Méthodes Empiriques en Linguistique Informatique

Paola Merlo

University of Geneva

année académique 2005-2006

Intuition

- Soit une pièce d'un franc. Quelle est la probabilité qu'elle tombe pile ? Justifiez votre réponse.
- Est-il plus probable que
  (a) une personne donnée ait son anniversaire un jour spécifique de l'année, le 3 mars par exemple, ou
  (b) que deux personnes aient leurs anniversaires le même jour de l'année, un jour quelconque. Justifiez votre réponse.

Intuition

- Soit une pièce d'un franc. Quelle est la probabilité qu'elle tombe pile ? Justifiez votre réponse.
- Est-il plus probable que
  (a) une personne donnée ait son anniversaire un jour spécifique de l'année, le 3 mars par exemple, ou
  (b) que deux personnes aient leurs anniversaires le même jour de l'année, un jour quelconque. Justifiez votre réponse.

Vers une définition

Comment peut-on définir la probabilité d'un événement ?

- Une mesure subjective des chances que l'événement aura lieu.
  Par exemple, si vous dites que vous aller probablement réussir les examens, quel sont les facteurs que vous prenez en compte ?
- Le pari que l'événement aura lieu.
  Par exemple, si dans les courses on donne la cote 3 contre 1, c'est qu'on suppose que ce cheval a une chance sur 4 de gagner.
- La fréquence relative avec laquelle il a lieu par rapport aux autres événements possibles dans la même situation.

Vers une définition

Comment peut-on définir la probabilité d'un événement ?

- Une mesure subjective des chances que l'événement aura lieu.
  Par exemple, si vous dites que vous aller probablement réussir les examens, quel sont les facteurs que vous prenez en compte ?
- Le pari que l'événement aura lieu.
  Par exemple, si dans les courses on donne la cote 3 contre 1, c'est qu'on suppose que ce cheval a une chance sur 4 de gagner.
- La fréquence relative avec laquelle il a lieu par rapport aux autres événements possibles dans la même situation.

Vers une définition

Comment peut-on définir la probabilité d'un événement ?

- Une mesure subjective des chances que l'événement aura lieu.
  Par exemple, si vous dites que vous aller probablement réussir les examens, quel sont les facteurs que vous prenez en compte ?
- Le pari que l'événement aura lieu.
  Par exemple, si dans les courses on donne la cote 3 contre 1, c'est qu'on suppose que ce cheval a une chance sur 4 de gagner.
- La fréquence relative avec laquelle il a lieu par rapport aux autres événements possibles dans la même situation.

Définition fréquentiste de la probabilité

Si vous analysez vos réponses, vous verrez que l'estimation de la probabilité de l'issue d'une expérience (jeter des dés, choisir un mot) est la fréquence relative de l'issue en question par rapport au total des issues possibles, sa *fréquence relative escomptée* (*expected relative frequency*).
Plus précisément, la probabilité d'un événement est définie comme étant la fréquence relative avec laquelle il a lieu par rapport aux autres événements possibles dans la même situation si l'expérience est répétée un nombre infini de fois.

Terminologie 1

- EXPÉRIENCE (ALÉATOIRE) : une procédure quelconque
  (a) pouvant être répétée un nombre illimité de fois et
  (b) ayant un ensemble bien défini d'issues.
  Par exemple, un jet des dés.
- ÉPREUVE : la réalisation d'une expérience aléatoire.
  Par exemple, si l'expérience consiste à jeter un dé, 60 jets du dé sont 60 épreuves.
- ISSUE : l'un des résultats possibles d'une expérience.
  Par exemple, si l'expérience consiste à jeter un dé, un résultat parmi : un, deux, trois, quatre, cinq ou six.

Terminologie 1

- EXPÉRIENCE (ALÉATOIRE) : une procédure quelconque (a) pouvant être répétée un nombre illimité de fois et (b) ayant un ensemble bien défini d'issues.
  Par exemple, un jet des dés.
- ÉPREUVE : la réalisation d'une expérience aléatoire.
  Par exemple, si l'expérience consiste à jeter un dé, 60 jets du dé sont 60 épreuves.
- ISSUE : l'un des résultats possibles d'une expérience.
  Par exemple, si l'expérience consiste à jeter un dé, un résultat parmi : un, deux, trois, quatre, cinq ou six.

