# Eléments de Probabilités Discrètes

**Roberto M. Amadio**

Université Paris Diderot (Paris 7)

AA 2013-2014

# Introduction

## Calculer des probabilités ≈ Mesurer

– Soit $\Omega$ un ensemble qui est un modèle mathématique des **résultats** possibles d'une certaine **expérience**.

– Par exemple, si l'expérience est le **jet de deux dés** on pourrait avoir :

$$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$$

– Un premier objectif consiste à **mesurer** (certains) sous-ensembles de $\Omega$ qu'on appelle **événements**. Par exemple, les jets dont la somme donne un nombre pair.

– Dans la suite on appelle la mesure associée à un événement la (mesure de) **probabilité**.

– Les éléments de $\Omega$ correspondent aux sous-ensembles singletons. S'ils sont mesurables on parle d'**événements élémentaires**.

## Exemples

– Quelle est la probabilité que parmi $n$ personnes il y en aient au moins 2 qui ont le **même anniversaire** ?

– Alice et Bob **lancent alternativement un dé** à 6 faces numérotées $1, \ldots, 6$. Le premier qui tire un 6 gagne. Si Alice commence, quelle est la probabilité qu'elle gagne ?

– En tirant un couple $(x, y)$ de **réels** dans l'intervalle $[0, 1]$ quelle est la probabilité que $(x, y)$ soit dans le cercle de centre $(0, 0)$ et rayon 1 ?

- Quelle est la probabilité qu'un message qui contient les mots *Gift certificate* soit un **spam** ?

- Le fait de connaître le **texte chiffré** d'un message affecte-il notre connaissance du **texte clair** qui pourrait lui correspondre ?

- En supposant que toutes les entrées ont la même probabilité, quelle est le **coût moyen de l'algorithme de tri rapide** (*quicksort*) ?

- En cas de collision sur un canal de communication (type *Ethernet*) quelle est la **meilleure stratégie pour retransmettre** le message ?

## Note méthodologique

Le calcul des probabilités est une sorte de **logique quantitative** qui s'applique dans des nombreux domaines de l'informatique (ainsi qu'en physique, biologie, économie,...)

La **méthodologie** consiste en :

1. Construction d'un **modèle** (techniquement, un **espace de probabilité**) qui est cohérent avec les données/hypothèses dont on dispose.
2. Application de **méthodes analytiques** et/ou de **méthodes de simulation** pour determiner certaines caracteristiques du modèle : la **moyenne** (ou espérance), l'**écart type** (ou déviation),...

## Terminologie

– Calcul des **Probabilités** : construction d'un modèle mathématique de ce qui est mesurable.

– **Statistique** : interpréter des données disponibles (exemple sondage) en utilisant un modèle probabiliste.

# Plan

**Notions de base**

- Exemples de calcul des probabilités.
- Définition axiomatique éspace de probabilité.
- Indépendance et Probabilité conditionnelle.
- Variables aléatoires discrètes et leur espérance (ou moyenne).
- Déviation de la moyenne et Bornes sur la déviation.

**Exemples d'application en informatique**

- Conception et analyse d'algorithmes et de systèmes.
- Cryptographie, codage,...

**Compléments** (en fonction du temps et de l'intérêt...)

- Notions sur les suites (pseudo-)aléatoires.
- Exemples de simulation de systèmes à événements discrets.
- Quelques résultats spectaculaires (sans preuve mais illustrés par des simulations).

## Prérequis

**Combinatoire** Permutations, combinaisons, théorème du binôme.

**Analyse** Séries, Limites, Dérivation et développement de Taylor, Intégrale de Riemann.

**NB** On dispose de **systèmes de calcul formel en ligne** qui sont utiles pour vérifier les calculs. Exemple : `http://www.wolframalpha.com/widgets/gallery/?category=math`.

**Algorithmique et programmation** Récurrences, algorithmes de base (tri rapide par exemple), structures de données, mise en oeuvre dans un langage de programmation (Java).

## Références pour le cours

– Ces **transparents** et planches de **travaux dirigés**. : `http://www.pps.univ-paris-diderot.fr/ amadio/Ens/Proba/`

– **Initiation aux probabilités**. Sheldon Ross. Presses Polytechniques. *Un classique, niveau élémentaire, très complet. Existe aussi en anglais.*

- **Introduction au Calcul des Probabilités**. Charles Suquet. Université Lille, Notes de cours. `http://math.univ-lille1.fr/ suquet/ens/ICP/Cmd060902.pdf`. *Niveau comparable au livre de Ross, couvre une grande partie du cours.*

- **Aleátoire : une introduction aux probabilités.** S. Méléard *et al.*. `https://class.coursera.org/probas-001/class/index`. *Un cours en ligne développé à l'Ecole Polytechnique sur la base d'un cours de première année ; un peu plus avancé que ce qu'on va faire.*

– **Mathematics for computer science.** T. Leighton *et al.* `http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-042j-mathematics-for-computer-science-fall-2010/` *Un cours en ligne développé à l'MIT. La dernière partie du cours est une introduction au calcul des probabilités dont le niveau est comparable à celui de ce cours.*

– Michael Mitzenmacher, Eli Upfal. **Probability and computing : randomized algorithms and probabilistic analysis**. Cambridge University Press. *Un livre plus avancé orienté vers l'application du calcul des probabilités à la conception et à l'analyse d'algorithmes.*

## Crédit

Ces 'transparents' fournissent une **trace des sujets traités en cours** et ils sont basés en grande partie sur les textes cités (entre autres).

**NB** La **lecture** du livre de Ross ou des notes de cours de Souquet (ou d'un autre texte d'un niveau équivalent) est fortement conseillée.

## Contrôle des connaissances

– Une note CC de **contrôle continu** basée sur 3-4 tests en classe annoncés en cours une sémaine à l'avance).

– Une note E d'**examen**.

– **Note finale** :

$$N = max(E, (2E + CC)/3)$$

**NB** Pendant les tests et pendant l'examen les documents et les dispositifs éléctroniques sont interdits.

# Exemples de calcul des probabilités

## Problème

On dispose d'un ensemble $\Omega$ de **résultats**, d'un ensemble $A \quad 2^{\Omega}$ d'**événements** et d'une **probabilité** $P : A \to [0, 1]$.

Quelles sont les **propriétés** attendues ?

**Pour** $A$ Si $\{A_i\}_{i \ \mathbf{N}} \quad A$ alors

$$\bigcup_{i \ \mathbf{N}} A_i, \quad \bigcap_{i \ \mathbf{N}} A_i, \quad A_i^c \quad A$$

Stable par rapport aux opérations ensemblistes **dénombrables**

**Pour** $P$ Si $\{A_i\}_{i \ \mathbf{N}} \quad A$ et $i = j$ implique $A_i \cap A_j = \quad$ alors

$$P(\bigcup_{i \ \mathbf{N}} A_i) = \Sigma_{i \ \mathbf{N}} P(A_i)$$

Additivité **dénombrable**

## Rappel sur les séries

Soit $\Sigma_{n \in \mathbf{N}} a_n$ une série dans $\mathbf{R}$.

– La série **converge** si

$$\lim_{n \to +\infty} S_n \in \mathbf{R} \qquad \text{où } S_n = \Sigma_{i=0,\ldots,n} a_i$$

– La série **converge absolument** si la série des valeurs absolues $\Sigma_{n \in \mathbf{N}} |a_n|$ converge.

Si une série converge absolument alors elle converge à une valeur qui ne dépend pas de l'ordre d'addition

**NB** Pour une série de nombres **non-négatifs** ceci implique que la convergence à une valeur est invariante par permutation des éléments de la série.

## Exemple : jet d'une pièce

- Résultats : $\Omega = \{0, 1\}$.
- Evénements : $A = 2^{\Omega}$.
- Probabilité : $P(\{1\}) = p$, $P(\{0\}) = (1 - p)$, $0 \leq p \leq 1$.

On peut définir $P : \{0, 1\} \to \mathbf{R}$ telle que :

$$P(i) = \begin{cases} p & \text{si } i = 1 \\ (1 - p) & \text{si } i = 0 \end{cases}$$

On appelle cette fonction **loi de Bernoulli**.

## Cas $\Omega$ fini avec probabilité uniforme

– Si $\Omega$ est **fini** et la mesure est **uniforme** alors **mesurer revient à compter**.

– Dans ce cas, on peut prendre $A = 2^{\Omega}$ et la probabilité d'un événement $A \quad \Omega$ est simplement :

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

où $X$ est la **cardinalité** de $X$.

## Rappels de Combinatoire

– Le nombre de **sous-ensembles** de cardinalité $k$ d'un ensemble de cardinalité $n$ :

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

**Conventions :** $0! = 1$, $\binom{n}{k} = 0$ si $n < 0$ ou $k < 0$ ou $n < k$.

– Une **identité utile** pour $n \geq k+1$ (identité du triangle) :

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

– Une autre **identité utile** pour $1 \leq k \leq n$ :

$$\binom{n}{k} = \frac{1 \cdots k(k+1) \cdots k + (n-k)}{1 \cdots k \; 1 \cdots (n-k)} = (1+\frac{k}{1})(1+\frac{k}{2}) \cdots (1+\frac{k}{(n-k)})$$

– Indentité **binomiale** :

$$(a+b)^n = \Sigma_{k=0,\ldots,n} \binom{n}{k} a^k b^{(n-k)}$$

– Le nombre de **partitions** d'un ensemble de cardinalité $n$ en ensembles de cardinalité $n_1, \ldots, n_k$ (donc $n = n_1 + \cdots + n_k$) :

$$\frac{n!}{n_1! \cdots n_k!} = \binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdots \binom{n-n_1-\cdots-n_{k-1}}{n_k}$$

## Exemple : tirage sans ou avec remise

Une urne contient $N$ boules dont $N_1$ sont **rouges** et $(N - N_1)$ **vertes**.

**Tirage avec remise** On repète $n$ fois l'opération suivante : on prend une boule on note sa couleur et on la remet dans l'urne.

Quelle est la probabilité $P(k)$ de tirer $k$ boules rouges $(0 \leq k \leq n)$ ?

$$P(k) = \binom{n}{k} \frac{N_1^k (N - N_1)^{n-k}}{N^n}$$

## Argument pour le tirage avec remise

– Soient $1, \ldots, N_1$ les boules **rouges** et $N_1 + 1, \ldots, N$ les boules **vertes**.

– Un suite de $n$ tirages avec remise est une **fonction** :

$$f : \{1, \ldots, n\} \to \{1, \ldots, N\}$$

On a donc $N^n$ tirages possibles.

– Un tirage avec $k$ boules rouges est determiné par :
  – Un **sous-ensemble de cardinalité** $k$ de $n$.
  – Une **fonction** de $\{1, \ldots, k\}$ à $\{1, \ldots, N_1\}$.
  – Une **fonction** de $\{1, \ldots, n - k\}$ à $\{1, \ldots, N - N_1\}$.

  On a donc :

$$\binom{n}{k} \cdot N_1^k \cdot (N - N_1)^{(n-k)}$$

## Loi binomiale

Si l'on pose $p = \frac{N_1}{N}$, on a pour $0 \leq k \leq n$ :

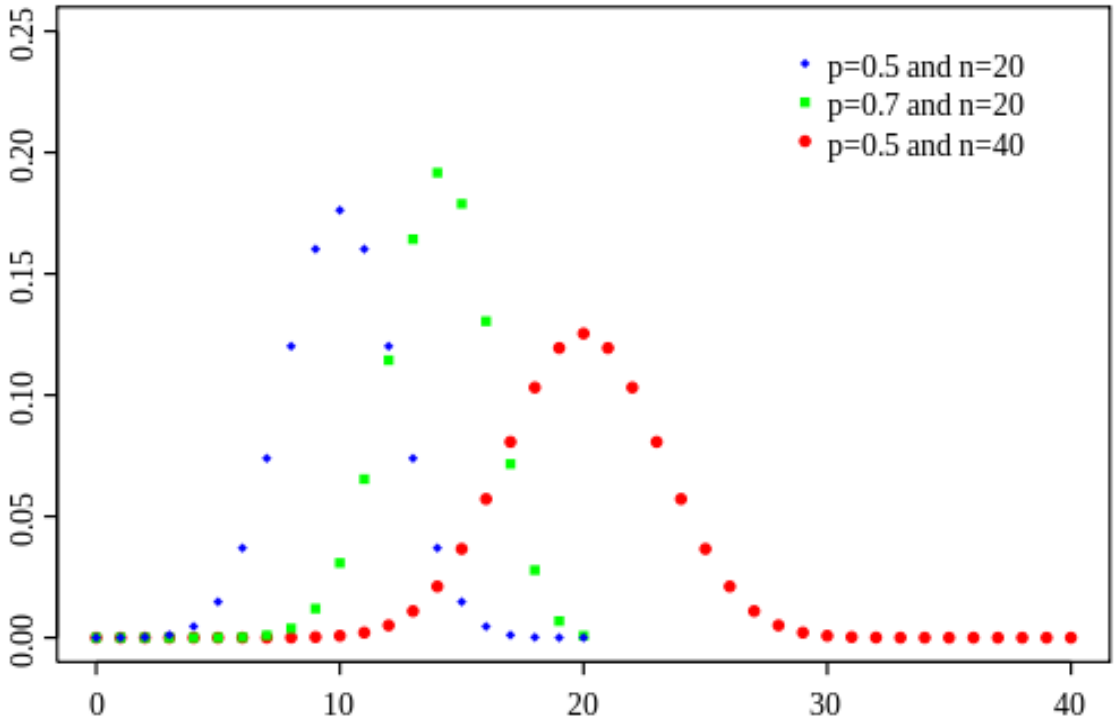
$$P(k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

Cette fonction s'appelle **loi binomiale** avec paramètres $p$ $[0,1]$ et $n$ $\mathbf{N}$.

L'**identité binomiale** assure :

$$\Sigma_{k=0,\ldots,n} P(k) = 1$$

$P(k)$ est la probabilité de tirer $k$-fois pile dans $n$ tirages.


p=0.5 and n=20
p=0.7 and n=20
p=0.5 and n=40
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40

## Tirage sans remise

On prend $n \leq N$ boules dans l'urne. Quelle est la probabilité $Q$ de tirer $k$ boules rouges ?

$$Q(k) = \frac{\binom{N_1}{k} \cdot \binom{N - N_1}{n-k}}{\binom{N}{n}}$$

Notez que tirer $n$ boules dans un coup ou tirer $n$ fois 1 boule sans remise revient au même.

## Argument pour tirage sans remise

- Soient $1, \ldots, N_1$ les boules **rouges** et $N_1 + 1, \ldots, N$ les boules **vertes**.

- Un tirage est un sous-ensemble de cardinalité $n$ de $\{1, \ldots, N\}$.

- Un tirage avec $k$ boules rouges et ($n - k$ boules vertes) est determiné par :
  - Un sous-ensemble de cardinalité $k$ de $\{1, \ldots, N_1\}$.
  - Un sous-ensemble de cardinalité $n - k$ de $\{1, \ldots, N - N_1\}$.

