

# Préface

Ce livre est issu du polycopié du cours de probabilités et statistique de première année de l'école des Ponts ParisTech dont je suis responsable depuis 2002. Du fait de la nature inconnue ou chaotique de leur évolution, de nombreux phénomènes (météorologie, cours de bourse, volume de vente d'une pièce détachée automobile...) font naturellement l'objet d'une modélisation aléatoire, ce qui explique la place de plus en plus grande accordée aux probabilités dans les formations d'ingénieurs et dans les cursus universitaires. L'objectif du cours est de permettre aux étudiants de comprendre comment construire de tels modèles aléatoires et comment identifier leurs paramètres à partir de données.

À la différence de nombreux cours de probabilités de niveau licence, il ne fait pas appel à la théorie de la mesure. En effet, dans la pédagogie développée à l'école des Ponts, les notions de tribu et de mesurabilité ne sont étudiées qu'en deuxième année pour pouvoir introduire les martingales. En conséquence, le prérequis pour la lecture de ce livre est léger : maîtrise des notions de série et d'intégrale et du calcul matriciel. Et l'accent est mis sur les notions centrales en probabilités et statistique que sont la loi, l'indépendance, l'espérance, la variance, la fonction caractéristique, les convergences, l'estimateur du maximum de vraisemblance, les intervalles de confiance et les tests d'hypothèses plutôt que sur les fondements théoriques de ces disciplines.

Des exercices sont insérés au cœur des chapitres pour permettre aux étudiants de mettre en application les différents concepts au fur et à mesure de leur introduction. Mais des exercices et problèmes en nombre plus important sont également réunis à la fin de chaque chapitre. Certains font l'objet d'une correction dans le chapitre 10. Enfin, après chaque chapitre, un résumé d'une page environ reprend les notions importantes qui viennent d'être développées.

Après un chapitre introductif sur les espaces de probabilité finis où les calculs se ramènent à du dénombrement, l'espérance et ses propriétés, dont la linéarité, sont présentées en détail au chapitre 2 dans le cadre des variables aléatoires discrètes. La généralisation de la linéarité de l'espérance au cas des vecteurs aléatoires à densité est énoncée sans preuve dans le chapitre 3. Ces deux chapitres précisent également la notion de loi d'une variable aléatoires et fournissent les outils nécessaires (loi marginale, formule de changement de variable pour les intégrales multidimensionnelles) pour déterminer la loi d'une variable aléatoire d'intérêt dans un modèle probabiliste spécifique.

Le chapitre 4 est consacré aux techniques permettant de simuler sur ordinateur les variables aléatoires discrètes et à densité introduites auparavant. La simulation sur ordinateur permet de mieux appréhender les deux grands théorèmes limites de la théorie des probabilités qui forment le cœur du chapitre 5 : la loi forte des grands nombres et

le théorème de la limite centrale. Les différentes notions de convergence de variables aléatoires et leurs liens font également l'objet d'un traitement détaillé dans ce chapitre. L'objectif est que les étudiants acquièrent suffisamment de maîtrise sur les théorèmes limites pour bien comprendre ensuite les propriétés asymptotiques des estimateurs, intervalles de confiance et tests d'hypothèses dans la partie statistique du cours. Le théorème de la limite centrale explique le rôle fondamental en théorie des probabilités de la loi gaussienne et plus généralement des vecteurs gaussiens, auxquels le chapitre 6 est consacré.

En raison de son importance, le modèle gaussien sert d'exemple clé dans toute la partie statistique du livre. Le chapitre 7 introduit l'estimation de paramètres dans le modèle statistique paramétrique. L'accent est mis sur l'estimateur du maximum de vraisemblance et ses propriétés et sur la construction d'intervalles de confiance permettant de mesurer la précision de l'estimation. La notion de test d'hypothèses est présentée sur l'exemple du modèle gaussien dans le chapitre 8 qui explique également comment vérifier si des données sont issues d'une loi de probabilité fixée. Le livre s'achève sur un chapitre consacré à la régression linéaire qui fournit un cadre pour l'étude de l'influence de certains facteurs explicatifs sur des grandeurs mesurées ou des données expérimentales.

