

# Éléments de statistiques pour les data sciences

## Cours 3 : probabilités continues

Dr. Ph. Müllhaupt

IGM - EPFL

—

# Éléments de statistiques pour les data sciences

## Cours 3 : probabilités continues

Dr. Ph. Müllhaupt

IGM - EPFL

—

# Plan

- 1 densité de probabilité univariée
  - distribution uniforme
  - distribution normale
- 2 espérance mathématique
  - moyenne
  - variance
- 3 la distribution exponentielle
- 4 la distribution  $\Gamma$ 
  - la distribution du  $\chi^2$
- 5 densité de probabilité
  - jointe
  - marginale
- 6 somme de variables aléatoires
- 7 fonction génératrice
- 8 produit de convolution

## densité de probabilité continue univariée

Soit  $X$  une variable aléatoire définie sur l'espace  $(\Omega, F, P)$  avec la fonction de distribution cumulée

$$F(x) = P(X \leq x)$$

### définition

S'il existe une fonction  $f$  telle que

$$F(x) = \int_{-\infty}^x f(y) dy$$

alors la fonction  $f$  est appelée la densité de probabilité (également appelée la distribution). Elle satisfait l'égalité

$$\int_{-\infty}^{+\infty} f(y) dy = 1$$

# distribution uniforme

## définition

La densité de probabilité uniforme (distribution uniforme) est définie par

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$$

Elle représente la possibilité qu'un nombre situé dans l'intervalle  $[a, b]$  puisse apparaître comme valeur de la variable aléatoire  $X$  avec autant de chance qu'un autre nombre de ce même intervalle.

## densité de probabilité uniforme

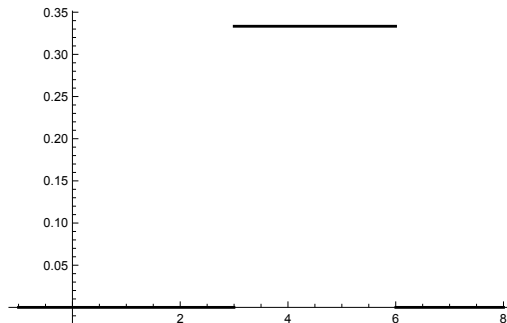


Figure – la distribution uniforme avec  $a = 3$  et  $b = 6$

# la distribution normale (gaussienne, Laplace Gauss)

centrée à l'origine et de variance unité  $\mathcal{N}(0, 1)$

## définition

La distribution normale de moyenne nulle et de variance unité notée  $\mathcal{N}(0, 1)$  est donnée par

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

Un moyen pour s'en souvenir est de prendre le logarithme naturel et on obtient la somme d'une constante et d'une forme quadratique négative standard  $-\frac{1}{2}x^2$ . Le facteur devant l'exponentiel est obtenu pour forcer  $\int_{-\infty}^{+\infty} f(x) dx = 1$ .

# la distribution normale (gaussienne, Laplace-Gauss)

centrée à l'origine et de variance unité  $\mathcal{N}(0, 1)$

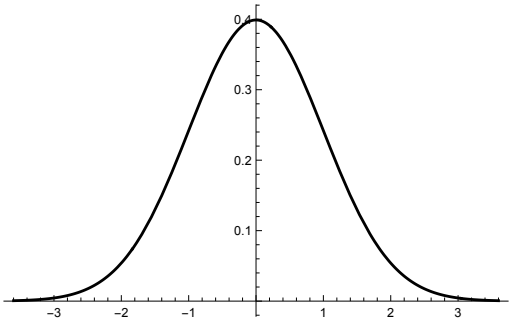


Figure – la distribution normale de moyenne nulle et de variance unité  $\mathcal{N}(0, 1)$



# la distribution normale

non centrée et de variance quelconque

## définition

La distribution normale de moyenne  $\mu$  et de variance  $\sigma^2$   $\mathcal{N}(\mu, \sigma^2)$  est donnée par

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

Le centrage est autour de la moyenne  $\mu = \mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx$  et "l'épaisseur" est proportionnelle à l'écart type  $\sigma$ .

