Fonctions de densité de probabilité fondamentales

Un grand nombre de variables aléatoires décrivant des processus relatifs à tous les domaines de l'observation expérimentale sont distribuées exactement ou approximativement suivant des fonctions de densité de probabilité (ou des convolutions de fdp) aux propriétés mathématiques simples associées à trois types de processus définis par des règles très générales:

les processus de Bernoulli : distributions binomiale et multinomiale, ...

les processus de Poisson : distributions de Poisson, exponentielle, ...

les processus de Gauss: distribution gaussienne ou normale.

Jakob Bernoulli
1654 - 1705

Siméon Denis Poisson
1781 - 11840

Carl Friedrich Gauss
1777 - 18555

Processus de Bernoulli pour des variables aléatoires discrètes à valeurs entières positives ou nulles

Distribution multinomiale

- Un processus a k résultats possibles $A_1, A_2, \ldots, A_k$.
- Chaque résultat A_i a une probabilité p_i de se réaliser.
- Les A_i représentent l'ensemble des résultats possibles $\Rightarrow \quad \sum_{i=1}^{k} p_i = 1$

- Probabilité d'observer $\underline{r} = r_1, r_2, \ldots r_k$ résultats de type $A_1, A_2, \ldots A_k$

sur un ensemble de $n = \sum_{i=1}^{k} r_i$ essais étant donné $\underline{p}$:

$$f\left(\underline{r} \mid \underline{p}, n\right) = \frac{n!}{\prod_{i=1}^{k} r_i !} \prod_{i=1}^{k} p_i^{r_i}$$

nombre de combinaisons réalisant la configuration

probabilité de réaliser la configuration

- **Moyenne** $\mu_i = E\left[r_i\right] = \sum_{r_i=0}^{n} \frac{r_i \; n!}{\prod_{j=1}^{k} r_j !} \prod_{j=1}^{k} p_i^{r_j} = n \; p_i$

- **Variance** $\sigma_i^2 = E\left[\left(r_i - \mu_i\right)^2\right] = n \; p_i \left(1 - p_i\right)$

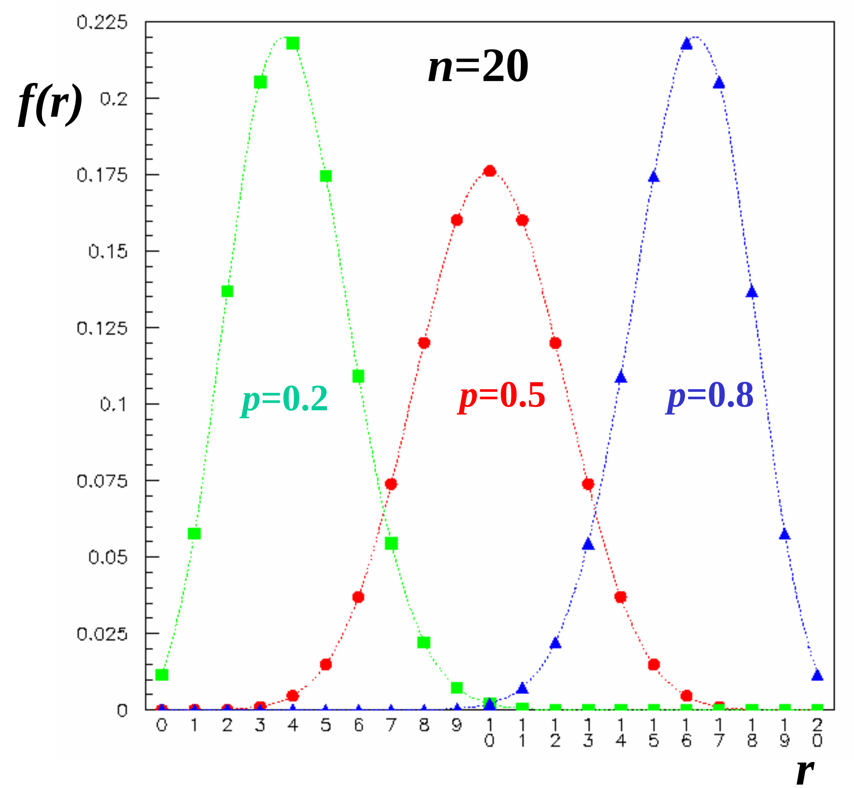
- Variables corrélées; correlation négative: $r_k = n - \sum_{i=1}^{k-1} r_i \quad \Rightarrow \quad \sigma_{ij} = -n \; p_i \; p_j \quad \Leftrightarrow \quad \rho_{ij} = -\sqrt{\frac{p_i \; p_j}{(1-p_i)(1-p_j)}}$

Distribution binomiale

- **Cas particulier où k=2 : $r_1 = r$ $\quad p_1 = p$ $\quad$ succès**
  $r_2 = n-r$ $\quad p_2 = 1-p$ $\quad$ échec