## Remerciements

Je tiens à remercier les membres de l'équipe enseignante du cours de probabilités et statistique de l'école des Ponts, Aurélien Alfonsi, Mohamed Ben Alaya, Anne Dutfoy, Michel de Lara, Julien Guyon, Tony Lelièvre, Jean-Michel Marin, Mohamed Sbair et Alain Toubol qui ont apporté de nombreuses améliorations à ce livre par leurs remarques et qui ont contribué à la compilation de corrigés d'exercices du chapitre 10. Je suis également très reconnaissant à Jean-François Delmas pour les emprunts qu'il m'a permis de faire au polycopié [6] de son cours de première année à l'ENSTA et au recueil d'exercices de son cours de statistique de seconde année à l'école des Ponts. Je dois beaucoup à Jean-Philippe Chancelier pour son aide précieuse concernant l'utilisation des logiciels Latex et Scilab. Ma gratitude va encore à tous les membres de l'équipe de probabilités appliquées du CERMICS et en particulier à Bernard Lapeyre pour nos discussions sur la pédagogie, qui, je l'espère, ont trouvé leur prolongement dans ce livre. Je tiens à remercier tous mes collègues du CERMICS pour l'ambiance de travail conviviale et stimulante qui règne au sein de ce laboratoire. Mes pensées vont enfin à Anne, Erwan et Alexy pour le bonheur que je partage avec eux au quotidien.

Benjamin Jourdain

# Chapitre 1

## Introduction : probabilité sur un espace fini

Historiquement, le calcul des probabilités s'est développé à partir du XVII<sup>e</sup> siècle autour des problèmes de jeux dans des situations où le nombre de cas possibles est fini. Les développements plus récents concernant des espaces non nécessairement finis nécessitent les outils techniques de la théorie de la mesure. Mais on peut introduire simplement sur les espaces finis toutes les notions importantes de probabilités sans avoir besoin de cet outillage.

### 1.1 Probabilité sur un espace fini, événements

#### 1.1.1 Définitions

On s'intéresse à une expérience aléatoire qui conduit à la réalisation d'un seul résultat parmi un nombre fini de résultats possibles  $\omega_1, \omega_2, \dots, \omega_n$ . On note  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  l'ensemble de ces résultats.

**Exemple 1.1.1.** – Jet d'une pièce à pile où face :  $\Omega = \{P, F\}$ .  
– Jet d'un dé :  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

Si on mesure la fréquence d'apparition du résultat  $\omega_k$  au cours d'un grand nombre de répétitions de l'expérience i.e. on calcule le rapport  $F_k = \frac{N_k}{N}$  du nombre  $N_k$  d'expériences dont le résultat est  $\omega_k$  sur le nombre total d'expériences  $N$ , on constate qu'elle fluctue de moins en moins. La limite  $p_k \geq 0$  de  $F_k$  lorsque  $N \rightarrow +\infty$  correspond à la notion intuitive de probabilité.

On appelle événement une partie  $A$  de  $\Omega$ . La fréquence de  $A$  c'est-à-dire la proportion d'expériences dont le résultat est dans  $A$  est égale à  $\sum_{k:\omega_k \in A} F_k$ . On est donc amené à associer la probabilité  $\sum_{k:\omega_k \in A} p_k$  à l'événement  $A$ . Comme la fréquence de  $\Omega$  vaut 1, en passant à la limite, on obtient  $\sum_{k=1}^n p_k = 1$ .

**Définition 1.1.2.** Une probabilité  $\mathbb{P}$  sur un ensemble fini  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$  est une pondération  $p_1, p_2, \dots, p_n$  des éléments de cet ensemble t.q.

$$\forall 1 \leq k \leq n, p_k \geq 0 \quad \text{et} \quad \sum_{k=1}^n p_k = 1.$$

On attribue à tout événement  $A \subset \Omega$  le nombre

$$\mathbb{P}(A) = \sum_{k:\omega_k \in A} p_k$$

qui est appelé probabilité de l'événement  $A$ .

