

*UE4 : Biostatistiques*

---

# Chapitre 2 : Probabilités

Professeur Philippe CINQUIN

---

Année universitaire 2010/2011

Université Joseph Fourier de Grenoble - Tous droits réservés.

# Plan

- A) Introduction
- B) Statistiques descriptives
- C) Probabilités
- D) Estimation
- E) Intervalles de confiance
- F) Problème récapitulatif
- G) Résumé des objectifs

# C.1 Modèle Probabiliste

- Le modèle probabiliste décrit des expériences (*ou des systèmes*) **aléatoires**,
- ... c'est à dire dont le résultat (*ou l'état*) ne peut être prédit avec certitude.
- **Espace probabilisé = triplet ( $\Omega$ ,  $\mathcal{a}$ ,  $P$ )**
- **Construire successivement :**
  - l'univers  $\Omega$
  - la tribu des événements possibles  $\mathcal{a}$
- **la probabilité  $P$**

# *L'exemple du jet d'un dé cubique*

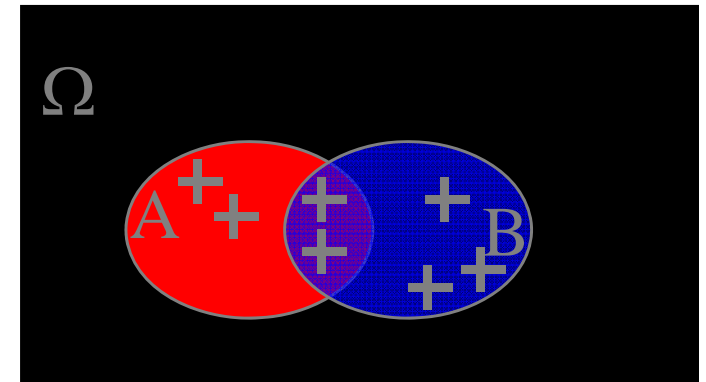
- **1) Construction de l'univers  $\Omega$** 
  - Associer une épreuve abstraite à une épreuve concrète
    - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Dénombrer  $\Omega$
- **2) Construire la tribu des événements possibles**
  - caractériser une propriété d'un événement concret
  - **l'événement possible** correspondant est une **partie** de  $\Omega$  : le sous-ensemble de  $\Omega$  dont les éléments vérifient cette propriété

***exemple :***

*l'événement concret « **une face paire apparaît** »  
se traduit par l'événement possible  $\{2, 4, 6\}$*

### 3) Construire la loi de probabilité

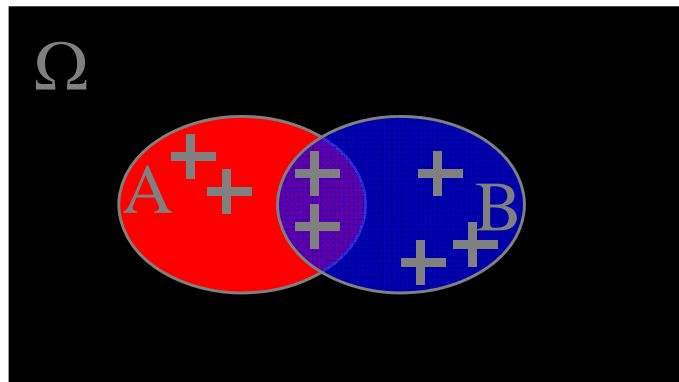
- événements immédiatement probabilisables (*par exemple, sortie de la face portant le “ 6 ”*)
- événements “ complexes ”
  - $P(\Omega) = 1$
  - $P(A^c) = 1 - P(A)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - **NB** :  $P(A \cap B) = 0$  si  $P(A \cap B) = \emptyset$  (A et B sont disjoints, ou encore incompatibles)



