

- 0/ Probabilités, statistiques comme outils informatiques.
- 1/ Rappels de probabilités ( **ICI**)
- 2/ Génération aléatoire (principes et méthodes).
- 3/ Rappels et éléments de statistiques.
- 4/ Evaluation de performances.
- 5/ Modèles mathématiques et analyse.
- 6/ Simulation.
- 7/ Méthodologie de l'évaluation de performances

# Rappels de probabilités

# Rappels de probabilités

- ❶ Evènements et expériences aléatoires.
- ❷ Axiomes des probabilités.
- ❸ Probabilités conditionnelles.
- ❹ Inégalités de Markov, Tchebychev & Chernoff.

On va appeler *expérience aléatoire* une expérience dont on ne peut ou ne veut pas prévoir complètement le résultat. Autrement dit même si les conditions de répétitions de l'expérience sont identiques, les résultats d'une telle expérience peuvent être différents. L'ensemble des résultats possibles d'une expérience aléatoire est simplement un ensemble de code noté  $\Omega$  comme dans les exemples qui suivent :

On va appeler *expérience aléatoire* une expérience dont on ne peut ou ne veut pas prévoir complètement le résultat. Autrement dit même si les conditions de répétitions de l'expérience sont identiques, les résultats d'une telle expérience peuvent être différents. L'ensemble des résultats possibles d'une expérience aléatoire est simplement un ensemble de code noté  $\Omega$  comme dans les exemples qui suivent :

Expérience	$\Omega$
Lancer une pièce	{Pile, Face}
Etat d'un bit de mémoire	{0, 1}
Electeur d'un référendum	{Oui, Non}
Lancer un dé	{1, 2, 3, ..., 6}
Nombre de clients dans une file	$\mathbb{N}$

En informatique, tous les langages (et les systèmes d'exploitation) ont leur fonction **RANDOM** qui est en fait un **générateur pseudo-aléatoire** (voir plus loin ...).

Une *loi de probabilité* ou *distribution de probabilité* est une fonction  $P$  qui à un évènement  $A$  associe un nombre  $P[A]$ , sa probabilité. Ce nombre traduit les “chances” que l’évènement  $A$  se produise.

Une *loi de probabilité* ou *distribution de probabilité* est une fonction  $P$  qui à un évènement  $A$  associe un nombre  $P[A]$ , sa probabilité. Ce nombre traduit les “chances” que l’évènement  $A$  se produise.

Intuitivement, on peut associer à toute répétition d’une expérience aléatoire sa fréquence expérimentale. Si  $n$  est le nbr. de répétitions,  $n_A$  est le nbr. de fois où  $A$  s’est produit, la *fréquence expérimentale* de  $A$  est donnée par  $\frac{n_A}{n}$ . Cette fréquence expérimentale peut changer d’une suite d’expériences à d’autres.

Une *loi de probabilité* ou *distribution de probabilité* est une fonction  $P$  qui à un évènement  $A$  associe un nombre  $P[A]$ , sa probabilité. Ce nombre traduit les “chances” que l’évènement  $A$  se produise.

Intuitivement, on peut associer à toute répétition d’une expérience aléatoire sa fréquence expérimentale. Si  $n$  est le nbr. de répétitions,  $n_A$  est le nbr. de fois où  $A$  s’est produit, la *fréquence expérimentale* de  $A$  est donnée par  $\frac{n_A}{n}$ . Cette fréquence expérimentale peut changer d’une suite d’expériences à d’autres. Les propriétés d’une probabilité sont celles de fréquences expérimentales et considérées comme des *axiomes de définition*.



Une *loi de probabilité* ou *distribution de probabilité* est une fonction  $P$  qui à un évènement  $A$  associe un nombre  $P[A]$ , sa probabilité. Ce nombre traduit les “chances” que l’évènement  $A$  se produise.

Intuitivement, on peut associer à toute répétition d’une expérience aléatoire sa fréquence expérimentale. Si  $n$  est le nbr. de répétitions,  $n_A$  est le nbr. de fois où  $A$  s’est produit, la *fréquence expérimentale* de  $A$  est donnée par  $\frac{n_A}{n}$ . Cette fréquence expérimentale peut changer d’une suite d’expériences à d’autres. Les propriétés d’une probabilité sont celles de fréquences expérimentales et considérées comme des *axiomes de définition*.