**Exemple 1.1.3.** Jet de deux dés à six faces :  $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$  où  $i$  désigne la valeur de la face supérieure du premier dé et  $j$  celle du second.

Pour des raisons de symétrie (si les dés ne sont pas pipés), on munit  $\Omega$  de la pondération suivante :

$$\forall 1 \leq i, j \leq 6, p_{(i,j)} = \frac{1}{36}.$$

Soit  $A$  l'événement : les valeurs des deux dés sont identiques.

$$A = \{(1, 1), (2, 2), \dots, (6, 6)\} \quad \text{et} \quad \mathbb{P}(A) = \sum_{i=1}^6 p_{(i,i)} = \frac{6}{36} = \frac{1}{6}.$$

On note  $S$  la somme des deux dés et  $\{S = k\}$  l'événement  $\{(i, j) : S(i, j) = k\}$ . On a  $S(i, j) = i + j$ . Donc

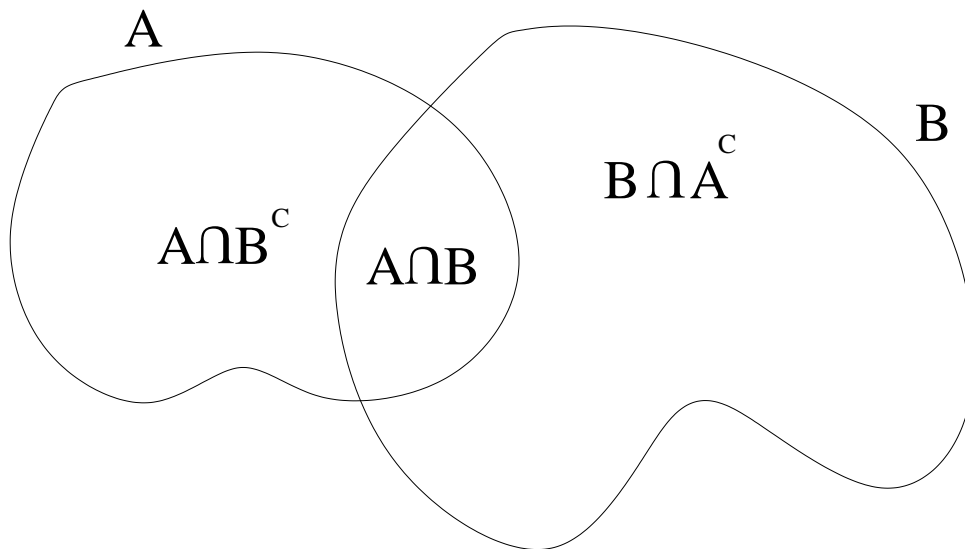
$\{S = 2\}$	$= \{(1, 1)\}$	$\mathbb{P}(S = 2) = 1/36$
$\{S = 3\}$	$= \{(1, 2), (2, 1)\}$	$\mathbb{P}(S = 3) = 1/18$
$\{S = 4\}$	$= \{(1, 3), (2, 2), (3, 1)\}$	$\mathbb{P}(S = 4) = 1/12$
$\{S = 5\}$	$= \{(1, 4), (2, 3), (3, 2), (4, 1)\}$	$\mathbb{P}(S = 5) = 1/9$
$\{S = 6\}$	$= \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$	$\mathbb{P}(S = 6) = 5/36$
$\{S = 7\}$	$= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$	$\mathbb{P}(S = 7) = 1/6$
$\{S = 8\}$	$= \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$	$\mathbb{P}(S = 8) = 5/36$
$\{S = 9\}$	$= \{(3, 6), (4, 5), (5, 4), (6, 3)\}$	$\mathbb{P}(S = 9) = 1/9$
$\{S = 10\}$	$= \{(4, 6), (5, 5), (6, 4)\}$	$\mathbb{P}(S = 10) = 1/12$
$\{S = 11\}$	$= \{(5, 6), (6, 5)\}$	$\mathbb{P}(S = 11) = 1/18$
$\{S = 12\}$	$= \{(6, 6)\}$	$\mathbb{P}(S = 12) = 1/36$

### Terminologie concernant les événements :

- Si  $\mathbb{P}(A) = 0$ , l'événement  $A$  est dit négligeable.
- Si  $\mathbb{P}(A) = 1$ , il est dit presque sûr.
- On appelle événement contraire de  $A$  et on note  $A^c$  l'événement  $\Omega \setminus A$ .
- Si  $A, B \subset \Omega$ , l'événement  $A$  et  $B$  (réalisé lorsque  $A$  et  $B$  le sont) est noté  $A \cap B$ .
- L'événement  $A$  ou  $B$  (réalisé lorsque  $A$  ou  $B$  le sont) est noté  $A \cup B$ .

