Variables Aléatoires, Lois de Probabilité Discrète & Continue

I- Définition

En application de ce cours, on va toujours décrire une opération précise qui mène à un résultat aléatoire. Cette opération sera appelée épreuve, et ses résultats aléatoires seront des évènements élémentaires.

Une **variable aléatoire** est donc une **épreuve** aboutissant à des **évènements élémentaires** *(qui sont des nombres)*

➔ *Exemple : « Lancer un dé » est une variable aléatoire (car les évènements élémentaires sont des nombres, de 1 à 6). Par contre, « tirer une carte au hasard » n'est pas une variable aléatoire car les différentes cartes ne sont pas des nombres*

Si le résultat fait partie d'un ensemble fini, alors c'est une **variable aléatoire discrète** *(ex : le nombre de QRU réussis au concours)*

Si le résultat fait partie de ℝ ou d'un intervalle de ℝ, alors c'est une **variable aléatoire continue,** ou variable à densité *(ex : le dosage de la glycémie)*

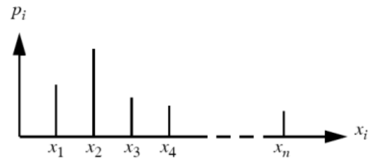
II- Variable Aléatoire Discrète

La loi de probabilité d'une variable aléatoire X discrète est définie en donnant l'ensemble des valeurs *(p1, p2, pn)* qui sont les probabilités de ses différentes éventualités *(x1, x2, xn)*

soit $p_i = P(X = x_i)$ donc $0 \leq p_i \leq 1$ *et* $\sum_{i=1}^{n} p_i = 1$

On peut représenter cette loi par une table ou un diagramme en bâtons :

x_1	x_2	$\dots$	x_n	$\dots$
p_1	p_2	$\dots$	p_n	$\dots$



A) Paramètres

Moyenne : La moyenne **µ** de la variable aléatoire X est la valeur moyenne des résultats que l'on obtiendrait en répétant indéfiniment l'épreuve. C'est un indicateur de **position.**

Espérance : L'espérance **E(X)** c'est la moyenne, mais en statistiques. Elle traduit la tendance centrale de la variable aléatoire. C'est un indicateur de **position**.

Théorèmes de l'espérance :

Soit X une variable aléatoire et k un nombre réel

$\boldsymbol{E(X + k) = E(X) + k}$

$\boldsymbol{E(kX) = k\,E(X)}$

Soient X et Y deux variables aléatoires

$\boldsymbol{E(X + Y) = E(X) + E(Y)}$

Variance & Écart-type : La variance σ^2 est la moyenne des carrés des écarts entre X et sa moyenne **μ**. L'écart-type **σ** de la distribution est la racine carrée de la variance. Ce sont des indicateurs de **dispersion**

Variable Centrée Réduite :

« Centrée » consiste à soustraire la moyenne **μ** de la variable à chacune de ses valeurs initiales

« Réduite » consiste à diviser toutes les valeurs que prend la variable par son écart-type **σ**

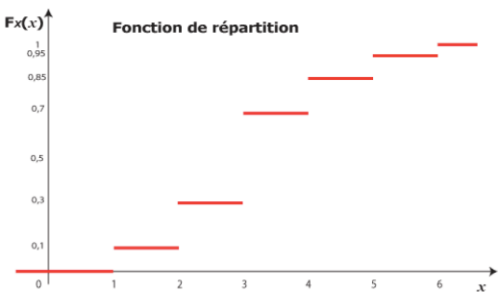
Soit X, variable aléatoire de moyenne **μ** et d'écart-type **σ.** Alors Y est la variable centrée réduite définie comme :

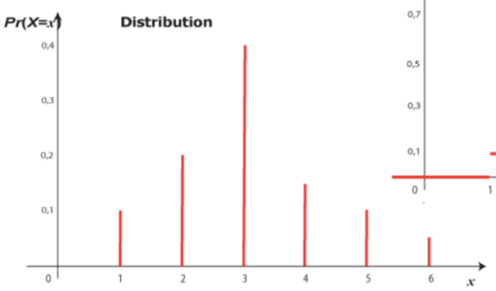
$$Y = \frac{X-\mu}{\sigma}$$

Elle a 2 propriétés : E(Y) = 0 ; Var(Y) = 1

Fonction de Répartition : C'est une fonction cumulative car on somme l'ensemble des probabilités (p_i) des x_i survenus avant x. F(x) sera toujours une fonction monotone croissante. À bien différencier de la fonction de distribution qui a la forme d'un diagramme en bâtons.

