

PROBABILITÉS

19, 22 et 23 décembre

PROGRAMME DE LA SÉANCE

1. Premières définitions
2. Premiers indicateurs
3. Probabilités
4. Evènements
5. Variables aléatoires
6. Bernoulli et loi binomiale
7. Echantillonnage

1. PREMIÈRES DÉFINITIONS

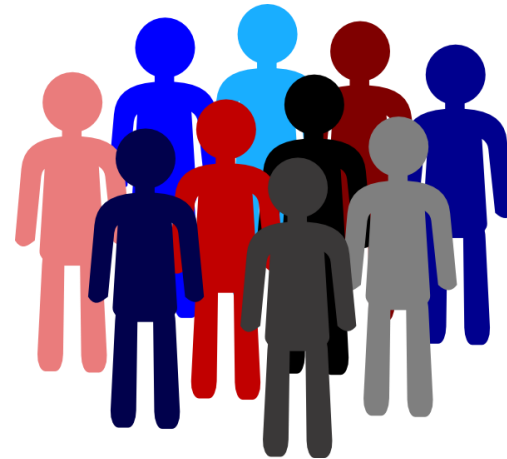
SÉRIE STATISTIQUE

- Une **série statistique** est une liste de "mesures", généralement obtenues par l'observation et/ou l'expérimentation.
- Exemples :
 - Liste des âges des élèves
 - Liste des prénoms des élèves
 - Liste des résultats obtenus en lançant 200 dés à six faces
 - Liste des résultats obtenus en jouant à pile ou face 10000 fois
 - etc.



POPULATION ET INDIVIDUS

- La **population** d'une série statistique est l'ensemble des éléments étudiés.
- Les membres de cette population sont appelés **individus**, même s'il ne s'agit pas forcément d'être humains !
- Exemples
 - Elèves d'une classe
 - 200 dés à six faces
 - 10000 lancers de pièce
 - etc.



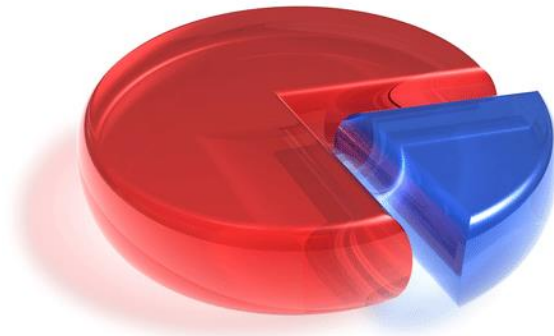
CARACTÈRE

- Le **caractère** d'une série statistique est la propriété étudiée pour chaque individu
- On distingue plusieurs types de caractères :
 - Un caractère **quantitatif discret** ne peut prendre qu'un nombre fini de valeurs numériques
 - Un caractère **quantitatif continu** peut prendre une infinité de valeurs numériques
 - Un **caractère qualitatif** peut prendre des valeurs non numériques
- Remarque : On peut observer plusieurs caractères sur une même population
- Exemples pour une population d'élèves :
 - Couleur des yeux
 - Age (en années)
 - Moyenne générale
 - Gaucher
 - Taille en centimètres, arrondie à l'inférieur

EFFECTIF ET FRÉQUENCE

- L'**effectif** d'une valeur est le nombre d'individus de la population qui présente cette valeur.
 - En d'autres termes, c'est le nombre de fois où cette valeur apparaît dans la série statistique
- La **fréquence** d'une valeur est le quotient de l'effectif de cette valeur par l'effectif total :

$$frequency(x) = \frac{effectif(x)}{effectif\ total}$$



EFFECTIF ET FRÉQUENCE CUMULÉS

- Pour un caractère quantitatif (discret ou continu), on note généralement x_i le i -ème terme de la série statistique
- L'**effectif cumulé croissant** (resp. décroissant) de x_i est la sommes des effectifs des valeurs inférieures ou égales (resp. supérieures ou égales) à x_i .
- La **fréquence cumulée croissante** (resp. décroissante) de x_i est la sommes des fréquences des valeurs inférieures ou égales (resp. supérieures ou égales) à x_i .

EXEMPLE

- Population : classe
- Caractère : âge (en années)
- Série :

x_i	
Effectif	
Fréquence	
Effectif cumulé croissant	
Fréquence cumulée décroissante	

2. PREMIERS INDICATEURS

MOYENNE

- La **moyenne** d'une série statistique indique la valeur qu'aurait chacun des éléments de l'ensemble s'ils étaient tous identiques, sans changer leur somme.
- Si on note n_i l'effectif de la valeur x_i , et N l'effectif total, on a

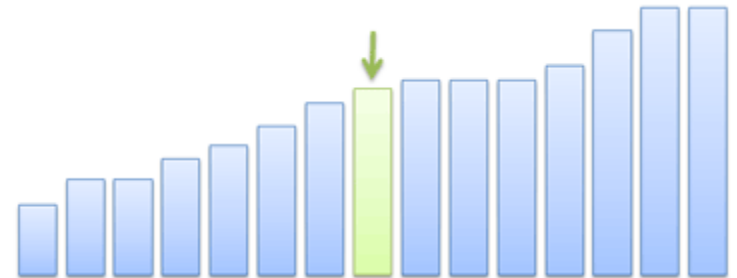
$$\text{moyenne} = \frac{n_1 \times x_1 + n_2 \times x_2 + \cdots + n_k \times x_k}{N}$$

$$\text{moyenne} = \sum_{i=1}^k \frac{n_i \times x_i}{N}$$

MÉDIANE

- La **médiane** d'une série statistique est la valeur qui permet de partager la série en deux parties avec le même nombre d'éléments.
- Formellement, la médiane d'une série statistique est une valeur **M** telle que
 - Au moins 50% des individus ont une valeur du caractère inférieure ou égale à **M**
 - Au moins 50% des individus ont une valeur du caractère supérieure ou égale à **M**

- **Question** : Pourquoi *Au moins 50 %* ?