**Axiome 1:** Pour tout évènement  $A$ ,  $0 \leq P[A] \leq 1$ .

**Axiome 2:** La proba. d’un évènement certain est 1.  $P[\Omega] = 1$ .

**Axiome 3:** Si  $(A_i)_{i \in \mathbb{N}}$  est une suite d’évènements disjoints ( $A_i$  et  $A_j$  ne peuvent pas se produire en même temps si  $i \neq j$ ) alors

$$P\left[\bigcup_{i \in \mathbb{N}} A_i\right] = \sum_{i \in \mathbb{N}} P[A_i].$$

L'ensemble des éventualités  $\Omega$  est **fini** ou **dénombrable** :

$$\Omega = \{\omega_i, i \in I \subset \mathbb{N}\}.$$

Toutes les parties de  $\Omega$  sont des évènements. Comme tout évènement est une réunion finie ou dénombrable de singletons, il suffit de définir la probabilité de chaque singleton : pour tout  $i$ ,

$$P[\{\omega_i\}] = p_i.$$

Pour tout  $A \subset \Omega$ , la probabilité de  $A$  sera alors déterminée par l'axiome 3:

$$P[A] = \sum_{\omega_i \in A} P[\{\omega_i\}] = \sum_{\omega_i \in A} p_i.$$

**Exemple combinatoire.** Souvent les calculs peuvent se ramener à des problèmes de **dénombrements** :

$$\text{proba.} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

## Exemple issu de l'informatique

Le problème de l'élection dans un système distribué de processeurs consiste à élire (ou à distinguer) un processeur des autres.

Le problème de l'élection dans un système distribué de processeurs consiste à élire (ou à distinguer) un processeur des autres.

- **Entrée** :  $n$  processeurs identiques (non nécessairement identifiés, i.e. sans adresses IP connues)
- **Sortie** : un processeur ELU (et le sait),  $n - 1$  processeurs PERDANT (et le savent).

Le problème de l'élection dans un système distribué de processeurs consiste à élire (ou à distinguer) un processeur des autres.

- **Entrée** :  $n$  processeurs identiques (non nécessairement identifiés, i.e. sans adresses IP connues)
- **Sortie** : un processeur ELU (et le sait),  $n - 1$  processeurs PERDANT (et le savent).
- **Modèle de communication**. On suppose que les processeurs utilisent un **canal de communication partagé** qui peut transmettre le message de chacun d'entre eux.

## Exemple issu de l'informatique

Le problème de l'élection dans un système distribué de processeurs consiste à élire (ou à distinguer) un processeur des autres.

- **Entrée** :  $n$  processeurs identiques (non nécessairement identifiés, i.e. sans adresses IP connues)
- **Sortie** : un processeur ELU (et le sait),  $n - 1$  processeurs PERDANT (et le savent).
- **Modèle de communication**. On suppose que les processeurs utilisent un **canal de communication partagé** qui peut transmettre le message de chacun d'entre eux.
- Si deux processeurs tentent d'envoyer un message en même temps sur le canal, ce message est **perdu!**

Le problème de l'élection dans un système distribué de processeurs consiste à élire (ou à distinguer) un processeur des autres.

- **Entrée** :  $n$  processeurs identiques (non nécessairement identifiés, i.e. sans adresses IP connues)
- **Sortie** : un processeur ELU (et le sait),  $n - 1$  processeurs PERDANT (et le savent).
- **Modèle de communication**. On suppose que les processeurs utilisent un **canal de communication partagé** qui peut transmettre le message de chacun d'entre eux.
- Si deux processeurs tentent d'envoyer un message en même temps sur le canal, ce message est **perdu!**
- On suppose de plus que le temps est **discret** et qu'à chaque instant  $t \in \mathbb{N}$  chaque processeur peut envoyer un message à destination des autres.

## Exemple issu de l'informatique

Le problème de l'élection dans un système distribué de processeurs consiste à élire (ou à distinguer) un processeur des autres.

