

WERNER SÖRENSEN

CALCUL  
DES  
PROBABILITÉS

Manuscrit 1981/82

CALCUL DES PROBABILITES

§0. Introduction. page 1

§1. Modélisation d'une expérience. page 3

1. Expériences.
2. Evénements liés à une expérience.
3. Opérations logiques sur les événements.
4. Tribus.
5. Expériences aléatoires.

§2. Le modèle classique. page 7

1. Définition.
2. Deux problèmes à l'origine du calcul des probabilités.
3. Rappel d'analyse combinatoire.
4. Urnes.
5. Loi hypergéométrique.
6. Echantillons: loi binomiale.
7. Problèmes d'application.
8. Insuffisance du modèle classique.

§3. Axiomatique du calcul des probabilités. page 25

1. Définition d'un espace probabilisé.
2. Exemples.
3. Conséquences élémentaires des axiomes.

§4. Probabilités conditionnelles - Evénements indépendants. 3

1. Exemples et définitions.
2. Propriétés des probabilités conditionnelles.
3. Evénements indépendants.

§5. Schéma de Bernoulli. page 41

1. Modélisation.
2. Loi binomiale.
3. Loi multinomiale.

4. Approximation normale de la loi binomiale.
5. Théorème de Poisson.

§6. Variables aléatoires. page 55

1. Définition ; exemples.
2. Loi de probabilité d'une variable aléatoire.
3. Fonction de répartition d'une variable aléatoire.
4. Variables aléatoires vectorielles.
5. Fonction d'une ou plusieurs variables aléatoires.
6. Application 1 : Erreurs de mesure.
7. Application 2 : Rayonnement radioactif.

§7. Moments d'une variable aléatoire. page 87

1. Espérance d'une v.a.
2. Variance et covariance.
3. Fonction génératrice d'une v.a. à valeurs dans  $\mathbb{N}$ .
4. Application : processus de ramification.
5. Fonctions caractéristiques : résultats.

§8. Théorèmes limites. page 112

1. Deux inégalités.
2. Lois faibles des grands nombres.
3. Théorème central-limite.

§9. Mouvement brownien et processus de Wiener-Lévy. p. 126

1. Introduction et définition.
2. Quelques propriétés : loi du max. - oscillations - loi de l'Arc sinus - temps d'attente du niveau 1.

§0. Introduction

L'étude de probabilités se présente sous 3 aspects :

1. Philosophie des probabilités : on étudie le concept de probabilité. Le terme de "probabilité" a depuis toujours deux sens, ce qu'on peut illustrer par les exemples suivants :

Exemple A. Deux personnes X et Y considèrent une pièce de monnaie ; X dit que la probabilité de "pile" est  $\frac{1}{2}$ , Y dit que c'est  $\frac{3}{4}$ .

L'expérience permet de trancher : soit en mesurant la symétrie, l'homogénéité de la pièce, soit en la lançant un grand nombre de fois. La croyance en la possibilité d'un accord prouve que l'on attribue à la probabilité de "pile" une existence objective, indépendante de l'obscurité.

Exemple B. La équipe de football Xamax et Malup se rencontrent mardi prochain. Deux personnes X et Y discutent du match ; X dit que la probabilité que Xamax gagne est  $\frac{1}{2}$ , Y dit que c'est  $\frac{3}{4}$ . Il est impossible de départager X et Y. Quel que soit le résultat du match X et Y pourront continuer de penser qu'ils avaient raison.

On voit que le terme "probabilité" a deux acceptations :

A) la probabilité est une propriété intrinsèque des choses ; on peut la déterminer soit par des raisons de symétrie, soit en répétant un grand nombre de fois une expérience (probabilité statistique).

B) la probabilité est le degré de crédibilité ; c'est une mesure de la confiance d'un observateur en la réalisation d'un événement.

Le point de vue A) a donné lieu à des tentatives d'axiomatisation basées sur la stabilité des fréquences : théorie des "collectifs" de von Mises, puis Martin-Löf.

Le point de vue B) a donné naissance à des théories "personnelles" de probabilités (de Finetti) et à la "théorie de la décision dans l'incertain" (Savage, Wald).

2. Le Calcul des probabilités mathématique : c'est la création, l'étude et l'application des modèles mathématiques des phénomènes aléatoires (modèles stochastiques). Dans ces modèles, les probabilités apparaissent comme des symboles, des paramètres (cf. masses en mécanique).

Remarque. Les deux premiers aspects ont été longtemps indissociés et les traités de calcul des probabilités commençaient tous par les définitions possibles du terme "probabilité". Ce n'est que depuis l'axiomatisation de Kolmogorov (1933) que la séparation est nette.

3. la Statistique : son but est de déterminer expérimentalement les probabilités qui figurent dans les modèles stochastiques. La statistique mathématique peut être considérée comme une idéalisation de la notion d'expérience.

A lire : Encyclopaedia Universalis : Statistique ; Probabilités.

## §1. Modélisation d'une expérience.

### 1.1. Expérience.

Expérience =  $\left\{ \begin{array}{l} \text{— déclenchement d'un phénomène observé} \\ \text{— observation du résultat parmi un ensemble} \\ \text{précisé à l'avance de possibilités qui s'excluent} \\ \text{deux à deux.} \end{array} \right.$

Une expérience est décrite mathématiquement par un ensemble  $\Omega$  (espace des épreuves ou univers des éventualités) qui schématise les résultats possibles; à chaque réalisation de l'expérience correspond son résultat, appelé épreuve<sup>(\*)</sup> ou éventualité, qui est associé à un point  $\omega \in \Omega$ .

Image : Expérience = machine à fournir des points de  $\Omega$ .

- Exemples.
1. Jet d'un dé :  $\Omega = \{1, 2, 3, 4, 5, 6\}$
  2. Battage d'un jeu de 52 cartes :  $\Omega = \mathcal{C}_{52}^2$ .
  3. Mesure d'une durée de vie :  $\Omega = \mathbb{N}$  ou  $\mathbb{R}_+$ .
  4. Suite infinie de lancers d'une pièce :  $\Omega = \{0, 1\}^{\mathbb{N}}$ .
  5. Enregistrement de la température dans l'intervalle de temps  $[0, T]$  :  $\Omega = \mathcal{C}_{\mathbb{R}}([0, T])$ .

### 1.2. Événements liés à une expérience.

C'est un événement dont on peut dire, ou on du seul résultat de l'expérience, s'il est réalisé ou non.

On identifie des événements qui, bien que définis par des phrases différentes, sont réalisés ou non réalisés en un même temps. Par ex. "obtenir 0" et "obtenir 7" en lançant un dé.

On peut alors identifier un événement  $A$  au sous-ensemble de  $\Omega$  formé des épreuves qui le réalisent :

$$A \approx \{\omega \in \Omega \mid A \text{ est réalisé si le résultat de l'exp. est } \omega\}$$

(\*) "épreuve" a le sens de "résultat d'un essai", comme en typographie.

Ex. 1. "obtenir un nombre pair" = {2, 4, 6}

Ex. 5. "la température reste inférieure à 10°C"  
= { f ∈ ℚ<sub>R</sub>([0,T]) | f(t) < 10, ∀ t ∈ [0,T] }

Deux événements particuliers :

- l'événement impossible ≈ ∅
- l'événement certain ≈ Ω.

1.3. Opérations logiques sur les événements.

a) On associe à tout événement A son contraire "non A", noté  $\bar{A}$ .  
Donc  $\bar{\bar{A}}$  est réaliséssi A n'est pas réalisé. Donc  $\bar{\bar{A}} \approx A$ .

b) A, B étant deux événements, "A et B" désigne l'événement qui est réaliséssi A est réalisé et B est réalisé.  
On a: "A et B" ≈ A ∩ B (Noté aussi: AB).

Remarque : A et B sont incompatiblesssi A ∩ B = ∅.

c) A, B étant deux événements, "A ou B" désigne l'événement qui est réaliséssi l'un au moins de A ou B est réalisé.  
On a: "A ou B" ≈ A ∪ B.

Remarque : si A ∩ B = ∅ on note aussi A ∪ B = A + B.

Généralisation à une suite (finie ou infinie) d'événements:  
 $\bigcap A_n, \bigcup A_n, \sum A_n$  si  $A_n \cap A_m = \emptyset$  ( $\forall n \neq m$ ).

d) A, B étant deux événements, on dit que A implique B si la réalisation de A entraîne la réalisation de B.  
On a: A implique Bssi A ⊂ B.

Terminologie probabiliste	Terminologie ensembliste	Notation
Espace des épreuves ou univers	Référentiel	Ω
Épreuve	Point ou élément	ω
Événement	Partie	A
Événement contraire	Complémentaire	$\bar{A}$ ( $A^c, A'$ )
Événement certain	Espace entier	Ω
Événement impossible	Partie vide	∅
Événement élémentaire	Singleton (!)	{ω}
Événements incompatibles	Parties disjointes	A ∩ B = ∅
Et	Intersection	∩ (ou ·)
Ou (non exhaustif)	Réunion	∪
	(Somme si A ∩ B = ∅)	(+)
Implication	Inclusion	A ⊂ B
Système exhaustif	Partition	$\sum A_i = \Omega$
Ou (exclusif)	Différence symétrique	Δ

1.4. Tribe.

A une expérience donnée on a associé le couple (Ω, P(Ω)), formé de l'ensemble des épreuves et de l'ensemble des événements.

Souvent, soit pour des raisons pratiques, soit pour des raisons de technique mathématique, on ne considère qu'une sous-famille  $\mathcal{O} \subset P(\Omega)$  d'événements.

Mais alors il est logique de demander que  $\mathcal{O}$  contienne ∅ et Ω, et qu'elle soit stable pour les opérations fondamentales: contraire, et, ou.

Pour le développement de la théorie, il est en outre indispensable (lorsque Ω n'est pas fini) de supposer que  $\mathcal{O}$  soit stable par réunion et intersection dénombrable. Autrement dit,  $\mathcal{O}$  est une tribu.

Définition:  $\mathcal{O} \subset \mathcal{P}(\Omega)$  est une tribu (ou  $\sigma$ -algèbre)

si (1)  $\emptyset, \Omega \in \mathcal{O}$  (2)  $A \in \mathcal{O} \rightarrow \bar{A} \in \mathcal{O}$

(3)  $A_n \in \mathcal{O}, \forall n \in \mathbb{N} \rightarrow \bigcup A_n \in \mathcal{O}$

NB. Par (3) on a aussi  $\bigcap A_n \in \mathcal{O}$ .

En résumé, une expérience se décrit mathématiquement par un couple  $(\Omega, \mathcal{O})$  où  $\Omega$  est un ensemble et  $\mathcal{O}$  une tribu de parties de  $\Omega$ .

### 1.5. Expériences aléatoires.

On dit qu'une expérience est déterministe si, lorsqu'on la répète, on s'aperçoit que le résultat dépend de paramètres qu'on peut contrôler. On pourra donner des lois et prévoir les résultats (Ex. Température d'ébullition de l'eau).

Les autres expériences sont aléatoires (Ex: Jet d'un dé).

Remarques. 1) Question d'échelle: un courant électrique dans un métal est déterministe à l'échelle macroscopique et aléatoire à l'échelle atomique.

2) Une expérience déterministe est un peu aléatoire par le fait qu'on ne peut pas contrôler exactement tous les paramètres. Inversement une expérience aléatoire pourrait être considérée comme déterministe si tous les paramètres étaient connus.

→ Pour décrire une expérience aléatoire, il faut ajouter au modèle  $(\Omega, \mathcal{O})$  un objet qui permette de parler de la probabilité des événements  $A \in \mathcal{O}$ .

Commençons par un cas particulier.

C'est donc à l'observateur à décider s'il considère son expérience comme aléatoire ou déterministe.

## §2 Le modèle classique.

### 2.1. Définition du modèle classique.

C'est le modèle où  $\Omega$  est un ensemble fini,  $\mathcal{O} = \mathcal{P}(\Omega)$ , et où la probabilité  $P(A)$  d'un événement  $A$  est définie par

$$P(A) = \frac{N(A)}{N(\Omega)}$$

où  $N(\cdot)$  désigne le nombre d'éléments.

Remarques 1) Pour tout  $\omega \in \Omega$  on a donc  $P(\{\omega\}) = \frac{1}{N(\Omega)}$ . On voit donc que le modèle classique sera applicable lorsqu'on peut supposer que tous les événements élémentaires ont la même probabilité (la phrase classique est: "lorsque tous les cas possibles sont également possibles"). Cela concerne la plupart des jeux de hasard (dés, cartes, ...).

2) Lorsque, dans le langage courant, on dit que l'on choisit au hasard un élément dans un ensemble fini, on sous-entend presque toujours que les choix possibles sont équi-probables; on se place donc dans le modèle classique.

3) L'espace  $\Omega$  étant défini (c'est souvent la <sup>seule</sup> difficulté !), les calculs de probabilités sont ramenés à des calculs de puissances d'ensembles; ces calculs sont le plus souvent de nature combinatoire.

### 2.2. Deux problèmes à l'origine du calcul des probabilités.

Ce sont les jeux de hasard qui ont donné naissance au Calcul des probabilités. M. Loève: "le premier théoricien fut sans doute celui qui le premier pipa ses dés."

Un des premiers problèmes dont on a le témoignage écrit fut posé à Galilée (1564-1642) par un de ses amis; c'est le problème du jeu de passe-dix (cf. J. Bertrand - Calcul des probabilités - Paris 1888).

Problème du jeu de passe-dix.

On jette 3 dés et on gagne si la somme des points dépasse 10 (i.e. vaut 11, 12, ..., 18).

L'ami de Galilée s'étonne de gagner plus souvent avec 11 points qu'avec 12, ce qui lui semble en contradiction avec ses calculs! Il raisonne en effet de la façon suivante:

- pour obtenir 11 points, il faut:

- |                     |                     |
|---------------------|---------------------|
| 1) un 6, un 4, un 1 | 4) un 5, un 4, un 2 |
| 2) un 6, un 3, un 2 | 5) un 5, un 3, un 3 |
| 3) un 5, un 5, un 1 | 6) un 4, un 4, un 3 |

- pour obtenir 12 points, il faut:

- |                     |                      |
|---------------------|----------------------|
| 1) un 6, un 5, un 1 | 4) un 5, un 5, un 2  |
| 2) un 6, un 4, un 2 | 5) un 5, un 4, un 3  |
| 3) un 6, un 3, un 3 | 6) un 4, un 4, un 4. |

Donc dans les 2 cas il y a 6 possibilités. Or l'ami de Galilée a obtenu dans une série d'expériences 1080 fois le total 11 et 1000 fois le total 12. D'où son étonnement.

Réponse de Galilée - exprimée dans notre langage:

$$\Omega = \{1, 2, 3, 4, 5, 6\}^3 \quad (\text{on suppose les 3 dés discernables!})$$

$$\text{i.e. } \Omega = \{\omega = (\omega_1, \omega_2, \omega_3) \mid 1 \leq \omega_i \leq 6 \text{ (N)}\}. \quad N(\Omega) = 6^3 = 216.$$

$$A = \text{"obtenir le total 11"} = \{\omega \in \Omega \mid \omega_1 + \omega_2 + \omega_3 = 11\}$$

$$= \{(6, 4, 1); (6, 1, 4); (4, 6, 1); (4, 1, 6); \dots\}$$

i.e. on voit qu'il faut tenir compte des permutations, ce que l'ami de Galilée n'avait pas fait!

$$N(A) = 6 + 6 + 3 + 6 + 3 + 3 = 27 \quad \text{donc } P(A) = \frac{27}{216} = 0,125$$

$$B = \text{"obtenir le total 12"} = \{\omega \in \Omega \mid \omega_1 + \omega_2 + \omega_3 = 12\}$$

$$N(B) = 6 + 6 + 3 + 3 + 6 + 1 = 25 \quad \text{donc } P(B) = \frac{25}{216} \approx 0,116$$

Remarque: On constate que  $P(A) = \frac{1080}{8640}$  et  $P(B) = \frac{1000}{8640}$ ;

- l'ami de Galilée a donc jeté environ 8640 fois ses 3 dés; on comprend que ces joueurs, avant en l'absence du hasard!
- l'exactitude de ces résultats est surprenante: quelle est la probabilité que J. Bertrand ait "retouché" l'histoire?

Le deuxième problème est plus célèbre; il est déjà mentionné dans un ouvrage de 1494, mais toutes les solutions proposées étaient fausses. Le problème fut posé à Pascal par le chevalier de Méné. Pascal a trouvé une solution en 1654, l'a soumise à Fermat, qui en a trouvé une autre (correspondance célèbre).

Problème du partage.

Pierre et Paul jouent à pile ou face. Ils misent la même somme (32 pitoles). Le premier des joueurs qui aura obtenu 3 points (consecutifs ou non) gagnera la somme des mises. Ils jouent une première partie, puis doivent se séparer. Comment vont-ils se partager équitablement l'enjeu?

Solution (pour la solution de Pascal voir le Que sais-je n° 1571)

Supposons pour fixer les idées que Pierre gagne avec pile (P) et que c'est Pierre qui a gagné la 1<sup>ère</sup> partie.

Tout sera décidé dans les 4 parties suivantes; on peut donc prendre  $\Omega = \{P, F\}^4$ , ce qui donne  $N(\Omega) = 16$ .

Explicitons:

$$\Omega = \{(P, P, P, P); (P, P, P, F); (P, P, F, P); (P, F, P, P); (F, P, P, P); (P, P, F, F); (P, F, P, F); (P, F, F, P); (F, P, P, F); (F, P, F, P); (F, F, P, P); (P, F, F, F); (F, P, F, F); (F, F, P, F); (F, F, F, P); (F, F, F, F)\}$$

$$A = \text{"Pierre gagne l'enjeu"} = \text{l'ens. de } \omega \in \Omega \text{ qui comptent au moins 2 P}; \quad N(A) = 11.$$

$$B = \text{"Paul gagne l'enjeu"} = \Omega - A; \quad N(B) = 5.$$

Donc Pierre prendra les  $\frac{11}{16}$  de l'engin (soit 44 pistols) et Paul les  $\frac{5}{16}$  (soit 20 pistols).

Remarque. La première publication traitant du calcul des probabilités fut écrite par Huyghens en 1657. Il accorde l'honneur de l'invention du calcul des probabilités à Pascal et à Fermat. Cinq problèmes sont posés par Huyghens, dont celui de la suite des joueurs, que nous verrons plus tard.

2.3. Rappels d'analyse combinatoire.

A. r-uples ordonnés.

Soit  $S$  un ensemble fini de  $n$  éléments. Un  $r$ -uplet ordonné est un élément de  $S^r$  ( $r$  entier  $\geq 1$ ).

Lemme 1. 1) Il y a  $n^r$   $r$ -uplets ordonnés distincts i.e.  $N(S^r) = n^r$ .

2) Il y a  $(n)_r = n(n-1)\dots(n-r+1)$   $r$ -uplets ordonnés formés d'éléments distincts 2 à 2.

Exemple 1) On jette 4 dés distincts. Quelle est la probabilité d'obtenir 4 résultats distincts ?

Ici  $S = \{1, 2, 3, 4, 5, 6\}$ ,  $\Omega = S^4$ ,  $N(\Omega) = 6^4 = 1296$ ,  
 $A = \text{"obtenir 4 résultats distincts"} = \{4\text{-uplets formés d'él. distincts 2 à 2}\}$   
donc  $N(A) = (6)_4$  et  $P(A) = \frac{(6)_4}{6^4} = \frac{360}{1296} \approx 0,28$ .

2) Dans une classe de 23 élèves, quelle est la probabilité que tous les élèves aient des anniversaires différents ?

Comme ci-dessus on trouve  $P = \frac{(365)_{23}}{365^{23}} \approx 0,5$  (A)  
donc la probabilité que 2 (au moins) des élèves aient leur anniversaire le même jour est  $\approx \frac{1}{2}$  !!

3) En général, la probabilité d'obtenir  $r$  éléments distincts en les tirant (avec remise) d'un ensemble de  $n$  éléments est

$$P = \frac{(n)_r}{n^r} = \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{r-1}{n}\right)$$

(A) Calcul approx.  $\log P = \sum_{i=1}^{23} \log\left(1 - \frac{i}{365}\right) \approx -\sum_{i=1}^{23} \frac{i}{365} = -\frac{27 \cdot 23}{730}$

B. sous-ensembles.

Lemme 2. Un ensemble  $S$  de  $n$  éléments contient  $\binom{n}{r} = \frac{n!}{r!(n-r)!}$  sous-ensembles de  $r$  éléments.

Preuve. Il y a  $(n)_r$   $r$ -uplets d'éléments de  $S$  avec des éléments distincts 2 à 2. Chaque sous-ensemble de  $r$  éléments admet  $r!$  uplets ordonnés (par permutation), et on obtient ainsi tous les  $r$ -uplets : donc le nombre de sous-ensembles de  $r$  éléments est égal à  $\frac{(n)_r}{r!} = \binom{n}{r}$ .

Exemple. Au bridge (52 cartes) quelle est la probabilité d'avoir une main (13 cartes) donnée ?

$\Omega =$  ens. des sous-ensembles à 13 éléments d'un ensemble à 52 c.  
 $N(\Omega) = \binom{52}{13} \approx 6,35 \cdot 10^{11}$ ;  $P = \frac{1}{N(\Omega)} \approx 1,57 \cdot 10^{-12}$ .

Généralisation. On voit donc que  $\binom{n}{r}$  désigne le nombre de partitions d'un ensemble  $S$  de  $n$  éléments en un sous-ensemble de  $r$  éléments et un sous-ensemble de  $n-r$  éléments. Plus généralement,

Lemme 3. le nombre de partitions de  $S$  en  $h$  sous-ensembles de  $v_1, v_2, \dots, v_h$  ( $\sum_{i=1}^h v_i = n$ ) éléments est  $\frac{n!}{v_1! v_2! \dots v_h!}$  (coefficient multinomial)

Preuve. Il y a  $\binom{n}{v_1}$  façons de choisir le 1<sup>er</sup> sous-ensemble,  $\binom{n-v_1}{v_2}$  de choisir le 2<sup>ème</sup>, puis ..., d'où finalement le nombre de partitions est égal à

$$\begin{aligned} & \binom{n}{v_1} \binom{n-v_1}{v_2} \binom{n-v_1-v_2}{v_3} \dots \binom{n-v_1-v_2-\dots-v_{h-1}}{v_h} \\ &= \frac{n!}{v_1! (n-v_1)!} \cdot \frac{(n-v_1)!}{v_2! (n-v_1-v_2)!} \cdot \frac{(n-v_1-v_2)!}{v_3! (n-v_1-v_2-v_3)!} \dots \frac{(n-v_1-v_2-\dots-v_{h-1})!}{v_h! (n-v_1-v_2-\dots-v_h)!} \\ &= \frac{n!}{v_1! v_2! \dots v_h!} \end{aligned}$$



Exemple. Au bridge il y a  $\frac{52!}{(13!)^4} \approx 5,3645 \cdot 10^{28}$  répartition<sup>12</sup> de cartes, donc la probabilité d'obtenir une répartition donnée est  $P \approx 1,8641 \cdot 10^{-29}$ .

## 2.4. Urnes.

Beaucoup d'expériences aléatoires peuvent être assimilées au jeu suivant : on extrait une ou plusieurs boules d'une urne contenant des boules de différentes couleurs et on observe la couleur des boules tirées.

Il y a plusieurs façons d'extraire r boules d'une urne de n boules :

a) si les r boules sont tirées en une seule fois, le résultat est un sous-ensemble de r éléments ( $r \leq n$ ) ; on parle de sous-population de taille n.

Modèle :  $\Omega = \text{ens. des sous-ensembles de } r \text{ éléments de } S$  ( $S$  est l'ensemble des n boules de l'urne) ; P comme en 2.1.

b) si les r boules sont tirées séquentiellement, le résultat est un r-uple ordonné : on parle d'échantillon de taille r. On distingue encore

b') le tirage avec remise (avec remplacement)

Modèle :  $\Omega = S^r$  ; P comme en 2.1. ( $N(\Omega) = n^r$ )

b'') le tirage sans remise (sans remplacement)

Modèle :  $\Omega = \{ \omega = (s_1, s_2, \dots, s_r) \mid s_i \neq s_j \ \forall i \neq j \}$  ; P comme en 2.1. ( $N(\Omega) = (n)_r$ ).

Exemples : 8 boules : 4 blanches et 4 noires.

a) Une sous-population de taille 2 est extraite ; quelle est la probabilité qu'elle soit formée de 2 boules blanches ?

$N(\Omega) = \binom{8}{2} = 28$  ;  $A = \text{ens. des sous-ens. à 2 éléments tirés de l'ens. des boules blanches}$  :  $N(A) = \binom{4}{2} = 6$  ;  $P(A) = \frac{6}{28} = \frac{3}{14}$ .

b') Un échantillon de taille 2 est extrait avec remise ; quelle est la probabilité qu'il soit formé de 2 boules blanches ?  
 $N(\Omega) = 8^2 = 64$  ;  $N(A) = 4^2$  ;  $P(A) = \frac{4^2}{8^2} = \frac{1}{4}$ .

b'') Comme b') mais sans remise :  $N(\Omega) = (8)_2 = 56$  ;  
 $N(A) = (4)_2 = 12$  ;  $P(A) = \frac{12}{56} = \frac{3}{14}$ .

Proposition. Une urne contient n boules : m blanches, n-m noires. On extrait un échantillon de taille r avec ou sans remise. Pour tout i,  $1 \leq i \leq r$ , soit  $A_i$  l'événement "la i<sup>ème</sup> boule extraite est blanche". Alors

$$P(A_i) = \frac{m}{n}$$

quel que soit i.

Preuve. Soit S l'ensemble des boules de l'urne.

1) Tirage avec remise (intuitivement clair !) :  $\Omega = S^r$  ;  $N(\Omega) = n^r$  ;  
 $A_i = \{ \omega = (s_1, s_2, \dots, s_r) \mid s_i \text{ est blanche} \}$  ;  $N(A_i) = n^{r-1} \cdot m$  ;  
 donc  $P(A_i) = \frac{n^{r-1} \cdot m}{n^r} = \frac{m}{n}$ .

2) Tirage sans remise (plus surprenant, non ?) (\*)

$\Omega = \{ \omega = (s_1, s_2, \dots, s_r) \mid s_i \neq s_j \ \forall i \neq j \}$  ;  $N(\Omega) = (n)_r$ .

$i=1$ .  $A_1 = \{ \omega \in \Omega \mid s_1 \text{ est blanche} \}$  ; il est clair que  
 $N(A_1) = m(m-1)\dots(m-r+1)$  , donc  $P(A_1) = \frac{m(m-1)\dots(m-r+1)}{n(n-1)\dots(n-r+1)} = \frac{m}{n}$ .

i quelconque : supposons que les m boules blanches soient numérotées de 1 à m et les noires de m+1 à n.

$A_i = \sum_{k=1}^m \{ \omega \in \Omega \mid s_i = \text{boule } n^{\circ} k \}$  (NB. les  $\{-\}$  sont disjoints 2 à 2)

et  $N(\{ \omega \mid s_i = \text{boule } n^{\circ} k \}) = (n-1)_{r-1}$  , donc  $N(A_i) = m \cdot (n-1)_{r-1}$

et  $P(A_i) = \frac{m (n-1)_{r-1}}{(n)_r} = \frac{m}{n}$ .

\*) Le contenu de l'urne est évidemment modifié par ces i-2 premiers tirages, mais on ne sait pas comment i.e. on n'a pas plus d'information qu'avant le i<sup>ème</sup> tirage, cela revient à dire que les

2.5. Loi hypergéométrique.

Proposition. On considère une urne contenant  $n$  boules, dont  $n_1$  sont blanches et  $n_2 = n - n_1$  sont noires. On extrait de l'urne une sous-population de taille  $r$ . La probabilité que cette sous-population contienne  $k$  boules blanches et  $r-k$  l. noires

$$P_k = \frac{\binom{n_1}{k} \binom{n-n_1}{r-k}}{\binom{n}{r}}$$

si  $\max(0, r-n_2) \leq k \leq \min(n_1, r)$ ; sinon  $P_k = 0$ .

Définition: La probabilités  $P_k$  forment la loi hypergéométrique. (Le nom vient du fait que  $f(s) = \sum_k P_k s^k$  est une fct. hypergeom.)

Démonstration: on prendra pour  $\Omega$  l'ensemble des sous-populations de taille  $r$  de l'ensemble  $S$  des boules de l'urne.  $N(\Omega) = \binom{n}{r}$ . soit  $A_k = \{\omega \in \Omega \mid \omega \text{ contient exactement } k \text{ boules blanches}\}$ . Pour extraire un  $\omega \in A_k$ , on peut extraire  $k$  boules blanches de l'ensemble des boules blanches de l'urne - il ya  $\binom{n_1}{k}$  possibilités - puis extraire  $r-k$  boules noires de l'ensemble des  $n-n_1$  boules noires de l'urne - il ya  $\binom{n-n_1}{r-k}$  possibilités. On a donc  $N(A_k) = \binom{n_1}{k} \binom{n-n_1}{r-k}$  et on a bien

$$P_k = P(A_k) = \frac{\binom{n_1}{k} \binom{n-n_1}{r-k}}{\binom{n}{r}}$$

Il est clair que pour que  $A_k$  ne soit pas impossible il faut (et il suffit que)  $0 \leq k \leq n_1$ ,  $k \leq r$ ,  $r-k \leq n-n_1$  i.e. que  $\max(0, r-(n-n_1)) \leq k \leq \min(n_1, r)$ .

Notation: on étend la définition de  $\binom{a}{b}$  en posant  $\binom{a}{b} = 0$  si  $b < 0$  ou si  $b > a$ . Avec ces conventions on voit que  $P_k = \frac{\binom{n_1}{k} \binom{n-n_1}{r-k}}{\binom{n}{r}}$  quel que soit  $k \geq 0$ .

Exemple. Au bridge chaque joueur reçoit 13 cartes. La probabilité qu'un joueur spécifique possède  $k$  as est donc:

$$P_k = \frac{\binom{4}{k} \binom{48}{13-k}}{\binom{52}{13}} \quad k=0,1,2,3,4$$

$$P_0 = \frac{\binom{48}{13}}{\binom{52}{13}} = \frac{39 \cdot 38 \cdot 37 \cdot 36}{52 \cdot 51 \cdot 50 \cdot 49} = 0,304$$

$$P_1 = \frac{\binom{4}{1} \binom{48}{12}}{\binom{52}{13}} = 4 \cdot \frac{17 \cdot 37 \cdot 36 \cdot 37}{52 \cdot 51 \cdot 50 \cdot 49} = 0,439$$

$$P_2 = \frac{\binom{4}{2} \binom{48}{11}}{\binom{52}{13}} = 6 \cdot \frac{13 \cdot 12 \cdot 39 \cdot 38}{52 \cdot 51 \cdot 50 \cdot 49} = 0,213$$

$$P_3 = \frac{\binom{4}{3} \binom{48}{10}}{\binom{52}{13}} = 4 \cdot \frac{13 \cdot 12 \cdot 11 \cdot 39}{52 \cdot 51 \cdot 50 \cdot 49} = 0,041$$

$$P_4 = \frac{\binom{4}{4} \binom{48}{9}}{\binom{52}{13}} = \frac{13 \cdot 12 \cdot 11 \cdot 10}{52 \cdot 51 \cdot 50 \cdot 49} = 0,003$$

(Noter que  $\sum_k P_k = 1$  comme il se doit!)

Application: Estimation de la taille d'une population par recapture.

On aimerait connaître le nombre de poissons contenus dans un lac. Désignons par  $n$  ce nombre. On capture 1000 poissons qu'on marque d'un point blanc puis qu'on relâche. Après un certain temps, on recapture 1000 poissons et on compte que 100 d'entre eux sont marqués d'un point blanc. Que peut-on en déduire en ce qui concerne  $n$ ?

Admis: i) le nombre des poissons ne change pas entre les deux captures; ii) chaque capture peut être considérée comme faite au hasard i.e. chaque sous population de 1000 poissons extraite du lac est équiprobable.

On peut donc supposer qu'on a une urne avec  $n$  boules, dont  $n_1 = 1000$  sont blanches et  $n_2 = n - 1000$  sont noires. On extrait  $r = 1000$  boules de l'urne (sans-population). La probabilité d'obtenir  $k = 100$  boules blanches est donc

$$p_{100} = \frac{\binom{1000}{100} \binom{n-1000}{900}}{\binom{n}{1000}} = f(n)$$

Il est théoriquement possible qu'il n'y ait que 100 poissons dans le lac (moins, c'est impossible!). Cependant c'est très peu probable, car si  $n = 1900$  on a  $p_{100} \approx 10^{-430}$  ... Même considération pour  $n$  très grand.

En fait il est naturel de chercher la valeur de  $n$  pour laquelle  $p_{100} = f(n)$  est la plus grande: on la note  $\hat{n}$ ; c'est l'estimateur de  $n$  donné par le maximum de vraisemblance (méthode déjà utilisée par Gauss (1777-1855) dans certains cas particuliers, puis généralisée et largement utilisée par R.A. Fisher (1922)).

Cherchons  $\hat{n}$ . On a  $f(n) = \frac{\binom{n_1}{k} \binom{n-n_1}{r-k}}{\binom{n}{r}}$  donc

$$\frac{f(n)}{f(n-1)} = \dots = \frac{(n-n_1)(n-r)}{n(n-n_1-r+k)}$$

d'où on voit que  $f(n) > f(n-1) \iff n < \frac{rn_1}{k}$

i.e.  $f(n)$  croît avec  $n$  jusqu'à ce que  $n = \frac{rn_1}{k}$  puis décroît.

On a donc  $\hat{n} =$  l'entier le plus proche de  $\frac{rn_1}{k}$ . (\*)

Dans notre exemple on trouve  $\hat{n} = 10'000$ .

Remarque: On peut calculer que si  $n = 8500$ , la proba. d'obtenir moins de 100 poissons marqués est 9,04. De même si  $n = 12'000$  la proba. d'obtenir plus de 100 poissons marqués est 9,03. Donc  $n$  "doit" se trouver entre 8500 et 12'000. (cf Feller page 46).

(\*) résultat vraisemblable par valeurs:  $\frac{k}{r} = \frac{n_1}{n}$  i.e. proportion de blancs dans la sous-popul. = prop. de blancs dans la pop.

Généralisation de la Proposition: On considère une urne contenant  $n$  boules dont  $n_1$  sont de la 1<sup>ère</sup> couleur, ...,  $n_g$  sont de la  $g$ <sup>ème</sup> couleur ( $\sum_{i=1}^g n_i = n$ ). On extrait une sous-population de  $r$  boules. La probabilité qu'il y ait  $k_i$  boules de la  $i$ <sup>ème</sup> couleur ( $1 \leq i \leq g$ ) ( $\sum_{i=1}^g k_i = r$ ) est

$$P(k_1, \dots, k_g) = \frac{\binom{n_1}{k_1} \dots \binom{n_g}{k_g}}{\binom{n}{r}}$$

la preuve est la même que celle de la Proposition.

Exemple: Au bridge la probabilité qu'un joueur spécifié reçoive 5 coeurs, 4 piques, 3 carreaux, 1 trefle est

$$p = \frac{\binom{13}{5} \binom{13}{4} \binom{13}{3} \binom{13}{1}}{\binom{52}{13}} \approx 0,0054$$

2.6. Echantillons: loi binomiale.

Lemme: Une urne contient  $n$  boules dont  $n_1$  sont blanches et  $n_2 = n - n_1$  sont noires. On extrait un échantillon de taille  $r$ . La probabilité que  $k$  tirages spécifiés donnent une boule blanche et les  $r-k$  autres tirages une boule noire est

- 1)  $p = \frac{n_1^k n_2^{r-k}}{n^r}$  si les tirages sont faits avec remplacement;
- 2)  $p = \frac{\binom{n_1}{k} \binom{n_2}{r-k}}{\binom{n}{r}}$  si les tirages sont faits sans remplacement.

Preuve: soit  $S$  l'ensemble des boules de l'urne.

Fixons-nous un sous-ensemble  $I$  de  $\{1, 2, \dots, r\}$  avec  $N(I) = k$ .

1) Dans le 1<sup>er</sup> cas,  $\Omega = S^r$ , et on s'intéresse à l'événement

$$A_I = \{ \omega = (s_1, s_2, \dots, s_r) \in \Omega \mid s_i \text{ est blanche } \forall i \in I, s_j \text{ noire } \forall j \in \bar{I} \}$$

on a  $N(A_I) = n_1^k n_2^{r-k}$ , donc  $p = P(A_I) = \frac{n_1^k n_2^{r-k}}{n^r}$ .

2) Dans le 2<sup>ème</sup> cas  $\Omega = \{\omega = (s_1, s_2, \dots, s_n) \in S^n \mid s_i \neq s_j, \forall i \neq j\}$   
et l'événement qui nous intéresse est

$$B_I = \{\omega \in \Omega \mid s_i \text{ blanche } \forall i \in I, s_j \text{ noire } \forall j \in \bar{I}\}$$

On a  $N(B_I) = \binom{m_1}{k} \binom{m_2}{n-k}$ , donc  $p = P(B_I) = \frac{\binom{m_1}{k} \binom{m_2}{n-k}}{\binom{m}{n}}$ .

Proposition. Mêmes hypothèses que pour le lemme. La probabilité que l'échantillon extrait contienne k boules blanches et r-k boules noires est

1)  $p_k = \binom{r}{k} \frac{\binom{m_1}{k} \binom{m_2}{r-k}}{m^r}$  si les tirages sont faits avec remplacement;

2)  $p_k = \binom{r}{k} \frac{\binom{m_1}{k} \binom{m_2}{r-k}}{\binom{m}{r}}$  si les tirages sont faits sans remplacement.

Preuve: Notation du lemme.

1)  $\Omega = S^r$ ; on s'intéresse à l'événement

$$A_k = \{\omega \in \Omega \mid \text{il y a k noirs } i \text{ avec } s_i \text{ blanche}\}$$

$$= \sum_I A_I \quad \text{où } I \text{ parcourt l'ans. des parties à k éléments}$$

Par le lemme on a  $N(A_k) = \sum_I N(A_I) = \binom{r}{k} \binom{m_1}{k} \binom{m_2}{r-k}$  <sup>de  $\{1, 2, \dots, r\}$</sup>   
d'où le résultat  $p_k = P(A_k) = \binom{r}{k} \frac{\binom{m_1}{k} \binom{m_2}{r-k}}{m^r}$ .

2) Même preuve.

Définition: Les probabilités  $p_k = \binom{r}{k} \frac{\binom{m_1}{k} \binom{m_2}{r-k}}{m^r} = \binom{r}{k} \left(\frac{m_1}{m}\right)^k \left(\frac{m_2}{m}\right)^{r-k}$  définissent la loi binomiale.

Notons que  $\frac{m_1}{m} =$  probabilité de tirer une boule blanche (en un tirage) et  $\frac{m_2}{m} = 1 - \frac{m_1}{m}$ .

Exemple: On jette 2n fois une pièce équilibrée; la probabilité d'obtenir n fois "pile" est  $\binom{2n}{n} \frac{1}{2^{2n}}$  (modèle d'urne: 2 boules: 1 blanche, 1 noire; 2n tirages, avec remplacement)

NB. Avec Stirling  $\binom{2n}{n} \frac{1}{2^{2n}} = \frac{1}{\sqrt{\pi n}} \rightarrow 0$  si  $n \rightarrow \infty$  !! Expliquer...

