Probabilités et statistiques : un bref aperçu

Alexis Devulder
Laboratoire de Mathématiques de Versailles - UMR 8100
Université de Versailles Saint-Quentin en Yvelines
45 avenue des Etats-Unis
78035 Versailles cedex
devulder@math.uvsq.fr

14 janvier 2009

Qu'est ce que le hasard ?

Question philosophique...

- **Hasard** : mot d'origine arabe (dé, chance) lié à l'incertitude, l'inexplicable
- **Aléa** : mot d'origine latine (=dé)

Qu'est ce que le hasard ?

Question philosophique...

- **Hasard** : mot d'origine arabe (dé, chance) lié à l'incertitude, l'inexplicable
- **Aléa** : mot d'origine latine (=dé)
- **Expérience aléatoire** : Expérience dont on ne peut pas prévoir le résultat à l'avance

  Souvent lié à notre *manque de connaissances* (cf livre d'Ekeland)

Un peu de vocabulaire

Notion d'évènement

Un ensemble de résultats parmi les résultats possibles. Ex : $\{2, 4, 6\}$ pour un lancer d'un dé.

Un peu de vocabulaire

Notion d'évènement

Un ensemble de résultats parmi les résultats possibles. Ex : $\{2, 4, 6\}$ pour un lancer d'un dé.

Notion de probabilité

On réalise la même expérience de nombreuses fois.
Si A est un évènement, le quotient

$$\boxed{\frac{\text{Nombre de réalisations de } A}{\text{nombre d'expériences}}}$$

tend expérimentalement vers une limite, appelée **probabilité de A**

Les paris du chevalier de Méré (1607-1684)

- **Pari 1** : On jette 4 fois un dé à 6 faces. Méré pense que le pari «on obtient au moins un 6» est plus souvent gagnant que perdant.
  Explication de Méré : 4 lancers, 6 possibilités par lancer, et $4/6 > 1/2$.

Les paris du chevalier de Méré (1607-1684)

- **Pari 1** : On jette 4 fois un dé à 6 faces. Méré pense que le pari «on obtient au moins un 6» est plus souvent gagnant que perdant.
  Explication de Méré : 4 lancers, 6 possibilités par lancer, et $4/6 > 1/2$.
- **Pari 2** : On jette 24 fois deux dés à 6 faces. Méré pense que le pari «on obtient au moins un double 6» est plus souvent gagnant que perdant.
  Explication de Méré : 24 lancers, 36 possibilités par lancer, et $24/36 = 4/6$.

Les paris du chevalier de Méré (1607-1684)

- **Pari 1** : On jette 4 fois un dé à 6 faces. Méré pense que le pari «on obtient au moins un 6» est plus souvent gagnant que perdant.
  Explication de Méré : 4 lancers, 6 possibilités par lancer, et $4/6 > 1/2$.
- **Pari 2** : On jette 24 fois deux dés à 6 faces. Méré pense que le pari «on obtient au moins un double 6» est plus souvent gagnant que perdant.
  Explication de Méré : 24 lancers, 36 possibilités par lancer, et $24/36 = 4/6$.
- **Problème** : Méré est perdant sur le pari 2. Pourquoi ?
  Question posée à Pascal et Fermat en 1654.

Réponses de Pascal (1623-1662)

- **Pari 1** : On jette un dé à 6 faces. «On obtient au moins un 6» est plus souvent gagnant que perdant.

$$\boxed{1 - (5/6)^4 \approx 0,518}$$

Réponses de Pascal (1623-1662)

- **Pari 1** : On jette un dé à 6 faces. «On obtient au moins un 6» est plus souvent gagnant que perdant.

$$\boxed{1 - (5/6)^4 \approx 0,518}$$

- **Pari 2** : On jette 24 fois deux dés à 6 faces. «On obtient au moins un double 6».

$$\boxed{1 - (35/36)^{24} \approx 0,492}$$

Définition de Laplace (1749-1827)

Dans une situation à l'issue incertaine, mais pour laquelle toutes les issues ont la même chance de se produire (*équiprobabilité*) :

Cas équiprobabile

La **probabilité** d'un évènement est alors :

$$\boxed{\frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}}$$

Le calcul de probabilités revient alors à compter le nombre de cas favorables : c'est de la **combinatoire**.