- **Remarque** : La moyenne et la médiane peuvent-être très différentes

MÉDIANE

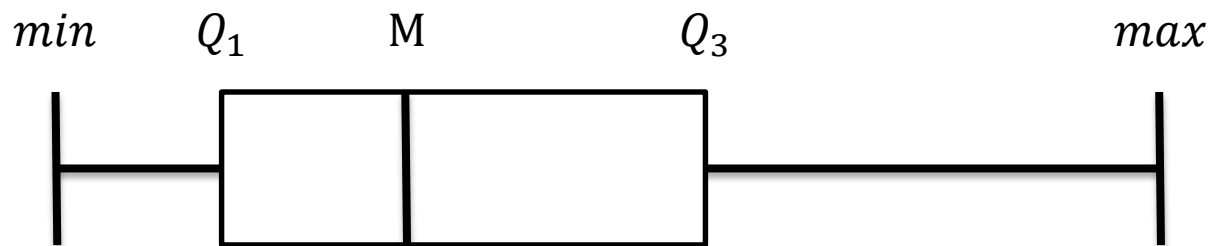
- Pour déterminer la médiane, une méthode simple consiste à trier la liste par ordre croissant
- Si la série est de taille impaire ($N = 2p + 1$), la médiane est le terme de rang $p + 1$.
- Si la série est de taille paire ($N = 2p$), la médiane est le terme de rang p .
- **Remarque** : En pratique, il existe des méthodes plus rapides (mais plus complexes) pour le calcul de la médiane

QUARTILES

- Le **premier quartile** est la plus petite valeur Q_1 appartenant à la série et telle qu'au moins 25 % des individus de la population ont une valeur supérieure ou égale à Q_1 .
- Le **troisième quartile** est la plus petite valeur Q_3 appartenant à la série et telle qu'au moins 75 % des individus de la population ont une valeur supérieure ou égale à Q_3 .
- La médiane est le deuxième quartile.

DIAGRAMME EN BOITE

- Pour représenter une série statistique, on utilise généralement un **diagramme en boîte** :



- Remarque : On parle aussi de **boîtes à moustaches**.



ÉTENDUE, ÉCART ET INTERVALLE INTERQUARTILE

- L'**étendue** d'une série statistique est la différence entre la plus grande valeur et la plus petite valeur : **$\max - \min$**
- L'**écart interquartile** est la différence entre le troisième quartile et le premier quartile : **$Q_3 - Q_1$**
- L'**intervalle interquartile** est l'intervalle **$[Q_1; Q_3]$**

EXEMPLE

- Population : classe
- Caractère : âge (en années)
- Série :
- Diagramme en boîte :
- Etendue :
- Ecart interquartile :
- Intervalle interquartile :

3. PROBABILITÉS

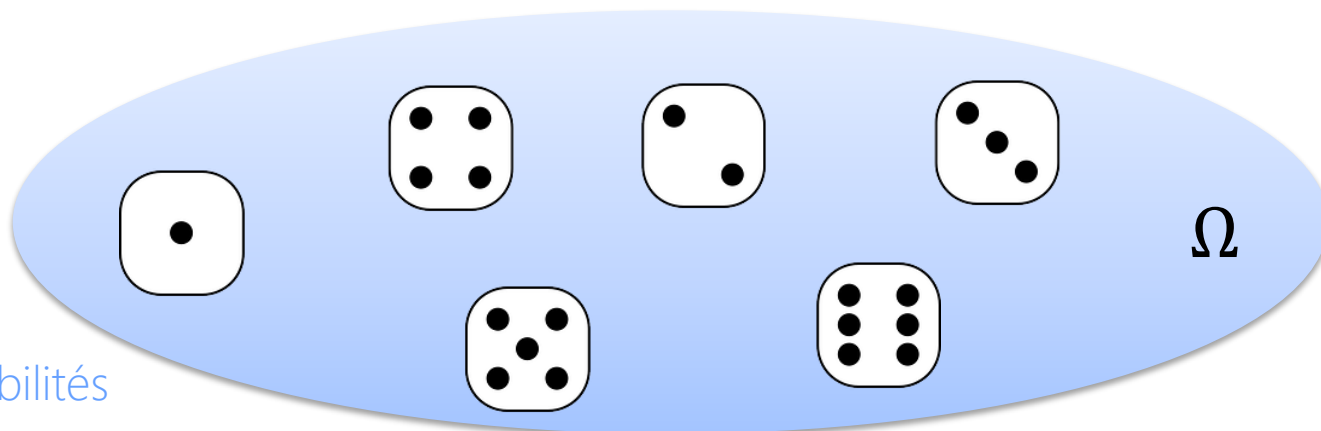
PROBABILITÉS ET STATISTIQUES

- **Question** : *Existe-t-il une différences entres les statistiques et les probabilités ?*
- Les statistiques s'intéressent aux résultats d'expériences passées, dont on étudie les différentes caractéristiques : on parle d'approche **empirique**.
- Les probabilités s'intéressent à des évènements à venir, dont on essaye d'estimer les résultats (ou plus exactement les chances que chacun de ces résultats arrive) : c'est une approche plus **théorique**.

EXPÉRIENCE ALÉATOIRE

- Une **expérience aléatoire** est une expérience dont on ne peut pas prévoir à l'avance le résultat avec certitude.
- Les résultats d'une expérience aléatoire sont appelées des **éventualités** (ou des **issues**), et sont généralement notés $e_1, e_2, \dots, e_n$.
- L'ensemble des éventualités, noté Ω , est appelé l'**univers** :

$$\Omega = \{e_1, e_2, \dots, e_n\}$$



EXEMPLES

- Lancer un dé à 6 faces
- Lancer une pièce de monnaie
- Choisir au hasard une boule dans un sac opaque qui contient des 10 boules blanches et 10 boules noires.
- Lancer simultanément 2 dés à 6 faces

PROBABILITÉS

- A chaque éventualité e_i , on associe une **probabilité** p_i , qui représente ses chances d'apparaître.
- Cette probabilité est comprise entre 0 et 1 (inclus) :
$$0 \leq p_i \leq 1$$
- La somme des probabilités de toutes les éventualités donne 1 :
$$p_1 + p_2 + \cdots + p_n = 1$$
- La donnée de ces p_i est appelée une **loi de probabilité**.

LOI EQUIPROBABLE

- Si toutes les éventualités d'une expérience aléatoire ont les mêmes chances d'apparaître, on parle de **loi équiprobable**.