Remarques: 1) la formule 2) de la Proposition redonne la loi hypergéométrique. En effet:

$$\begin{aligned} \binom{r}{k} \frac{\binom{m_1}{k} \binom{m_2}{r-k}}{\binom{m}{r}} &= \frac{r!}{k!(r-k)!} \frac{m_1(m_1-1)\dots(m_1-k+1) \cdot m_2(m_2-1)\dots(m_2-r+k+1)}{m(m-1)\dots(m-r+1)} \\ &= \frac{\binom{m_1}{k} \binom{m_2}{r-k}}{\binom{m}{r}} \quad (\text{cf. page 14}). \end{aligned}$$

Intuitivement c'est clair: si on ne s'intéresse qu'au nombre de boules blanches extraites, il n'est pas important de savoir dans quel ordre on sort les boules (le modèle des échantillons sans remise est en fait trop fin pour ce problème...).

2) D'après l'exercice 3 de la série 4, si n est très grand par rapport à r, on voit que la loi hypergéométrique est très voisine de la loi binomiale:

$$\frac{\binom{m_1}{k} \binom{m_2}{r-k}}{\binom{m}{r}} \approx \binom{r}{k} \left(\frac{m_1}{m}\right)^k \left(\frac{m_2}{m}\right)^{r-k}$$

On comprend bien ce qui se passe: si on n'extrait qu'une faible proportion des boules de l'urne, la probabilité d'obtenir k boules blanches est peu influencée par le fait que les tirages soient avec ou sans remplacement.

Dans le cas des échantillons un autre problème peut être abordé: c'est celui du temps d'attente avant d'obtenir une boule blanche.

Proposition: Hypothèses du lemme. La probabilité que la 1<sup>ère</sup> boule blanche soit obtenue au r<sup>ème</sup> tirage est

1)  $p = \frac{m_1 m_2^{r-1}}{m^r}$  si les tirages sont faits avec remplacement,

2)  $p = \frac{m_1 (m_2)^{r-1}}{\binom{m}{r}}$  si les tirages sont faits sans remplacement.

Preuve. C'est le lemme avec  $I = \{r\} \subset \{1, 2, \dots, r\}$ .

Remarque: Pour décrire correctement l'expérience "extraire un échantillon avec remplacement jusqu'à ce qu'on obtienne une boule blanche" il est nécessaire d'introduire un espace des épreuves infini (car  $r$  peut être arbitrairement grand). On verra que les probabilités  $p_r = \frac{m \cdot m^{r-1}}{m^r}$  définissant alors une loi (dite loi géométrique): voir plus loin.

Exemples. 1) On jette une pièce équilibrée ( $p = \frac{1}{2}$  pour "pile"). La probabilité que "pile" apparaisse la 1<sup>ère</sup> fois au  $r$ <sup>ème</sup> lancer est  $\frac{1}{2^r}$ .

2) Pierre a  $n$  clés: une seule ouvre sa porte. Il les essaie au hasard, mais une seule fois chacune. La probabilité que la bonne clé soit trouvée au  $r$ <sup>ème</sup> essai est  $p_r = \frac{1 \cdot (n-1) \cdot \dots \cdot (n-r+1)}{(n)_r} = \frac{1}{n}$  i.e. Pierre a autant de chances d'essayer 1 clé que 2 clés, 3 clés, 4 clés, ...

NB. la surprise est due au fait qu'on confond cette probabilité avec celle d'obtenir la bonne clé dans les  $n-1, n-2, n-3, \dots$  clés restantes (cette proba. augmente!).

## 2.7. Problèmes d'affectation.

De nombreux problèmes concrets peuvent être énoncés de la façon suivante: "on place  $r$  boules au hasard dans  $n$  boîtes".

Il y a 4 cas à distinguer selon que:

- les boules sont discernables (p.ex. numérotées) ou indiscernables,
- la répartition se fait avec répétition ou sans répétition (i.e. une boîte peut contenir plusieurs boules ou une boule au maximum).

Remarque: Il existe un lien entre ce problème et celui des tirages d'une urne: la répartition des  $r$  boules dans les  $n$  boîtes peut être considérée comme  $r$  tirages d'une

urne contenant  $n$  boîtes. On a les correspondances suivantes:

boules discernables  $\leftrightarrow$  échantillon

boules indiscernables  $\leftrightarrow$  sous-population

répétition  $\leftrightarrow$  remplacement

Le fait nouveau ici est la répartition de boules indiscernables avec répétition (on ne pouvait pas parler d'extraire une sous-population avec remplacement!).

Combien y a-t-il, dans chaque cas, de configurations possibles?

Proposition: On répartit  $r$  boules dans  $n$  boîtes.

- 1) boules discernables, avec répétition:  $n^r$  configurations possibles.
- 2) boules discernables, sans répétition ( $r \leq n$ ):  $(n)_r$  configurations possibles.
- 3) boules indiscernables, sans répétition ( $r \leq n$ ):  $\binom{n}{r}$  configurations possibles.
- 4) boules indiscernables, avec répétition:  $\binom{n+r-1}{r}$  configurations possibles.

Preuve: 1), 2) et 4) sont connus (cf. Remarque ci-dessus).

3) Introduisons une notation:

Si  $r=8, n=6$  alors  $|000|0|||0000|$  représente la répartition suivante: 3 boules dans la 1<sup>ère</sup> boîte, 1 dans la 2<sup>ème</sup>, 0 dans la 3<sup>ème</sup>, 0 dans la 4<sup>ème</sup>, 0 dans la 5<sup>ème</sup>, 4 dans la 6<sup>ème</sup>.

Toute répartition commence et finit par une barre; entre ces deux barres il y a  $n-1$  barres et  $r$  boules qui peuvent être placés de toutes les façons possibles! Il y a donc

$$\binom{n-1+r}{r} = \binom{n-1+r}{n-1}$$

possibilités. Cqfd.

Illustration des 4 cas:  $n=3, r=2$

1)  $|ab| | |$ ;  $| | ab| |$ ;  $| | | ab|$ ;  $|a|b| |$ ;  $|a| |b|$ ;  $|b|a| |$ ;  $| | |a|$ ;  $| | |b|$ ;  $| | |$ .

2) les 6 derniers cas de 1).

3)  $|0|0| |$ ;  $|0| |0|$ ;  $| |0|0|$ .

4) comme 3) avec en plus:  $|00| | |$ ;  $| |00|$ ;  $| | |00|$ .

Proposition. On répartit  $r$  boules dans  $n$  boîtes. On suppose que toutes les répartitions sont équiprobables. Si  $r_1, r_2, \dots, r_n$  sont des entiers  $\geq 0$  avec  $r_1 + r_2 + \dots + r_n = r$ , alors la probabilité d'avoir  $r_i$  boules dans la  $i$ -ième boîte, est:

- a)  $\frac{r!}{r_1! r_2! \dots r_n! n^r}$  dans le cas des boules discernables, avec répétition
- b)  $\frac{1}{\binom{n+r-1}{r}}$  dans le cas des boules indiscernables, avec répétition
- c)  $\frac{1}{\binom{n}{r}}$  dans le cas des boules indiscernables, sans répétition  
NB. dans ce cas on a  $r_i = 0$  ou  $1$  ( $\forall i$ ).

Preuve: pour a) voir lemme 3 page 11; b) et c): clairs.

Application à la mécanique statistique.

On considère un système de  $r$  particules. L'espace de phase est décomposé en un grand nombre  $n$  de cellules, appelées états du système. La distribution des  $r$  particules dans les  $n$  états décrit l'état du système entier.

a) modèle classique: les particules sont discernables et plusieurs particules peuvent être simultanément dans le même état. C'est la statistique (\*) de Maxwell-Boltzmann. Elle s'applique à un gaz de molécules.

En mécanique quantique, le principe d'incertitude implique que les particules sont indiscernables (la notion de trajectoire n'a plus de sens). On distingue deux cas:

b) si plusieurs particules peuvent occuper simultanément le même état, c'est la statistique de Bose-Einstein. Elle s'applique aux photons p.ex.

c) si deux particules ne peuvent se trouver simultanément dans un même état (principe de Pauli), c'est la statistique de Fermi-Dirac. Elle s'applique aux électrons p.ex.

(\*) Ici, "statistique" veut dire "modèle probabiliste".

NB. Dans les 3 cas on fait l'hypothèse que les répartitions sont équiprobables!

Remarque concernant les statistiques classique et quantique.

cf. série 6.

Prenons l'exemple de  $n=3$  boîtes et  $r=3$  boules.

a) boules discernables: il y a 27 répartitions possibles:

abc	a   bc	ab     c	abc
a bc	a   c b	a c     b	
b ac	a   c b	b c     a	
c ab	a   b c	b c     a	
a b   c	b   ac	ac     b	
ac   b	b   a c	ac     b	
b c   a	b   c   a		
a b c	b   a c		
	c   a b		
	c   a   b		
	c   b   a		
	c   a b		

b) boules indiscernables: il y a 10 répartitions possibles:

000	0     00	00     0	000
0   00	0   0   0	00   0	
00   0	0   00		
000			

Lorsqu'il s'agit effectivement de boules à répartir dans des boîtes (au sens concret des termes) il est naturel - et conforme à l'expérience - de supposer que tous les cas du modèle a) sont équiprobables. (statistique classique). Si les boules sont indiscernables, il est également naturel de supposer que l'expérience n'est pas affectée par notre capacité de distinguer les boules, ou, autrement dit, que la nature sait-elle-distinguer les boules. Alors les 10 événements élémentaires de b) ont pour probabilité respectivement:

1/27	3/27	3/27	1/27
3/27	6/27	3/27	
3/27	3/27		
1/27			

Le fait que les "bosons" obéissent à la statistique de Bose-Einstein signifie que - pour eux - il faut supposer les 10 répartitions ci-dessus comme équiprobables. C'est un fait expérimental.

qu'il ne faut pas chercher à "comprendre" en regardant des boules de ping-pong! Il faut plutôt utiliser ce fait pour mieux "imaginer" ce que sont les boules (et le moins qu'on puisse dire, c'est qu'ils ne se comportent pas comme des boules de ping-pong...)

Il serait intéressant de trouver une situation concrète qui obéit à la statistique de B-E.  
des taches sur des pommes?

Exemple: on jette 5 dés indiscernables; on considère les ds comme des boules indiscernables qu'on répartit au hasard dans 6 boîtes (correspondant aux faces 1,2,3,4,5,6).  
← Il y a donc  $\binom{6+5-1}{5} = \binom{10}{5} = 252$  résultats possibles.

## 2.8. Insuffisances du modèle classique.

Résumons: pour décrire et étudier certaines expériences aléatoires on a utilisé le modèle classique  $(\Omega, \mathcal{P}(\Omega), P)$  où  $\Omega$  est un ensemble fini et  $P(A) = \frac{N(A)}{N(\Omega)}$  pour tout  $A \in \mathcal{P}(\Omega)$ .  
Nous avons surtout calculé des probabilités. Les propriétés suivantes de la probabilité ont été souvent utilisées:

- (1)  $0 \leq P(A) \leq 1$  pour tout  $A \in \mathcal{P}(\Omega)$
- (2) si  $A \cap B = \emptyset$ ,  $P(A+B) = P(A) + P(B)$ . Aussi:  $P(\bigcup_i A_i) = \sum_i P(A_i)$ .
- (3)  $P(\bar{A}) = 1 - P(A)$  pour tout  $A \in \mathcal{P}(\Omega)$
- (4)  $P(\bigcup_{i=1}^r A_i) = \sum_i P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) - \dots - (-1)^{r+1} P(A_1 \cap A_2 \cap \dots \cap A_r)$  (formule de Poincaré)

Ces propriétés permettent souvent de ramener le calcul de la probabilité d'un événement complexe au calcul de probabilités de plusieurs événements plus simples (voir p.ex. série 3 et série 5).

Cependant le modèle classique manque de souplesse et de généralité:

- souplesse: il oblige à considérer parfois des ensembles  $\Omega$  trop riches; par exemple si on extrait un échantillon de taille  $r$  d'une urne contenant  $n$  boules dont  $n_1$  blanches et  $n_2 = n - n_1$  noires et si on ne s'intéresse qu'au nombre  $k$  de boules blanches obtenues, il serait agréable de pouvoir considérer  $\Omega = \{0, 1, 2, \dots, r\}$  avec  $P(\{k\}) = p_k$  donné par la loi hypergéométrique ou binomiale (selon que les tirages sont effectués sans ou avec remplacement).

- généralité : (a) le jet d'une pièce non équilibrée :  $Pr(\text{face}) = p$ ,  $Pr(\text{pile}) = q = 1-p$  avec  $p+q$  (et  $p$  irrationnel !!) ne peut pas être représenté par le modèle classique (bien que  $\Omega$  soit fini).  
 (b) le problème du temps d'attente d'une boule blanche dans le cas des tirages avec remplacement ne peut être représenté que par un espace  $\Omega$  infini.

Nous allons remédier à ces défauts en définissant axiomatiquement la notion de probabilité. Nous développerons ensuite mathématiquement la théorie des probabilités. Et ce sont certains résultats de la théorie qui nous diront de manière précise comment la notion de probabilité peut être atteinte par l'expérience (fréquences  $\rightarrow$  probabilités).

### §3. Axiomatique du calcul des probabilités.

#### 3.1. Définition d'un espace probabilisé.

Un espace probabilisé est un triple  $(\Omega, \mathcal{O}, P)$  où

- 1)  $\Omega$  est un ensemble non vide (espace des épreuves);
- 2)  $\mathcal{O}$  est une tribu de parties de  $\Omega$  (tribu des événements);
- 3)  $P$  est une probabilité c'est-à-dire une application de  $\mathcal{O}$  dans  $\mathbb{R}$  qui vérifie la condition suivante:
  - a)  $0 \leq P(A) \leq 1$  pour tout  $A \in \mathcal{O}$ ;
  - b)  $P(\Omega) = 1$ ;
  - c<sub>1</sub>) si  $A, B \in \mathcal{O}$ ,  $A \cap B = \emptyset$ , alors  $P(A+B) = P(A) + P(B)$ ;
  - c<sub>2</sub>) si  $A_n \in \mathcal{O}$ ,  $\forall n \in \mathbb{N}^*$ ,  $A_n \cap A_m = \emptyset$  si  $n \neq m$ , alors  $P(\sum_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ .

Remarques: 1) Les événements de  $\mathcal{O}$  sont dits probabilisables.  
 2) L'axiome c<sub>1</sub>) est l'axiome d'additivité, c<sub>2</sub>) est l'axiome de  $\sigma$ -additivité ou additivité dénombrable; il est sans objet si  $\mathcal{O}$  est finie, mais il est nécessaire pour avoir une théorie suffisamment riche dans le cas infini.

Remarque que si  $(A_n)$  est comme dans c<sub>2</sub>) alors c<sub>2</sub>) implique  $\sum_{n=1}^{\infty} P(A_n) = P(\sum_{n=1}^{\infty} A_n) \leq 1$  ( $\forall n \geq 1$ ) donc  $\sum_{n=1}^{\infty} P(A_n)$  converge, mais on ne sait pas vers quoi!

#### 3.2. Exemples.

##### 1) Espaces probabilisés finis.

Soit  $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ . On définit un espace probabilisé  $(\Omega, \mathcal{P}(\Omega), P)$  en se donnant une fonction  $f: \Omega \rightarrow \mathbb{R}_+$  vérifiant  $\sum_{i=1}^m f(\omega_i) = 1$  et en posant

$$P(A) = \sum_{\omega_i \in A} f(\omega_i) \quad \text{pour tout } A \subset \Omega$$

(donc  $P(\{\omega_i\}) = f(\omega_i)$   $\forall i$ ). Tout espace probabilisé fini s'obtient de cette façon.



Exemple 1. Si  $f(\omega_i) = \frac{1}{n}$  ( $\forall i$ ) on retrouve la modélisation classique du f2. Cette probabilité est aussi appelée la "probabilité uniforme" sur  $\Omega$ .

Exemple 2. Jet de 2 pièces équilibrées indiscernables :  
 $\Omega = \{PP, PF, FF\}$  avec  $P(\{PP\}) = P(\{FF\}) = \frac{1}{4}$ ,  $P(\{PF\}) = \frac{1}{2}$ .  
 Voir la remarque page 22 bis.

## 2) Espaces probabilisés infinis dénombrables (discrets).

Soit  $\Omega$  un ensemble infini dénombrable; soit  $f: \Omega \rightarrow \mathbb{R}_+$  une fonction vérifiant  $\sum_{\omega \in \Omega} f(\omega) = 1$ .

[Remarque: Que signifie  $\sum_{\omega \in \Omega} f(\omega) = 1$ ? On choisit une numérotation de  $\Omega$ :  $\omega_1, \omega_2, \dots$  et on demande que la série  $\sum_{i=1}^{\infty} f(\omega_i) = 1$ ; comme les  $f(\omega_i)$  sont  $\geq 0$  la série est absolument convergente, donc commutativement convergente et donc la valeur de  $\sum_{i=1}^{\infty} f(\omega_i)$  ne dépend pas de la numérotation choisie; c'est cette valeur qu'on notera  $\sum_{\omega \in \Omega} f(\omega)$ ].

On définit alors une probabilité sur  $(\Omega, \mathcal{P}(\Omega))$  en posant, pour tout  $A \in \mathcal{P}(\Omega)$ ,  $P(A) = \sum_{\omega \in A} f(\omega)$ .

[Il est clair que les axiomes a) et b) sont vérifiés; vérifions c): si  $A = \sum_n A_n$  on a

$$P(A) = \sum_{\omega \in A} f(\omega) = \sum_{\omega \in \sum_n A_n} f(\omega) \stackrel{!}{=} \sum_n \left( \sum_{\omega \in A_n} f(\omega) \right) = \sum_n P(A_n)$$

où l'égalité importante résulte d'un théorème sur l'associativité des séries abs. conv. (séries de séries). cf. Remarque ci-dessus].

Exemple: Loi géométrique: on considère une urne de  $n$  boules, dont  $n_1$  sont blanches et  $n_2$  sont noires ( $n_1 + n_2 = n$ ). On extrait une boule avec remplacement jusqu'à ce qu'on obtienne une boule blanche.

$$\Omega = \{(n), (n, b), (n, m, b), \dots\} \cong \{1, 2, 3, \dots\}$$

$$\text{avec } P(\underbrace{\{(n, m, \dots, m, b)\}}_r) = P(\{r\}) = \frac{n_1 \cdot n_2^{r-1}}{n^r} = p(1-p)^{r-1}$$

si  $p = \frac{n_1}{n}$ .

C'est un exemple du cas ci-dessus à  $f(r) = p(1-p)^{r-1}$ .  
 Notons qu'on a bien  $\sum_{r=1}^{\infty} p(1-p)^{r-1} = \frac{p}{1-(1-p)} = 1$ .

Les probabilités  $p(1-p)^{r-1}$  ( $r=1, 2, \dots$ ) définissent la loi géométrique; cette loi apparaît comme loi du temps d'attente (ou de la durée de vie) dans des processus "sans mémoire" (ou "sans vieillissement"). Analogie continue: loi exponentielle (voir plus loin).

## 3) Espaces probabilisés continus.

Soient  $\Omega$  un intervalle de  $\mathbb{R}$  (fini ou non),  $f: \Omega \rightarrow \mathbb{R}_+$  une fonction telle que  $\int_{\Omega} f(\omega) d\omega$  existe et  $\int_{\Omega} f(\omega) d\omega = 1$ .

Pour tout intervalle  $I \subset \Omega$  on pose

$$P(I) = \int_I f(\omega) d\omega.$$

On a  $0 \leq P(I) \leq 1$  et  $P(\Omega) = 1$ , mais l'ensemble des intervalles  $I \subset \Omega$  n'est pas une tribu!

Notons que  $P(\{\omega\}) = 0$  ( $\forall \omega \in \Omega$ ).

Le problème ici est de savoir quels seront les ensembles dont la probabilité sera définie i.e. quelle sera la tribu des événements.

Définition: On notera  $\mathcal{B}$  la plus petite tribu de parties de  $\Omega$  qui contient tous les intervalles  $I \subset \Omega$ . Elle est appelée la tribu borélienne sur  $\Omega$  et ses éléments sont les boréliens de  $\Omega$ .

Remarques. 1) La définition repose sur les constatations suivantes:

a) l'intersection d'une famille  $\{\mathcal{A}_i\}$  de tribus de parties de  $\Omega$  est une tribu de parties de  $\Omega$ ; b)  $\mathcal{B}$  est définie comme l'intersection des tribus qui contiennent tous les intervalles ( $\mathcal{P}(\Omega)$  en est une!); c)  $\mathcal{B}$  ainsi définie est bien la "plus petite"...

2) On peut montrer que  $\mathcal{B} \neq \mathcal{P}(\Omega)$ .

3) Cependant  $\mathcal{B}$  contient tous les ensembles de  $\Omega$  dont on peut avoir besoin en probabilité et en analyse.

Théorème: Il existe une unique probabilité  $P$  sur  $(\Omega, \mathcal{B})$  telle que  $P(I) = \int_I f(\omega) d\omega$  pour tout intervalle  $I \subset \Omega$ .

1<sup>ère</sup> partie de la preuve: Prenons le cas de  $\Omega = ]0, 1[$ . Soit  $\mathcal{C}$  l'ensemble des réunions finies d'intervalles  $I = ]a, b[ \subset \Omega$ . Les propriétés suivantes sont immédiates:

- (1)  $\emptyset, \Omega \in \mathcal{C}$ ; (2)  $A \in \mathcal{C} \Rightarrow \bar{A} \in \mathcal{C}$ ; (3)  $A, B \in \mathcal{C} \Rightarrow A \cup B, A \cap B \in \mathcal{C}$   
 (4) tout  $A \in \mathcal{C}$  peut s'écrire  $A = \bigcup_{j=1}^n I_j$  avec  $I_j = ]a_j, b_j[$ .

On définit alors  $P(A) = \sum_{j=1}^n P(I_j)$  pour  $A = \bigcup_{j=1}^n I_j$  et on vérifie que c'est indépendant de la décomposition choisie.  $P$  est ainsi prolongée à  $\mathcal{C}$ : c'est le 1<sup>er</sup> pas de la preuve, le plus facile...

2<sup>ème</sup> partie: Voir par exemple Rudin, Principles of Mathematical Analysis, Chap. sur l'intégrale de Lebesgue.

Exemple 1.  $\Omega =$  intervalle de longueur  $L$ ;  $f(\omega) = \frac{1}{L}$ ; alors  $P$  est déterminée par  $P(I) = \frac{\text{longueur de } I}{L}$ ; on dit que c'est la loi uniforme sur  $\Omega$  (c'est l'équivalent continu du modèle classique du §2). La probabilité  $P$  sur  $\mathcal{B}$  est la mesure de Lebesgue. Notons une conséquence de l'axiome  $C_2$ ): comme  $A = \mathbb{Q} \cap \Omega$  est dénombrable on a  $P(A) = \sum_{\omega \in A} P(\{\omega\}) = 0$ .

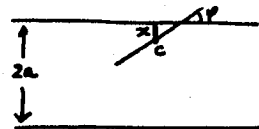
Exemple 2.  $\Omega = \mathbb{R}$ ,  $f(\omega) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\omega^2/2\sigma^2}$ : la probabilité associée est la loi normale sur  $\mathbb{R}$  (centrée, réduite).

Généralisation:  $\Omega$  région de  $\mathbb{R}^n$ ,  $f: \Omega \rightarrow \mathbb{R}_+$  avec  $\int_{\Omega} f(\omega) d\omega = 1$ . On remplace les intervalles ci-dessus par les parties de  $\Omega$ . (i.e. rectangles dans  $\mathbb{R}^2$ , ...).

Lorsque  $\Omega$  est de "volume" fini,  $f(\omega) = \frac{1}{\text{vol. de } \Omega}$  définit la loi uniforme sur  $\Omega$ , donnée par

$$P(A) = \frac{\text{vol. de } A}{\text{vol. de } \Omega}$$

### Exemple 3: L'aiguille de Buffon



Des lignes parallèles distantes de  $2a$  sont tracées dans le plan.

Une aiguille de longueur  $2l$  ( $l < a$ ) est lancée "au hasard" sur le plan.

On cherche la probabilité que l'aiguille coupe une des parallèles.

A chaque épreuve (i.e. jet de l'aiguille) correspond un couple  $(\varphi, x)$  où:  $\varphi$  est l'angle formé par l'aiguille et une parallèle  
 $x$  est la distance du centre de l'aiguille à la parallèle la plus proche. On a  $0 \leq \varphi < \pi$  et  $0 \leq x \leq a$ .

Donc  $\Omega = \{(\varphi, x) \mid 0 \leq \varphi < \pi, 0 \leq x \leq a\}$

Le jet de l'aiguille "au hasard" signifie ici que  $(\Omega, \mathcal{B})$  est muni de la loi uniforme i.e. si  $A \in \mathcal{B}$ ,



$$P(A) = \frac{\text{surface de } A}{\text{surface de } \Omega} = \frac{\text{surface de } A}{a\pi}$$

Considérons l'événement  $A =$  "l'aiguille coupe une parallèle"; on a:

$$A = \{(\varphi, x) \mid x \leq l \sin \varphi\}$$

donc

$$P(A) = \frac{1}{a\pi} \int_0^{\pi} l \sin \varphi d\varphi = \frac{2l}{a\pi}$$

Remarques: 1) Buffon (1707-1788) a étudié ce problème (ainsi que d'autres problèmes de probabilités et de applications aux tables de mortalité) dans son "Histoire naturelle". Le problème de l'aiguille est en relation avec des problèmes d'arbitraire.

2) Cette expérience peut être utilisée pour une détermination expérimentale de  $\pi$ . Si  $a = 2l$ ,  $P(A) = \frac{1}{\pi}$ .

Si on admet que la fréquence tend vers la probabilité on aura  $\pi \approx \frac{\text{Nb. total d'expériences}}{\text{Nb. de celles où l'aiguille coupe}}$ .

En 1850, Wolf en 5000 essais obtient  $\pi \approx 3,1596$ .

Des expériences sont toujours (?) faites au Palais de la Découverte à Paris pour déterminer  $\pi$  de cette façon (Encycl. Universalis).

3) Ce problème est à l'origine de ce qu'on appelait la "probabilité géométrique" et qu'on appelle aujourd'hui la "géométrie intégrale" (cf. le livre de Santaló).

NB. Dans tous les problèmes de ce type il est important de préciser ce que veut dire "au hasard" (cf. paradoxes de Bertrand - série 7).

3.3. Cousséquences élémentaires des axiomes.

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé.

Proposition 1. 1) Si  $A, B \in \mathcal{A}$ ,  $A \subset B$ , on a  $P(B-A) = P(B) - P(A)$ .

2) Si  $A, B \in \mathcal{A}$ ,  $A \subset B$ , alors  $P(A) \leq P(B)$ .

3) Si  $A \in \mathcal{A}$ , alors  $P(\bar{A}) = 1 - P(A)$ .

4) (formule de Poincaré) Si  $A_1, A_2, \dots, A_n \in \mathcal{A}$  on a

$$P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n).$$

5) (intégralité de Boole) Si  $A_1, A_2, \dots, A_n \in \mathcal{A}$  on a

$$P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$$

Preuve: 1) On a  $B = A + (B-A)$ .

2) découle directement de 1)

3) découle de 1) avec  $B = \Omega$

4) se démontre par induction sur  $n$  (cf série 1)

$n=2$ : on écrit  $A_1 \cup A_2 = A_1 + (A_2 - A_1 \cap A_2)$

d'où  $P(A_1 \cup A_2) = P(A_1) + P(A_2 - A_1 \cap A_2)$   
 $= P(A_1) + P(A_2) - P(A_1 \cap A_2)$  par 1).

On suppose la formule vraie pour  $n$ . On écrit

$$P(\bigcup_{i=1}^{n+1} A_i) = P(\bigcup_{i=1}^n A_i \cup A_{n+1})$$

$$= P(\bigcup_{i=1}^n A_i) + P(A_{n+1}) - P(\bigcup_{i=1}^n A_i \cap A_{n+1})$$

Puis on utilise l'hyp. d'induction :

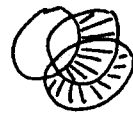
$$P(\bigcup_{i=1}^n A_i) = \sum_{1 \leq i \leq n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n)$$

$$+ P(A_{n+1})$$

$$- \sum_{1 \leq i \leq n} P(A_i \cap A_{n+1}) + \sum_{1 \leq i < j \leq n} P(A_i \cap A_j \cap A_{n+1}) - \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n \cap A_{n+1})$$

$$= \text{cf. fol.}$$

5) On pose  $B_1 = A_1$ ,  $B_2 = \bar{A}_1 \cap A_2$ ,  $B_3 = \bar{A}_1 \cap \bar{A}_2 \cap A_3$ , ...,  $B_n = \bar{A}_1 \cap \dots \cap \bar{A}_{n-1} \cap A_n$ . Les  $B_i$  sont disjoints  $\forall i \geq 2$  et  $\sum_{i=1}^n B_i = \bigcup_{i=1}^n A_i$ , donc



$$P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(B_i) \leq \sum_{i=1}^n P(A_i) \text{ car } B_i \subset A_i \text{ (ii').}$$

Notations:  $A_n \uparrow A$  si  $A_1 \subset A_2 \subset A_3 \subset \dots$  et  $\bigcup_{i=1}^{\infty} A_i = A$   
 $A_n \downarrow A$  si  $A_1 \supset A_2 \supset A_3 \supset \dots$  et  $\bigcap_{i=1}^{\infty} A_i = A$ .

Proposition 2. 1) (Continuité par dessous): si les  $A_n \in \mathcal{A}$ , si  $A_n \uparrow A$  on a  $P(A) = \lim_{n \rightarrow \infty} P(A_n) = \sup P(A_n)$ .

2) (Continuité par dessus): si les  $A_n \in \mathcal{A}$ , si  $A_n \downarrow A$  on a  $P(A) = \lim_{n \rightarrow \infty} P(A_n) = \inf P(A_n)$ .

3) (somme-additivité dénombrable): si les  $A_n \in \mathcal{A}$  on a  $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{n=1}^{\infty} P(A_n)$ .

Preuve: 1) On pose  $B_1 = A_1$ ,  $B_2 = A_2 - A_1$ ,  $B_3 = A_3 - A_2$ , ...

on a  $A = \bigcup_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} B_i$

donc  $P(A) = \sum_{n=1}^{\infty} P(B_n) = \lim_{k \rightarrow \infty} \sum_{n=1}^k P(B_n)$

$$= \lim_{k \rightarrow \infty} P(\sum_{i=1}^k B_i) = \lim_{k \rightarrow \infty} P(A_k).$$

2) Idem.

3) On a  $P(\bigcup_{i=1}^k A_i) \leq \sum_{i=1}^k P(A_i) \leq \sum_{i=1}^{\infty} P(A_i) \quad (\forall k)$

Mais  $\bigcup_{i=1}^k A_i \uparrow \bigcup_{i=1}^{\infty} A_i$  (lorsque  $k \rightarrow \infty$ ) donc par 1)

$$P(\bigcup_{i=1}^{\infty} A_i) = \lim_{k \rightarrow \infty} P(\bigcup_{i=1}^k A_i) \leq \sum_{i=1}^{\infty} P(A_i).$$

Exemple :  $\Omega = ]0,1[$  avec la loi uniforme.

Soit  $\{r_n | n \geq 1\}$  l'ens. des nb. rationnels de  $]0,1[$ .

Fixons un  $\varepsilon > 0$ . Pour tout  $n \geq 1$ , il existe un intervalle  $]a_n, b_n[ \subset ]0,1[$  avec  $r_n \in ]a_n, b_n[$  et  $b_n - a_n < \frac{\varepsilon}{2^n}$ .

Alors  $\{r_n | n \geq 1\} \subset \bigcup_{n=1}^{\infty} ]a_n, b_n[ = G$  ouvert de  $]0,1[$ .

D'après le point 3) du thm :  $0 < P(G) \leq \varepsilon$ .

Et pourtant  $G$  est dense dans  $]0,1[$ .

Difficile à imaginer !

#### §4. Probabilités conditionnelles - Événements indépendants.

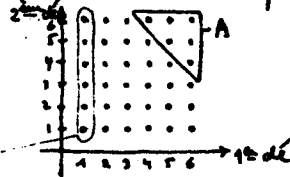
##### 4.1. Exemples et définition des probabilités conditionnelles.

Remarque : Jusqu'à maintenant, les propriétés d'une probabilité qu'on a étudiées sont en fait les propriétés d'une mesure  $\sigma$ -additive comme on en rencontre dans plusieurs domaines (intégration, dimension des sous-espaces dans un Hilbert, ...). Dans ce § nous introduisons une idée propre à la théorie des probabilités : la probabilité conditionnelle, qui s'introduit naturellement chaque fois que l'on acquiert une information partielle sur le résultat d'une expérience aléatoire.

Exemple 1 : On jette un dé deux fois de suite. L'espace des épreuves est  $\Omega = \{1,2,3,4,5,6\}^2$  avec  $\mathcal{O} = \mathcal{P}(\Omega)$  et  $P$  la probabilité uniforme (i.e. le modèle classique).

On s'intéresse à l'événement  $A = \{\omega = (i,j) \mid i+j \geq 10\}$  i.e.

"la somme des points est  $\geq 10$ ".



$$\text{On a : } P(A) = \frac{6}{36} = \frac{1}{6}$$

Si on connaît le résultat du 1<sup>er</sup> lancer, il est clair que la probabilité de  $A$  va changer.

Soit  $B_i$  l'événement "le 1<sup>er</sup> lancer donne  $i$ " i.e.  $B_i = \{(i,j) \mid 1 \leq j \leq 6\}$

Notons  $P(A|B_i)$  la probabilité de  $A$  sachant que  $B_i$  s'est déjà réalisé ou - comme on dit - la probabilité conditionnelle de  $A$  sachant  $B_i$  ; on a ici

$$P(A|B_1) = P(A|B_2) = P(A|B_3) = 0, \quad P(A|B_4) = \frac{1}{6}, \quad P(A|B_5) = \frac{2}{6}$$

$$\text{et } P(A|B_6) = \frac{3}{6}.$$

Sur cet exemple on voit que  $P(A|B) = \frac{N(A \cap B)}{N(B)}$   
 Pour pouvoir exprimer  $P(A|B)$  dans le modèle général - et pas seulement dans le modèle classique - on remarque aussi que

$$P(A|B) = \frac{N(A \cap B)}{N(B)} = \frac{\frac{N(A \cap B)}{N(\Omega)}}{\frac{N(B)}{N(\Omega)}} = \frac{P(A \cap B)}{P(B)}$$

Donc :

Définition : Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Etant donné un événement  $B \in \mathcal{A}$  tel que  $P(B) > 0$ , on appelle probabilité conditionnelle sachant B l'application  $A \mapsto P(A|B) = \frac{P(A \cap B)}{P(B)}$  définie sur  $(\Omega, \mathcal{A})$ .

Remarque :  $P(\cdot|B)$  est une probabilité sur  $(\Omega, \mathcal{A})$  : la vérification est immédiate.

Exemple 2. On considère une famille de 2 enfants ; l'espace des épreuves est  $\Omega = \{FF, FG, GF, GG\}$  où FG signifie que l'aîné est une fille et le cadet un garçon, etc.  
 A chaque événement élémentaire on attribue la probabilité  $1/4$ .  
 On considère les éléments :

A = "les 2 enfants sont des garçons" =  $\{GG\}$  ;  $P(A) = 1/4$

B = "la famille a un garçon" =  $\{FG, GF, GG\}$  ;  $P(B) = 3/4$

C = "l'aîné est un garçon" =  $\{GF, GG\}$  ;  $P(C) = 2/4$ .

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = 1/3$$

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{1/4}{2/4} = 1/2$$

Donc attention à la précision de la formulation !

Remarque. L'événement B étant réalisé on pourrait modéliser l'expérience par  $(B, \mathcal{A}_B, P_B)$  où  $\mathcal{A}_B = \{A \cap B | A \in \mathcal{A}\}$  et  $P_B(A) = P(A|B)$ . Mais la schématisation  $(\Omega, \mathcal{A}, P(\cdot|B))$  est agréable en ce qu'elle ne change pas l'espace des épreuves.

### 4.2. Propriétés des probabilités conditionnelles.

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé.

Prop. 1. Si  $A_1, A_2, \dots, A_n \in \mathcal{A}$  on a si  $P(\bigcap_{i=1}^n A_i) \neq 0$  :

$$P(\bigcap_{i=1}^n A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

Preuve.  $n=2$  :  $P(A \cap B) = P(B)P(A|B)$  découle de la définition de la probabilité conditionnelle. Puis par récurrence : on a  $P(\bigcap_{i=1}^n A_i) = P(\bigcap_{i=1}^{n-1} A_i \cap A_n) = P(\bigcap_{i=1}^{n-1} A_i)P(A_n|\bigcap_{i=1}^{n-1} A_i)$  d'où cf. cd. car  $P(\bigcap_{i=1}^{n-1} A_i) \neq 0 \rightarrow P(\bigcap_{i=1}^n A_i) \neq 0$ .

Exemple : Soit  $p_k$  la probabilité pour une personne d'âge  $k$  d'être en vie à l'âge  $k+n$ . Soit  $A_n$  l'événement "une personne d'âge  $k$  est en vie à l'âge  $k+n$ " ; on a  $A_i \supset A_{i+1}$  ( $\forall i$ ) et

$$P(A_n) = P(\bigcap_{i=1}^n A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

$$= p_k \cdot p_{k+1} \cdot p_{k+2} \dots p_{k+n-1}$$

Prop. 2. Soit  $\{B_j\}$  une partition finie ou dénombrable de  $\Omega$ .

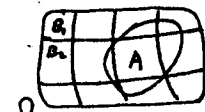
1) Pour tout  $A \in \mathcal{A}$  on a  $P(A) = \sum_j P(A|B_j)P(B_j)$   
 (règle des probabilités composées)

2) Si de plus  $P(A) > 0$  on a, pour tout  $i$ ,

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$$

(formule de Bayes)

Preuve : 1) On a  $\Omega = \sum_j B_j$  donc  $A = \sum_j (A \cap B_j)$



donc  $P(A) = \sum_j P(A \cap B_j) = \sum_j P(A|B_j)P(B_j)$

2)  $\forall i$  :  $P(A)P(B_i|A) = P(A \cap B_i) = P(A|B_i)P(B_i)$  d'où :

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$$

Remarque. La "règle de probabilités composées" est souvent très utile pour calculer des probabilités : si la partition  $\{B_j\}$  est bien choisie il est facile de calculer la  $P(A|B_j)$ .

Exemple 1. On considère l'ensemble de compositions possibles en filles et garçons des familles qui ont un nombre quelconque d'enfants. On prendra

$$\Omega = \{O, G, F, GG, GF, FG, FF, GGG, \dots\}$$

Soit A l'événement "la famille n'a pas de fille".

Considérons  $B_n$  l'événement "la famille a n enfants" et posons  $p_n = P(B_n)$ . On a  $\Omega = \sum_{n=0}^{\infty} B_n$ , donc  $\sum_{n=0}^{\infty} p_n = 1$ .

Supposons que  $\forall n$ , les compositions d'une famille de n enfants sont équiprobables (il y a  $2^n$  possibilités), donc

$$P(A|B_n) = \frac{1}{2^n}$$

$$\text{Alors } P(A) = \sum_{n=0}^{\infty} P(A|B_n)P(B_n) = \sum_{n=0}^{\infty} \frac{p_n}{2^n}$$

On peut aussi appliquer la formule de Bayes à cet exemple. Quelle est la probabilité qu'une famille sans fille n'ait qu'un seul enfant? C'est

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{\frac{1}{2} p_1}{\sum_{n=0}^{\infty} \frac{p_n}{2^n}}$$

Exemple 2. On dispose d'un test T pour une maladie M.