Exemple : jet d'un dé. Probabilité d'avoir un nombre pair : $\frac{3}{6} = \frac{1}{2}$

Le loto

Simplification : On ne s'intéresse qu'au gros lot : 6 bons numéros parmi 49.
Nombre de grilles de 6 numéros différents :

$$\boxed{\frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13983816}$$

- **Question 1** : Si le 7 n'est pas sorti depuis 100 tirages, ai-je plus de chances de gagner si je joue le 7 ?

Le loto

Simplification : On ne s'intéresse qu'au gros lot : 6 bons numéros parmi 49.
Nombre de grilles de 6 numéros différents :

$$\boxed{\frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13983816}$$

- **Question 1** : Si le 7 n'est pas sorti depuis 100 tirages, ai-je plus de chances de gagner si je joue le 7 ?

Le loto

Simplification : On ne s'intéresse qu'au gros lot : 6 bons numéros parmi 49.
Nombre de grilles de 6 numéros différents :

$$\boxed{\frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13983816}$$

- **Question 1** : Si le 7 n'est pas sorti depuis 100 tirages, ai-je plus de chances de gagner si je joue le 7 ?
- **Réponse** : **NON**

Le loto - Indépendance

Deux évènements sont **indépendants** quand l'un n'a rien à voir avec l'autre.

Indépendance

Mathématiquement, A et B sont indépendants si

$$\boxed{\mathbb{P}(A \text{ sachant } B) = \mathbb{P}(A)}$$

(si $\mathbb{P}(B) \neq 0$) soit dans tous les cas

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)}$$

Au loto : indépendance garantie par la machine qui mélange les boules.

Loto - astuce ?

Lu au hasard sur un site

« Jouez toujours les mêmes numéros. La probabilité que votre bulletin soit gagnant augmente chaque fois. Si vous changez un numéro, vous repartez avec la probabilité 1 contre 14 millions. »
FAUX

Loto - astuce ?

Lu au hasard sur un site

«Jouez toujours les mêmes numéros. La probabilité que votre bulletin soit gagnant augmente chaque fois. Si vous changez un numéro, vous repartez avec la probabilité 1 contre 14 millions.»
FAUX

Question 2 : Si je joue 1, 2, 3, 4, 5, 6, ai-je moins de chances de gagner qu'avec une autre combinaison ?

FAUX car toutes les grilles ont autant de chances de gagner.

Par contre, les grilles beaucoup jouées auront un gain inférieur si elles sont tirées (dates d'anniversaires, etc).

Exemple de l'alcootest

Un laboratoire a mis au point un **alcootest**.

- On sait que 2% des personnes controlées par la police sont réellement **en état d'ébriété**. Les premiers essais ont conduit aux résultats suivants :
- Lorsqu'une personne est réellement **en état d'ébriété**, 99 fois sur 100, l'alcootest est **positif**.
- Lorsqu'une personne **n'est pas en état d'ébriété**, 97 fois sur 100, l'alcootest est **négatif**.

Que penser de cet alcootest ?

Objet des statistiques

- **Statistiques descriptives** (= *analyse de données*)
  On cherche à analyser une série de données, aléatoires ou non.
  Ex : liste de notes à un examen

Objet des statistiques

- **Statistiques descriptives** (= *analyse de données*)
  On cherche à analyser une série de données, aléatoires ou non.
  Ex : liste de notes à un examen
- **Statistique mathématique** (= *statistiques inférentielles*)
  On ne connait pas la loi de probabilité $\mathbb{P}$ d'un phénomène aléatoire, on cherche à l'«estimer».
  Ex : Lancer d'une pièce ; $\mathbb{P}(\text{Pile}) = p$. On ne connait pas p. Que dire de p ?

Objet des statistiques

- **Statistiques descriptives** (= *analyse de données*)
  On cherche à analyser une série de données, aléatoires ou non.
  Ex : liste de notes à un examen
- **Statistique mathématique** (= *statistiques inférentielles*)
  On ne connait pas la loi de probabilité $\mathbb{P}$ d'un phénomène aléatoire, on cherche à l'«estimer».
  Ex : Lancer d'une pièce ; $\mathbb{P}(\text{Pile}) = p$. On ne connait pas p. Que dire de p ?
- **Probabilités** :
  On connait la loi de probabilités $\mathbb{P}$, on cherche à comprendre le comportement du système («prédictif»)
  Ex : Lancer d'une pièce ; $\mathbb{P}(\text{Pile}) = p$. On connait p.

Statistiques descriptives

Extraire des informations partielles pertinentes d'une liste de données (nombres...) difficiles à interpréter par simple lecture.

Statistiques descriptives

Extraire des informations partielles pertinentes d'une liste de données (nombres...) difficiles à interpréter par simple lecture.