- On a alors, pour chaque éventualité e_i , la probabilité

$$p_i = \frac{1}{n}$$

(où n désigne le nombre total d'éventualités)

- **Remarque** : En pratique, ce cas est plus une approximation qu'une juste représentation de la réalité.

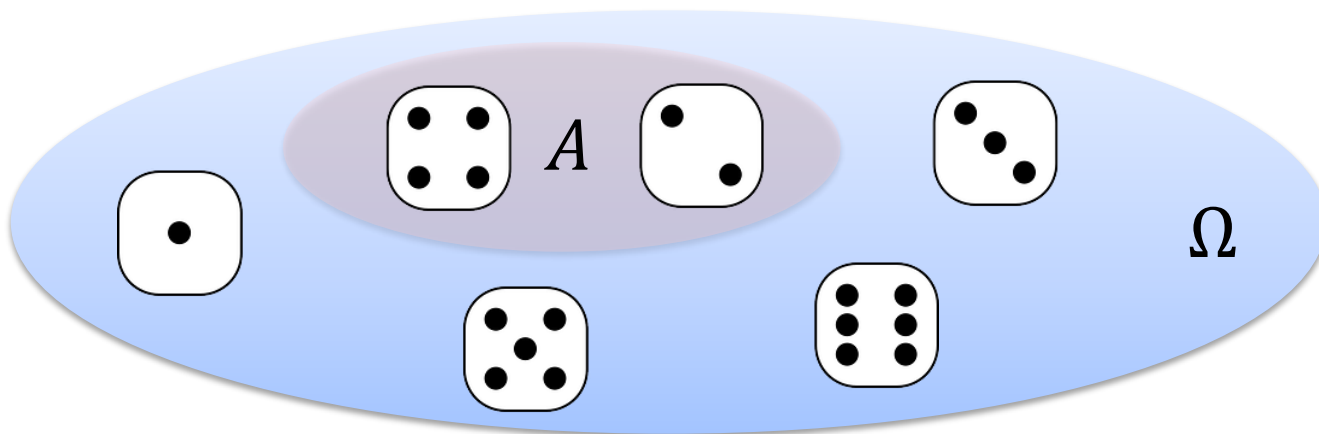
EXEMPLES

- Lancer un dé à 6 faces
- Lancer une pièce de monnaie
- Choisir au hasard une boule dans un sac opaque qui contient des 10 boules blanches et 10 boules noires.
- Lancer simultanément 2 dés à 6 faces

4. EVÈNEMENTS

DÉFINITION

- Un **évènement** A est une partie de l'univers Ω : c'est donc un ensemble d'éventualités ($A \subset \Omega$).
- L'ensemble vide $\emptyset$ est l'évènement **impossible**.
- L'évènement Ω (qui contient toutes les éventualités) est l'évènement **certain**.



EXEMPLES

- Lancer un dé à 6 faces
- Lancer une pièce de monnaie
- Choisir au hasard une boule dans un sac opaque qui contient des 10 boules blanches et 10 boules noires.
- Lancer simultanément 2 dés à 6 faces

PLUSIEURS ÉVÈNEMENTS

Si A et B sont deux événements d'un univers Ω ,

- L'évènement $A \cap B$ est l'évènement " A et B ".
- L'évènement $A \cup B$ est l'évènement " A ou B ".
- L'évènement $\bar{A}$ est l'évènement " $\text{contraire de } A$ " (ou " $\text{non } A$ ")
- On dit que A et B sont **incompatibles** s'ils ne peuvent pas se réaliser en même temps (c'est-à-dire que $A \cap B = \emptyset$)

PROBABILITÉ D'UN ÉVÈNEMENT

- On définit la **probabilité d'un événement** comme la somme des probabilités des éventualités qui le composent :

$$p(A) = \sum_{e_i \in A} p_i$$

- Premières propriétés :

Pour tout évènement A , $0 \leq p(A) \leq 1$

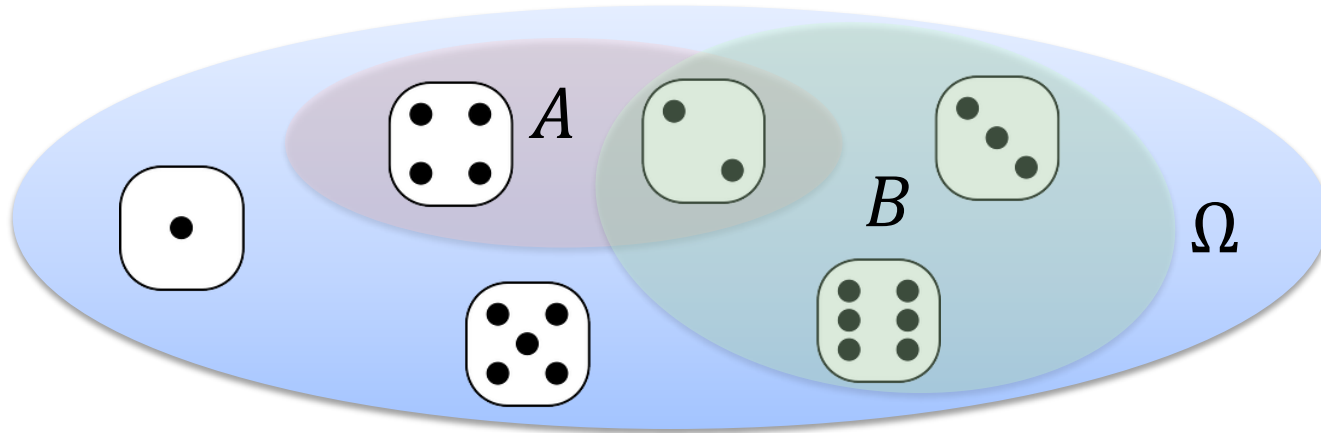
$$p(\emptyset) = 0$$

$$p(\Omega) = 1$$

QUELQUES PROPRIÉTÉS INTÉRESSANTES

- Lien entre union et intersection :