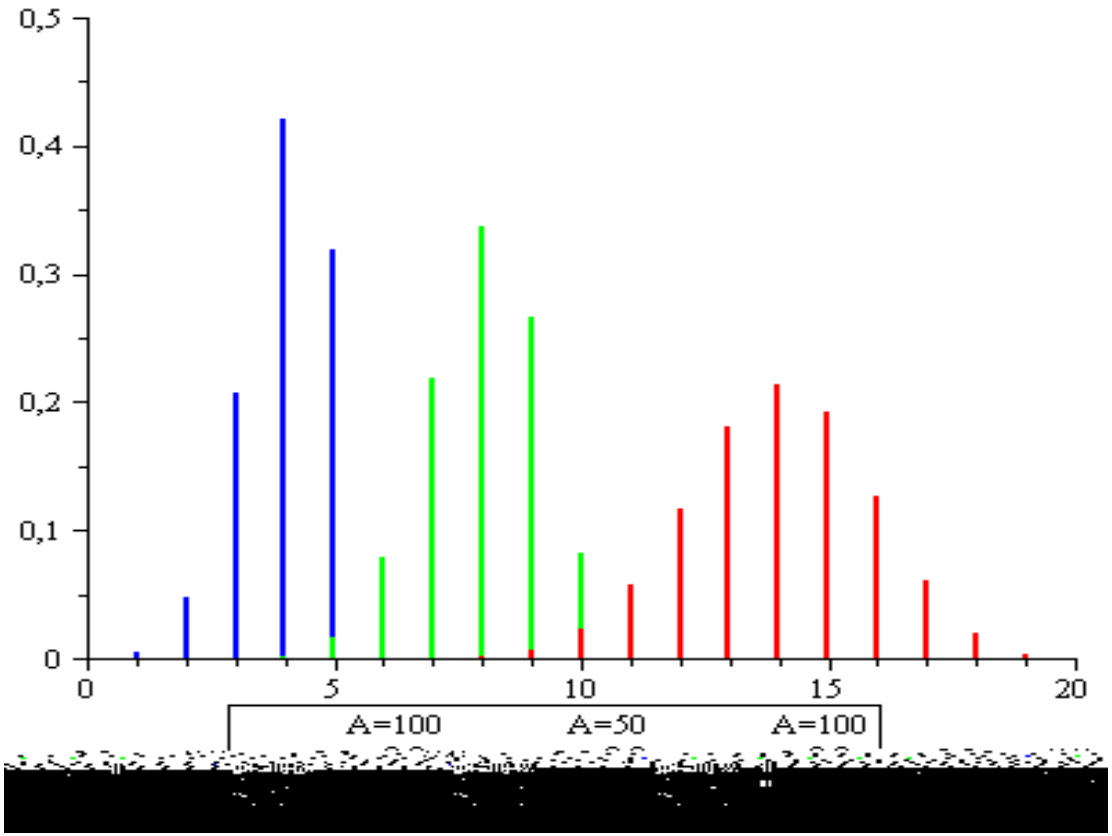
## Loi hypergéométrique

Si l'on pose $N_1 = pN$, on a pour $0 \leq k \leq n$ :

$$Q(k) = \frac{\binom{pN}{k} \cdot \binom{(1-p)N}{n-k}}{\binom{N}{n}}$$

Cette fonction s'appelle **loi hypergéométrique** avec paramètres $p \quad [0,1]$, $N \quad \mathbf{N}$ et $n, \quad Nat$.

| $Q(k)$ est la probabilité d'avoir $k$ boules rouges dans un tirage de $n$ boules sans remise. |
|---|




Distribution hypergéométrique

## Avec ou sans remise ?

– **Intuitivement**, si on fait tendre $N$ vers l'infini on s'attend à que la différence entre tirage avec remise et sans remise devienne **négligeable** car la probabilité de tirer deux fois la même boule est petite.

– **En pratique**, si $N$ et beaucoup plus grand que $n$ on utilise plutôt le **tirage avec remise** car la manipulation de la formule est plus aisée.

– Par ailleurs, à son tour la loi binomiale peut être **approximée** par des lois plus simples (Poisson, loi normale, ...).

## Convergence hypergéométrique vers binomiale

$$Q(k) = \frac{\binom{N_1}{k}\cdot\binom{N-N_1}{n-k}}{\binom{N}{n}}$$

$$= \frac{N_1!}{k!(N_1-k)!}\,\frac{(N-N_1)!}{(n-k)!(N-N_1-(n-k))!}\,\frac{(N-n)!n!}{N!}$$

$$= \binom{n}{k}\,\frac{N_1\cdots(N_1-k+1)}{N\cdots(N-k+1)}\,\frac{(N-N_1)\cdots(N-N_1)-(n-k)+1}{(N-k)\cdots(N-n+1)}$$

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

Si $pN = N_1$ et $N \to \infty$.

## Rappel : exponentiel et logarithme

On définit la fonction **exponentielle** sur $\mathbf{R}$

$$\boxed{e^x = \Sigma_{k=0,\ldots,\infty} \frac{x^k}{k!}}$$

En particulier $e = \Sigma_{k=0,\ldots,\infty} \frac{1}{k!} \approx 2,71$. Le **logarithme (naturel)** est la fonction inverse définie sur les réels positifs :

$$\boxed{\ln(x) = y \text{ si } x = e^y}$$

On a :

$$\begin{cases} \frac{\mathrm{d}e^x}{\mathrm{d}x} = e^x & \frac{\mathrm{d}\ln(x)}{\mathrm{d}x} = \frac{1}{x} \\ \int e^x \mathrm{d}x = e^x & \int \ln(x)\mathrm{d}x = x\ln(x) - x \end{cases}$$

## Rappel : approximations et développement de Taylor

On est souvent amené à **simplifier** un calcul avec les fonctions exponentiel ou logarithme. Par exemple, pour tout $x \quad \mathbf{R}$ :

$$1 + x \leq e^x \qquad \ln(1 + x) \leq x \quad (\text{si } x > -1)$$

**Technique de preuve** On peut utiliser le **développement de Taylor** (et la variante Taylor-Lagrange) :

$$\begin{aligned} f(x) &= \Sigma_{k=0,\ldots,\infty} f^{[k]}(x_0) \frac{(x-x_0)^k}{k!} \\ &= (\Sigma_{k=0,\ldots,n} f^{[k]}(x_0) \frac{(x-x_0)^k}{k!}) + f^{[n+1]}(y) \frac{(x-x_0)^{(n+1)}}{(k+1)!} \end{aligned}$$

où $f^{[k]}$ est la $k$-ème dérivée de $f$ (qui doit exister...) et $x_0 < y < x$.

## Exemples

– Pour la fonction **exponentielle** en prénant $x_0 = 0$ on a :

$$e^x = 1 + x + (e^y)\frac{x^2}{2} \geq 1 + x$$

– Pour la fonction **logarithme** en prénant $x_0 = 0$ on a avec $|x| < 1$ :

$$\ln(1 + x) = x - \frac{x^2}{2(y + 1)^2} \leq x$$

**NB** Ici on peut aussi dériver l'inégalité pour le logarithme de celle pour l'exponentiel (et inversement).

## Exemple : approximation d'une série par une intégrale

En utilisant l'interprétation de l'**intégrale** comme la surface sous une curve on approxime la **série harmonique** :

$$\ln(n) = \int_1^n \frac{1}{x}\mathrm{d}x \;\leq\; \Sigma_{k=1,\ldots,n}\frac{1}{k} \;\leq\; 1+\ln(n)$$

## Exemple : approximation du factoriel

$$\left(\frac{n}{e}\right)^n \leq n! \leq e\sqrt{n}\left(\frac{n}{e}\right)^n$$

**Borne inférieure** Par **définition** du factoriel :

$$e^n = \Sigma_{i=0\ldots,\infty}\frac{n^i}{i!} > \frac{n^n}{n!}$$

**Borne supérieure** On remarque :

$$\ln(n!) = \Sigma_{i=1,\ldots,n}\ln(i)$$

Pour $i \geq 1$, on approxime l'intégrale par la **surface d'un trapèze** (ceci marche pour les fonctions **concaves** !) :

$$\int_i^{i+1}\ln(x)\mathrm{d}x \geq \frac{\ln(i)+\ln(i+1)}{2}$$

On dérive :

$$\begin{aligned}
&\int_1^n \ln(x)\mathrm{d}x \\
&= \Sigma_{i=1,\ldots,n-1} \int_i^{i+1} \ln(x)\mathrm{d}x \\
&\geq \Sigma_{i=1\ldots,n-1} \frac{\ln(i)+\ln(i+1)}{2} \\
&= (\Sigma_{i=1\ldots,n}\ln(i)) - \frac{\ln(n)}{2}
\end{aligned}$$

Comme :

$$\int_1^n \ln(x) = n\ln(n) - n + 1$$

On a :

$$n\ln(n) - n + 1 \geq \ln(n!) - \frac{\ln(n)}{2}$$

On obtient la **borne supérieure** en prenant l'exposant.

**Remarque**

La borne supérieure est une assez bonne approximation, en effet la **formule de Stirling** prédit :

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

et le cofficient $\sqrt{2\pi} \approx 2,50$ est proche de $e \approx 2,71$.

## Quelques approximations à retenir

– Hypergéométrique $\approx$ Binomiale (pour $N$ qui va à l'infini).

– $1 + x \leq e^x$ (assez bonne pour $0 \leq x \leq 1$).

– Série et Intégrale (par exemple pour la série harmonique).

– Encadrement du factoriel :

$$(\frac{n}{e})^n \leq n! \leq (\frac{n}{e})^n \sqrt{ne}$$

## Exemple : paradoxe des anniversaires

Combien de personnes faut-il réunir pour que la probabilité de trouver deux personnes avec la même date d'anniversaire soit supérieure ou égale à $1/2$ ?

– Soient $n$ le nombre de **jours** dans une année et $k$ le nombre de **personnes**.

– Un **événement élémentaire** est un élément de $\{1, \ldots, n\}^k$.

– Tous les événements ayant la même probabilité, on **compte** le nombre d'événements $(d_1, \ldots, d_k)$ tels que $d_i = d_j$ si $i = j$.

– Ce nombre est :

$$n(n-1)\cdots(n-k+1) = \Pi_{i=0,\ldots,k-1}(n-i)$$

– Donc la probabilité $q$ qu'il **n'y ait pas deux personnes** avec la même date d'anniversaire est

$$\begin{aligned} q &= (\tfrac{1}{n^k})\Pi_{i=0,\ldots,k-1}(n-i) \\ &= \Pi_{i=0,\ldots,k-1}\tfrac{(n-i)}{n} \\ &= \Pi_{i=1,\ldots,k-1}(1-\tfrac{i}{n}) \end{aligned}$$

– On **borne supérieurement** $1 + x$ par $e^x$ et on dérive :

$$q \leq \Pi_{i=1,\ldots,k-1} e^{-\frac{i}{n}} = e^{-\Sigma_{i=1,\ldots,k-1}(\frac{i}{n})} = e^{\frac{-k(k-1)}{2n}}$$

– Pour avoir $q \leq 1/2$, il suffit que

$$\frac{-k(k-1)}{2n} \leq -\ln(2)$$

On étudie l'**inégalité quadratique** :

$$k^2 - k - 2n\ln(2) \geq 0$$

on dérive que pour avoir $q \leq 1/2$, il suffit que :

$$k \geq \frac{1 + \sqrt{(1 + 8n\ln(2))}}{2}$$

– En particulier, pour $n = 365$, il **suffit** d'avoir $k = 23$ (en général il suffit de prendre $k \approx (1,2)\sqrt{n}$).

## Implications de l'analyse en cryptographie

– En cryptographie on utilise des **fonctions de hachage** qui prennent une suite de bits et produisent une suite de longueur $n$ de bits :

$$h : 2 \quad \rightarrow 2^n$$

– Par exemple, dans les schémas de signature éléctronique, on **signe** le hachage du message.

– La fonction $h$ doit être **facile** à calculer mais il doit être difficile de trouver une **collision**, c'est-à-dire $x, y \quad 2 \quad$ tels que $x = y$ et $h(x) = h(y)$.

– L'analyse du paradoxe des anniversaires permet de déterminer une **borne inférieure** à $n$.

– Si le nombre de **valeurs de hachage** (les dates d'anniversaire) est $2^n$, l'inégalité devient :

$$k \geq \frac{1 + \sqrt{1 + 8\ln(2)2^n}}{2}$$

– Ceci signifie que si on calcule environ $\sqrt{2^n}$ valeurs alors la probabilité de trouver une collision est plus grande que $1/2$.

– Actuellement, on **recommande** $n \geq 160$, ce qui implique que le coût d'une attaque par force brute est environ $2^{80}$.

## Exemple : $\Omega$ fini avec probabilité non-uniforme

– Il y a trois portes et un prix derrière une porte. Le **modérateur** du jeu sait où se trouve le prix. Le **joueur** gagne le prix s'il devine la porte qui cache le prix.

– Le jouer choisit une porte **sans l'ouvrir**.

– Le modérateur du jeu **ouvre** une des deux portes qui n'ont pas été choisies par le modérateur et qui ne cache pas le prix (une telle porte existe toujours).

**Problème** (inspiré par un jeu télévisé) :

| Après l'ouverture de la porte, le joueur a-t-il intérêt à **changer** son choix ? |
|---|

## Analyse

– Représentons les **portes** avec un éléments de $\{1, 2, 3\}$.

– L'ensemble des **événements élémentaires** est de la forme :

$$\Omega = \{(i, j, k) \quad (\{1, 2, 3\})^3 \mid k \quad \{1, 2, 3\} \setminus \{i, j\}\}$$

avec l'interprétation :
– $i$ est la porte qui cache le **prix**,
– $j$ est la porte **choisie** par le joueur et
– $k$ est la porte **ouverte** par le modérateur.

**NB** Si $i = j$ alors la porte ouverte par le modérateur est **uniquement déterminée** alors que si $i = j$ le modérateur a **deux choix** (ceci implique que dans ce cas la probabilité n'est pas uniforme).

## Représentation des possibilités par un arbre

- $i \quad \{1, 2, 3\}$. En supposant un choix uniforme, la probabilité associée à chaque branche est $1/3$.

- $j \quad \{1, 2, 3\}$. En supposant un choix uniforme, la probabilité associée à chaque branche est $1/3$.

- On peut supposer que le choix de la porte qui cache le prix est 'indépendant' du choix du joueur. Donc chaque combinaison $(i, j)$ devrait avoir la même probabilité à savoir $1/9$.

- Si $i = j$ le moderateur doit choisir la porte dans $\{1, 2, 3\} \setminus \{i, j\}$. Sinon il peut choisir une porte dans $\{1, 2, 3\} \setminus \{i\}$. On suppose encore une fois le choix uniforme.

## En conclusion...

– L'événement où le jouer **ne change pas de porte et gagne** est :

$$A = \{(1,1,2),(1,1,3),(2,2,1),(2,2,3),(3,3,1),(3,3,2)\}$$

qui a probabilité $P(A) = 3 \quad (1/9) = 1/3$.

– D'autre part, l'événement où le jouer **change de porte et gagne** est :

$$B = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}$$

qui a probabilité $P(B) = 6 \quad (1/9) = 2/3$.

## Simulation

On peut écrire un programme qui simule l'expérience qu'on vient de décrire. Ceci suppose que la classe `Random` fait correctement son travail !

```
static boolean experience(){
        int i=gen.nextInt(3)+1;  //prix
        int j=gen.nextInt(3)+1;  //choix joueur
        int k;                   //porte ouverte par le moderateur
        if (i==j) {k=choix(i);}  //  k dans {1,2,3}\{i}
        else {k=troisieme(i,j);} // {k}={1,2,3}\{i,j}
        if (i==troisieme(j,k)){return true;}
        else {return false;}
    }
```

Que se passe-t-il si le choix de $i$ (la porte qui cache le prix) et/ou de $k$ (la porte ouverte) **n'est pas uniforme** ?

## Exemple : $\Omega$ infini

- Deux joueurs $J_1$ et $J_2$ jouent avec un dé **non-biaisé**.
- $J_1$ commence et le **premier** qui obtient un 6 gagne.
- Quelle est la **probabilité** que $J_1$ gagne, que $J_2$ gagne, que personne gagne ?

## Analyse

– Si $\Sigma$ est un ensemble alors $\Sigma$ est l'ensemble des **mots finis** sur $\Sigma$ et $\Sigma^{\mathbf{N}}$ l'ensemble des **mots infinis** (dénombrables). Soit $T = \{1, 2, 3, 4, 5\}$.

– Un jeu qui **termine** est décrit par un mot de la forme $w6$ où $w \quad T$ .

– Un jeu qui **ne termine pas** est un mot dans $T^{\mathbf{N}}$.

– On peut donc **définir** :

$$\Omega = \{w6 \mid w \quad T \ \} \quad \{w \mid w \quad T^{\mathbf{N}}\}$$

– Notez que si $\Omega$ est **infini** on **ne peut pas** associer aux événements élémentaires une probabilité **uniforme non nulle**.

– Par ailleurs, notez que $\Omega$ **n'est pas dénombrable** ! Cependant tous les événements élémentaires dans $T^{\mathbf{N}}$ vont recevoir une probabilité nulle.