- **Probabilité d'observer r succès et $n-r$ échecs sur n essais**

$$\boxed{f(r \mid p,n) = \frac{n!}{r!\,(n-r)!}\, p^r\,(1-p)^{n-r}}$$



- **Moyenne $\mu = E[r] = n\,p$**
- **Variance $\sigma^2 = E\left[(r-\mu)^2\right] = n\,p\,(1-p)$**
- **Relation linéaire entre les variables: $r + (n-r) = n \;\Rightarrow\; \rho_{12} = -1 \;\Rightarrow\; \sigma_{12} = -n\,p\,(1-p)$**

Distribution binomiale négative

• Probabilité d'un r^{eme} succès précisément au n^{eme} essais

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

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

• Moyenne $\mu=E[n]=r\,/\,p$

• Variance $\sigma^2=E\left[(n-\mu)^2\right]=r\,(1-p)\,/\,p^2$

Distribution géométrique

• Cas particulier où r=1

$$\boxed{P(n\,|\,p)=p(1-p)^{n-1}}$$

• Moyenne $\mu=E[n]=1\,/\,p$

• Variance $\sigma^2=E\left[(n-\mu)^2\right]=(1-p)\,/\,p^2$

Exemples de distributions binomiales et multinomiales

- **L'efficacité quantique (*QE*) de la photocathode d'un photo-senseur exposé à un flux de photons est la probabilité d'émission d'un électron par effet photoélectrique par photon incident. A λ = 500 nm ($E_\gamma = hc/\lambda$), *QE* = 20%.**

Probabilité d'émission de 15 photo-électrons pour un signal incident de 50 photons?

$$f(15\,|\,0.2, 50) = \frac{50!}{15!\,35!}\,0.2^{15}\ 0.8^{35} = 0.03$$

- **Le méson K^+ a 6 modes de désintégration principaux:**

$K^+ \to \mu^+ \nu_\mu$	$p_1 = 0.63$
$K^+ \to \mu^+ \nu_\mu \pi^0$	$p_2 = 0.03$
$K^+ \to e^+ \nu_e \pi^0$	$p_3 = 0.05$
$K^+ \to \pi^+ \pi^0$	$p_4 = 0.21$
$K^+ \to \pi^+ \pi^+ \pi^-$	$p_5 = 0.05$
$K^+ \to \pi^+ \pi^0 \pi^0$	$p_6 = 0.03$

La probabilité d'observer $r_1, \ldots r_6$ événements dans chacun des 6 modes pour un total de *n* événements a la forme multinomiale $f(\underline{r}\,|\,\underline{p}, n)$

La probabilité d'observer r_i événements du type *i* pour un total de *n* événements a la forme binomiale $f(r_i\,|\,p_i, n)$

Processus de Poisson pour des variables aléatoires discrètes à valeurs entières positives ou nulles

- **Un processus est dit Poissonien si:**
  - **le nombre de succès sur un intervalle (de temps, d'espace, ...) Δx petit est 0 ou 1**
  - **la probabilité de succès sur l'intervalle $[x, x+\Delta x]$ est**

    - proportionnelle à Δx : $P_1([x, x+\Delta x]) = \Delta x/\beta$

    $$P_0([x, x+\Delta x]) = 1 - \Delta x/\beta$$

    - indépendante de x

- **La constante de proportionnalité β est l'intervalle moyen entre deux succès et $1/\beta$ est la densité de succès par unité d'intervalle.**

Démonstration de la fdp de la distribution de Poisson

β est l'intervalle moyen entre deux succès

$1/\beta$ est le nombre moyen de succès par unité d'intervalle

x est un intervalle exprimé dans les unités de β

$\mu = \dfrac{x}{\beta}$ est un nombre réel sans dimension, c'est le nombre moyen de succès sur l'intervalle x

$$P_0(\Delta x) = 1 - P_1(\Delta x) = 1 - \frac{\Delta x}{\beta}$$

$$P_0(x+\Delta x) = P_0(x) \times P_0(\Delta x)$$

$$P_0(x) - \frac{\Delta x}{\beta} P_0(x)$$

$$\frac{P_0(x+\Delta x) - P_0(x)}{\Delta x} = -\frac{1}{\beta} P_0(x)$$

$$\left.\begin{array}{l} \dfrac{dP_0(x)}{dx} = -\dfrac{1}{\beta} P_0(x) \\ P_0(0) = 1 \end{array}\right\} \rightarrow P_0(x) = e^{-x/\beta}$$

$$P_r(x+\Delta x) = P_{r-1}(x) \times P_1(\Delta x) + P_r(x) \times P_0(\Delta x)$$

$$= P_{r-1}(x)\frac{\Delta x}{\beta} + P_r(x)\left(1 - \frac{\Delta x}{\beta}\right)$$

$$\frac{P_r(x+\Delta x) - P_r(x)}{\Delta x} = -\frac{1}{\beta}\left(P_r(x) - P_{r-1}(x)\right)$$

$$\left.\begin{array}{l} \dfrac{dP_r(x)}{dx} = -\dfrac{1}{\beta}\left(P_r(x) - P_{r-1}(x)\right) \\ P_0(x) = e^{-x/\beta} \end{array}\right\} \rightarrow P_r(x) = \frac{1}{r!}\left(\frac{x}{\beta}\right)^r e^{-x/\beta}$$

$$\boxed{P_r(x) = f(r \mid \mu) = \frac{1}{r!}\mu^r e^{-\mu} \qquad avec \quad \mu = \frac{x}{\beta}}$$