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$



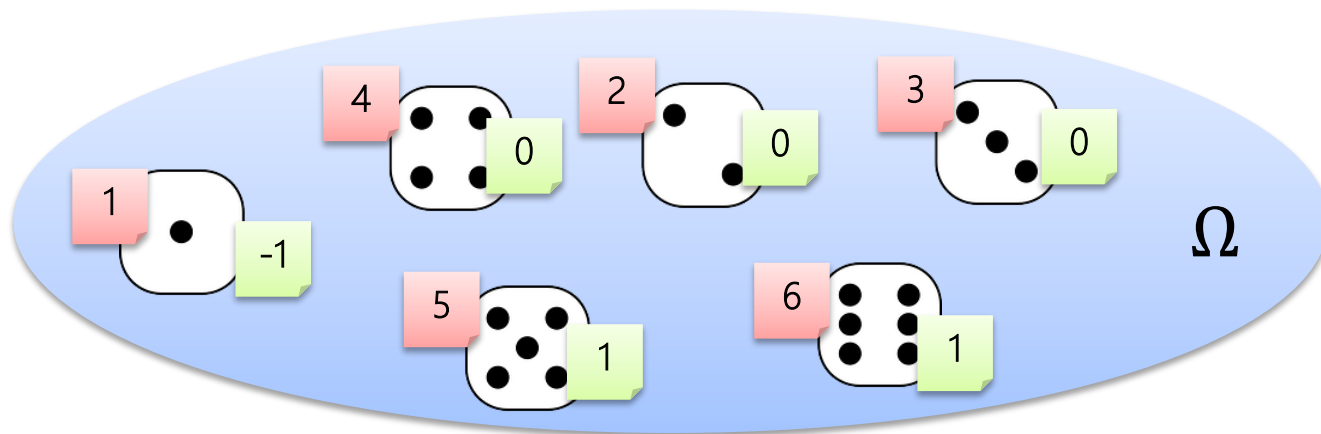
- Contraire

$$p(\overline{A}) = 1 - p(A)$$

5. VARIABLES ALÉATOIRES

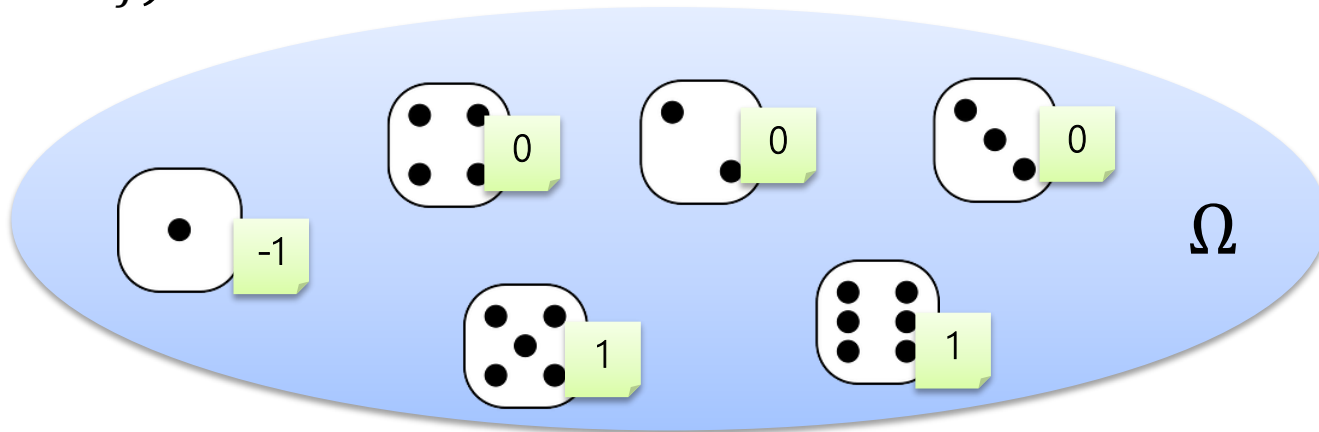
VARIABLE ALÉATOIRE

- Une **variable aléatoire** est une fonction définie sur Ω et à valeurs dans $\mathbb{R}$.
- Les valeurs prises par une variable aléatoire X sont généralement notées $x_1, x_2, \dots, x_k$
- La valeur de X est appelée le **gain**.



LOI DE PROBABILITÉ D'UNE VARIABLE ALÉATOIRE

- La **loi de probabilité** d'une variable aléatoire est un tableau qui donne, pour chaque gain possible x_j , la probabilité $p(X = x_j)$



gain x_j	-1	0	1
$p(X = x_j)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

ESPÉRANCE

Notons $p_j = p(X = x_j)$

- L'**espérance** mathématique d'une variable aléatoire X est définie par

$$E(X) = p_1 \times x_1 + p_2 \times x_2 + \cdots + p_k \times x_k = \sum_{j=1}^k p_j \times x_j$$

- Cette grandeur représente le gain que l'on peut "espérer en moyenne" dans le cadre de cette expérience.

VARIANCE ET ECART-TYPE

- On définit ensuite la **variance** de X par

$$V(X) = p_1 \times (x_1 - E(X))^2 + \dots + p_k \times (x_k - E(X))^2$$

$$V(X) = \sum_{j=1}^k p_j \times (x_j - E(X))^2$$

- Et l'**écart-type** de X par

$$\sigma(X) = \sqrt{V(X)}$$

- Ces grandeurs caractérisent la capacité de cette variable aléatoire à prendre des valeurs plus ou moins éloignées de l'espérance.

6. BERNOULLI ET LOI BINOMIALE

JACQUES BERNOULLI

- a.k.a. Jacob Bernoulli, Jakob Bernoulli
- Mathématicien et physicien suisse du 17^e siècle
- De nombreux travaux :
 - Séries infinies

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \sum_{i=1}^n \frac{1}{i}$$

- Constante mathématique e

$$\lim_{n \rightarrow +\infty} \left(1 + \frac{1}{n}\right)^n = e$$

- Probabilités



EPREUVE DE BERNOULLI

- Une **épreuve de Bernoulli** est une expérience aléatoire comportant exactement deux issues.
- Ces deux issues sont appelées :
 - le succès S
 - l'échec $\bar{S}$
- On note p la probabilité du succès, et on a donc une probabilité $1 - p$ pour l'échec
 - On parle d'épreuve de Bernoulli de **paramètre p**