– L'événements un **joueur gagne au coup** $n+1$ est :

$$G_n = \{w6 \mid w \quad T \ , |w| = n\}$$

– On associe à $G_n$ la **probabilité** :

$$P(G_n) = (5/6)^n(1/6)$$

– Si $n = m$ on a $G_n \cap G_m =$ . On peut s'attendre à que :

$$P(\bigcup_{n=0,\ldots,\infty} G_n) = \Sigma_{n=0,\ldots,\infty} P(G_n)$$

– Donc la **probabilité** de $G = \bigcup_{n=0,\ldots,\infty} G_n$ (un joueur gagne) est :

$$P(G) = \Sigma_{n=0,\ldots,\infty} (5/6)^n (1/6) = 1$$

– Comme $G \cap T^{\mathbf{N}} =$ la **seule possibilité** est maintenant d'avoir :

$$P(T^{\mathbf{N}}) = 1 - P(G) = 0$$

## Loi géométrique

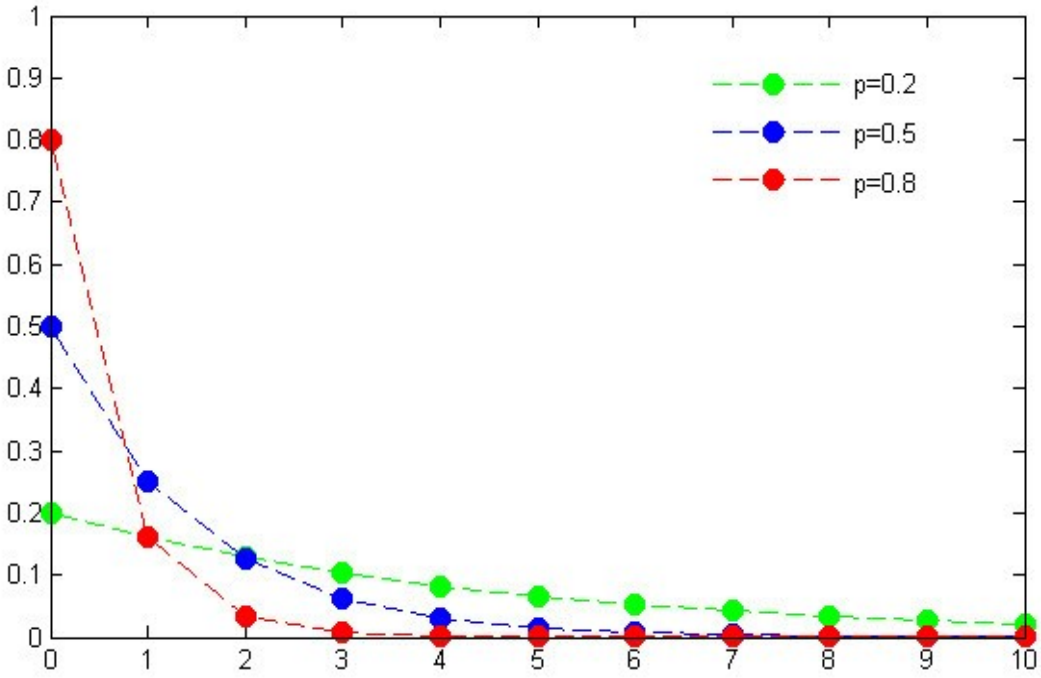
Pour $p \quad [0,1]$ et $k \quad \mathbf{N}^+$, la fonction définie par :

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

est appelée **loi géométrique**. Les propriétés de la série géométrique assurent :

$$\Sigma_{k=0,\ldots,\infty}(1-p)^k p = \frac{p}{1-(1-p)} = 1$$

| $P(k)$ est la probabilité de tirer pile pour la première fois après $k$ tirages. |
|---|




Distribution Géométrique

## Exemple : les nombres réels

– On veut définir une mesure de **probabilité uniforme** sur $[0, 1]$.

– On peut montrer qu'**on ne peut pas mesurer tous les sous-ensemble** de $[0, 1]$ (voir TD).

– Cependant on peut construire le plus petit ensemble d'événements qui contient les **intervalles** (ce qui est cohérent avec notre intuition physique de ce qui est mesurable !) et est stable par opérations ensemblistes denombrables. On appelle ces ensembles **boreliens**.

– Ceci explique pourquoi la définition de la probabilité est sur les **événements** plutôt que sur les **événements élémentaires**.

– Ensuite on peut définir une mesure de probabilité qui intuitivement **mesure la longueur**.

– Dans cette approche, les **sous-ensemble dénombrables** de $[0, 1]$ (par exemple les rationnels dans $[0, 1]$) ont probabilité 0.

**NB** Une définition rigoureuse est **hors de la portée** de ce cours.

# Sommaire

**Ingredients du modèle**

- Un ensemble de **résultats** $\Omega$.
- Un ensemble d'**événements** $A \quad \Omega$ stable par opérations ensemblistes dénombrables.
- Une mesure de **probabilité** $P : A \to [0, 1]$ avec une propriété d'additivité dénombrable.

**Exemples d'analyse à mémoriser**

- Pile ou face. Loi de **Bernoulli**.
- $n$ tirages avec remise. Loi **binomiale**.
- $n$ tirages sans remise. Loi **hypergéométrique**.
- Tirage itéré jusqu'à succès. Loi **géométrique**.

**Techniques d'approximation** issues de la combinatoire et/ou de l'analyse.

# Une définition axiomatique

## Espace de probabilité

Un **espace de probabilité** est un triplet $(\Omega, A, P)$ où :

- $\Omega$ est un ensemble non-vide de **résultats**.

- $A$ est un sous-ensemble non vide d'**événements** de $2^{\Omega}$, stable par complément et unions (et intersections) dénombrables. Des noms pour $A$ sont : **$\sigma$-algèbre** et **tribu** (tradition Bourbakiste).

- $P : A \to [0, 1]$ est une fonction qu'on appelle (mesure de) **probabilité** telle que $P(\Omega) = 1$ et telle que $P$ est **additive** par rapport à des unions dénombrables d'événements **disjoints** entre eux :

$$P(\bigcup_{i \geq 0} A_i) = \Sigma_{i \geq 0} P(A_i)$$

## Propriétés

1. $P(A^c) = 1 - P($

## Preuve de (1)

On prend : $A_1 = A, A_2 = A^c, A_{n+2} = \quad .$

## Preuve de (2)

Il suffit de noter que :

$$\begin{array}{lll} A & = (A \cap B^c) & (A \cap B) \\ B & = (B \cap A^c) & (A \cap B) \\ A \quad B & = (A \cap B^c) \quad (B \cap A^c) & (A \cap B) \end{array}$$

et appliquer la propriété d'additivité :

$$\begin{aligned} P(A \quad B) &= P(A \cap B^c) + P(B \cap A^c) + P(A \cap B) \\ &= P(A) - P(A \cap B) + P(B) \end{aligned}$$

## Preuve de (3)

– Supposons que $x$ appartient exactement a $m$ ensembles parmi $A_1, \ldots, A_n$ ; donc $1 \leq m \leq n$.

– Dans la **partie gauche** on compte la probabilité de $x$ **1 fois**.

– e gauc

on con fois

## Exemple (points-fixes d'une permutation)

On dénote par $Perm_n$ l'ensemble des **permutation** sur $\{1, \ldots, n\}$. Un **point fixe** de $\pi$ $Perm_n$ est un élément $i$ $\{1, \ldots, n\}$ tel que $\pi(i) = i$.

Quelle est la probabilité de tirer une permutation avec **exactement $k$ points fixes** ?

On peut reformuler, par exemple, en demandant la probabilité que parmi $n$ personnes ayant mélangé leurs chapeaux identiques il y en ait exactement $k$ qui retrouvent le leur.

## Analyse

On définit :

$$A_k = \{\pi \quad Perm_n | \pi(k) = k\}$$

Ce n'est pas l'événement qui nous intéresse mais certaines probabilités sont faciles à calculer :

$$\begin{aligned} P(A_k) &= \frac{1}{n} \\ P(A_{i_1} \cap \cdots \cap A_{i_k}) &= \frac{(n-k)!}{n!} \end{aligned}$$

On remarque que $\bigcup_{k=1,\ldots,n} A_k$ est l'ensemble des permutations qui ont au moins un point fixe. On applique le **principe d'inclusion-exclusion**...

$$\begin{array}{lll} P(\bigcup_{k=1,\ldots,n} A_k) = & + & \Sigma_{k=1,\ldots,n} P(A_k) \\ & - & \Sigma_{i_1<i_2} P(A_{i_1} \cap A_{i_2}) \\ & \ldots & \ldots \\ & (-1)^{n+1} & P(A_1 \cap \cdots \cap A_n) \end{array}$$

On remarque :

$$\Sigma_{i_1<\cdots<i_k} P(A_{i_1} \cap \cdots \cap A_{i_k}) = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}$$

Donc :

$$P(\bigcup_{k=1,\ldots,n} A_k) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1} \frac{1}{n!}$$

– La probabilité de l'événement complémentaire (= **pas de points fixes**) est

$$p_n = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} - \cdots + (-1)^n \frac{1}{n!}$$

et par définition de l'**exponentiel**, $\lim_{n \to +\infty} p_n = \frac{1}{e}$.

– Fixons $k$ **points** : $1 \leq i_1 < \cdots < i_k \leq n$. Combien de permutations peut-on construire qui ont **exactement** ces points comme points fixes ?

$$p_{n-k}(n-k)!$$

– On voit donc l'**intérêt** d'avoir calculé $p_n \ldots$

## Remarque

– On a $\frac{1}{e} \approx 0,367$ et pour $i \geq 6$, $0,366 \leq p_i \leq 0,3680\ldots$.

– Donc la **convergence est rapide** et si $n >> k$ on peut considérer que $p_{n-k} \approx \frac{1}{e}$.

– De plus on a :

$$\Sigma_{k=0,\ldots,5} \frac{1}{k!e} \approx 0,9994\ldots$$

Ce qui veut dire que la probabilité d'avoir plus que 5 points fixes est moins que $10^{-3}$ et ceci ne dépend pas vraiment de la taille de la permutation !

## Construction de σ-algèbres

Soit $\Omega$ un ensemble non-vide. Les **σ-algèbres** (ou **tribus**) sur $\Omega$ peuvent être **ordonnées par inclusion**. On a les propriétés suivantes :

1. $\{\ , \Omega\}$ est la **plus petite** et $2^\Omega$ la **plus grande** σ-algèbre.
2. Pour tout $A$ $\Omega$, $\{\ , A, A^c, \Omega\}$ est une σ-algèbre.
3. Stabilité par **intersection** : si $A_i$ $2^\Omega$ sont des σ-algèbres pour $i$ $I$ alors $\cap_i\ _I A_i$ est une σ-algèbre.
4. Pour tout $B$ $2^\Omega$ il existe une **plus petite** σ-algèbre $A$ telle que $B$ $A$.

## Note historique

Cette présentation des espaces de probabilité (qui est maintenant standard) est due à :

**A. Kolmogorov**.
*Grundbegri□e Wahrscheinlichkeitsrechnung*.
Springer 1933.

Il s'agit d'une synthèse d'une longue histoire qui suit en particulier le développement de la **théorie de la mesure** entre le XIX et le XX siècle.

## Le cas discret

Dans **ce cours**, étant donné $\Omega$ :

– On prend toujours : $A = 2^{\Omega}$.

– On suppose que :

$$\Omega \ = \{\omega \quad \Omega \mid P(\{\omega\}) = 0\}$$

est **fini ou dénombrable**.

– Ainsi si $A$ $\Omega$, la probabilité de $A$ est **déterminée** par :

$$\boxed{P(A) = \Sigma_{\omega \ A \cap \Omega'} P(\{\omega\})}$$

En d'autres termes, dans le **cas discret**, un espace de probabilité est déterminé par :

– un ensemble $\Omega$ de **résultats**,

– un sous-ensemble **dénombrable** $\Omega \quad \Omega$,

– une mesure de **probabilité** $P : \Omega \rightarrow [0, 1]$ telle que

$$\Sigma_{\omega \ \ \Omega'} P(\omega) = 1$$

On peut donc oublier de définir la $\sigma$-algèbre (=tribu) $A$ et définir $P$ directement sur $\Omega$ .

## Probabilité d'une suite croissante

Soit $\{A_i\}_{i \in \mathbf{N}}$ une suite **croissante** d'événements ($A_i \subseteq A_{i+1}$).
Alors

$$P(\bigcup_{i \in \mathbf{N}} A_i) = \lim_{n \to +\infty} P(A_n)$$

Par **dualité**, si la suite est **décroissante** on a :

$$P(\bigcap_{i \in \mathbf{N}} A_i) = \lim_{n \to +\infty} P(A_n)$$

## Preuve

On définit :

$$B_0 = A_0 \qquad B_{i+1} = (A_{i+1} \setminus A_i)$$

On remarque :

1. $B_i \cap B_j =$ si $i = j$.
2. $A_n = \bigcup_{i=0,\ldots,n} B_i$.

On observe :

$$\begin{aligned} P(\bigcup_{i \in \mathbf{N}} A_i) &= P(\bigcup_{i \in \mathbf{N}} B_i) \\ &= \Sigma_{i \in \mathbf{N}} P(B_i) \\ &= \lim_{n \to +\infty} (\Sigma_{i=0,\ldots,n} P(B_i)) \\ &= \lim_{n \to +\infty} P(\bigcup_{i=0,\ldots,n} B_i) \\ &= \lim_{n \to +\infty} P(A_n) \end{aligned}$$

## Application

On reprend le **jeu de dé** où le premier qui tire 6 gagne et on montre que **la probabilité que le jeu ne termine pas est** 0. Soit :

$$A_n = \{\omega \mid \omega \text{ on ne tire pas 6 dans le premiers } n \text{ jets}\}$$

On a :

$$A_n \quad A_{n+1} \qquad \text{et} \qquad P(A_n) = (\frac{5}{6})^n$$

$\bigcap_{n \ \mathbf{N}} A_n$ est l'événement où 'personne tire 6' et on peut conclure par **le passage à la limite sur une suite décroissante** :

$$P(\bigcap_{n \ \mathbf{N}} A_n) = \lim_{n \to +\infty} P(A_n) = 0$$

## Borel-Cantelli (loi 0)

Soit $\{A_i\}_{i \in \mathbf{N}}$ une suite d'événements. On peut toujours construire l'événement :

$$L = \bigcap_{i \in \mathbf{N}} (\bigcup_{j \geq i} A_i)$$

qui est caractérisé par la propriété :

$\omega \in L$ ssi $\omega$ apparaît **infiniment souvent** dans la suite

La **loi 0** affirme :

Si $\Sigma_{i \in \mathbf{N}} P(A_i) < \infty$ alors $P(L) = 0$.

Si la somme des probabilités **converge** alors la probabilité que un événement se produise infiniment souvent est 0.

## Remarque préliminaire

– Soit

## Preuve de Borel-Cantelli (loi 0)

On définit :

$$B_i = \bigcup_{j \geq i} A_j$$

Par **hypothèse** :

$$P(B_i) = P(\bigcup_{j \geq i} A_j) \leq \Sigma_{j \geq i} P(A_j) = S_{\geq i}$$

Comme $\{B_i\}_{i \in \mathbf{N}}$ est une suite **décroissante** on a :

$$P(L) = P(\bigcap_{i \in \mathbf{N}} B_i) = \lim_{n \to +\infty} P(B_n)$$

Mais $P(B_n) \leq S_{\geq n}$ et $S_{\geq n}$ tend vers 0 par la **remarque préliminaire**.

## Application

On considère une **suite de tirages** de 0 ou 1 où à chaque tirage 1 est tiré avec probabilité $\mathbf{p < \frac{1}{2}}$.

On définit une **suite d'événements** $\{A_i\}_{i \in \mathbf{N}}$ où $A_i$ est l'événement que dans la suite dénombrable de tirages dans l'intervalle entre $2^i$ et $2^{i+1} - 1$ on a tiré consécutivement $i$ fois 1.

| | | | | | |
|---|---|---|---|---|---|
| Tirages : | $1,$ | $2, 3,$ | $4, 5, 6, 7,$ | $8, 9, 10, 11, 12, 13, 14, 15,$ | $16, 17, \ldots$ |
| 1 consécutifs : | 0 | 1 | 2 | 3 | 4 |

La probabilité que ces événements se produisent infiniment souvent est 0.