- $p$  = proba. qu'un individu soit atteint de M.
- $0,95$  = proba. que le test soit positif pour un individu atteint de M
- $0,95$  = "

- A = "l'individu répond positivement au test"
- $B_1$  = "l'individu est atteint de M"
- $B_2 = \bar{B}_1$

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = \frac{0,95 \cdot p}{0,95 \cdot p + 0,05(1-p)}$$

Remarques: 1) Les problèmes qu'on résout par la formule de Bayes sont souvent appelés problèmes de "probabilités des causes". Ils concernent des expériences qui sont faites en 2 étapes : d'abord on tire au hasard une classe  $B_j$ , puis on prend au hasard un élément de  $B_j$ . Au vu de ce résultat on cherche à connaître le résultat de la 1<sup>ère</sup> étape.

2) La formule de Bayes (XVIII<sup>ème</sup> siècle) est devenue célèbre à cause de l'usage quasi métaphysique qu'on en a fait. Interprétons les  $B_j$  comme les états dans lesquels peut se trouver la nature et A comme le résultat possible d'une expérience faite dans le but de déterminer dans lequel des états  $B_j$  se trouve la nature. Alors

$P(B_j)$  = probabilité a priori = opinion de l'observateur avant l'expérience.

$P(A|B_j)$  = probabilité d'observer A si la nature est dans l'état  $B_j$

$P(B_j|A)$  = probabilité a posteriori = nouvelle opinion de l'observateur après l'expérience, sachant que A s'est réalisé.

Ainsi la formule de Bayes apparaît comme une règle méthodologique : elle dit comment on doit modifier son opinion en tenant compte du résultat d'une expérience.

Le abus de cette formule provient de l'impossibilité qu'il y a à attribuer une probabilité  $P(B_j)$  à une affirmation qq. concernant la nature.

Cependant, depuis une vingtaine d'années, ce principe connaît un regain d'intérêt, surtout en sciences économiques, en sciences actuelles, où l'expérimentation est difficile et l'observation prend du temps... ; il est à la base de la théorie de la décision et de la théorie de la crédibilité.

### 4.3. Evénements indépendants.

Définition. Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Deux événements  $A, B \in \mathcal{A}$  sont indépendants si  $P(A \cap B) = P(A)P(B)$ .

Remarques: 1)  $A, B$  indépendants  $\Leftrightarrow A, \bar{B}$  indépendants  
 $\Leftrightarrow \bar{A}, B$  indépendants  $\Leftrightarrow \bar{A}, \bar{B}$  indépendants.

$$[A, B \text{ indép.} \Rightarrow P(A \cap \bar{B}) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)P(\bar{B})]$$

2) Si  $P(A) = 0$  ou  $1$  alors  $A$  et  $B$  sont indépendants, car  $A \cap B \subset A$  implique que  $P(A \cap B) = 0$  si  $P(A) = 0$ .  
 Dans le cas  $P(A) = 1$ , utiliser 1).

3) Ne pas confondre " $A, B$  indépendants" et " $A$  et  $B$  incompatibles".

Le sens intuitif d'"indépendants" apparaît mieux dans le résultat suivant:

Proposition: Si  $0 < P(A) < 1$  et  $0 < P(B) < 1$  les affirmations suivantes sont équivalentes:

- (1)  $A$  et  $B$  sont indépendants,
- (2)  $P(A) = P(A|B) = P(A|\bar{B})$ ,
- (3)  $P(B) = P(B|A) = P(B|\bar{A})$ .

Preuve: (1)  $\rightarrow$  (2):  $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$ ; d'après la Remarque 1)  $A$  et  $\bar{B}$  sont aussi indépendants d'où  $P(A|\bar{B}) = P(A)$ .  
 (2)  $\rightarrow$  (1): immédiat.

Exemple 1. Jet de 2 pièces discernables;  $A$  = la 1<sup>ère</sup> pièce montre pile;  $B$  = la 2<sup>ème</sup> pièce montre face; on a  $P(A) = P(B) = \frac{1}{2}$  et  $P(A \cap B) = \frac{1}{4} = P(A)P(B)$ , donc  $A$  et  $B$  sont indépendants!

Exemple 2. Jet d'un dé;  $A$  = on obtient un nombre pair;  $B$  = on obtient 1 ou 6;  $P(A) = \frac{1}{2}$ ,  $P(B) = \frac{1}{3}$ .  
 $A \cap B$  = on obtient 6;  $P(A \cap B) = \frac{1}{6} = P(A)P(B)$  donc  $A$  et  $B$  sont indépendants. NB. Ce n'est plus nécessairement vrai si le dé est pipé!!

Exemple 3. Considérons l'ensemble des compositions possibles en filles et garçons d'une famille de 3 enfants:  
 $\Omega = \{GGG, GGF, GFG, FGG, GFF, FGF, FFG, FFF\}$  avec la probabilité uniforme.

$A$  = il y a des enfants des deux sexes;  $P(A) = \frac{6}{8} = \frac{3}{4}$   
 $B$  = il y a au plus une fille;  $P(B) = \frac{4}{8} = \frac{1}{2}$   
 $A \cap B = \{GGF, GFG, FGG\}$ ;  $P(A \cap B) = \frac{3}{8} = P(A)P(B)$ , donc  $A$  et  $B$  sont indépendants.

Les mêmes événements pour une famille de 4 enfants donnent  $P(A) = \frac{14}{16} = \frac{7}{8}$ ;  $P(B) = \frac{5}{16}$ ;  $P(A \cap B) = \frac{4}{16} \neq P(A)P(B)$ , donc  $A$  et  $B$  ne sont pas indépendants!

Donc attention: si l'"indépendance a priori" des événements  $A$  et  $B$  implique l'indépendance stochastique, elle ne lui est pas équivalente! Noter aussi que si  $A$  et  $B$  ne sont pas indépendants il n'y a pas nécessairement de relation directe entre  $A$  et  $B$ , mais  $A$  et  $B$  peuvent simplement être influencés par un autre événement [p.ex. la production de 2 pommes est influencée par le temps].

Exemple 4: cf. pages 39, 39".

Généralisation à  $n$  événements  $A_1, \dots, A_n \in \mathcal{A}$ .

Définitions. 1)  $A_1, \dots, A_n$  sont indépendants 2 à 2 si  $P(A_i \cap A_j) = P(A_i)P(A_j)$  ( $i \neq j$ )

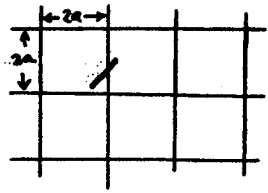
2)  $A_1, \dots, A_n$  sont (mutuellement) indépendants si  $P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$

pour tout  $k, 1 \leq k \leq n$ , et toute suite  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ .

Remarque: 1)  $\not\rightarrow$  2) comme le montre l'exemple suivant:  
 $\Omega = \{1, 2, 3, 4\}$  avec la probabilité uniforme;  $A_i = \{i, 4\}$ ,  $i=1, 2, 3$ ;  
 on a  $P(A_i) = \frac{2}{4} = \frac{1}{2}$ ,  $P(A_i \cap A_j) = P(\{4\}) = \frac{1}{4} = P(A_i)P(A_j)$  ( $i \neq j$ )  
 mais  $P(A_1 \cap A_2 \cap A_3) = P(\{4\}) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3)$ .

### Exemple 4 : Aiguille de Buffon (suite).

On considère un double réseau de parallèles orthogonales



à distance  $2a$  les uns des autres. Une aiguille de longueur  $2l$  ( $l < a$ ) est lancée au hasard. On considère les événements

A : "l'aiguille coupe une horizontale"

B : "l'aiguille coupe une verticale"

on a déjà vu que  $P(A) = P(B) = \frac{2l}{a\pi}$ .

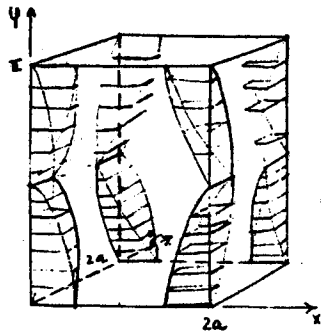
Question: Les événements A et B sont-ils indépendants ?

Pour répondre il faut calculer  $P(A \cap B)$ .

Le centre de l'aiguille tombe dans un carré ; la probabilité de  $A \cap B$  ne dépend pas des choix de ce carré ; on peut donc le supposer fixé (cf. formule de probabilités composées).

Une épreuve est donc complètement déterminée par :

- un point  $(x, y)$  du carré  $[0, 2a] \times [0, 2a]$  (= le centre de l'aiguille)
- un angle  $\varphi \in [0, \pi]$  (= l'angle de l'aiguille avec Ox).

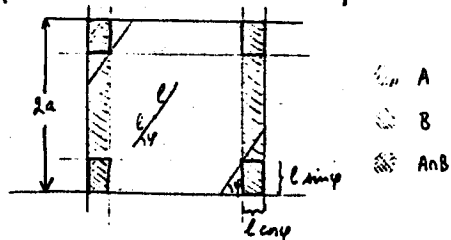


$$\Omega = [0, 2a] \times [0, 2a] \times [0, \pi]$$

avec la probabilité uniforme.

on cherche  $A \cap B \subset \Omega$ .

Coupe de  $\Omega$  à la hauteur  $\varphi$  :



- A
- B
- ⊗  $A \cap B$

Avec ce modèle on obtient :

$$P(A) = \frac{1}{4a^2\pi} \int_0^\pi 2 \cdot 2a \cdot l \sin \varphi \, d\varphi = \frac{2l}{a\pi}$$

$$P(B) = \frac{1}{4a^2\pi} \int_0^\pi 2 \cdot 2a \cdot l |\cos \varphi| \, d\varphi = \frac{2l}{a\pi}$$

ce qui correspond à ce qu'on avait calculé.

$$P(A \cap B) = \frac{1}{4a^2\pi} \int_0^\pi 4 \cdot l^2 \sin \varphi |\cos \varphi| \, d\varphi = \frac{l^2}{a^2\pi} \int_0^{\pi/2} 2 \sin \varphi \cos \varphi \, d\varphi$$

$$\text{i.e. } P(A \cap B) = \frac{l^2}{a^2\pi}$$

alors que  $P(A)P(B) = \frac{4l^2}{a^2\pi^2}$  ; donc A et B ne sont pas indép.

[Par ex. si  $2l = a$  on a  $P(A) = P(B) = \frac{1}{\pi}$  ;  $P(A \cap B) = \frac{1}{4\pi}$ ]

Remarque :

1) La dépendance de A et B se comprend : si  $\varphi$  est voisin de 0 ou  $\pi$ , A est très peu probable alors que la probabilité de B est à son maximum ; c'est le contraire si  $\varphi$  est voisin de  $\frac{\pi}{2}$ . Donc si A se produit il est probable que  $\varphi$  est voisin de  $\frac{\pi}{2}$ , donc B a moins de chances de se produire.

[Analogie : on cultive deux sortes de légumes ; la 1<sup>ère</sup> sorte a besoin de sec, la 2<sup>ème</sup> a besoin d'eau ; les productions annuelles ne sont pas indépendantes : si la 1<sup>ère</sup> sorte a donné une bonne récolte il est probable que le temps a été sec, donc que la 2<sup>ème</sup> sorte n'a pas donné une bonne récolte...]

2) lorsque  $\varphi$  est fixé, les événements A et B sont indépendants (cf. page précédente) car  $P_\varphi(A) = \frac{4al \sin \varphi}{4a^2}$ ,  $P_\varphi(B) = \frac{4al |\cos \varphi|}{4a^2}$  et  $P_\varphi(A \cap B) = \frac{4l^2 \sin \varphi |\cos \varphi|}{4a^2} = P_\varphi(A)P_\varphi(B)$ .



Proposition 2. Soient  $A_1, \dots, A_n \in \mathcal{O}$  des événements indépendants.

- 1) Les événements  $B_1, \dots, B_m$  où  $B_i = A_i$  ou  $\bar{A}_i$  ( $\bar{A}_i$ ) sont indép.
- 2) Les événements  $C_1, \dots, C_s$  où les  $C_i$  sont des réunions ou des intersections de sous-familles disjointes de  $\{A_1, \dots, A_n\}$ , sont indépendants.

Preuve: 1) Il suffit de le voir pour  $A_1, \dots, A_{i-1}, \bar{A}_i, A_{i+1}, \dots, A_n$ , car en répétant le procédé on obtiendra toutes les répartitions possibles de  $A_i$  et  $\bar{A}_i$ . Il suffit encore de le faire pour  $\bar{A}_1, A_2, \dots, A_n$ . Mais alors si  $1 < i_1 < \dots < i_k < n$  on a

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$$

et

$$\begin{aligned} P(\bar{A}_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) &= P(A_{i_1} \cap \dots \cap A_{i_k}) - P(A_2 \cap A_{i_2} \cap \dots \cap A_{i_k}) \\ &= P(A_{i_1}) \dots P(A_{i_k}) - P(A_2) P(A_{i_2}) \dots P(A_{i_k}) \\ &= P(\bar{A}_2) P(A_{i_2}) \dots P(A_{i_k}). \end{aligned}$$

- 2) En particulier:  $k=2$ ,  $C_1 = \bigcup_{i=1}^r A_i$ ,  $C_2 = \bigcap_{i=1}^s A_i$  avec  $r < s$ . Il suffit de voir que  $\bar{C}_1$  et  $C_2$  sont indépendants, mais  $\bar{C}_1 = \bigcap_{i=1}^r \bar{A}_i$  et  $\bar{C}_1 \cap C_2 = \bar{A}_1 \cap \dots \cap \bar{A}_r \cap A_{r+1} \cap \dots \cap A_n$ ; Comme les  $A_i$  sont indépendants on a par 1)

$$P(\bar{C}_1 \cap C_2) = P(\bar{A}_1) \dots P(\bar{A}_r) P(A_{r+1}) \dots P(A_n) = P(\bar{C}_1) P(C_2).$$

Même preuve pour le cas général !

Proposition 3. Si  $A = \bigcup_{i=1}^m A_i$  et  $B = \bigcap_{j=1}^n B_j$  et si  $A_i$  et  $B_j$  sont indépendants pour tout couple  $(i, j)$  alors  $A$  et  $B$  sont indép.

Preuve:  $A \cap B = \sum_{i,j} A_i \cap B_j$  donc  $P(A \cap B) = \sum_{i,j} P(A_i \cap B_j)$   
 $= \sum_{i,j} P(A_i) P(B_j) = (\sum_i P(A_i)) (\sum_j P(B_j)) = P(A) P(B).$

## §5. Schéma de Bernoulli

But: Etudier une suite de  $n$  répétitions "indépendantes" d'une même expérience aléatoire. "Indépendantes" signifie que le résultat d'une expérience n'influence pas les résultats des autres.

### 5.1. Modélisation.

On considère une expérience aléatoire  $E_0$ , modélisée par l'espace probabilisé  $(\Omega_0, \mathcal{O}_0, P_0)$ . On se fixe un entier  $n$  et on considère l'expérience aléatoire  $E$  qui consiste en  $n$  répétitions "indépendantes" de  $E_0$ . Quel est l'espace probabilisé qui modélise  $E$ ?

Hypothèse simplificatrice (mais non essentielle): L'espace probabilisé  $(\Omega_0, \mathcal{O}_0, P_0)$  est discret i.e.  $\Omega_0$  est fini ou dénombrable,  $\mathcal{O}_0 = \mathcal{P}(\Omega_0)$ ,  $P_0(A) = \sum_{\omega \in A} P_0(\{\omega\})$ .

Définissons alors un espace probabilisé (discret)  $(\Omega, \mathcal{O}, P)$  par:

- (i)  $\Omega = \Omega_0 \times \dots \times \Omega_0 = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in \Omega_0, \forall i\}$
- (ii)  $\mathcal{O} = \mathcal{P}(\Omega)$
- (iii) Si  $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ ,  $P(\{\omega\}) = \prod_{i=1}^n P_0(\{\omega_i\})$ .

Il suffit de vérifier que  $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$ ; or:

$$\sum_{\omega \in \Omega} P(\{\omega\}) = \sum_{\omega_1, \dots, \omega_n \in \Omega_0} P_0(\{\omega_1\}) \dots P_0(\{\omega_n\}) = \left( \sum_{\omega_1 \in \Omega_0} P_0(\{\omega_1\}) \right) \dots \left( \sum_{\omega_n \in \Omega_0} P_0(\{\omega_n\}) \right) = 1.$$

Alors, pour tout  $A \subset \Omega$ ,  $P(A) = \sum_{\omega \in A} P(\{\omega\})$ .

$(\Omega, \mathcal{P}(\Omega), P)$  est l'espace probabilisé produit de  $n$  copies de  $(\Omega_0, \mathcal{P}(\Omega_0), P_0)$ . On dit aussi que c'est un schéma de Bernoulli.

Remarque: Ce n'est évidemment pas la seule façon de définir une probabilité sur  $(\Omega, \mathcal{P}(\Omega))$  (cf. p. ex. série 9). Celle

qu'on choisit ici tient compte de l'hypothèse d'"indépendance" des expériences répétées. Voyons comment.

Lemme 1. Si  $A_1, \dots, A_n \subset \Omega_0$ ,  $P(A_1 \times \dots \times A_n) = P_0(A_1) \dots P_0(A_n)$ .

$$\begin{aligned} \text{Preuve: } P(A_1 \times \dots \times A_n) &= \sum_{\omega \in A_1 \times \dots \times A_n} P(\{\omega\}) = \sum_{\substack{\omega_1 \in A_1, \dots, \omega_n \in A_n}} P_0(\{\omega_1\}) \dots P_0(\{\omega_n\}) \\ &= \left( \sum_{\omega_1 \in A_1} P_0(\{\omega_1\}) \right) \dots \left( \sum_{\omega_n \in A_n} P_0(\{\omega_n\}) \right) = P_0(A_1) \dots P_0(A_n). \end{aligned}$$

Etant donné un événement  $A \subset \Omega_0$ , on peut lui associer les événements  $A^{(i)} \subset \Omega$ ,  $i \in \{1, \dots, n\}$ , définis par

$$A^{(i)} = "A \text{ se réalise à la } i\text{-ième expérience}"$$

$$= \Omega_0 \times \dots \times \Omega_0 \times A \times \Omega_0 \times \dots \times \Omega_0$$

Lemme 2.  $P(A^{(i)}) = P_0(A)$ , quel que soit  $i$ ,  $i \in \{1, \dots, n\}$ .

Preuve: Immédiat par le lemme 1.

Etant donné  $n$  événements  $A_1, \dots, A_n \subset \Omega_0$  (ils peuvent être distincts ou non), considérons les  $n$  événements  $A_1^{(i)}, \dots, A_n^{(n)} \subset \Omega$ .

Proposition. Les événements  $A_1^{(i)}, \dots, A_n^{(n)}$  sont indépendants (dans  $(\Omega, \mathcal{P}(\Omega), P)$ ).

Preuve: soit  $I \subset \{1, 2, \dots, n\}$ ; il faut voir que  $P(\bigcap_{i \in I} A_i^{(i)}) = \prod_{i \in I} P(A_i^{(i)})$ .

soit  $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ ; on a:

$$\omega \in \bigcap_{i \in I} A_i^{(i)} \iff \forall i \in I, \omega \in A_i^{(i)} \iff \forall i \in I, \omega_i \in A_i$$

$$\text{donc } \bigcap_{i \in I} A_i^{(i)} = B_I \times \dots \times B_n \quad \text{où } B_i = \begin{cases} A_i & i \in I \\ \Omega_0 & i \notin I \end{cases}$$

$$\text{Alors } P\left(\bigcap_{i \in I} A_i^{(i)}\right) = \prod_{i \in I} P_0(B_i) = \prod_{i \in I} P_0(A_i) = \prod_{i \in I} P(A_i^{(i)}).$$

En résumé: le choix de la probabilité  $P$  sur  $(\Omega, \mathcal{P}(\Omega))$  a traduit notre hypothèse d'"indépendance" par le résultat suivant:

"quels que soient les événements  $A_1, \dots, A_n \subset \Omega_0$ , les événements  $A_1^{(1)}, \dots, A_n^{(n)}$  sont indépendants". Ce modèle est confirmé par l'expérience.

## 5.2. Loi binomiale.

Considérons un schéma de Bernoulli et soit  $A \subset \Omega_0$  avec  $P_0(A) = p$  et  $P_0(\bar{A}) = 1 - p = q$ .

Proposition. La probabilité que  $A$  se réalise exactement  $h$  fois au cours de  $n$  expériences [i.e. que  $h$  exactement des événements  $A^{(i)}$ ,  $i \in \{1, \dots, n\}$ , se réalisent] est égale à

$$b(h; n, p) = \binom{n}{h} p^h q^{n-h} \quad \forall h, 0 \leq h \leq n.$$

Preuve: soit  $K \subset \{1, 2, \dots, n\}$  un sous-ensemble de  $h$  éléments.

soit  $B_K = \left(\bigcap_{i \in K} A^{(i)}\right) \cap \left(\bigcap_{i \notin K} \bar{A}^{(i)}\right)$ ; par indépendance (cf §4.3) on a

$$P(B_K) = \prod_{i \in K} P(A^{(i)}) \prod_{i \notin K} P(\bar{A}^{(i)}) = p^h q^{n-h}.$$

soit  $E_h$  l'événement " $h$  exactement des  $A^{(i)}$  sont réalisés";

on a  $E_h = \sum_K B_K$  (somme sur l'ens. des parties  $K$  à  $h$  élém.) donc

$$P(E_h) = \sum_K P(B_K) = \binom{n}{h} p^h q^{n-h}.$$

Remarques: 1) On a  $\sum_{h=0}^n b(h; n, p) = (p+q)^n = 1$ , donc les  $b(h; n, p)$  définissent une loi de probabilité sur  $\{0, 1, \dots, n\}$ : c'est la loi binomiale (cf. §2.6).

2) Posons  $b_h = b(h; n, p)$ ; on a

$$\frac{b_h}{b_{h-1}} = \frac{\frac{n!}{h!(n-h)!} p^h q^{n-h}}{\frac{n!}{(h-1)!(n-h+1)!} p^{h-1} q^{n-h+1}} = \frac{(n-h+1)p}{hq} = 1 + \frac{(n+1)p-h}{hq}$$

donc  $b_{h-1} < b_h \iff h < (n+1)p$ .

Si on pose  $m = [(n+1)p]$  on a  $b_0 < b_1 < \dots < b_{m-1} \leq b_m > b_{m+1} > \dots > b_n$  (l'égalité a lieu si  $(n+1)p$  est entier).

NB.  $m$  est le nombre le plus probable de réalisations de  $A$ ; cependant  $b(m; n, p)$  peut être très petit!

Exemple: Jet d'une pièce:  $n=100$ ,  $p=\frac{1}{2}$ ,  $m=50$   $b_{50} = \binom{100}{50} \frac{1}{2^{100}} \approx 0,04847$  (cf. page 18).

5.3. Loi multinomiale.

Considérons une partition de  $\Omega_0$  :  $A_1 + A_2 + \dots + A_r = \Omega_0$   
et posons  $P(A_i) = p_i$  ; on a  $p_1 + p_2 + \dots + p_r = 1$ .

Proposition. La probabilité que  $A_1$  se produise  $k_1$  fois, ...,  $A_r$  se produise  $k_r$  fois ( $k_1 + k_2 + \dots + k_r = n$ ) est

$$m(k_1, \dots, k_r; p_1, \dots, p_r) = \frac{n!}{k_1! \dots k_r!} p_1^{k_1} \dots p_r^{k_r}$$

Preuve: soit  $\alpha$  une partition de  $\{1, 2, \dots, n\}$  en  $r$  parties,  $\{1, 2, \dots, n\} = K_1 + \dots + K_r$  avec  $N(K_i) = k_i$ . Soit

$$B_\alpha = \left( \prod_{i \in K_1} A_1^{(i)} \right) \cap \dots \cap \left( \prod_{i \in K_r} A_r^{(i)} \right) ;$$

on a  $P(B_\alpha) = p_1^{k_1} \dots p_r^{k_r}$ . Puis on a

$$E_{k_1, \dots, k_r} = \text{" } A_i \text{ se réalise } k_i \text{ fois, } i=1, \dots, r \text{"}$$
$$= \sum_{\alpha} B_\alpha \text{ (somme sur toutes les partitions adéquates)}$$

Comme il y a  $\binom{n}{k_1} \binom{n-k_1}{k_2} \dots \binom{n-k_1-k_2-\dots-k_{r-1}}{k_r} = \frac{n!}{k_1! \dots k_r!}$  telles partitions, on a

$$P(E_{k_1, \dots, k_r}) = \frac{n!}{k_1! \dots k_r!} p_1^{k_1} \dots p_r^{k_r}$$

Remarque:  $\sum_{k_i} m(k_1, \dots, k_r; p_1, \dots, p_r) = (p_1 + \dots + p_r)^n = 1$  ; les probabilités  $m(k_1, \dots, k_r; p_1, \dots, p_r)$  définissent donc une loi de probabilité : c'est la loi multinomiale.

Exemple: on jette 12 fois un dé ; quelle est la probabilité d'obtenir 2 fois chaque face ?

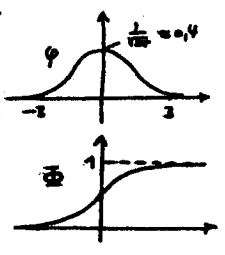
On a :  $n=12$ ,  $r=6$ ,  $k_i=2 \forall i$  ; la probabilité cherchée est

$$m(2,2,2,2,2,2; \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}) = \frac{12!}{2!2!2!2!2!2!} \left(\frac{1}{6}\right)^{12}$$
$$= \frac{12!}{2^6} \cdot \frac{1}{6^{12}} = 0,0034 \dots$$

5.4. Approximation normale de la loi binomiale.

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

et  $\Phi(x) = \int_{-\infty}^x \varphi(t) dt$ .



On dit que  $\varphi$  est la densité de la loi normale et  $\Phi$  la fonction de répartition de la loi normale.

Le but de ce n° est d'utiliser  $\varphi$  et  $\Phi$  pour calculer les probabilités données par la loi binomiale.

Étant donné un entier  $k$ ,  $0 \leq k \leq n$ , posons  $x_{n,k} = \frac{k-np}{\sqrt{npq}}$  et

$$\sqrt{npq} \ell(k; n, p) = \varphi(x_{n,k}) + r_{n,k}$$

On montrera que  $\lim_{n \rightarrow \infty} (\max_{\text{où } k} |r_{n,k}|) = 0$  (\*)

On a d'abord tout d'abord le théorème local de DeMoivre-Laplace (\*\*)

$$\ell(k; n, p) \approx \frac{1}{\sqrt{npq}} \varphi(x_{n,k}) \text{ (si } n \text{ est grand)}$$

Notons  $S_n$  le nombre de réalisations de l'événement  $A$  en  $n$  répétitions de  $\mathcal{E}_0$  (cf 5.2) ; notons  $Pr(\cdot)$  la probabilité de l'événement décrit dans la parenthèse (et qu'on pourrait écrire  $P(B)$  pour une partie  $B$  adéquate de  $\Omega$ ) ; on a par exemple  $Pr(S_n = k) = \ell(k; n, p)$ . On verra qu'on peut déduire de (\*) le théorème intégral de DeMoivre-Laplace :

$$Pr\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) \approx \Phi(b) - \Phi(a) \text{ (si } n \text{ est grand)}$$

Notation:  $a_n$  et  $b_n$  sont deux suites de nombres

$$a_n \sim b_n \text{ signifie } \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$$

On démontre en Math. I la formule de Stirling :

$$n! \sim \sqrt{2\pi n} n^n e^{-n}$$

(\*\*) DeMoivre (1667-1754) a traité le cas  $p=1/2$   
Laplace (1749-1825) a traité le cas général

Théorème 1. Soient  $0 < p < 1$  et  $(k_n)$  une suite d'entiers avec  $0 \leq k_n \leq n$ .  
 Posons  $x_n, k_n = \frac{k_n - np}{\sqrt{npq}}$  pour tout  $n$ . Si  $\lim_{n \rightarrow \infty} \frac{x_n, k_n}{n^{1/6}} = 0$  alors  
 $b(k_n; n, p) \sim \frac{1}{\sqrt{npq}} \varphi(x_n, k_n)$ .

Preuve: Pour simplifier posons  $k = k_n, x = x_n, j = n - k$ .

On a :

$$k = np + x\sqrt{npq} \quad \text{et} \quad j = nq - x\sqrt{npq}$$

Comme  $\frac{x}{n^{1/6}} \rightarrow 0$  on a  $\frac{x}{\sqrt{n}} \rightarrow 0$  donc  $\frac{k}{n} \rightarrow p, \frac{j}{n} \rightarrow q$  et  $k, j \rightarrow +\infty$ .

Par Stirling on a

$$b(k; n, p) = \frac{n!}{k!j!} p^k q^j \sim \frac{\sqrt{2\pi n} n^n e^{-n}}{\sqrt{2\pi k} k^k e^{-k} \sqrt{2\pi j} j^j e^{-j}} p^k q^j = \sqrt{\frac{n}{2\pi k j}} \left(\frac{k}{np}\right)^k \left(\frac{j}{nq}\right)^j$$

donc

$$\frac{b(k; n, p)}{\frac{1}{\sqrt{npq}} \varphi(x)} \sim \frac{\sqrt{\frac{n}{2\pi k j}} \left(\frac{k}{np}\right)^k \left(\frac{j}{nq}\right)^j}{\frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}} = \underbrace{\sqrt{\frac{n^2 pq}{k j}}}_{A_n} \underbrace{\left(\frac{k}{np}\right)^k \left(\frac{j}{nq}\right)^j}_{B_n} e^{\frac{x^2}{2}}$$

Il faut voir que  $A_n \rightarrow 1$  et  $B_n \rightarrow 1$  si  $n \rightarrow \infty$ .

On a  $A_n \rightarrow 1$  car  $A_n = \sqrt{\frac{pq}{\frac{k}{n} \frac{j}{n}}}$  et  $\frac{k}{n} \rightarrow p, \frac{j}{n} \rightarrow q$ .

Pour montrer que  $B_n \rightarrow 1$  on montrera que  $\log B_n \rightarrow 0$ . On a

$$\begin{aligned} \log B_n &= -k \log\left(\frac{k}{np}\right) - j \log\left(\frac{j}{nq}\right) + \frac{x^2}{2} \\ &= -(np + x\sqrt{npq}) \log\left(1 + x\sqrt{\frac{q}{np}}\right) - (nq - x\sqrt{npq}) \log\left(1 - x\sqrt{\frac{q}{np}}\right) + \frac{x^2}{2} \end{aligned}$$

Comme  $\frac{x}{\sqrt{n}} \rightarrow 0$  on a  $x\sqrt{\frac{q}{np}} \rightarrow 0$  et  $x\sqrt{\frac{p}{nq}} \rightarrow 0$ ; il existe donc un entier  $n_0$  tel que pour  $n \geq n_0$

$$\left|x\sqrt{\frac{q}{np}}\right| \leq \frac{1}{2} \quad \text{et} \quad \left|x\sqrt{\frac{p}{nq}}\right| \leq \frac{1}{2}$$

Suffisons de maintenant que  $n \geq n_0$ .

On a le développement de Taylor  $\log(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} \left(\frac{1}{1+\theta}\right)^3$  47  
 où  $|\theta| < t$ . Donc

$$\log\left(1 + x\sqrt{\frac{q}{np}}\right) = x\sqrt{\frac{q}{np}} - \frac{1}{2}x^2\frac{q}{np} + R_n \quad \text{où} \quad R_n = \frac{1}{3}\left(\frac{1}{1+\theta}\right)^3 \left(x\sqrt{\frac{q}{np}}\right)^3, |\theta| < \frac{1}{2}$$

$$\log\left(1 - x\sqrt{\frac{p}{nq}}\right) = -x\sqrt{\frac{p}{nq}} - \frac{1}{2}x^2\frac{p}{nq} + R'_n \quad \text{où} \quad R'_n = \frac{1}{3}\left(\frac{1}{1+\theta}\right)^3 \left(x\sqrt{\frac{p}{nq}}\right)^3, |\theta| < \frac{1}{2}$$

Ainsi

$$\begin{aligned} \log B_n &= -(np + x\sqrt{npq}) \left[ x\sqrt{\frac{q}{np}} - \frac{1}{2}x^2\frac{q}{np} + R_n \right] \\ &\quad - (nq - x\sqrt{npq}) \left[ -x\sqrt{\frac{p}{nq}} - \frac{1}{2}x^2\frac{p}{nq} + R'_n \right] + \frac{x^2}{2} \\ &= -x\sqrt{npq} + \frac{1}{2}x^2q - x^2q + \frac{1}{2}x^3q\sqrt{\frac{q}{np}} - nR_n \\ &\quad + x\sqrt{npq} + \frac{1}{2}x^2p - x^2p - \frac{1}{2}x^3p\sqrt{\frac{p}{nq}} - jR'_n + \frac{x^2}{2} \\ &= -nR_n - jR'_n + \frac{1}{2}x^3 \left[ q\sqrt{\frac{q}{np}} - p\sqrt{\frac{p}{nq}} \right] \\ &\quad \rightarrow 0 \quad \text{car} \quad \frac{x^3}{n} = \left(\frac{x}{n^{1/6}}\right)^3 \rightarrow 0. \end{aligned}$$

De plus  $|nR_n| \leq n|R_n| = n \frac{1}{3} \left(\frac{1}{1+\theta}\right)^3 \left|x\sqrt{\frac{q}{np}}\right|^3$   
 $\leq \frac{8}{3} \frac{|x|^3}{\sqrt{n}} \left(\frac{q}{p}\right)^{3/2} \rightarrow 0$  comme ci-dessus

Il en va de même pour  $jR'_n$ . Donc  $\log B_n \rightarrow 0$ .

Remarque: La condition  $\frac{x_n, k_n}{n^{1/6}} = \frac{k_n - np}{n^{1/6} \sqrt{npq}} \rightarrow 0$  signifie que  $k_n$  n'est pas trop éloigné de  $np$ .

Théorème 2. Posons  $\sqrt{npq} b(k; n, p) = \varphi(x_n, k) + r_{n,k} \quad \forall k, 0 \leq k \leq n$ .  
 Alors on a :

$$\lim_{n \rightarrow \infty} \left( \max_{0 \leq k \leq n} |r_{n,k}| \right) = 0$$

Preuve: On a vu que la  $b(k; n, p)$  croissant pour  $k=0, 1, 2, \dots, m_n$  puis décroissant ( $m_n = [n\sqrt{pq}]$ ).

Choisissons  $i_n < m_n$  tel que  $x_{n,i_n} \rightarrow -\infty$  et  $\frac{x_{n,i_n}}{n^{1/6}} \rightarrow 0$

puis  $j_n > m_n$  tel que  $x_{n,j_n} \rightarrow +\infty$  et  $\frac{x_{n,j_n}}{n^{1/6}} \rightarrow 0$ .

(par exemple  $i_n = np - n^{3/2}, j_n = np + n^{3/2}$ )

Alors pour tout  $k, 0 \leq k \leq i_n$ , on a

$$\begin{aligned} |r_{n,k}| &= |\sqrt{\Delta x} b(k; n, p) - \varphi(x_{n,k})| \\ &\leq \sqrt{\Delta x} b(k; n, p) + \varphi(x_{n,k}) \\ &\leq \sqrt{\Delta x} b(i_n; n, p) + \varphi(x_{n,i_n}) \quad \text{car } \varphi \text{ est croissante sur } (-\infty, 0) \end{aligned}$$

donc  $\max_{0 \leq k \leq i_n} |r_{n,k}| \leq \sqrt{\Delta x} b(i_n; n, p) + \varphi(x_{n,i_n})$   
 $= \varphi(x_{n,i_n}) \left[ \frac{\sqrt{\Delta x} b(i_n; n, p)}{\varphi(x_{n,i_n})} + 1 \right]$   
 $\rightarrow 0$  car  $x_{n,i_n} \rightarrow -\infty$  et [...]  $\rightarrow 2$  (Thm.1).

Même raison pour  $\max_{j_n \leq k \leq n} |r_{n,k}| \rightarrow 0$ .

Enfin il existe  $l_n$  avec  $i_n \leq l_n \leq j_n$  et  $|r_{n,k}| \leq |r_{n,l_n}|$  pour tout  $k, i_n \leq k \leq j_n$ . Alors

$$\begin{aligned} \max_{i_n \leq k \leq j_n} |r_{n,k}| &= |r_{n,l_n}| = |\sqrt{\Delta x} b(l_n; n, p) - \varphi(x_{n,l_n})| \\ &= \varphi(x_{n,l_n}) \left| \frac{\sqrt{\Delta x} b(l_n; n, p)}{\varphi(x_{n,l_n})} - 1 \right| \rightarrow 0 \end{aligned}$$

$\leq \varphi(0)$                        $\rightarrow 0$  (Thm.1)

donc le Thm.2 est démontré.

Théorème 3. Pour tout  $n$ , soient  $j_n, l_n$  avec  $0 \leq j_n < l_n \leq n$ .

Posez  $a_n = \frac{j_n - np - 1/2}{\sqrt{\Delta x}}$ ,  $b_n = \frac{l_n - np + 1/2}{\sqrt{\Delta x}}$ .

Supposons qu'il existe  $c < \infty$  avec  $-c \leq a_n \leq b_n \leq c$ .

Alors on a :

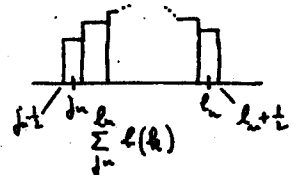
$$\sum_{k=j_n}^{l_n} b(k; n, p) - (\Phi(b_n) - \Phi(a_n)) \rightarrow 0 \quad \text{si } n \rightarrow \infty.$$

NB. Le théorème est vrai sans supposer que  $a_n$  et  $b_n$  restent bornés, mais en supposant que  $\frac{a_n}{n^{1/2}} \rightarrow 0, \frac{b_n}{n^{1/2}} \rightarrow 0$ .

Preuve: on a :

$$\sum_{k=j_n}^{l_n} b(k; n, p) = \underbrace{\frac{1}{\sqrt{\Delta x}} \sum_{k=j_n}^{l_n} \varphi(x_{n,k})}_{I_n} + \underbrace{\frac{1}{\sqrt{\Delta x}} \sum_{k=j_n}^{l_n} (\sqrt{\Delta x} b(k; n, p) - \varphi(x_{n,k}))}_{R_n}$$

i) Comme  $x_{n,k+1} - x_{n,k} = \frac{k+1-np}{\sqrt{\Delta x}} - \frac{k-np}{\sqrt{\Delta x}} = \frac{1}{\sqrt{\Delta x}}$  on voit que  $I_n$  est une somme de Riemann pour  $\int_{a_n}^{b_n} \varphi(y) dy$ .



$$\begin{aligned} \text{ii) } |R_n| &\leq \frac{1}{\sqrt{\Delta x}} \sum_{k=j_n}^{l_n} |r_{n,k}| \leq \frac{l_n - j_n + 1}{\sqrt{\Delta x}} \max_k |r_{n,k}| \\ &= \left( \frac{l_n - np + \frac{1}{2}}{\sqrt{\Delta x}} - \frac{j_n - np + \frac{1}{2}}{\sqrt{\Delta x}} \right) \max_k |r_{n,k}| \\ &\leq 2c \cdot \max_k |r_{n,k}| \rightarrow 0 \end{aligned}$$

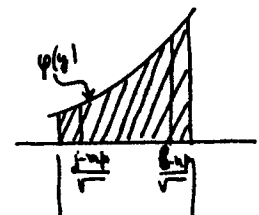
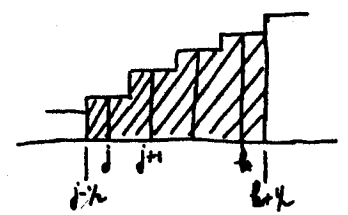
d'où le théorème.