VARIABLE ET LOI DE BERNOULLI

- Pour une épreuve de Bernoulli de paramètre p , on peut définir une variable aléatoire X telle que :

$$X(S) = 1$$

$$X(\bar{S}) = 0$$

- On a donc

x	0	1
$p(X = x)$	$1 - p$	p

- On dit alors que
 - X est une **variable de Bernoulli** de paramètre p
 - X suit la **loi de Bernoulli** de paramètre p

VARIABLE DE BERNOULLI : ESPÉRANCE ET VARIANCE

- Espérance

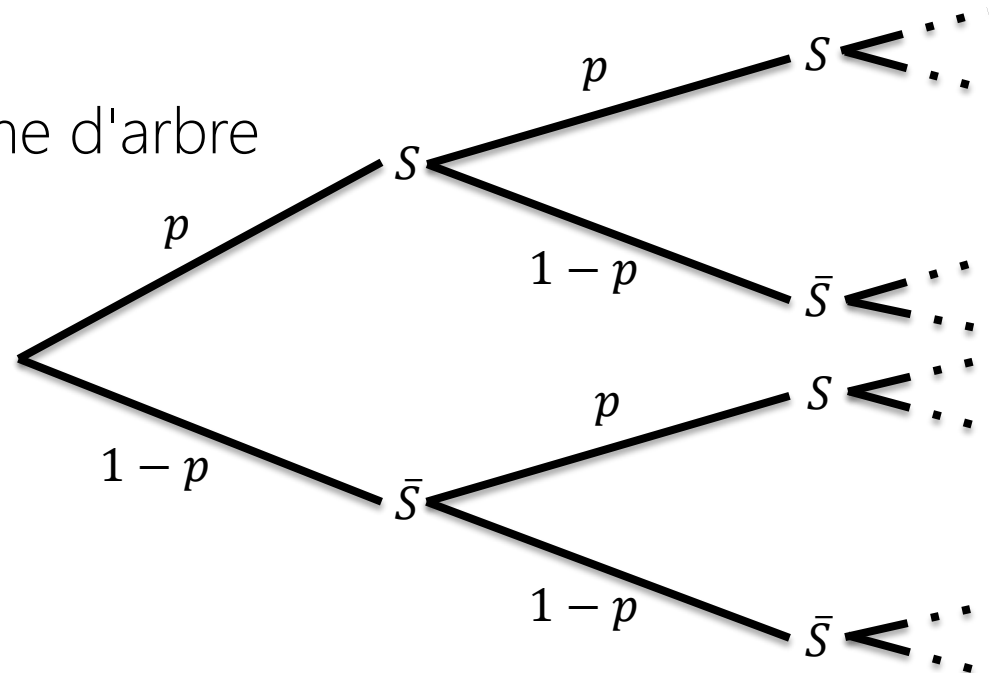
$$E(X) = p$$

- Variance

$$V(X) = p(1 - p)$$

SCHÉMA DE BERNOULLI

- Un **schéma de Bernoulli** de paramètres n et p est une expérience aléatoire consistant à répéter n fois de façon indépendante une épreuve de Bernoulli de paramètre p .
- Contrainte d'indépendance : l'issue d'une épreuve ne doit pas influencer sur les autres
- Représentation sous forme d'arbre



EXEMPLES

Quelles sont les situations qui correspondent à un schéma de Bernoulli ? Quels en sont les paramètres ?

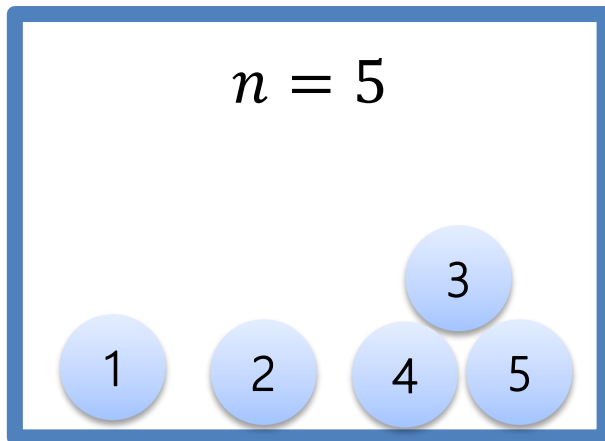
- Jouer à pile ou face 10 fois de suite
- Choisir trois cartes au hasard dans un jeu de 32 cartes, en espérant trouver un cœur
- Tirer 5 balles au hasard dans une urne contenant 10 balles blanches et 20 balles rouges

COEFFICIENTS BINOMIAUX

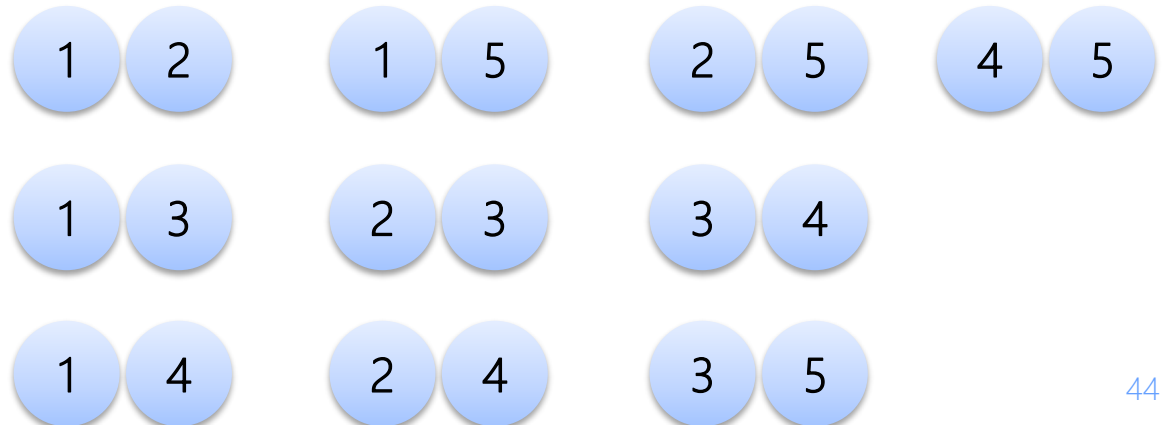
- On considère un schéma de Bernoulli de paramètres n et p .
- Pour tout entier k tel que $0 \leq k \leq n$, on note $\binom{n}{k}$ le nombre de chemin dans l'arbre comportant exactement k succès lors des n répétitions.
- Le nombre $\binom{n}{k}$ est appelé un **coefficient binomial**
- $\binom{n}{k}$ se lit " k parmi n ".