# Remarque sur « l'indépendance »

- Deux événements **A** et **B** sont dits « indépendants » si et seulement si

$$P(A \cap B) = P(A) \cdot P(B)$$



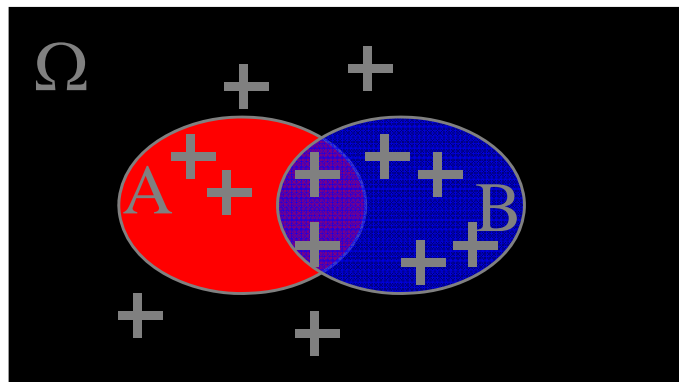
$$\text{Card}(\Omega) = 7$$

$$\text{Card}(A) = 4; p(A) = 4/7$$

$$\text{Card}(B) = 5; p(B) = 5/7$$

$$\text{Card}(A \cap B) = 2; p(A \cap B) = 2/7$$

$$p(A) \cdot P(B) = 20/49$$



$$\text{Card}(\Omega) = 12$$

$$\text{Card}(A) = 4; p(A) = 4/12 = 1/3$$

$$\text{Card}(B) = 6; p(B) = 6/12 = 1/2$$

$$\text{Card}(A \cap B) = 2; p(A \cap B) = 2/12 = 1/6$$

$$p(A) \cdot P(B) = 1/3 \times 1/2 = 1/6$$

# Un cas particulier

- **Si**
  - $\Omega$  fini ou infini dénombrable
  - tous les “ singletons ” sont immédiatement probabilisables
- **Alors** la plus petite tribu les possédant est l'ensemble des parties de  $\Omega$ , noté  $\mathbf{P}(\Omega)$ , et la probabilité d'un événement quelconque est la somme des probabilités des singletons qui le composent.
- Si les singletons sont équi-probables (formule des “ cas favorables ” sur les “ cas possibles ”),

$$P(A) = \text{cardinal}(A)/\text{cardinal}(\Omega)$$

*NB dans cette formule  $P(A)$  représente la probabilité de l'événement  $A$ , alors que  $\mathbf{P}(\Omega)$  dénote l'ensemble des parties de  $\Omega$ .*

## C.2 Variable aléatoire

- **1 Définition**

- Soit  $(\Omega, \mathcal{a}, P)$  un espace probabilisé
- $X : \Omega \rightarrow \mathbb{R}$  est une variable aléatoire si
$$\forall x \in \mathbb{R}, \{\omega \in \Omega, X(\omega) \leq x\} \in \mathcal{a}$$

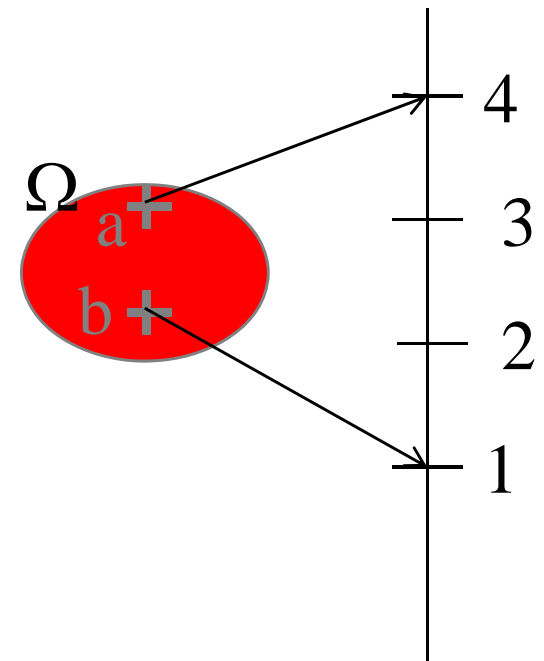
Par exemple,  $\Omega = \{a, b\}$   $X(a) = 4$ ;  $X(b) = 1$

$$\{\omega \in \Omega, X(\omega) \leq 5\} = \{a, b\}$$

$$\{\omega \in \Omega, X(\omega) \leq 3\} = \{b\}$$

$$\{\omega \in \Omega, X(\omega) \leq 0\} = \emptyset$$

- ***NB en pratique, toutes les variables qui nous intéresseront seront des variables aléatoires, vous n'aurez jamais à vérifier cette condition...***





# Exemples

- *somme des points obtenus lors du lancer de deux dés cubiques,*
- *nombre de filles dans un ensemble de fratries de 3 enfants,*
- *...*