Distribution de Poisson comme cas limite de la distribution binomiale

- **Cas limite de la binomiale quand :** $n \to \infty$ **grand nombre d'essais**

  $p \to 0$ **petite probabilité de succès**

  $np \to \mu$ **nombre moyen de succès fini**

- **Probabilité d'avoir r (entier) succès étant donné le nombre moyen μ (réel) de succès**

$$\boxed{f(r \mid \mu) = \frac{1}{r!}\mu^r e^{-\mu}}$$

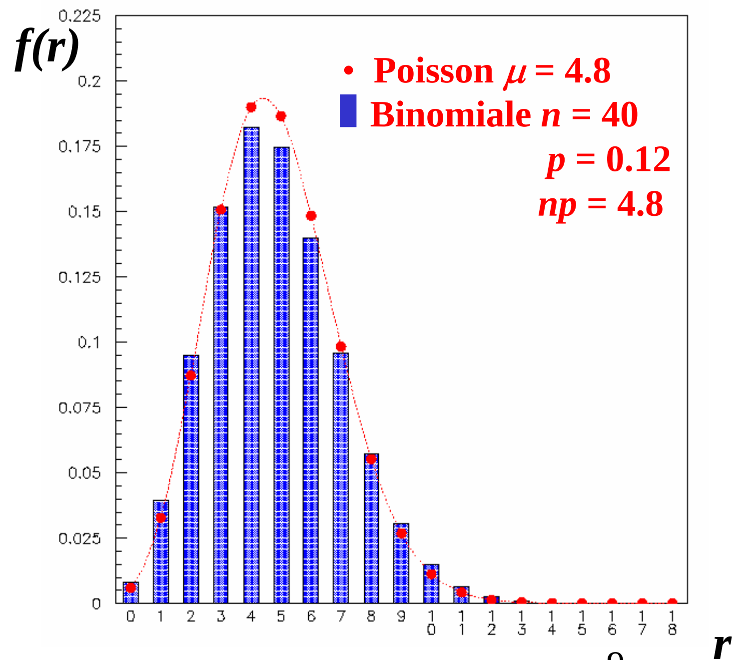
s'obtient à partir de la fdp binomiale et de la relation $lim_{n\to\infty} \; n! = \sqrt{2\pi n}\; n^n \; e^{-n}$

- **Moyenne μ**
- **Variance** $\sigma^2 = lim_{p\to 0}\, np(1-p) = np = \mu$

Binomiale: $0 \le r \le n$

Poisson: $0 \le r \le \infty \quad (lim\; n \to \infty)$

mais $f(r \gg \mu) \approx 0$



Exemples de distributions de Poisson

•1-Emission de photons par ionisation d'un plastique scintillant

C'est in phénomène poissonnien : la probabilité qu'un photon soit émis sur une distance Δl est proportionnelle à Δl et indépendante de la distance l parcourue par la particule ionisante depuis l'émission précédente d'un photon.

-La section efficace d'émission de photons de scintillation par une particule au minimum d'ionisation est d'environ $7 \cdot 10^{-20}\, cm^2 \quad (p \to 0)$

–La densité de centre diffuseurs (électrons) est de

$$Z\,\frac{N_A \rho}{A} \approx 6\,\frac{6 \cdot 10^{23} \times 1}{12} \approx 3 \cdot 10^{23}\, cm^{-3} \quad (n \to \infty)$$

- Le nombre moyen de photons émis est $\approx\ 7 \cdot 10^{-20}\, cm^2 \times 3 \cdot 10^{23}\, cm^{-3} \approx 20\,000\ cm^{-1}$

$(np \to \mu \text{ fini})$

- Probabilité d'emission de 15 photons par une couche de 10 μm **de polyvinyltoluène:**

$\mu = 20\,000 / 1000 = 20$

$$P(15 \mid 20) = \frac{1}{15!}\ 20^{15} e^{-20} \approx 0.05$$

•2-Désintégration de noyaux radioactifs

C'est in phénomène poissonnien : la probabilité qu'à un noyau radioactif de se désintégrer pendant un laps de temps Δt est proportionnel à Δt et indépendant du temps écoulé depuis sa création

-Vie moyenne de ^{234}U **est de** $\approx 3.4 \cdot 10^6$ **ans** $\approx 10^{14}\, s$

$\rightarrow$ **taux de désintégration** $\approx 10^{-14}\, Hz \quad (p \rightarrow 0)$