- **Moyenne** de $X_1, \ldots, X_n$: $\dfrac{1}{n}\sum_{i=1}^{n} X_i$

Statistiques descriptives

Extraire des informations partielles pertinentes d'une liste de données (nombres...) difficiles à interpréter par simple lecture.

- **Moyenne** de $X_1, \dots, X_n$: $\frac{1}{n}\sum_{i=1}^{n} X_i$
- **Médiane** : valeur qui sépare les données en deux groupes de taille égale

Statistiques descriptives

Extraire des informations partielles pertinentes d'une liste de données (nombres...) difficiles à interpréter par simple lecture.

- **Moyenne** de $X_1, \ldots, X_n$: $\frac{1}{n}\sum_{i=1}^{n} X_i$
- **Médiane** : valeur qui sépare les données en deux groupes de taille égale
- **Dispersion** : étendue, variance, écart-type, etc.

Statistiques descriptives

Extraire des informations partielles pertinentes d'une liste de données (nombres...) difficiles à interpréter par simple lecture.

- **Moyenne** de $X_1, \ldots, X_n$: $\frac{1}{n}\sum_{i=1}^{n} X_i$
- **Médiane** : valeur qui sépare les données en deux groupes de taille égale
- **Dispersion** : étendue, variance, écart-type, etc.
- **Représentations graphiques** des données

Statistiques mathématiques

On observe un phénomène **aléatoire**. On ne **connait pas** la loi de probabilité $\mathbb{P}$. Ex : on lance une pièce biaisée. On peut écrire $\mathbb{P}(\text{pile}) = p$, mais on ne connait pas le paramètre p.

Statistiques mathématiques

On observe un phénomène **aléatoire**. On ne **connait pas** la loi de probabilité $\mathbb{P}$. Ex : on lance une pièce biaisée. On peut écrire $\mathbb{P}(\text{pile}) = p$, mais on ne connait pas le paramètre p.

- On peut **estimer** le paramètre. Ex : $\dfrac{\text{Nombre de succès}}{\text{Nombre de lancers}} \approx p$

Statistiques mathématiques

On observe un phénomène **aléatoire**. On ne **connait pas** la loi de probabilité $\mathbb{P}$. Ex : on lance une pièce biaisée. On peut écrire $\mathbb{P}(\text{pile}) = p$, mais on ne connait pas le paramètre p.

- On peut **estimer** le paramètre. Ex : $\dfrac{\text{Nombre de succès}}{\text{Nombre de lancers}} \approx p$
- On peut donner un **intervalle de confiance** pour les paramètres estimés. Ex : sondages.

Statistiques mathématiques

On observe un phénomène **aléatoire**. On ne **connait pas** la loi de probabilité $\mathbb{P}$. Ex : on lance une pièce biaisée. On peut écrire $\mathbb{P}(\text{pile}) = p$, mais on ne connait pas le paramètre p.

- On peut **estimer** le paramètre. Ex : $\dfrac{\text{Nombre de succès}}{\text{Nombre de lancers}} \approx p$
- On peut donner un **intervalle de confiance** pour les paramètres estimés. Ex : sondages.
- On peut **tester une hypothèse**
  Ex 1 : $p = \frac{1}{2}$?
  Ex 2 : un médicament testé cliniquement est-il efficace ?
  Ex 3 : valider ou non un modèle, par exemple les lois de Mendel.

Axiomatisation de Kolmogorov, 1933

Insuffisance des probabilités de Pascal

Axiomatisation de Kolmogorov, 1933

Insuffisance des probabilités de Pascal

Idée de Kolmogorov : utiliser la *théorie de la mesure* et l'*intégrale de Lebesgue*

On définit un **espace de probabilités** par un triplet $(\Omega, \mathcal{A}, \mathbb{P})$.

Axiomatisation de Kolmogorov, 1933

Univers Ω

$\Omega \neq \emptyset$ est l'ensemble des résultats possibles de l'expérience. Un résultat de l'expérience est un élément $\omega \in \Omega$.