## Analyse

On **borne** :

$$P(A_i) \leq (2^i - i)p^i \leq (2p)^i$$

Ici on additionne les probabilités de tirer $i$ fois 1 à commencer du tirage $2^i, 2^i + 1, \ldots, 2^i - i$ (ces événements recouvrent $A_i$ mais ne sont pas disjoints !)

Comme $2p < 1$ on a une série géométrique qui **converge** :

$$\Sigma_{i \in \mathbf{N}} P(A_i) \leq \Sigma_{i \in \mathbf{N}} (2p)^i = \frac{1}{1 - 2p}$$

On peut donc **appliquer la loi 0** :

$$P(\bigcap_{i \in \mathbf{N}} \bigcup_{j \geq i} A_j) = 0$$

## Sommaire

– Définition d'**espace de probabilité**.

– Principe d'**inclusion-exclusion**.

–

# Probabilité Conditionnelle et Indépendance

## Probabilité conditionnelle

Soient $A, B$ deux événements sur un espace de probabilité $(\Omega, A, P)$ tel que $P(A) > 0$. Alors :

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$

est la **probabilité conditionnelle** de $B$ étant donné $A$. A noter :

- Si $P(A) = 0$ alors $P(B \mid A)$ n'est pas définie.
- Il y a un **abus de notation** car $B \mid A$ **n'est pas** un événement. On écrit aussi $P_A(B)$.

Chaque fois qu'on écrit $P(A \mid B)$ on fait l'hypothèse que $P(B) > 0$.

## Remarque (probabilité dérivée)

Soit $(\Omega, A, P)$ un espace de probabilité. Alors pour tout événement $B \quad A$ avec **probabilité positive**, on a une **nouvelle mesure de probabilité** $P$ :

$$P\ (A) = P(A \mid B)$$

On vérifie

$$P\ (\Omega) = P(\Omega \mid B) = \frac{P(\Omega \cap B)}{P(B)} = 1$$

et si $\{A_i\}_{i\ \mathbf{N}}$ est une séquence d'événements **disjoints** alors :

$$\begin{aligned} P\ (\textstyle\bigcup_{i\ \mathbf{N}} A_i) &= \frac{P(\bigcup_{i\in\mathbf{N}} A_i \cap B)}{P(B)} \\ &= \Sigma_{i\ \mathbf{N}} \frac{P(A_i \cap B)}{P(B)} \\ &= \Sigma_{i\ \mathbf{N}} P(A_i \mid B) \\ &= \Sigma_{i\ \mathbf{N}} P\ (A_i) \end{aligned}$$

## Indépendance

Deux événements $A, B$ sont **indépendants** si :

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

Si $A$ ou $B$ a probabilité 0 alors $A$ et $B$ sont indépendants. Sinon, l'indépendance est équivalente à demander :

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

ou de façon plus suggestive :

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

La proportion de $A$ dans $B$ et la même que
la proportion de $A$ dans $\Omega$.

## Indépendance d'un ensemble d'événements

La notion d'indépendance se généralise à un ensemble fini ou dénombrable d'événements. Dans ce cas il faut distinguer **deux notions**.

Les événements $\{A_i\}_{i \in \mathbf{N}}$ sont :

**mutuellement indépendants** si pour tout $I \subseteq \mathbf{N}$ :

$$P(\bigcap_{i \in I} A_i) = \Pi_{i \in I} P(A_i)$$

**indépendants deux à deux** si pour tout $i, j \in \mathbf{N}$ :

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

**NB** La première condition est **strictement plus forte**. Par défaut **indépendance = indépendance mutuelle**.

## Exemple

Soit $\mathbf{Z}_n = \{0, 1, \ldots, n-1\}$ les entiers modulo $n$ avec addition $+$ et

$$\Omega = \{(x, y, z) \quad (\mathbf{Z}_n)^3 \mid z = (x + y) \mod n\}$$

Les événements de la forme :

$$\begin{aligned} A_{1,i} &= \{(x, y, z) \quad \Omega \mid x = i\} \\ A_{2,j} &= \{(x, y, z) \quad \Omega \mid y = j\} \\ A_{3,k} &= \{(x, y, z) \quad \Omega \mid z = k\} \end{aligned}$$

sont indépendants deux à deux, mais **pas** mutuellement indépendants.

## Identités avec Probabilité conditionnelle

1. Règle de **multiplication** :

$$P(\bigcap_{i=1,\ldots,n} A_i) = P(A_1) \cdot P(A_2 \mid A_1) \cdots P(A_n \mid (A_1 \cap \cdots \cap A_{n-1}))$$

2. En utilisant le **complémentaire** :

$$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$$

3. Règle de **Bayes** (en supposant $P(A), P(B) > 0$) :

$$P(B)P(A \mid B) = P(A)P(B \mid A)$$

4. Encore une règle de **Bayes** en supposant que $A_1, \ldots, A_n$ est une **partition** de $\Omega$ :

$$P(A_i \mid B) = \frac{P(B \mid A_i) \cdot P(A_i)}{\Sigma_{j=1,\ldots,n} P(B \mid A_j) \cdot P(A_j)}$$

## Exemple : application Bayes

– On dispose de 3 pièces dont 1 biaisée donne pile avec probabilité 2/3.

– On tire les 3 pièces et on obtient le résultat : pile,pile,face.

– Quelle est la probabilité que la pièce biaisée soit la première ?

– Les **événements élémentaires** sont :

$$\Omega = \{(b, x, y, z) \mid b \quad \{1, 2, 3\} \text{ et } x, y, z \quad \{\text{pile}, \text{face}\}\}$$

– Soient :

$$\begin{aligned} E_i &= \{\omega \quad \Omega \mid \pi_1(\omega) = i\} \\ O &= \{\omega \quad \Omega \mid \pi_2(\omega) = \pi_3(\omega) = \text{pile}, \pi_4(\omega) = \text{face}\} \end{aligned}$$

– On sait que $P(E_i) = 1/3$ pour $i = 1, 2, 3$ et on cherche à calculer :

$$P(E_1 \mid O)$$

– On sait aussi que $\{E_1, E_2, E_3\}$ est une partition de $\Omega$ et que :

$$P(O \mid E_1) = P(O \mid E_2) = 1/6 \quad P(O \mid E_3) = 1/12$$

On peut donc appliquer la **formule de Bayes** :

$$P(E_1 \mid O) = \frac{P(O \mid E_1)P(E_1)}{\Sigma_{i=1,2,3} P(O \mid E_i)P(E_i)} = \frac{2}{5}$$

## Exemple : sécurité parfaite

Un **système cryptographique** (dans une version symétrique simplifiée) est défini par :

– Trois **ensembles finis** :

$P$ Textes clairs

$C$ Textes chiffrés

$K$ Clés.

– Pour toute clé $k \quad K$ une :

$E_k : P \to C$ fonction chiffrement et

$D_k : C \to P$ fonction de déchiffrement

telles que pour tout texte clair $p \quad P$ :

$$D_k(E_k(p)) = p$$

## Événements à considérer

On prend :

$$\Omega = P \times K$$

et on utilise les **abréviations** :

- $(p, k) \equiv \{(p, k)\}$ : le texte clair $p$ est chiffré avec la clé $k$.
- $p \equiv \{(p, k) \mid k \quad K\}$ : le texte clair $p$ est chiffré.
- $k \equiv \{(p, k) \mid p \quad P\}$ : un texte clair est chiffré avec la clé $k$.
- $c \equiv \{(p, k) \mid E_k(p) = c\}$ : le texte chiffré est $c$.

## Hypothèses

– Les événements $p$ et $k$ sont **indépendants**, c'est-à-dire

$$P(p \cap k) = P(p, k) = P(p)P(k) \ .$$

– L'attaquant connaît $P(p)$, pour tout $p \quad P$.

**NB**

## Sécurité parfaite

– On dit qu'un système cryptographique est **parfaitement sûr** si le fait d'observer un texte chiffré $c$ ne fournit aucune information (probabiliste) sur le texte clair correspondant $p$.

– En d'autres termes, le système est parfaitement sûr si pour tout $p \quad P$ et $c \quad C$ avec $P(c) > 0$

$$P(p \mid c) = P(p)$$

– Toujours en supposant $P(c) > 0$ ceci est équivalent à demander que $p$ et $c$ sont **indépendants** :

$$P(p \mid c) = P(p) \qquad \Leftrightarrow \qquad P(p \cap c) = P(p)P(c)$$

## Exemple

$P = \{0, 1\}$, $K = \{A, B\}$, $C = \{a, b\}$.

Chiffrement

| $p$ | $k$ | $E_k(p)$ |
|---|---|---|
| 0 | $A$ | $a$ |
| 1 | $A$ | $b$ |
| 0 | $B$ | $b$ |
| 1 | $B$ | $a$ |

Probabilités

| $p$ | $P$ | $k$ | $P$ |
|---|---|---|---|
| 0 | 1/4 | $A$ | 1/4 |
| 1 | 3/4 | $B$ | 3/4 |

Proba. texte chiffré

| $c$ | $P$ |
|---|---|
| $a$ | 10/16 |
| $b$ | 6/16 |

Autres proba. dérivées

| $p$ | $c$ | $P(p \cap c)$ | $P(p \cap c)/P(c)$ |
|---|---|---|---|
| 0 | $a$ | 1/16 | 1/10 |
| 1 | $a$ | 9/16 | 9/10 |
| 0 | $b$ | 3/16 | 1/2 |
| 1 | $b$ | 3/16 | 1/2 |

Ce système est-il parfaitement sûr ?

**Non !**

– Par exemple, supposons que l'attaquant voit le texte chiffré $a$.

– La probabilité que le texte clair est 1 est 9/10.

– Ce qui est supérieur à 3/4 qui est la probabilité qu'un texte clair est 1.

## Un système parfaitement sûr (masque jetable)

– On prend $P = C = K = 2^n$ (mots de $n$ bits).

– On définit $E_k(p) = p \quad k$ et $D_k(c) = c \quad k$.

– On choisit les clés avec **probabilité uniforme**.

Ce système est **parfaitement sûr**
Mais la clé est **aussi longue** que le message !

## Analyse

On considère le cas $n = 1$ (la généralisation est directe). Soit $P(p = 0) = q$ et $P(p = 1) = (1 - q)$. Le tableau du est :

| $p$ | $k$ | $c = p$ $c$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 1 | 0 |
| 1 | 0 | 1 |

On remarque : $P(c = 0) = P(c = 1) = \frac{1}{2}$. Il suit par exemple :

$$P(p = 0 \mid c = 1) = \frac{q\frac{1}{2}}{\frac{1}{2}} = q = P(p = 0)$$

## Exemple : problème du max en ligne

– On peut lire successivement jusqu'à $n$ entiers **différents** et on cherche à sélectionner **le plus grand entier**.

– La contrainte non-standard est donc que l'on peut sélectionner seulement le **dernier entier lu** (on travaille en ligne). Donc :
  – si l'on décide de lire l'entier en position $i$ on ne pourra pas sélectionner les entiers en position $1, \ldots, i-1$.
  – si l'on sélectionne l'entier en position $i$, il est possible qu'un entier plus grand se trouve dans les positions $i+1, \ldots, n$.

**NB** On retrouve le même problème avec des situations comme mariage, recrutement,...

## Stratégie

La **stratégie** est paramétrée sur un entier $k$ $(1 \leq k < n)$ :

– On lit les premiers $k$ entiers et on mémorise l'entier $m$ le plus grand .

– On lit les entiers en position $k+1, k+2, \ldots, n$ et on sélectionne le premier qui est plus grand que $m$.

En `Java` :

```
static int max_enligne(int p[], int n, int k){
    int m=p[1];
    for (int i=2;i<=k;i++)  {if p[i]>m {m=p[i];}}
    for (int i=k+1;i<n;i++) {if p[i]>m {return p[i];}}
    return p[n];}
```

## Choix optimale de $k$

Une analyse élémentaire permet de conclure qu'il convient de prendre

$$k \approx \frac{n}{e}$$

et que dans ce cas la **probabilité** de sélectionner le maximum est au moins $\frac{1}{e}$.

| Si on lit d'abord 37% des entrées on a une chance de 37% de sélectionner le maximum. |
|---|

## Analyse

– Pour un $n$ **fixé**, soit $Perm_n$ l'**ensemble des permutations** sur $\{1, \ldots, n\}$ et soit $A_k$ l'**algorithme décrit**.

– On peut toujours transformer une suite de $n$ **entiers différents** dans la suite de leurs **positions relatives** (avec 1 pour le plus petit et $n$ pour le plus grand). On a donc :

$$A_k : Perm_n \to \{1, \ldots, n\}$$

– On cherche a calculer la probabilité de l'événement où $A_k$ **sélectionne le maximum** :

$$M = \{\pi \quad Perm_n \mid A_k(\pi) = n\}$$

On pose pour $i = 1, \ldots, n$

$$B_i = \{\pi \quad Perm_n \mid \pi(i) = n\}$$

Comme les $B_i$ forment une **partition**, on a $M = \bigcup_{i=1,\ldots,n}(M \cap B_i)$ et :

$$\begin{aligned} P(M) &= \Sigma_{i=1,\ldots,n} P(M \cap B_i) \\ &= \Sigma_{i=1,\ldots,n} P(M \mid B_i) P(B_i) \end{aligned}$$

La mise en évidence de $P(M \mid B_i)$ va nous permettre de mener à bien l'analyse. . .

En effet on remarque :

$$P(B_i) = \frac{1}{n}$$

$$P(M \mid B_i) = \begin{cases} 0 & \text{si } i \leq k \\ \frac{k}{i-1} & \text{si } i > k \end{cases}$$

**Justification :** si le maximum est en position $i$ alors

– Si $i \leq k$, $A_k$ ne peut pas le sélectionner.

– Si $i > k$, $A_k$ le sélectionne ssi le maximum des éléments en position $1, \ldots, i-1$ est dans les positions $1, \ldots, k$ (**le point clé de l'analyse !**)

On a donc :

$$\begin{aligned} P(M) &= (\Sigma_{i=k+1,\ldots,n} \tfrac{k}{i-1}) \tfrac{1}{n} \\ &= \tfrac{k}{n}(\Sigma_{i=k,\ldots,n-1} \tfrac{1}{i}) \\ &\geq \tfrac{k}{n} \int_k^n \tfrac{1}{x} \mathrm{d}x \\ &= \tfrac{k}{n}(\ln(n) - \ln(k)) \\ &= f(k) \end{aligned}$$

On **étudie la fonction** $f(k)$ pour $1 \leq k \leq (n-1)$ et $n \geq 3$ :

$$\begin{aligned} f'(k) &= \tfrac{1}{n}(\ln(n) - \ln(k) - 1), f'(1) > 0, f'(n-1) < 0 \\ f''(k) &= -\tfrac{1}{nk} < 0 \end{aligned}$$

Il y a donc un **maximum** quand $f'(k) = 0$, à savoir pour $k = \frac{n}{e}$, et dans ce cas $P(M) \geq \frac{1}{e}$.

## Exemple : marche aléatoire avec barrières absorbantes

– On considère une **marche aléatoire en dimension** 1 où l'on démarre d'un point $x \quad \mathbf{N}$ tel que $0 < x < m$.

– A chaque étape du jeu on va soit dans $x + 1$ soit dans $x - 1$ M e ´e t e ´e

## Analyse

- L'ensemble $\Omega$ des **résultats** est constitué de **trajectoires** en zigzag qui commencent par $x$ et s'arrêtent à 0 ou $m$ ou vont à l'infini sans jamais toucher 0 et $m$.
- On considère les **événements** :

$$\begin{array}{ll} A_x & \text{le joueur gagne à partir de } x \text{ E} \\ B & \text{le joueur gagne le prochain jeu} \end{array}$$

- Pour $0 < x < m$ on remarque :

$$A_x = (B \cap A_{x+1}) \quad (B^c \cap A_{x-1})$$

- Soit $p^+(x)$ la probabilité de l'événement $A_x$ on a :

$$\begin{array}{lll} p^+(x) & = \frac{1}{2}(p^+(x+1) + p^+(x-1)) & 0 < x < m \\ p^+(m) & = 1 & \\ p^+(0) & = 0 & \end{array}$$

– On a donc une **relation de récurrence** :

$$p^+(x+1) - 2p^+(x) + p^+(x-1) = 0$$

dont on peut trouver une **solution** en posant $p^+(x) = ax + b$ et en utilisant les **conditions initiales** pour conclure que $a = \frac{1}{m}$ et $b = 0$.