Correction de continuité:

on utilisera l'approximation

$$Pr[j \leq S_n \leq l] \approx \Phi\left(\frac{l+1/2-np}{\sqrt{\Delta x}}\right) - \Phi\left(\frac{j-1/2-np}{\sqrt{\Delta x}}\right)$$

qui est meilleure que  $\Phi\left(\frac{l-np}{\sqrt{\Delta x}}\right) - \Phi\left(\frac{j-np}{\sqrt{\Delta x}}\right)$ , comme on le voit sur le dessin!



surface hachurée =  $\sum_{j} b(k; n, p)$   
 au dénominateur  $\rightarrow \Phi\left(\frac{l+1/2-np}{\sqrt{\Delta x}}\right) - \Phi\left(\frac{j-1/2-np}{\sqrt{\Delta x}}\right)$

Exemple: on jette 90 fois un dé; quelle est la probabilité

- d'obtenir 15 fois le 6 ?
- d'obtenir un nombre de 6 compris entre 13 et 17 ?

On a  $n=90$ ,  $p=\frac{1}{6}$ ,  $q=\frac{5}{6}$ ;  $np=15$ ;  $\sqrt{npq} = \sqrt{12,5} = 3,5355$

$$a) \quad \mathbb{P}(15; 90, \frac{1}{6}) \approx \frac{1}{\sqrt{npq}} \varphi(0) = \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} = 0,113$$

$$b) \quad \mathbb{P}[13 \leq S \leq 17] \approx \Phi\left(\frac{17 + \frac{1}{2} - 15}{3,5355}\right) - \Phi\left(\frac{13 - \frac{1}{2} - 15}{3,5355}\right) = 2\Phi(0,7071) - 1 = 0,587$$

Remarque 1. L'estimation de l'erreur commise en appliquant les théorèmes de De Moivre-Laplace est très difficile: cf Johnson-Kotz, "Discrete distributions", pages 62-63 (N 159 I).

D'après Kiezyzig l'approximation est bonne si  $np > 5$  lorsque  $p \leq \frac{1}{2}$  et  $nq > 5$  si  $p > \frac{1}{2}$ . Plus précis, Mosteller dit que si

$$3\sqrt{npq} < np < n - 3\sqrt{npq}$$

alors l'erreur dans l'application du théorème local est  $\leq \frac{11}{1000}$  et l'erreur dans l'application du théorème intégral ( $\int_{-\infty}^x$ ) est  $\leq \frac{25}{1000}$ .

Mais l'erreur relative peut être grande lorsque l'intervalle  $[a, b]$  n'est plus centré (cf Feller VIII 4 (d)).

Remarque 2. Les valeurs suivantes sont souvent utilisées:

$$\mathbb{P}(np - \sqrt{npq} \leq S_n \leq np + \sqrt{npq}) = 0,6826$$

$$\mathbb{P}(np - 2\sqrt{npq} \leq S_n \leq np + 2\sqrt{npq}) = 0,9545$$

$$\mathbb{P}(np - 3\sqrt{npq} \leq S_n \leq np + 3\sqrt{npq}) = 0,9973$$

$$\mathbb{P}(np - 1,96\sqrt{npq} \leq S_n \leq np + 1,96\sqrt{npq}) = 0,95$$

On n'a pas tenu compte ici de la correction de continuité.

Exemple (Feller) Deux  $C^1$  de transport font chacune partir un train au même moment du même endroit et vers la même destination, ceci chaque jour. On suppose aussi que chaque jour  $n$  voyageurs prennent l'un ou l'autre train au hasard ( $p = \frac{1}{2}$ ). Si un train a  $s$  places il y a une probabilité  $f(s)$  que des voyageurs ne trouvent pas de place dans ce train. En admettant qu'un schéma de Bernoulli est applicable, on a

$$f(s) = \mathbb{P}[S_n > s] \approx 1 - \Phi\left(\frac{s - \frac{n}{2}}{\sqrt{\frac{n}{4}}}\right) = 1 - \Phi\left(\frac{2s - n}{\sqrt{n}}\right)$$

Si une  $C^1$  desire offrir assez de places 99 fois sur 100, de combien de places doit-elle disposer dans son train ?

$$\text{On a } f(s) = 0,01 \Leftrightarrow \Phi\left(\frac{2s - n}{\sqrt{n}}\right) = 0,99 \Leftrightarrow \frac{2s - n}{\sqrt{n}} = 2,326 \Leftrightarrow$$

$$s = \frac{1}{2}(n + 2,326\sqrt{n})$$

Ex. si  $n = 100$   $s = 62$  i.e. avec 124 places } chaque voyageur  
si  $n = 1000$   $s = 537$  — 1074 — } trouvera une place  
99 fois sur 100 !!

Application: Fréquences relatives et probabilités.

Une des conceptions intuitives de la notion de probabilité est basée sur l'hypothèse suivante: "si on répète un grand nombre  $n$  de fois une expérience aléatoire et si l'événement aléatoire se produit  $k$  fois, alors la probabilité de  $A$  doit être voisine de  $\frac{k}{n}$ ".

On a modélisé cette situation par un schéma de Bernoulli. Si ce modèle est adéquat on doit avoir un théorème qui précise l'hypothèse ci-dessus.

Si  $p = P_0(A)$ , si  $S_n$  est le nombre de réalisations de  $A$  en  $n$  expériences,  $\frac{S_n}{n}$  représente la fréquence relative d'apparition de  $A$ , si  $\varepsilon > 0$ , on a

$$\left| \frac{S_n}{n} - p \right| < \varepsilon \Leftrightarrow p - \varepsilon < \frac{S_n}{n} < p + \varepsilon \Leftrightarrow -\frac{\sqrt{n}\varepsilon}{\sqrt{pq}} < \frac{S_n - np}{\sqrt{npq}} < \frac{\sqrt{n}\varepsilon}{\sqrt{pq}}$$

d'où

$$(*) \quad \mathbb{P}\left[\left|\frac{S_n}{n} - p\right| < \varepsilon\right] \approx 2\Phi\left(\frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right) - 1$$

a) On a déduit tout d'abord que, quel que soit  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \Pr\left[\left|\frac{S_n}{n} - p\right| < \varepsilon\right] = 1 \quad \text{ou} \quad \lim_{n \rightarrow \infty} \Pr\left[\left|\frac{S_n}{n} - p\right| > \varepsilon\right] = 0$$

C'est la loi faible des grands nombres; elle dit que si  $n$  est grand, une différence importante entre  $\frac{S_n}{n}$  et  $p$  est très peu probable. L'intérêt de cette loi est plus théorique que pratique: elle renforce et précise notre conviction que les fréquences relatives se stabilisent et donnent la probabilité.

NB. La loi faible des grands nombres dit:

$$\forall \varepsilon > 0, \forall \delta > 0, \exists m_0 = m_0(\varepsilon, \delta) \text{ t.q. } \forall n \geq m_0 \quad \Pr\left[\left|\frac{S_n}{n} - p\right| < \varepsilon\right] > 1 - \delta$$

$$\text{i.e.} \quad \Pr\left[\left|\frac{S_{200}}{200} - p\right| < \varepsilon\right] > 1 - \delta$$

$$\Pr\left[\left|\frac{S_{400}}{400} - p\right| < \varepsilon\right] > 1 - \delta$$

etc.

mais elle ne dit pas que  $\Pr\left[\left|\frac{S_n}{n} - p\right| < \varepsilon, \forall n \geq m_0\right] > 1 - \delta$

(d'ailleurs cette affirmation n'a pas de sens dans notre modèle qui ne peut considérer qu'un nombre fini de répétitions de l'expérience de base). La loi forte des grands nombres améliorera cela!

b) L'expression (\*) permet de déterminer expérimentalement des probabilités, en prenant comme valeur approchée de  $p$  la fréquence relative observée  $\frac{S_n}{n}$  en  $n$  expériences. Se pose alors la question suivante: combien faut-il faire d'expériences pour que  $\Pr\left[\left|\frac{S_n}{n} - p\right| < \varepsilon\right] \geq \gamma$ ? En appliquant l'approximation (\*) on a

$$n \geq \frac{pq \left(\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\right)^2}{\varepsilon^2}$$

Comme  $pq = p(1-p) \leq \frac{1}{4}$  il suffit de prendre  $n \geq \frac{\left(\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\right)^2}{4\varepsilon^2}$ .

Exemple: on veut déterminer la probabilité qu'une personne d'une certaine population soit un fumeur, en étudiant un échantillon extrait de la population avec remplacement. Quelle doit être la taille de l'échantillon si on admet

une erreur sur la probabilité de  $\varepsilon = 1\%$  et un seuil de confiance  $\gamma = 0,95$ ? On a  $\frac{1+\gamma}{2} = 0,975$ ,  $\Phi^{-1}(0,975) = 1,96$ , donc

$$n \geq \frac{(1,96)^2}{4(0,01)^2} = 9604$$

Pour  $\varepsilon = 0,02$  on obtient  $n \geq 2401$ .

5.5. Théorème de Poisson.

But: Obtenir une approximation de  $b(k; n, p)$  lorsque  $n$  est grand,  $p$  petit et  $np$  pas trop grand.

Théorème. Soit  $p_1, p_2, p_3, \dots$  une suite de nombres telle que  $0 < p_n < 1$ ,  $\forall n$ , et  $np_n \rightarrow \lambda$  si  $n \rightarrow \infty$ . Alors:

$$\lim_{n \rightarrow \infty} b(k; n, p_n) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (\forall k=0,1,2,\dots)$$

Preuve:

$$b(k; n, p_n) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad \text{où } \lambda_n = np_n.$$

$$= \frac{1}{k!} \cdot \frac{n(n-1)\dots(n-k+1)}{n^k} \cdot \lambda^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

$$\underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\rightarrow 1} \cdot \underbrace{\lambda^k}_{\rightarrow \lambda^k} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow 1}$$

$$\rightarrow \frac{1}{k!} \lambda^k e^{-\lambda} \quad \text{si } n \rightarrow \infty.$$

Pour  $p(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$  ( $\forall k=0,1,2,\dots$ ); on a  $\sum_{k=0}^{\infty} p(k; \lambda) = 1$  donc la  $p(k; \lambda)$  définissant une loi de probabilité sur  $\mathbb{N}$ ; c'est la loi de Poisson. Cette loi apparaît très souvent pour donner les probabilités d'apparition d'un événement rare.

En pratique, si  $n$  est grand,  $p$  petit,  $\lambda=np$  pas trop grand, on utilisera l'approximation  $b(k; n, p) \approx p(k; \lambda)$ .

Exemple: Quelle est la probabilité que parmi 500 personnes prises au hasard,  $k$  aient leur anniversaire le jour de Noël? C'est  $b(k; 500, \frac{1}{365}) \approx p(k; \frac{500}{365})$ ; on a (cf Feller):

$k$	0	1	2	3	4
$b_k$	0,2537	0,3484	0,2388	0,1089	0,0372
$p_k$	0,2541	0,3487	0,2385	0,1089	0,0373
	0,16	0,37	0,32	0,11	0,03

si  $\lambda$  est grand on utilise l'approx. normale ( $\lambda > 4$ )

§6. Variables aléatoires.

6.1. Définition

Soit  $\mathcal{E}$  une expérience aléatoire. Il arrive souvent qu'on soit d'avantage intéressé à une caractéristique numérique (\*) du résultat de l'expérience qu'à sa description complète.

Exemple 1. Répétitions indépendantes d'une même expérience; à chaque expérience on s'intéresse à l'apparition d'un événement A (succès). On peut s'intéresser au nombre de succès obtenus ou au rang du 1<sup>er</sup> succès, plutôt que de savoir exactement quand ont eu lieu les succès et les échecs.

Exemple 2. On choisit un point au hasard dans un disque de rayon 1. On peut s'intéresser à la distance du point choisi au centre du disque.

Si l'expérience  $\mathcal{E}$  est modélisée par l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ , une telle caractéristique numérique apparaît comme une fonction  $X: \Omega \rightarrow \mathbb{R}$ .

Exemple 1.  $\Omega = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i = 1 \text{ si A se réalise à la } i^{\text{ème}} \text{ exp.}, \omega_i = 0 \text{ sinon}\}$

$$S_n: \Omega \rightarrow \mathbb{R} : S_n(\omega) = \text{nb. de "1" dans } \omega$$

$$R_n: \Omega \rightarrow \mathbb{R} : R_n(\omega) = \text{rang du 1}^{\text{er}} \text{ "1" dans } \omega$$

Exemple 2.  $\Omega = \{\omega = (x, y) \mid x^2 + y^2 \leq 1\}$ ;  $\mathcal{A} = \mathcal{B}$ ;  $P$  uniforme

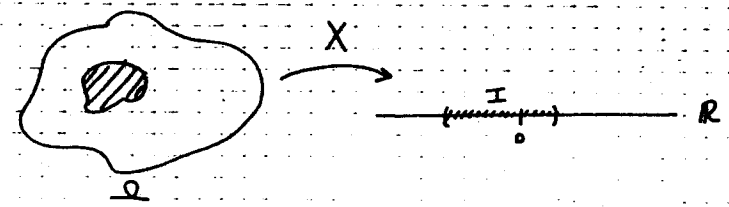
$$D: \Omega \rightarrow \mathbb{R} : D(\omega) = \sqrt{x^2 + y^2}$$

En fait seuls vont nous intéresser les fonctions  $X: \Omega \rightarrow \mathbb{R}$  pour lesquelles la affirmation du genre "il y a une probabilité de 0,7 pour que  $X$  prenne ses valeurs entre -1 et  $\frac{1}{2}$ " ont un sens. D'où la définition:

(\*) cf. les "observables" en physique.



Définition. Etant donné une expérience  $E$  modélisée par l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ , une variable aléatoire (v.a.) associée à  $E$  est une fonction  $X: \Omega \rightarrow \mathbb{R}$  telle que, pour tout intervalle  $I \subset \mathbb{R}$ , l'événement  $X^{-1}(I) = \{\omega \in \Omega \mid X(\omega) \in I\}$  appartienne à  $\mathcal{A}$  (pour qu'on puisse parler de sa probabilité).



Notations: On notera  $[X \in I]$  l'événement  $X^{-1}(I)$  et plus particulièrement  $[a < X < b]$ ,  $[X < a]$ , ... lorsque  $I = ]a, b[$ ,  $I = ]-\infty, a[$ , ...  
On notera aussi  $P[X \in I]$  plutôt que  $P([X \in I])$ .

Remarque: Si  $\mathcal{A} = \mathcal{P}(\Omega)$  - ce qui est le cas dans les espaces probabilisés discrets - toute fonction  $X: \Omega \rightarrow \mathbb{R}$  est une variable aléatoire. Ainsi dans l'exemple 1  $S_n$  et  $R_n$  sont des v.a. Dans l'exemple 2,  $\Delta$  est une v.a. car si  $a < t$

$$[a < \Delta < t] = [\omega \in \Omega \mid a < \Delta(\omega) < t] =$$

et ces couronnes sont des boréliens de  $\Omega$ .  
Dans cet exemple - et dans du même type - il existe des fonctions  $\Omega \rightarrow \mathbb{R}$  qui ne sont pas des v.a. Mais il est difficile d'en exhiber et on a très peu de chances d'en rencontrer une... [cf. série 13]

Exemple 1. 1) L'indicateur d'un événement  $A \in \mathcal{A}$  est la v.a.  $I_A$  définie par  $I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$ .  
Alors  $[I_A = 1] = A$ ;  $[I_A = 0] = \bar{A}$ .  
 $I_A$  indique donc si  $A$  s'est réalisé ou non.

2) Si  $A_1, \dots, A_n \in \mathcal{A}$ ,  $X = \sum_{j=1}^n I_{A_j}$  est une v.a.; elle compte combien d'éléments  $A_1, \dots, A_n$  se sont réalisés.

Par exemple, dans un schéma de Bernoulli, si  $A_j^{(i)}$  désigne (comme à la page 42) l'événement "A se réalise à la j<sup>ème</sup> expérience" on a  $S_n = \sum_{j=1}^n I_{A_j^{(1)}}$ .

3) Si  $\Omega = A_1 + \dots + A_n$  est une partition de  $\Omega$ , alors  $X = \sum_{j=1}^n j \cdot I_{A_j}$  indique lequel des  $A_j$  s'est réalisé.

6.2. Loi de probabilité d'une v.a.

Si  $X: \Omega \rightarrow \mathbb{R}$  est une v.a. on peut considérer  $P[X \in I]$  pour tous les intervalles  $I \subset \mathbb{R}$ . Peut-on considérer  $P[X \in B]$  pour d'autres parties  $B \subset \mathbb{R}$ ? La réponse est oui; on procède comme suit:

- i) soit  $\mathcal{C} = \{E \subset \mathbb{R}; [X \in E] \in \mathcal{A}\}$ ; on montre immédiatement que  $\mathcal{C}$  est une tribu de parties de  $\mathbb{R}$ ;
- ii) comme  $X$  est une v.a.  $\mathcal{C}$  contient tous les intervalles; donc  $\mathcal{C}$  contient la tribu borélienne  $\mathcal{B}$  sur  $\mathbb{R}$  (cf. page 27).

Ainsi pour tout borélien  $B \subset \mathbb{R}$ ,  $[X \in B] \in \mathcal{A}$  et on peut considérer  $P[X \in B]$ ; on posera

$$P_X(B) = P[X \in B]$$

Lemme:  $P_X$  est une probabilité sur  $(\mathbb{R}, \mathcal{B})$ .

La preuve est immédiate (cf. série 12).

Définition: la probabilité  $P_X$  sur  $(\mathbb{R}, \mathcal{B})$  est appelée la loi de probabilité de  $X$ .

Exemple 1. 1) la loi de probabilité  $P_{S_n}$  de la v.a.  $S_n$  est la loi binomiale; 2) la loi de probabilité  $P_{R_n}$  de la v.a.  $R_n$  est la loi géométrique.

Exemple 2.  $P_\Delta(I)$  est la surface de la couronne (de son intersection avec  $\Omega$ ).

Remarque: Souvent dans les applications ce n'est pas l'espace probabilisé  $(\Omega, \mathcal{G}, P)$  qui nous intéresse, mais des v.a.  $X, Y, Z, \dots$  et les lois de probabilité  $P_X, P_Y, P_Z, \dots$  sur  $\mathbb{R}$ . La connaissance de l'expérience aléatoire qu'on considère permet souvent de faire directement des hypothèses sur ces lois de probabilité.

Exemple: On observe une substance radioactive pendant  $t$  secondes. On suppose qu'elle émet en moyenne  $\beta$  particules par seconde ( $\beta =$  intensité de la radiation). Le nombre de particules émises pendant  $t$  secondes est une v.a.  $X$ . Si  $t$  est petit par rapport à la demi-vie de la substance, il est admis que  $X$  a une loi de probabilité de Poisson de paramètre  $\beta t$ :

$$P[X=k] = \frac{(\beta t)^k}{k!} e^{-\beta t} \quad \text{pour } k=0,1,2,\dots$$

Dans ce cas on n'explique pas le modèle de l'expérience.

Définition 1: Une v.a.  $X$  est dite discrète si  $X(\Omega)$  est fini ou dénombrable. Alors on écrit  $X(\Omega) = \{x_1, x_2, \dots\}$ .

La loi  $P_X$  d'une v.a. discrète est déterminée par les valeurs

$$P_X(\{x_i\}) = P[X=x_i] \quad (i=1,2,\dots)$$

si  $B \subset \mathbb{R}$  on a  $P_X(B) = P[X \in B] = \sum_{x_i \in B} P[X=x_i]$ .

Toutes les v.a. définies sur des espaces  $\Omega$  finis ou dénombrables sont discrètes.

Définition 2: Une v.a.  $X$  est dite absolument continue si  $X(\Omega)$  est un intervalle (ou une réunion d'intervalles...) et s'il existe une fonction  $f: \mathbb{R} \rightarrow \mathbb{R}$ ,  $f \geq 0$ ,  $\int_{\mathbb{R}} f(x) dx = 1$ , avec

$$P_X(B) = P[X \in B] = \int_B f(x) dx$$

[NB: L'intégral de Lebesgue est nécessaire si on veut considérer tous les boréliens!]

On dit que  $f$  est la densité de la loi  $P_X$  (ou de la v.a.  $X$ ). (notée  $f_X$  si c'est nécessaire).

Remarque 1) Dans ce cas on a  $P[X=a] = 0$  pour tout  $a \in \mathbb{R}$ .

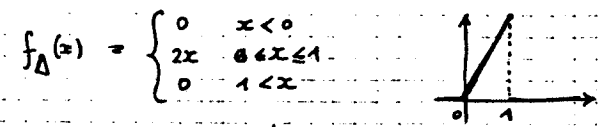
2) Si  $f$  est continue au point  $a$  on a  $\frac{1}{2h} P[a-h < X < a+h] = \frac{1}{2h} \int_{a-h}^{a+h} f(x) dx \rightarrow f(a)$  si  $h \rightarrow 0$

donc si  $h$  est assez petit on a  $P[a-h < X < a+h] \approx 2h \cdot f(a)$

On peut dire aussi  $P[a \leq X < a+da] \approx f(a) da$

Exemple: La v.a.  $\Delta$  de l'exemple 2 ci-dessus est absolument continue. On a

$$P_{\Delta}([a,b]) = P[a \leq \Delta \leq b] = \begin{cases} 0 & \text{si } a < b < 0 \\ b^2 & \text{si } a < 0 < b < 1 \\ b^2 - a^2 & \text{si } 0 < a < b < 1 \\ 1 - a^2 & \text{si } 0 < a < 1 < b \\ 1 & \text{si } a < 0, 1 < b \end{cases}$$



Il existe aussi des v.a. de type mixte; une v.a.  $X$  est de type mixte s'il existe un ensemble fini ou dénombrable  $\{x_1, x_2, \dots\} \subset \mathbb{R}$  avec  $P[X=x_i] > 0$  ( $i=1,2,3,\dots$ ) et un intervalle (ou un borélien) sur lequel  $X$  admet une densité  $f$ . Alors pour tout borélien  $B \subset \mathbb{R}$  on a

$$P_X(B) = P[X \in B] = \sum_{x_i \in B} P[X=x_i] + \int_B f(x) dx$$

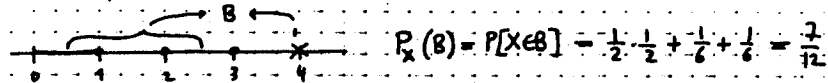
Exemple (un peu académique...) on jette une pièce (équilibrée); si c'est pile, on choisit un nombre au hasard dans  $\{1,2,3\}$ ; si c'est face, on choisit un nombre au hasard dans  $[0,1]$ . Soit  $X$  le point choisi. On a

$$P[X=1] = P[X=2] = P[X=3] = \frac{1}{6}$$

$$P[a < X < b] = \frac{1}{2}(b-a) \quad \text{si } 0 < a < b < 1$$

↑  
strict!

Par exemple si  $B = [\frac{1}{2}, \frac{5}{6}] \cup \{4\}$  on a



### 6.3. Fonction de répartition d'une variable aléatoire

Soit  $X$  une v.a. sur  $(\Omega, \mathcal{A}, P)$ . Pour tout  $x \in \mathbb{R}$  on pose

$$F(x) = P_X(-\infty, x[) = P[X < x]$$

Définition.  $F$  est la fonction de répartition de  $X$  (notée  $F_X$  s'il le faut!).

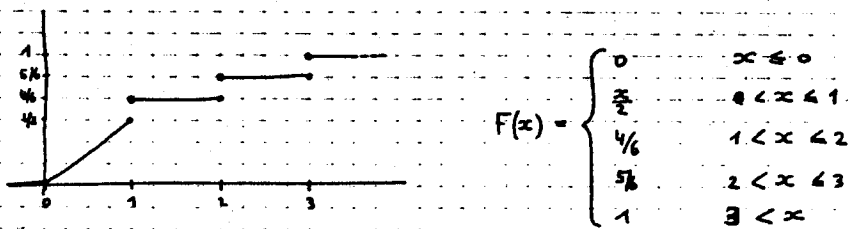
Remarques: 1) Si  $X$  est discrète on a  $F(x) = \sum_{x_i < x} P[X = x_i]$

2) Si  $X$  est absolument continue on a  $F(x) = \int_{-\infty}^x f(t) dt$

(ce qui peut s'écrire avec la Rem 2) page 59 :  $F(x) = \int_{-\infty}^x P[t \leq X < t+dt]$ .)

Exemples. 1) Dans les cas discret et continu on a déjà vu les fonctions de rép. des v.a. binomiale et normale.

2) Déterminons la fonction de répartition de l'exemple de la v.a. mixte de la page précédente. On obtient



$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{6} & 0 < x \leq 1 \\ \frac{1}{6} & 1 < x \leq 2 \\ \frac{5}{6} & 2 < x \leq 3 \\ 1 & 3 < x \end{cases}$$

Proposition. Soit  $F$  la fonction de répartition d'une v.a.  $X$ .

- 1)  $0 \leq F(x) \leq 1$  ( $\forall x \in \mathbb{R}$ )
- 2)  $x \leq y \Rightarrow F(x) \leq F(y)$ .
- 3)  $F$  est continue à gauche (en tout point).
- 4)  $F(a+) = P[X \leq a]$  ( $\forall a \in \mathbb{R}$ )
- 5)  $F(a+) - F(a) = P[X = a]$  ( $\forall a \in \mathbb{R}$ )
- 6)  $\lim_{x \rightarrow -\infty} F(x) = 0$  et  $\lim_{x \rightarrow \infty} F(x) = 1$ .

Preuve. 1) Clair car  $F(x)$  est une probabilité.

2)  $x \leq y \Rightarrow [X < x] \subset [X < y] \Rightarrow P[X < x] \leq P[X < y]$ .

3)  $F$  étant monotone croissante possède en chaque point une limite à gauche et une limite à droite (à gauche est une suite croissante vers  $a$ ,  $(F(x_n))$  est une suite croissante majorée par  $F(a)$ , donc  $\lim_{n \rightarrow \infty} F(x_n)$  existe et vaut  $F(a-) \leq F(a)$ )

Pour démontrer que  $F$  est continue à gauche il suffit donc de démontrer que  $F(a-) = F(a)$  pour tout  $a$ .

Posez  $A_n = [X > a - \frac{1}{n}] = \overline{[X < a - \frac{1}{n}]}$ ; on a  $A_1 \supset A_2 \supset A_3 \dots$

et  $A = \bigcap_{n=1}^{\infty} A_n = [X > a] = \overline{[X \leq a]}$

On a (Prop. 2 § 3.3 page 31) :  $P(A) = \lim_{n \rightarrow \infty} P(A_n)$ .

Mais  $P(A) = 1 - P[X \leq a] = 1 - F(a)$  et  $P(A_n) = 1 - F(a - \frac{1}{n})$ , donc

$\lim_{n \rightarrow \infty} F(a - \frac{1}{n}) = F(a)$  i.e.  $F(a-) = F(a)$ .

4) Posez  $B_n = [X < a + \frac{1}{n}]$ ; on a  $B_1 \supset B_2 \supset B_3 \dots$  et  $B = \bigcup_{n=1}^{\infty} B_n = [X < a+]$ .

Comme à-demos on en déduit que

$$P[X < a+] = \lim_{n \rightarrow \infty} P[X < a + \frac{1}{n}] = \lim_{n \rightarrow \infty} F(a + \frac{1}{n}) = F(a+)$$

5)  $F(a+) = P[X < a+] = P[X = a] + P[X < a] = P[X = a] + F(a)$

6) Comme à 3) et 4) avec  $\bigcap_{n=1}^{\infty} [X < -n] = \emptyset$  et  $\bigcup_{n=1}^{\infty} [X < n] = \Omega$ .

Remarques. 1) Si  $a \leq b$  on a

$$P[a \leq X < b] = F(b) - F(a)$$

$$P[X \leq a] = F(a+)$$

$$P[a < X \leq b] = F(b+) - F(a)$$

$$P[X > a] = 1 - F(a)$$

$$P[a < X < b] = F(b) - F(a+)$$

$$P[X > a] = 1 - F(a+)$$

$$P[a < X \leq b] = F(b+) - F(a+)$$

$$P[X = a] = F(a+) - F(a)$$

2) Si  $X$  est absolument continue,  $F$  est continue. Mais la réciproque est fautive! (cf. théorie de l'intégration)  
En les points  $x$  où  $f$  est continue on  $F'(x) = f(x)$ .

3) Deux v.a.  $X$  et  $Y$  ont la même loi de probabilité (i.e.  $P_X = P_Y$ ) ssi elles ont les mêmes fct. de répartition (i.e.  $F_X = F_Y$ ).

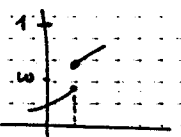
Justification Si  $P_X = P_Y$  il est clair que  $F_X = F_Y$ . Inversement si  $F_X = F_Y$  on a  $P_X(I) = P_Y(I)$  pour tout intervalle  $I \subset \mathbb{R}$  (cf. Remarque 2); on peut "prolonger" cette égalité à tous les boréliens, d'où  $P_X = P_Y$ . (cf. le résultat adms page 28)

Proposition Soit  $F: \mathbb{R} \rightarrow \mathbb{R}$  une fonction telle que

- a)  $F$  est croissante;
- b)  $F$  est continue à gauche en tout point;
- c)  $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$ .

Il existe une v.a.  $X$  telle que  $F_X = F$ .

Preuve: On prend  $\Omega = ]0, 1[$ ,  $\mathcal{A} = \mathcal{B}_\Omega$  les boréliens de  $\Omega$  et  $P$  la probabilité uniforme (i.e. la longueur).



Pour tout  $\omega \in \Omega$ , posons  $X(\omega) = \sup \{x \in \mathbb{R} \mid F(x) \leq \omega\}$

Notons que  $\{x \in \mathbb{R} \mid F(x) \leq \omega\} \neq \emptyset$  à cause de  $F(-\infty) = 0$ , et que  $\{x \in \mathbb{R} \mid F(x) \leq \omega\}$  est majoré car  $F(+\infty) = 1$ , donc que  $X(\omega)$  est bien défini.

D'autre part la borne supérieure est atteinte i.e.

$$X(\omega) = \max \{x \in \mathbb{R} \mid F(x) \leq \omega\},$$

à cause de la continuité à gauche de  $F$ . Ainsi  $F(X(\omega)) \leq \omega$ . A cause de sa et de la croissance de  $F$  on a si  $a < b$

$$a \leq X(\omega) < b \iff a < \max \{x \in \mathbb{R} \mid F(x) \leq \omega\} < b$$

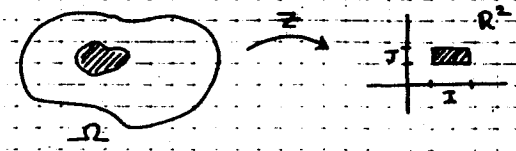
$$\iff F(a) \leq \omega < F(b)$$

donc  $[a \leq X < b] = [F(a), F(b)[ \subset \Omega$ ; ainsi  $X$  est une v.a.

Enfin  $[X < t] = ]0, F(t)[$  donc  $F_X(t) = P(]0, F(t)[) = F(t)$  ( $\forall t$ )  
i.e.  $F_X = F$ .

6.4. Variable aléatoire vectorielle.

Etant donné deux variables aléatoires  $X$  et  $Y$  sur  $(\Omega, \mathcal{A}, P)$  on peut considérer le couple  $Z = (X, Y)$  comme une v.a. à valeurs dans  $\mathbb{R}^2$ :



en effet si  $I$  et  $J$  sont des intervalles, on voit que

$$[Z \in I \times J] = \{\omega \in \Omega \mid X(\omega) \in I, Y(\omega) \in J\} = [X \in I] \cap [Y \in J]$$

est un élément de  $\mathcal{A}$ . De même si  $B \in \mathcal{B}_{\mathbb{R}^2}$ .

La loi de probabilité sur  $\mathbb{R}^2$   $P_{(X,Y)} = P_Z$  définie par  $P_{(X,Y)}(B) = P[(X,Y) \in B]$  est appelée loi de probabilité du couple  $(X, Y)$  ou loi de probabilité conjointe de  $X$  et  $Y$ .

Car discret: Si  $X$  et  $Y$  sont des v.a. discrètes avec  $X(\Omega) = \{x_1, x_2, \dots\}$  et  $Y(\Omega) = \{y_1, y_2, \dots\}$  la loi  $P_{(X,Y)}$  est entièrement définie par le nombre

$$P[X=x_i, Y=y_j] \quad i=1,2,\dots; j=1,2,\dots$$

Exemple: Considérons une urne avec  $n_1$  boules blanches,  $n_2$  boules noires et  $n_3$  boules rouges. On extrait de l'urne une sous-population de  $k$  boules. On considère les variables aléatoires

- $X$  = Nb. de boules blanches extraites
- $Y$  = Nb. de boules noires extraites

on a:

$$P[X=i] = \frac{\binom{n_1}{i} \binom{n_2+n_3}{k-i}}{\binom{n_1+n_2+n_3}{k}}, \quad P[Y=j] = \frac{\binom{n_2}{j} \binom{n_1+n_3}{k-j}}{\binom{n_1+n_2+n_3}{k}}$$

et

$$P[X=i, Y=j] = \frac{\binom{n_1}{i} \binom{n_2}{j} \binom{n_3}{k-i-j}}{\binom{n_1+n_2+n_3}{k}}$$

Cas absolument continu: C'est le cas où il existe une fonction  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  telle que  $f(x,y) \geq 0$  et  $\int_{\mathbb{R}^2} f(x,y) dx dy = 1$ , et telle que

$$P_{(X,Y)}(B) = P[(X,Y) \in B] = \int_B f(x,y) dx dy$$

On dit que  $f$  est la densité du couple  $(X,Y)$  - on a la loi  $P_{(X,Y)}$

Exemple: Loi normale à 2 dimensions de paramètre  $r$

C'est la loi de probabilité du couple  $(X,Y)$  définie par la densité

$$f(x,y) = C_r e^{-\frac{1}{2}Q(x,y)} \quad \forall (x,y) \in \mathbb{R}^2$$

où  $C_r = \frac{1}{2\pi\sqrt{1-r^2}}$  et  $Q(x,y) = \frac{x^2 - 2rxy + y^2}{1-r^2}$ ,  $-1 < r < 1$ .

Il faut vérifier que  $\int \int f(x,y) dx dy = 1$ . Or:

$$\int \int = \int_{-\infty}^{+\infty} dx \left[ \int_{-\infty}^{+\infty} C_r e^{-\frac{1}{2}Q(x,y)} dy \right] = \int_{-\infty}^{+\infty} dx \left[ \int_{-\infty}^{+\infty} C_r e^{-\frac{1}{2}(x^2+z^2)} \frac{1}{\sqrt{1-r^2}} dz \right]$$

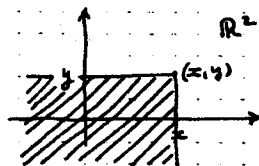
en posant  $z = \frac{y-rx}{\sqrt{1-r^2}}$   $dz = \frac{dy}{\sqrt{1-r^2}}$  et  $Q(x,y) = x^2 + z^2$

Donc: 
$$\int \int = \int_{-\infty}^{+\infty} dx \frac{1}{2\pi} e^{-x^2/2} \left[ \int_{-\infty}^{+\infty} e^{-z^2/2} dz \right] = \int_{-\infty}^{+\infty} dx \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = 1$$

Fonction de répartition du couple  $(X,Y)$ : elle est définie par

$$F_{(X,Y)}(x,y) = P[X < x, Y < y] = P_{(X,Y)}((-\infty, x] \times (-\infty, y])$$

Propriétés: en chaque variable  $F_{(X,Y)}$  est croissante, continue à gauche, et  $F_{(X,Y)}(-\infty, y) = F(x, -\infty) = 0$ ;  $F(+\infty, +\infty) = 1$  (ie  $\lim_{x \rightarrow +\infty} \dots$  !)



on a: 
$$P[a \leq X < b, c \leq Y < d] = F_{(X,Y)}(b,d) - F_{(X,Y)}(a,d) - F_{(X,Y)}(b,c) + F_{(X,Y)}(a,c)$$

etc.

Remarques: 1) Les fonctions de répartition sont beaucoup moins utiles que dans le cas à 1 dimension car  
- l'emploi en est plus difficile,  
- beaucoup de bornes intéressants ne sont pas des rectangles!

2) Dans le cas où  $(X,Y)$  admet une densité  $f(x,y)$  on a

$$F(x,y) = \int_{-\infty}^x \int_{-\infty}^y f(s,t) ds dt$$

et si  $f$  est continue au point  $(a,b)$ :

$$f(a,b) = \frac{\partial^2 F}{\partial x \partial y}(a,b)$$

si  $F$  est deux fois différentiable alors  $\frac{\partial^2 F}{\partial x \partial y}$  est une densité.

Probabilités marginales

Etant donné deux v.a.  $X$  et  $Y$  et le couple  $(X,Y)$  quelles relations existe-t-il entre  $P_X, P_Y$  et  $P_{(X,Y)}$ ?

Il est clair qu'on a:

$$P_X(I) = P[X \in I] = P[X \in I, Y < +\infty] = P_{(X,Y)}(I \times \mathbb{R})$$

et de même:  $P_Y(J) = P_{(X,Y)}(\mathbb{R} \times J)$

A l'aide des fonctions de répartition, cela s'exprime par

$$F_X(x) = F_{(X,Y)}(x, +\infty) = \lim_{y \rightarrow +\infty} F_{(X,Y)}(x,y)$$

$$F_Y(y) = F_{(X,Y)}(+\infty, y) = \lim_{x \rightarrow +\infty} F_{(X,Y)}(x,y)$$

On dit que  $F_X$  et  $F_Y$  sont les fonctions de répartition marginales de  $X$  et  $Y$ . De plus on a

- cas discret:  $P[X=x_i] = \sum_{j \geq 1} P[X=x_i, Y=y_j]$ ;  $P[Y=y_j] = \dots$

- cas absolument continu: si  $f(x,y)$  est la densité de  $(X,Y)$  alors

$$g(x) = \int_{-\infty}^{+\infty} f(x,y) dy \quad \text{est une densité pour } X$$

et 
$$h(y) = \int_{-\infty}^{+\infty} f(x,y) dx \quad \text{est une densité pour } Y$$

Ce sont les densités marginales de X et Y.

Exemple 1 : cf. exemple page 63.

Exemple 2 : Supposons que la loi de  $(X, Y)$  soit la loi uniforme sur le disque unité i.e. admette la densité

$$f(x, y) = \begin{cases} \frac{1}{\pi} & \text{si } x^2 + y^2 \leq 1 \\ 0 & \text{si } x^2 + y^2 > 1 \end{cases}$$

Alors la densité marginale de  $X$  est

$$g(x) = \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} f(x, y) dy = \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi} \sqrt{1-x^2}$$

pour  $-1 \leq x \leq 1$  et  $g(x) = 0$  sinon. Idem pour  $Y$ .

Exemple 3 : Loi normale à 2 dimensions de paramètre (cf. page 64). Le calcul fait page 64 montre que la densité marginale de  $X$  est

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

Idem pour  $Y$ . Donc les lois marginales de  $X$  et  $Y$  sont des lois normales (à 1 dim.).