Terminologie 1

- EXPÉRIENCE (ALÉATOIRE) : une procédure quelconque (a) pouvant être répétée un nombre illimité de fois et (b) ayant un ensemble bien défini d'issues. Par exemple, un jet des dés.
- ÉPREUVE : la réalisation d'une expérience aléatoire. Par exemple, si l'expérience consiste à jeter un dé, 60 jets du dé sont 60 épreuves.
- ISSUE : l'un des résultats possibles d'une expérience. Par exemple, si l'expérience consiste à jeter un dé, un résultat parmi : un, deux, trois, quatre, cinq ou six.

Terminologie 2

- ENSEMBLE FONDAMENTAL que l'on note Ω ou S : l'ensemble de toutes les issues de base.
  Par exemple, { un, deux, trois, quatre, cinq, six }.
- ÉVÉNEMENT (ALÉATOIRE) : N'importe quel sous-ensemble de l'ensemble fondamental Ω, y compris l'ensemble vide et l'ensemble fondamental lui-même.
  Par exemple, l'événement comprenant toutes les issues paires : { deux, quatre, six }.
  On note l'ensemble des événements $\mathcal{P}(\Omega)$, puisque c'est l'ensemble des parties (*power set*) de Ω.

Terminologie 2

- ENSEMBLE FONDAMENTAL que l'on note Ω ou S :
  l'ensemble de toutes les issues de base.
  Par exemple, { un, deux, trois, quatre, cinq, six }.
- ÉVÉNEMENT (ALÉATOIRE) : N'importe quel sous-ensemble de l'ensemble fondamental Ω, y compris l'ensemble vide et l'ensemble fondamental lui-même.
  Par exemple, l'événement comprenant toutes les issues paires : { deux, quatre, six }.
  On note l'ensemble des événements $\mathcal{P}(\Omega)$, puisque c'est l'ensemble des parties (*power set*) de Ω.

Ensemble fondamental : exemple 1

Ensemble fondamental de deux jets consécutifs d'une pièce

1ère pièce	2ème pièce
F	F
F	P
P	F
P	P

Écrivez l'événement « la première pièce montre pile. »

Ensemble fondamental : exemple 2

Ensemble fondamental d'un jet de deux dés

		dé 2					
		1	2	3	4	5	6
	1						
	2						
dé 1	3						
	4						
	5						
	6						

Écrivez l'événement « la somme des dés est 7. »

Opérations sur les événements

On utilise une algèbre des événements, un ensemble de règles pour les manipuler. Souvent ces règles sont utiles, voir nécessaires, car le problème est trop complexe et on ne peut pas le réduire à un seul événement $A \subseteq \Omega$.
Soient A et B deux événements définis dans Ω.

- L'**intersection** de A et B (notée $A \cap B$) est l'événement dont les issues sont à la fois dans A et dans B.
- L'**union** de A et B (notée $A \cup B$) est l'événement dont les issues appartiennent à A, à B ou aux deux.
- On dit que A et B sont **mutuellement exclusif** si $A \cap B = \emptyset$.
- Le **complément** de A (noté $\overline{A}$) est l'événement composé de toutes les issues en Ω qui ne sont pas dans A.

Opérations sur les événements

On utilise une algèbre des événements, un ensemble de règles pour les manipuler. Souvent ces règles sont utiles, voir nécessaires, car le problème est trop complexe et on ne peut pas le réduire à un seul événement $A \subseteq \Omega$.
Soient A et B deux événements définis dans Ω.

- L'**intersection** de A et B (notée $A \cap B$) est l'événement dont les issues sont à la fois dans A et dans B.
- L'**union** de A et B (notée $A \cup B$) est l'événement dont les issues appartiennent à A, à B ou aux deux.
- On dit que A et B sont **mutuellement exclusif** si $A \cap B = \emptyset$.
- Le **complément** de A (noté $\overline{A}$) est l'événement composé de toutes les issues en Ω qui ne sont pas dans A.