Ex : $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Axiomatisation de Kolmogorov, évènements

Ex : $\mathcal{A} = \mathcal{P}(\Omega)$

$\mathcal{A}$ est l'ensemble des évènements

On demande à $\mathcal{A}$ de vérifier les propriétés suivantes (tribu) :

- Le **vide** est un évènement $\emptyset \in \mathcal{A}$
- $\mathcal{A}$ est stable par **complémentaire**

$$\forall A \in \mathcal{A}, \qquad A^c \in \mathcal{A}$$

- $\mathcal{A}$ est stable par **union dénombrable**

$$\forall (A_0, A_1, \dots, A_n \dots) \in (\mathcal{A})^{\mathbb{N}}, \qquad \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$$

Axiomatisation de Kolmogorov, probabilité

Probabilité

Une probabilité est une application de $\mathcal{A} \to [0,1]$, telle que

- $\mathbb{P}(\Omega) = 1$ et $\mathbb{P}(\emptyset) = 0$.
- Pour toute famille dénombrable (A_n) d'éléments de $\mathcal{A}$, disjoints deux à deux,

$$\mathbb{P}(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} \mathbb{P}(A_n)$$

Ex : $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) = \mathbb{P}(\{6\})$

Axiomatisation de Kolmogorov, exemples de Ω

- Ω **ensemble fini** : cartes, dés, etc

Axiomatisation de Kolmogorov, exemples de Ω

- Ω **ensemble fini** : cartes, dés, etc
- Ω **infini dénombrable** ; ex : loi géométrique
  $\mathbb{P}(\{k\}) = (1-p)p^{k-1}$, $k \in \mathbb{N}^*$

Axiomatisation de Kolmogorov, exemples de Ω

- Ω **ensemble fini** : cartes, dés, etc
- Ω **infini dénombrable** ; ex : loi géométrique $\mathbb{P}(\{k\}) = (1-p)p^{k-1}$, $k \in \mathbb{N}^*$
- $\Omega = \mathbb{R}$ ex : loi normale $\mathbb{P}([a,b]) = \displaystyle\int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathrm{d}x$

Axiomatisation de Kolmogorov, exemples de Ω

- Ω **ensemble fini** : cartes, dés, etc
- Ω **infini dénombrable** ; ex : loi géométrique $\mathbb{P}(\{k\}) = (1-p)p^{k-1}$, $k \in \mathbb{N}^*$
- $\Omega = \mathbb{R}$ ex : loi normale $\mathbb{P}([a,b]) = \displaystyle\int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathrm{d}x$
- Ω ensemble de **fonctions**

Axiomatisation de Kolmogorov, exemples de Ω

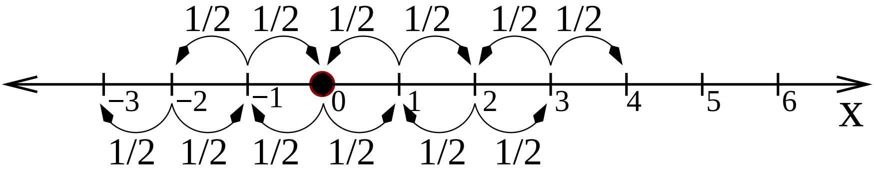
- Ω **ensemble fini** : cartes, dés, etc
- Ω **infini dénombrable** ; ex : loi géométrique $\mathbb{P}(\{k\}) = (1-p)p^{k-1}$, $k \in \mathbb{N}^*$
- $\Omega = \mathbb{R}$ ex : loi normale $\mathbb{P}([a,b]) = \displaystyle\int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x$
- Ω ensemble de **fonctions**
- etc

Observations de Brown

- Observation au microscope de mouvements de particules dans les grains de Pollen par Brown (1827)
- Une simulation du mouvement de particules de gaz
- Questions : comment faire pour comprendre un mouvement aussi désordonné ?
- Un peu de physique : température et agitation thermique
- Modèle introduit par Pearson (1905)