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{et} \quad \lim_{x\to+\infty} F(x) = 1$$





III- Lois de Probabilité Discrète

A) Loi de Bernoulli B(p)

Épreuve de Bernoulli : Expérience aléatoire aboutissant soit à un succès, soit à un échec

Paramètres:

p : probabilité d'un succès
q = (1 – p) : probabilité d'un échec
X : v-a « nb de succès » pendant l'épreuve

$$P(X = k) = p^k(1-p)^{1-k} = p^k q^{1-k} \qquad \text{pour} \qquad k \in \{0,1\}$$

- $\mu = p$
- $\sigma^2 = p(1-p) = pq$

➔ *Exemple : On a une boite avec 2 boules rouges et 6 noires. Soit l'événement « tirer une boule noire » le succès avec une probabilité* $p = \frac{6}{8} = \frac{3}{4}$
(⚠ ici le succès est bien tirer une boule).

Si on tire une boule dans la boite au hasard on suit la loi de Bernoulli et on a :
$P(X=0) = \frac{3^0}{4}\frac{1^1}{4} = \frac{1}{4}$ $\quad P(X=1) = \frac{3^1}{4}\frac{1^0}{4} = \frac{3}{4}$

B) Loi Binomiale B(n,p)

C'est une épreuve répétée de Bernoulli : n essais indépendants conduisant à « succès » ou « échec »

Paramètres:

n : nombre d'essais indépendants
p : probabilité d'un succès
q = (1 – p) : probabilité d'un échec
X : v-a « nb de succès » pendant l'épreuve

$$P(X=k) = C_n^k p^k (1-p)^{n-k} = C_n^k p^k q^{n-k} \qquad \text{pour} \qquad 0 \leq k \leq n$$

- $\mu = np$
- $\sigma^2 = np(1-p) = npq$

Rappel : $C_n^k = \dfrac{n!}{k!\,(n-k)!}$

- Pour p = 0,5 la forme du diagramme des probabilités d'une distribution normale est symétrique.
  Si p > 0,5 on dit que la distribution est « asymétrique positive ».
  Si p < 0,5 elle est « asymétrique négative ».

  Quand n est grand, et p pas trop proche de 0 ou 1, la forme du diagramme de distribution devient symétrique et la loi binomiale tend vers la loi normale.

- Pour étudier la totalité d'une population on procède à un échantillonnage de celle-ci. De ce fait, grâce à un tirage au sort (=randomisation), on constitue plusieurs échantillons de la population auxquels on appliquera notre loi binomiale

.

Le TAS peut être :
Non exhaustif (=indépendant) : Les éléments sélectionnés sont remis dans l'échantillon après le tirage. p reste constant.
Exhaustif (=dépendant des autres tirages) : il n'y a pas de remise. p varie au fils des tirages.

- On peut définir **le Taux de sondage** = $\frac{n}{N}$

(n taille de l'échantillon, N effectif de la population)

Si $\frac{n}{N} \leq 0{,}10$ on applique la loi Binomiale

Si $\frac{n}{N} \geq 0{,}10$ on applique la loi Hypergéométrique

➔ *Exemple : On reprend une boite avec cette fois 4 boules rouges et 6 noires. Soit l'événement « tirer une boule noire » le succès. On tire trois fois une boule dans cette boite et chaque tirage est indépendant des autres. On cherche la probabilité de tirer une boule noire soit P(X=1).*
On a $p = \frac{6}{10}$ et n = 3

$$P(X=1) = C_3^1 \left(\frac{6}{10}\right)^1 . \left(\frac{4}{10}\right)^2 = \frac{3!}{2!\,1!} \,.\, 0{,}6 \,.\, 0{,}16 = 3 \,.\, 0{,}096 = 0{,}288$$

C) Loi Hypergéométrique H(N,D,n)

Soit une population de **N** individus parmi lesquels **D** ont un caractère donné. On prélève un échantillon **n** de cette population **N**. Les individus de l'échantillon sont tirés simultanément (l'ordre de tirage n'a pas de sens) et sans remise.