Variables aléatoires indépendantes :

Les v.a.  $X$  et  $Y$  sont dites indépendantes si les événements  $[X \in I]$  et  $[Y \in J]$  sont indépendants, quel que soient les intervalles  $I$  et  $J$  i.e. si on a

$$P[X \in I, Y \in J] = P[X \in I] P[Y \in J].$$

Cela revient aussi à dire que la loi  $P_{(X, Y)}$  du couple est le produit des lois  $P_X$  et  $P_Y$  (i.e.  $P_{(X, Y)} = P_X \otimes P_Y$ ) soit

$$P_{(X, Y)}(I \times J) = P_X(I) P_Y(J).$$

Proposition : Pour que  $X$  et  $Y$  soient indépendantes, il faut et il suffit que :

$$F_{(X, Y)}(x, y) = F_X(x) F_Y(y) \quad \forall (x, y) \in \mathbb{R}^2.$$

Preuve : ( $\Rightarrow$ ) est clair.

( $\Leftarrow$ ) : Prenons  $I = [a, b[$  et  $J = [c, d[$ . On a

$$\begin{aligned} P[a \leq X < b, c \leq Y < d] &= F_{(X, Y)}(b, d) - F_{(X, Y)}(a, d) - F_{(X, Y)}(a, c) + F_{(X, Y)}(a, c) \\ &= F_Y(b) F_Y(d) - F_X(a) F_Y(d) - F_X(a) F_Y(c) + F_X(a) F_Y(c) \\ &= [F_Y(b) - F_Y(c)] [F_Y(d) - F_Y(c)] = P[a \leq X < b] P[c \leq Y < d] \end{aligned}$$

et de même pour les autres types d'intervalles.

Cas discret :  $X$  et  $Y$  sont indépendantes ssi

$$P[X = x_i, Y = y_j] = P[X = x_i] P[Y = y_j] \quad \forall i, j$$

Cas absolument continu : si  $X$  est absolument continue de densité  $g$  et  $Y$  absolument continue de densité  $h$  alors

$X$  et  $Y$  sont indépendantes

$\Leftrightarrow (X, Y)$  admet pour densité la fct.  $f$  où  $f(x, y) = g(x)h(y)$

C'est immédiat à partir de l'égalité :

$$\int_I g(x) dx \int_J h(y) dy = \iint_{I \times J} g(x)h(y) dx dy$$

Exemple 1 : Dans l'exemple de la page 63  $X$  et  $Y$  ne sont pas indépendantes. Voir aussi l'ex. 2 de la séance 15.

Exemple 2 : Dans l'exemple 2 page 66,  $X$  et  $Y$  ne sont pas indépendantes. Mais si  $(X, Y)$  suit une loi uniforme sur le carré  $[0, 1] \times [0, 1]$  les lois marginales de  $X$  et  $Y$  respectivement sont les lois uniformes sur  $[0, 1]$  et  $X$  et  $Y$  sont indépendantes.

Exemple 3. Loi normale à 2 dimensions de paramètre  $r$  (cf. p. 64)

C'est la loi définie par la densité  $f(x,y) = C_r e^{-\frac{1}{2}Q(x,y)}$

où:  $-1 < r < 1$ ,  $C_r = \frac{1}{2\pi\sqrt{1-r^2}}$ ,  $Q(x,y) = \frac{x^2 - 2rxy + y^2}{1-r^2}$

On a vu que:  $\int_{-\infty}^{+\infty} f(x,y) dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \varphi(x)$

De même on a:  $\int_{-\infty}^{+\infty} f(x,y) dx = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \varphi(y)$

Donc  $X$  et  $Y$  ont une loi <sup>marginales</sup> normale (réduite). On voit que  $X$  et  $Y$  sont indépendants, car  $r=0$ .

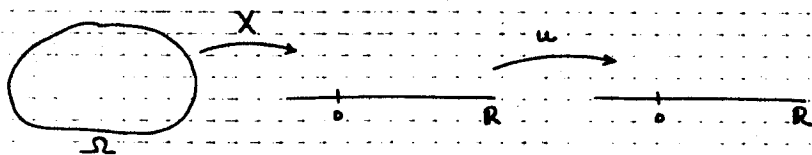
Les considérations de ce n° se généralisent sans peine au cas de  $n$  v.a.  $X_1, \dots, X_n$  définies sur un même espace probabilisé. Nous précisons seulement la définition:

$X_1, \dots, X_n$  sont indépendants si:

$$P[X_1 \in I_1, \dots, X_n \in I_n] = P[X_1 \in I_1] \cdots P[X_n \in I_n]$$

quel que soient les intervalles  $I_1, \dots, I_n$  de  $\mathbb{R}$ .

### 6.5. Fonction d'une variable aléatoire, de plusieurs v.a.



Considérons une v.a.  $X: \Omega \rightarrow \mathbb{R}$  et une fonction  $u: \mathbb{R} \rightarrow \mathbb{R}$ . Posons  $Y = u \circ X = u(X)$ . Nous supposons toujours que  $Y$  est une v.a.; il suffit pour cela que  $u^{-1}(I) \in \mathcal{B}_{\mathbb{R}}$  pour tout intervalle  $I \subset \mathbb{R}$ . Le problème est ici d'étudier la loi de probabilité  $P_Y$  à fonction de  $P_X$ . Il est clair que

$$P[Y \in I] = P[X \in u^{-1}(I)]$$

quel que soit  $I$ .

En particulier la fonction de répartition de  $Y$  est donnée par

$$F_Y(y) = P[Y \leq y] = P[X \in u^{-1}(-\infty, y]] \quad (\forall y)$$

Exemple 1.  $Y = aX + b$

- si  $a > 0$ :  $Y < y \Leftrightarrow X < \frac{y-b}{a}$  donc  $F_Y(y) = F_X\left(\frac{y-b}{a}\right)$

- si  $a < 0$ :  $Y < y \Leftrightarrow X > \frac{y-b}{a}$  donc  $F_Y(y) = 1 - F_X\left(\frac{y-b}{a} +\right)$

Exemple 2:  $Y = |X|$

On a:  $Y < y$  n'est possible si  $y \geq 0$ , donc  $F_Y(y) = 0$  si  $y < 0$ , et si  $y \geq 0$ ,  $Y < y \Leftrightarrow -y < X < y$ , donc  $F_Y(y) = F_X(y) - F_X(-y+)$  si  $y > 0$ .

Cas discret: si  $X$  est discrète alors  $Y = u(X)$  est aussi discrète et

$$P[Y = y_j] = \sum_{u(x_i) = y_j} P[X = x_i] \quad (\forall j)$$

Cas absolument continu: si  $X$  possède une densité  $f$  continue et si  $u$  possède une dérivée continue telle que  $u'(x) \neq 0$  ( $\forall x$ ) alors  $Y = u(X)$  possède une densité  $g$  donnée par

$$g(y) = \begin{cases} f(u^{-1}(y)) |u^{-1}'(y)| & \text{si } y \in u(\mathbb{R}) \\ 0 & \text{si } y \notin u(\mathbb{R}) \end{cases}$$

Preuve: Ou bien  $u'(x) > 0$  ( $\forall x$ ) ou bien  $u'(x) < 0$  ( $\forall x$ ) i.e.  $u$  est strictement croissante ou décroissante. Dans les deux cas  $u^{-1}(-\infty, y[)$  est un intervalle: soit  $]-\infty, u^{-1}(y)[$  soit  $]u^{-1}(y), +\infty[$ ; dans le 1<sup>er</sup> cas on a (cf. haut de la page):

$$F_Y(y) = F_X(u^{-1}(y))$$

et dans le 2<sup>ème</sup> cas:  $F_Y(y) = 1 - F_X(u^{-1}(y))$  ( $F_X$  continue!)

Par dérivation on obtient la densité  $g$  cherchée.

Exemple: Si la loi de  $X$  est la loi normale (réduite) de densité  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ , alors  $Y = \sigma X + \mu$  ( $\sigma > 0, \mu \in \mathbb{R}$ ) a la densité

$$g(y) = \frac{1}{\sigma} \varphi\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

c'est la densité de la loi normale de paramètres  $\mu$  et  $\sigma$ .

On peut également considérer une fonction de plusieurs variables aléatoires: si  $X_1, \dots, X_n$  sont des v.a. et  $u: \mathbb{R}^n \rightarrow \mathbb{R}$  est une fonction (borélienne) alors  $Y = u(X_1, \dots, X_n)$  est une variable aléatoire. Sa loi de probabilité est donnée par

$$P[Y \in I] = P[(X_1, \dots, X_n) \in u^{-1}(I)]$$

pour tout intervalle  $I \subset \mathbb{R}$ .

Exemple: Somme de 2 v.a. indépendantes: convolution.

Soient  $X$  et  $Y$  deux v.a. indépendantes et  $Z = X+Y$ .

a) Cas où  $X$  et  $Y$  sont à valeurs dans  $\mathbb{Z}$ .

Alors pour tout  $k \in \mathbb{Z}$  on a

$$P[Z=k] = \sum_{i=-\infty}^{+\infty} P[X=i, Y=k-i] = \sum_{i=-\infty}^{+\infty} P[X=i]P[Y=k-i]$$

b) Cas où  $X$  et  $Y$  sont absolument continues.

Soient  $f$  une densité pour  $X$ ,  $g$  une densité pour  $Y$ .

Cherchons une densité pour  $Z = X+Y$ .

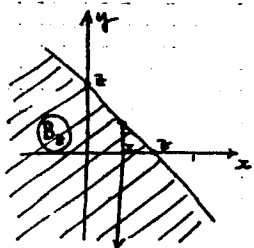
$X$  et  $Y$  étant indépendantes  $(X, Y)$  admet la densité  $f(x)g(y)$ ;

on a donc:

$$F_Z(z) = P[X+Y \leq z] = P[(X, Y) \in B_z]$$

$$\text{où } B_z = \{(x, y) \in \mathbb{R}^2 \mid x+y \leq z\}$$

$$\begin{aligned} \text{donc } F_Z(z) &= \iint_{B_z} f(x)g(y) dx dy \\ &= \int_{-\infty}^{+\infty} dx f(x) \int_{-\infty}^{z-x} g(y) dy \quad y' = y+x \\ &= \int_{-\infty}^{+\infty} dx f(x) \int_{-\infty}^z g(y-x) dy \end{aligned}$$



$$F_Z(z) = \int_{-\infty}^z dy \int_{-\infty}^{+\infty} f(x)g(y-x) dx$$

ce qui montre que  $h(y) = \int_{-\infty}^{+\infty} f(x)g(y-x) dx$  est une densité pour  $Z = X+Y$ . On dit que  $h$  est le produit de convolution de  $f$  et  $g$ , noté  $h = f * g$ .

Exemple 1. Si  $X$  est de loi binomiale  $(n, p)$  et  $Y$  de loi binomiale  $(m, p)$  et  $X, Y$  sont indépendantes, alors  $X+Y$  est de loi binomiale  $(n+m, p)$ , comme il est intuitivement clair (penser à  $X$  = nombre de succès dans la  $n$  premières expériences,  $Y$  = idem dans les  $m$  suivantes ...). On le vérifie facilement:

$$\begin{aligned} P[X+Y=k] &= \sum_{i=-\infty}^{+\infty} P[X=i]P[Y=k-i] = \sum_{i=0}^k \binom{n}{i} p^i q^{n-i} \binom{m}{k-i} p^{k-i} q^{m-k+i} \\ &= p^k q^{n+m-k} \sum_{i=0}^k \binom{n}{i} \binom{m}{k-i} = \binom{n+m}{k} p^k q^{n+m-k} \end{aligned}$$

Exemple 2. Si  $X$  est de loi de Poisson de paramètre  $\alpha$  et  $Y$  de loi de Poisson de paramètre  $\beta$  et si  $X$  et  $Y$  sont indépendantes, alors  $X+Y$  suit une loi de Poisson de paramètre  $\alpha+\beta$ .

$$\begin{aligned} P[X+Y=k] &= \sum_{i=0}^k P[X=i]P[Y=k-i] = \sum_{i=0}^k \frac{\alpha^i}{i!} e^{-\alpha} \frac{\beta^{k-i}}{(k-i)!} e^{-\beta} \\ &= \frac{e^{-(\alpha+\beta)}}{k!} \sum_{i=0}^k \binom{k}{i} \alpha^i \beta^{k-i} = \frac{(\alpha+\beta)^k}{k!} e^{-(\alpha+\beta)} \end{aligned}$$

Exemple 3. Si  $X$  et  $Y$  sont des v.a. indépendantes de loi normale réduite, alors  $X+Y$  est une v.a. normale de paramètres  $0$  et  $\sqrt{2}$ . En effet  $X+Y$  admet pour densité

$$\begin{aligned} h(y) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-x)^2} dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp\left[-\frac{1}{4}y^2 - (x-\frac{y}{2})^2\right] dx \\ &= \frac{1}{2\sqrt{\pi}} e^{-y^2/4} \underbrace{\int_{-\infty}^{+\infty} \frac{e^{-(x-y/2)^2}}{\sqrt{\pi}} dx}_1 = \frac{1}{\sqrt{2}\sqrt{2\pi}} e^{-(y/\sqrt{2})^2} \end{aligned}$$



Plus généralement on montre par un calcul direct analogue que si  $X$  est une v.a. normale  $(\mu, \sigma)$ ,  $Y$  une v.a. normale  $(\nu, \rho)$  et  $X$  et  $Y$  sont indépendantes, alors  $Z = X + Y$  est une v.a. normale  $(\mu + \nu, \sqrt{\sigma^2 + \rho^2})$ .

Remarque: les exemples 1, 2 et 3 ci-dessus se généralisent (par induction) à la somme  $Z = X_1 + \dots + X_n$  de  $n$  v.a. indépendantes.

Nous allons voir, pour terminer ce paragraphe, comment les résultats précédents - et leurs généralisations naturelles - permettent de résoudre deux problèmes concrets.

### Application 1: Erreurs de mesure - Intervalle de confiance

Problème: On cherche à déterminer une grandeur  $\mu$  par une série de  $n$  mesures. Soient  $X_1, \dots, X_n$  les résultats de ces  $n$  mesures; ce sont des variables aléatoires. On supposera qu'elles sont indépendantes et que

$$X_i = \mu + \sigma Z_i \quad (i=1, 2, \dots, n)$$

où  $Z_1, \dots, Z_n$  sont des v.a. normales  $(0, 1)$  et  $\sigma > 0$  représente la qualité des mesures ( $\sigma Z_i$  représente l'erreur commise à la  $i$ -ième mesure). Pour estimer  $\mu$  il est naturel de considérer la v.a.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Question: Quelle erreur commet-on en prenant  $\bar{X}$  comme estimation de  $\mu$ ?

La réponse que nous donnerons sera de la forme: "il y a une probabilité  $\gamma$  que  $\mu$  se trouve dans l'intervalle  $(\bar{X} - a(\gamma), \bar{X} + a(\gamma))$ ". Cet intervalle est appelé intervalle de confiance au seuil de confiance  $\gamma$ .

1<sup>er</sup> cas:  $\sigma$  est connu.

Posez  $\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$ ; on a  $X_i = \mu + \sigma Z_i$ , donc  $\bar{X} = \mu + \sigma \bar{Z}$ .

Comme  $\sum_{i=1}^n Z_i$  est une v.a. normale  $(0, \sqrt{n})$ ,  $\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$  est une v.a. normale  $(0, 1)$ . Mais

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i = \sqrt{n} \bar{Z} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

On a donc

$$P\left[-c < \sqrt{n} \frac{\bar{X} - \mu}{\sigma} < c\right] = P\left[-\frac{c\sigma}{\sqrt{n}} < \bar{X} - \mu < \frac{c\sigma}{\sqrt{n}}\right] \\ = \Phi(c) - \Phi(-c) = 2\Phi(c) - 1$$

[Il y a donc une probabilité  $2\Phi(c) - 1$  pour que  $\mu$  se trouve dans l'intervalle  $(\bar{X} - \frac{c\sigma}{\sqrt{n}}, \bar{X} + \frac{c\sigma}{\sqrt{n}})$ .

2<sup>ème</sup> cas:  $\sigma$  est inconnu.

Alors  $\sigma$  doit être estimé à partir des  $X_i$ ; on utilise en général comme estimateur de  $\sigma^2$  la valeur prise par la v.a.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Ce qui remplace la v.a. normale  $\sqrt{n} \frac{\bar{X} - \mu}{\sigma}$  du cas ci-dessus est la v.a.

$$T = \sqrt{n} \frac{\bar{X} - \mu}{S};$$

sa loi de probabilité est donnée par le théorème suivant.

Théorème: Posez  $k = n - 1$ ;  $T$  admet pour densité la fonction:

$$f_k(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi} \Gamma\left(\frac{k}{2}\right)} \frac{1}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} \quad (t \in \mathbb{R})$$

(C'est la loi de Student à  $k$  degrés de liberté).

Alors, si  $G_R(t) = \int_{-\infty}^t g_R(s) ds$  est la fonction de répartition de la v.a.  $T$ , on a

$$P[|T| < c] = P\left[-\frac{sc}{\sqrt{n}} < \bar{X} - \mu < \frac{sc}{\sqrt{n}}\right] = 2G_R(c) - 1$$

i.e. il y a une probabilité  $2G_R(c) - 1$  pour que  $\mu$  se trouve dans l'intervalle  $(\bar{X} - \frac{sc}{\sqrt{n}}, \bar{X} + \frac{sc}{\sqrt{n}})$

La preuve du théorème est longue :

1) Posons  $\bar{Z} = \frac{1}{n} \sum Z_i$  ; on a vu que  $Y = \sqrt{n} \bar{Z}$  est une v.a. normale  $(0, 1)$ .

Posons  $R = \sum (Z_i - \bar{Z})^2$  ; on a

$$(n-1)S^2 = \sum (X_i - \bar{X})^2 = \sum (\mu + \sigma Z_i - \mu - \sigma \bar{Z})^2 = \sigma^2 \sum (Z_i - \bar{Z})^2 = \sigma^2 R$$

donc  $S = \sigma \sqrt{\frac{R}{n-1}}$

Alors  $T = \sqrt{n} \frac{\bar{X} - \mu}{S} = \sqrt{n} \frac{\sigma \bar{Z}}{\sigma \sqrt{\frac{R}{n-1}}} = \frac{\sqrt{n} \bar{Z}}{\sqrt{R/(n-1)}} = \frac{Y}{\sqrt{R/(n-1)}}$

On va d'abord étudier la v.a.  $R$ .

2) On remarque que dans  $R = \sum (Z_i - \bar{Z})^2$ , les v.a.  $Z_i - \bar{Z}$  ne sont pas indépendantes car  $\sum (Z_i - \bar{Z}) = 0$ .

On va représenter  $R$  sous la forme  $\sum W_i^2$  où les  $W_i$  sont des v.a. normales  $(0, 1)$  indépendantes.

Cela se fait par la transformation de Helmert :

$$\begin{cases} W_1 = \frac{1}{\sqrt{1-2}} (Z_1 - Z_2) \\ W_2 = \frac{1}{\sqrt{2-3}} (Z_1 + Z_2 - 2Z_3) \\ \dots \\ W_{n-1} = \frac{1}{\sqrt{(n-1)n}} (Z_1 + Z_2 + \dots + Z_{n-1} - (n-1)Z_n) \\ W_n = \frac{1}{\sqrt{n}} (Z_1 + Z_2 + \dots + Z_{n-1} + Z_n) \end{cases}$$

- Lemme 1. a)  $R = \sum_{i=1}^{n-1} W_i^2$   
 b) les  $W_i$  ( $1 \leq i \leq n$ ) sont normales  $(0, 1)$  indépendantes,  
 c)  $R$  et  $Y$  sont indépendantes.

Preuve. a) Posons  $\vec{Z} = (Z_1, \dots, Z_n)$  et  $\vec{W} = (W_1, \dots, W_n)$  ; alors  $\vec{W} = A \vec{Z}$  où  $A$  est la matrice  $n \times n$  :

$$A = \begin{pmatrix} \frac{1}{\sqrt{1-2}} & \frac{1}{\sqrt{1-2}} & 0 & \dots & 0 & 0 \\ \frac{1}{\sqrt{2-3}} & \frac{1}{\sqrt{2-3}} & \frac{-2}{\sqrt{2-3}} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \dots & \frac{1}{\sqrt{(n-1)n}} & \frac{-(n-1)}{\sqrt{(n-1)n}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \end{pmatrix}$$

On vérifie immédiatement que  $A$  est orthogonale :  $A^t A = 4A A = I$ .  
 Donc on a  $\sum Z_i^2 = \sum W_i^2 = \sum_{i=1}^{n-1} W_i^2 + W_n^2$ . Mais  $W_n = \sqrt{n} \bar{Z}$   
 et

$$R = \sum (Z_i - \bar{Z})^2 = \sum Z_i^2 - 2\bar{Z} \sum Z_i + n\bar{Z}^2 = \sum_{i=1}^{n-1} W_i^2 - n\bar{Z}^2 = \sum_{i=1}^{n-1} W_i^2 + n\bar{Z}^2 - n\bar{Z}^2 = \sum_{i=1}^{n-1} W_i^2$$

b) Comme les  $Z_i$  sont indépendantes, une densité pour  $\vec{Z}$  est donnée par

$$f(\vec{Z}) = f(z_1, \dots, z_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \sum z_i^2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \|\vec{Z}\|^2}$$

Si  $B$  est un borélien de  $\mathbb{R}^n$  on a

$$P[\vec{W} \in B] = P[\vec{Z} \in A^{-1}(B)] = \int_{A^{-1}(B)} f(\vec{Z}) d\vec{Z} = \int_B f(A\vec{w}) |\det A^{-1}| d\vec{w}$$

donc

$$g(\vec{w}) = f(A\vec{w}) \left| \frac{1}{\det A} \right| = f(A\vec{w})$$

est une densité pour  $\vec{W}$  ; mais  $A^{-1}$  est aussi orthogonale

$$g(\vec{w}) = f(A\vec{w}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \|A\vec{w}\|^2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2} \|\vec{w}\|^2}$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-w_i^2/2}, \text{ d'où b.}$$

c) C'est clair car  $R = \sum_{i=1}^{m+1} W_i^2$  et  $Y = W_m$ , et les  $W_i$  sont indépendantes.

Lemme 2  $R$  admet pour densité la fonction

$$h(r) = \begin{cases} 0 & \text{si } r \leq 0 \\ \frac{1}{2^{\frac{m+1}{2}} \Gamma(\frac{m+1}{2})} r^{\frac{m}{2}-1} e^{-\frac{r}{2}} & \text{si } r > 0 \end{cases}$$

(c'est la loi du  $\chi^2$  à  $m+1$  degrés de liberté).

Preuve, Induction sur  $m$ :

$m=2$ :  $R = W_1^2$  où  $W_1$  est une v.a. normale  $(0,1)$ . On a:

$$R < r \Leftrightarrow W_1^2 < r \Leftrightarrow -\sqrt{r} < W_1 < \sqrt{r} \quad \text{donc si}$$

$H_1$  désigne la fonction de répartition de  $R$  on a

$$H_1(r) = \Phi(\sqrt{r}) - \Phi(-\sqrt{r}) = 2 \frac{1}{\sqrt{2\pi}} \int_0^{\sqrt{r}} e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^r e^{-x/2} y^{-1/2} dy \quad \begin{matrix} x^2 = y \\ dx = \frac{dy}{2\sqrt{y}} \end{matrix}$$

d'où  $h_1(r) = H_1'(r) = \frac{1}{\sqrt{2\pi}} e^{-r/2} r^{-1/2} = \frac{1}{2^{1/2} \Gamma(1/2)} r^{-1/2} e^{-r/2}$  (si  $r > 0$ )

On a  $h_1(r) = 0$  si  $r \leq 0$  (car  $R = W_1^2$  est toujours  $\geq 0$ ).

Il suffit de montrer que  $\sum_{i=1}^m W_i^2$  a la densité annoncée et étudier  $\sum_{i=1}^{m+1} W_i^2$ ; comme  $W_1, \dots, W_m, W_{m+1}$  sont indépendantes, alors  $\sum_{i=1}^m W_i^2$  et  $W_{m+1}^2$  sont indépendantes, donc

$$h_{m+1} = h_m * h_1$$

où  $h_m(r) = \frac{1}{2^{\frac{m}{2}} \Gamma(\frac{m}{2})} r^{\frac{m}{2}-1} e^{-r/2}$  et  $h_1$  comme ci-dessus.

On a donc:

$$h_{m+1}(r) = \int_{-\infty}^{+\infty} h_m(x) h_1(r-x) dx = \int_0^r \text{idem}$$

car  $x \geq 0$   
et  $r-x \geq 0$

$$= \int_0^r \frac{1}{2^{\frac{m}{2}} \Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}} \frac{1}{2^{1/2} \Gamma(1/2)} (r-x)^{-1/2} e^{-\frac{r-x}{2}} dx$$

$$= \frac{e^{-r/2}}{2^{\frac{m+1}{2}} \Gamma(\frac{m}{2}) \Gamma(1/2)} \int_0^r x^{\frac{m}{2}-1} (r-x)^{-1/2} dx$$

$$= \frac{e^{-r/2} r^{\frac{m}{2}}}{2^{\frac{m+1}{2}} \Gamma(\frac{m}{2}) \Gamma(1/2)} \int_0^1 y^{\frac{m}{2}-1} (1-y)^{-1/2} dy \quad y = \frac{x}{r}$$

$$= \frac{1}{2^{\frac{m+1}{2}} \Gamma(\frac{m+1}{2})} r^{\frac{m+1}{2}-1} e^{-r/2}$$

en vertu de l'égalité:

$$\int_0^1 y^{a-1} (1-y)^{b-1} dy = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)}$$

$\Gamma(a) \Gamma(b) = \int_0^\infty \int_0^\infty s^{a-1} t^{b-1} e^{-(s+t)} ds dt = \dots$

changement de var.  $\begin{cases} s = uv \\ t = u(1-v) \end{cases} \quad \begin{cases} u = s+t \\ v = \frac{s}{s+t} \end{cases}$

$ds dt = \left| \frac{\partial(s,t)}{\partial(u,v)} \right| du dv = u du dv$  car  $\frac{\partial(s,t)}{\partial(u,v)} = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u$

$\dots = \int_0^\infty du \int_0^1 dv u u^{a-1} v^{a-1} u^{b-1} (1-v)^{b-1} e^{-u}$

$= \int_0^\infty du e^{-u} u^{a+b-1} \int_0^1 dv v^{a-1} (1-v)^{b-1}$

$= \Gamma(a+b) \int_0^1 v^{a-1} (1-v)^{b-1} dv$

Donc le lemme 2 est démontré.

3) D'après les lemmes 1 et 2 le couple  $(Y, R)$  admet pour densité la fonction

$$f(y, r) = \frac{1}{\sqrt{2\pi} 2^{\frac{m+1}{2}} \Gamma(\frac{m+1}{2})} r^{\frac{m-3}{2}} e^{-\frac{1}{2}(y^2+r)}$$

pour  $r > 0$  et  $-\infty < y < +\infty$ .

Considérons la transformation:  $\Psi: (y, r) \mapsto (t, u)$  où

$$\begin{cases} t = y\sqrt{\frac{m-1}{r}} \\ u = r \end{cases}$$

La transformation inverse est  $\Psi^{-1} \begin{cases} y = \sqrt{\frac{u}{m-1}} t \\ r = u \end{cases}$ ; on a

$$J_{\Psi^{-1}}(t, u) = \frac{\partial(y, r)}{\partial(t, u)} = \begin{vmatrix} \sqrt{\frac{u}{m-1}} & \dots \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{u}{m-1}}$$

la densité du couple  $(T, U) = \Psi(Y, R)$  est donnée par:

$$\begin{aligned} P[(T, U) \in B] &= P[(Y, R) \in \Psi^{-1}(B)] \\ &= \iint_{\Psi^{-1}(B)} f(y, r) dy dr = \iint_B f(\Psi^{-1}(t, u)) \left| \frac{\partial(y, r)}{\partial(t, u)} \right| dt du \end{aligned}$$

i.e. c'est

$$\begin{aligned} g(t, u) &= f(\Psi^{-1}(t, u)) \left| \frac{\partial(y, r)}{\partial(t, u)} \right| \\ &= \frac{\sqrt{u}}{\sqrt{m-1}} \frac{1}{\sqrt{2\pi} 2^{\frac{m-1}{2}} \Gamma(\frac{m-1}{2})} u^{\frac{m-3}{2}} e^{-\frac{1}{2}[\frac{u}{m-1}t^2 + u]} \\ &= \frac{1}{\sqrt{2\pi} \sqrt{m-1} 2^{\frac{m-1}{2}} \Gamma(\frac{m-1}{2})} u^{\frac{m-2}{2}} e^{-\frac{u}{2}[1 + \frac{t^2}{m-1}]} \end{aligned}$$

On obtient la densité de T comme densité marginale du couple  $(T, U)$ :

$$g_{m1}(t) = \int_{-\infty}^{+\infty} g(t, u) du = \text{ct.} \int_{-\infty}^{+\infty} u^{\frac{m-2}{2}} e^{-\frac{u}{2}[1 + \frac{t^2}{m-1}]} du = \dots$$

on pose  $v = \frac{u}{2}[1 + \frac{t^2}{m-1}]$ ,  $dv = \frac{1}{2}[1 + \frac{t^2}{m-1}] du$

$$\begin{aligned} \dots &= \text{ct.} \frac{2^{\frac{m-2}{2}+1}}{[1 + \frac{t^2}{m-1}]^{\frac{m-2}{2}+1}} \int_{-\infty}^{+\infty} v^{\frac{m-2}{2}} e^{-v} dv \\ &= \frac{2^{\frac{m-2}{2}} \Gamma(\frac{m}{2})}{\sqrt{2\pi} \sqrt{m-1} 2^{\frac{m-1}{2}} \Gamma(\frac{m-1}{2})} \frac{1}{[1 + \frac{t^2}{m-1}]^{\frac{m}{2}}} = \text{cfd.} \end{aligned}$$

1<sup>er</sup> cas:

Table 12.1.1. Determination of a Confidence Interval for the Mean  $\mu$  of a Normal Distribution with Known Variance  $\sigma^2$

1st step. Choose a confidence level  $\gamma$  (95%, 99%, or the like).  
 2nd step. Determine the corresponding  $c$ :

$\gamma$	0.90	0.95	0.99	0.999
$c$	1.645	1.960	2.576	3.291

3rd step. Compute the mean  $\bar{x}$  of the sample  $x_1, \dots, x_n$ .  
 4th step. Compute

(1)  $k = c\sigma/\sqrt{n}$ .

The confidence interval for the mean  $\mu$  of the population is

(2) CONF  $(\bar{x} - k \leq \mu \leq \bar{x} + k)$ .

2<sup>ème</sup> cas:

Table 12.3.1. Determination of a Confidence Interval for the Mean  $\mu$  of a Normal Distribution with Unknown Variance

1st step. Choose a confidence level  $\gamma$  (95%, 99%, or the like).  
 2nd step. Determine the solution  $c$  of the equation

(1)  $F(c) = \frac{1}{2}(1 + \gamma)$

from the table of the  $t$ -distribution with  $n - 1$  degrees of freedom (Table 8 in Appendix 4;  $n$  = sample size). Compute  $c/\sqrt{n}$ . For  $\gamma = 0.95$  and 0.99 and some values of  $n$  the values of  $c/\sqrt{n}$  are as follows.

$n$	$c/\sqrt{n}$ when		$n$	$c/\sqrt{n}$ when		$n$	$c/\sqrt{n}$ when	
	$\gamma = 0.95$	$\gamma = 0.99$		$\gamma = 0.95$	$\gamma = 0.99$		$\gamma = 0.95$	$\gamma = 0.99$
2	8.99	43.0	8	0.836	1.24	30	0.373	0.503
3	2.48	5.73	9	0.769	1.12	40	0.320	0.428
4	1.59	2.92	10	0.715	1.03	50	0.284	0.379
5	1.24	2.06	15	0.554	0.769	100	0.198	0.263
6	1.05	1.65	20	0.468	0.640	200	0.139	0.184
7	0.925	1.40	25	0.413	0.559	500	0.088	0.116

3rd step. Compute the mean  $\bar{x}$  and the variance  $s^2$  of the sample  $x_1, \dots, x_n$ .  
 4th step. Compute

(2)  $k = sc/\sqrt{n}$ .

The confidence interval is

(3) CONF  $(\bar{x} - k \leq \mu \leq \bar{x} + k)$ .

6.7. Application 2. Rayonnement radioactif.

Situation: Une certaine quantité de substance radioactive contient  $n$  atomes instables. Chacun d'eux peut émettre une particule ( $\alpha$  ou  $\beta$ ) puis il devient stable.

On numérote les atomes instables de 1 à  $n$  et on déclenche une montre en  $t=0$ . Soit pour tout  $i=1, \dots, n$

$X_i$  = temps auquel le  $i^{\text{ème}}$  atome émet son rayonnement

$X_i$  est une v.a. à valeurs dans  $[0, \infty[$ .

Hypothèses:

(I)  $X_1, \dots, X_n$  sont des v.a. indépendantes et de même loi; on notera  $F$  la fct. de répartition de chaque  $X_i$  i.e.

$$F(t) = P[X_i \leq t] \quad \forall i=1, 2, \dots, n$$

(II)  $F(t) = 0$  si  $t \leq 0$  et  $0 < F(t) < 1$  si  $t > 0$ .

(III) Pour tout  $s=1, \dots, n$   $P[X_i \geq t+s | X_i \geq t] = P[X_i \geq s]$   $\forall t > 0, \forall s > 0$

i.e. l'émission du rayonnement ne devient pas plus ou moins probable lorsque le temps passe (spontanéité).

1) Lemme 1. Les  $X_i$  suivent une loi exponentielle i.e.

$$F(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ 1 - e^{-\alpha t} & \text{si } t > 0 \end{cases} \quad (\alpha \text{ est } > 0)$$

et

$$f(t) = F'(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \alpha e^{-\alpha t} & \text{si } t > 0 \end{cases}$$

Preuve (cf. ex. 14). On pose  $G(t) = 1 - F(t) = P[X_i \geq t]$

(III) signifie que  $\frac{P[X_i \geq t+s \text{ et } X_i \geq t]}{P[X_i \geq t]} = P[X_i \geq s]$

donc :  $G(t+s) = G(t)G(s) \quad \forall t > 0, \forall s > 0$ .

On a :  $G\left(\frac{m}{n}\right) = G\left(\frac{1}{n} + \dots + \frac{1}{n}\right) = G\left(\frac{1}{n}\right)^m$  pour  $m, n \in \mathbb{N}_+$ .

En particulier ( $m=n$ ) :  $G(1) = G\left(\frac{1}{n}\right)^n$  donc  $G\left(\frac{1}{n}\right) = G(1)^{\frac{1}{n}}$

Ainsi  $\forall r \in \mathbb{Q}, r > 0, G(r) = G(1)^r$

Par hypo. on a  $0 < G(1) < 1$ , donc il existe  $\alpha > 0$  t.q.  $G(1) = e^{-\alpha}$

On a donc  $G(r) = e^{-\alpha r}$ . Comme  $G$  est continue à gauche (car  $F$  l'est!) on en déduit que  $G(t) = e^{-\alpha t} \quad \forall t > 0$ , donc que  $F(t) = 1 - e^{-\alpha t} \quad \forall t > 0$ .

Déf. Demi-vie (de la substance) = solution de  $F(z) = \frac{1}{2}$

On trouve  $\tau = \frac{\log 2}{\alpha}$

2) Intéressons-nous maintenant au nombre d'émissions dans l'intervalle  $[0, t[$ . Posons, pour tout  $t > 0$ ,

$$N_t = \text{nb. d'émissions dans } [0, t[$$

$N_t$  est une v.a. à valeurs dans  $\mathbb{N}$

Lemme 2.  $N_t$  suit une loi binomiale de par.  $n, p = 1 - e^{-\alpha t}$ .

Dém. On a :

$$N_t = k \iff \text{exactement } k \text{ des événements } [X_i < t], \text{ seism, ont eu lieu.}$$

Par hypo. (I) les  $n$  événements  $[X_i < t]$  sont indépendants et ont tous pour proba.  $p = F(t) = 1 - e^{-\alpha t}$ . C'est donc bien la loi binomiale qui s'afflige

$$P[N_t = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

Remarque: Supposons  $n$  très grand et  $t$  très petit par rapport à la demi-vie de la substance  $\tau = \frac{\log 2}{\alpha}$ . Alors  $\alpha t$  est très petit et

$$p = 1 - e^{-\alpha t} \approx \alpha t.$$

L'approximation de Poisson est alors légitime :

$$P[N_t = k] \cong \frac{(mt)^k}{k!} e^{-mt} = \frac{(\beta t)^k}{k!} e^{-\beta t}$$

en posant  $\beta = m\alpha$  :  $\beta$  est l'intensité de la radiation.

Ainsi  $N_t$  a approximativement une distribution de Poisson de paramètre  $\beta t$  (si  $t \ll \tau$ ).

Expérimentalement, Rutherford, Chadwick et Ellis (1926) ont montré que ce modèle était très bon (cf. Feller p. 160).

On voit que  $\{N_t\}_{t \geq 0}$  est un processus de Poisson de paramètre  $\beta$ . Ces processus sont de bons modèles pour de nombreux phénomènes, par exemple :

- radioactivité [cf. ci-dessus],
- nb. d'appels téléphoniques reçus d'une ville,
- nb. d'accidents à un croisement,
- nb. de pannes d'une machine,
- nb. de clients arrivant à un comptoir.

[cf. série 17].

3) Reprenons notre substance radioactive ! Si on s'intéresse au moment auquel se produit la  $k$ <sup>ème</sup> émission on introduit la v.a.

$$Y_1 = \text{temps auquel se produit la 1<sup>ère</sup> émission} = \min(X_1, \dots, X_n).$$

$$Y_2 = \text{-----} \quad 2^{\text{ème}} \text{-----}$$

$$Y_m = \text{-----} \quad m^{\text{ème}} \text{-----} = \max(X_1, \dots, X_n).$$

On a donc  $Y_1 \leq Y_2 \leq \dots \leq Y_m$  (ce qui montre tout de suite que la v.a.  $Y_i$  ne sont pas indépendantes ; on dit que  $Y_1, \dots, Y_m$  est la "statistique d'ordre" des  $X_i$ ).

Lemme 3.  $Y_k$  admet pour densité la fonction

$$h_k(y) = k \binom{m}{k} F(y)^{k-1} [1-F(y)]^{m-k} f(y) \\ = k \binom{m}{k} (1-e^{-\alpha y})^{k-1} e^{-\alpha y(m-k)} \alpha e^{-\alpha y} \quad \forall y > 0$$

et  $h_k(y) = 0$  si  $y \leq 0$ .