Opérations sur les événements

On utilise une algèbre des événements, un ensemble de règles pour les manipuler. Souvent ces règles sont utiles, voir nécessaires, car le problème est trop complexe et on ne peut pas le réduire à un seul événement $A \subseteq \Omega$.
Soient A et B deux événements définis dans Ω.

- L'**intersection** de A et B (notée $A \cap B$) est l'événement dont les issues sont à la fois dans A et dans B.
- L'**union** de A et B (notée $A \cup B$) est l'événement dont les issues appartiennent à A, à B ou aux deux.
- On dit que A et B sont **mutuellement exclusif** si $A \cap B = \emptyset$.
- Le **complément** de A (noté $\overline{A}$) est l'événement composé de toutes les issues en Ω qui ne sont pas dans A.

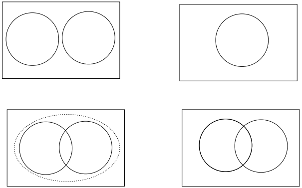
Opérations sur les événements

On utilise une algèbre des événements, un ensemble de règles pour les manipuler. Souvent ces règles sont utiles, voir nécessaires, car le problème est trop complexe et on ne peut pas le réduire à un seul événement $A \subseteq \Omega$.
Soient A et B deux événements définis dans Ω.

- L'**intersection** de A et B (notée $A \cap B$) est l'événement dont les issues sont à la fois dans A et dans B.
- L'**union** de A et B (notée $A \cup B$) est l'événement dont les issues appartiennent à A, à B ou aux deux.
- On dit que A et B sont **mutuellement exclusif** si $A \cap B = \emptyset$.
- Le **complément** de A (noté $\overline{A}$) est l'événement composé de toutes les issues en Ω qui ne sont pas dans A.

Les diagrammes de Venn

Les relations entre événements sont parfois difficiles à concevoir si on n'utilise que ces définitions.
Mais comme pour les ensembles, on peut utiliser une méthode visuelle pour simplifier la manipulation d'événements complexes : les diagrammes de Venn.
La figure ci-basse montre des diagrammes de Venn représentant une intersection, une union, un complément et deux ensembles mutuellement exclusif.



Exemple

Un club très exclusif a 100 membres, dont 30 sont des avocats. La rumeur veut que 25 des membres du club soient des menteurs et que 55 ne soient ni menteurs ni avocats. Quelle proportion d'avocats sont des menteurs ?

Mesure de probabilité

Considérons une expérience dont l'ensemble fondamental est Ω. Une *mesure de probabilité* est une fonction $\mathrm{P} : \mathcal{P}(\Omega) \rightarrow [0,1]$ des sous-ensembles de Ω (c'est-à-dire tous les événements possibles) aux nombres réels dans l'intervalle fermé $[0,1]$, avec les propriétés suivantes :

1. $0 \leq \mathrm{P}(A) \leq 1 \qquad \forall A \subseteq \Omega$
2. $\mathrm{P}(\Omega) = 1$
3. $A \cap B = \emptyset \Rightarrow \mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) \qquad \forall A, B \subseteq \Omega$

Ce sont les trois axiomes des probabilités.

Odd man out

L'exemple suivant utilise la propriété que la probabilité d'un événement qui est l'union de deux (ou plusieurs) événements est la somme des événements individuels (axiome 3). Il montre aussi, indirectement, que toutes les issues possibles d'une expérience sont des événements mutuellement exclusif (autrement dit, ils définissent une partition de l'ensemble fondamental.)
Soit le jeu suivant (*odd man out*) : chaque joueur jette une pièce non-pipée. Si toutes les pièces *sauf une* tombent du même coté, le joueur ayant jeté la pièce minoritaire est éliminé.

Supposons que trois personnes jouent ce jeu, quelle est la probabilité que quelqu'un soit éliminé au premier jet des pièces ?