Paramètres:

N : effectif de la population
D : nb de personnes présentant le caractère étudié dans la population
D/N : probabilité **p** d'avoir le caractère étudié

$$P(X=k)=\frac{C_D^k \times C_{N-D}^{n-k}}{C_N^n} \quad \text{avec} \quad min(0;n-D) \le k \le max(n;D)$$

- **Moyenne** : $\mu = \frac{nD}{N} = np$
- **Variance** : $\sigma^2 = \frac{nD}{N} \times \frac{N-D}{N} \times \frac{N-n}{N-1} = (\frac{N-n}{N-1})np(1-p) = (\frac{N-n}{N-1})npq$

➔ *Exemple : Dans une usine, il y a 1000 machines dont 200 ont des défauts. On tire au sort 300 machines dans cette population. Quelle est la probabilité que la moitié de cet échantillon ait des défauts ?*

$$P(X=150)=\frac{C_{200}^{150}C_{800}^{150}}{C_{1000}^{300}}$$

D) Loi Géométrique G(p)

On répète des épreuves de Bernoulli jusqu'à l'obtention d'un succès.

Paramètres:

X : v-a « nb d'essais nécessaires » à l'obtention du 1er succès
p : probabilité d'un succès
q : probabilité d'un échec

$$P(X=k)=p(1-p)^{k-1}=pq^{k-1}$$

- $\mu = 1/p$
- $\sigma^2 = \frac{1-p}{p^2}$

➔ *On lance un dé à six faces jusqu'à obtenir un « 6 ». La probabilité d'obtenir un 6 au bout de 3 essais est* $P(X=3)=\frac{1}{6}.(\frac{5}{6})^2$

E) Loi de Poisson P(λ)

Utilisée pour déterminer la probabilité qu'un certain nombre d'évènements se réalisent par unité de temps (ou de volume, de surface)

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!} \qquad \lambda > 0$$

- $\mu = \sigma^2 = \lambda$

Paramètres :

λ : taux moyen avec lequel un évènement particulier se produit en général.
X : variable aléatoire qui donne le nombre d'évènement particulier qui se produisent dans la situation étudiée.

Cette loi est très utilisée dans les domaines de la qualité, la sécurité et la fiabilité.

➔ *Exemple : Dans le cabinet d'un ophtalmologue, on a en moyenne quatre consultations en deux heures. Quelle est la probabilité d'avoir une consultation au cours d'une heure ? On a λ = 4 hospitalisations pour 2h or ici on cherche une probabilité relative à 1h donc on prend λ = 2 hospitalisations pour 1h.*

$$P(X=1)=\frac{2^1.e^{-2}}{1!}=2e^{-2}$$

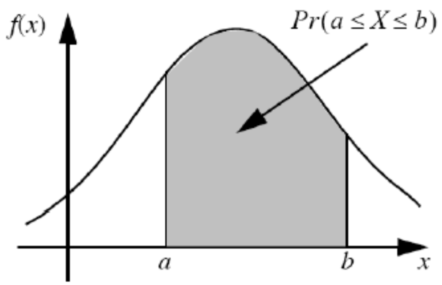
IV- Variable Aléatoire Continue

Ce qui caractérise une variable aléatoire continue, c'est qu'elle a une probabilité nulle d'être égale à un nombre donné