## espérance mathématique

Soit une variable aléatoire continue  $X$  de densité de probabilité  $f$  et une fonction  $\psi$  de la variable aléatoire  $X$  dans les nombres réels  $\mathbb{R}$

$$\psi : X \rightarrow \mathbb{R}$$

L'espérance mathématique est définie par

$$\mathbb{E}[\psi(X)] = \int_{-\infty}^{+\infty} \psi(x) f(x) dx$$

## moyenne

Soit une variable aléatoire continue  $X$  de densité de probabilité  $f$ . La moyenne  $\mu$  est l'espérance mathématique de la fonction identité  $\psi(X) = X$

définition

$$\mu \triangleq \mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

## variance

Soit une variable aléatoire continue  $X$  de densité de probabilité  $f$ . La variance  $\sigma^2$  est l'espérance mathématique de la fonction  $\psi(X) = (X - \mu)^2$

### définition

$$\sigma^2 \triangleq \mathbb{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Remarque : la définition  $\mathbb{E}[(X - \mu)^2]$  est identique que  $X$  soit discrète ou  $X$  continue. Ce qui change est son calcul, par série lorsque  $X$  est discrète et par une intégrale lorsque  $X$  est continue.

# la distribution normale

non centrée et de variance quelconque  $\mathcal{N}(\sigma, \mu)$

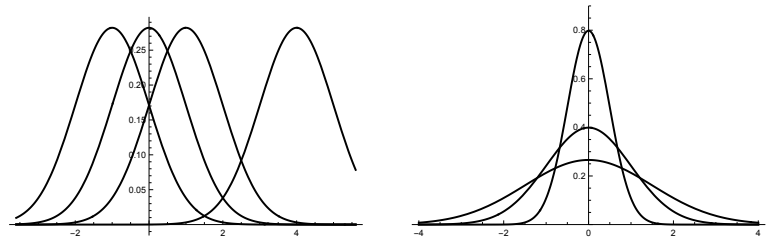


Figure – La distribution normale de moyenne nulle variable à gauche pour  $\mu = -1, 0, 1, 4$  et de moyenne nulle mais de variance  $\sigma^2$  variable, avec  $\sigma = 0.5, 1, 1.5$

# la distribution exponentielle

## définition

La distribution exponentielle est définie par

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Il est immédiat de vérifier que  $\int_{-\infty}^{+\infty} \lambda e^{-\lambda x} = 1$ . De plus, la valeur à l'origine donne le paramètre  $\lambda$ .

## la distribution exponentielle

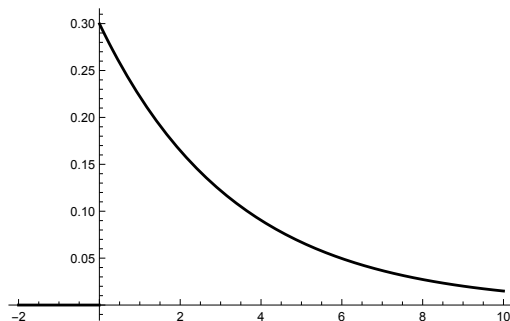


Figure – la distribution la distribution expoentielle de paramètre  $\lambda = 0.3$ .

# la distribution d'Erlang

continue de paramètres  $k \in \mathbb{N}$  et  $\lambda \in \mathbb{R}, \lambda > 0$

## définition

La distribution d'Erlang de paramètres  $k \in \mathbb{N}$  et  $\lambda \geq 0$  est donnée par

$$f(x; k, \lambda) \triangleq \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \quad x, \lambda \geq 0$$



## Remarques

intuitives, sans trop de rigueur...

- ① En comparant les courbes, on constate que la distribution exponentielle est monotone décroissante.
- ② Les distributions d'Erlang ont une tendance à croître initialement pour atteindre une valeur maximale et ensuite décroissent comme une exponentielle.
- ③ Il y a un 'combat' entre le facteur de croissance  $x^{k-1}$  la croissance monomiale et la décroissance exponentielle : pour de petite valeur de  $x$  le facteur  $x^{k-1}$ , croissance monomiale, l'emporte sur  $e^{-x}$ , la décroissance exponentielle. Pour de grande valeur de  $x$ , le facteur de décroissance exponentielle  $e^{-x}$  l'emporte sur la croissance monomiale  $x^{k-1}$ .