Dém. Soit  $H_k$  la fct. de répartition de  $Y_k$  :  $H_k(y) = P[Y_k \leq y]$

On a :

$$Y_k \leq y \Leftrightarrow \text{au moins } k \text{ des événements } [X_i \leq y] \text{ sont réalisés} \\ \Leftrightarrow k \text{ ou } k+1 \text{ ou } \dots \text{ ou } n \text{ des év. } [X_i \leq y] \text{ sont réalisés}$$

Mais on a vu (preuve du lemme 2) que la probabilité que exactement  $l$  des événements  $[X_i \leq y]$  soient réalisés est

$$\binom{m}{l} F(y)^l [1-F(y)]^{m-l}$$

$$\text{Donc } H_k(y) = \sum_{l=k}^m \binom{m}{l} F(y)^l [1-F(y)]^{m-l}$$

On obtient  $h_k(y)$  en dérivant  $H_k(y)$  :

$$h_k(y) = \sum_{l=k}^m \binom{m}{l} l F(y)^{l-1} [1-F(y)]^{m-l} f(y) \\ - \sum_{l=k}^m \binom{m}{l} (m-l) F(y)^l [1-F(y)]^{m-l-1} f(y) \\ = k \binom{m}{k} F(y)^{k-1} [1-F(y)]^{m-k} f(y) + \sum_{l=k}^{m-1} \binom{m}{l+1} (l+1) F(y)^l [1-F(y)]^{m-l-1} f(y) \\ - \sum_{l=k}^{m-1} \binom{m}{l} (m-l) F(y)^l [1-F(y)]^{m-l-1} f(y)$$

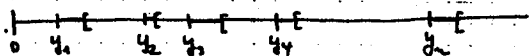
$$\text{d'où le résultat car } \binom{m}{l+1} (l+1) = \frac{m!}{l!(m-l-1)!} = \binom{m}{l} (m-l).$$

Remarque :  $h_1(y) = m\alpha e^{-\alpha y} = \beta e^{-\beta y}$  si  $y > 0$  donc  $Y_1 = \min(X_1, \dots, X_m)$  suit une loi exponentielle de param.  $\beta$ .

Lemme 3' la v.a.  $\vec{Y} = (Y_1, \dots, Y_n)$  admet pour densité

$$h(y_1, \dots, y_n) = \begin{cases} n! f(y_1) \dots f(y_n) & \text{si } y_1 < y_2 < \dots < y_n \\ 0 & \text{sinon} \end{cases}$$

Dém. Prenons  $y_1 < y_2 < \dots < y_n$  et choisissons  $\delta_1 > 0, \dots, \delta_n > 0$  avec petits pour que les intervalles  $[y_1, y_1 + \delta_1], \dots, [y_n, y_n + \delta_n]$  ne se couvrent pas.



$$\begin{aligned} & P[y_i \leq Y_i < y_i + \delta_i, \forall i=1,2,\dots,n] \\ &= P[\text{il existe } \sigma \in \mathcal{S}_n \text{ t.q. } y_i \leq X_{\sigma(i)} < y_i + \delta_i, \forall i=1,2,\dots,n] \\ &= \sum_{\sigma \in \mathcal{S}_n} P[y_i \leq X_{\sigma(i)} < y_i + \delta_i, \forall i=1,2,\dots,n] \quad (\text{intervalles disjoints!}) \\ &= \sum_{\sigma \in \mathcal{S}_n} \prod_{i=1}^n (F(y_i + \delta_i) - F(y_i)) \quad (\text{indépendance des } X_i) \\ &= n! \prod_{i=1}^n (F(y_i + \delta_i) - F(y_i)) \end{aligned}$$

Par passage à la limite lorsque  $\delta_1 \rightarrow 0, \dots, \delta_n \rightarrow 0$  on obtient le résultat.

4) Si on s'intéresse à la durée de l'attente entre deux émissions successives, on introduit les v.a.

$$Z_i = Y_i - Y_{i-1} \quad (\text{avec } Y_0 = 0)$$

On a donc  $Z_1 = Y_1$ , dont la loi est exponentielle de paramètre  $\alpha$ . Intuitivement, la 1<sup>ère</sup> émission ayant eu lieu, on repart avec  $n-1$  atomes instables et on devrait trouver que  $Z_2$  suit une loi exponentielle de paramètre  $\beta_2 = (n-1)\alpha$ , etc. C'est ce qu'on va voir.

Lemme 4. a) Les v.a.  $Z_1, \dots, Z_n$  sont indépendantes.  
 b)  $Z_i$  suit une loi exponentielle de paramètre  $\beta_i = (n-i+1)\alpha$ .

Dém. Posons  $\vec{Y} = (Y_1, \dots, Y_n)$  et considérons la transformation linéaire  $\vec{W} = A\vec{Y}$  donnée par la matrice

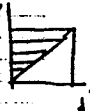
$$A = \begin{bmatrix} n & 0 & 0 & \dots & 0 \\ -(n-1) & n-1 & 0 & \dots & 0 \\ 0 & -(n-2) & n-2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \quad \det A = n!$$

i.e.  $W_i = (n-i+1)(Y_i - Y_{i-1}) = (n-i+1)Z_i \quad (1 \leq i \leq n)$

Notons que  $Y_i = (Y_i - Y_{i-1}) + (Y_{i-1} - Y_{i-2}) + \dots + (Y_2 - Y_1) + (Y_1 - Y_0)$

$$= \sum_{j=1}^i \frac{W_j}{n-j+1}$$

d'où  $\sum_{i=1}^n Y_i = \sum_{i=1}^n \sum_{j=1}^i \frac{W_j}{n-j+1} = \sum_{j=1}^n \sum_{i=j}^n \frac{W_j}{n-j+1} = \sum_{j=1}^n W_j$



Cherchons la densité de la loi de  $\vec{W}$ :

$$\begin{aligned} P[\vec{W} \in B] &= P[\vec{Y} \in A^{-1}(B)] = \int_{A^{-1}(B)} h(\vec{y}) d\vec{y} \\ &= \int_B h(A^{-1}\vec{w}) |\det A^{-1}| d\vec{w} \\ &= \int_B n! \alpha^n e^{-\alpha(\sum_{i=1}^n Y_i)} \frac{1}{n!} dw_1 \dots dw_n \quad (\text{cf. lemme 3'}) \\ &= \int_B \alpha^n e^{-\alpha \sum_{i=1}^n W_i} dw_1 \dots dw_n \end{aligned}$$

si  $B$  est une région dans  $w_1 > 0, \dots, w_n > 0$ . Donc la densité de  $\vec{W}$  est

$$g(w_1, \dots, w_n) = \begin{cases} \prod_{i=1}^n \alpha e^{-\alpha w_i} & \text{si } w_1 > 0, \dots, w_n > 0 \\ 0 & \text{sinon} \end{cases}$$

Autrement dit  $g(w_1, \dots, w_n)$  est le produit de  $n$  densités de lois exponentielles de paramètre  $\alpha$ .  
On en déduit que  $W_1, \dots, W_n$  sont des v.a. indép. de loi exponentielle de paramètre  $\alpha$ .

Comme  $Z_i = \frac{W_i}{n-i+1}$ , le lemme 4 est démontré.

## §7. Moments d'une variable aléatoire.

But: Définir des paramètres et des fonctions associés à une v.a. et qui servent à décrire la loi de probabilité de cette v.a.

### 7.1. Espérance d'une v.a.

1) Soit  $X$  une v.a. (définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  qui prend un nombre fini de valeurs: posons

$$X(\Omega) = \{x_1, x_2, \dots, x_m\} \quad \text{et} \quad P[X=x_i] = p_i \quad 1 \leq i \leq m.$$

Définition: L'espérance <sup>(moyenne)</sup> de  $X$  est le nombre

$$E(X) = \sum_{i=1}^m x_i p_i.$$

Exemple: Considérons  $n$  copies notées de 1 à 6; soit  $m_i$  le nombre de copies qui ont la note  $i$ . On tire une copie "au hasard"; soit  $X$  la note de la copie tirée;  
 $P[X=i] = \frac{m_i}{n} = p_i$  pour  $i=1, 2, \dots, 6$ . La note moyenne est

$$\frac{1}{n} \sum_{i=1}^6 i \cdot m_i = \sum_{i=1}^6 i \cdot p_i = E(X)$$

2) Soit  $X$  une v.a. discrète:

$$X(\Omega) = \{x_1, x_2, \dots\}, \quad P[X=x_i] = p_i, \quad \forall i \geq 1$$

Définition: On dit que  $X$  admet une espérance si  $\sum_{i=1}^{\infty} |x_i| p_i < +\infty$ .  
Cette espérance est alors le nombre

$$E(X) = \sum_{i=1}^{\infty} x_i p_i.$$

NB. Comme la série  $\sum_{i=1}^{\infty} x_i p_i$  est absolument convergente sa valeur ne dépend pas de la numérotation choisie.



3) soit  $X$  une v.a. absolument continue i.e. admettant une densité  $f$ .

Définition: On dit que  $X$  admet une espérance si on a  $\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty$ . Cette espérance est alors le nombre  $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$ .

Exemples

1)  $X$  v.a. de loi binomiale  $(n, p)$  i.e.  $P[X=i] = p_i = \binom{n}{i} p^i (1-p)^{n-i}$   
 $E(X) = \sum_0^n i \binom{n}{i} p^i (1-p)^{n-i} = np \sum_1^n \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} (1-p)^{n-i}$   
 $= np [p + (1-p)]^{n-1} = np$

$E(X) = np$

2)  $X$  v.a. de loi de Poisson de param.  $\lambda$  i.e.  $P[X=k] = \frac{\lambda^k}{k!} e^{-\lambda}$

$E(X) = \sum_0^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_1^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$

$E(X) = \lambda$

3)  $X$  v.a. de loi normale  $(\mu, \sigma)$  i.e.  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$   $y = \frac{x-\mu}{\sigma} \Rightarrow \sigma dy = dx$

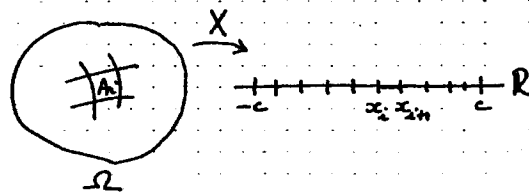
$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\mu + \sigma y) e^{-\frac{1}{2}y^2} dy$

$= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} dy + \frac{\sigma}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} y e^{-\frac{1}{2}y^2} dy}_0$

$= \mu$

$E(X) = \mu$

Remarque: Pour définir  $E(X)$  pour des v.a. qui ne sont ni discrètes ni absolument continues, on peut procéder comme suit:



Supposons  $X$  bornée i.e.  $-c \leq X \leq c$

On décompose  $[-c, c]$  en  $n$  sous-intervalles  $[x_i, x_{i+1}]$  avec  $|x_{i+1} - x_i| \leq \frac{2c}{n}$

On considère les événements  $A_i = [x_i \leq X < x_{i+1}]$ ,  $1 \leq i \leq n$ . Ils forment une partition de  $\Omega$ . On introduit deux v.a.

$X_n^-(\omega) = x_i$  si  $\omega \in A_i$

$X_n^+(\omega) = x_{i+1}$  si  $\omega \in A_i$

On a donc:  $X_n^- \leq X \leq X_n^+$ . Comme  $X_n^-$  et  $X_n^+$  ne prennent qu'un nombre fini de valeurs  $E(X_n^-)$  et  $E(X_n^+)$  sont définies:

$E(X_n^-) = \sum_1^n x_i P(A_i)$ ,  $E(X_n^+) = \sum_1^n x_{i+1} P(A_i)$

On a:  $0 \leq E(X_n^+) - E(X) \leq \frac{2c}{n} \rightarrow 0$ . On définit alors

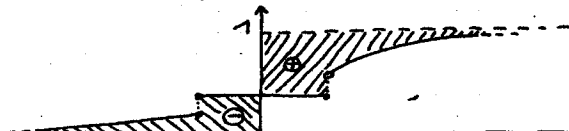
$E(X) = \lim_{n \rightarrow \infty} E(X_n^+) = \lim_{n \rightarrow \infty} E(X_n^-)$

C'est l'idée qui est à la base de l'intégrale de Lebesgue.

Une interprétation géométrique de  $E(X)$ :

Proposition: Soit  $F$  la fonction de répartition de  $X$ ; on a  $E(X) = -\int_{-\infty}^0 F(x) dx + \int_0^{+\infty} [1-F(x)] dx$

(de façon précise: si l'un des membres existe, l'autre aussi et ils sont égaux).



Preuve. Supposons que  $X$  possède une densité  $f$ . on a

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \lim_{a \rightarrow -\infty} \int_a^0 x f(x) dx + \lim_{b \rightarrow +\infty} \int_0^b x f(x) dx$$

On a:

$$\begin{aligned} \int_0^b x f(x) dx &= x F(x) \Big|_0^b - \int_0^b F(x) dx = b F(b) - \int_0^b F(x) dx \\ &= b [F(b) - 1] + \int_0^b [1 - F(x)] dx \end{aligned}$$

$$\text{mais } 0 \in b [1 - F(b)] = b \int_b^{+\infty} f(x) dx \leq \int_b^{+\infty} x f(x) dx \xrightarrow{b \rightarrow \infty} 0$$

car l'intégrale converge! Donc

$$\lim_{b \rightarrow \infty} \int_0^b x f(x) dx = \int_0^{\infty} [1 - F(x)] dx;$$

meilleurs: si la limite existe, l'intégrale de droite existe et

$$\text{on a } \int_0^{\infty} x f(x) dx = \int_0^{\infty} [1 - F(x)] dx.$$

Inversement comme  $b [F(b) - 1] \leq 0$  on a

$$\int_0^b x f(x) dx \leq \int_0^b [1 - F(x)] dx,$$

donc si  $\int_0^{\infty} [1 - F(x)] dx$  existe, il en est de même pour  $\int_0^{\infty} x f(x) dx$ ; on a alors l'égalité cherchée.

Même raisonnement pour  $\int_{-\infty}^0 x f(x) dx$ .

Proposition: Soit  $Y = u(X)$  une v.a. obtenue comme fonction d'une v.a.  $X$ . Alors

$$E(Y) = \begin{cases} \sum_i u(x_i) P[X=x_i] & \text{si } X \text{ est discrète} \\ \int u(x) f(x) dx & \text{si } X \text{ est abs. cont.} \end{cases}$$

(pourvu que la série et l'intégrale soient absolument conv.)

Preuve. Supposons que  $X$  est discrète:  $X(\Omega) = \{x_1, x_2, \dots\}$ .

Alors  $Y$  est aussi discrète:  $Y(\Omega) = \{y_1, y_2, \dots\}$  et

$$P[Y=y_j] = P[X \in \tilde{u}(y_j)] = \sum_{x_i \in \tilde{u}(y_j)} P[X=x_i]$$

Comme la série  $\sum y_j P[Y=y_j]$  est absolument convergente on a:

$$\begin{aligned} E(Y) &= \sum_j y_j P[Y=y_j] = \sum_j y_j \sum_{x_i \in \tilde{u}(y_j)} P[X=x_i] \\ &= \sum_j \sum_{x_i \in \tilde{u}(y_j)} u(x_i) P[X=x_i] = \sum_i u(x_i) P[X=x_i] \end{aligned}$$

On admettra ce résultat dans les autres cas.

cf. ex. ←

Remarque: Le résultat est valable si  $X = (X_1, \dots, X_n)$  est une v.a. vectorielle.

Corollaire: 1)  $E$  est linéaire:  $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$   
2)  $E$  est positive:  $X \leq Y \Rightarrow E(X) \leq E(Y)$ .

Preuve: 1) Cas absolument continu: soit  $h(x, y)$  une densité pour le couple  $(X, Y)$ ; on a par la Prop.

$$\begin{aligned} E(\alpha X + \beta Y) &= \iint (\alpha x + \beta y) h(x, y) dx dy \\ &= \alpha \int dx \cdot x \underbrace{\int h(x, y) dy}_{f(x)} + \beta \int dy \cdot y \underbrace{\int h(x, y) dx}_{g(y)} \\ &= \alpha E(X) + \beta E(Y) \end{aligned}$$

2) Découle de 1) et de  $Z \geq 0 \Rightarrow E(Z) \geq 0$  qui est clair!

Corollaire:  $E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n)$ .

Exemples. 1) Si  $X$  est une v.a. binomiale  $(n, p)$  on peut écrire  $X = X_1 + \dots + X_n$  où  $X_i$  prend les valeurs 0 et 1 avec les probabilités  $1-p$  et  $p$  respectivement. On a  $E(X_i) = p$ . Donc  $E(X) = np$ .

2)  $X$  v.a.  $\chi^2$  à  $n$  d.l.; alors  $E(X) = n$ . (cf. ci-contre).

Remarque 1) le corollaire ci-dessus est obtenu sans hypothèse sur le couple  $(X, Y)$ .

2) En général  $E(XY) \neq E(X)E(Y)$  : on a vu que si  $X$  est normale réduite, alors  $E(X)=0$ ,  $E(X^2)=1$ .  
Mais lorsque  $X$  et  $Y$  sont indépendants :

Proposition : si  $X$  et  $Y$  sont indépendants on a  
 $E(XY) = E(X)E(Y)$

Preuve : si  $X$  et  $Y$  sont indépendants et admettent une densité  $f(x)$  et  $g(y)$ , alors  $(X, Y)$  admet pour densité  $h(x, y) = f(x)g(y)$  et on a

$$E(XY) = \iint xy f(x)g(y) dx dy = \int x f(x) dx \int y g(y) dy = E(X)E(Y)$$

Même chose dans le cas discret.

← Remarque :

7.2. Variance d'une v.a.

Soit  $X$  une v.a. telle que  $\mu = E(X)$  existe. Comme  $\mu$  indique le "centre de gravité" de  $X$ , il est naturel d'écrire

$$X = \mu + X - \mu$$

et de considérer  $X - \mu$  comme la fluctuation de  $X$  autour de son espérance. On évitera les compensations en considérant  $(X - \mu)^2$ .

Notons que si  $E(X^2) < \infty$  alors  $E(X)$  existe car  $|X| \leq X^2 + 1$  et  $E[(X - \mu)^2] < \infty$  car  $(X - \mu)^2 \leq 2X^2 + 2\mu^2$ .

Définition : soit  $X$  une v.a. telle que  $E(X^2)$  existe ; posons  $E(X) = \mu$ . La variance de  $X$  est le nombre

$$\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$$

L'écart type (ou écart quadratique moyen) de  $X$  est

$$\sigma = \sqrt{\text{Var}(X)}$$

Exemple : 1)  $X$  v.a. de Poisson de paramètre  $\lambda$  ; on a vu  $E(X) = \lambda$

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} (j+1) \frac{\lambda^j}{j!} e^{-\lambda} \\ &= \lambda \underbrace{\sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda}}_{E(X)=\lambda} + \lambda \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda}}_1 = \lambda^2 + \lambda \end{aligned}$$

d'où  $\text{Var}(X) = E(X^2) - E(X)^2 = \lambda$ .

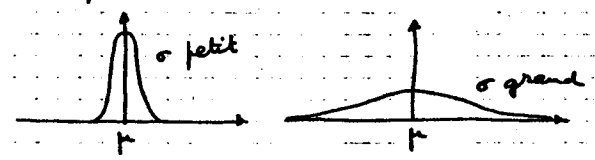
2)  $X$  v.a. normale réduite ; on a vu que  $E(X) = 0$  et que  $E(X^2) = 1$ . on a donc  $\text{Var}(X) = 1$ .

Si  $X$  est une v.a. normale de paramètres  $\mu$  et  $\sigma$  on a vu que  $E(X) = \mu$ . Comme la v.a.  $X^* = \frac{X - \mu}{\sigma}$  est alors normale réduite et on a

$$X - E(X) = X - \mu = \sigma X^*$$

d'où  $\text{Var}(X) = E[(X - E(X))^2] = \sigma^2 \underbrace{E(X^{*2})}_1 = \sigma^2$

[En résumé : si  $X$  est une v.a. normale de paramètres  $\mu$  et  $\sigma$  alors  $E(X) = \mu$  et  $\text{Var}(X) = \sigma^2$ .



Définition : si  $X$  est une v.a. d'espérance  $\mu$  et de variance  $\sigma^2$  alors

$$X^* = \frac{X - \mu}{\sigma}$$

est la variable aléatoire réduite (ou normalisée) associée à  $X$ . On voit que  $X^*$  a une espérance 0 et une variance 1.

Comment calculer  $\text{Var}(X_1 + \dots + X_n)$  ?

Commençons par  $\text{Var}(X+Y)$  ; par déf. on a

$$\begin{aligned} \text{Var}(X+Y) &= E[(X+Y - E(X+Y))^2] = E[(X - E(X) + Y - E(Y))^2] \\ &= E[(X - \mu_X)^2 + 2(X - \mu_X)(Y - \mu_Y) + (Y - \mu_Y)^2] \quad \text{où } \mu_X = E(X) \\ &\quad \mu_Y = E(Y) \end{aligned}$$

donc  $\text{Var}(X+Y) = \text{Var}(X) + 2 E[(X-\mu_X)(Y-\mu_Y)] + \text{Var}(Y)$

Définition: on appelle covariance de X et Y le nombre  

$$\text{Cov}(X, Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X \mu_Y$$
  
 (la dernière égalité par linéarité de E).

Notons que si X et Y sont indépendants, alors  $\text{Cov}(X, Y) = 0$ .

On voit donc que

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$$

et que:

$$X, Y \text{ indép.} \Rightarrow \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

De façon générale:

Proposition: 1)  $\text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j)$   
 2)  $X_1, \dots, X_n$  indép.  $\Rightarrow \text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i)$

Preuve: Comme ci-dessus en utilisant  $(a_1 + \dots + a_n)^2 = \sum a_i^2 + 2 \sum_{i < j} a_i a_j$ .

Exemple: 1) soit X une v.a. binomiale (n, p). Ecrivons  
 $X = \sum_{i=1}^n X_i$  où chaque  $X_i$  est une v.a. de Bernoulli de paramètre p  
 et  $X_1, \dots, X_n$  sont indépendants. On a  $E(X_i) = p$ ,  $X_i = X_i^2$   
 donc  $\text{Var}(X_i) = p - p^2 = pq$ . Donc  $\text{Var}(X) = npq$ .

On voit donc que la v.a. réduite associée à X est  $X^* = \frac{X - np}{\sqrt{npq}}$   
 souvenons-nous du thm. de De Moivre-Laplace:

$$P\left[\frac{X - np}{\sqrt{npq}} < \infty\right] \xrightarrow{n \rightarrow \infty} \Phi(x)$$

i.e. si n est grand, binomiale réduite  $\approx$  normale réduite.

2) On a vu que  $X, Y$  indép.  $\Rightarrow \text{Cov}(X, Y) = 0$ , mais la réciproque est fautive:

si X v.a. uniforme sur  $[-1, 1]$ .  $E(X) = 0$   
 et  $Y = X^2$

$$\text{alors } \text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X^3) = \int_{-1}^1 x^3 \frac{dx}{2} = 0,$$

bien que X et Y ne soient pas indépendants!

Mais alors que mesure  $\text{Cov}(X, Y)$ ?

soient  $X^* = \frac{X - \mu_X}{\sigma_X}$  et  $Y^* = \frac{Y - \mu_Y}{\sigma_Y}$  la v.a. réduites associées à X et Y

Définition: le coefficient de corrélation de X et Y est le nombre

$$\rho(X, Y) = \text{Cov}(X^*, Y^*) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Proposition: 1)  $|\rho(X, Y)| \leq 1$

2)  $\rho(X, Y) = \pm 1 \Rightarrow Y = aX + b$  (avec une prob. 1)

Preuve: 1) on a

$$0 \leq \text{Var}(X^* \pm Y^*) = \text{Var}(X^*) \pm 2 \text{Cov}(X^*, Y^*) + \text{Var}(Y^*) = 2[1 \pm \rho(X, Y)]$$

d'où  $1 \pm \rho(X, Y) \geq 0$  i.e.  $-1 \leq \rho(X, Y) \leq 1$ .

2) si  $\rho(X, Y) = 1$  on a  $\text{Var}(X^* - Y^*) = 0$  donc  $X^* = Y^*$  (avec prob. 1)  
 d'où  $Y = \frac{\sigma_Y}{\sigma_X} X + \mu_Y - \frac{\sigma_Y}{\sigma_X} \mu_X$ . Idem si  $\rho(X, Y) = -1$ .

Remarque

1) Du point 1) on déduit:  $|\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y$ : c'est un cas particulier de l'inégalité de Schwarz:  $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$

2)  $\rho(X, Y)$  est une mesure (standardisée) du degré de dépendance linéaire entre X et Y. De façon précise:

parmi toutes les v.a.  $Y_p = aX + b$ , celle pour laquelle  $E[(Y - Y_p)^2]$  est minimum est

$$Y_p = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (X - \mu_X) \quad \rho = \rho(X, Y)$$

et alors

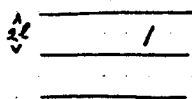
$$E[(Y - Y_p)^2] = \sigma_Y^2 (1 - \rho^2)$$

(l'équation ci-dessus définit la droite de régression de Y en X)

Exemple (où on voit qu'une dépense peut être profitable).

On reprend l'expérience de l'aiguille de Buffon.

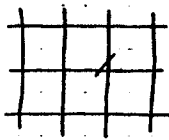
a) 1<sup>ère</sup> procédure: si l'aiguille est de longueur  $l$  la probabilité qu'elle coupe une parallèle est  $\frac{1}{\pi}$ .



On veut estimer  $\frac{1}{\pi}$  en lançant 200 fois l'aiguille:

$$\text{Estimateur}_1 = \frac{\text{nb. de succès}}{200}$$

b) 2<sup>ème</sup> procédure:



On jette 100 fois l'aiguille et on compte les intersections avec les horizontales et les verticales:

$$\text{Estimateur}_2 = \frac{\text{nb. de succès}}{200}$$

Questions: 1) Les deux procédures sont-elles équivalentes?  
2) Sinon, quelle est la meilleure?

1) Elles ne sont pas équivalentes car les événements  $A =$  "l'aiguille coupe une horizontale" et  $B =$  "l'aiguille coupe une verticale" ne sont pas indépendants (cf §4. n°23):

$$P(A) = P(B) = \frac{1}{\pi} \quad \text{alors que} \quad P(A \cap B) = \frac{1}{4\pi}$$

2) Examinons la deux procédures:

a) 1<sup>ère</sup> procédure: soit  $X_i$  l'indicatrice de la  $i$ <sup>ème</sup> expérience  
 $P[X_i=1] = \frac{1}{\pi}$ ,  $P[X_i=0] = 1 - \frac{1}{\pi}$ ;  $E(X_i) = \frac{1}{\pi}$ ,  $\text{Var}(X_i) = \frac{1}{\pi} \left(1 - \frac{1}{\pi}\right)$ .  
Si on fait  $n$  expériences (indépendantes) alors

$$\text{Estimateur}_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad ; \quad E(\bar{X}) = \frac{1}{\pi}$$

$$\text{Var}(\bar{X}) = \frac{1}{n} \cdot \frac{1}{\pi} \left(1 - \frac{1}{\pi}\right) = \frac{1}{n} \cdot \frac{\pi-1}{\pi^2} = \frac{0,217}{n}$$

b) 2<sup>ème</sup> procédure: Considérons la v.a.  
$$U_i = \begin{cases} 1 & \text{si l'aiguille coupe une horizontale à la } i\text{ème exp.} \\ 0 & \text{sinon.} \end{cases}$$
  
$$V_i = \begin{cases} 1 & \text{si l'aiguille coupe une verticale à la } i\text{ème exp.} \\ 0 & \text{sinon.} \end{cases}$$

Loi du couple  $(U_i, V_i)$ :  $P[U_i=1, V_i=1] = \frac{1}{4\pi}$

Comme  $P[U_i=1] = \frac{1}{\pi} = P[U_i=1, V_i=1] + P[U_i=1, V_i=0]$

on a:

$$P[U_i=1, V_i=0] = \frac{3}{4\pi}$$

Par symétrie:  $P[U_i=0, V_i=1] = \frac{3}{4\pi}$

Enfin  $P[U_i=0, V_i=0] = 1 - \frac{1}{4\pi} - \frac{3}{4\pi} - \frac{3}{4\pi} = 1 - \frac{7}{4\pi}$

Si on jette  $n$  fois l'aiguille sur le réseau alors

$$\text{Estimateur}_2 = T = \frac{1}{2n} \sum_{i=1}^n (U_i + V_i) = \frac{1}{2} (\bar{U} + \bar{V})$$

on a:  $E(T) = \frac{1}{2} E(\bar{U}) + \frac{1}{2} E(\bar{V}) = \frac{1}{\pi}$

$$\text{Var}(T) = \frac{1}{4} \text{Var}(\bar{U} + \bar{V}) = \frac{1}{4} \text{Var}(\bar{U}) + \frac{1}{4} \text{Var}(\bar{V}) + \frac{1}{2} \text{Cov}(\bar{U}, \bar{V})$$

où:  $\text{Var}(\bar{U}) = \text{Var}(\bar{V}) = \frac{1}{n} \cdot \frac{\pi-1}{\pi^2}$

$$\begin{aligned} \text{Cov}(\bar{U}, \bar{V}) &= \text{Cov}\left(\frac{1}{n} \sum_{i=1}^n U_i, \frac{1}{n} \sum_{j=1}^n V_j\right) = \frac{1}{n^2} \left\{ E\left(\sum_{i=1}^n U_i \cdot \sum_{j=1}^n V_j\right) - E\left(\sum_{i=1}^n U_i\right) E\left(\sum_{j=1}^n V_j\right) \right\} \\ &= \frac{1}{n^2} \sum_{i,j} \text{Cov}(U_i, V_j) \end{aligned}$$

$$\text{Cov}(U_i, V_j) = \begin{cases} 0 & \text{si } i \neq j \text{ ( indép. des exp. )} \\ E(U_i V_i) - E(U_i) E(V_i) = \frac{1}{4\pi} - \frac{1}{\pi^2} & \text{si } i=j \\ = -\frac{4-\pi}{4\pi^2} \end{cases}$$

Finalement:

$$\begin{aligned} \text{Var}(T) &= 2 \cdot \frac{1}{4} \cdot \frac{1}{n} \cdot \frac{\pi-1}{\pi^2} - \frac{1}{2} \cdot \frac{1}{n^2} \cdot n \cdot \frac{4-\pi}{4\pi^2} \\ &= \frac{1}{n} \cdot \frac{5\pi-8}{8\pi^2} = \frac{0,0976}{n} \end{aligned}$$

Amor avec  $n=200$  et  $m=100$

$$\text{Var}(\text{Estimateur}_1) = \frac{0,217}{200} = 0,001085$$

$$\text{Var}(\text{Estimateur}_2) = \frac{0,0976}{100} = 0,000976$$

donc l'Estimateur<sub>2</sub> est meilleur.

Remarque: De  $0,000976 = \frac{0,217}{n}$  on tire  $n = 222$ .

Donc 100 mesures sur la grille contiennent autant d'information que 222 mesures sur un système de parallèles. C'est là un exemple d'estimation par des variables anticorrélées: si des variables ont une covariance négative, la variance de leur moyenne est plus petite que la variance de chacune, les variations se compensant.

[cf. Hammersley-Handscomb, Monte-Carlo Methods, Methuen]

### 7.3. Fonction génératrice d'une v.a. à valeurs dans $\mathbb{N}$ .

Définition. Soit  $X$  une v.a. à valeurs dans  $\mathbb{N}$ ; posons  $p_k = P[X=k]$  pour tout  $k=0,1,2,\dots$ . La fonction génératrice de  $X$  est la fonction  $G_X$  de  $t \in \mathbb{R}$  définie par

$$G_X(t) = E(t^X) = \sum_{k \geq 0} p_k t^k$$

Remarques: 1) Comme  $\sum_{k \geq 0} p_k = 1$ , la série  $\sum p_k t^k$  est normalement convergente dans  $[-1,1]$ . Mais son domaine de convergence peut être plus grand: si  $X(\Omega)$  est fini,  $G_X$  est un polynôme!

2)  $G_X$  ne dépend que de la loi de probabilité de  $X$  et la caractérise car on a

$$p_k = P[X=k] = \frac{1}{k!} G_X^{(k)}(0)$$

Donc si  $X$  et  $Y$  sont des v.a. à valeurs dans  $\mathbb{N}$ , on a:

$$G_X = G_Y \Leftrightarrow X \text{ et } Y \text{ ont la même loi de probabilité.}$$

Proposition: 1) Si  $X$  et  $Y$  sont indépendantes  $G_{X+Y}(t) = G_X(t)G_Y(t)$ ,  
 2) Plus généralement si  $X_1, \dots, X_n$  sont indép.  $G_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n G_{X_i}(t)$ .

Preuve: 1) Si  $X$  et  $Y$  sont indép. alors les v.a.  $t^X$  et  $t^Y$  sont aussi indépendantes, donc  $E(t^X t^Y) = E(t^X)E(t^Y)$ ; mais  $E(t^X t^Y) = E(t^{X+Y}) = G_{X+Y}(t)$ .

2) Induction.

Remarque: L'introduction des fonctions génératrices permet de traiter des problèmes "discrets" de probabilité par des méthodes analytiques. En voici un exemple:

Problème: Peut-on piper deux dés de telle façon que la somme des points suive une loi uniforme?

Introduisons les deux v.a.  $X$  et  $Y$  par:

$$X = \text{résultat du 1er dé}; \quad P[X=i] = p_i \quad (1 \leq i \leq 6)$$

$$Y = \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---}; \quad P[Y=i] = q_i \quad (1 \leq i \leq 6)$$

$S = X+Y$  prend les 11 valeurs  $2, 3, \dots, 12$ . Tout revient à savoir s'il est possible de trouver  $p_i, q_i$  ( $1 \leq i \leq 6$ ) de telle façon que  $P[S=k] = \frac{1}{11}$  pour tout  $k, 2 \leq k \leq 12$ ?

Il y a 10 équations indépendantes (car  $\sum p_i = 1 = \sum q_i$ ) et on a les 11 équations ( $X$  et  $Y$  sont indépendantes):

$$\begin{cases} P[S=2] = \frac{1}{11} = p_1 q_1 \\ P[S=3] = \frac{1}{11} = p_1 q_2 + p_2 q_1 \\ \dots \\ P[S=12] = \frac{1}{11} = p_6 q_6 \end{cases}$$

Le système possède-t-il une solution??

Les fonctions génératrices permettent de résoudre ce problème de façon très élégante:

On a:  $G_X(t) = p_1 t + \dots + p_n t^n = t[p_1 + \dots + p_n t^{n-1}]$

$$G_Y(t) = q_1 t + \dots + q_m t^m = t[q_1 + \dots + q_m t^{m-1}]$$

on voit:  $G_{X+Y}(t) = \frac{1}{n} t^2 + \dots + \frac{1}{n} t^n = \frac{t^2}{n} [1 + t + \dots + t^{n-2}]$

Comme  $G_{X+Y}(t) = G_X(t)G_Y(t)$  on doit avoir

$$[p_1 + \dots + p_n t^{n-1}][q_1 + \dots + q_m t^{m-1}] = \frac{1}{n} [1 + t + \dots + t^{n-2}] = \frac{1}{n} \frac{t^n - 1}{t - 1}$$

Le polynôme de droite n'a pas de racine réelle (racines  $n^{\text{es}}$  de 1,  $n$ )  
alors que les polynômes de gauche en possèdent, car ils  
sont de degré impair!

Le problème posé ne possède pas de solution: on ne peut pas  
lancer deux dés pour que la somme des points soit de loi uniforme.

Exemples:

1) V.a. de Bernoulli de paramètre  $p$  i.e.  $P[X=0]=q, P[X=1]=p, p+q=1$ .

Alors  $G_X(t) = q + pt$ .

2) V.a. binomiale de paramètres  $n, p$  i.e.  $P[X=k] = \binom{n}{k} p^k q^{n-k}$

Alors  $G_X(t) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} t^k = (q+pt)^n$

NB. On peut obtenir ce résultat autrement: on écrit  
 $X = X_1 + \dots + X_n$  où les  $X_i$  sont des v.a. de Bernoulli de param.  $p$   
et indépendantes. Alors  $G_X(t) = \prod_{i=1}^n G_{X_i}(t) = (q+pt)^n$ .

3) V.a. de Poisson de paramètre  $\lambda$ :  $P[X=k] = \frac{\lambda^k}{k!} e^{-\lambda}$

Alors  $G_X(t) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} t^k = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}$

NB. On retrouve ainsi facilement le fait qu'une somme  
de v.a. indépendantes de paramètre  $\lambda_i$  est une v.a.  
de Poisson de paramètre  $\sum \lambda_i$ .

4) V.a. géométrique de paramètre  $p$ :  $P[X=k] = pq^{k-1}$  ( $\forall k \geq 1$ )

Alors

$$G_X(t) = \sum_{k=1}^{\infty} pq^{k-1} t^k = \frac{pt}{1-qt}$$

Toujours par le même raisonnement, on en déduit (cf. série 19)  
que la f.g. d'une v.a. binomiale négative de paramètres  
 $p$  et  $r$  est

$$G_{Y_r}(t) = \left( \frac{pt}{1-qt} \right)^r$$

Il est clair que la f.g.  $G_X$  doit fournir tous les paramètres  
de la loi de  $X$ . En effet:

Proposition: 1) Si  $G_X'$  a une limite à gauche en  $t=1$  alors  
 $E(X)$  est finie et

$$E(X) = G_X'(1-)$$

2) Si  $G_X''$  a une limite à gauche en  $t=1$  alors  $X$  possède  
une variance et

$$\text{Var}(X) = G_X''(1-) + G_X'(1-) - G_X'(1-)^2$$

NB. Les réciproques sont vraies.

Preuve: 1) Pour tout  $t$ ,  $|t| < 1$ , on a  $G_X'(t) = \sum_{k=1}^{\infty} k p_k t^{k-1}$   
si  $G_X'(1-) < +\infty$  alors  $G_X'(t)$  est bornée sur  $]0, 1[$ .  
Donc  $\forall n \geq 0$ ,  $\sum_{k=1}^n k p_k t^{k-1} \leq M$ ; donc  $(t \rightarrow 1): \sum_{k=1}^n k p_k \leq M$ ;  
donc  $\sum_{k=1}^{\infty} k p_k \leq M$  i.e.  $E(X)$  existe ( $< +\infty$ ). L'égalité  
est claire.

2) Pour tout  $t$ ,  $|t| < 1$ , on a  $G_X''(t) = \sum_{k=2}^{\infty} k(k-1) p_k t^{k-2}$   
Comme ci-dessus on déduit de  $G_X''(1-) < +\infty$  que  
 $\sum k(k-1) p_k$  converge et que  $E(X(X-1)) = G_X''(1-)$ .  
Alors  $\text{Var}(X) = E(X^2) - E(X)^2 = G_X''(1-) + G_X'(1-) - G_X'(1-)^2$ .

Théorème de continuité: Soient  $X_1, X_2, \dots$  une suite de v.a. à valeurs dans  $\mathbb{N}$ ,  $G_n(t) = \sum_{k \geq 0} p_{n,k} t^k$  la f.g. de  $X_n$  pour tout  $n \geq 1$ . Les affirmations suivantes sont équivalentes:

(1)  $\forall k \geq 0, \lim_{n \rightarrow \infty} p_{n,k} = p_k$  existe.

(2)  $\lim_{n \rightarrow \infty} G_n(t) = G(t)$  existe pour tout  $t, 0 < t < 1$ .

Alors  $G(t) = \sum_{k \geq 0} p_k t^k$ .

NB. Il n'est pas dit que  $\sum p_k = 1$ , donc que  $G(t)$  est la fonction génératrice d'une v.a.

Preuve: (1)  $\rightarrow$  (2). Notons tout d'abord que  $0 < p_k \leq 1$

et que  $\sum_{k=0}^m p_{n,k} \rightarrow \sum_{k=0}^m p_k$  ( $\forall m$ ) d'où  $\sum_{k=0}^m p_k \leq 1$  ( $\forall m$ )

et  $\sum_{k=0}^{\infty} p_k \leq 1$ .