P(X=3) = 0 ; P(X=6,66) = 0 ; P(X=$\sqrt{2}$) = 0

On utilisera donc des intervalles : P(a ≤ X ≤ b) ≠ 0

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$



La densité de probabilité

C'est une fonction utilisée pour définir la loi de probabilité de X

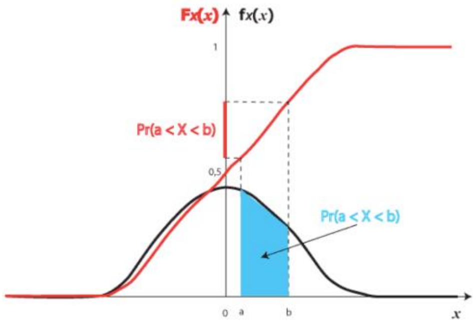
P(a ≤ X ≤ b) est l'aire sous la courbe

Fonction de Répartition

Elle est toujours croissante, monotone et continue.

Partant de 0 pour x → -∞ .
Atteignant 1 pour x → +∞ .

(en rouge la fonction de répartition, en noir la fonction de densité)



V- Lois de Probabilité Continue

A) Loi Exponentielle E(λ)

Fonction de densité : $f(x) = \lambda e^{-\lambda x}$ avec $\lambda > 0$ et $x \geq 0$

Paramètres : λ = taux de défaillance instantané

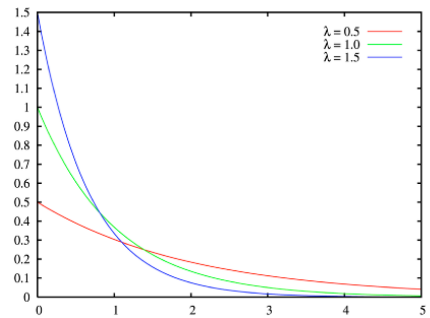
- $\mu = 1 / \lambda$
- $\sigma^2 = 1 / \lambda^2$

Fonction de répartition: $f(x) = P(X \geq x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$

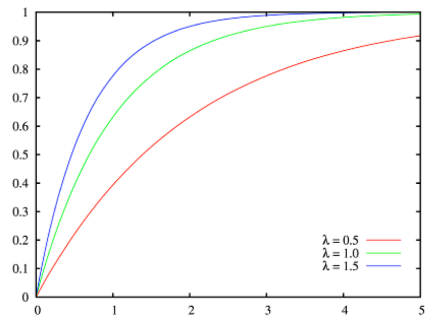
Si un évènement se réalise selon une loi de Poisson de paramètre λ, le temps entre deux réalisations consécutives de l'évènement considéré est distribué selon une loi exponentielle de paramètre 1/λ

Elle est utilisée pour décrire un processus de mortalité dans lequel le « risque instantané » de décès (ou taux de défaillance) est constant (durée de vie de composants, d'équipements)

Densité de probabilités (f):



Fonction de répartition (F):



➔ *Exemple : La radioactivité peut être décrite ainsi et à chaque instant le taux de radioactivité (la probabilité de désintégration) est constant. C'est-à-dire que peu importe le temps écoulé ou le nombre d'atomes déjà désintégrés la probabilité que le noyau se désintègre reste la même pour un instant donné.*

B) Loi Uniforme U(a ;b)

Fonction de densité : $f(x) = \frac{1}{b-a}$ si $x \in [a, b]$

$f(x) = 0$ si $x \notin [a, b]$ avec $\lambda > 0$ et $x \geq 0$

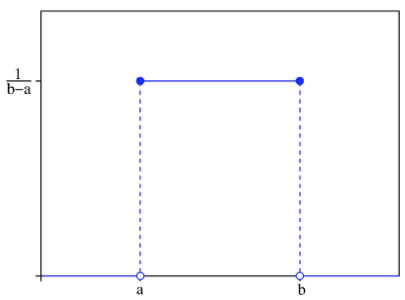
Paramètres : intervalle [a,b] ∈ ℝ

- $\mu = (a+b) / 2$
- $\sigma^2 = (b-a)^2 / 12$

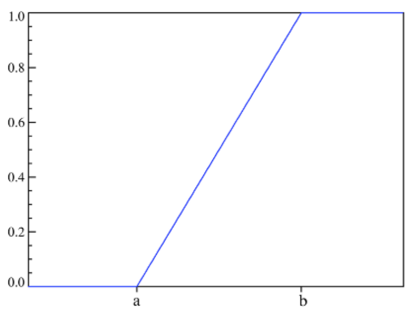
Toutes les valeurs ont la même probabilité *(par exemple un lancer de dé, on a 1/6 de chance pour chaque valeur)*