# la fonction $\Gamma(\alpha)$

la fonction gamma généralise la factorielle

## définition

La fonction  $\Gamma(n)$  discrète est la fonction pour  $n \in \mathbb{N}$

$$\Gamma(n) \triangleq (n-1)!$$

Pour une valeur  $\alpha \in \mathbb{R}$ , on a la définition

$$\Gamma(\alpha) \triangleq \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad \alpha > 0$$

# la fonction $\Gamma$

## propriétés

### propriétés

Pour tout  $\alpha > 0$ ,  $\alpha \in \mathbb{R}$

$$\textcircled{1} \int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^{\alpha}} \quad \lambda > 0$$

$$\textcircled{2} \Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$$

$$\textcircled{3} \Gamma(n) = (n-1)! \quad n \in \mathbb{N}$$

$$\textcircled{4} \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

## la distribution $\Gamma$ (gamma)

C'est une autre façon de décrire une distribution du type de celle de Poisson où s'interpose l'effet de la croissance de la fonction d'un facteur  $x^\alpha$  avec la décroissance de  $e^{-x}$ . L'exposant  $\alpha$  évolue selon la fonction  $\Gamma$ .

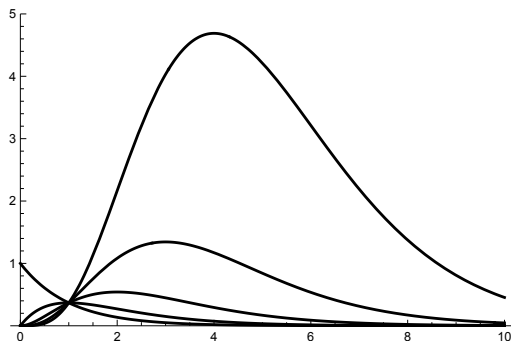


Figure – Graphiques de  $x^{\alpha-1} e^{-x}$  pour  $\alpha = 1, 2, 3, 4, 5$

## la distribution gamma $\Gamma(\alpha, \lambda, x)$

### définition

$$\alpha > 0, \alpha \in \mathbb{R}$$

$$f(x) = \Gamma(\alpha, \lambda, x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

## la distribution gamma $\Gamma(\alpha, \lambda, x)$

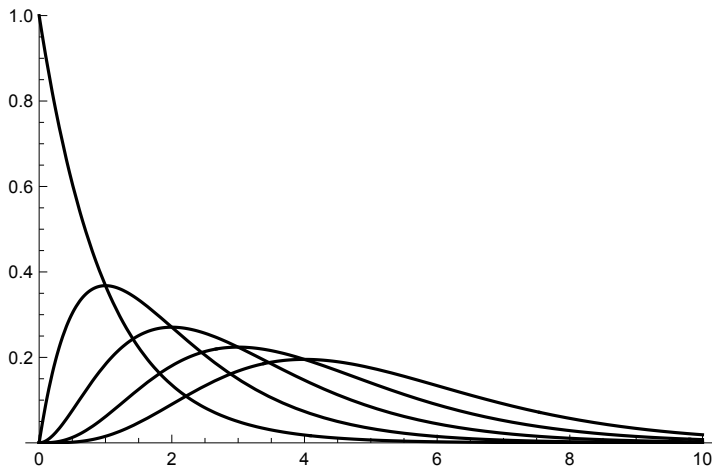


Figure – Représentation graphique de la densité de probabilité  $\Gamma(\alpha, 1, x)$  pour les valeurs  $\alpha = 1, 2, 3, 4, 5$ . Remarque  $\alpha = 1$  correspond à la distribution exponentielle.

## cas particulier : la distribution du $\chi^2$

$$X \sim \chi^2_\nu \Leftrightarrow X \sim \Gamma\left(\alpha = \frac{\nu}{2}, \lambda = \frac{1}{2}\right)$$

$\nu$  désigne les degrés de liberté (grosso modo le nombre de façons que possède le hasard de perturber de manière indépendante le résultat de l'expérience). On étudiera  $\chi^2_\nu$  à la leçon 4.