- Nombre de noyaux dans une source de 1 $\mu g \approx 10^{15} \quad (n \rightarrow \infty)$

- Taux moyen de désintégartion dans la source: $\approx 10^{15} \times 10^{-14}\, Hz \quad \approx 10\, Hz$

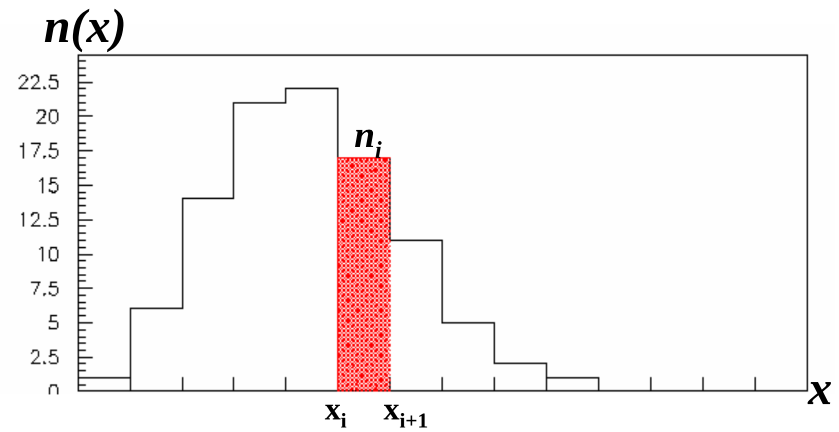
$(np \rightarrow \mu \text{ fini})$

- Probabilité d'observer 15 désintégrations en 1 s **:**

$$P(15\,|\,10) = \frac{1}{15!}\, 10^{15} e^{-10} \approx 0.035$$

•3-Contenu d'un histogramme

- **Soit:**
  - **- n événements repartis dans k classes**
  - **- n_i événements dans la classe i définie par $X_i \leq x < X_{i+1}$**
  - **- $n = \sum_{i=1}^{k} n_i$**
  - **- p_i probabilité pour que $X_i \leq x < X_{i+1}$**



- **La fdp des n_i est une multinomiale $f\left(\underline{n} \mid n, \underline{p}\right)$ avec $\mu_i = np_i$ $\quad \forall i = 1, k$**

$$\sigma_i^2 = np_i\left(1 - p_i\right)$$

La forme multinomiale assure la conservation du nombre total d'événements: tout événement appartient à une et une seule classe.

Pour une classe particulière i, la fdp de n_i est une binomiale $f\left(n_i \mid n, p_i\right)$ de même μ_i et σ_i^2

- **Si le nombre de classes k est grand de sorte que $p_i \ll 1 \quad \forall i = 1, k$**

La fdp binomiale de n_i peut être approximée par une fdp de Poisson $f\left(n_i \mid \mu_i\right)$ avec $\mu_i = np_i$

$$\sigma_i^2 = \mu_i = np_i$$

Le contenu n_i de chaque boite fluctue indépendamment : $n \neq \sum_{i=1}^{k} n_i$

Les variations sur n_i sont surestimée par l'approximation : $\sigma_i^2\left(\text{Poisson}\right) = \sigma_i^2\left(\text{binomiale}\right) / \left(1 - p_i\right)$

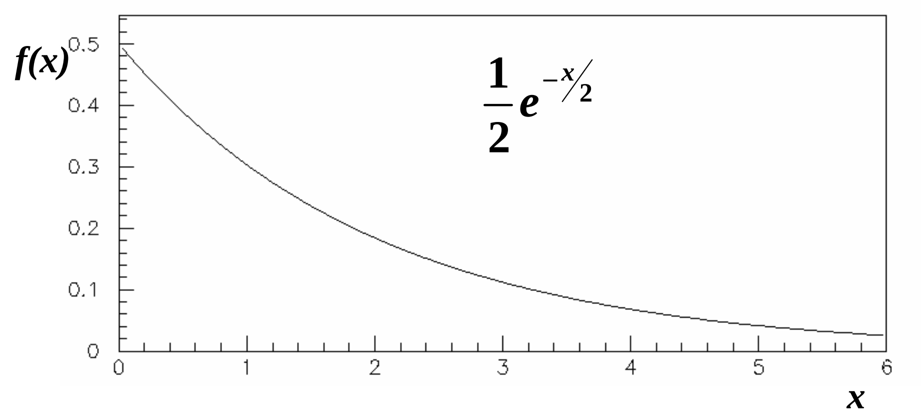
Lim $_{p_i \to 0 \ \forall i}$ $\sigma_i^2\left(\text{Poisson}\right) = \sigma_i^2\left(\text{binomiale}\right)$

Distribution exponentielle

• **Probabilité d'un premier succès sur l'intervalle $[x, x+dx]$ lors d'un processus poissonien.**
Soit β la distance moyenne entre deux succès.