Fonction de densité : f(x) est donc constante sur [a ; b] et nulle en dehors.
Fonction de répartition : F(x) part de 0 pour atteindre 1 de manière linéaire

Densité de probabilités (f):



Fonction de répartition (F):



➔ *Exemple : Le boulanger prépare toujours ses croissants avant l'ouverture de la boulangerie. La préparation des croissants se déroule de 06 :00 à 07 :00. Soit l'événement « les croissants sont prêts ». Sa probabilité est définie par une loi uniforme [06 :00 ; 07 :00]. La probabilité que les croissants soient prêts au moins quinze minutes avant l'ouverture est :*

$$\int_{6}^{6,75} \frac{1}{7-6}\,dx = \frac{6,75-6}{7-6} = 0,75$$

Il y a trois quarts de chance qu'ils soient prêts 1/4 d'heure avant !

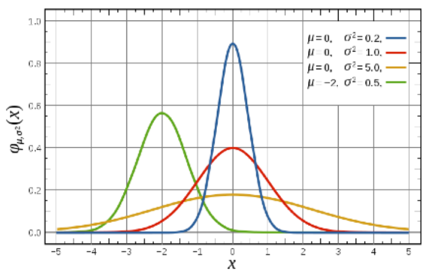
C) Loi Normale N(μ ; σ)

Fonction de densité : $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ pour pour $-\infty \geq x \geq +\infty$

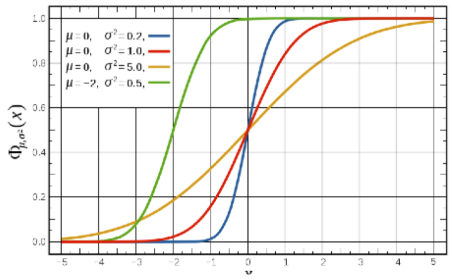
Paramètres : μ et σ , moyenne et écart-type de X

La densité de probabilité d'une v-a normale est symétrique autour de μ et a deux points d'inflexion aux abscisses μ - σ et μ + σ

Densité de probabilités (f):



Fonction de répartition (F):



Valeurs limites importantes à savoir

- il y a 10 chances sur 100 pour que $X < \mu - 1,65\sigma$ ou $X > \mu + 1,65\sigma$
- il y a 5 chances sur 100 pour que $X < \mu - 1,96\sigma$ ou $X > \mu + 1,96\sigma$
- il y a 1 chance sur 100 pour que $X < \mu - 2,58\sigma$ ou $X > \mu + 2,58\sigma$
- il y a 1 chance sur 1000 pour que $X < \mu - 3,30\sigma$ ou $X > \mu + 3,30\sigma$

La loi normale sert tout le temps ! Notamment pour les appareils de mesure *(comme en laboratoire d'analyses médicales)*, ça permet de quantifier et les erreurs de mesure.

➔ Exemple :
Le poids des PACES se répartit selon une loi Normale N(μ ;σ) avec μ=60Kg et σ=10Kg. Ainsi 95% des PACES pèsent entre (μ-1,96σ) = (60- 1,96x10) = (60-19.6) = 40,4 et (μ+1,96σ) = (60+ 1,96x10) = (60+19.6) = 79,6.
La taille « X » des hommes adultes suit une loi Normale de moyenne μ = 180 cm et d'écart type σ = 6 cm.
La proportion d'homme dont la taille est comprise entre 174 cm (μ-σ) et 186 cm (μ+σ) est de 68%.
La proportion d'homme dont la taille est inférieure à 168,2 cm (μ - 1,96σ) ou supérieure à 191,8 cm (μ + 1,96σ) est de 2,5% + 2,5% = 5%
La proportion d'homme dont la taille est supérieure à 189,9 cm (μ + 1,65σ) est de 5%.
La proportion d'homme dont la taille est inférieure à 189,9 cm (μ + 1,65σ) est de 95%.