## fonction d'une variable aléatoire

$$Y = h(X)$$

Soit  $X$  une variable aléatoire avec une densité de probabilité  $f$  et de fonction de répartition  $F$ . On aimerait déterminer la densité de probabilité de  $Y = h(X)$  où  $h$  est une fonction réelle d'une variable réelle.

### méthode recommandée

Il est commode de travailler avec la fonction de répartition

$$G(y) = P(Y \leq y)$$

à partir de la fonction de répartition  $F(x) = P(X \leq x)$ . Ensuite on utilise

$$g(y) = \frac{dG(y)}{dy}$$



## exemple d'une fonction de variable aléatoire

$$Y = e^{-X}$$

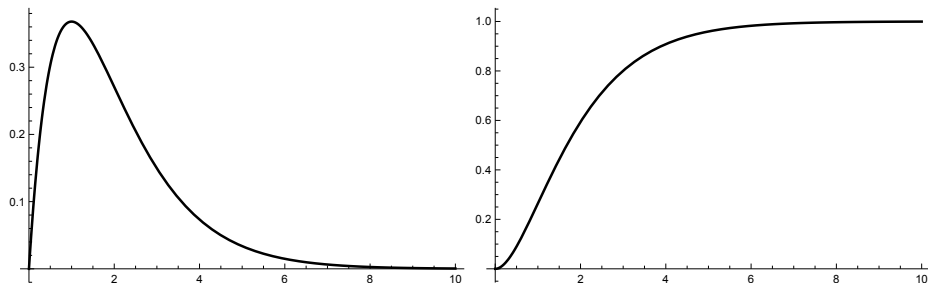


Figure – à gauche la densité de probabilité  $f(x)$  et à droite la fonction de répartition  $F(x)$

## exemple d'une fonction de variable aléatoire

$$Y = e^{-X}$$

Soit  $X$  une variable aléatoire continue de densité de probabilité

$$f(x) = \begin{cases} x e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

la fonction de répartition correspondante est

$$F(x) = (1 - (x + 1)e^{-x})$$

Etudions

$$Y = e^{-X}$$

## exemple d'une fonction de variable aléatoire

$$Y = e^{-X}$$

Soit  $G$  la fonction de répartition de  $Y$ .

$$\begin{aligned} G(y) &= P(Y \leq y) = P(e^{-X} \leq y) = P(-X \leq \ln y) \\ &= P(X \geq -\ln y) = 1 - P(X < -\ln y) \\ &= 1 - F(-\ln y) \end{aligned}$$

en substituant  $x = -\ln y$  cela donne

$$G(y) = y(1 - \ln y)$$

## exemple d'une fonction de variable aléatoire

$$Y = e^{-X}$$

On obtient  $g$  par dérivation

$$g(y) = \frac{dG}{dy} = \begin{cases} -\ln y & 0 < y < 1 \\ 0 & \text{sinon} \end{cases}$$

## exemple d'une fonction de variable aléatoire

$$Y = e^{-X}$$

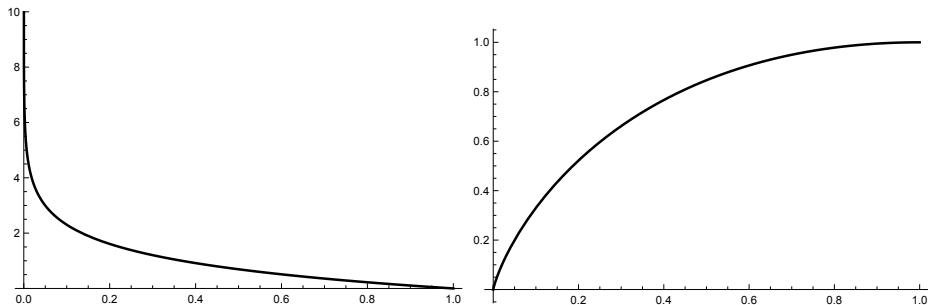


Figure – à gauche la densité de probabilité  $g(y)$  et à droite la fonction de répartition  $F(y)$

## fonction monotone d'une variable aléatoire

croissante ou décroissante

Le résultat est général quel que soit la fonction monotone de  $h(x)$  avec  $Y = h(X)$ .