– On obtient donc pour $0 \leq x \leq m$ :

$$p^+(x) = \frac{x}{m} \qquad \text{(probabilité de gagner)}$$

et par un raisonnement similaire on peut dériver que la probabilité que le jouer perd est $p^-(x) = 1 - \frac{x}{m}$. Il suit que la probabilité que le jeu continue à l'infini est 0.

**NB** Si $p < \frac{1}{2}$ (exemple roulette) les chances pour le joueur deviennent très petites... on parle de **ruine du joueur**.

## Borel-Cantelli (loi 1)

Soit $\{A_i\}_{i \in \mathbf{N}}$ une suite d'événements **indépendants** et soit $L = \bigcap_{i \in \mathbf{N}} (\bigcup_{j \geq i} A_i)$

$$\boxed{\text{Si } \Sigma_{i \in \mathbf{N}} P(A_i) = +\infty \text{ alors } P(L) = 1.}$$

## Preuve Borel-Cantelli (loi 1)

Si $B_i = \bigcup_{j \geq i} A_i$, **on sait déjà** que : $P(L) = \lim_{i \to +\infty} P(B_i)$. On **montre** $P(B_i) = 1$. Par **dualité** et **indépendance** :

$$\begin{aligned} P(\bigcup_{j=i,\ldots,m} A_i) &= 1 - P(\bigcap_{j=i,\ldots,m} A_i^c) \\ &= 1 - \Pi_{j=i,\ldots,m}(1 - P(A_i)) \end{aligned}$$

Par l'**inégalité** $(1 + x) \leq e^x$ avec $x = -P(A_i)$ on dérive (attention au changement de direction de l'inégalité) :

$$\begin{aligned} P(\bigcup_{j=i,\ldots,m} A_i) &\geq 1 - \Pi_{j=i,\ldots,m} e^{-P(A_i)} \\ &= 1 - e^{-\Sigma_{j=i,\ldots,m} P(A_i)} \end{aligned}$$

Comme $\Sigma_{i \ \mathbf{N}} P(A_i) = +\infty$ on doit avoir

$$\lim_{m \to +\infty} \Sigma_{j=i,\ldots,m} P(A_i) = +\infty$$

Donc $1 \geq P(B_i) = P(\bigcup_{j=i,\ldots,+\infty} A_i) \geq 1$.

## Application loi 1

– On considère un tirage d'une suite dénombrable de $0, 1$.

– A chaque tirage on obtient 1 avec probabilité $p \quad ]0, 1[$.

– Fixons un mot fini $w \quad 0, 1$ de longueur .

| La probabilité que $w$ soit tiré une infinité de fois est 1. |
|---|

## Argumentation

– Soit $A_i$ l'**événement** où l'on tire le mot $w$ dans les tirages consécutifs : $i\ , i\ +1, \ldots , i\ +(\ \ -1)$.

– Les événements $\{A_i\}_{i\ \ \mathbf{N}}$ sont **indépendants** !

– Par ailleurs, $P(A_i) \geq (min(p, (1-p)))^{\ell} > 0$ (qui ne dépend pas de $i$). Donc $\Sigma_{i\ \ \mathbf{N}} P(A_i) = +\infty$.

– Par la loi 1 on peut **conclure** :

$$P(L) = P(\bigcap_{i\ \ \mathbf{N}} (\bigcup_{j \geq i} A_j)) = 1$$

Tout mot fini est donc tiré infinement souvent avec probabilité 1 mais si le mot est long il faudra être très très patient. . .

## Sommaire

- Probabilité conditionnelle.
- Indépendance (mutuelle et deux par deux).
- Règles de transformation (Bayes).
- Exemple : sécurité parfaite.
- Exemple : max en ligne.
- Exemple : marche aléatoire.
- Borel-Cantelli : loi 1.

# Variables aléatoires discrètes

## Variables aléatoires discrètes

– Soit $(\Omega, A, P)$ un **espace de probabilité**.

– Une variable aléatoire discrète (v.a.d.) est une **fonction** $X : \Omega \to \mathbf{R}$ telle que :

1. $im(X)$ est **dénombrable**.
2. pour tout $x \quad \mathbf{R}$, $X^{-1}(x) \quad A$.

**NB** Dans ce cours le co-domaine d'une v.a.d. est $\mathbf{R}$, mais la notion peut être généralisée à d'autres espaces... D'ailleurs, comme on prend $A = 2^{\Omega}$, la deuxième condition est toujours satisfaite !

## Notation pour v.a.d.

Soit $B \quad \mathbf{R}$ et $X$ v.a.d.

– On écrit $X \quad B$ pour l'**événement** (union dénombrable) :

$$\bigcup\{X^{-1}(x) \mid x \quad im(X) \cap B\}$$

– Comme les $X^{-1}(x)$ sont **disjoints**, on doit avoir :

$$P(X \quad B) = \Sigma_{x \quad im(X)\cap B} P(X^{-1}(x))$$

– On écrit aussi $P(X = x)$ pour $P(X^{-1}(\{x\}))$.

## Exemples de v.a.d.

On reprend le jeu de dé où **le premier qui tire** 6 **gagne**. Rappel :

$$\Omega = \{w6 \mid w \quad T \ \} \quad T^{\mathbf{N}}, \qquad \text{où : } T = \{1, 2, 3, 4, 5\}$$

Si $\omega \quad \Omega$ soit $\omega(i)$ le $i$-ème caractère de $\omega$ (s'il existe). On définit pour $i \quad \mathbf{N}^+$ la v.a.d :

$$X_i(\omega) = \begin{cases} 1 & \text{si } \omega(i) \text{ existe et } \omega(i) = 3 \\ 0 & \text{autrement} \end{cases}$$

On peut aussi définir la v.a.d :

$$Y(\omega) = \begin{cases} \{i \mid \omega(i) = 3\} & \text{si } \omega \text{ fini} \\ -1 & \text{autrement} \end{cases}$$

**Exercice** Calculez $P(X_i = 1)$ et $P(Y = 5)$.

## Sur la terminologie : pourquoi v.a.d. ?

– Une **variable** est un élément/un point d'un ensemble (par exemple $\mathbf{R}$).

– On peut voir une variable réelle comme une **fonction** :

$$x : 1 \to \mathbf{R}$$

où 1 est un ensemble singleton.

– Par analogie, on peut voir une **variable aléatoire** comme une fonction

$$X : \Omega \to \mathbf{R}$$

où $(\Omega, A, P)$ est un espace de probabilité.

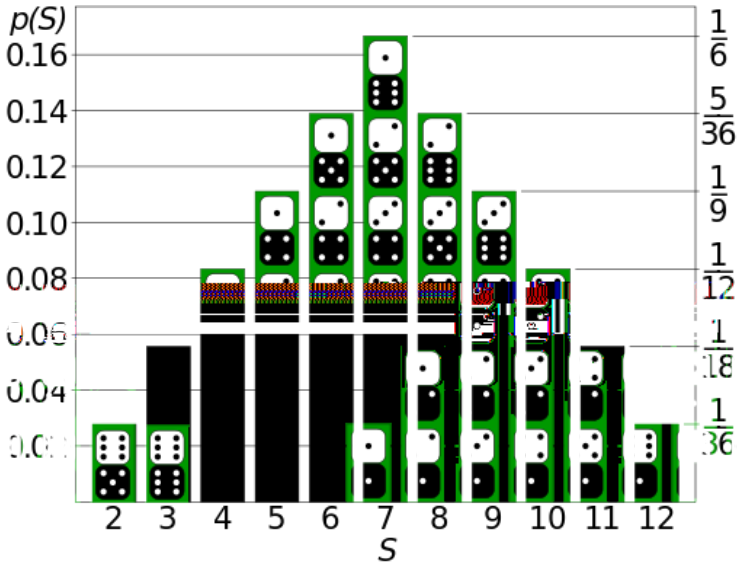
– Par exemple, prenons $\Omega = \{0, 1\}$, $A = 2^{\Omega}$ et $P(\{0\}) = P(\{1\}) = 1/2$, $X(0) = \pi$ et $X(1) = e$. Le 'point' $X$ est $\pi$ avec probabilité $1/2$ et $e$ avec probabilité $1/2$.

## Loi de distribution d'une v.a.d.

Soit $X : \Omega \to \mathbf{R}$ une v.a.d. On définit une fonction $f_X : \mathbf{R} \to \mathbf{R}$ par :

$$f_X(x) = \begin{cases} P(X^{-1}(x)) & \text{si } x \quad im(X) \\ 0 & \text{autrement} \end{cases}$$

On appelle $f_X$ la **loi de distribution** (ou simplement distribution) de $X$. Sa représentation typique est un **histogramme**.



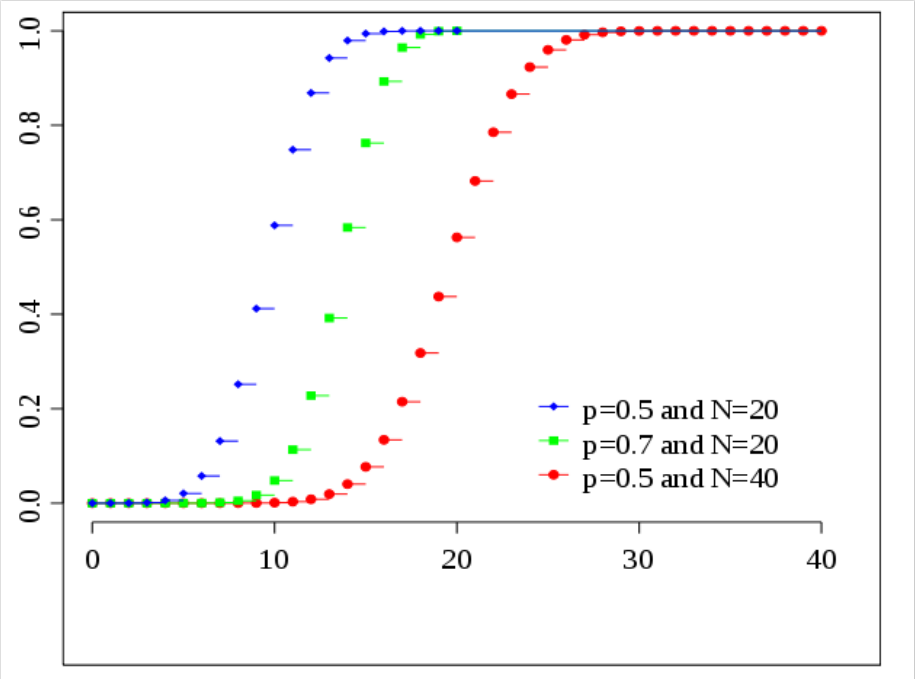

Exemple de distribution : somme jet de deux dés

## Loi de répartition d'une v.a.d.

Soit $X : \Omega \to \mathbf{R}$ une v.a.d. On définit une fonction $F_X : \mathbf{R} \to \mathbf{R}$ par :

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

On appelle $F_X$ la **fonction de répartition**




Exemple de fonction de répartition (pour loi binomiale)

## Remarque

– Dans les calculs analytiques souvent on oublie la v.a.d. $X$ et les détails de l'espace de probabilité et on travaille directement avec la fonction de **distribution** (ou de répartition).

– D'autre part, le **modèle** qui correspond à la distribution est une source d'intuition et parfois permet de simplifier un calcul qui semble compliqué.

On trouve une dialectique similaire, par exemple, entre géométrie analytique et géométrie synthétique.

## Espérance

L'espérance ramène les points d'une v.a.d. à 1 point en prenant une somme pondérée des points par les probabilités. Ceci revient à calculer le **barycentre** (en dimension 1) où la probabilité joue le rôle de la masse.

– Soit $X : \Omega \to \mathbf{R}$ une **v.a.d.**.
– On définit l'**espérance** de $X$ par :

$$E[X] = \Sigma_{x \ \ im(X)} x \cdot P(X = x) = \Sigma_{x \ \ im(X)} x \cdot f_X(x)$$

en supposant que la série en question **converge absolument**.

L'espérance est le barycentre d'une distribution

## Exemple : v.a.d. de Bernoulli

– On tire une pièce à pile ou face (1 ou 0). La probabilité de pile est $p$ et chaque tirage est indépendant.

– On peut prendre $\Omega = \{0, 1\}$ et $X : \Omega \to \{0, 1\}$ où :

$$X(\omega) = \begin{cases} 1 & \text{si } \omega = 1 \\ 0 & \text{autrement} \end{cases}$$

On obtient la **v.a.d. de Bernoulli** avec espérance $E[X] = p$.

## Exemple de v.a.d. avec espérance non-définie

– On prend $X : \mathbf{N}^+ \to \mathbf{R}$ avec :

$$X(i) = 2^i \qquad P(X = 2^i) = 1/(2^i)\ .$$

– Il s'agit bien d'une **probabilité** car :

$$\Sigma_{i=1,\ldots,\infty}\,(1/2)^i = 1$$

– Mais l'espérance va à l'**infini** :

$$E[X] = \Sigma_{i\ \mathbf{N}^+}(2^i)(1/2^i) = \Sigma_{i\ \mathbf{N}^+} 1 = +\infty$$

## Un autre exemple avec espérance non-définie

– On prend $X : \mathbf{N}^+ \to \mathbf{R}$ avec :

$$X(i) = i \qquad P(X = i) = \frac{c}{i\sqrt{i}}$$

On vérifie que $\Sigma_{i=1,\ldots,\infty} \frac{1}{i\sqrt{i}}$ **converge** en approximant avec une intégrale. Ensuite on choisi $c$ de façon à normaliser la somme à 1.

– D'autre part l'espérance va à l'**infini** :

$$E[X] = \Sigma_{i \geq 1} \frac{c\, i}{i\sqrt{i}} = c\ \Sigma_{i \geq 1} \frac{1}{\sqrt{i}} = +\infty$$

*Cette exemple apparaît effectivement dans une variante de la* **marche aléatoire** *avec barrières absorbantes où la barrière supérieure va à* $+\infty$ *. Dans ce cas,* **avec probabilité** 1 **le jeu termine** *mais la* **moyenne du temps de jeu est** $+\infty$ *!*

**Interprétation physique** Si l'on revient à l'interprétation de l'espérance comme le barycentre en dimension 1, ceci veut dire qu'il est possible (en théorie) de distribuer une masse finie sur une droite de façon à que le **point d'équilibre aille à l'infini**.

**Convention** Dans les calculs, on fait l'hypothèse que les espérances en question existent, c.a.d. les séries **convergent absolument**.

## Distribution conjointe et distribution marginale

Soient $X, Y : \Omega \to \mathbf{R}$ v.a.d. On définit la **distribution conjointe de X et Y** par :

$$f(x,y) = P(X = x, Y = y)$$

Notez que $f$ est définie sur $im(X) \times im(Y)$ qui est dénombrable. A partir de la distribution conjointe on dérive les **distributions marginales de X et Y** par :

$$f_X(x) = \Sigma_{y \ \ im(Y)} f(x,y) \quad f_Y(y) = \Sigma_{x \ \ im(X)} f(x,y)$$

**N B** On utilise souvent :

$$\begin{aligned} P(X = x) &= \Sigma_{y \in im(Y)} P(X = x, Y = y) \\ P(Y = y) &= \Sigma_{x \in im(X)} P(X = x, Y = y) \end{aligned}$$

## Exemple

Soient $X$ et $Y$ les v.a.d. qui enregistrent les résultat du jet de deux dés non biaisés. La **distribution conjointe** de $X$ et $Y$ est :

$$f(x,y) = \frac{1}{36} \quad (x,y) \quad \{1,\ldots,6\}^2$$

Si maintenant l'on pose $Z = 7 - X$ alors la **distribution conjointe** de $X$ et $Z$ est :

$$g(x,z) = \begin{cases} \frac{1}{6} & \text{si } (x,z) \quad \{1,\ldots,6\}^2, 7 - x = z \\ 0 & \text{autrement} \end{cases}$$

Notez que : $f_X = g_X = f_Y = g_Z$. Donc la distribution conjointe **détermine** les distributions marginales, mais la **réciproque est fausse**.