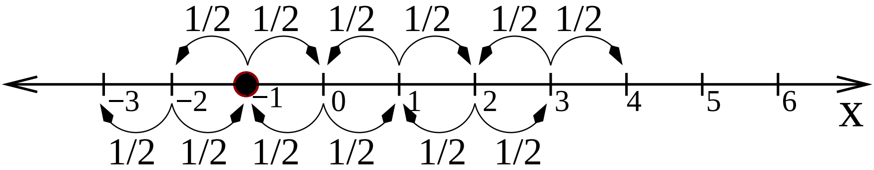
Marches aléatoires - Définition

Exemple pour $p = 1/2$, $n = 0$



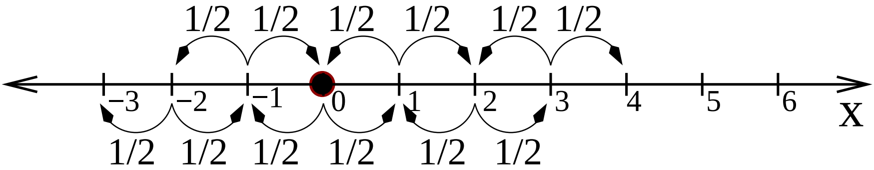
Marches aléatoires - Définition

Exemple pour $p = 1/2$, $n = 1$



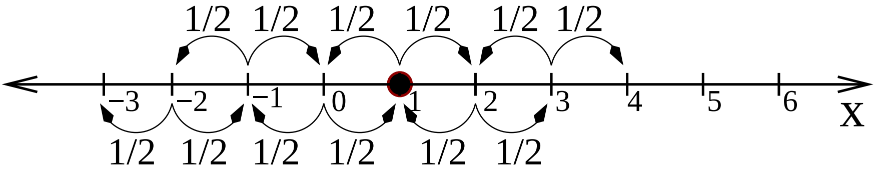
Marches aléatoires - Définition

Exemple pour $p = 1/2$, $n = 2$



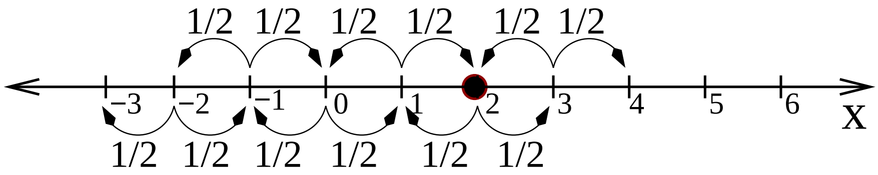
Marches aléatoires - Définition

Exemple pour $p = 1/2$, $n = 3$



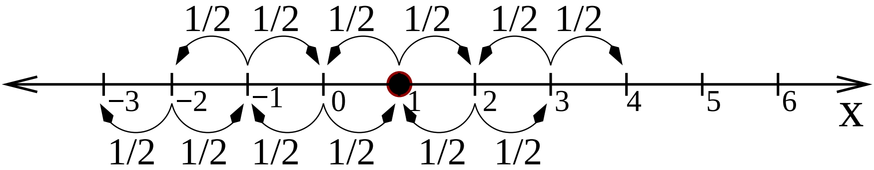
Marches aléatoires - Définition

Exemple pour $p = 1/2$, $n = 4$



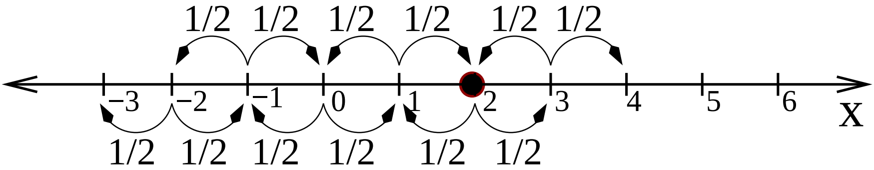