L'expérience—trois personnes qui jettent une pièce chacune—donne lieu à un ensemble fondamental S de huit issues équiprobables. Six des issues satisfont l'événement A : il y a exactement une pièce différente.

• FFF	• PPF
• FPF	• PFP
• FFP	• FPP
• PFF	• PPP

$\mathrm{P}(s) = \frac{1}{8}$ pour chaque $s \in S$ (puisque les pièces ne sont pas pipées),

$$\begin{aligned}\mathrm{P}(A) = \mathrm{P}(\textit{qq sera limin}) &= \textstyle\sum_{s \in A} \mathrm{P}(s) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= \frac{6}{8}\end{aligned}$$

Dé pipé

Un dé est pipé de façon telle que la probabilité qu'une face quelconque apparaisse est directement proportionnelle au nombre sur la face en question. Quelle est la probabilité qu'un nombre pair apparaisse ?
Pour la solution de ce problème on utilisera l'axiome disant que la somme $\mathrm{P}(s)$ sur S est nécessairement 1 (Axiome 2).
L'expérience—jeter un dé—donne lieu à un ensemble fondamental contenant six issues qui ne sont pas équiprobables.

Dé pipé

Par hypothèse

$$P(A) = P(\textit{face i apparait}) = P(i) = ki, i = 1, 2, 3, 4, 5, 6$$

où k est une constante. Sur la base de l'axiome 2,

$$P(S) = \sum_{s \in S} P(s) = \sum_{i=1}^{6} P(i) = \sum_{i=1}^{6} ki = \frac{6(6+1)}{2} k = 21k = 1$$

ce qui implique que $k = 1/21$ et $P(i) = i/21$. Alors

$$\begin{aligned} P(A) = P(\textit{nombre pair}) = \sum_{s \in A} P(s) &= P(2) + P(4) + P(6) \\ &= \frac{2}{21} + \frac{4}{21} + \frac{6}{21} \\ &= \frac{12}{21} \end{aligned}$$

Probabilités conditionnelles

Les probabilités conditionnelles entrent en jeu quand on veut calculer des probabilités mais que seulement une partie de l'information concernant le résultat est disponible.
Par exemple, soit l'expérience d'un jet de dé non-pipé et soit A l'événement défini comme « le dé est tombé sur le 6. »
Évidemment, $\mathrm{P}(A) = \frac{1}{6}$.

Supposons maintenant que le dé ait été jeté, mais qu'on refuse de nous dire s'il est tombé sur 6. On nous dit par contre que l'événement B, « le dé est tombé sur un nombre pair, » a eu lieu. Quelle est maintenant la probabilité de A ?
L'événement B est composé de trois nombres pairs équiprobables, dont le 6, donc $\mathrm{P}(A|B) = \frac{1}{3}$. $\mathrm{P}(A|B)$ est la notation de la probabilité de *A étant donné que B a eu lieu*. On l'appelle la *probabilité conditionnelle* de A étant donné B, A sachant B.

Probabilités conditionnelles

C'est utile d'avoir une formule qui nous permet d'exprimer la probabilité $\mathrm{P}(A|B)$ par rapport à l'ensemble fondamental initial.

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \tag{1}$$

Cette formule est la définition de la probabilité conditionnelle.
Si on utilise cette formule, on obtient la même réponse pour l'exemple précédent :

$$\mathrm{P}(6|pair) = \frac{\mathrm{P}(6 \cap pair)}{\mathrm{P}(pair)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Probabilités conditionnelles et indépendance

Si $\mathrm{P}(A|B) = \mathrm{P}(A \cap B) / \mathrm{P}(B)$ alors, en multipliant les deux cotés par $\mathrm{P}(B)$:

$$\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\,\mathrm{P}(B) \tag{2}$$

C'est le théorème de la multiplication (*chain rule*).
Si la probabilité d'un événement A reste inchangée, après qu'un événement B a eu lieu, alors les deux événements sont *indépendants* et les deux équations suivantes s'appliquent :

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B) \tag{3}$$

$$\mathrm{P}(A|B) = \mathrm{P}(A) \tag{4}$$

Indépendance—commentaires

Attention, si nous savons que A et B sont indépendants, alors (3) et (4) s'appliquent. Mais pour prouver l'indépendance de A et B, il ne suffit pas de montrer que c'est vrai pour une issue ou l'autre, mais bien que c'est vrai en théorie pour toutes les issues !