### Probabilité de l'événement $A \cup B$ :

Par définition,  $\mathbb{P}(A \cup B) = \sum_{k:\omega_k \in A \cup B} p_k$ . Comme  $A \cup B$  est égal à l'union disjointe



$$(A \cap B^c) \cup (A \cap B) \cup (A^c \cap B),$$

$$\begin{aligned} \mathbb{P}(A \cup B) &= \sum_{k:\omega_k \in A \cap B^c} p_k + \sum_{k:\omega_k \in A \cap B} p_k + \sum_{k:\omega_k \in A^c \cap B} p_k \\ &= \left( \sum_{k:\omega_k \in A \cap B^c} p_k + \sum_{k:\omega_k \in A \cap B} p_k \right) + \left( \sum_{k:\omega_k \in A^c \cap B} p_k + \sum_{k:\omega_k \in A \cap B} p_k \right) - \sum_{k:\omega_k \in A \cap B} p_k \\ &= \sum_{k:\omega_k \in A} p_k + \sum_{k:\omega_k \in B} p_k - \sum_{k:\omega_k \in A \cap B} p_k \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \end{aligned}$$

Ainsi

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).}$$

**Fonction indicatrice :**

On appelle fonction indicatrice de l'événement  $A$  la fonction  $1_A : \Omega \rightarrow \{0, 1\}$  définie par

$$\forall \omega \in \Omega, 1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon.} \end{cases}$$

**Exercice 1.1.4.** Quel est l'événement  $\{\omega : 1_A(\omega) \times 1_B(\omega) = 1\}$  que l'on note aussi de façon condensée  $\{1_A \times 1_B = 1\}$  ?

Conclure que

$$\boxed{1_{A \cap B} = 1_A \times 1_B.}$$

Montrer également que

$$\boxed{1_{A^c} = 1 - 1_A \quad \text{et} \quad 1_{A \cup B} = 1_A + 1_B - 1_{A \cap B}.}$$

### 1.1.2 Probabilités uniformes

Dans le cas où les symétries font que tous les résultats possibles  $\omega_1, \omega_2, \dots, \omega_n$  jouent le même rôle, ces résultats doivent avoir la même pondération  $1/\text{Card}(\Omega)$ . On dit alors qu'il sont équiprobables.

On a alors pour tout événement  $A \subset \Omega$ ,

$$\mathbb{P}(A) = \sum_{\omega_k \in A} \frac{1}{\text{Card}(\Omega)} = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

Cette probabilité  $\mathbb{P}$  s'appelle *probabilité uniforme* sur  $\Omega$ .

**Exemple 1.1.5.** Dans le cas du jet de deux dés non pipés,  $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$  est muni de la probabilité uniforme.

**Remarque 1.1.6.** Si on s'intéresse à la somme des deux dés, on peut choisir  $\Omega = \{2, 3, 4, \dots, 12\}$ , ensemble des valeurs prises par cette somme. Mais faute de propriétés de symétrie, on ne sait pas munir cet espace d'une probabilité naturelle.

Dans l'exemple 1.1.3, en travaillant sur l'espace plus gros  $\{(i, j) : 1 \leq i, j \leq 6\}$  des couples des valeurs des deux dés muni de la probabilité uniforme, nous avons pu construire la pondération naturelle sur les valeurs de la somme des deux dés. Cette pondération n'a rien d'uniforme.

Cet exemple permet de bien comprendre l'importance du choix de l'espace de probabilité sur lequel on travaille.