Définissons  $G(t) = \sum_{k \geq 0} p_k t^k$  (convergence normale sur  $[t, 1]$ ).

quel que soit  $t, 0 < t < 1$ , on a:

$$\begin{aligned} |G_n(t) - G(t)| &= \left| \sum_{k \geq 0} (p_{n,k} - p_k) t^k \right| \leq \sum_{k=0}^m |p_{n,k} - p_k| + \sum_{k=m+1}^{\infty} t^k \\ &\leq \sum_{k=0}^m |p_{n,k} - p_k| + \frac{t^{m+1}}{1-t} \end{aligned}$$

soit, on choisit  $m$  assez grand pour que  $\frac{t^{m+1}}{1-t} < \varepsilon$  ( $t < 1$ ).

puis on choisit  $n_0$  tel que  $\forall n \geq n_0, \sum_{k=0}^m |p_{n,k} - p_k| < \varepsilon$ .

Alors  $|G_n(t) - G(t)| < 2\varepsilon$  dès que  $n \geq n_0$ .

(2)  $\rightarrow$  (1). (i) Montrons que  $(p_{n,0})$  converge vers  $G(0)$ .

Tout d'abord comme  $G_n(t) = \sum p_{n,k} t^k$  est croissante en  $t$ , il en est de même pour  $G(t)$ , donc  $G(0) = \lim_{t \rightarrow 0} G(t)$  existe.

Pour tout  $t, 0 < t < 1$ , on a:

$$p_{n,0} = G_n(0) \leq G_n(t) = p_{n,0} + \sum_{k \geq 1} p_{n,k} t^k \leq p_{n,0} + \frac{t}{1-t}$$

d'où  $G_n(t) - \frac{t}{1-t} \leq p_{n,0} \leq G_n(t)$

Ainsi toute valeur d'adhérence de la suite  $(p_{n,0})$  appartient à l'intervalle  $[G(t) - \frac{t}{1-t}, G(t)]$  et ceci pour tout  $t > 0$ .  
Donc  $G(0)$  est la seule valeur d'adhérence. Mais alors on a  $p_{n,0} \rightarrow G(0)$ .

(ii) Passons à  $(p_{n,1})$ . Considérons  $H(t) = \frac{G(t) - G(0)}{t}$

et  $H_n(t) = \frac{G_n(t) - G_n(0)}{t} = \sum_{k \geq 1} p_{n,k} t^{k-1}$ .

On a  $H_n(t) \rightarrow H(t)$  pour tout  $t, 0 < t < 1$ .

On en déduit comme en (i) que  $H(0) = G'(0)$  existe et que  $p_{n,1} \rightarrow G'(0)$ .

(iii) Induction.

Exemple: Généralisation du théorème de Poisson.

Soit  $X_1, X_2, \dots, X_n$  des v.a. indépendantes de Bernoulli,  $X_i$  étant de paramètre  $p_i$  ( $\forall i=1, \dots, n$ ).

Soit  $S_n = X_1 + \dots + X_n$ . On a  $G_{S_n}(t) = \prod_{i=1}^n (q_i + p_i t)$ .

Supposons que lorsque  $n \rightarrow \infty$ , les  $p_i = p_i(n)$  vérifient

$$\sum_{i=1}^n p_i = \lambda \text{ est et } \max_{i \leq n} p_i \rightarrow 0$$

Alors  $\log G_{S_n}(t) = \sum_{i=1}^n \log(q_i + p_i t) = \sum_{i=1}^n \log(1 - p_i(1-t))$

Comme on a  $\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots = -x - o_2$  avec  $o_2 \rightarrow 0$  si  $x \rightarrow 0$  on peut écrire

$$\begin{aligned} \log G_{S_n}(t) &= -(1-t) \sum_{i=1}^n [p_i + p_i o_i] \\ &= -(1-t)\lambda - (1-t) \sum_{i=1}^n p_i o_i \end{aligned}$$

où  $\sum_{i=1}^n p_i o_i \leq \lambda \max_{i \leq n} o_i \rightarrow 0$  si  $n \rightarrow \infty$ . D'où on tire

$$G_{S_n}(t) \rightarrow e^{\lambda(t-1)}$$

i.e.  $S_n$  "converge en loi" vers une v.a. de Poisson de paramètre  $\lambda$ .



Par exemple, si  $p_i$  est la probabilité <sup>annuelle</sup> de sinistre pour le  $i^{\text{ème}}$  risque, le nombre total de sinistres pendant 1 an est (approx.) donné par une loi de Poisson dont le paramètre est  $\lambda = \sum p_i = \text{nb. moyen de sinistres par an}$  [on suppose que les  $p_i$  sont tous petits, et qu'un seul sinistre par assuré peut se produire et que les risques sont indépendants].

Les fonctions génératrices permettent aussi d'étudier la somme d'un nombre aléatoire de v.a. aléatoires indépendantes et de même loi.

Soient  $X_1, X_2, \dots$  des v.a. indépendantes et de même loi; notons  $G_X$  la f.g. génératrice de cette loi. Soit  $N$  une v.a. indépendante des  $X_i$ ; notons  $G_N$  sa f.g. Poisson

$$S_N = X_1 + \dots + X_N$$

(i.e.  $\forall \omega \in \Omega$   $S_N(\omega) = X_1(\omega) + \dots + X_{N(\omega)}(\omega)$ ).

Exemples: 1)  $N$  est le nombre de sinistres dans une année,  $X_i$  le coût du  $i^{\text{ème}}$  sinistre,  $S_N$  le total des prestations versées.

2) Cf. séries n°15 et n°20.

Proposition:  $G_{S_N} = G_N \circ G_X$ .

Preuve: on a pour tout entier  $j \geq 0$ :

$$\begin{aligned} P[S_N = j] &= \sum_{m=0}^{\infty} P[S_N = j | N=m] P[N=m] \\ &= \sum_{m=0}^{\infty} P[N=m] P[X_1 + \dots + X_m = j] \end{aligned}$$

donc:

$$\begin{aligned} G_{S_N}(t) &= \sum_{j=0}^{\infty} P[S_N = j] t^j = \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} P[N=m] P[X_1 + \dots + X_m = j] t^j \\ &= \sum_{m=0}^{\infty} P[N=m] \sum_{j=0}^{\infty} P[X_1 + \dots + X_m = j] t^j \end{aligned}$$

$$\begin{aligned} &= \sum_{m=0}^{\infty} P[N=m] G_{X_1 + \dots + X_m}(t) = \sum_{m=0}^{\infty} P[N=m] G_X(t)^m \\ &= G_N(G_X(t)). \end{aligned}$$

Corollaire: 1)  $E(S_N) = E(N)E(X)$ .  
2)  $\text{Var}(S_N) = \text{Var}(N)E(X)^2 + E(N)\text{Var}(X)$ .

Preuve: 1)  $E(S_N) = G'_{S_N}(1)$ ; or  $G'_{S_N}(t) = G'_N(G_X(t)) G'_X(t)$   
d'où  $G'_{S_N}(1) = G'_N(1) G'_X(1)$  car  $G_X(1) = 1$ .

2)  $G''(t) = G''_N(G_X(t)) G'_X(t)^2 + G'_N(G_X(t)) G''_X(t)$  d'où:  
 $G''_{S_N}(1) = G''_N(1) E(X)^2 + E(N) G''_X(1)$ . on a vu que

$$\begin{aligned} \text{Var}(S_N) &= G''_{S_N}(1) + G'_{S_N}(1) - G'_{S_N}(1)^2 = G''_N(1) E(X)^2 + E(N) G''_X(1) + E(S_N) - E(S_N)^2 \\ &= (\text{Var}(N) - E(N) + E(N)^2) E(X)^2 + E(N) (\text{Var}(X) - E(X) + E(X)^2) + E(S_N) - E(S_N)^2 \\ &= \text{Var}(N) E(X)^2 + E(N) \text{Var}(X) \text{ en utilisant 1).} \end{aligned}$$

En particulier:

1) si les  $X_i$  sont des v.a. de Bernoulli de paramètre  $p$ , on a  $G_X(t) = q + pt$  avec  $q = 1 - p$ , donc  $G_{S_N}(t) = G_N(q + pt)$ .

2) si  $N$  est une v.a. de Poisson de paramètre  $\lambda$  on a

$$G_{S_N}(t) = e^{\lambda[G_X(t) - 1]}$$

Définition: Lorsque  $N$  est une v.a. de Poisson, la loi de  $S_N$  est appelée loi de Poisson composée.

Exemple: si  $X$  est une v.a. de Bernoulli de paramètre  $p$  on a  $G_X(t) = e^{\lambda[q + pt - 1]} = e^{\lambda p[t - 1]}$ , donc  $S_N$  est une v.a. de Poisson de paramètre  $\lambda p$ . [cf séries n°15 et n°20].

Remarque: Pour une v.a.  $Z$  à valeurs dans  $\mathbb{N}$  on a la caractérisation suivante (P. Lévy): les affirmations suivantes sont équivalentes

- $Z$  suit une loi de Poisson composée
- pour tout  $n > 0$ , on peut écrire  $Z$  comme une somme de  $n$  v.a. indépendantes de même loi. [cf. Feller I p.200].

7.4. Application: Processus de ramification

On considère une population formée d'individus identiques. Chacun d'eux peut donner naissance à  $k$  nouveaux individus avec une probabilité  $p_k$  ( $k=0,1,2,\dots$ ); on suppose  $p_0 > 0, 1$ .  
 la génération initiale - d'ordre 0 - compte un seul individu; la  $n+1$  ème génération est formée des descendants directs des individus de la  $n$  ème génération. A chaque génération les individus agissent indépendamment des autres. On s'intéresse à la taille de générations successives, à la probabilité d'extinction, au nombre total d'individus produits, etc.

Modèle: Le processus est décrit par la suite  $(X_n)$  où  $X_n$  est le nombre d'individus de la  $n$  ème génération.

Exemples:

1) Survivance des noms de famille: les individus sont les hommes et on ne tient compte que des fils qu'ils peuvent avoir (!). Ce problème fut étudié d'abord par Galton et Watson (1874), puis par Steffenson (1930). D'où le nom de processus de Galton-Watson donné aussi aux processus de ramification.  
 [cf. Harris, T. - The Theory of Branching Processes - Springer]

2) Multiplicateur d'électrons: Pour amplifier un faible courant d'électrons, on dresse sur leur trajectoire une série de plaques. Chaque électron, lorsqu'il frappe la  $i$  ème plaque, lui arrache un nombre aléatoire de nouveaux électrons, qui à leur tour...

3) Réaction de fission nucléaire: Un noyau est détruit par un choc avec un neutron; d'où un nombre aléatoire de nouveaux neutrons, etc. Il y a explosion si la taille de générations successives augmente indéfiniment. Schroedinger a utilisé cette description (1945).

4) Survivance de gènes mutants: cf. Karlin et Feller.

5) Fila d'attente: lorsqu'un client arrive à un guichet de poste (par exemple) et qu'il est immédiatement servi on dit qu'il continue la génération initiale. Ses descendants directs sont tous les clients qui arrivent pendant qu'il est servi. le processus continue tant que la file d'attente dure.

a) Fonction génératrice de  $X_n$

Notons  $G$  la f.g. de  $X_1$  i.e.  $G(t) = \sum_{k=0}^{\infty} p_k t^k$   
 et  $G_n$  la f.g. de  $X_n$ . On peut écrire

$$X_n = \sum_{i=1}^{X_{n-1}} Y_i$$

où  $Y_i$  est le nombre de descendants du  $i$  ème individu de la  $(n-1)$  ème génération; chaque  $Y_i$  a la même loi que  $X_1$  et les  $Y_i$  sont indépendantes. Par la Prop. page 104 on a

$$G_n = G_{n-1} \circ G$$

d'où:

$$G_n = G \circ G \circ \dots \circ G \quad (n \text{ fois})$$

On en déduit que (cf. Corollaire p. 105):

$$E(X_n) = G_n'(1) = E(X_{n-1})E(X_1) = \dots = E(X_1)^n$$

Pourvu  $E(X_1) = \mu$ ; on a donc  $E(X_n) = \mu^n$ .

Si  $\mu > 1$  on a  $E(X_n) \rightarrow +\infty$ ; si  $\mu < 1$  on a  $E(X_n) \rightarrow 0$ ; dans ce dernier cas le processus cessera tôt ou tard.

Dans l'autre cas quelle est la probabilité qu'il s'arrête? Que se passe-t-il si  $\mu = 1$ ?

b) Probabilité d'extinction.

Posons  $d_n = G_n(0) = P[X_n = 0]$ .

On a  $d_{n+1} = G_{n+1}(0) = G(G_n(0)) = G(d_n)$  et  $G$  est une fonction strictement croissante [car  $G(t) = p_0 + p_1 t + p_2 t^2 + \dots$  avec  $0 < p_0 < 1$ ] donc on a:

$$0 < d_1 = p_0 < d_2 < d_3 < \dots \leq 1$$

Donc  $(d_n)$  converge. Soit  $d = \lim_{n \rightarrow \infty} d_n$ . Comme  $G$  est continue sur  $[0,1]$  (la série est normalement convergente sur  $[0,1]$ ) on a :

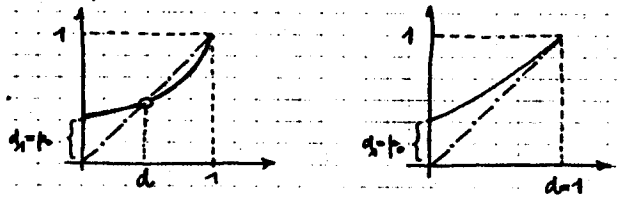
$$G(d) = G(\lim_{n \rightarrow \infty} d_n) = \lim_{n \rightarrow \infty} G(d_n) = \lim_{n \rightarrow \infty} d_{n+1} = d$$

i.e.  $d$  est solution de l'équation  $G(t) = t$ .

De plus si  $t_0$  vérifie  $G(t_0) = t_0$  on a  $t_0 > 0$  car  $G(0) = p_0 > 0$ , donc  $d_1 = G(0) < G(t_0) = t_0$ , puis  $d_2 = G(d_1) < G(t_0) = t_0$ , etc. d'où  $d_n < t_0$  pour tout  $n$ , d'où  $d \leq t_0$ .

Ainsi  $d$  est le plus petit nombre de  $[0,1]$  tel que  $G(d) = d$ .

Comme  $G'(t) = \sum k(k-1)p_k t^{k-2} \geq 0$  sur  $[0,1]$ , la fonction  $G$  est convexe sur  $[0,1]$ . Son graphe ne peut couper la bissectrice qu'en deux points au plus. Comme  $G(1) = 1$  il n'y a que deux cas possibles :



1<sup>er</sup> cas: Si  $\mu = G'(1) > 1$  la pente de la tangente en 1 est plus grande que 1, et on a :  $0 < d < 1$ .

2<sup>ème</sup> cas: Si  $\mu = G'(1) \leq 1$ , on a  $d = 1$ .

En résumé :

**Théorème :** Soit  $\mu$  le nombre moyen de descendants par individu. Si  $\mu \leq 1$  le processus s'éteint avec une probabilité 1. Si  $\mu > 1$ , la probabilité d'extinction est l'unique solution  $< 1$  de l'équation  $G(d) = d$ .

Montrons que  $\lim_{n \rightarrow \infty} G_n(t) = d$  pour tout  $t, 0 \leq t < 1$ .

- i) Si  $0 \leq t < d$ , on a  $d_1 = G(0) \leq G(t) \leq G(d) = d$ , puis  $d_2 = G(d_1) \leq G_2(t) \leq G(d) = d$ , etc.  $d_n \leq G_n(t) \leq d$ .  $\forall n$  d'où  $\lim G_n(t) = d$ .
- ii) Si  $d \leq t < 1$  on a  $d = G(d) \leq G(t) \leq t$ , puis  $d = G(d) \leq G_2(t) \leq G(t) \leq t$ , etc. d'où la suite décroissante  $t \geq G(t) \geq G_2(t) \geq \dots \geq d$ .  
 Donc  $G_n(t) \rightarrow a_t \geq d$ . Si  $a_t > d$  on a  $G(a_t) < a_t$  alors que  $G(a_t) = G(\lim G_n(t)) = \lim G_{n+1}(t) = a_t$ ; donc  $a_t = d$ .

Notons que  $G_n(1) = 1 \forall n$ . Ainsi on a donc

$$G_n(t) = \sum_{j=0}^{\infty} P[X_n = j] t^j \rightarrow d \quad \forall t, 0 \leq t < 1.$$

Par le thm. de continuité on en déduit

$$P[X_n = j] \rightarrow \begin{cases} d & \text{si } j=0 \\ 0 & \text{si } j \geq 1 \end{cases}$$

Donc quel que soit  $k \geq 1$  on a

$$\lim_{n \rightarrow \infty} P[1 \leq X_n \leq k] = 0$$

**Conclusion :** de façon un peu imprécise on dira :  
 - ou bien  $X_n \rightarrow 0$  avec la probabilité  $d$   
 - ou bien  $X_n \rightarrow \infty$  avec la probabilité  $1-d$

**Remarque :**  $(X_n)$  est un exemple de chaîne de Markov car  $X_n$  dépend de  $X_{n-1}$  mais pas de  $X_{n-2}, X_{n-3}, \dots, X_0$  i.e.

$$P[X_n = j \mid X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_1 = i_1] = P[X_n = j \mid X_{n-1} = i_{n-1}]$$

7.5. Fonctions caractéristiques - Résultats.

Lorsque  $X$  n'est plus à valeurs dans  $\mathbb{N}$  on ne peut plus parler de sa fonction génératrice. Si on cherche une fonction qui jouisse des mêmes propriétés que  $G_X$  mais qui soit définie pour tous les v.a.  $X$ , on est conduit à poser

$$\varphi_X(t) = E(e^{itX}); \quad (i = \sqrt{-1}, t \in \mathbb{R})$$

$\varphi_X$  est la fonction caractéristique de  $X$ .

Exemples: 1) Si  $X$  est à valeurs dans  $\mathbb{Z}$ , si  $P[X=n] = p_n$ , on a

$$\varphi_X(t) = \sum_{n \in \mathbb{Z}} p_n e^{int}$$

i.e.  $\varphi_X$  est la somme de la série de Fourier (nom. cour.) de coefficients  $p_n$ .

2) Si  $X$  possède une densité  $f$ , alors

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx$$

i.e.  $\varphi_X$  est la transformée de Fourier de  $f$ .

Propriétés de la fonction caractéristique (\*)

A) la fct. caractéristique de  $X$  caractérise la loi de  $X$

i.e.  $\varphi_X = \varphi_Y \Leftrightarrow F_X = F_Y$  (ou:  $P_X = P_Y$  comme déjà vu)

C'est un théorème de P. Lévy.

B) Si  $E(|X|^k) < +\infty$  alors  $\varphi$  est  $k$  fois cont. dérivable et

$$\varphi^{(k)}(0) = i^k E(X^k)$$

$E(X^k)$  est le  $k$ -ème moment de  $X$ .

C) Si  $X_1, \dots, X_m$  sont indépendantes alors

$$\varphi_{\sum_{i=1}^m X_i}(t) = \prod_{i=1}^m \varphi_{X_i}(t) \quad (\forall t \in \mathbb{R})$$

(\*) Cf. Rényi - Calcul des probabilités ou Métivier - Notions fondamentales de la théorie des proba.

D) Théorème de continuité (P. Lévy): Soient  $X_1, X_2, \dots$  des v.a.

i) s'il existe une v.a.  $X$  telle que  $F_{X_n}(x) \rightarrow F_X(x)$  en chaque point de continuité de  $F_X$  alors  $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$  pour tout  $t \in \mathbb{R}$  (et même unif. sur tout compact).

ii) s'il existe une fonction  $g$  continue en  $t=0$  telle que  $\varphi_{X_n}(t) \rightarrow g(t)$ ,  $\forall t \in \mathbb{R}$ , alors il existe une v.a.  $X$  telle que  $g = \varphi_X$  et  $F_{X_n}(x) \rightarrow F_X(x)$  en chaque point de continuité de  $F_X$ .

E) Si  $N$  est une v.a. à valeurs dans  $\mathbb{N}$  et si  $X_1, X_2, \dots$  sont des v.a. indépendantes de même loi,  $N$  étant indépendante des  $X_i$ , on a pour  $S_N = X_1 + \dots + X_N$ ,

$$\varphi_{S_N}(t) = G_N(\varphi_{X_1}(t)) \quad (\forall t \in \mathbb{R})$$

Voir une application série 23 et page 115

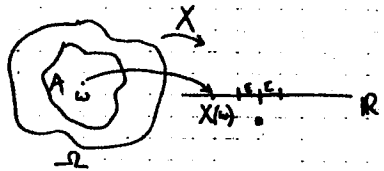
§8. Théorèmes limites.

8.1. Deux inégalités.

Inégalité de Markov: Si  $X$  est une v.a. on a  $\forall \varepsilon > 0, \forall r > 0$

$$P[|X| \geq \varepsilon] \leq \frac{E(|X|^r)}{\varepsilon^r}$$

Preuve:



Soit  $A = [X \leq \varepsilon] = \{\omega \mid |X(\omega)| \leq \varepsilon\}$

et soit  $I_A$  l'indicatrice de  $A$

i.e.

$$I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

On a:  $\varepsilon^r I_A \leq |X|^r$ . En effet si  $\omega \notin A$  le membre de gauche est nul et si  $\omega \in A$  on a  $\varepsilon \leq |X(\omega)|$  par définition de  $A$ , donc  $\varepsilon^r I_A \leq |X(\omega)|^r$ .

Donc (E croissante):  $E(\varepsilon^r I_A) \leq E(|X|^r)$

Mais  $E(\varepsilon^r I_A) = \varepsilon^r E(I_A) = \varepsilon^r P(A)$ . D'où le résultat.

Inégalité de Tchebychev: Soit  $X$  une v.a. telle que  $\mu = E(X)$  et  $\sigma^2 = \text{Var}(X)$  soient finies. Pour tout  $\varepsilon > 0$ , on a

$$P[|X - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2}$$

Preuve: C'est l'inégalité de Markov avec  $X - \mu$  et  $r = 2$ .

Remarque 1. Lorsqu'on connaît la loi de  $X$  et qu'on peut calculer  $P[|X - \mu| \geq \varepsilon]$  on constate en général que l'inégalité de Tchebychev est très grossière.

Par exemple si  $X$  est normale réduite on voit que

- si  $\varepsilon \leq 1$ , l'inégalité de T. n'affrond rien!
- si  $\varepsilon = 2$ ,  $P[|X| \geq 2] = 1 - P[-2 < X < 2] = 2(1 - P[X < 2]) = 0,054 \leq \frac{1}{20}$   
alors que Tchebychev donne  $P[|X| \geq 2] \leq \frac{1}{4}$ .

- si  $\varepsilon = 3$ ,  $P[|X| \geq 3] = 2(1 - P[X < 3]) = \frac{2}{1000}$   
alors que Tchebychev donne  $P[|X| \geq 3] \leq \frac{1}{9} \leq \frac{112}{1000}$

∑ L'intérêt de l'inégalité de Tchebychev réside dans sa généralité

Remarque 2. L'inégalité de Tchebychev confirme l'interprétation de la variance comme mesure de la déviation de  $X$  par rapport à sa moyenne. En particulier, cas extrême:

$$\text{Var}(X) = 0 \rightarrow \forall \varepsilon > 1 \quad P[|X - \mu| \geq \frac{1}{\varepsilon}] = 0$$

$$\rightarrow P[|X - \mu| > 0] = 0, \text{ en utilisant Prop. 2 § 3.3, p. 3}$$

$$\text{i.e. } P[X = \mu] = 1$$

8.2. Lois faibles des grands nombres.

Rappel: Considérons un schéma de Bernoulli de paramètres  $n, p$ . Soit  $S_n = X_1 + \dots + X_n$  la v.a. binomiale associée. On a vu au § 5.4. Application (page 52) que, pour tout  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P\left[\left|\frac{S_n}{n} - p\right| < \varepsilon\right] = 1$$

C'est la loi des grands nombres de Bernoulli.

Nous allons généraliser ce résultat. Précisons le mode de convergence

Définition: Soient  $X_1, X_2, \dots$  et  $X$  des v.a. définies sur  $\Omega$ . On dira que la suite  $(X_n)$  converge en probabilité vers  $X$  si, pour tout  $\varepsilon > 0$ , on a

$$\lim_{n \rightarrow \infty} P[|X_n - X| \geq \varepsilon] = 0$$

ou de façon équivalente

$$\lim_{n \rightarrow \infty} P[|X_n - X| < \varepsilon] = 1$$

Notation:  $X_n \xrightarrow{P} X$  si  $n \rightarrow \infty$ .

La loi des grands nombres de Bernoulli peut s'énoncer: "la v.a.  $\frac{S_n}{n}$  (fréquences relatives) convergent en probabilité vers  $p$ ".

**Théorème 1.** Soient  $X_1, X_2, \dots$  des v.a. indépendantes, telles que  $E(X_i) = E(X_1) = \dots = \mu$  et  $\text{Var}(X_i) = \text{Var}(X_1) = \dots = \sigma^2 < \infty$ .  
Posons  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$  pour tout  $n \geq 1$ . On a :

$$\bar{X}_n \xrightarrow{P} \mu \quad \text{si } n \rightarrow \infty.$$

(C'est une des formes de la loi faible des grands nombres).

Preuve : On a  $E(\bar{X}_n) = \frac{E(X_1) + \dots + E(X_n)}{n} = \mu$

$$\text{et } \text{Var}(\bar{X}_n) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}.$$

Donc  $\text{Var}(\bar{X}_n) \rightarrow 0$ , donc on doit avoir  $\bar{X}_n$  de plus en plus proche de  $\mu$ . Formellement, par l'inégalité de Tchebychev :

$$\forall \varepsilon > 0, \quad P[|\bar{X}_n - \mu| \geq \varepsilon] \leq \frac{\sigma^2}{n \varepsilon^2},$$

$$\text{donc } \lim_{n \rightarrow \infty} P[|\bar{X}_n - \mu| \geq \varepsilon] = 0. \quad \text{Donc } \bar{X}_n \xrightarrow{P} \mu.$$

Remarque : Il suffit en fait que  $E(X_i) \rightarrow \mu$  et  $\text{Var}(\bar{X}_n) \rightarrow 0$ , cette dernière condition étant vérifiée par exemple si  $\text{Cov}(X_i, X_j) = 0, \forall i \neq j$ , et  $\text{Var}(X_i) \leq K, \forall i$ .

Application : Démonstration probabiliste du théorème de Weierstrass.

Théorème : Soit  $g$  une fonction continue sur  $[0, 1]$ .

Il existe une suite  $(g_n)$  de polynômes, telle que  $g_n \rightarrow g$  uniformément sur  $[0, 1]$ .

Preuve : Soient  $S_n = X_1 + \dots + X_n$  une v.a. binomiale  $(n, p)$ .

$$\text{et } \bar{X}_n = \frac{1}{n} S_n; \text{ on a } E(\bar{X}_n) = p \text{ et } \text{Var}(\bar{X}_n) = \frac{p(1-p)}{n}.$$

Idee de la preuve :  $\bar{X}_n \xrightarrow{P} p \rightarrow g(\bar{X}_n) \xrightarrow{P} g(p) \rightarrow E(g(\bar{X}_n)) \rightarrow E(g(p)) = g(p)$

$$\text{Posons } E(g(\bar{X}_n)) = g_n(p)$$

a)  $g_n(p)$  est un polynôme en  $p$ ; en effet :

$$g_n(p) = E(g(\bar{X}_n)) = \sum_{k=0}^n g\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k}.$$

b)  $g_n \rightarrow g$  uniformément sur  $[0, 1]$ . Soit  $\varepsilon > 0$ .

Nous utiliserons deux conséquences de l'hyp. "g continue" :

- i)  $g$  est uniformément continue :  $\exists \delta > 0, |p_1 - p_2| < \delta \Rightarrow |g(p_1) - g(p_2)| < \varepsilon$   
ii)  $g$  est bornée :  $\exists B > 0, |g(p)| \leq B \quad \forall p \in [0, 1]$ .

On a :

$$\begin{aligned} |g_n(p) - g(p)| &= |E(g(\bar{X}_n) - g(p))| \leq E(|g(\bar{X}_n) - g(p)|) \\ &= E(|g(\bar{X}_n) - g(p)| Y_n) + E(|g(\bar{X}_n) - g(p)| (1 - Y_n)) \end{aligned}$$

où  $Y_n$  est la v.a. définie par  $Y_n = \begin{cases} 1 & \text{si } |\bar{X}_n - p| < \delta \\ 0 & \text{sinon} \end{cases}$ ,  
donc :

$$\begin{aligned} |g_n(p) - g(p)| &\leq \underbrace{\varepsilon P[|\bar{X}_n - p| < \delta]}_{< 1} + 2B \underbrace{P[|\bar{X}_n - p| \geq \delta]}_{\leq \frac{p(1-p)}{n\delta^2} \leq \frac{1}{4n\delta^2}} \\ &\leq \varepsilon + \frac{B}{2n\delta^2} \leq 2\varepsilon \quad \text{si } n \geq \frac{B}{2\varepsilon^2} \end{aligned}$$

et cela a lieu uniformément en  $p$  dans  $[0, 1]$ .

Déf.  $g_n(p) = \sum_{k=0}^n g\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k}$  est le polynôme de Bernstein de degré  $n$  associé à  $g$ .

Une autre forme de la loi faible des grands nombres est la suivante (Khinchine, 1929) :

**Théorème 2** : Soient  $X_1, X_2, X_3, \dots$  des v.a. indépendantes, de même loi et qui possèdent une moyenne  $\mu$ .  
Posons  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$  pour tout  $n \geq 1$ . Alors  $\bar{X}_n \xrightarrow{P} \mu$  si  $n \rightarrow \infty$ .

Preuve : voir série 23.

## 8.3. Théorème central limite.

Rappel: Considérons un schéma de Bernoulli de param.  $p$ ; soit  $S_n = X_1 + \dots + X_n$  la v.a. "nombre de succès"; on a  $E(S_n) = np$  et  $\text{Var}(S_n) = npq$ ; on a vu au §5 le théorème de de Moivre-Laplace: si  $n \rightarrow \infty$  alors

$$P\left[a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right] \rightarrow \Phi(b) - \Phi(a)$$

où  $\Phi$  est la fonction de répartition de la loi normale réduite.

Nous allons généraliser ce résultat. Précisons d'abord le mode de convergence.

Définition: Soient  $X_1, X_2, \dots$  et  $X$  des variables aléatoires et  $F_{X_1}, F_{X_2}, \dots, F_X$  les fonctions de répartition associées. On dira que la suite  $(X_n)$  converge en loi vers  $X$  (si  $n \rightarrow \infty$ ) si  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$  pour tout  $x$  en lequel  $F_X$  est continue.

Notation:  $X_n \xrightarrow{L} X$ .

Le théorème de de Moivre-Laplace peut s'énoncer: "les v.a.  $\frac{S_n - np}{\sqrt{npq}}$  convergent en loi vers une v.a. normale réduite!"

Proposition: 1)  $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$

2) Réciproque vraie si  $X = c$  (cst.).

Démonstration: 1) Soient  $x \in \mathbb{R}$  et  $\varepsilon > 0$ . Il est clair que

$$X \geq x + \varepsilon \text{ et } |X_n - X| < \varepsilon \Rightarrow X_n > x$$

donc  $X_n < x \Rightarrow X < x + \varepsilon$  ou  $|X_n - X| \geq \varepsilon$ .

$$P[X_n < x] \leq P[X < x + \varepsilon] + P[|X_n - X| \geq \varepsilon]$$

Ainsi par somm. add.

De même:  $P[X < x - \varepsilon] \leq P[X_n < x] + P[|X_n - X| \geq \varepsilon]$ .

Ainsi on a l'encadrement

$$P[X < x - \varepsilon] - P[|X_n - X| \geq \varepsilon] \leq P[X_n < x] \leq P[X < x + \varepsilon] + P[|X_n - X| \geq \varepsilon]$$

ie. si  $F_n = F_{X_n}$  et  $F = F_X$ :

$$(*) \quad F(x - \varepsilon) - P[|X_n - X| \geq \varepsilon] \leq F_n(x) \leq F(x + \varepsilon) + P[|X_n - X| \geq \varepsilon]$$

La suite  $(F_n(x))_n$  est bornée; soit  $\xi$  une valeur d'adhérence de cette suite i.e.  $\xi = \lim_{k \rightarrow \infty} F_{n_k}(x)$ ; on a en passant à la limite dans (\*) et en utilisant l'hypothèse  $X_{n_k} \xrightarrow{P} X$ :

$$F(x - \varepsilon) \leq \xi \leq F(x + \varepsilon) \quad \forall \varepsilon > 0$$

Si  $F$  est continue en  $x$  on en déduit que  $\xi = F(x)$ .

Mais alors  $F(x)$  est la seule valeur d'adhérence de la suite  $(F_n(x))$ , donc  $F_n(x) \rightarrow F(x)$ . Donc  $X_n \xrightarrow{L} X$ .

2) Il suffit que  $X_n \xrightarrow{L} c$ . La fonction de répartition de la cst.  $c$  est

$$F(x) = \begin{cases} 0 & \text{si } x < c \\ 1 & \text{si } x \geq c \end{cases}$$

donc par hypo. on a:  $F_n(x) \rightarrow 0$  si  $x < c$  et  $F_n(x) \rightarrow 1$  si  $x > c$ .

Si  $x = c$  on a

$$\begin{aligned} P[|X_n - c| \geq \varepsilon] &= P[X_n < c - \varepsilon] + P[X_n > c + \varepsilon] \\ &\leq P[X_n < c - \varepsilon] + P[X_n > c + \varepsilon] \\ &= F_n(c - \varepsilon) + 1 - F_n(c + \varepsilon) \xrightarrow{n \rightarrow \infty} 0 + 1 - 1 = 0 \end{aligned}$$

Donc  $X_n \xrightarrow{P} c$ .

Remarque: D'après le théorème de continuité de Lévy (p. 114) on voit que  $X_n \xrightarrow{L} X \Leftrightarrow \varphi_{X_n}(t) \rightarrow \varphi_X(t) (\forall t)$  où les  $\varphi$  sont les fonctions caractéristiques de v.a. (cf. série 23).

Le problème posé par la généralisation du théorème de de Moivre-Laplace peut s'énoncer ainsi :

Soit  $(X_i)$  une suite de v.a. indépendantes possédant des moyennes <sup>(qui peuvent être nulles)</sup> et des variances. Soit  $S_n = X_1 + \dots + X_n$  pour  $n \geq 1$ . Trouver les conditions "minimales" sous lesquelles

$$(*) \quad S_n^* = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \xrightarrow{L} N^* \quad (n \rightarrow \infty)$$

où  $N^*$  désigne une v.a. normale réduite.

Depuis Laplace ce problème a été un des moteurs du calcul des probabilités ; les pas importants ont été franchis par

Tchelychev (1887) incomplet  
Markov (1898) complète T } méthode des moments ;  
 suffit  $E|X_i|^k| < \infty \forall k$ .  
Liapounov (1900) amélioration nette : pour avoir (\*), il suffit que  
 $\exists \delta > 0$  avec  $\frac{E|X_i|^{2+\delta} + \dots + E|X_n|^{2+\delta}}{s_n^{2+\delta}} \rightarrow 0$  ( $s_n^2 = \text{Var}(S_n)$ )  
 Méthode de fct. caractéristiques (peu rigoureuse).

Lindeberg (1922) et Feller (1937) : trouvant une cond. néc. et suff.  
 pour avoir (\*) et  $\max_{k \geq n} \frac{\text{Var}(X_k)}{s_n^2} \rightarrow 0$  ; c'est-à-dire  
 $\forall \varepsilon > 0, \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x| > \varepsilon s_n} x^2 dF_{X_k}(x) \rightarrow 0$ .

Dans le même ordre d'idées, il faut citer le résultat dû à

Lévy (1934) : soit  $(X_{n,k})_{1 \leq k \leq n}$  ( $n=1,2,\dots$ ) une suite de suites finies de v.a. indépendantes ( $n \rightarrow \infty$ ) ;  
 supposons que  $\sum_{k=1}^n X_{n,k} \xrightarrow{L} X$  si  $n \rightarrow \infty$  ; alors  
 $X$  est normale  $\iff \max_{1 \leq k \leq n} |X_{n,k}| \xrightarrow{L} 0$

On va démontrer un théorème du type Liapounov, mais avec une estimation de la vitesse de convergence (résultat de Berry-Esseen) :

Théorème (central-limite). Soit  $\{X_1, X_2, \dots, X_n \text{ de } \}$  une suite de v.a. indépendantes telles que  $E|X_i|^3| < +\infty$  pour tout  $i$ . Soit  $S_n = X_1 + \dots + X_n$  pour tout  $n \geq 1$ . Posons  $s_n^2 = \text{Var}(S_n)$  et  $S_n^* = \frac{S_n - E(S_n)}{s_n}$  (ou suppose  $s_n > 0$ ).

Alors pour tout  $x \in \mathbb{R}$  on a :

$$|P[S_n^* < x] - \Phi(x)| \leq 3 \left\{ \frac{\sum_{i=1}^n E|X_i - E(X_i)|^3}{s_n^3} \right\}^{1/4}$$

Remarque 1. La condition  $E|X_i|^3| < \infty$  implique  $E|X_i|^2| < \infty$  qui implique elle-même l'existence de  $\text{Var}(X_i)$  et  $E(X_i)$ . Donc  $E(S_n)$  et  $\text{Var}(S_n)$  existent.

[en effet on a  $|X_i|^2 \leq |X_i|^3 + 1$  donc  $E|X_i|^2| \leq E|X_i|^3| + 1$ ]

Remarque 2. Supposons que toutes les  $X_i$  suivent la même loi ; posons  $\mu = E(X_i)$  et  $\sigma^2 = \text{Var}(X_i)$  ; on a  $s_n^2 = n\sigma^2$  et

$$\frac{X_i - E(X_i)}{s_n} = \frac{1}{\sqrt{n}} X_i^*$$

donc  $\{ \dots \} = n E\left( \left| \frac{X_1^*}{\sqrt{n}} \right|^3 \right) = \frac{E|X_1^*|^3|}{\sqrt{n}}$ . On peut énoncer :

Corollaire 1. Hypothèses du thm. Supposons de plus que toutes les  $X_i$  suivent la même loi. Alors :

i)  $|P[S_n^* < x] - \Phi(x)| \leq 3 \left( \frac{E|X_1^*|^3|}{\sqrt{n}} \right)^{1/4}$  ;

ii)  $S_n^* \xrightarrow{L} N^*$  si  $n \rightarrow \infty$ .

Remarque 3. Dans le cas général on peut énoncer le corollaire suivant :



Corollaire 2: Hypothèse du Théorème. Supposons qu'il existe une constante  $M > 0$  telle que

$$E(|X_i - E(X_i)|^3) \leq M \text{Var}(X_i) \quad (\forall i \geq 1)$$

Alors

i)  $|P[S_n^* < x] - \Phi(x)| \leq 3 \left\{ \frac{M}{S_n} \right\}^{1/4}$

ii) si  $S_n \rightarrow \infty$  lorsque  $n \rightarrow \infty$  alors  $S_n^* \xrightarrow{d} N^*$ .

En effet :

$$\{ \dots \} \leq \frac{M \sum_{i=1}^n \text{Var}(X_i)}{S_n^3} = \frac{M}{S_n}$$

Noter que la condition du corollaire 2 est vérifiée si les  $X_i$  sont bornées, car si  $|X_i| \leq c$  on a

$$|X_i - E(X_i)|^3 = |X_i - E(X_i)| (X_i - E(X_i))^2$$

d'où  $E(|X_i - E(X_i)|^3) \leq 2c \text{Var}(X_i)$ .