$$P_0([0,x])\times P_1([x,x+dx]) = \frac{1}{0!}\left(\frac{x}{\beta}\right)^0 e^{-x/\beta}\times\frac{dx}{\beta} = \frac{1}{\beta}e^{-x/\beta}dx$$

$$\boxed{f(x\,|\,\beta) = \frac{1}{\beta}e^{-x/\beta}}$$



• **Moyenne $\mu = \beta$**

• **Variance $\sigma^2 = \beta^2$**

• **Loi de décroissance exponentielle de toute population par processus poissonnien**

• **Probabilité pour que la valeur de x où a lieu le premier succès soit :**

$$x \geq x_0 \qquad \frac{1}{\beta}\int_{x_0}^{\infty} e^{-y/\beta}dy = e^{-x_0/\beta}$$

$$x < x_0 \qquad = 1 - e^{-x_0/\beta}$$

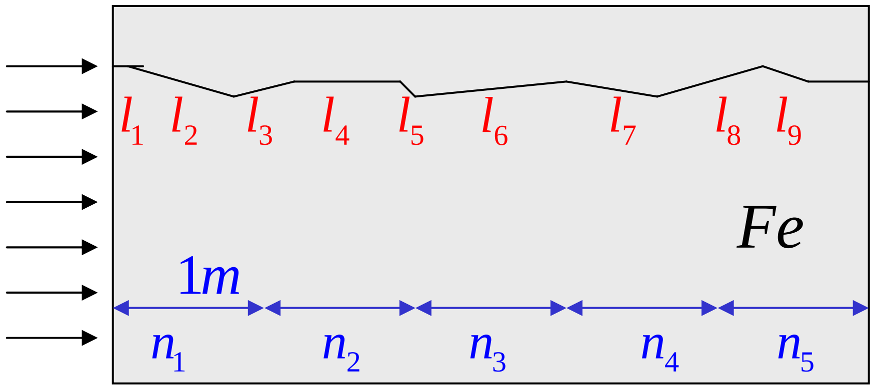
Distribution du nombre d'événements r parmi un total de n tels que $x \geq x_0$:

fdp binomiale $f\left(r\,|\,n, p = e^{-x_0/\beta}\right)$

Relation entre distributions exponentielle et de Poisson

Longueur d'interaction d'une particule dans un milieu de longueur moyenne d'interaction λ

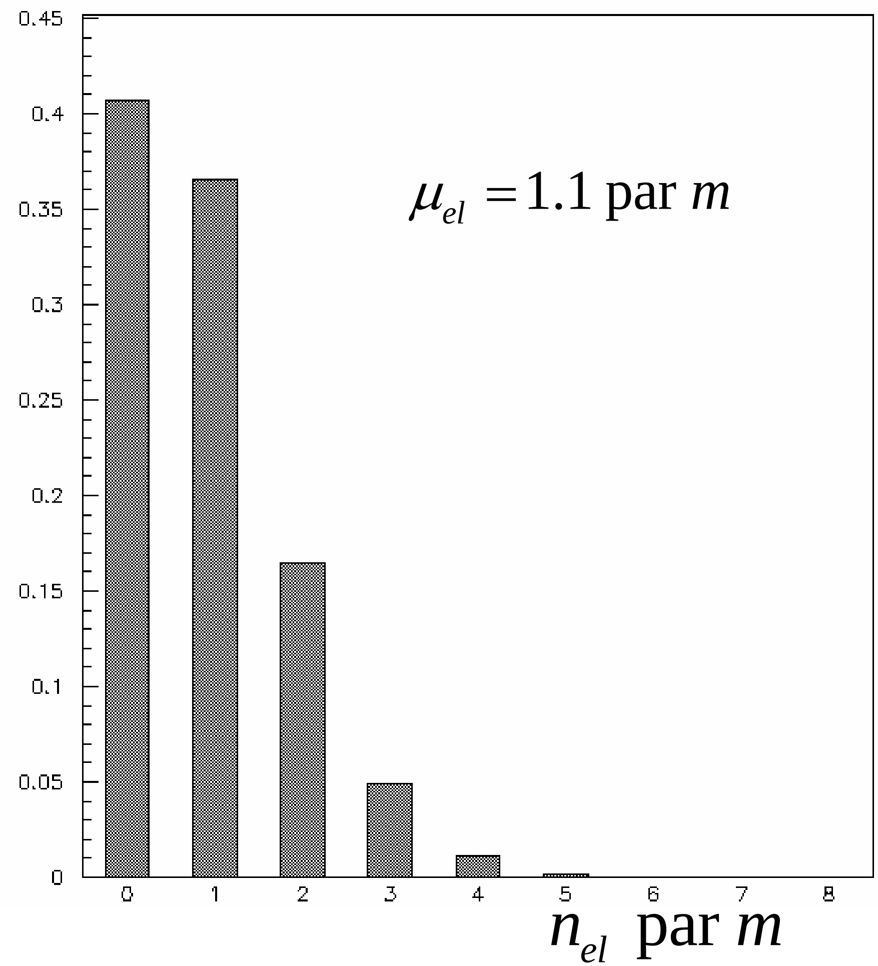
Faisceau de protons de très haute énergie