- En pratique en médecine, toutes les grandeurs biologiques peuvent être considérées comme des variables aléatoires.
- *Intérêt des variables aléatoires : la manipulation de réels est plus facile que la manipulation d'éléments de la tribu  $\mathcal{A}$ ...*

## 2 Propriétés des Variables aléatoires

- Soient  $X$  et  $Y$  deux variables aléatoires. Les variables suivantes sont également des v. a. :
  - $Z = X + Y$
  - $Z = k X$  ( $k$  réel)
  - $Z = X^n$  ( $n$  entier)
  - $Z = XY$
- **3 Comment représenter le lien entre  $X$  et  $P$  ?**
- *La réponse dépend des caractéristiques de  $X$  ...*

## C.3 Variable Aléatoire discrète

- **Définition**

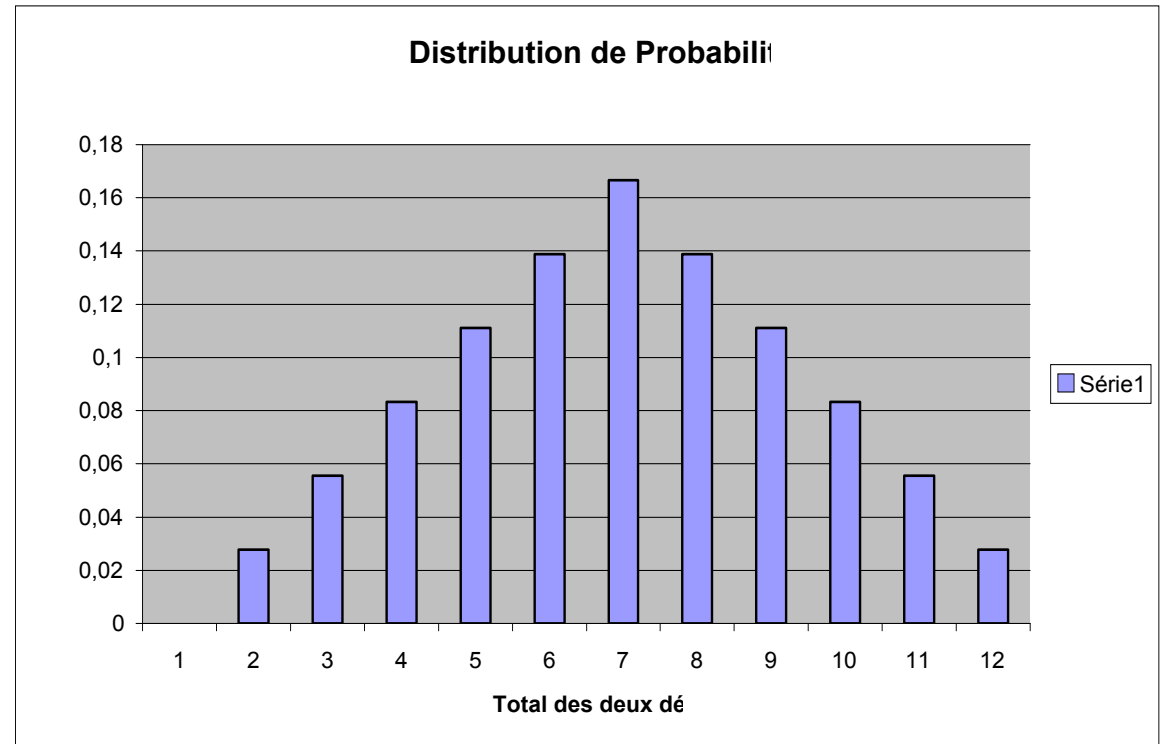
- Une v. a.  $X$  est **discrète** si l'ensemble des valeurs qu'elle prend est **fini** ou **dénombrable** (on peut indexer ses valeurs par des entiers).

## 2 Distribution des probabilités d'une v. a. discrète

- Soient  $\{x_i, i = 1, \dots, n\}$  les valeurs de  $X$ .
- **L'expression  $P(X=x_i)$  a un sens.**
- Soit  $a_i = \{\omega_k, X(\omega_k) = x_i\}$
- On notera  $f(x_i) = P(X=x_i) = P(a_i)$
- **$f$  est la fonction de distribution des probabilités de  $X$ .**