COEFFICIENTS BINOMIAUX

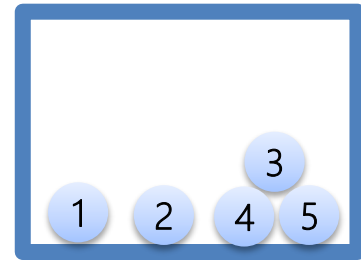
- On peut voir $\binom{n}{k}$ comme le nombre de façons de choisir k éléments parmi n , sans considération d'ordre.
- Exemple :
 - On a une urne avec n balles numérotées de 1 à n
 - $\binom{n}{k}$ désigne le nombre de façon de choisir k balles au hasard, sans considération d'ordre



$$\binom{5}{2} = 10$$



COEFFICIENTS BINOMIAUX – PROPRIÉTÉS



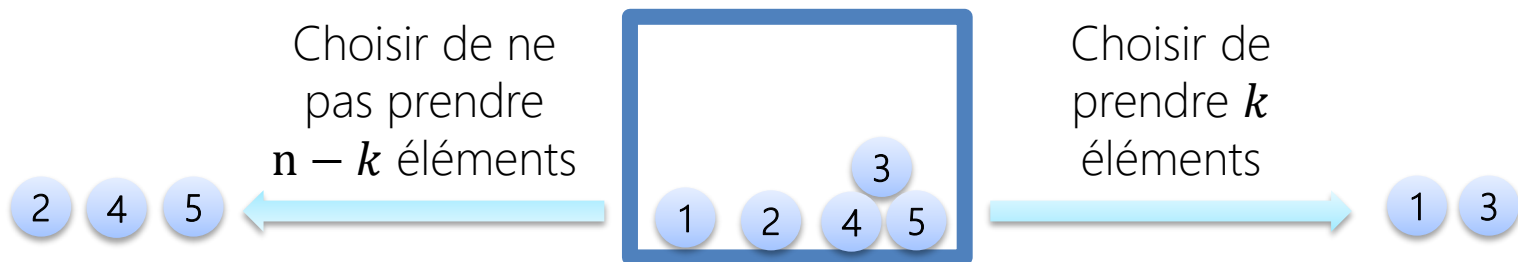
- Pour tout entier $n \geq 0$, on a $\binom{n}{0} = 1$
 - Nombre de façons de choisir 0 balle parmi n : 1 (on n'en prend aucune)
- Pour tout entier $n \geq 0$, on a $\binom{n}{n} = 1$
 - Nombre de façons de choisir n balles parmi n : 1 (on les prend toutes)
- Pour tout entier $n \geq 1$, on a $\binom{n}{1} = n$
 - Nombre de façons de choisir 1 balle parmi n : n (chaque solution est définie par la balle qu'on prend)
- Pour tout entier $n \geq 1$, on a $\binom{n}{n-1} = n$
 - Nombre de façons de choisir $n - 1$ balles parmi n : n (chaque solution est définie par la balle qu'on ne prend pas)

COEFFICIENTS BINOMIAUX – PROPRIÉTÉS

- Pour tous entiers naturels n et k tels que $0 \leq k \leq n$, on a

$$\binom{n}{k} = \binom{n}{n-k}$$

A chaque façon de choisir k éléments (à conserver) correspond une façon de choisir $n - k$ éléments (à ne pas conserver)



Il y a donc autant de façons de choisir k éléments (à conserver) que de façons de choisir $n - k$ éléments (à ne pas conserver).

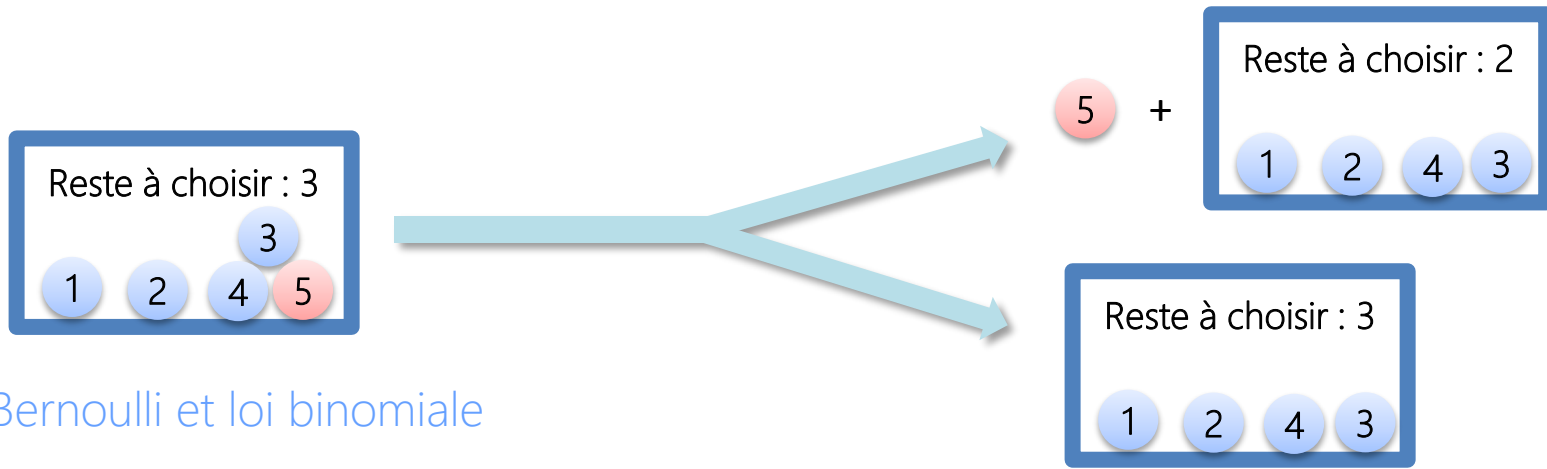
COEFFICIENTS BINOMIAUX – PROPRIÉTÉS

- Pour tous entiers naturels n et k tels que $0 \leq k \leq n$, on a

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

Lorsqu'on choisit k éléments parmi n :

- soit on choisit l'élément n :
 - on choisit alors $k-1$ éléments parmi les $n-1$ restants, soit $\binom{n-1}{k-1}$ possibilités.
- soit on ne choisit pas l'élément n :
 - on choisit alors k éléments parmi les $n-1$ restants, soit $\binom{n-1}{k}$ possibilités.