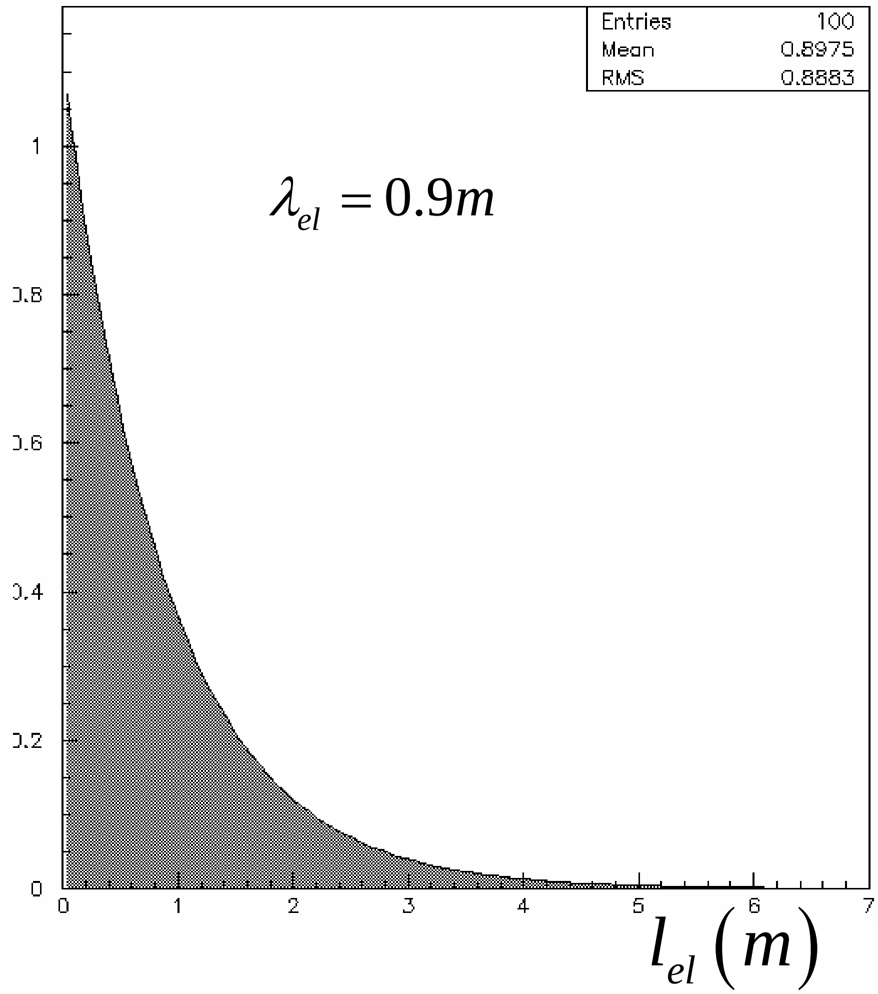


$$\sigma_{pFe}^{elast} \approx 0.110 \cdot 10^{24}\, cm^2 \text{ pour } E \geq 50\, GeV$$

$$\lambda_{el} = \frac{A_{Fe}}{\sigma_{pFe}^{elast} N_A \rho_{Fe}} \approx 0.9\, m\text{: moyenne de la distribution exponentielle des } l_i$$

$$\mu_{el} = \frac{1\, m}{0.9\, m} \approx 1.1\text{:moyenne de la distribution poisonniènne des } n_i$$





$$f\left(n_{el} \mid \mu_{el} = 1.1\right) = \frac{1}{n_{el}\,!} 1.1^{n_{el}} e^{-1.1}$$

$$f\left(l_{el} \mid \lambda_{el} = 0.9\,m\right) = \frac{1}{0.9} e^{-l_{el}/0.9}$$

Distribution d'Erlang

- **Généralisation de la distribution exponentielle**

  Probabilité d'un k^{eme} succès sur l'intervalle $[x, x+dx]$ lors d'un processus poissonien.

  Soit β la distance moyenne entre deux succès

$$\boxed{f\left(x \mid k,\beta\right)=\frac{1}{\beta\ \left(k-1\right)!}\left(\frac{x}{\beta}\right)^{k-1} e^{-x/\beta}}$$

- $\mu = k\beta$
- $\sigma^2 = k\beta^2$

Distribution gamma

- **Généralisation de la distribution d'Erlang.**
- k **entier** $\rightarrow$ α **réel**
- $\left(k-1\right)! \rightarrow \Gamma\left(\alpha\right)$

Convolution de distributions binomiales et poissoniennes

Somme de variables distribuées suivant des fdp poissoniennes

Soit r_1 et r_2 deux variables distribuées suivant des fdp de Poisson de moyenne μ_1 et μ_2.