- **Entrée** :  $n$  processeurs identiques (non nécessairement identifiés, i.e. sans adresses IP connues)
- **Sortie** : un processeur ELU (et le sait),  $n - 1$  processeurs PERDANT (et le savent).
- **Modèle de communication**. On suppose que les processeurs utilisent un **canal de communication partagé** qui peut transmettre le message de chacun d'entre eux.
- Si deux processeurs tentent d'envoyer un message en même temps sur le canal, ce message est **perdu!**
- On suppose de plus que le temps est **discret** et qu'à chaque instant  $t \in \mathbb{N}$  chaque processeur peut envoyer un message à destination des autres.
- Un message correctement envoyé est reçu de **manière instantanée** par tous les autres processeurs.



- **Si**  $n$  est connu, avec une proba.  $\frac{1}{n}$  chaque participant tente d'annoncer qu'il est le leader sur le canal.
- **Sinon** nous pouvons supposer que  $n \in [2^k, 2^{k+1}]$  et on incrémente  $k$ .
- On a alors un algorithme avec un temps d'exécution logarithmique :

## SimpleLeaderElection

```
k := 0;
while election = FALSE do
  with proba.  $\frac{1}{2^k}$  each station  $u$  broadcasts on the channel ;
  if  $u$  hears a message then election := TRUE ;
  else k := k + 1;
endwhile
```

**Si  $n$  est connu**, chaque participant tente d'annoncer qu'il est le leader aux autres. Cette tentative peut se faire par un tirage aléatoire avec une proba. de  $\frac{1}{n}$  par processeur. La probabilité d'élection lors d'une tentative est donnée par

$$\text{Proba.} = \binom{n}{1} \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1}.$$

Pour  $n$  grand, cette proba. est très proche de  $\frac{1}{e} = 0.3678794412 \dots$ .  
On a en effet numériquement

$n$	100	250	500	1000	5000
Proba	.369 729 637	.368 616 921	.368 247 750	.368 063 488	.367 916 233

Par la suite, l'algorithme a donc une probabilité **strictement positive** de réussir (à élire un leader). Au bout d'un temps d'espérance finie, il converge vers cette élection.

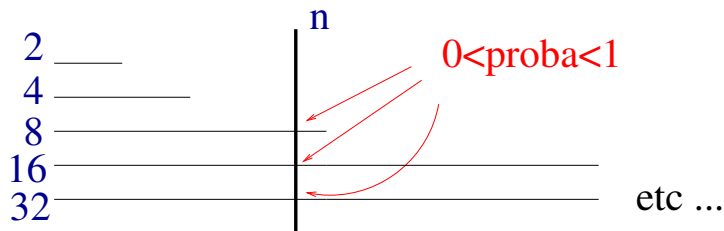
**Si  $n$  n'est pas connu**, cette valeur de  $n$  est comprise entre  $2^k$  et  $2^{k+1}$ .

**Si  $n$  n'est pas connu**, cette valeur de  $n$  est comprise entre  $2^k$  et  $2^{k+1}$ .  
Pour chaque valeur de  $k$ , chaque participant va annoncer sa candidature (il y a élection si et seulement si il y a une unique candidature) avec proba.  $\frac{1}{2^k}$  puis on va augmenter  $k$  au fur et à mesure.

**Si  $n$  n'est pas connu**, cette valeur de  $n$  est comprise entre  $2^k$  et  $2^{k-1}$ . Pour chaque valeur de  $k$ , chaque participant va annoncer sa candidature (il y a élection si et seulement si il y a une unique candidature) avec proba.  $\frac{1}{2^k}$  puis on va augmenter  $k$  au fur et à mesure. Cependant, il faut “revenir sur ses pas” après chaque itération. En effet, au mieux avec une probabilité de  $\sim 0.37$  on réussit l'élection. Donc, il faut répéter ces tentatives pour faire converger l'algorithme. La figure ci-dessous schématise cette approche :

## Analyse du second protocole

**Si  $n$  n'est pas connu**, cette valeur de  $n$  est comprise entre  $2^k$  et  $2^{k+1}$ . Pour chaque valeur de  $k$ , chaque participant va annoncer sa candidature (il y a élection si et seulement si il y a une unique candidature) avec proba.  $\frac{1}{2^k}$  puis on va augmenter  $k$  au fur et à mesure. Cependant, il faut “revenir sur ses pas” après chaque itération. En effet, au mieux avec une probabilité de  $\sim 0.37$  on réussit l'élection. Donc, il faut répéter ces tentatives pour faire converger l'algorithme. La figure ci-dessous schématise cette approche :



L'ensemble des éventualités  $\Omega$  est  $\mathbb{R}$ .