## théorème

Soit  $X$  une variable aléatoire continue de densité de probabilité  $f(x)$  et  $Y$  une variable aléatoire continue de densité de probabilité  $g(y)$ . Soit également une fonction  $Y = h(X)$  monotone dérivable (et donc continue) qui relie les deux variables aléatoires. On a le résultat

$$g(y) = f(x) \cdot \left| \frac{dx}{dy} \right|$$

$$x = h(y) \quad y = h^{-1}(x) \quad dx = \frac{dh}{dy} dy$$

## densité de répartition jointe

Soit un couple de variables aléatoires  $X$  et  $Y$ .

A tout évènement exprimé par  $X$  et  $Y$  correspond une région  $A$  du plan. On a

$$P(X, Y) = P(A)$$

### hypothèse

L'hypothèse est que  $P(A)$  puisse se représenter par une intégrale double

$$P(A) \triangleq \int \int h(x, y) dx dy$$

alors la fonction  $h(x, y)$  est appelée la densité de probabilité jointe.

# Poker

$x$ : nombre d'aces  
 $y$ : " de rois

parmi les  
52 cartes

dans 5  
cartes

4 aces  
4 rois.

entout.

5 cartes  
en main

$$f(x, y) = \binom{4}{x} \binom{4}{y} \binom{44}{5-x-y} / \binom{52}{5}$$

$$f_1(x) = \sum_{y=0}^{\infty} f(x, y) = \binom{4}{x} \left[ \sum_{y=0}^{\infty} \binom{4}{y} \binom{44}{5-x-y} \right] / \binom{52}{5}$$
$$= f_1(x) = \binom{4}{x} \binom{48}{5-x} / \binom{52}{5}$$



|                                   |  | $f(x,y) = \binom{4}{x} \binom{4}{y} \binom{44}{5-x-y} / \binom{52}{5}$ |   |   |          |                        | $\text{marge} \downarrow$<br>$f_1(x)$                   |  |
|-----------------------------------|--|--|---|---|----------|------------------------|---|--|
| $x \backslash y$                  |  | $y=0$  | $y=1$   | $y=2$   | $y=3$    | $y=4$                  |   |  |
| 0                                 |  | 0,417863   | 0,20893                                       | 0,030575                                      | 0,001456 | $1,69 \cdot 10^{-5}$   | 0,658842  |  |
| 1                                 |  | 0,20893  | 0,081534                                      | 0,008736                                      | 0,000271 | $0,000002$             | 0,299474  |  |
| 2                                 |  | 0,030575   | 0,008736                                      | 0,000271                                      | 0,000009 | 0                      | 0,039929  |  |
| 3                                 |  | 0,001456   | 0,000271<br>$(2,708 \cdot 10^{-4})$           | $\underbrace{0,000009}_{9,234 \cdot 10^{-6}}$ | 0        | 0                      | 0,01736   |  |
| 4                                 |  | $\underbrace{0,000017}_{1,69 \cdot 10^{-5}}$                           | $\underbrace{0,000002}_{1,539 \cdot 10^{-6}}$ | 0   | 0        | 0                      | $7,3875 \cdot 10^{-5}$                                  |  |
| $\text{marge} \rightarrow f_2(y)$ |  | 0,658842   | 0,299474                                      | 0,039929                                      | 0,01736  | $7,3875 \cdot 10^{-5}$ | $f_1(x) = \binom{4}{x} \binom{48}{5-x} / \binom{52}{5}$ |  |

## fonction de répartition jointe

Soit l'évènement

$$\{\omega \in \Omega | X(\omega) \leq x\} \cap \{\omega \in \Omega | Y(\omega) \leq y\}$$

que l'on peut écrire avec un "abus de notation"

$$(X \leq x) \cap (Y \leq y)$$

### définition

S'il existe une fonction  $H(x, y)$  telle que

$$P(X \leq x) \cap P(Y \leq y) \triangleq H(x, y)$$

alors  $H(x, y)$  est appelée la fonction de répartition jointe.