Marches aléatoires - Définition

Exemple pour $p = 1/2$, $n = 5$



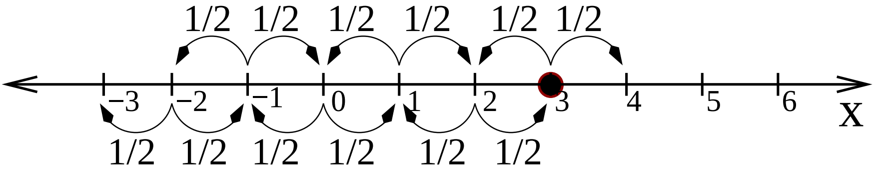
Marches aléatoires - Définition

Exemple pour $p = 1/2$, $n = 6$



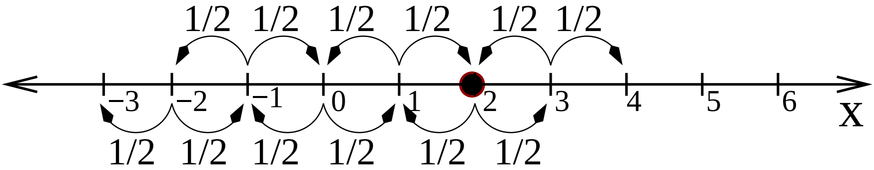
Marches aléatoires - Définition

Exemple pour $p = 1/2$, $n = 7$

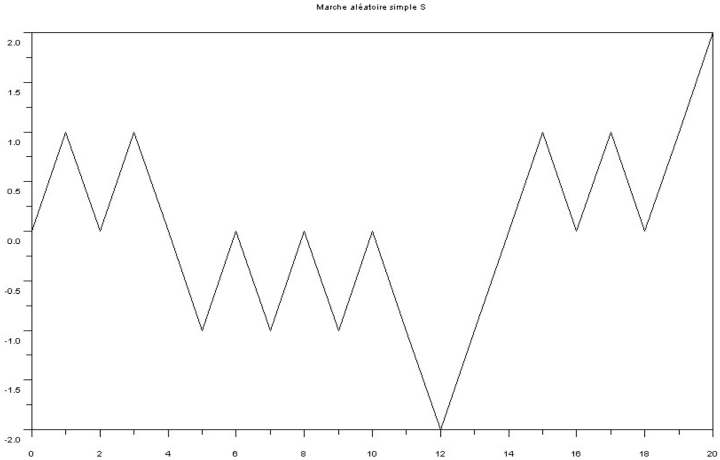


Marches aléatoires - Définition

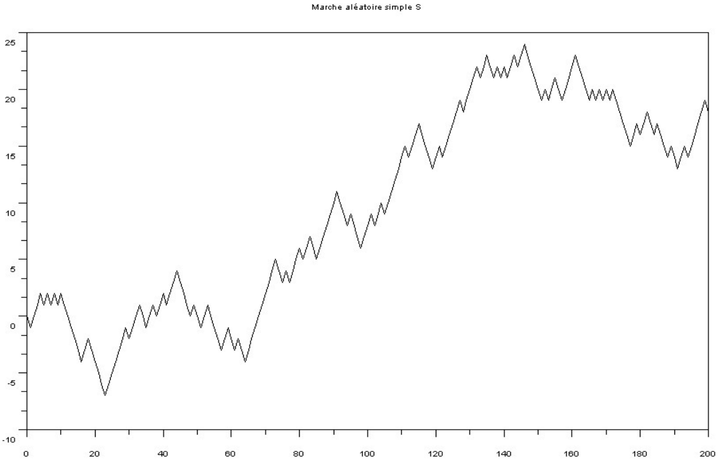
Exemple pour $p = 1/2$, $n = 8$



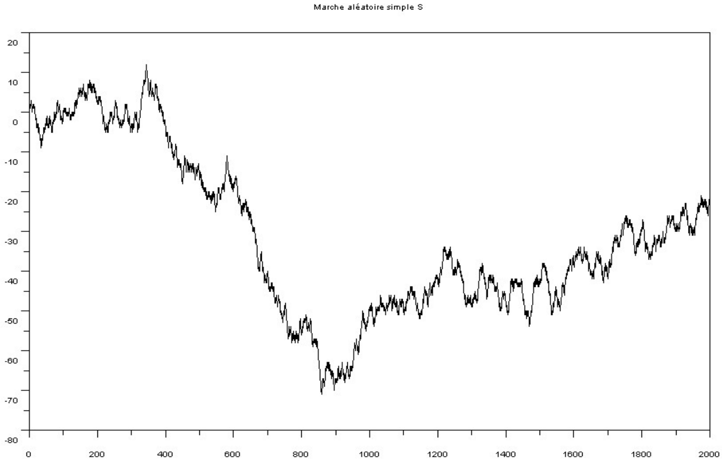
Marches aléatoires - Simulation - 20 pas



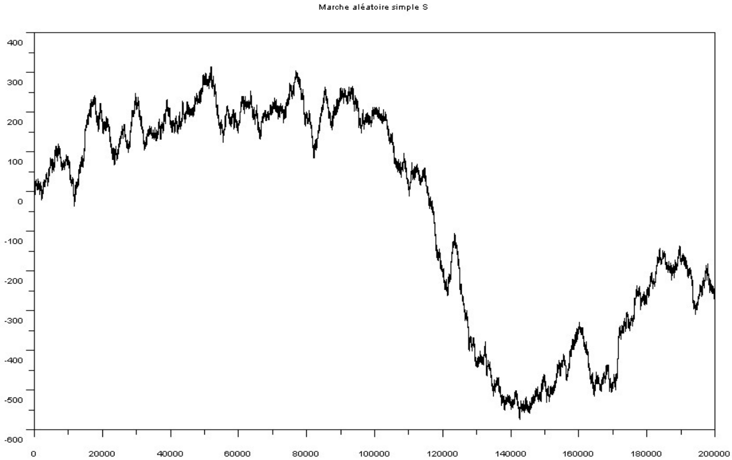
Marches aléatoires - Simulation - 200 pas