Dans le cas des probabilités uniformes, les calculs se ramènent à du dénombrement.

#### Rappels de dénombrement

On se donne  $n, k \in \mathbb{N}^*$  avec  $k \leq n$ .

- Le nombre de permutations d'un ensemble à  $n$  éléments est  $n!$ .
- De façon plus générale, le nombre d'injections d'un ensemble à  $k$  éléments dans un ensemble à  $n$  éléments est

$$A_n^k = \frac{n!}{(n-k)!} = n(n-1)\dots(n-k+1).$$

Le facteur  $n$  (resp.  $n-1, \dots$ , resp.  $n-k+1$ ) vient du choix de l'image du 1<sup>er</sup> (resp. 2<sup>e</sup>, ...,  $k^e$ ) élément.

- Le nombre de parties à  $k$  éléments d'un ensemble à  $n$  éléments est

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

**Exercice résolu 1.1.7.** Dans une classe de  $n \leq 365$  élèves, quelle est la probabilité de l'événement : "2 élèves au moins sont nés le même jour" que l'on note  $A$  ?

On choisit comme espace de probabilité  $\Omega = \{f : [1, n] \rightarrow [1, 365]\}$  où pour  $1 \leq i \leq n$ ,  $f(i)$  représente le jour d'anniversaire du  $i^{\text{ème}}$  élève dans l'ordre alphabétique.

Même si les naissances ne sont pas vraiment équiréparties au long de l'année, on munit  $\Omega$  de la probabilité uniforme. On a  $\text{Card}(\Omega) = 365^n$ .

Pour calculer la probabilité de  $A$ , on peut calculer la probabilité de l'événement contraire  $A^c$  : "tous les élèves ont des dates d'anniversaire différentes". En effet comme  $A \cup A^c = \Omega$  et  $A \cap A^c = \emptyset$ ,

$$\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) - \mathbb{P}(A \cap A^c) \Rightarrow \mathbb{P}(A) = 1 - \mathbb{P}(A^c).$$

On a  $A^c = \{f : [1, n] \rightarrow [1, 365] \text{ injective}\}$ . Donc  $\text{Card}(A^c) = A_{365}^n$  et

$$\mathbb{P}(A^c) = \frac{\text{Card}(A^c)}{\text{Card}(\Omega)} = \frac{365!}{(365-n)!365^n} = \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{365-n+1}{365},$$

et

$$\mathbb{P}(A) = 1 - \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{365-n+1}{365}.$$

On peut vérifier que dès que  $n \geq 23$ , cette probabilité est supérieure à  $1/2$ .

## 1.2 Probabilité conditionnelle et indépendance

### 1.2.1 Probabilité conditionnelle

La notion de probabilité conditionnelle permet de prendre en compte l'information dont on dispose (à savoir qu'un événement  $B$  est réalisé) pour actualiser la probabilité que l'on donne à un événement  $A$  :

**Définition 1.2.1.** Soit  $\Omega$  muni d'une probabilité  $\mathbb{P}$  et  $A, B \subset \Omega$ . La probabilité conditionnelle de l'événement  $A$  sachant l'événement  $B$  est notée  $\mathbb{P}(A|B)$  et définie par

$$\mathbb{P}(A|B) = \begin{cases} \mathbb{P}(A \cap B)/\mathbb{P}(B) & \text{si } \mathbb{P}(B) > 0 \\ \mathbb{P}(A) & \text{sinon.} \end{cases}$$

**Remarque 1.2.2.** Lorsque l'on sait que l'événement  $B$  est réalisé, il est naturel d'affecter à l'événement  $A$  un poids proportionnel à  $\mathbb{P}(A \cap B)$ , ce qui justifie le choix du numérateur dans la définition précédente. Le dénominateur  $\mathbb{P}(B) = \mathbb{P}(\Omega \cap B)$  est une constante de normalisation qui assure que  $\mathbb{P}(\Omega|B) = 1$ .

**Exercice résolu 1.2.3.** 1. Dans une famille qui comporte deux enfants, l'un est une fille. On cherche la probabilité que l'autre soit un garçon.