## Sur le calcul de l'espérance

Soient $X, Y : \Omega \to \mathbf{R}$ v.a.d. et $f : \mathbf{R}^2 \to \mathbf{R}$ une fonction. Alors :

1. $(f(X,Y) = z) = \bigcup_{x \ \ im(X), y \ \ im(Y), z = f(x,y)} (X = x, Y = y)$.

2. $E[f(X,Y)] = \Sigma_{x \ \ im(X), y \ \ im(Y)} \ f(x,y) \cdot P(X = x, Y = y)$.

## Preuve (2)

$$
\begin{aligned}
& E[f(X,Y)] \\
&= \Sigma_{z \;\; im(f(X,Y))} \; z \cdot P(f(X,Y) = z) \\
&= \Sigma_{z \;\; im(f(X,Y))} \; z \cdot (\Sigma_{x \;\; im(X), y \;\; im(Y), z=f(x,y)} \; P(X = x, Y = y)) \quad (\text{par } (1)) \\
&= \Sigma_{x \;\; im(X), y \;\; im(Y)} \; f(x,y) \cdot P(X = x, Y = y)
\end{aligned}
$$

## Linéarité de l'espérance

Si $X_i : \Omega \to \mathbf{R}$ pour $i = 1, \ldots, n$ sont des v.a.d. avec espérance $E[X_i]$ alors toute **combinaison linéaire** $\Sigma_{i=1,\ldots,n} c_i \cdot X_i$ est une v.a.d. avec espérance :

$$E[\Sigma_{i=1,\ldots,n} c_i \cdot X_i] = \Sigma_{i=1,\ldots,n} c_i \cdot E[X_i]$$

**Méthode** pour calculer $E[X]$ on cherche à exprimer $X$ comme combinaison linéaire de v.a.d. $X_i$ dont on sait calculer l'espérance.

## Preuve pour $n = 2$ et $c_1 = c_2 = 1$

$$
\begin{aligned}
& E[X + Y] \\
&= \Sigma_{x \ \ im(X), y \ \ im(Y)} (x + y) \cdot P(X = x, Y = y) \qquad \text{(par (2))} \\
&= \Sigma_{x \ \ im(X), y \ \ im(Y)} \ x \cdot P(X = x, Y = y) + \\
&\quad \Sigma_{x \ \ im(X), y \ \ im(Y)} \ y \cdot P(X = x, Y = y) \\
&= \Sigma_{x \ \ im(X)} \ x \cdot P(X = x) + \Sigma_{y \ \ im(X)} \ y \cdot P(Y = y) \\
&= E[X] \ + \ E[Y]
\end{aligned}
$$

## Indépendance et produit de v.a.d.

Deux v.a.d. $X, Y : \Omega \to \mathbf{R}$ sont **indépendantes** si pour tout $x, y \quad \mathbf{R}$ :

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

où $X = x, Y = y$ est une notation pour l'événement $X^{-1}(\{x\}) \cap Y^{-1}(\{y\})$.

Si $X$ et $Y$ sont **indépendantes** alors :

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

## Preuve

$$
\begin{aligned}
& E[X \cdot Y] \\
&= \Sigma_{x \ \ im(X), y \ \ im(Y)} \ x \cdot y \cdot P(X = x, Y = y) && \text{(par (2))} \\
&= \Sigma_{x \ \ im(X), y \ \ im(Y)} \ x \cdot y \cdot P(X = x) \cdot P(Y = y) && \text{(par indépendance)} \\
&= (\Sigma_{x \ \ im(X)} \ x \cdot P(X = x)) \cdot (\Sigma_{y \ \ im(Y)} \ y \cdot P(Y = y)) \\
&= E[X] \cdot E[Y]
\end{aligned}
$$

## Conditionnement d'une v.a.d. et son espérance

Soit $X : \Omega \to \mathbf{R}$ une v.a.d. et $B$ un événement tel que $P(B) > 0$. On définit une distribution de $X$ conditionnée sur $B$ par :

$$f_{X|B}(x) = P(X = x \mid B) = \frac{P(X = x \cap B)}{P(B)}$$

On peut vérifier que $f_{X|B}(x) = 0$ si $x \not\in im(X)$ et

$$\Sigma_{x \in im(X)} f_{X|B}(x) = 1$$

Donc il s'agit bien d'une **distribution** et on peut définir son **espérance** par :

$$E[X \mid B] = \Sigma_{x \in im(X)} x f_{X|B}(x)$$

Si $\{B_i\}_{i \in I}$ est une partition au plus dénombrable de $\Omega$ on a l'identité suivante (**exercice !**) :

$$E[X] = \Sigma_{i \in \mathbf{N}} E[X \mid B_i] P(B_i)$$

# Distributions et leur espérance

## Exemple : v.a.d. binomiale

On prend $\Omega = \{0,1\}^n$ et $X : \omega \to \{0,1,\ldots,n\}$ où

$$P(X = k) = \text{ probabilité que il y ait } k \text{ piles dans } n \text{ essais}$$

Ceci revient à compter les combinaisons de $k$ éléments parmi $n$ :

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

La **loi binomiale** assure que :

$$\Sigma_{k=0,\ldots,n} \binom{n}{k} \cdot p^k \cdot (1-p)^{(n-k)} = (p + (1-p))^n = 1$$

$X$ est **v.a.d. binomiale** avec paramètres $n$ et $p$.

## Binomiale = Somme Bernoulli

Sur le même espace de probabilité $\Omega = \{0,1\}^n$ on peut définir $n$ v.a.d. (indépendantes) de **Bernoulli** $X_1, \ldots, X_n$ où :

$$X_i(\omega) = \begin{cases} 1 & \text{si le i-ème tirage est 1} \\ 0 & \text{autrement} \end{cases}$$

Alors pour tout $\omega$ et $0 \leq k \leq n$ :

$$X(\omega) = X_1(\omega) + \cdots + X_n(\omega)$$

Par la **linéarité de l'espérance** on peut conclure :

$$E[X] = E[\Sigma_{i=1,\ldots,n} X_i] = \Sigma_{i=1,\ldots,n} E[X_i] = np$$

## Exemple : v.a.d. géométrique

On pose maintenant $\Omega = \{0,1\}^{\mathbf{N}}$. On définit :

$$X(\omega) = \begin{cases} k & \text{si le premier 1 apparaît après } k \text{ essais} \\ 0 & \text{autrement} \end{cases}$$

On suppose $P(X = 0) = 0$ et pour $k \geq 1$

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

On a bien (**série géométrique**) :

$$\Sigma_{i=0,\ldots,\infty} (1 - p)^k \cdot p = \frac{1}{1 - (1 - p)} \cdot p = 1$$

## Espérance v.a.d. géométrique

On pose $q = (1 - p)$ et on utilise une propriété de commutation entre dérivation et série :

$$\begin{aligned}
E[X] &= \Sigma_{k=1,\ldots,\infty}\, k \cdot q^{(k-1)} \cdot p \\
&= p \cdot (\Sigma_{k=0,\ldots,\infty}\, (\tfrac{\mathrm{d}}{\mathrm{d}x} x^k)(q)) \\
&= p \cdot (\tfrac{\mathrm{d}}{\mathrm{d}x} \Sigma_{k=0,\ldots,\infty}\, x^k)(q) \\
&= p \cdot (\tfrac{\mathrm{d}}{\mathrm{d}x} (\tfrac{1}{(1-x)}))(q) \\
&= p \cdot (\tfrac{1}{(1-x)^2})(q) \\
&= p \cdot \tfrac{1}{p^2} \\
&= \tfrac{1}{p}
\end{aligned}$$

Si la pièce n'est pas biaisée ($p = 1/2$) en moyenne il faut 2 tirages pour avoir pile.

## Propriété remarquable

Soit $X : \mathbf{N} \to \mathbf{R}$ une v.a.d. On dit que $X$ est **sans mémoire** si pour tout $n, k \quad \mathbf{N}$ :

$$P(X > n + k \mid X > n) = P(X > k)$$

Une v.a.d. sur $\mathbf{N}$ est **sans mémoire** ssi elle est **géométrique**.

## Preuve : une variable géométrique est sans mémoire

Soit $X$ géométrique avec paramètre $p$. En utilisant la série géométrique on déduit :

$$P(X > n) = (1-p)^n$$

Il suit que :

$$\begin{aligned} & P(X > n+k \mid X > n) \\ &= \frac{P(X>n+k)}{P(X>n)} \\ &= \frac{(1-p)^{(n+k)}}{(1-p)^n} \\ &= (1-p)^k \\ &= P(X > k) \end{aligned}$$

## Preuve : une v.a.d. sur N sans mémoire est géométrique

Soit $X : \mathbf{N} \to \mathbf{R}$ une v.a.d. sans mémoire. On a donc pour tout $n, k \quad \mathbf{N}$ :

$$P(X > n + k) = P(X > n)P(X > k)$$

Pour $n = k = 0$, ceci force $P(X > 0) \quad \{0, 1\}$. Si $P(X > 0) = 0$ alors $P(X = 0) = 1$ : un **cas dégénéré**.

Sinon, $P(X > 0) = 1$ et soit $p = 1 - P(X > 1) < 1$. Alors pour $k \quad \mathbf{N}$ :

$$P(X > k) = (1 - p)^k$$

On remarque pour $k \quad \mathbf{N}$ :

$$\begin{aligned} P(X = k + 1) &= P(X > k) - P(X > k + 1) \\ &= (1 - p)^k - (1 - p)^{k+1} \\ &= (1 - p)^k(1 - (1 - p)) \\ &= (1 - p)^k p \end{aligned}$$

## Tirage jusqu'à $r$ succès

On tire dans $\{0, 1\}$ où la probabilité de tirer 1 est $p$.

## Loi binomiale négative

Soit $\Omega$ l'ensemble des suites de tirages jusqu'à $r$ succès. (Il est facile de voir que de toute façon la probabilité de ne jamais arriver à $r$ succès est 0).

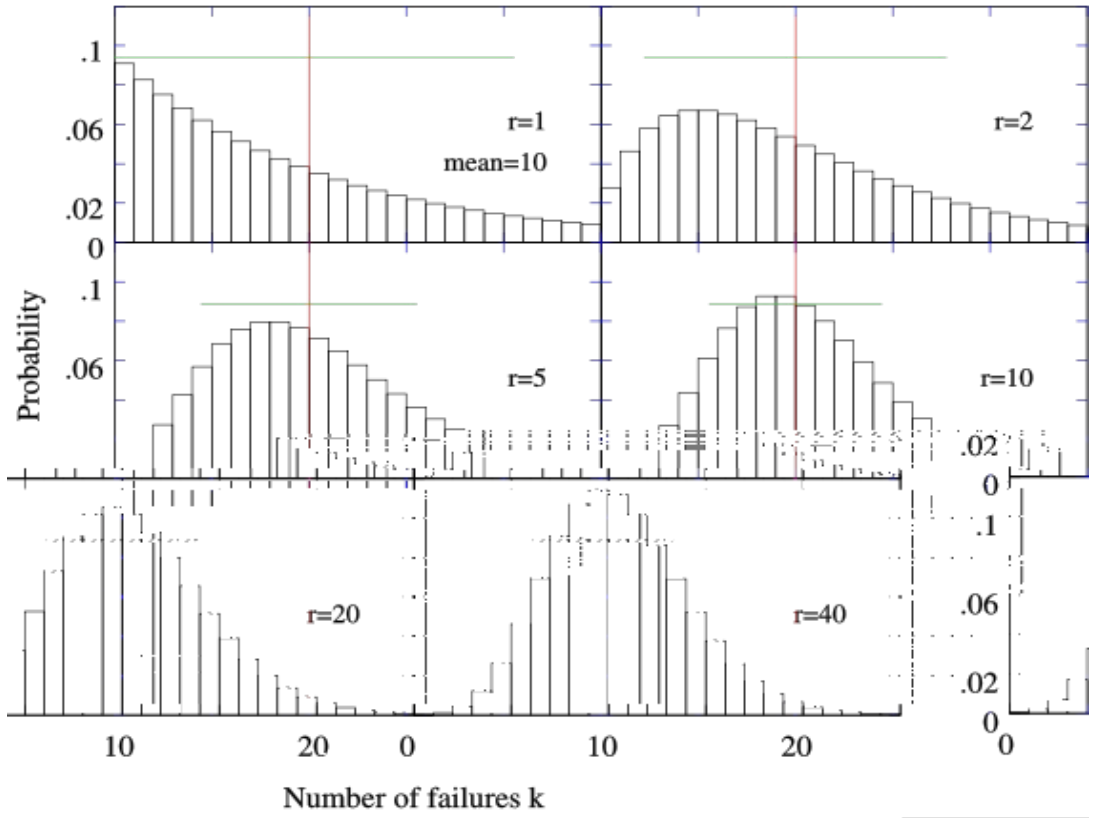
On définit $X : \Omega \to \mathbf{N}$ par

$$X(\omega) = n \text{ si } |\omega| = n$$

On a pour $n \quad \mathbf{N}$ et $n \geq r$ :

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

**NB** Si $r = 1$ on retrouve la **loi géométrique**.




Dis i i i i i s iss s s i

## Binomiale négative = Somme géométriques

On définit $r$ v.a.d. $X_1, \ldots, X_r$ (indépendantes) sur le même espace de probabilité par $X_i(\omega) = n_i$ si

> $n_i$ est le nombre de tirages pour passer dans $\omega$ du $(i-1)$-ème 1 au $i$-ème 1

Alors $X = \Sigma_{i=1,\ldots,r} X_i$ et

$$E[X] = \frac{r}{p}$$

**Terminologie :** l'adjectif 'négative' vient du fait que parfois la loi est formalisée en utilisant un 'coefficient binomial négatif'.

## Variante : problème des urnes

– On dispose de $n$ **urnes**.

– On a l'**expérience élémentaire** suivante : on sélectionne une urne de façon aléatoire et on y dépose une boule.

– Combien de fois faut-il répéter l'expérience élémentaire en **moyenne** pour que toutes les urnes contiennent **au moins** une boule ?

## Analyse

- Soit $\Omega$ l'ensemble des suites de distribution de boules jusqu'à que toutes les urnes contiennent au moins une boule.
- On définit $X : \Omega \to \mathbf{N}$ par

$$X(\omega) = k \text{ si } |\omega| = k$$

  Par exemple si $n = 2$ et $\omega = 112$ on aura $X(\omega) = 3$.
- On considère aussi les v.a.d. $X_i$ pour $i = 1, \ldots, n$, où $X_i(\omega) = k$ si $k$ est le nombre de boules nécessaires pour passer de $(i - 1)$ à $i$ urnes remplies.
- Par exemple, si $n = 2$ et $\omega = 112$ on aura :

$$X(\omega) = 3, \qquad X_1(\omega) = 1 \qquad X_2(\omega) = 2$$

## On applique la linéarité de l'espérance

– On a :

$$X = \Sigma_{i=1,\ldots,n} X_i \qquad E[X] = \Sigma_{i=1,\ldots,n} E[X_i]$$

– On observe que $X_i$ correspond à une v.a.d. **géométrique** avec coefficient :

$$p_i = 1 - \frac{(i-1)}{n}$$

Donc $E[X_i] = \frac{1}{p_i}$. **NB** Le paramètre $p_i$ est fonction de $i$ (ce qui n'est pas le cas pour la binomiale négative).

$$\begin{aligned} E[X] &= \Sigma_{i=1,\ldots,n} \frac{n}{(n-i+1)} \\ &= n \cdot \Sigma_{i=1,\ldots,n} \frac{1}{i} \\ &\leq n \cdot (\ln(n) + 1) \end{aligned}$$

En moyenne pour remplir $n$ urnes il faut $n\ln(n)$ boules

## Variante : problème du collectionneur de coupons

– Il y a $n$ **coupons**.