COEFFICIENTS BINOMIAUX – PROPRIÉTÉS (RÉSUMÉ)

- Par convention, $\binom{0}{0} = 1$
- Pour tout entier $n \geq 0$, on a $\binom{n}{0} = 1$ et $\binom{n}{n} = 1$
- Pour tout entier $n \geq 1$, on a $\binom{n}{1} = n$ et $\binom{n}{n-1} = n$
- Pour tous entiers naturels n et k tels que $0 \leq k \leq n$, on a
$$\binom{n}{k} = \binom{n}{n-k}$$
- Pour tous entiers naturels n et k tels que $0 \leq k \leq n$, on a
$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

EXERCICES

Calculer les nombres suivants

1. $\binom{5}{0}$

2. $\binom{5}{1}$

3. $\binom{42}{41}$

4. $\binom{6}{5}$

5. $\binom{4}{2}$

6. $\binom{6}{7}$

COEFFICIENTS BINOMIAUX – TRIANGLE DE PASCAL

- Ces propriétés permettent de tracer le **triangle de Pascal** :

$n \backslash k$	0	1	2	3	4	5
0	1					
1	1	1				
2	1	2	1			
3	1	3	3	1		
4	1	4	6	4	1	
5	1	5	10	10	5	1

LOI BINOMIALE

- Pour un schéma de Bernoulli de paramètres n et p , on peut définir une variable aléatoire X associée au nombre de succès lors de ces n épreuves :

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- La loi de probabilité de la variable aléatoire X est appelée la **loi binomiale** de paramètres n et p , que l'on note $\mathcal{B}(n, p)$
- Propriétés :
 - Espérance : np
 - Variance : $np(1 - p)$
 - Ecart-type : $\sqrt{np(1 - p)}$

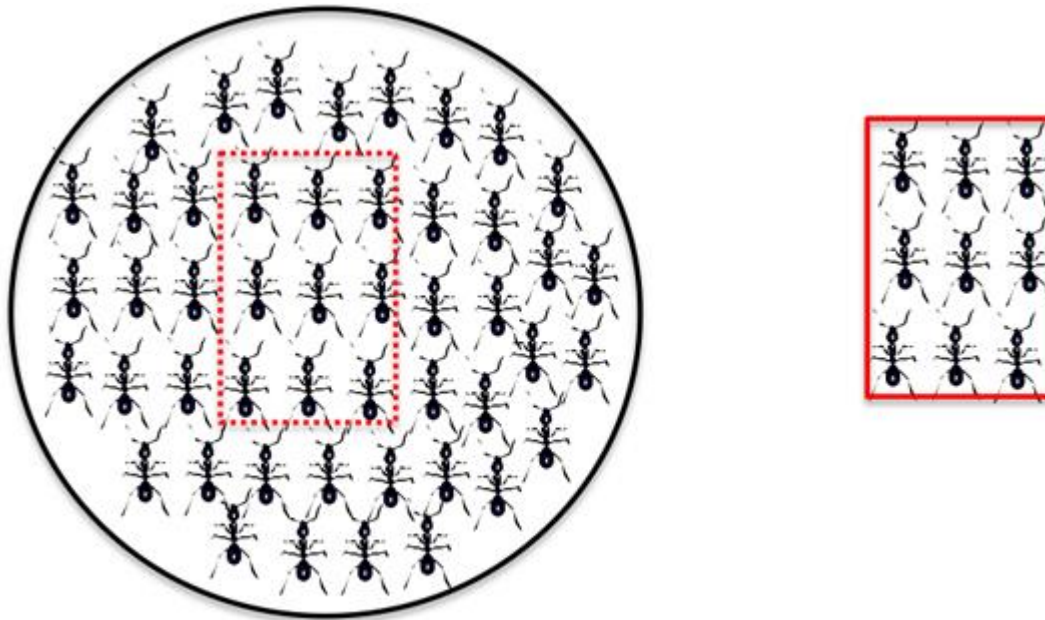
LOI BINOMIALE - EXEMPLE

- On dispose d'une urne contenant
 - 5 boules bleues
 - 10 boules rouges
- On définit une épreuve de Bernoulli comme le fait de tirer au hasard une boule bleue.
- **Question 1.** Si X_1 est la variable aléatoire associée à cette épreuve, quelles sont son espérance et sa variance ?
- On répète cette expérience 2 fois (avec remise)
- **Question 2.**
 - Quelle est la loi binomiale associée à ce schéma de Bernoulli ?
 - Quelles sont les valeurs des $P(X_2 = k)$?
 - Quelles sont son espérance et sa variance ?

8. ÉCHANTILLONNAGE

ECHANTILLON

- Un **échantillon de taille n** est une sélection de n individus choisis au hasard dans une population.



- Un échantillon est choisi pour étudier un ou plusieurs caractères sur une population

AVEC OU SANS REMISE

- Dans un échantillon **sans remise**, les n individus sont choisis simultanément : on a la garantie qu'on obtient n individus différents.
- Dans un échantillon **avec remise**, les n individus sont choisis simultanément : on a la garantie qu'on obtient n individus différents.
- Remarque : Si l'effectif global est nettement supérieur à la taille de l'échantillon, les deux méthodes donnent des résultats très similaires.

FRÉQUENCE ET FLUCTUATION

- On peut alors définir la **fréquence** d'un caractère dans cet échantillon. Pour une valeur x , on a ainsi

$$f(x) = \frac{\text{nombre d'individus avec la valeur } x \text{ dans l'échantillon}}{\text{taille de l'échantillon}}$$

- Si l'on effectue plusieurs échantillonnages de même taille sur une même population, on risque d'obtenir des fréquences légèrement différentes : on parle de **fluctuation d'échantillonnage**.

INTERVALLE DE FLUCTUATION AU SEUIL DE 95%

- **Hypothèse** : On suppose connue la proportion p d'un caractère dans une population donnée.
- **Théorème** : On prélève un échantillon de taille n dans cette population, et on note f la fréquence du caractère dans l'échantillon.