Réduction: En posant  $U_i = \frac{X_i - E(X_i)}{S_n}$  pour tout  $i \geq 1$  on a  $S_n^* = \sum_{i=1}^n U_i$  et  $\{ \dots \} = \sum_{i=1}^n E(|U_i|^3)$ , donc le théorème s'énonce:

Théorème: Soient  $U_1, U_2, \dots, U_n$  des v.a. indépendantes telles que  $E(U_i) = 0$  et  $E(|U_i|^3) < +\infty$  pour tout  $i$ . Soit  $U = \sum_{i=1}^n U_i$ ; supposons que  $\text{Var}(U) = 1$ . Alors

$$|P[U < x] - \Phi(x)| \leq 3 \left\{ \sum_{i=1}^n E(|U_i|^3) \right\}^{1/4}$$

pour tout  $x \in \mathbb{R}$ .

Nous allons démontrer ce théorème de façon élémentaire i.e. sans utiliser la notion de fct. caractéristique.

Remarque: Pour une preuve "élémentaire" - i.e. sans utiliser la transformée de Fourier - du théorème de Lindeberg, voir H.F. Trotter - An Elementary Proof of the Central Limit Theorem - Archiv der Math. 10 (1959) pages 226-234.

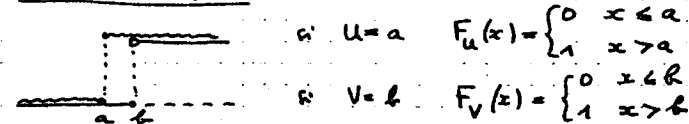
Def: Une famille d'écartés sur la fct. de répartition.

Soient  $U$  et  $V$  deux v.a.,  $F_U$  et  $F_V$  leur f.r. Posons

$$d(U, V) = \sup_{x \in \mathbb{R}} |F_U(x) - F_V(x)| = \|F_U - F_V\|_{\infty}$$

la conclusion du thm. s'énonce:  $d(U, N^*) \leq 3 \{ \dots \}^{1/4}$ .

Exemple avec  $d(U, V)$ :



on a  $d(U, V) = 1$  si  $a \neq b$ ,  $d(U, V) = 0$  si  $a = b$ .

on aurait bien que si  $b \rightarrow a$  alors  $d(U, V) \rightarrow 0$ .

Remarque: Soit  $f_x$  la fonction définie par  $f_x(x) = \begin{cases} 0 & x < \alpha \\ 1 & x \geq \alpha \end{cases}$  ( $\alpha \in \mathbb{R}$ ).  
Alors  $f_x(U)$  est une v.a. qui prend 2 valeurs:

$$P[f_x(U) = 0] = P[U < \alpha]; \quad P[f_x(U) = 1] = P[U \geq \alpha]$$

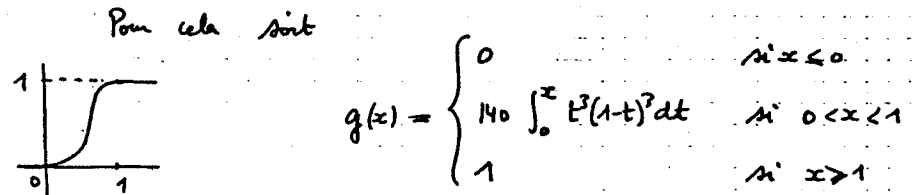
et on a  $E(f_x(U)) = P[U \geq \alpha] = 1 - F_U(\alpha)$

i.e. on connaît la bi d'une v.a. lorsqu'on connaît le moyenn. de certaines fonctions de cette v.a.

On a déduit aussi l'égalité:

$$d(U, V) = \sup_{x \in \mathbb{R}} |F_U(x) - F_V(x)| = \sup_{x \in \mathbb{R}} |E(f_x(U)) - E(f_x(V))|$$

Pour corriger le défaut mentionné ci-dessus on va remplacer  $f_x$  par une fonction de la forme  $\int_0^x \dots dt$



pour  $\forall x \in \mathbb{R}, \forall \epsilon > 0, \quad g_{\alpha, \epsilon}(x) = g\left(\frac{x - \alpha}{\epsilon}\right)$

Calcul:  $|g''(x)| \leq 60 \quad \forall x \in \mathbb{R}$ .

$$d_E(U, V) = \sup_{\alpha \in \mathbb{R}} |E(g_{\alpha, \epsilon}(U)) - E(g_{\alpha, \epsilon}(V))|$$

Les propriétés suivantes sont immédiates :

- 1)  $d_E(U, V) \leq d_E(U, W) + d_E(W, V)$
- 2)  $d_E(U, V) = d_E(V, U)$
- 3)  $F_U = F_V \Rightarrow d_E(U, V) = 0$
- 4)  $d_E(U+a, V+a) = d_E(U, V) \quad (V, a \in \mathbb{R})$

Proposition : Soient  $U$  une v.a. quelconque et  $N$  une v.a. normale de variance  $\sigma^2$ . Pour tout  $\epsilon > 0$  on a

$$d(U, N) \leq \frac{\epsilon}{2\sigma} + d_E(U, N)$$

Preuve : Notons que  $f_{\alpha+\epsilon} \leq g_{\alpha, \epsilon} \leq f_\alpha$  dans un a

$$E(f_{\alpha+\epsilon}(U)) \leq E(g_{\alpha, \epsilon}(U)) \leq E(f_\alpha(U))$$

d'où  $P[\alpha+\epsilon \leq U] \leq E(g_{\alpha, \epsilon}(U)) \leq P[\alpha \leq U]$

Quelles que soient les v.a.  $U$  et  $V$  on a donc

$$P[\alpha+\epsilon \leq V] \leq E(g_{\alpha, \epsilon}(V)) = E(g_{\alpha, \epsilon}(U)) + E(g_{\alpha, \epsilon}(V)) - E(g_{\alpha, \epsilon}(U)) \leq P[\alpha \leq U] + d_E(U, V)$$

Comme  $P[\alpha \leq V] = P[\alpha \leq V < \alpha+\epsilon] + P[\alpha+\epsilon \leq V]$  on a d'où

$$P[\alpha \leq V] - P[\alpha \leq U] \leq P[\alpha \leq V < \alpha+\epsilon] + d_E(U, V)$$

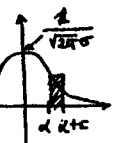
Si  $V$  est normale de variance  $\sigma^2$  on a  $P[\alpha \leq V < \alpha+\epsilon] \leq \frac{\epsilon}{\sqrt{2\pi}\sigma} < \frac{\epsilon}{2\sigma}$  d'où

$$P[\alpha \leq V] - P[\alpha \leq U] \leq \frac{\epsilon}{2\sigma} + d_E(U, V)$$

En remplaçant  $U$  par  $V$  et  $V$  par  $U$  et  $\alpha$  par  $\alpha-\epsilon$  dans (\*) on obtient

$$P[\alpha \leq U] - P[\alpha \leq V] \leq P[\alpha-\epsilon \leq V < \alpha] + d_E(U, V) < \frac{\epsilon}{2\sigma} + d_E(U, V)$$

d'où la Proposition.



Stratégie : on va utiliser deux propriétés :

Propriété 1 : Si  $U, V, Z$  sont indép. alors  $d_E(U+Z, V+Z) \leq d_E(U, V)$

Propriété 2 : Si  $E(U) = E(V), E(U^2) = E(V^2)$  alors  $d_E(U, V) \leq \frac{10}{\epsilon} \{E(U^3) + E(V^3)\}$

Soit  $U = U_1 + \dots + U_n$  comme dans le théorème. Écrivons  $U = U_1 + Z$ , si on remplace  $U_1$  par une v.a. normale  $V_1$  on altere peu  $U = U_1 + Z$  si  $d_E(U_1, V_1)$  est petit (propriété 1), c'est à dire si  $E(U_1^3)$  et  $E(V_1^3)$  sont petits (propriété 2).

On remplacera ensuite  $U_2$  par  $V_2$ , etc. en prenant  $V_1, V_2, \dots$  indépendants, alors  $V = V_1 + \dots + V_n$  sera normale, et peu "éloignée" de  $U$ .

Lemme 1. Soient  $U, V$  et  $Z$  des v.a. indépendantes avec  $E(U) = E(V), E(U^2) = E(V^2)$  et  $E(U^3) < \infty, E(V^3) < \infty$ . Alors on a

$$|E(g(U+Z)) - E(g(V+Z))| \leq 10 \{E(U^3) + E(V^3)\}$$

Preuve : Formule de Taylor :

$$g(z+u) = g(z) + g'(z)u + \frac{g''(z)}{2}u^2 + v_3(z, u)u^3$$

où  $v_3(z, u) = \frac{g'''(\xi)}{6}$  avec  $z < \xi < z+u$ , donc  $|v_3(z, u)| \leq \frac{60}{6} = 10$

On en déduit

$$g(U+Z) - g(V+Z) = g'(Z)(U-V) + \frac{g''(Z)}{2}(U^2 - V^2) + v_3(Z, U)U^3 - v_3(Z, V)V^3$$

et par indépendance

$$E(g(U+Z)) - E(g(V+Z)) = E(g'(Z))E(U-V) + \frac{1}{2}E(g''(Z))E(U^2 - V^2) + E\{v_3(Z, U)U^3 - v_3(Z, V)V^3\}$$

d'où

$$|E(g(U+Z)) - E(g(V+Z))| \leq E\{|v_3(Z, U)U^3 - v_3(Z, V)V^3|\} \leq 10\{E(U^3) + E(V^3)\}$$

En remplaçant  $U$  par  $\frac{U}{\epsilon}$ ,  $V$  par  $\frac{V}{\epsilon}$ ,  $Z$  par  $\frac{Z-\alpha}{\epsilon}$  on obtient  
 $|E(g_{a,c}(U+Z)) - E(g_{a,c}(V+Z))| \leq \frac{10}{\epsilon^3} \{E(U^3) + E(V^3)\}$  d'où

Corollaire: sous les hypothèses du lemme on a:  
 $d_\epsilon(U+Z, V+Z) \leq \frac{10}{\epsilon^3} \{E(U^3) + E(V^3)\}$

Lemme 2. Soient  $U_1, U_2, \dots, U_n, V_1, V_2, \dots, V_n$  des v.a. indépendantes avec  $E(U_i) = E(V_i)$ ,  $E(U_i^2) = E(V_i^2)$ ,  $E(U_i^3) < +\infty$ ,  $E(V_i^3) < +\infty$  pour tout  $i$ . Alors on a

$$d_\epsilon(\sum_1^n U_i, \sum_1^n V_i) \leq \frac{10}{\epsilon^3} \sum_1^n \{E(U_i^3) + E(V_i^3)\}$$

Preuve: Posons:  $W_0 = U_1 + U_2 + \dots + U_n$   
 $W_k = U_1 + \dots + U_k + V_{k+1} + \dots + V_n$   $1 \leq k \leq n-1$   
 $W_n = V_1 + \dots + V_n$

On a:  $d_\epsilon(W_0, W_n) \leq d_\epsilon(W_0, W_1) + d_\epsilon(W_1, W_2) + \dots + d_\epsilon(W_{n-1}, W_n)$   
 et  $d_\epsilon(W_{k+1}, W_k) = d_\epsilon(U_{k+1} + Z_k, V_k + Z_k) \leq \frac{10}{\epsilon^3} \{E(U_{k+1}^3) + E(V_k^3)\}$

Preuve du théorème:  
 Soient donc  $U_1, \dots, U_n$  des v.a. indépendantes avec  $E(U_i) = 0$  et  $E(U_i^3) < +\infty$ ,  $\forall i$ . Posons  $U = \sum_1^n U_i$  et supposons de plus que  $E(U^2) = 1$ .

Soient  $V_1, \dots, V_n$  des v.a. indépendantes entre elles et indép. des  $U_i$  avec  $V_i$  normale,  $E(V_i) = 0$ ,  $E(V_i^2) = E(U_i^2)$ ,  $\forall i$ . Posons  $V = \sum_1^n V_i$ ;  $V$  est normale réduite.

Calcul:  $E(V^3) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |v|^3 e^{-\frac{1}{2}v^2} dv = \frac{4}{\sqrt{2\pi}} < 2$  ( $t = \frac{v^2}{2}$ )

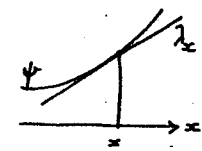
Pour tout  $i$  on a  $V_i = E(V_i^2)^{\frac{1}{2}} \cdot \frac{V_i}{E(V_i^2)^{\frac{1}{2}}}$  où  $\frac{V_i}{E(V_i^2)^{\frac{1}{2}}}$  est normal centré et réduite.

Donc  $E(V_i^3) \leq 2 E(V_i^2)^{\frac{3}{2}} = 2 E(U_i^2)^{\frac{3}{2}}$

Inégalité de Jensen: Si  $\psi$  est une fonction convexe et si  $X$  est une v.a. telle que  $E(|X|) < +\infty$ , on a

$$\psi(E(X)) \leq E(\psi(X))$$

Preuve:  $\psi$  est convexe, donc pour tout  $x$  il existe une fonction "biaisée"  $\lambda_x$  t.q.  $\lambda_x \leq \psi$  et  $\lambda_x(x) = \psi(x)$ .



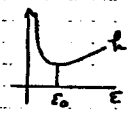
Alors pour  $x = E(X)$  on a  
 $\psi(E(X)) = \lambda_x(E(X)) = E(\lambda_x(X)) \leq E(\psi(X))$

Un particulier:  $\psi(x) = |x|^{3/2}$  d'où  $E(X^2)^{3/2} \leq E(|X|^3)$

Ainsi on a:  $E(V_i^3) \leq 2 E(U_i^3)$  pour tout  $i$ .

Finalement quel que soit  $\epsilon > 0$ .

$$\begin{aligned} d(U, V) &\leq \frac{\epsilon}{2} + d_\epsilon(U, V) && \text{Proposition} \\ &\leq \frac{\epsilon}{2} + \frac{10}{\epsilon^3} \sum_1^n \{E(U_i^3) + E(V_i^3)\} && \text{Lemme 2} \\ &\leq \frac{\epsilon}{2} + \frac{10}{\epsilon^3} \cdot 3 \sum_1^n E(U_i^3) && \text{ci-dessus} \\ &= \frac{\epsilon}{2} + \frac{30B}{\epsilon^3} = h(\epsilon) \text{ où } B = \sum_1^n E(U_i^3) \end{aligned}$$



Le choix de  $\epsilon = \epsilon_0 = (180B)^{1/4}$  minimise  $h(\epsilon)$  et donne

$$d(U, V) \leq h(\epsilon_0) = \frac{2}{3} (180B)^{1/4} < 3 B^{1/4}$$

d'où le théorème.

Voir une application au schéma de Bernoulli généralisé à la série 24.

§9. Mouvement brownien et processus de Wiener-Lévy.9.1. Introduction.

Pour décrire l'évolution d'un système aléatoire au cours du temps on introduit la notion de processus stochastique.

Définitions: 1. Un processus stochastique (ou aléatoire) est une famille  $\{X_t\}_{t \in T}$  de v.a. définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , l'indice  $t$  parcourant un sous-ensemble  $T$  de  $\mathbb{R}$ .

2. Une trajectoire du processus stochastique  $\{X_t\}_{t \in T}$  est une fonction  $t \in T \mapsto X_t(\omega) \in \mathbb{R}$  où  $\omega \in \Omega$ .

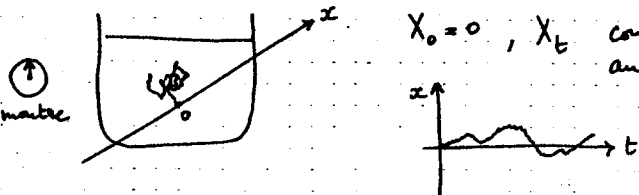
On a déjà rencontré de tels processus:

- le processus de Poisson [cf. radioactivité et série 17],
- le processus de ramification [cf. §7 et série 22].

A côté du processus de Poisson, le processus le plus important est celui de Wiener-Lévy qui décrit le mouvement brownien (linéaire).

1827 le botaniste Robert Brown observe le mouvement irrégulier d'une petite particule plongée dans un liquide (\*).

1905 Einstein explique ce mouvement par les collisions perpétuelles de la particule avec les molécules du milieu environnant (des millions de choc par seconde...) et donne les premiers résultats quantitatifs:



(\*) cf. Encyclopaedia Universalis: Colloïdes

Soit  $f(x,t)$  la densité de la v.a.  $X_t$ . Einstein a montré que

$$\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2}$$

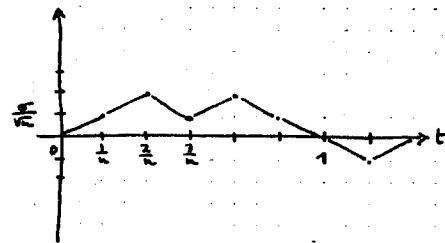
(cf. équation de diffusion, Rossel p. 306-307) où  $D$  est le coefficient de diffusion  $D = 2 \frac{RT}{Nf}$  où  $R$  est la constante des gaz parfaits,  $T$  température,  $N$  nb. d'Avogadro,  $f$  un coeff. de friction (qui dépend de  $k$  viscosité du fluide et de la dimension de la particule). Remarque: Perrin a utilisé cette relation pour calculer  $N$  à partir du mouvement brownien (Nobel 1926).

Pour  $\sigma^2 = 2D$ ; alors on vérifie que  $f(x,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{x^2}{2\sigma^2 t}}$  est solution de l'équation ci-dessus, donc  $X_t$  est une v.a. de loi normale de moyenne 0 et de variance  $\sigma^2 t$ .

1918-1930: Wiener trouve la 1<sup>ère</sup> construction rigoureuse d'un modèle mathématique décrivant le mouvement brownien.

1939-1948: Lévy donne une analyse exhaustive du mouvement brownien dans son livre "Processus stochastiques et mouvement brownien".

Le mouvement brownien comme limite d'une partie de pite ou face



Soit  $\sigma > 0$ . Soit  $n$  entier  $\geq 1$ .  
En chaque instant  $\frac{j}{n}$  ( $j=1,2,\dots$ ) la particule subit un déplacement  $\pm \frac{\sigma}{\sqrt{n}}$ , ou mieux un déplacement  $\frac{\sigma}{\sqrt{n}} Y_j$  où  $Y_j$  prend les valeurs  $+1$  et  $-1$  avec  $P[Y_j = +1] = P[Y_j = -1] = \frac{1}{2}$ .

On pose  $\forall k \geq 1$ :

$$X_{\frac{k}{n}}^{(n)} = \frac{\sigma}{\sqrt{n}} (Y_1 + Y_2 + \dots + Y_k)$$

et par interpolation linéaire:

$$\forall t > 0 \quad X_t^{(n)} = X_{\frac{k}{n}}^{(n)} + n(t - \frac{k}{n}) \frac{\sigma}{\sqrt{n}} Y_{k+1} \quad \text{si } \frac{k}{n} \leq t < \frac{k+1}{n}$$

On a:  $E(X_t^{(n)}) = 0 \quad \forall t$

et  $\text{Var}(X_t^{(n)}) = \text{Var}(X_{\frac{k}{n}}^{(n)}) = \frac{\sigma^2}{n} \sum_1^k \text{Var}(Y_j) = \frac{\sigma^2}{n} k = \sigma^2 t \quad \text{si } t = \frac{k}{n}$

NB. On a  $\text{Var}(X_1^{(n)}) = \sigma^2$  quel que soit  $n$ . C'est pour avoir cette propriété qu'on a choisi un déplacement de  $\frac{\sigma}{\sqrt{n}}$ .

Soit  $t = \frac{k}{n}$ ;  $X_t^{(n)} = \frac{\sigma}{\sqrt{n}} Y_1 + \dots + \frac{\sigma}{\sqrt{n}} Y_k$  est une somme de  $k$  v.a. indépendantes de même loi. D'après le théorème central-limite (Corollaire 1) on a

$$P\left[\frac{X_t^{(n)}}{\sigma\sqrt{t}} < x\right] - \Phi(x) \leq 3 \left\{ \frac{E(|\frac{\sigma}{\sqrt{n}} Y_1|^3)}{\sqrt{t}} \right\}^{1/4} = \frac{3}{(nt)^{1/4}}$$

(car  $(\frac{\sigma}{\sqrt{n}} Y_1)^2 = Y_1$  et  $E(|Y_1|^3) = 1$ ).

On en déduit (\*) que si  $n \rightarrow \infty$  et  $t$  est fixé alors  $\{X_t^{(n)}\}$  converge en loi vers une loi normale de moyenne 0 et de variance  $\sigma^2 t$ .

On peut aussi dire que si  $n$  est suffisamment grand pour que  $\frac{1}{n}$  et  $\frac{\sigma}{\sqrt{n}}$  soient plus petits que la précision des mesures (de temps et d'espace) alors  $\{X_t^{(n)}\}_t$  est un modèle mathématique du mouvement brownien.

Définition: Le processus de Wiener-Lévy de paramètre  $\sigma > 0$

est le processus stochastique  $\{X_t\}_{0 \leq t < \infty}$  où

- (1)  $X_0 = 0$ ;
- (2) pour tout  $t$  et tout  $h$  la loi de  $X_{t+h} - X_t$  ne dépend que de  $h$ ;
- (3) si  $0 \leq t_1 < t_2 < \dots < t_n$  les v.a.  $X_{t_1} - X_0, \dots, X_{t_n} - X_{t_{n-1}}$  sont indépendantes;
- (4) pour tout  $t$  et tout  $h$  la v.a.  $X_{t+h} - X_t$  est normale  $(0, \sigma^2 h)$ ;
- (5) chaque trajectoire du processus est continue.

(\*) Il faudrait faire attention avec  $nt = k$  entier... c'est un détail!

Remarques:

1) Justification des hypothèses: (2) est acceptable dans un milieu en équilibre macroscopique; (4) a été justifié de 2 façons; (5) est naturel: des discontinuités dans les trajectoires seraient choquantes; (3) est plus difficile à justifier: on peut l'admettre à une échelle pas trop petite (pour  $t$ ) mais cela devient absurde pour des intervalles de temps de l'ordre du temps entre deux collisions successives (\*). D'ailleurs on peut démontrer que les trajectoires du processus de W-L. n'ont de dérivée en aucun point, donc sont non rectifiables! L'étude infinitésimale des trajectoires n'a plus d'interprétation physique très exacte. On peut amplifier (3) pour être plus réaliste (cf. Doob).

2) On peut démontrer que le processus de W-L. existe (Wiener en a donné une construction en 1923; Lévy une autre).

3) la condition (3) implique la propriété de Markov.

4) la condition (5) peut être considérée comme une conséquence des autres (cf. Breiman Cor. 12.21)

5) Utilisations des processus de W-L.

- mouvement brownien,
- bruits thermiques dans les circuits électriques
- mécanique quantique (cf. Kac)
- gestion de stocks

Il apparaît d'autre part comme la limite de nombreux processus. C'est à la fois le plus simple et le plus pathologique des processus (le "monstre absolu" comme l'a dit Poincaré).

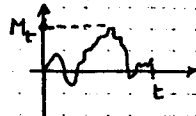
(\*) Citons Lévy (page 15): "Le processus stochastique... qui représente bien les propriétés du mouvement brownien à une échelle assez petite, mais non infiniment petite, et qui suffit que les mêmes propriétés existent à n'importe quelle échelle."

## 9.2. Quelques propriétés des processus de Wiener-Lévy (\*)

Soit  $\{X_t\}_{0 \leq t < \infty}$  le processus de W.-L. de paramètre  $\sigma > 0$

Fixons  $t > 0$ . Par continuité des trajectoires on peut considérer la v.a.

$$M_t = \max_{0 \leq s \leq t} X_s$$



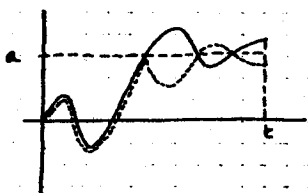
Théorème:  $P[M_t > a] = 2P[X_t > a]$  pour tout  $a > 0$ .

Justification par le principe de réflexion (André)

Si  $M_t > a$  on peut distinguer deux cas qui s'excluent :

①:  $M_t > a$  et  $X_t > a$

②:  $M_t > a$  et  $X_t < a$



on choisit le premier  $t_a$  pour lequel la trajectoire atteint  $a$ ; à partir de là on symétrise la trajectoire par rapport à l'horizontale  $a$ .

Alors à chaque trajectoire vérifiant ① on associe une trajectoire vérifiant ② et réciproquement.

A cause de la symétrie de la loi normale, les trajectoires correspondantes "ont même probabilité", donc

$$P[M_t > a \text{ et } X_t > a] = P[M_t > a \text{ et } X_t < a]$$

Ainsi on a bien :

$$\begin{aligned} P[M_t > a] &= P[M_t > a \text{ et } X_t > a] + P[M_t > a \text{ et } X_t < a] \\ &= 2P[M_t > a \text{ et } X_t > a] = 2P[X_t > a] \end{aligned}$$

NB. Le raisonnement ci-dessus n'est pas complet ! Mais le principe de réflexion est un outil efficace dans l'étude des processus à trajectoires continues.

(\*) cf. P. Lévy "Proc. stoch. et mart. brownien" Chap. VI.

Remarque: on a  $2P[X_t > a] = P[|X_t| > a]$  ( $\forall a > 0$ ) donc  $M_t$  et  $|X_t|$  suivent la même loi. Comme  $X_t$  suit une loi normale  $(0, \sigma\sqrt{t})$  on a :

Corollaire:  $M_t$  admet pour densité la fonction

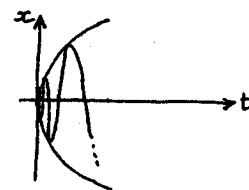
$$f_{M_t}(a) = \begin{cases} 0 & \text{si } a \leq 0 \\ \frac{2}{\sqrt{2\pi}\sigma\sqrt{t}} e^{-\frac{1}{2}\frac{a^2}{\sigma^2 t}} & \text{si } a > 0 \end{cases}$$

Preuve:  $P[M_t < a] = 1 - P[M_t > a] = 1 - 2P[X_t > a] = 1 - 2(1 - P[X_t < a])$   
 $= 2P[X_t < a] - 1 = 2\Phi\left(\frac{a}{\sigma\sqrt{t}}\right) - 1$ ; en dérivant on a  
 bien  $f_{M_t}(a) = 2\varphi\left(\frac{a}{\sigma\sqrt{t}}\right) \frac{1}{\sigma\sqrt{t}} = \frac{2}{\sqrt{2\pi}\sigma\sqrt{t}} e^{-\frac{1}{2}\frac{a^2}{\sigma^2 t}}$

Nous allons voir quelques applications du Théorème 1 qui montrent que le mouvement brownien possède des propriétés qui semblent tout d'abord contraires au bon sens.

A) Oscillation des trajectoires.

Il semblerait naturel que, sur un petit intervalle près de 0, la moitié des trajectoires soient  $> 0$  et la moitié soit  $< 0$ . Nous allons voir que c'est faux et que presque toutes les trajectoires "changent d'avis" une infinité de fois avant de "choisir" de rester un moment en dessous ou en dessus de l'axe  $Ox$ .



Cela se passe comme pour  $\sqrt{t} \sin \frac{1}{t}$ .

De façon précise

Proposition: Au voisinage de  $t=0$  les trajectoires du processus ont presque sûrement une infinité de zéros.

Démonstration: a) On a  $[M_t > 0] = \bigcup_{n=1}^{\infty} [M_t \geq \frac{1}{n}]$  donc

$$P[M_t > 0] = \lim_{n \rightarrow \infty} P[M_t \geq \frac{1}{n}] = 2 \lim_{n \rightarrow \infty} P[X_t \geq \frac{1}{n}] = 2P[X_t > 0] = 2 \cdot \frac{1}{2} = 1,$$

et cela quel que soit  $t > 0$ . Donc de tout voisinage  $[0, t]$  de l'origine, <sup>(la trajectoire)</sup> presque sûrement (i.e. avec une proba. 1) une valeur  $> 0$ .

Si  $M_t'$  désigne le minimum de  $X_s$  sur  $[0, t]$ , il est clair que  $-M_t'$  suit la même loi que  $M_t$ ; donc dans tout voisinage  $[0, t]$  de 0, la trajectoire presque sûrement prend une valeur  $< 0$ .

Par continuité,  $\forall t > 0, P[\exists s, 0 < s < t, \text{ avec } X_s = 0] = 1$ .

b) Considérons les événements  $Z_n = [\exists s, 0 < s < \frac{1}{n}, \text{ avec } X_s = 0]$ ; on a  $Z_1 \supset Z_2 \supset Z_3 \supset \dots$  et  $P(Z_n) = 1 (\forall n > 1)$  donc

$$P(\bigcap_{n=1}^{\infty} Z_n) = 1$$

Mais  $\bigcap_{n=1}^{\infty} Z_n = [\forall n > 1, X_s \text{ a un zéro dans } ]0, \frac{1}{n}[ ]$ , donc il est presque sûr que les trajectoires possèdent une suite de zéros qui s'accumulent à l'origine.

Remarque. Par translation on voit que si  $X_t = x$ , alors dans tout intervalle  $[t, t+h]$  ( $h > 0$ ) les trajectoires vont prendre p.s. une infinité de fois la valeur  $x$ . On comprend que les trajectoires soient p.s. non dérivables au tout point!

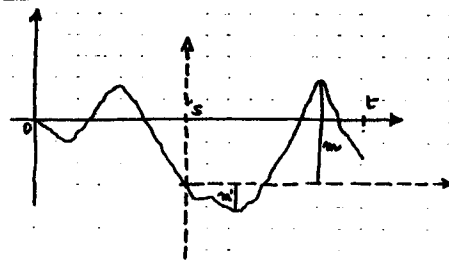
B) La loi de l'Arc sinus.

Proposition: Considérons  $0 \leq s < t$ . Alors on a:

$$P[\exists \tau, s \leq \tau \leq t, \text{ avec } X_\tau = 0] = \frac{2}{\pi} \text{Arccos} \sqrt{\frac{s}{t}}$$

NB. Pour  $s=0$  la proposition est évidente car  $X_0 = 0$ . On peut donc supposer  $s > 0$ .

Preuve.



On remarque tout d'abord que  $Y_h = X_{s+h} - X_s$  ( $h > 0$ ) est un processus de W.L. de même paramètre que  $\{X_t\}$ . Notons  $M$  et  $M'$  les v.a. max. et min. de  $Y_h$  sur  $[0, t-s]$

On a déjà remarqué que  $M$  et  $-M'$  suivent la même loi. D'autre part il est clair que  $M$  et  $X_s$  sont indépendantes, de même pour  $M'$  et  $X_s$ .

Comme  $[\exists \tau, s \leq \tau \leq t, \text{ tel que } X_\tau = 0] = [X_s \leq 0 \text{ et } M \geq -X_s] \cup [X_s > 0 \text{ et } M' \leq -X_s]$    
 on a: réunion disjointe

$$p = P[\exists \tau, s \leq \tau \leq t, X_\tau = 0] = P[X_s \leq 0 \text{ et } M \geq -X_s] + P[X_s > 0 \text{ et } M' \leq -X_s]$$

$$\begin{aligned} \text{mais } P[X_s > 0 \text{ et } M' \leq -X_s] &= P[X_s > 0] P[M' \leq -X_s] = P[X_s > 0] P[M \geq X_s] \\ &\quad \uparrow \quad \uparrow \\ &\quad X_s \text{ et } M' \text{ indép.} \quad M' \text{ et } -M \text{ ont la même loi} \\ &= P[X_s > 0 \text{ et } M \geq X_s] \end{aligned}$$

donc

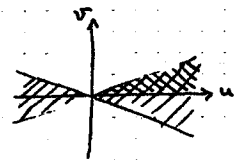
$$\begin{aligned} p &= P[X_s \leq 0 \text{ et } M \geq |X_s|] + P[X_s > 0 \text{ et } M \geq |X_s|] \\ &= P[M \geq |X_s|] \end{aligned}$$

- Alors:
- i) D'après le théorème 1,  $M$  suit la même loi que  $|X_t - X_s|$  et  $X_t - X_s$  suit une loi normale  $(0, \sigma\sqrt{t-s})$ ; donc on peut écrire  $M = \sigma\sqrt{t-s} |U|$  où  $U$  est une v.a. normale réduite.
  - ii) On peut écrire  $|X_s| = \sigma\sqrt{s} |V|$  où  $V$  est normale réduite.
  - iii)  $M$  et  $|X_s|$  sont indép.  $\Rightarrow U$  et  $V$  sont indép.

et on peut écrire:  $p = P\left[\left|\frac{V}{U}\right| \leq \frac{\sqrt{t-s}}{\sqrt{s}}\right]$

Le couple  $(U, V)$  admet pour densité la fonction

$$f(u, v) = \frac{1}{2\pi} e^{-\frac{1}{2}(u^2+v^2)} \quad (v(u, v) \in \mathbb{R}^2)$$



la région hachurée correspond à  $\left|\frac{v}{u}\right| \leq c$

Coord. polaires:  
$$\begin{cases} r^2 = u^2 + v^2 \\ \text{tg } \varphi = \frac{v}{u} \end{cases}$$

On a

$$p = 4 \cdot \frac{1}{2\pi} \int_0^{+\infty} dr \int_0^{\text{Arctg} \frac{\sqrt{t-s}}{\sqrt{s}}} d\varphi r e^{-\frac{1}{2}r^2}$$

$$= \frac{2}{\pi} \text{Arctg} \frac{\sqrt{t-s}}{\sqrt{s}} = \frac{2}{\pi} \text{Arctg} \sqrt{\frac{t-s}{s}}$$

$$= \frac{2}{\pi} \text{Arccos} \sqrt{\frac{s}{t}}$$

la dernière égalité vient de  $\text{Arctg } x = \text{Arccos} \frac{1}{\sqrt{1+x^2}}$

Remarque. En vertu de la relation  $\text{Arccos } x + \text{Arccos } \sqrt{1-x^2} = \frac{\pi}{2}$  on voit que:

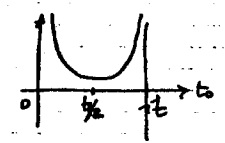
$$P[X_t \text{ ne s'annule pas entre } s \text{ et } t] = \frac{2}{\pi} \text{Arccos} \sqrt{\frac{s}{t}}$$

Soit  $T_0$  le dernier zéro de la trajectoire avant  $t$ .  
On a donc (loi de l'Arc finis pour la dernière visite):

$$P[T_0 < t_0] = \frac{2}{\pi} \text{Arccos} \sqrt{\frac{t_0}{t}}$$

d'où

$$f_{T_0}(t_0) = \frac{1}{\pi \sqrt{t_0(t-t_0)}}$$



Conclusions:

les valeurs les moins probables pour  $T_0$  sont les valeurs voisines de  $\frac{t}{2}$ ; les valeurs les plus probables sont les valeurs extrêmes (proches de 0 et de  $t$ ). Quel que soit  $t$ , il y a une probabilité  $\frac{1}{2}$  que la trajectoire ne coupe plus l'axe dans la 2<sup>ème</sup> moitié de la durée d'observation!

Exemple (cf. Feller I pages 78 et suivantes). On jette un fil à la vitesse de 1 partie par seconde, jour et nuit pendant 1 an. Alors

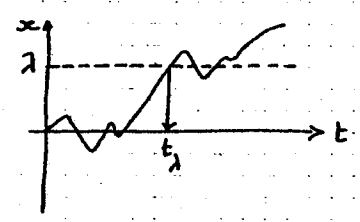
- 1 fois sur 10 la dernière égalisation aura lieu avant le 9<sup>ème</sup> jour
- 1 fois sur 20 elle aura lieu avant 2 1/4 jours.
- 1 fois sur 100 avant 2 h. 10 min.

C) Temps d'attente du niveau  $\lambda$ .

On a vu que  $P[M_t > \lambda] = 2P[X_t > \lambda] = 2(1 - P[X_t \leq \lambda])$   
 $= 2(1 - P[\frac{X_t}{\sigma\sqrt{t}} \leq \frac{\lambda}{\sigma\sqrt{t}}]) = 2(1 - \Phi(\frac{\lambda}{\sigma\sqrt{t}}))$

donc si  $t \rightarrow +\infty$   $P[M_t > \lambda] \rightarrow 2(1 - \Phi(0)) = 1$

i.e. quel que soit  $\lambda \geq 0$  il est presque certain que la trajectoire finira par dépasser le niveau  $\lambda$ .



Soit  $T_2$  le temps d'attente du niveau  $\lambda$  i.e.

$$T_2 = \inf\{t \mid M_t \geq \lambda\}$$

Il est clair que  $[T_2 \leq t] = [M_t \geq \lambda]$

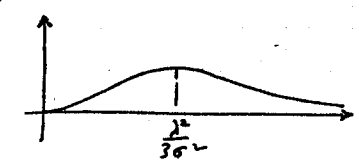
donc

$$F_{T_2}(t) = P[T_2 \leq t] = P[M_t \geq \lambda] = 2(1 - \Phi(\frac{\lambda}{\sigma\sqrt{t}}))$$

$[T_2 \leq t] \subset [M_t \geq \lambda]$  qui est de prob. nulle ci-dessus

d'où en dérivant:

Proposition:  $T_2$  admet pour densité  $f_{T_2}(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \frac{\lambda}{\sigma\sqrt{t}^3} e^{-\frac{\lambda^2}{2\sigma^2 t}} & \text{si } t > 0 \end{cases}$

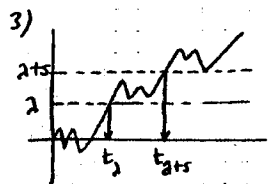




Remarques : 1)  $E(T_2) = +\infty$ .

2) on a  $f_{T_2}(t) = \frac{\lambda^2}{\sigma \sqrt{2\pi} (\frac{t}{\sigma})^{3/2}} e^{-\frac{1}{2} \frac{t^2}{\sigma^2}} = \frac{1}{\lambda^2} f_{T_1}(\frac{t}{\lambda^2})$

d'où on déduit que  $T_2$  suit la même loi que  $\lambda^2 T_1$ .



Il est assez clair que  $T_{2+5} - T_2$  suit la même loi que  $T_2$  (on peut justifier par le calcul!)

Autrement dit, si  $Y_1, \dots, Y_n$  sont des v.a. indépendantes,  $Y_i$  suivant la même loi de  $T_{\lambda_i}$ , alors  $Y = Y_1 + \dots + Y_n$  suit la même loi que  $T_\lambda$  où  $\lambda = \lambda_1 + \dots + \lambda_n$ .

En utilisant la Remarque 2) on peut encore dire : si  $X_1, \dots, X_n$  sont des v.a. indépendantes, chaque  $X_i$  suivant la même loi que  $T_{\lambda_i}$ , alors  $\sum_{i=1}^n \lambda_i^2 X_i$  suit la même loi que  $(\sum_{i=1}^n \lambda_i)^2 T_1$ .

Preons  $\lambda_i^2 = \frac{1}{n} (\lambda_i)$ ;  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  suit la même loi que  $(\sqrt{n})^2 X_1 = n X_1$ . Donc la moyenne arithmétique de 100 observations indépendantes de  $T_2$  est 100 fois plus "dispersée" que  $T_1$  (elle correspond à une observation de  $T_{10}$ ).

Quel contraste avec la loi faible des grands nombres!

[Rappel: si  $\mu = E(X_1)$  existe,  $\bar{X}_n \xrightarrow{p} \mu$  cf. p.115]