# Exemple : distribution des probabilités de la somme des points obtenus lors du lancer de deux dés cubiques

- $\Omega = \{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\} = \{(1,1), (1,2), \dots, (1,6), (2,1), (2,2), \dots, (2,6), \dots, (6,1), (6,2), \dots, (6,6)\}$
- $\text{Card}(\Omega) = 36$
- $X(d_1, d_2) = d_1 + d_2$
- $\{x_i, i = 1, \dots, n\} = \{2, 3, \dots, 12\}$
- Par exemple,  
 $a_6 = \{\omega_k, X(\omega_k) = 6\} = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$
- $f(6) = P(X=6) = P(a_6) = 5/36 = 0.14$



### 3) Paramètres caractérisant la distribution de probabilités d'une variable aléatoire discrète

- Paramètre = grandeur apportant une information résumée sur la loi de distribution d'une variable aléatoire

## 3.1 Espérance (Moyenne)

- $E(X) = \sum p_i x_i$ , avec  $p_i = P(X=x_i)$
- **Propriétés**
- Si pour tout  $i$ ,  $x_i = k$ , alors  $E(X) = k$
- $E(X + Y) = E(X) + E(Y)$
- $E(kX) = k E(X)$  ( $k$  réel)



## 3.2 Variance - Ecart-type

- $V(X) = E [(X - E(X))^2]$        $\sigma(X) = \sqrt{V(X)}$
- **Propriétés**
- *Formule de Huyghens*
- $V(X) = E [(X - E(X))^2] = V(X) = E (X^2) - E^2 (X)$
- *Pour tout k réel,*
  - $V(X + k) = V(X)$
- $V(kX) = k^2 V(X)$

# Remarques

- $E(X)$  et  $\sigma$  ont même “ *dimension* ” que  $X$  (ces grandeurs s’expriment numériquement avec la même *unité* que  $X$ ).
- **L’Espérance est un paramètre de *localisation*.**
- **Variance et Ecart-type sont des paramètres de *dispersion*.**

## C.4 Loi Binomiale

- **1 Epreuve de Bernoulli**
  - C'est une épreuve dont le résultat ne présente que deux éventualités incompatibles :
    - **Succès**, de probabilité  $p$ ,
    - **Echec**, de probabilité  $q = 1-p$

## 2 Distribution binomiale

- Réaliser  $n$  épreuves de Bernoulli :
  - indépendantes
  - $p$  reste constant pendant les  $n$  épreuves
- **Le nombre  $B$  de succès suit une loi binomiale.**

# Exemple : lancers d'une pièce truquée

- Une pièce est truquée. Sa probabilité de tomber sur face est de 0.6. On la lance 3 fois : quelle est la probabilité qu'elle tombe trois fois sur face ?
  - Premier lancer : Probabilité de succès = 0.6
  - Deuxième lancer (indépendant du premier!) : probabilité d'avoir 2 succès =  $0.6 \times 0.6$
  - Troisième lancer (indépendant des deux premiers!) : probabilité d'avoir 3 succès =  $0.6 \times 0.6 \times 0.6 = 0.6^3$
- Quelle est la probabilité qu'elle tombe une seule fois sur face ?
  - Choisir le lancer où elle tombera sur Face : 3 possibilités (1er, 2ème, 3ème)
    - La probabilité d'obtenir Face pour le lancer choisi est de 0.6
    - Il faudra également obtenir deux fois de suite Pile, soit une probabilité de  $0.4 \times 0.4$
    - En tout, compte tenu de l'indépendance des lancers, une fois qu'on a choisi le lancer où Face sortira, la probabilité que ce soit bien le seul où Face sortira est de  $0.6 \times 0.4 \times 0.4 = 0.6^1 \times 0.4^2$
    - Et donc la probabilité que la pièce tombe une seule fois sur Face est de  $3 \times 0.6^1 \times 0.4^2$

# Calcul de la loi de distribution de la loi binomiale

- Si on généralise ce raisonnement, pour obtenir la probabilité d'avoir  $k$  succès, il faut :
  - 1) Dénombrer le nombre de manières de choisir les  $k$  succès parmi les  $n$  lancers
  - 2) Calculer la probabilité d'obtenir les  $k$  succès et les  $(n-k)$  échecs aux épreuves retenues
- Le premier point correspond à la définition de  $C_n^k$ , nombre de combinaisons de  $k$  objets pris parmi  $n$  :
  - $C_n^k = n! / [k! (n-k)!]$  NB car  $k$  et  $(n-k)$  jouent des rôles symétriques
  - NB  $C_n^k = C_n^{n-k}$
  - En pratique, commencer la factorielle du numérateur, s'arrêter au  $k^{\text{ème}}$  terme, puis mettre au dénominateur la factorielle de  $k$ , par exemple

$$C_{20}^3 = 20! / [3! 17!] = 20 \times 19 \times 18 / (3 \times 2) \quad \text{car dans la formule}$$