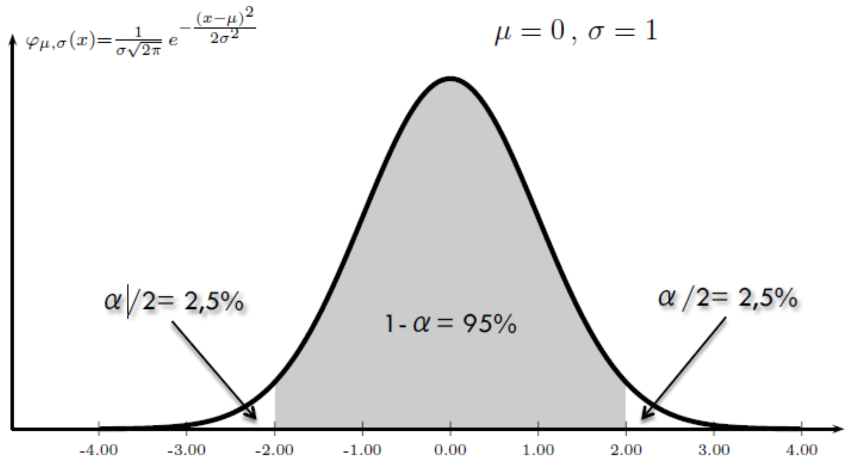
D) Loi Normale Centrée Réduite

La loi normale centrée réduite est une loi normale de moyenne 0 et de variance 1

Centrée → Autour de la moyenne μ = 0

Réduite → Ayant une variance σ = 1

N(μ ; σ) => N(0 ; 1)



Ce changement de variable est très utile en pratique, on peut ramener n'importe quel problème de probabilité à distribution normale (qui suit donc une loi normale) à un seul cas : celui de la loi normale centrée réduite.

E) Approximations

Certaines lois peuvent être approximées par d'autres selon certaines conditions :

La **loi Binomiale** peut être **approximée par** une **loi de Poisson** si :

- $N > 50$
- $p \leq 0{,}10$
- $np \leq 5$

Dans ce cas, **B(n;p) → P(λ=np)**

La **loi Binomiale** peut être **approximée par** une **loi Normale** si :

- $np \geq 5$
- $nq \geq 5$

Dans ce cas, **B(n;p) → N(np;√npq)**

La **loi de Poisson** peut être **approximée par** une **loi Normale** si :

- $\lambda > 25$

Dans ce cas, **P(λ) → N(λ;√ λ)**

Il est temps pour moi de passer à la dédicace !

À Melina, t'es une battante hors pair, j'ai grave confiance en ta réussite

À Anthony, fonce tout droit t'as la motiv t'as le talent, t'as tout ce qu'il faut

Kévin je sens l'année de qualité arriver, démonte tout !

Mes pioux, Ines, Timotey, Élodie, Terésa & Andréa, vous êtes des champions <3

Ma co-marraine Marion, love

Mes co-tuts de l'extrême Léa & Théo, vous êtes beaux, talentueux, je dirai même splendides <3

À des tuteurs exceptionnels, on va en nommer quelques-uns : Je nomme SLAASH (histo), Kanye Western-Blot (biocell), Léaccouchement & Grohl (biostats, double dédicace les frères), Tristampax (chimie o), Feenex (ssh), rambo (ue8). Ya pas à dire, zêtes chouettes

Mes vieilles si qualitatives love x3

Mr Poisson qui a inventé la loi de Poisson

Mr Bernoulli qui a inventé la loi de Bernoulli

Mme Normale-Centrée-Réduite qui a inventé la loi Normale centrée réduite

À tous les Hommes bons dans ce monde