Les évènements sont les intervalles, et tous les sous-ensembles de  $\mathbb{R}$  que l'on peut former en combinant des intervalles par intersections et réunions. En théorie de la mesure, on les appelle des *boréliens*. (cf. cours de proba.)

## Les lois continues

L'ensemble des éventualités  $\Omega$  est  $\mathbb{R}$ .

Les évènements sont les intervalles, et tous les sous-ensembles de  $\mathbb{R}$  que l'on peut former en combinant des intervalles par intersections et réunions. En théorie de la mesure, on les appelle des *boréliens*. (cf. cours de proba.)

On appelle *densité de probabilité* une fonction de  $\mathbb{R}$  dans  $\mathbb{R}$  , continue par morceaux et d'intégrale 1.

$$f(x) \in \mathbb{R}$$



L'ensemble des éventualités  $\Omega$  est  $\mathbb{R}$ .

Les évènements sont les intervalles, et tous les sous-ensembles de  $\mathbb{R}$  que l'on peut former en combinant des intervalles par intersections et réunions. En théorie de la mesure, on les appelle des *boréliens*. (cf. cours de proba.)

On appelle *densité de probabilité* une fonction de  $\mathbb{R}$  dans  $\mathbb{R}$ , continue par morceaux et d'intégrale 1.

$$f(x) \geq 0, \forall x \in \mathbb{R} \text{ et } \int_{\mathbb{R}} f(x) dx = 1.$$

Etant donnée une densité de probabilité, on définit une loi de probabilité sur  $\mathbb{R}$  en associant à tout évènement l'intégrale de la densité sur cet évènement.

$$P[A] = \int_{x \in A} f(x) dx.$$

La connaissance d'une information sur une expérience peut modifier l'idée qu'on se fait de la probabilité d'un évènement. Par exemple, la probabilité qu'une personne soit âgée de moins de 61 ans est grande si la personne est choisie à l'Université

La connaissance d'une information sur une expérience peut modifier l'idée qu'on se fait de la probabilité d'un évènement. Par exemple, la probabilité qu'une personne soit âgée de moins de 61 ans est grande si la personne est choisie à l'Université

Soient  $A$  et  $B$  deux évènements tels que  $P[B] \neq 0$ . La probabilité conditionnelle de  $A$  sachant  $B$  est :

$$P[A | B] = \frac{P[A \cap B]}{P[B]} .$$

La connaissance d'une information sur une expérience peut modifier l'idée qu'on se fait de la probabilité d'un évènement. Par exemple, la probabilité qu'une personne soit âgée de moins de 61 ans est grande si la personne est choisie à l'Université

Soient  $A$  et  $B$  deux évènements tels que  $P[B] \neq 0$ . La probabilité conditionnelle de  $A$  sachant  $B$  est :

$$P[A | B] = \frac{P[A \cap B]}{P[B]} .$$

**Interprétation :** Le fait de savoir que  $B$  est réalisé réduit l'ensemble des résultats possibles de  $\Omega$  à  $B$ . A partir de là, seules les éventualités de  $A \cap B$  ont une importance. La probabilité de  $A$  sachant  $B$  doit donc être proportionnelle à  $P[A \cap B]$ . Le coefficient de proportionnalité  $1/P[B]$  assure que l'application qui à  $A$  associe  $P[A|B]$  est bien une probabilité, pour laquelle  $B$  est l'évènement certain.

Deux évènements  $A$  et  $B$  sont indépendants ssi :

$$P[A \cap B] = P[A] P[B] .$$

Deux évènements  $A$  et  $B$  sont indépendants ssi :

$$P[A \cap B] = P[A] P[B] .$$

Deux expériences aléatoires sont indépendantes si tout évènement observable à l'issue de l'une est indépendant de tout évènement observable à l'issue de l'autre.