La variable $r_1 + r_2$ est distribuée suivant une fdp de Poisson de moyenne $\mu_1 + \mu_2$

Produit de variables distribuées suivant des fdp poissonienne et binomiale

Soit r une variable distribuée suivant une binomiale de moyenne np (n essais, probabilité p de succès).

Soit n, le nombre d'essais, distribué suivant une distribution de Poisson de moyenne μ.

Le nombre de succès r est distribué suivant une fdp de Poisson de moyenne $p\mu$.

Exemple de convolution: efficacité d'un compteur à détecter une particule ionisante

Le nombre moyen de photons émis par une couche de 0.1 mm de plastique scintillant traversé par une particule chargée au minimum d'ionisation est de $\mu_\gamma = 200$.

Le signal photonique doit traverser 1 m de scintillateur de longueur d'absorption λ=0.5 m pour atteindre la photocathode d'un photo-senseur.

La fraction de photons atteignant le photo-senseur est donc $p_\lambda = \int_1^\infty \frac{1}{0.5} e^{-x/0.5} dx = e^{-1/0.5} = 0.14$

et leur nombre est distribué suivant une fdp de Poisson de moyenne $p_\lambda \mu_\gamma = 0.14 \times 200 = 28$

L'efficacité quantique (nombre d'électron émis par photon incident) de la photocathode du photo-senseur est $p_{QE} = 0.2$

Le nombre d'électrons émis par la photocathode est distribué suivant une fdp de Poisson de moyenne $p_{QE}\, p_\lambda \mu_\gamma = 0.2 \times 0.14 \times 200 = 5.6.$

L'efficacité ε du détecteur à détecter des particules ionisantes est la probabilité qu'un signal d'au moins un photo-électron soit emis par la photocathode:

$$\varepsilon = 1 - P(0\,|\,5.6) = 1 - e^{-5.6} = 1. - 0.004 = 0.996$$

Processus de Gauss pour des variables aléatoires continues

- **La distribution normale ou gaussienne fondamentale en statistique comme cas limite des grands échantillons bien que pratiquement aucun processus naturel ne soit strictement gaussien.**

- **Bonne approximation des distributions symétriques « en cloche ».**

- **La distribution de Poisson (et donc la distribution binomiale) tend vers une distribution normale à la limite des grands échantillons.**

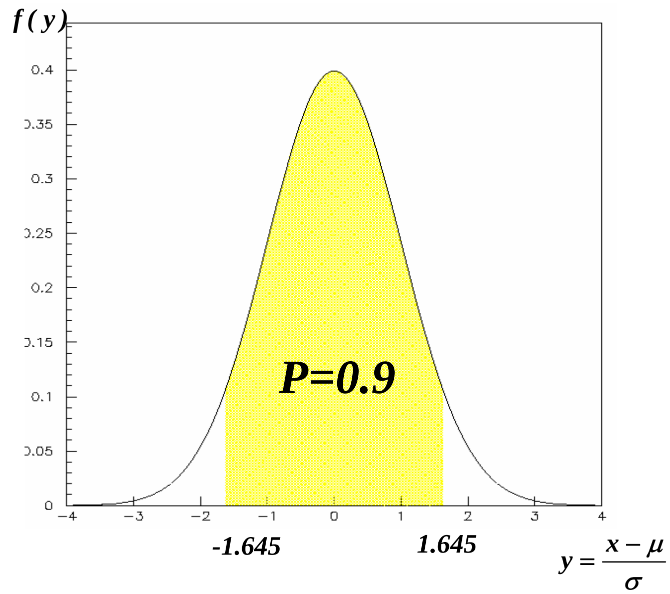
- **Théorème central limite:**
  La moyenne d'un échantillon est distribuée normalement à la limite des grands échantillons.
  De même pour la somme de variables aléatoires indépendantes: erreur statistique sur des mesures complexes résultant de mesures élémentaires indépendantes.

- **Les statistiques (moyenne, variance) des échantillons suivent des distributions (chi-carré, Student, Fisher) construites à partir de la distribution normale.**
  Ces distributions tendent vers la distribution normale à la limite des grands échantillons.