# densité marginale

## définition

Soit  $h(x, y)$  la fonction de densité jointe des deux variables aléatoires  $X$  et  $Y$ . Pour une valeur de  $x$  donnée, on peut définir la fonction de densité de probabilité marginale

$$f_X(x) \triangleq \int_{-\infty}^{+\infty} h(x, y) dy$$

de même la fonction de densité marginale selon  $y$

$$f_Y(y) \triangleq \int_{-\infty}^{+\infty} h(x, y) dx$$

## somme de variables aléatoires

Soit la somme de deux variables aléatoires continues réelles

$$Z = X + Y$$

et une fonction  $\phi(x, y)$  quelconque, on a

définition

$$\mathbb{E}[\phi(x, y)] \triangleq \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi(x, y) h(x, y) dx dy$$

## somme de variables aléatoires

en particulier pour  $\phi(x, y) = x + y$

$$\begin{aligned}
 \mathbb{E}[Z] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) h(x, y) dx dy \\
 &= \int_{-\infty}^{+\infty} x \left( \int_{-\infty}^{+\infty} h(x, y) dy \right) dy + \int_{-\infty}^{+\infty} y \left( \int_{-\infty}^{+\infty} h(x, y) dx \right) dx \\
 &= \int_{-\infty}^{+\infty} x f_X(x) dx + \int_{-\infty}^{+\infty} y f_Y(y) dy \\
 &= \mathbb{E}(X) + \mathbb{E}(Y)
 \end{aligned}$$

# corrélation

variables aléatoires  $X$  et  $Y$  continues

$$\sigma_{XY} \triangleq \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y) h(x, y) dx dy$$

Thm:

indépendance  $\Leftrightarrow h(x, y) = f(x)g(y)$

Thm.  $\Downarrow$

indépendance définition  $\equiv P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$

## corrélation

variables aléatoires  $X$  et  $Y$  discrètes

$$\sigma_{XY} \triangleq \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \sum_{k=0}^{+\infty} \sum_{l=0}^{+\infty} (k - \mu_X)(l - \mu_Y) p_X(k) p_Y(l)$$

$p(k, l)$

indépendance  $\Leftrightarrow p(k, l) = p_X(k) p_Y(l)$

## variable aléatoire discrète complexe

En introduisant une variable aléatoire discrète complexe

$$Z = X + j Y$$

avec  $j = \sqrt{-1}$  la variable imaginaire et  $X$  et  $Y$  des variables aléatoires discrètes à valeurs dans les entiers naturels  $\mathbb{N}$  on peut définir

$$\mathbb{E}[Z] \triangleq \mathbb{E}[X] + j \mathbb{E}[Y]$$

si, et seulement si,  $\mathbb{E}[X] \in \mathbb{R}$  et  $\mathbb{E}[Y] \in \mathbb{R}$  sont définis.



## fonction génératrice

En introduisant le paramètre  $s \in \mathbb{C}$ ,  $|s| < 1$ , on a comme cas particulier de variable aléatoire discrète complexe

$$Z = s^X$$

## définition

Soit une seule variable aléatoire discrète  $X$  qui donne des nombres naturels  $X \in \mathbb{N}$ . La fonction génératrice est la fonction de  $s \in \mathbb{C}$  et  $|s| \leq 1$  qui associe à ce nombre (le paramètre  $s \in \mathbb{C}$ ) l'espérance mathématique de la variable aléatoire discrète complexe  $Z = s^X$

$$g(s) \triangleq \mathbb{E}[s^X] = \sum_{k=0}^{\infty} s^k p_k \quad \forall s \in \mathbb{C}, |s| \leq 1$$

A l'intérieur du cercle unité, la série converge uniformément et de manière absolue pour  $s$  dans le cercle unité car  $\sum_{k=1}^{\infty} p_k |s^k| \leq \sum_{k=1}^{\infty} p_k = 1$ .