Soit  $X$  une variable aléatoire réelle définie positive ou nulle alors

$$\forall t > 0, \quad P[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

( $\mathbb{E}[Z]$  est l'espérance de  $Z$ .)

Soit  $X$  une variable aléatoire réelle définie positive ou nulle alors

$$\forall t > 0, \quad P[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

( $\mathbb{E}[Z]$  est l'espérance de  $Z$ .)

**Preuve.** Soit  $f$  la fonction de densité de  $X$ , i.e.,

$$\mathbb{P}[X \in [a, b]] = \int_a^b f(t) dt.$$

On écrit

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty xf(x)dx \geq \int_t^\infty xf(x)dx \geq t \times \int_t^\infty f(x)dx \\ &= t \times \mathbb{P}[X \geq t]. \end{aligned}$$



Soit  $X$  une v.a. d'esperance finie  $\mu = \mathbb{E}[(X)]$  et de variance  $\sigma^2$ . Pour tout  $k > 0$  on a

$$\mathbb{P}[|X - \mathbb{E}[(X)]| \geq k\sigma] \leq \frac{1}{k^2}.$$

Soit  $X$  une v.a. d'esperance finie  $\mu = \mathbb{E}[(X)]$  et de variance  $\sigma^2$ . Pour tout  $k > 0$  on a

$$\mathbb{P}[|X - \mathbb{E}[(X)]| \geq k\sigma] \leq \frac{1}{k^2}.$$

**Preuve.** La variance de  $X$  est donnée par

$$\begin{aligned}\sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\&= \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu - k\sigma}^{\mu + k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu + k\sigma}^{\infty} (x - \mu)^2 f(x) dx \\&\geq k^2 \sigma^2 \int_{-\infty}^{\mu - k\sigma} f(x) dx + k^2 \sigma^2 \int_{\mu + k\sigma}^{\infty} f(x) dx \\&= k^2 \sigma^2 (\mathbb{P}[X \leq \mu - k\sigma] + \mathbb{P}[X \geq \mu + k\sigma]) = k^2 \sigma^2 \mathbb{P}[|X - \mu| \geq k\sigma].\end{aligned}$$

(Nous avons utilisé pour la première intégrale  $(x - \mu)^2 \geq (\mu - k\sigma)^2 = k^2 \sigma^2$ .)

Il existe plusieurs versions. La version donnée ici est une version qui peut se généraliser. Soit  $(X_i)_{i \in 1, n}$   $n$  variables aléatoires indépendantes telles que  $\mathbb{P}[X_i = 1] = \frac{1}{2}$  et  $\mathbb{P}[X_i = -1] = \frac{1}{2}$ . Soit  $Y = \sum_{i=1}^n X_i$ . Alors  $\forall \Delta > 0$ :

$$\mathbb{P}[Y \geq \Delta] \leq \exp\left(-\frac{\Delta^2}{2n}\right).$$

**Idées de la preuve.** 1) Montrer que  $\mathbb{P}[Y \geq \Delta] = \mathbb{P}[e^{tY} \geq e^t]$  et utiliser l'inégalité de Markov pour m.q. pour tout  $i$

$$\mathbb{E}[\exp(tX_i)] = \frac{e^t}{2} + \frac{e^{-t}}{2}.$$

2) Montrer alors que

$$\mathbb{E}[\exp(tX_i)] \leq e^{\frac{t^2}{2}}.$$

3) Utiliser les indépendances des  $X_i$  pour prouver que

$$\mathbb{E}[\exp(tY)] = \prod_{i=1}^n \mathbb{E}[\exp(tX_i)].$$

## Inégalité de Chernoff (suite preuve)

Et donc,

$$\mathbb{E}[\exp(tY)] \leq e^{nt^2/2}.$$

4) En déduire

$$\mathbb{P}[Y \geq \Delta] \leq \exp\left(nt^2/2 - t\Delta\right).$$

5) **Application** : on tire pile ou face  $n$  fois. Soit  $Z$  la variable aléatoire du nombre de fois où on a tiré pile. Montrer alors pour  $\forall \Delta > 0$ :

$$\mathbb{P}\left[\left|Z - \frac{n}{2}\right| \geq \Delta\right]$$