On choisit  $\Omega = \{FF, FG, GF, GG\}$  où par exemple  $FG$  signifie que l'aîné des enfants est une fille et le second un garçon.

Cet espace est muni de la probabilité uniforme. On note

$$A = \{\text{un des enfants est un garçon}\} = \{FG, GF, GG\}$$

$$B = \{\text{un des enfants est une fille}\} = \{FF, FG, GF\}.$$

On a  $\mathbb{P}(B) = \frac{\text{Card}(B)}{\text{Card}(\Omega)} = \frac{3}{4}$ . Comme  $A \cap B = \{FG, GF\}$ ,  $\mathbb{P}(A \cap B) = \frac{\text{Card}(A \cap B)}{\text{Card}(\Omega)} = \frac{1}{2}$ . Donc la probabilité recherchée est

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/2}{3/4} = \frac{2}{3}.$$

2. On suppose maintenant que l'aîné des enfants est une fille. On veut alors connaître la probabilité pour que l'autre soit un garçon.

En reprenant la démarche ci-dessus, on obtient que cette probabilité vaut  $1/2$ .

Dans certains problèmes, ce sont les probabilités conditionnelles que l'on connaît naturellement et on est amené à utiliser la définition sous la forme

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

qui se généralise en

$$\begin{aligned} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_m) &= \mathbb{P}(A_m | A_1 \cap \dots \cap A_{m-1}) \\ &\quad \times \mathbb{P}(A_{m-1} | A_1 \cap \dots \cap A_{m-2}) \dots \mathbb{P}(A_2 | A_1) \mathbb{P}(A_1), \end{aligned}$$

pour  $m$  événements  $A_1, \dots, A_m$ .

**Exercice résolu 1.2.4.** Parmi 10 pièces mécaniques, 4 sont défectueuses. On prend successivement deux pièces au hasard dans le lot (sans remise). Quelle est la probabilité pour que les deux pièces soient correctes.

On note  $A_1$  l'événement la première pièce est bonne et  $A_2$  l'événement la seconde pièce est bonne.

Comme, au départ, il y a 6 pièces bonnes sur 10,  $\mathbb{P}(A_1) = 6/10 = 3/5$ . Lorsque l'on a retiré une pièce bonne, il reste 5 pièces bonnes sur 9. D'où  $\mathbb{P}(A_2|A_1) = 5/9$ . On conclut que la probabilité cherchée est

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2|A_1)\mathbb{P}(A_1) = \frac{5}{9} \times \frac{3}{5} = \frac{1}{3}.$$

On peut retrouver ce résultat en munissant l'espace

$$\Omega = \{\text{sous-ensembles comportant 2 pièces de l'ensemble des 10 pièces}\}$$

de la probabilité uniforme. L'événement dont on cherche la probabilité est

$$A = \{\text{sous-ensembles comportant 2 pièces de l'ensemble des 6 pièces correctes}\}.$$

On a alors

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{\binom{6}{2}}{\binom{10}{2}} = \frac{6! 8! 2!}{10! 4! 2!} = \frac{6 \times 5}{10 \times 9} = \frac{1}{3}.$$

Enfin le résultat suivant qui porte le nom de formule de Bayes est souvent utile.

**Proposition 1.2.5.** Soit  $B_1, \dots, B_m$  une partition de  $\Omega$  (i.e. des sous-ensembles disjoints de  $\Omega$  dont la réunion est  $\Omega$ ) et  $A \subset \Omega$  t.q.  $\mathbb{P}(A) > 0$ . Alors pour tout  $1 \leq i \leq m$ ,

$$\boxed{\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^m \mathbb{P}(A|B_j)\mathbb{P}(B_j)}}.$$

**Démonstration :** Le numérateur du second membre est égal à  $\mathbb{P}(A \cap B_i)$ . Le dénominateur vaut  $\sum_{j=1}^m \mathbb{P}(A \cap B_j)$  et comme les  $B_j$  forment une partition de  $\Omega$  il est égal à  $\mathbb{P}(A)$ . Donc le second membre est bien égal à  $\mathbb{P}(A \cap B_i)/\mathbb{P}(A)$ .  $\square$