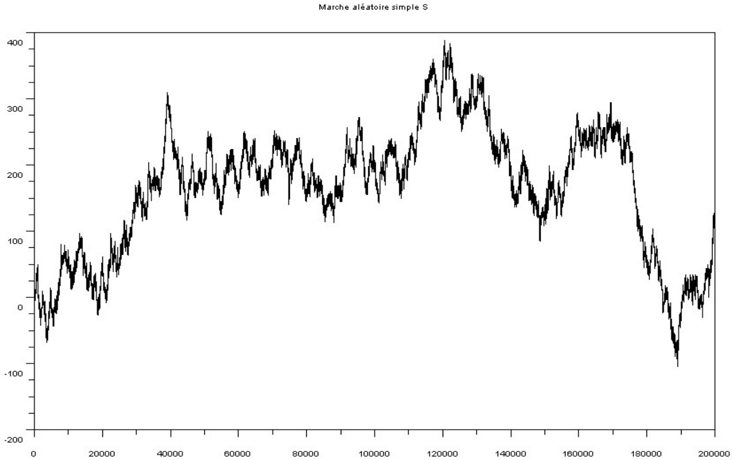
Marches aléatoires - Simulation - 20 000 pas



Marches aléatoires - Simulation - 200 000 pas



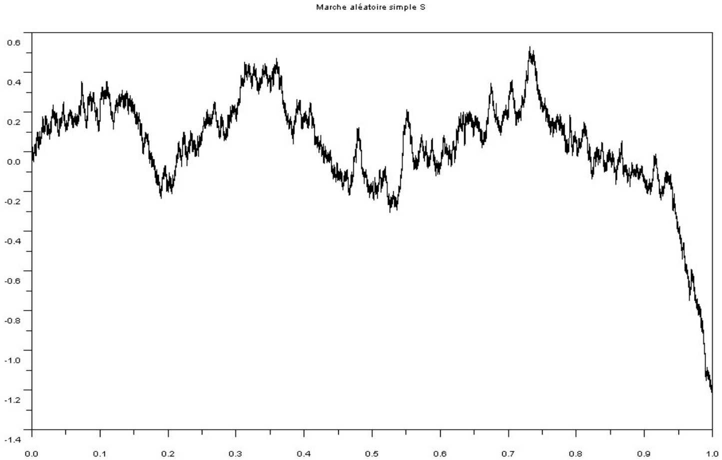
Marches aléatoires - Simulation - 200 000 pas



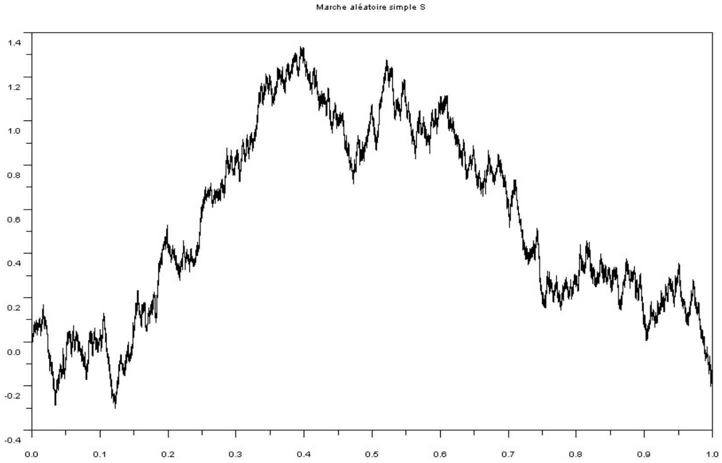
Mouvement Brownien

- Cf **Bachelier** 1900, **Einstein** 1905, **Perrin** 1906, **Wiener** 1923
- Le mouvement Brownien peut-être défini par la convergence «en loi» des marches aléatoires (avec changement d'échelle)
- Quelques simulations du Mouvement Brownien