# produit de convolution

## entre distributions discrètes

### définition

$$h(k) \triangleq f * g \triangleq \sum_{l=0}^k f(l)g(k-l)$$

## la convolution

Variables aléatoires  $X$  et  $Y$  et soit  $Z$  la somme des variables aléatoires

$$Z = X + Y$$

théorème de convolution

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} f_Y(y) f_X(z-y) dy$$

On note de manière compact cette formule par le produit de convolution

$$f_Z = f_X * f_Y$$

## fonction caractéristique d'une variable aléatoire

La définition de l'espérance mathématique peut-être étendue au cas où la variable  $X$  est une variable aléatoire complexe

$$X = X_1 + j X_2$$

où  $X_1$  et  $X_2$  sont des variables aléatoires réelles. On a

$$\mathbb{E}[X] = \mathbb{E}[X_1] + j\mathbb{E}[X_2]$$

### définition

La fonction caractéristique d'une variable aléatoire  $X$  avec une densité de probabilité  $f_X$  est définie par

$$\phi_X(u) \triangleq \mathbb{E}[e^{iux}] = \int_{-\infty}^{+\infty} e^{iux} f_X(x) dx$$

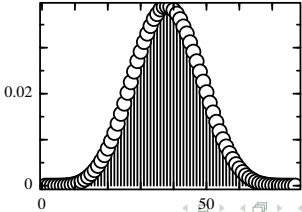
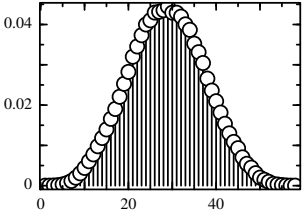
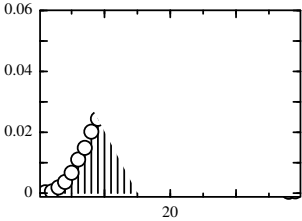
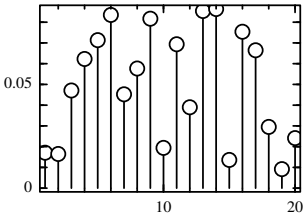
C'est la transformée de Fourier de la densité de probabilité.

# théorème des sommes de variables aléatoires

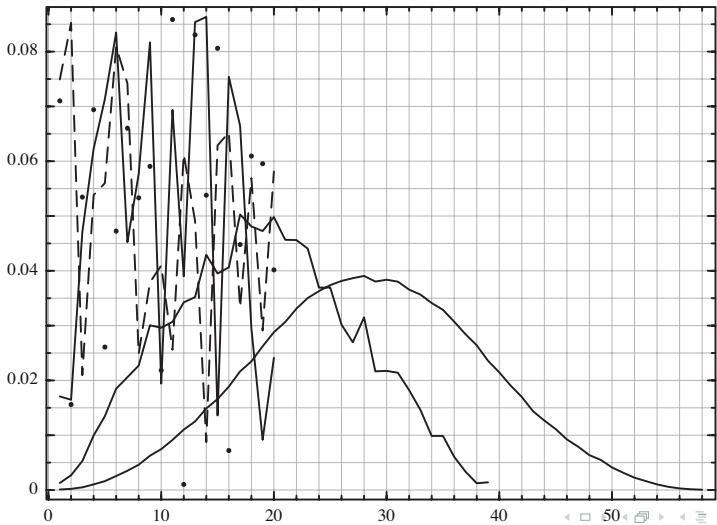
$$X = X_1 + X_2$$
$$f_X(x) = \int_{-\infty}^{\tau} f_{X_1}(\tau) f_{X_2}(x - \tau) d\tau$$

# convolution

## illustration



convolution  
illustration



# approximation de la distribution discrète binomiale

## Théorème de De Moivre - Laplace

Si la probabilité de  $n$  expériences de Bernouilli  $p$  est constante, la probabilité d'avoir  $k$  succès parmi  $n$  durant ces expériences tend vers une distribution normale, plus précisément

$$\sqrt{npq} \mathcal{B}in(n, k) : \frac{1}{2\pi} e^{-\frac{1}{2}x^2} \rightarrow 1$$

lorsque  $n \rightarrow \infty$  et de manière uniforme pour tout  $k$  pour lequel la variable  $x$  est contenue dans un certain intervalle fini. Ici la correspondance

$$x = \frac{k - np}{\sqrt{nkq}}$$

Rappel :  $q = 1 - p$



# approximation de la distribution discrète binomiale

$$\mathcal{B}in(n, k) \sim \frac{1}{\sqrt{2\pi kp(1-p)}} e^{-\frac{1}{2} \left( \frac{m-np}{\sqrt{np(1-p)}} \right)^2}$$