– Chaque **boite** contient un coupon (avec probabilité uniforme).

– Combien de **boites** faut-il acheter en moyenne pour avoir tous les coupons ?

Il suffit de voir les coupons comme des urnes et les boites comme des boules.

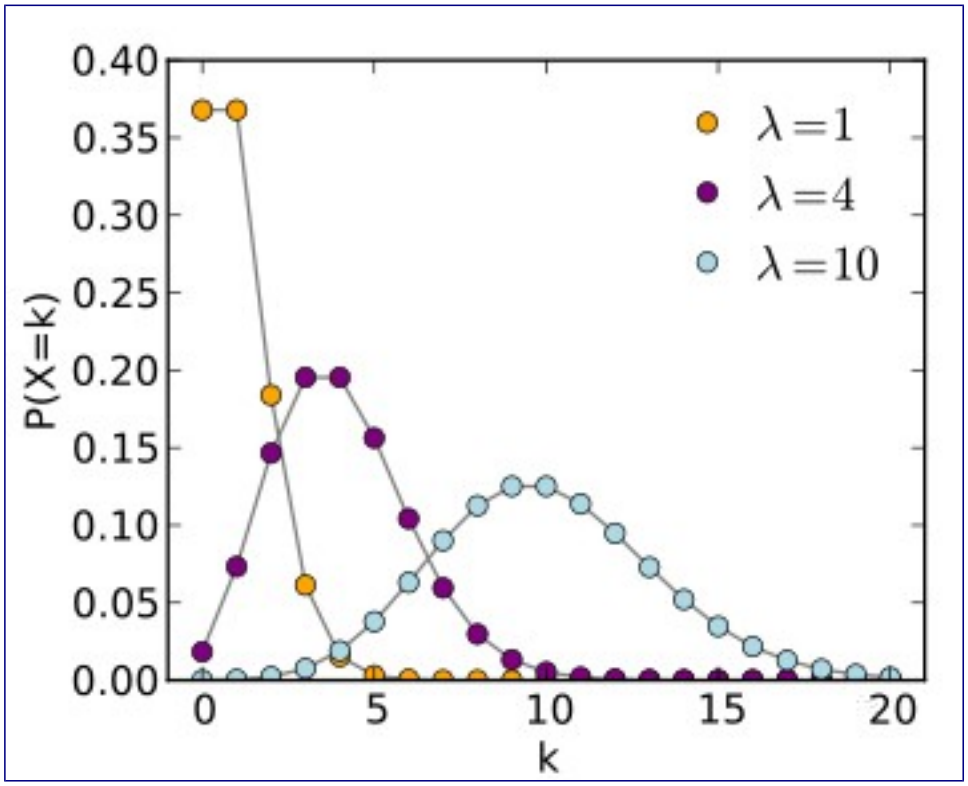
## Une autre distribution importante : Poisson

Une v.a.d. de Poisson de paramètre $\lambda$ $\mathbf{R}$ est une v.a.d. $X : \Omega \to \mathbf{N}$ telle que pour $k$ $\mathbf{N}$

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

En utilisant le **développement de Taylor** de la fonction exponentielle on vérifie qu'il s'agit bien d'une v.a.d. :

$$\begin{aligned}
&\Sigma_{k=0,\ldots,\infty} \tfrac{e^{-\lambda} \cdot \lambda^k}{k!} \\
&= e^{-\lambda} \cdot (\Sigma_{k=0,\ldots,\infty} \tfrac{\lambda^k}{k!}) \\
&= e^{-\lambda} \cdot e^{\lambda} \\
&= 1
\end{aligned}$$




Distribution Poisson

## Espérance v.a.d. de Poisson

On utilise encore le **développement de Taylor** :

$$\begin{aligned} E[X] &= \Sigma_{k=0,\ldots,\infty}\, k\, \frac{e^{-\lambda}\cdot\lambda^k}{k!} \\ &= e^{-\lambda}\cdot\left(\Sigma_{k=1,\ldots,\infty}\, \frac{\lambda^k}{(k-1)!}\right) \\ &= \lambda\cdot e^{-\lambda}\cdot\left(\Sigma_{k=0,\ldots,\infty}\, \frac{\lambda^k}{k!}\right) \\ &= \lambda \end{aligned}$$

## Approximation loi binomiale par loi de Poisson

– Une v.a.d binomiale $B$ de paramètres $n$ et $p$ peut être approximée par une v.a.d. de Poisson $P$ de paramètre $\lambda = np$.

– Ceci est cohérent avec le fait qu'on s'attend à que $E[B] = E[P]$.

– Cette approximation est **assez bonne** quand :

  – $n$ est grand ($n \approx 10^2$),
  – $p$ est petit ($p \approx 10^{-2}$) et
  – $B(k), P(k)$ ne sont pas trop petits/grands.

# Exemples

```
Paramètres: n=25, p=0.16, lambda=4
Input  Binomial>1%          Poisson              Erreur relatif<5%
3      0.20334433477152622  0.19536681481316465  0.03923158207148953
6      0.10820439280359379  0.10419563456702115  0.037048017485289145
Total avec erreur relatif <5% =0.31154872757512

---

Paramètres: n=50, p=0.08, lambda=4
Input Binomial>1%          Poisson              Erreur relatif<5%
2     0.14326225250667887  0.14652511110987348  0.022775424412948542
3     0.19932139479190122  0.19536681481316465  0.01984021827092522
4     0.20365446859172487  0.19536681481316465  0.04069468171196899
5     0.16292357487338008  0.15629345185053173  0.04069468171196899
6     0.10625450535220439  0.10419563456702115  0.0193767857500127
7     0.05807699671424839  0.05954036260972637  0.025196996715894
Total avec erreur relatif <5% =0.8734931928301377
```

```
Paramètres: n=100, p=0.04, lambda=4
Input Binomial>1%          Poisson                Erreur relatif<5%
1     0.07029299732853991  0.07326255555493674  0.04224543466993615
2     0.14497930699011333  0.14652511110987348  0.010662239679939787
3     0.19733294562543244  0.19536681481316465  0.009963520313530426
4     0.19938849714236417  0.19536681481316465  0.02017008195978342
5     0.15951079771389146  0.15629345185053173  0.020170081959784064
6     0.1052328179362476   0.10419563456702115  0.00985608282251646
7     0.05888026717861478  0.05954036260972637  0.011210809032322732
8     0.028520129414641555 0.029770181304863187 0.043830512549493696
Total avec erreur relatif <5% =0.9641377593298452
```

```
Paramètres: n=200, p=0.02, lambda=4
Input Binomial>1%          Poisson              Erreur relatif<5%
0     0.0175879466057215   0.0183156388888734186 0.04137448784248416
1     0.071787537166210003  0.07326255555493674  0.020546998085636994
2     0.14577265200077336  0.14652511110987348  0.005161867461230909
3     0.19634683738879719  0.19536681481316465  0.004991282715147288
4     0.19734860696731144  0.19536681481316465  0.010042088386643895
5     0.15787888557384933  0.15629345185053173  0.010042088386644494
6     0.10471558737041024  0.10419563456702115  0.004965381147601908
7     0.05922689198209791  0.05954036260972637  0.0052927077065468 3
8     0.029160178960573744 0.029770181304863187 0.020919019225300407
9     0.012695588118889248 0.013231191691050305 0.04218816545916092 4
Total avec erreur relatif <5% =0.992520712134634
```

**NB** Ces données contiennent des approximations dues à la représentation des réels avec le type `double` de `Java`.

## Poisson comme limite de Binomiale

– Soit $\lambda > 0$ et $P$ la **distribution de Poisson** associée.

– Soit $B$ une **distribution binomiale** avec paramètres $n$ et $p = \frac{\lambda}{n}$ (donc si $n \to +\infty$ alors $p \to 0$).

– **Fixons $k$**. Alors on peut montrer (preuve omise) que :

$$\begin{aligned}
&\lim_{n \to +\infty} B(k) \\
&= \lim_{n \to +\infty} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{1}{e^\lambda} \lambda^k k! \\
&= P(k)
\end{aligned}$$

Qui se réécrit comme :

$$\mu = \frac{nN_1}{N} \Sigma_{k=0,\ldots,n-1} \frac{\binom{N_1-1}{k}\binom{N-N_1}{n-1-k}}{\binom{N-1}{n-1}}$$

On retrouve donc la somme d'un distribution hypergéométrique avec paramètres $N-1$, $N_1-1$ et $n-1$ **qui vaut** 1.

Par ailleurs, en utilisant $N_1 = pN$ on peut conclure par : $\mu = \mathbf{np}$.

## Sommaire

- **V.a.d.** et fonctions de **distribution** et de **répartition** associées.
- **Espérance** et sa propriété de **linéarité**.
- Pour v.a.d. **indépendantes**, l'**espérance du produit** est le produit de l'espérance.
- **Distributions** : Bernoulli, binomiale, géométrique, binomiale négative, Poisson, hypergéométrique.
- **Absence de mémoire** de la loi géométrique.
- **Approximation** de la loi binomiale par la loi de Poisson.

# Mémento

| **Distribution avec paramètres** | **Espérance** |
|---|---|
| Bernoulli $p \quad [0,1]$ : $f_X(x) = \begin{cases} p & \text{si } x = 1 \\ (1-p) & \text{si } x = 0 \end{cases}$ | $E[X] = p$ |
| Binomiale $n \quad \mathbf{N}, p \quad [0,1]$ $f_X(k) = \binom{n}{k} p^k (1-p)^{(n-k)} \quad k \quad \{0,\ldots,n\}$ | $E[X] = np$ |
| Géométrique $p \quad [0,1]$ $f_X(k) = (1-p)^{k-1} p \quad k \quad \mathbf{N}^+$ | $E[X] = \frac{1}{p}$ |

**NB** Les distributions valent 0 la où elles ne sont pas spécifiées.

| **Distribution avec paramètres** | **Espérance** |
|---|---|
| Binomiale négative $r \quad \mathbf{N}^+, p \quad [0,1]$<br>$f_X(n) = \binom{n-1}{r-1} p^r (1-p)^{(n-r)} \quad n \quad \mathbf{N}, n \geq r$ | $E[X] = \frac{r}{p}$ |
| Poisson $\lambda \quad \mathbf{R}$<br>$f_X(n) = \frac{e^{-\lambda} \cdot \lambda^n}{n!} \quad n \quad \mathbf{N}$ | $E[X] = \lambda$ |
| Hypergéométrique $p \quad [0,1], N, pN, n \quad \mathbf{N}$<br>$f_X(k) = \frac{\binom{pN}{k} \cdot \binom{(1-p)N}{n-k}}{\binom{N}{n}} \quad k \quad \{0, \ldots, n\}$ | $E[X] = np$ |

# Algorithmes et Probabilités

## Conception et Analyse

On peut introduire une **composante aléatoire** dans la **conception** et/ou l'**analyse** d'un algorithme. On distingue :

**Algorithme probabiliste** On considère un algorithme qui à certains moments du calcul **joue à pile ou face** pour déterminer son prochain état. On définit une v.a.d. $X$ qui associe à chaque exécution possible (sur une entrée fixée) son **coût d'exécution** et on cherche à calculer $E[X]$.

**Analyse en moyenne** On fait une hypothèse sur la **probabilité des entrées**. On définit une v.a.d. $X$ qui associe à chaque entrée son **coût d'exécution** et on cherche à calculer $E[X]$.

## Générateurs dans les langages de programmation

– Dans un **algorithme probabiliste**, on peut invoquer une fonction qu'on appelle **générateur** qui produit un nombre dans $\{0, 1\}$ avec une **probabilité uniforme**.

– Dans `Java`, on utilise la classe `Random` pour créer un objet `g` et on génère un élément en invoquant la méthode `nextInt(2)`.

```
import java.util.Random;              // on importe la classe Random
...
Random g = new Random();              // création d'un objet de la classe
...
int x=g.nextInt(2);                   // x prend 0 ou 1
```

– **Hypothèse :** les résultats d'une suite d'invocations du générateur sont **indépendants**. Donc la probabilité d'obtenir une suite $w \quad \{0, 1\}^+$ est $2^{-|w|}$.

## Probabilité de terminaison d'un programme

– Fixons un **algorithme/programme** $A$ et une **entrée** $i$ du programme.

– Une **exécution** de $A(i)$ est une **suite de configurations** traversées par le programme à partir de la configuration initiale.

– Si l'exécution **termine** alors la suite est **finie** sinon elle **infinie** (dénombrable). Dans les algorithmes **déterministes** l'exécution est unique (à équivalence près), mais dans les algorithmes **probabilistes** on peut avoir plusieurs exécutions (pour la même entrée).

– Soit $\Omega$ l'ensemble des **exécutions finies** de $A(i)$. Pour tout $\omega \quad \Omega$ on définit :

| | |
|---|---|
| $rnd(\omega) \quad \{0,1\}$ | les bits aléatoires utilisés dans $\omega$ |
| $r(\omega) = \lvert rnd(\omega) \rvert$ | longueur $rnd(\omega)$ |
| $p(\omega) = 2^{-r(\omega)}$ | 'probabilité' de l'exécution de $\omega$ |

– La **'probabilité' que l'algorithme $A$ termine sur l'entrée $i$** est alors :

$$\Sigma_{\omega \ \ \Omega} p(\omega) = \Sigma_{\omega \ \ \Omega} 2^{-r(\omega)}$$

Par extension, on dit qu'un algorithme $A$ termine avec probabilité 1 si pour toute entrée il termine avec probabilité 1.

## Remarques

1. $p$ n'est pas forcement une probabilité, mais au moins on a :

$$\Sigma_{\omega \ \Omega} p(\omega) \leq 1$$

2. Si $\Sigma_{\omega \ \Omega} p(\omega) = 1$ alors on peut définir un **espace de probabilité discret** :

$$(\Omega, 2^{\Omega}, P)$$

avec pour $A \quad \Omega$, $P(A) = \Sigma_{\omega \ A} p(\omega)$.

3. Un algorithme qui termine avec probabilité 1 **n'est pas** un algorithme dont toutes les exécutions sont finies (mais les exécutions infinies ont probabilité 0).

## **Pourquoi** $\Sigma_{\omega \ \Omega} p(\omega) \leq 1$ **?**

– Si $w, w$ sont deux **mots**, on écrit $w \leq w$ si $w$ est un **préfixe** de $w$ .

– Si $\omega = \omega$ alors $rnd(\omega) \leq rnd(\omega)$.

– Donc :

$$R = \{rnd(\omega) \mid \omega \quad \Omega\}$$

est un ensemble de mots $\{0, 1\}$ qui sont **incomparables par rapport au préfixe**. Par exemple : $R = \{1, 01, 001, 0001, \ldots\}$.

– On sait (TD Axiomatique) que si en plus $R$ est **fini** alors

$$\Sigma_{w \ \ R} 2^{-|w|} \leq 1$$

– Si maintenant $R$ et **dénombrable** on pose :

$$R_n = \{w \quad R \mid |w| \leq n\}$$

comme $R_n \quad R_{n+1}$ :

$$\Sigma_{w \ \ R_n} 2^{-|w|} \leq \Sigma_{w \ \ R_{n+1}} 2^{-|w|} \leq 1$$

Donc :

$$\Sigma_{w \ \ R} 2^{-|w|} = \lim_{n \to +\infty} \Sigma_{w \ \ R_n} 2^{-|w|} \leq 1$$

## Algo. Probabiliste = Algo. Non-déterministe

Pour simplifier la comparaison, considérons des **algorithmes de décision** qui terminent toujours et qui rendent comme résultat 1 (accepte) ou 0 (refuse).

**Algorithme non-déterministe :** on rend 1 s'il y a une exécution qui rend 1. En pratique, on est obligé d'explorer toutes les exécutions possibles jusqu'à en trouver une qui accepte ou à épuiser toutes les possibilités.

**Algorithme probabiliste :** on explore un seul chemin d'exécution (guidé par le générateur aléatoire) et on s'arrange pour que la réponse finale soit correcte (presque toujours).