Si $0,2 \leq p \leq 0,8$ et si $n \geq 25$, alors dans au moins 95% des cas, f appartient à l'intervalle suivant

$$I = \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$$

- I est appelé l'intervalle de fluctuation au seuil de 95%

ATTENTION

- Ce théorème part du postulat qu'on connaît la probabilité p du caractère dans la population totale.
- En pratique, on peut donc se servir de ce théorème pour estimer la fiabilité d'une telle hypothèse sur p :
 - Si f appartient à l'intervalle de confiance, on ne peut pas en conclure grand-chose
 - Par contre, si f n'appartient pas à l'intervalle de confiance, il est fort probable que l'hypothèse sur p soit erronée
- Attention, il n'y a donc pas de résultat "absolu" : juste une indication sur la fiabilité d'une hypothèse.

EXEMPLE

- On dispose de 1000 pièces qu'on suppose parfaitement équilibrées pour le jeu "pile ou face" :

$$p(Pile) = p(Face) = 0,5$$

- On choisit au hasard 100 pièces, et on les lance chacune une fois. On obtient la distribution de fréquence suivante :

Valeur	<i>Pile</i>	<i>Face</i>
Effectif	75	25
Fréquence	0,75	0,25

- Or l'intervalle de confiance pour la valeur *Pile* est

$$I = \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right] = \left[0,5 - \frac{1}{\sqrt{100}}; 0,5 + \frac{1}{\sqrt{100}} \right] = [0,4; 0,6]$$

- On peut donc remettre en question l'hypothèse initiale "au risque de 5%" (on avait 95% de chances de tomber dans l'intervalle de confiance)

BONUS

LE PARADOXE DES ANNIVERSAIRES

- Question : Combien de personnes faut-il réunir dans une pièce pour avoir une chance sur deux que deux soient nées le même jour de l'année ?
- Réponse : 23
- Question : Combien de personnes faut-il réunir dans une pièce pour avoir 99% de chances que deux soient nées le même jour ?
- Réponse : 57

LE PARADOXE DES ANNIVERSAIRES

- Remarque : pour simplifier, on suppose toutes les années sont non bissextiles
- La probabilité qu'un événement E arrive peut s'écrire :
$$p(E) = \frac{\text{Nombre de cas qui vérifient } E}{\text{Nombre total de cas possibles}}$$
- S'il y a n personnes dans la pièce, le nombre total de cas possibles est 365^n (365 dates possibles pour chaque personne)
- Notre événement E correspond aux cas de figures où il y a au moins deux personnes dans la pièce qui sont nées le même jour de l'année
- On va s'intéresser à l'événement contraire $\bar{E}$, c'est-à-dire les cas de figures où toutes les personnes de la pièce sont nées des jours de l'année différents.
- Si on note $\bar{E}$ l'événement contraire de l'événement E , on a
$$p(E) = 1 - p(\bar{E})$$

LE PARADOXE DES ANNIVERSAIRES

- On a donc

$$p(\bar{E}) = \frac{\text{Nombre de cas qui vérifient } \bar{E}}{\text{Nombre total de cas possibles}} = 1 - p(E)$$

- Estimons le nombre de cas qui vérifient $\bar{E}$ s'il y a n personnes dans la pièce :
 - Il y a 365 dates possibles pour le premier invité
 - Il reste 364 dates pour le second invité
 - Il reste 363 dates pour le troisième invité
 - ...
 - Il reste $365 - n + 1$ dates pour le n -ième invité
 - Soit finalement un nombre de cas qui vérifient $\bar{E}$ qui vaut

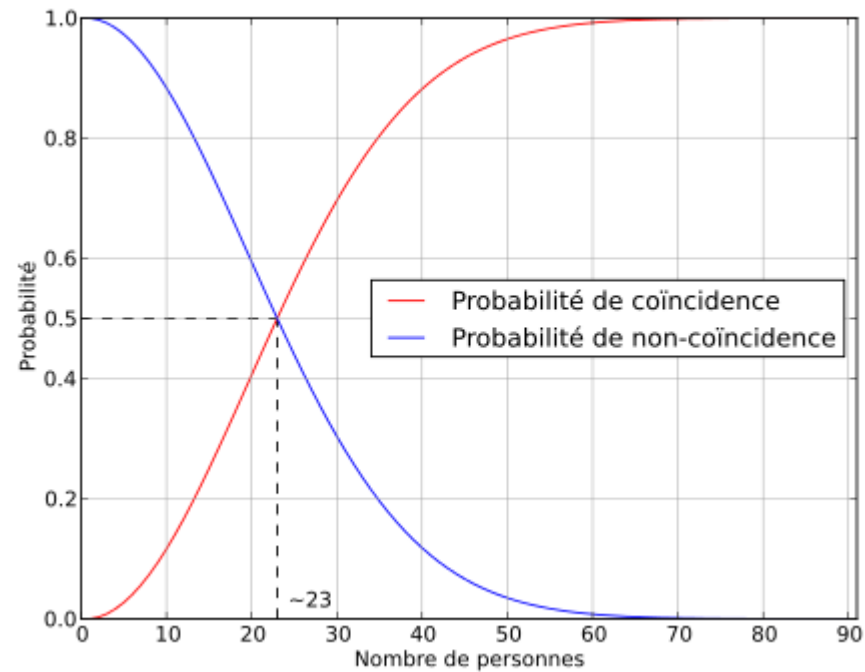
$$365 \times 364 \times 363 \times \dots \times (365 - n + 1) = \frac{365!}{(365 - n)!}$$

- D'où finalement

$$p(\bar{E}) = \frac{\left(\frac{365!}{(365-n)!}\right)}{365^n} \quad \text{et} \quad p(E) = 1 - \frac{\left(\frac{365!}{(365-n)!}\right)}{365^n}$$

LE PARADOXE DES ANNIVERSAIRES

n	5	10	20	23	30	50	57	100	366
$p(E)$	2,7 %	11,7 %	41,1 %	50,7 %	70,6 %	97 %	99 %	99,9997 %	100 %



MERCI !

andre-lovichi@ecole-alsacienne.org