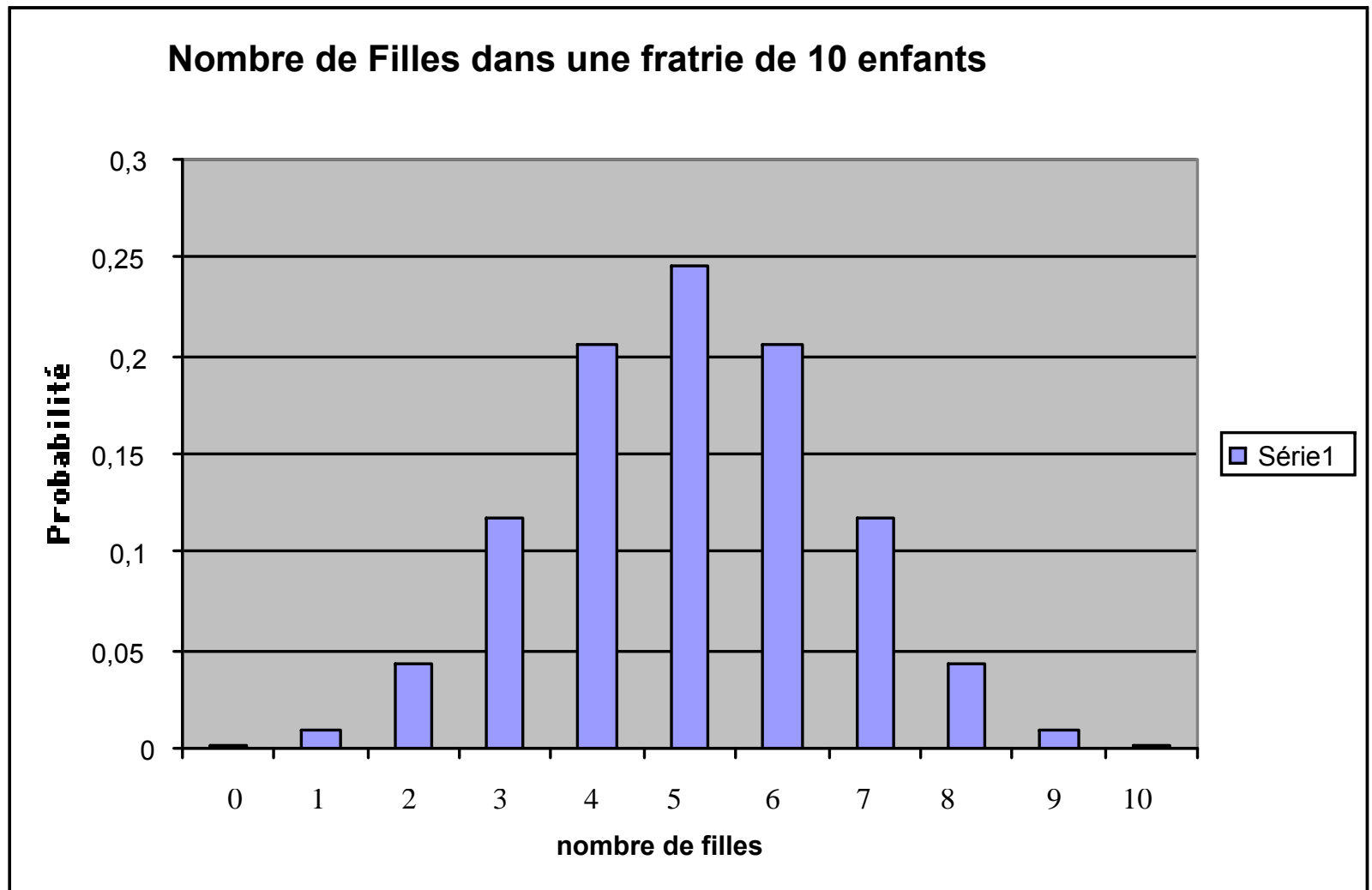
$$20 \times 19 \times 18 \times 17 \times 16 \times \dots \times 2 \times 1 / [(3 \times 2) \times (17 \times 16 \times \dots \times 2 \times 1)]$$