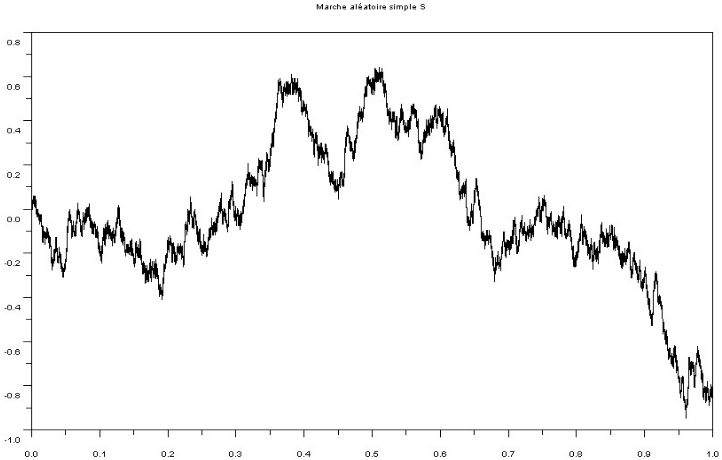
Mouvement Brownien - Simulation



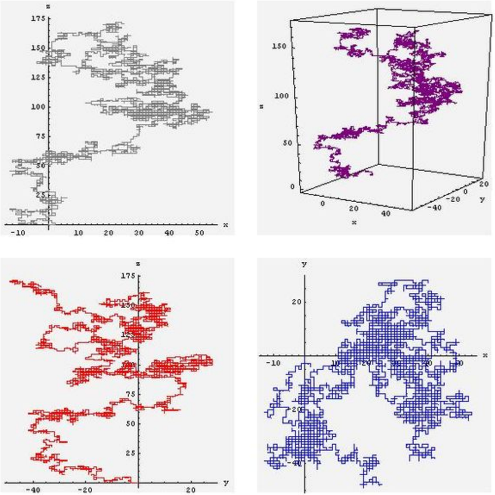
Mouvement Brownien - Simulation



Mouvement Brownien - Simulation



Mouvement Brownien en dimensions 2 et 3



Quelques applications des probabilités et statistiques

- Physique

Quelques applications des probabilités et statistiques

- Physique
- Fiabilité

Quelques applications des probabilités et statistiques

- Physique
- Fiabilité
- Assurances

Quelques applications des probabilités et statistiques

- Physique
- Fiabilité
- Assurances
- Finance

Quelques applications des probabilités et statistiques

- Physique
- Fiabilité
- Assurances
- Finance
- Télécommunications, Réseaux

Quelques applications des probabilités et statistiques

- Physique
- Fiabilité
- Assurances
- Finance
- Télécommunications, Réseaux
- Biologie (populations, génétique, chimiothérapie, etc)

Quelques applications des probabilités et statistiques

- Physique
- Fiabilité
- Assurances
- Finance
- Télécommunications, Réseaux
- Biologie (populations, génétique, chimiothérapie, etc)
- Reconnaissance de forme, de caractères, reconnaissance vocale

Quelques applications des probabilités et statistiques

- Physique
- Fiabilité
- Assurances
- Finance
- Télécommunications, Réseaux
- Biologie (populations, génétique, chimiothérapie, etc)
- Reconnaissance de forme, de caractères, reconnaissance vocale
- Moteurs de recherche

Quelques applications des probabilités et statistiques

- Physique
- Fiabilité
- Assurances
- Finance
- Télécommunications, Réseaux
- Biologie (populations, génétique, chimiothérapie, etc)
- Reconnaissance de forme, de caractères, reconnaissance vocale
- Moteurs de recherche
- Etc (neurosciences, climatologie, géophysique...)

Bibliographie

- **Livres scolaires**, collège, lycée
  Internet : wikipedia, `www.statistix.fr`

Bibliographie

- **Livres scolaires**, collège, lycée
  Internet : wikipedia, `www.statistix.fr`
- **Au hasard, la chance, la science et le monde**, I. Ekeland, Points Sciences
  Probabilités et statistiques aujourd'hui M. Quinio Benamo, L'Harmattan
  Cours de Probabilites, J. Neveu, cours de l'Ecole Polytechnique
  Mathématiques, Intégration et Probabilités, Licence 3e année, Auliac, Cocozza-Thivent, Mercier, Rossignol, Ediscience

Bibliographie

- **Livres scolaires**, collège, lycée
  Internet : wikipedia, `www.statistix.fr`
- **Au hasard, la chance, la science et le monde**, I. Ekeland, Points Sciences
  Probabilités et statistiques aujourd'hui M. Quinio Benamo, L'Harmattan
  Cours de Probabilites, J. Neveu, cours de l'Ecole Polytechnique
  Mathématiques, Intégration et Probabilités, Licence 3e année, Auliac, Cocozza-Thivent, Mercier, Rossignol, Ediscience
- **Statistiques, méfiez-vous !** Gauvrit, Ellipses
  Introduction à la méthode statistique (5e édition), B. Goldfarb et C. Pardoux
  Pratiques de la Statistique, C. Schwarz, Vuibert