## Temps moyen de calcul d'un algorithme probabiliste

– On suppose que $A(i)$ termine avec probabilité 1.

– Alors on peut définir une v.a.d. $C$ qui associe à chaque exécution finie un **coût**. Par exemple on peut prendre :

$$C(\omega) = |\omega|$$

en considérant que la **longueur de l'exécution** correspond en gros au **temps de calcul**.

– Ensuite on peut calculer l'espérance $E[C]$ qui est donc le **coût moyen de l'algorithme $A$ sur l'entrée $i$** :

$$E[C] = \Sigma_{\omega \ \Omega} |\omega| 2^{-r(\omega)}$$

## En résumé

Pour un algorithme probabiliste $A$ avec entrée $i$, soit $\Omega$ l'ensemble des **exécutions finies** et pour $\omega \quad \Omega$ soit $r(\omega)$ le nombre de bits aléatoires utilisés dans $\omega$. Alors :

– La **probabilité de terminaison** est :

$$\Sigma_{\omega \ \ \Omega} \ 2^{-r(\omega)}$$

– Le **temps moyen de calcul** est :

$$\Sigma_{\omega \ \ \Omega} \ |\omega| \ 2^{-r(\omega)}$$

## Exemple

La fonction `proba1` termine-t-elle ?

```
import java.util.Random;
...
public static void proba1(){
        Random g = new Random();
        while (true) {if (g.nextInt(2)==1){break;} }}
```

## Analyse

On **suppose** que :

$$P(\texttt{g.nextInt(2)} == 1) = \frac{1}{2}$$

La probabilité de **terminer exactement à la la n-ième itération** est :

$$\frac{1}{2^n}$$

Donc la probabilité de **terminer dans les premières $n$ itérations** est :

$$\Sigma_{i=1,\ldots,n} \frac{1}{2^n} = 1 - \frac{1}{2^n}$$

et la probabilité de **terminer tout court** est :

$$\Sigma_{i=1,\ldots,\infty} \frac{1}{2^n} = 1$$

On réconnaît ici une **distribution géométrique** avec paramètre $\frac{1}{2}$. Le **coût moyen** de l'algorithme est donc 2 (itérations).

## Exemple

La fonction `proba2` termine-t-elle ?

```
public static void proba2(){
        long n=1;
        boolean stop=false;
        Random g = new Random();
        while (!stop) {stop=true;
                       for (int i=0;i<n;i++){
                       if (g.nextInt(2)==1){stop=false; break;}}
                       n=n+1;}}
```

## Analyse

On se **souvient** que :

$$(a) \quad 1 + x \leq e^x \text{ et } (b) \quad \Sigma_{i=1,\ldots,n} \frac{1}{2^i} = 1 - \frac{1}{2^n}$$

La probabilité $p_n$ de **terminer exactement à la la n-ième itération** est :

$$p_n = (\Pi_{i=1,\ldots,(n-1)}(1 - \frac{1}{2^i})) \cdot \frac{1}{2^n}$$

En utilisant $(a)$ et $(b)$ on a pour $n \geq 1$ :

$$(\Pi_{i=1,\ldots,n}(1 - \frac{1}{2^i})) \leq \sqrt{\frac{1}{e}}$$

Donc pour $n = 1$ on a $p_1 = 1/2$ et pour $n \geq 2$ on a :

$$p_n \leq \sqrt{\frac{1}{e 2^n}}$$

Il suit que la probabilité de **terminer** est :

$$\begin{aligned}
&\Sigma_{n=1,\ldots,\infty}\, p_n \\
&\leq \tfrac{1}{2} + \tfrac{1}{\sqrt{e}} \cdot (\Sigma_{i=2,\ldots,\infty}\, \tfrac{1}{2^i}) \\
&= \tfrac{1}{2} + \tfrac{1}{\sqrt{e}} \cdot \tfrac{1}{2} = \tfrac{\sqrt{e}+1}{2\sqrt{e}} \approx 0,8 < 1
\end{aligned}$$

Et donc la probabilité de **boucler est significative**!

## Exemple

```
public static void proba3(int m){
        long k=0;
        Random g = new Random();
        while (k<m) {if (g.nextInt(2)==1){k=k+1;}} }
```

## Analyse

Pour terminer on doit tirer $m$ fois 1. On reconnaît ici la distribution **binomiale négative**. Donc :

1. `proba3` **termine avec probabilité** 1.
2. Le **coût moyen** est : $2m$.

**NB** Le coût est **exponentiel** dans le nombre de bits nécessaires à représenter l'entrée `m`.

## Exemple

```
public static void proba4(){
        int n=1;
        Random g = new Random();
        while (!(g.nextInt(2)==1)){n=n+1;}
        boolean stop=false;
        while (!stop) {
            stop=true;
            for (int i=0;i<n;i++){if (g.nextInt(2)==1){stop=false;}}}}
```

## Analyse

1. D'abord on suit une **distribution géométrique** et on affecte à la variable $n$ un entier $i \geq 1$ avec probabilité $2^{-i}$.
2. La boucle `for` laisse `stop` à `true` avec probabilité $p_n = \frac{1}{2^n}$. La deuxième boucle `while` correspond donc aussi à une **distribution géométrique** avec paramètre $p_n$.
3. La **probabilité de terminaison est donc** 1 :

$$\begin{aligned}
&\Sigma_{n=1\ldots,\infty}\, 2^{-n}(\Sigma_{k=1,\ldots,\infty}\,(1-p_n)^{(k-1)}p_n)\\
&= \Sigma_{n=1\ldots,\infty}\, 2^{-n}(1)\\
&= 1
\end{aligned}$$

4. On considère maintenant le **temps moyen de calcul** en comptant les itérations des 2 boucles `while` et en faisant abstraction du fait que les entiers représentables par le type `int` de Java sont bornés par `2,147,483,647`.

5. Soient $C_1$ et $C_2$ les v.a.d. **coût** qui correspondent à la première et à la deuxième boucle `while`. On a :

$$\begin{aligned} P(C_1 = n) &= \tfrac{1}{2^n} \\ E[C_1] &= 2 \\ E[C_2 | C_1 = n] &= 2^n \end{aligned}$$

On calcule (voir propriétés v.a.d. conditionnelle) :

$$\begin{aligned} E[C_2] &= \Sigma_{n=1,\ldots,\infty} E[C_2|C_1 = n]P(C_1 = n) \\ &= \Sigma_{n=1,\ldots,\infty} 2^n 2^{-n} \\ &= \Sigma_{n=1,\ldots,\infty} 1 \\ &= \infty \end{aligned}$$

Encore un barycentre qui va à l'infini...

## Un 'vrai' exemple : tri rapide (quicksort)

Pour trier un **tableau** $t$ compris entre $i$ et $j$ $(i < j)$ :

1. On sélectionne de façon **uniforme** un élément $k$ entre $i$ et $j$ (extrêmes inclus). On **échange** $t[k]$ et $t[j]$. **Le pivot est choisi de façon aléatoire !**
2. On pose le **pivot** $p = t[j]$. On compare les éléments dans $[i, j-1]$ à $p$ et on place dans $[i, \ ]$ les éléments plus petits et dans $[\ +1, j-1]$ ceux plus grands (ces intervalles peuvent être vides).
3. Si $\ +1 < j$ on **échange** $t[\ +1]$ et $t[j]$. Si nécessaire, on trie récursivement les tableaux $[i, \ ]$, $[\ +2, j]$.

## Analyse

– On suppose tous les **éléments différents**, disons

$$\{1, \ldots, n\}$$

– On veut compter le **nombre de comparaisons** effectuées **en moyenne** par l'algorithme. Ce nombre dépend du **choix aléatoire des pivots**.

– On peut représenter une calcul par la suite $p_1 \cdots p_k$ des pivots choisis. On prend comme espace de probabilité $\Omega$ l'ensemble de ces suites et on définit une **v.a.d.**

$$X : \Omega \to \mathbf{R}$$

qui associe à chaque suite le nombre de comparaisons effectuées par le tri rapide. Le **but** est de calculer $E[X]$.

## Utilisation de la linéarité

Soient $i, j \quad \{1, \ldots, n\}$ avec $i < j$.

$i$ et $j$ sont comparés au plus un fois.

L'algorithme compare un pivot aux autres éléments d'une partition. Donc pour comparer $i$ et $j$ il faut que l'un des deux soit un pivot et l'autre se trouve dans la même partition. Par ailleurs, dans la suite du calcul le pivot ne sera plus comparé à un autre élément.

On **définit** :

$$X_{i,j}(\omega) = \begin{cases} 1 & \text{si } i \text{ et } j \text{ sont comparés dans } \omega \\ 0 & \text{autrement} \end{cases}$$

On **observe** :

$$X = \Sigma_{1 \leq i < j \leq n} X_{i,j}$$

Et par **linéarité** :

$$E[X] = \Sigma_{1 \leq i < j \leq n} E[X_{i,j}]$$

## Calcul de $E[X_{i,j}]$

On note $P(i,j,n) = E[X_{i,j}]$ la probabilité que $i$ et $j$ sont **comparés** dans un tri rapide avec $n$ éléments, où $1 \leq i < j \leq n$.

**Idée 1** On observe :

$$\begin{aligned} P(1,2,2) &= 1 \\ P(i,j,n) &= \tfrac{2}{n} + \tfrac{1}{n} \cdot ( \quad \Sigma_{k=1,\ldots,(i-1)}\, P(i-k, j-k, n-k) + \\ &\qquad\qquad \Sigma_{k=(j+1),\ldots,n} P(i,j,k-1) \; ) \end{aligned}$$

| pour comparer $i$ à $j$, soit on prend le pivot dans $\{i,j\}$ soit on le prend avant $i$ ou après $j$. |
|---|

**Idée 2** $P(i,j,n)$ ne dépend pas de $n$. En effet on montre **par récurrence** sur $n$ :

$$P(i,j,n) = \frac{2}{(j-i+1)}$$

En particulier : $P(i,i+1,n) = 1$.

## Calcul de $E[X]$

$$\begin{aligned} E[X] &= \Sigma_{i=1,\ldots,(n-1)} \Sigma_{j=i+1,\ldots,n} \tfrac{2}{(j-i+1)} \\ &= 2 \cdot (\Sigma_{i=1,\ldots,n-1} (\Sigma_{k=1,\ldots,(n-i)} \tfrac{1}{(k+1)})) \\ &\leq 2 \cdot (\Sigma_{i=1,\ldots,n-1} (\Sigma_{k=1,\ldots,n} \tfrac{1}{k})) \end{aligned}$$

On se **souvient** que :

$$\Sigma_{x=2,\ldots,m} \frac{1}{x} < \int_1^m \frac{1}{x}\, \mathrm{d}x = \ln(m)$$

Donc $\Sigma_{k=1,\ldots,n} \frac{1}{k} \leq 1 + \ln(n)$ et :

$$E[X] \leq 2 \cdot (n-1)(\ln(n)+1)$$

soit $\mathbf{E[X] = O(n \ln(n))}$.

## **Preuve** $P(i,j,n) = \frac{2}{j-i+1}$

Pour $n = 2$ on a bien $P(1,2,2) = \frac{2}{2-1+1} = 1$. Pour $n+1 > 2$ on a :

$$\begin{aligned}
P(i,j,n+1) &= \tfrac{2}{n+1} + \tfrac{1}{n+1}\big( \quad \Sigma_{k=1,\ldots,(i-1)}\, P(i-k, j-k, n+1-k) + \\
&\qquad\qquad \Sigma_{k=(j+1),\ldots,n+1} P(i,j,k-1) \big) \\
&= \tfrac{2}{n+1} + \tfrac{1}{n+1}\big( \quad \Sigma_{k=1,\ldots,(i-1)}\, \tfrac{2}{j-i+1} + \\
&\qquad\qquad \Sigma_{k=(j+1),\ldots,n+1}\, \tfrac{2}{j-i+1} \big) \\
&= \tfrac{2}{n+1} + \tfrac{1}{n+1} \quad \tfrac{2(n-j+i)}{j-i+1} \\
&= \tfrac{2}{j-i+1}
\end{aligned}$$

## Variante : analyse en moyenne

Plutôt que choisir les pivots de façon aléatoire on peut considérer la situation suivante :

– l'**entrée est permutée de façon aléatoire** ($n!$ permutations possibles).

– le pivot est, par exemple, l'élément le **plus à droite** de la partition.

Dans ce cas aussi on peut définir une variable aléatoire qui correspond au nombre de comparaisons et estimer son **espérance** à $O(n\ln(n))$.

## Variété d'algorithmes probabilistes

– Dans le cas du tri rapide le nombre maximum d'itérations/comparaisons est borné par une valeur qui ne dépend pas de la suite de bits aléatoires. L'**algorithme termine**.

– En général, on peut concevoir des algorithmes probabilistes où cette propriété n'est pas satisfaite. Dans ce cas on cherchera à montrer que l'**algorithme termine avec probabilité** 1.

– Certains algorithmes probabilistes (dits de **Montecarlo**) **s'ils terminent peuvent fournir des réponses incorrectes**.

– Par exemple le **test de primalité de Miller-Rabin** peut affirmer qu'un nombre est premier alors qu'il ne l'est pas.

– En gros, le test de Miller-Rabin raffine le **test de Fermat** qui affirme que si $n$ est premier alors pour tout $a \quad \{1, \ldots, n-1\}$ :

$$(a^{n-1} \equiv 1) \mod n$$

– L'enjeu est alors de **borner la probabilité** $p$ de la réponse incorrecte (pour Miller-Rabin on a $p \leq 1/4$).

– Ensuite en répétant l'algorithme $n$ fois avec des choix **indépendants** on fait tomber la probabilité à $p^n$.

## Exemple : identité de polynômes

– Soient $p$ et $q$ deux **polynômes** dans une variables de **degré** $d$ (par exemple sur $\mathbf{R}$).

– On cherche à déterminer s'ils sont **identiques**, à savoir si $p - q = 0$.

– Une façon de résoudre ce problème est d'écrire les polynômes $p$ et $q$ comme **somme de monômes** et de comparer les coefficients. Cette approche demande un nombre de multiplications qui est **quadratique en** $d$.

– Par contre l'évaluation d'un polynôme sur un point $x$ demande un nombre d'opérations qui est **linéaire** en $d$. On cherche un algorithme probabiliste qui peut vérifier l'identité en temps linéaire avec une **probabilité d'erreur négligeable**.

## Méthode

– Par le **théorème fondamentale de l'algèbre** le polynôme $p - q$ a au plus $d$ racines s'il n'est pas 0.

– Fixons un **ensemble fini** $T \quad \mathbf{R}$ avec $T = kd$, $k > 1$.

– Si on tire un point de $T$ de façon **uniforme** la probabilité de tomber sur une racine de $p - q$ est au plus $\frac{1}{k}$.

- On tire $n$ points $x_1, \ldots, x_n$ dans $T$ de façon **indépendante** :
  - Si on trouve **un** $x_i$ tel que tel que $(p - q)(x_i) = 0$ alors on est **sûr que** $p = q$.
  - Si pour **tous les** $x_i$ on a que $(p - q)(x_i$

## Une généralisation : lemme de Schwartz et Zippel

– Schwartz et Zippel ($\approx$ 1979) ont montré que le résultat se généralise à des **polynômes avec plusieurs variables**.

– Dans ce cas, la réduction à une somme de monômes peut prendre un **temps exponentiel**. Par exemple considérez :

$$\Pi_{i=1,\ldots,n}(X_i + X_{i+1})$$

– En effet **on ne connaît pas** à ce jour un algorithme **déterministe efficace** pour résoudre le problème de l'identité de polynômes à plusieurs variables.

– Les bibliothèques de calcul formel utilisent donc un **algorithme probabiliste de type Montecarlo**.

## Sommaire

- Probabilité de **terminaison**.
- **Coût moyen** comme espérance.
- **Coût moyen** d'un algorithme par rapport à :
  - des **choix probabilistes**
  - une **distribution des entrées**.

  Exemple remarquable : **tri rapide**.
- Algorithmes probabilistes avec possibilité d'**erreur**. Exemples remarquables : **test de primalité, identité de polynômes**.