*17! disparaît, puisqu'il se trouve au numérateur ET au dénominateur]*

- Pour le deuxième point, compte tenu de l'indépendance des lancers, on obtient  $p^k q^{(n-k)}$
- Donc  **$B_k = P(B=k) = C_n^k p^k q^{(n-k)}$**

# Exemple : nombre de filles dans une fratrie de 10 enfant

Nombre de filles	Probabilité
0	0,00097656
1	0,00976563
2	0,04394531
3	0,1171875
4	0,20507813
5	0,24609375
6	0,20507813
7	0,1171875
8	0,04394531
9	0,00976563
10	0,00097656

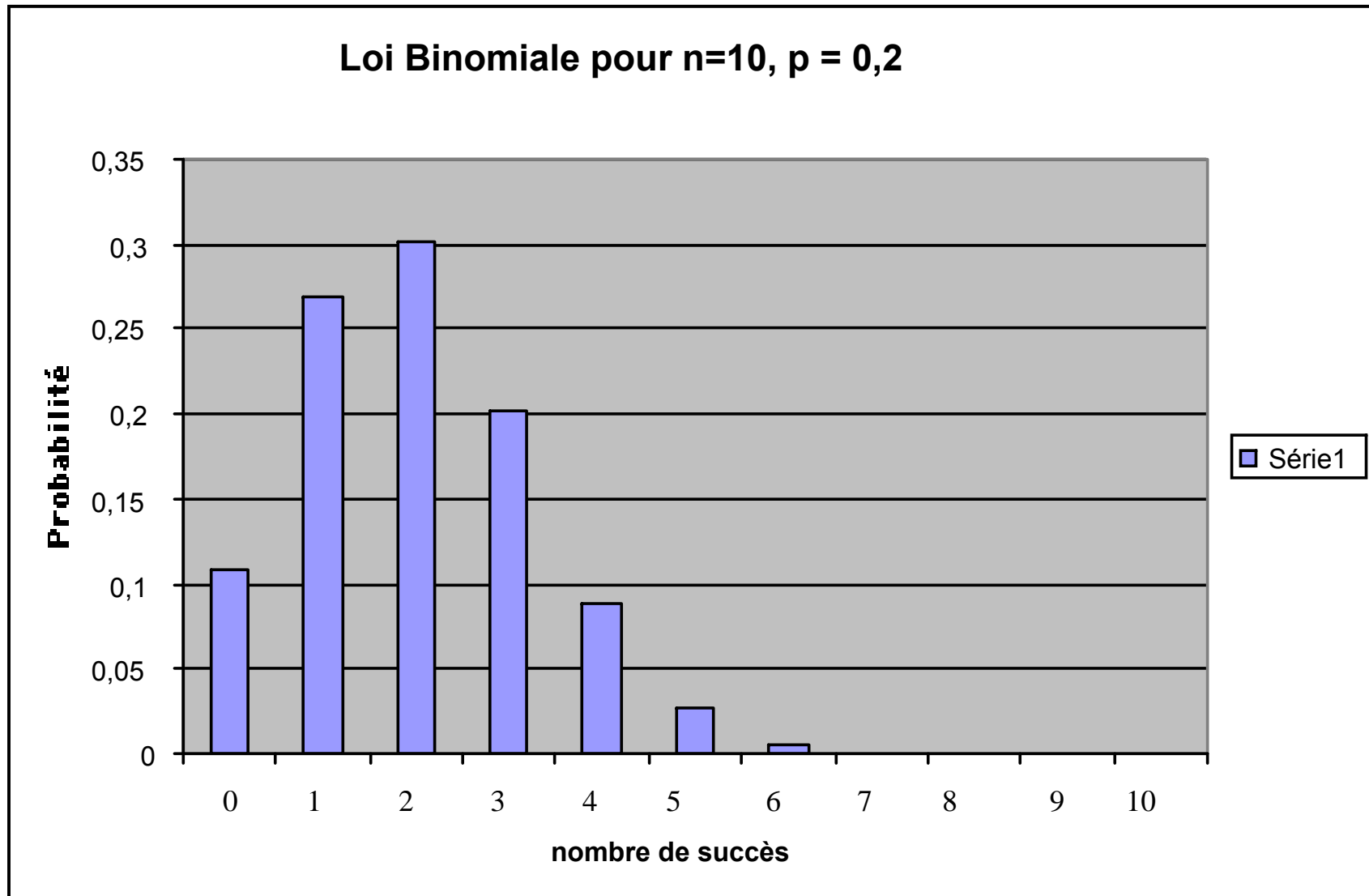


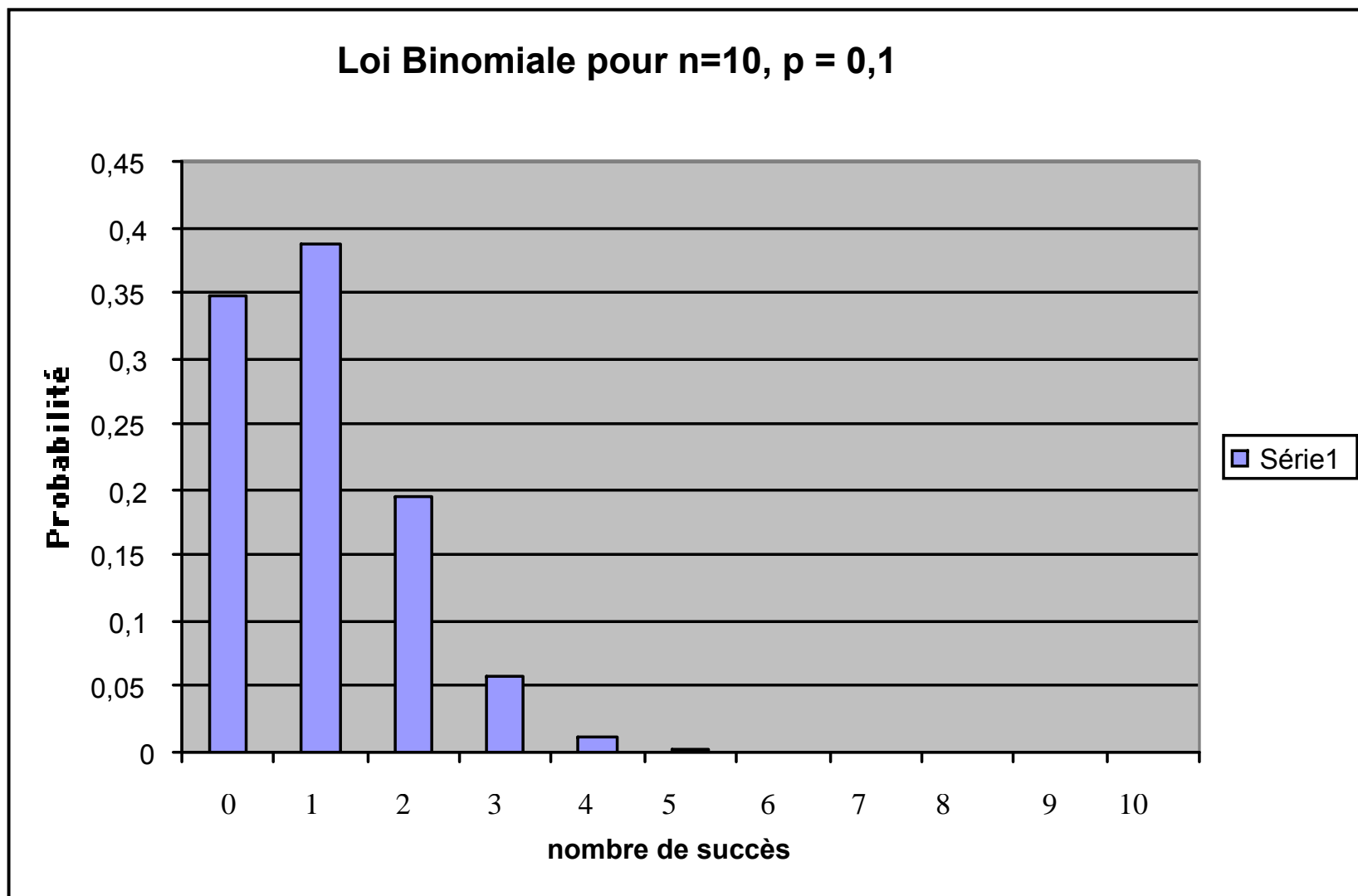
## *Exemple 2 Transmission dominante (un seul gène suffit) autosomique (non lié au chromosome X ou Y)*