Distribution normale ou gaussienne

- **Fonction de densité de probabilité**

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



- **Variable standardisée: centrée sur μ et normalisée à σ**

$$y = \frac{x-\mu}{\sigma} \rightarrow \mu_y = 0 \text{ et } \sigma_y^2 = 1$$

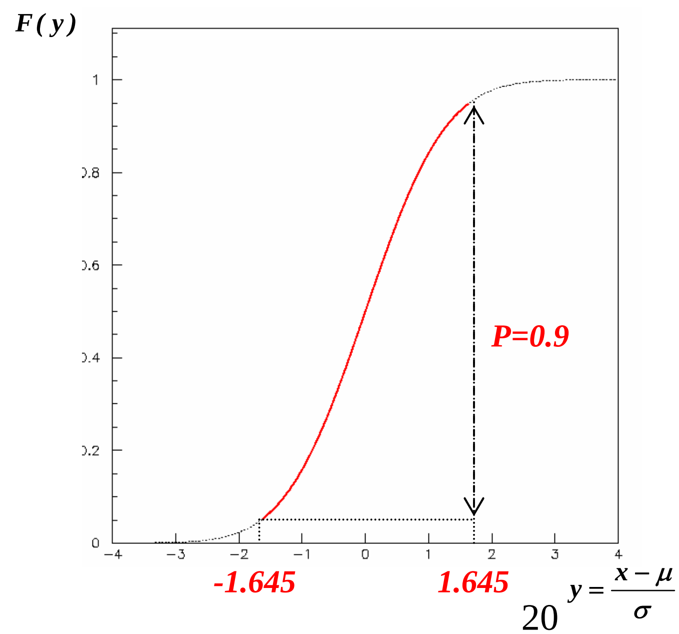
$$f(y|0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} = N(0,1)$$

$$x = \sigma y + \mu$$

- $$P(\mu - n\sigma \le x \le \mu + n\sigma) = \int_{\mu-n\sigma}^{\mu+n\sigma} f(x|\mu,\sigma^2)dx$$

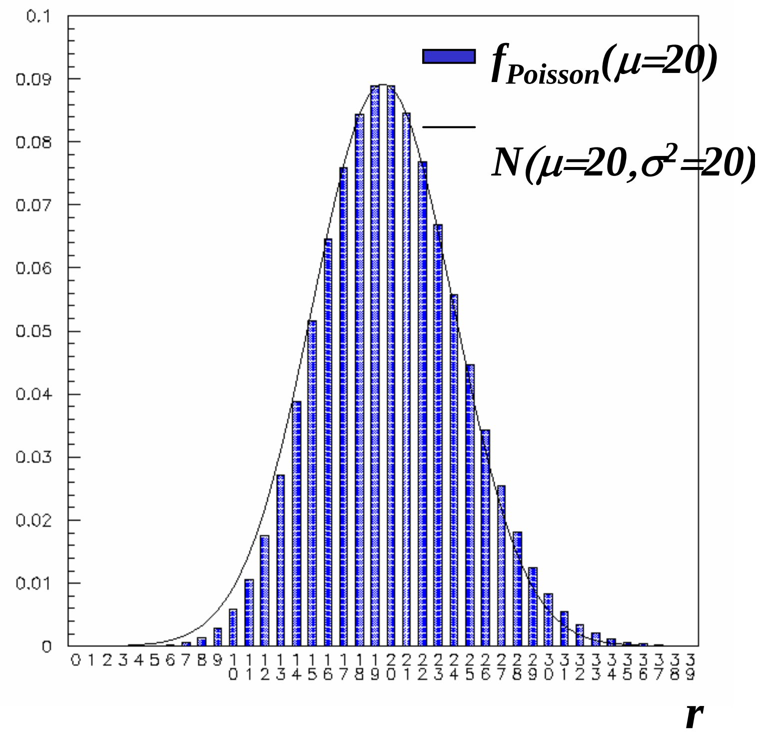
$$= 2F(\mu + n\sigma|\mu,\sigma^2) - 1 = 1 - 2F(\mu - n\sigma|\mu,\sigma^2)$$

n	$P(\mu - n\sigma < x < \mu + n\sigma)$
1	0.683
1.645	0.900
1.960	0.950
2	0.955
2.576	0.990
3	0.997
3.29	0.999



Distribution normale comme cas limite de la distribution de Poisson

$$lim_{\mu\to\infty} \frac{1}{r\,!}\mu^r e^{-\mu} = \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{1}{2}\frac{(r-\mu)^2}{\mu}}$$



Exemple: contenu d'un histogramme

Si le nombre d'événements n est grand et toutes les probabilités p_i par classe sont petites (le nombre de boites est grand) :
La distribution binomiale $f(n_i|n,p_i)$ du contenu de chaque classe peut être approximé par une distribution de Poisson $f(n_i|\, np_i)$

Si le nombre d'événements n_i dans chaque boite est grand:
La distribution de Poisson $f(n_i|\, np_i)$ du contenu de chaque boite peut être approximé par une distribution normale $N(n_i|\, np_i,\, np_i)$

Distributions binormale et multinormale

Distribution binormale et multinormale de variables indépendantes

$$f\left(x_1,x_2\right)=\frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}}e^{-\frac{1}{2}\left(\frac{\left(x_1-\mu_1\right)^2}{\sigma_1^2}+\frac{\left(x_2-\mu_2\right)^2}{\sigma_2^2}\right)}$$

$$f\left(\underline{x}\right)=\frac{1}{\left(2\pi\right)^{\frac{n}{2}}\prod_{i=1}^{n}\sqrt{\sigma_i^2}}e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{\left(x_i-\mu_i\right)^2}{\sigma_i^2}} \qquad \underline{x}=\left(x_1,\ldots,x_n\right)$$

Distribution binormale de variables corrélées

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