- Soit  $p$  la probabilité de recevoir un allèle A,
  - $q = 1-p$  la probabilité de recevoir un allèle a.
- La transmission des gènes par les parents suit bien une loi binomiale.
  - L'enfant « tire » deux fois le gène (une fois de son père, une fois de sa mère)
  - Les deux tirages sont indépendants
  - La probabilité de tirer l'allèle A (succès) est constante ( $p$ )
- Pour être malade
  - Avoir les 2 gènes (le paternel ET le maternel) porteur de l'allèle a :  
0 succès, probabilité =  $C_2^0 p^0 q^2 = q^2$
  - Avoir 1 seul gène porteur de l'allèle a, 1 succès,  
probabilité =  $C_2^1 pq = 2pq$
  - Soit en tout  $2pq + q^2$
- Si  $q$  est faible (par exemple  $10^{-2}$ ), on peut approximer cette valeur par  $2pq = 2 \times 0.99 \times 0.01 = 0.0198 (\approx 0.02)$ , bonne approximation de  $2pq + q^2 = 0.0198 + 0.0001 = 0.0199$



# *Exemple 3 : allure de la loi de distribution pour $p$ faible*





- Si  $n \rightarrow \infty$  et  $p \rightarrow 0$ , alors que  $np \rightarrow \lambda$ , alors  $C_n^k p^k q^{(n-k)} \rightarrow e^{-\lambda} \lambda^k / k!$  On dit que la loi binomiale tend vers la loi de Poisson.

## *Loi de Poisson*

- Données fictives (non validées par mes collègues physiciens!) :
  - Soit un élément radioactif de masse atomique 131 tel que la probabilité de décomposition en une seconde d'un atome donné soit de  $1/(6.02 \cdot 10^{23})$ . Calculer la probabilité qu'un détecteur capable de détecter toute décomposition découvre au moins un tel événement en une seconde dans une masse de 131 g composée exclusivement de cet élément (on approximera le nombre d'Avogadro par  $6.02 \cdot 10^{23}$ ).
- On a bien une loi binomiale
  - Chaque seconde, chaque atome « décide » indépendamment de se décomposer; l'expérience est répétée  $n$  fois, avec
    - $n = 6.02 \cdot 10^{23}$
    - $p = 1 / (6.02 \cdot 10^{23})$
- Pour répondre à la question, il faut calculer :  
 $P(B=1) + P(B=2) + P(B=3) + \dots + P(B=6.02 \cdot 10^{23})$   
Euh, c'est long...

## Loi de Poisson (suite)

- On observe que  $np = \lambda = 1$ . On est bien dans la situation dite « de la loi de Poisson » :  $n \rightarrow \infty$  et  $p \rightarrow 0$ , alors que  $np \rightarrow \lambda$ , pour laquelle on peut approximer  $C_n^k p^k q^{(n-k)}$  par  $e^{-\lambda} \lambda^k / k!$
- Calculons  $P(B=0)$

$$P(B=0) = \left(1 - \frac{1}{6.02 \cdot 10^{23}}\right)^{6.02 \cdot 10^{23}}$$

– Que l'on peut approximer par  $e^{-\lambda} \lambda^k / k! = e^{-1} 1^0 / 0! = e^{-1} = 0.37$

- Or,

$$\begin{aligned} P(B=1) + P(B=2) + P(B=3) + \dots + P(B=6.02 \cdot 10^{23}) &= 1 - P(B=0) \\ &= 1 - 0.37 = 0.63 \end{aligned}$$

# 3 Paramètres de distribution de la loi binomiale

- $E(B) = np$
- $V(B) = npq$

# Mentions légales

L'ensemble de ce document relève des législations française et internationale sur le droit d'auteur et la propriété intellectuelle. Tous les droits de reproduction de tout ou partie sont réservés pour les textes ainsi que pour l'ensemble des documents iconographiques, photographiques, vidéos et sonores.

Ce document est interdit à la vente ou à la location. Sa diffusion, duplication, mise à disposition du public (sous quelque forme ou support que ce soit), mise en réseau, partielles ou totales, sont strictement réservées à l'université Joseph Fourier de Grenoble.

L'utilisation de ce document est strictement réservée à l'usage privé des étudiants inscrits en 1<sup>ère</sup> année de Médecine ou de Pharmacie de l'Université Joseph Fourier de Grenoble, et non destinée à une utilisation collective, gratuite ou payante.