De plus, si on utilise des probabilités estimées à partir d'observations et qu'on constate que (3) et (4) s'appliquent, ça ne prouve pas que l'indépendance est réelle. On ne peut que conclure qu'elle est probable.

En TALN, on fait souvent des suppositions d'indépendance pour simplifier un modèle probabiliste trop complexe ou impossible à estimer pour cause de données insuffisantes. On essaie de faire des hypothèses d'indépendance basée sur des théories linguistiques ou notre intuition mais parfois on va beaucoup plus loin pour arriver à simplifier un modèle suffisamment.

Populations de locuteurs bilingues

Les deux tables suivantes montrent le nombre d'individus, hommes ou femmes, dans deux populations hypothétiques, mais réalistes, définies comme monolingue et bilingue.

Population A			
	Homme	Femme	Total
Bilingue	2080	1920	4000
Monolingue	3120	2880	6000
Total	5200	4800	10000

Population B			
	Homme	Femme	Total
Bilingue	2500	1500	4000
Monolingue	2700	3300	6000
Total	5200	4800	10000

Chaque population compte le même nombre d'effectifs, la même proportion d'hommes et de femmes et la même proportion d'individus bilingues et monolingues.
Dans la population A, la proportion d'hommes bilingues est de $\frac{2080}{5200} = 0.4$, la même que la proportion de femmes bilingues $(\frac{1920}{4800}.)$
Dans la population B, par contre, $\frac{2500}{5200} = 0.48$ des hommes sont bilingues, tandis que seulement $\frac{1500}{4800} = 0.31$ des femmes sont bilingues.
Dans la réalité, on observe souvent ce type de déséquilibre dans les populations d'immigrés de première génération, où les hommes apprennent une nouvelle langue au travail tandis que les femmes passent beaucoup plus de temps au foyer ou intégrées dans leur groupe linguistique et ethnique d'origine.
En général on dit que les deux populations ont des proportions (probabilités) marginales identiques, mais une structure différente.

Populations de locuteurs bilingues

Pour la population A :

$P(\textit{homme}, \textit{bilingue}) = 0.208$
$P(\textit{femme}, \textit{bilingue}) = 0.192$
$P(\textit{homme}, \textit{monolingue}) = 0.312$
$P(\textit{femme}, \textit{monolingue}) = 0.29$

Pour la population B :

$P(\textit{homme},\textit{bilingue}) = 0.25$
$P(\textit{femme},\textit{bilingue}) = 0.15$
$P(\textit{homme}, \textit{monolingue}) = 0.27$
$P(\textit{femme},\textit{monolingue}) = 0.33$

Populations de locuteurs bilingues

Quelle est la probabilité qu'un individu soit bilingue étant donné qu'il (ou qu'elle) est homme (ou femme) ? Il faut appliquer la notion de probabilité conditionnelle :

- Pour la population A :

  $P(bilingue|homme) = 2080/5200 = 0.4$

  $P(bilingue|femme) = 1920/4800 = 0.4$

- Pour la population B

  $P(bilingue|homme) = 2500/5200 = 0.48$

  $P(bilingue|femme) = 1500/4800 = 0.31$

Populations de locuteurs bilingues

Quelle est la probabilité qu'un individu soit bilingue étant donné qu'il (ou qu'elle) est homme (ou femme) ? Il faut appliquer la notion de probabilité conditionnelle :

- Pour la population A :

  $P(bilingue|homme) = 2080/5200 = 0.4$

  $P(bilingue|femme) = 1920/4800 = 0.4$

- Pour la population B

  $P(bilingue|homme) = 2500/5200 = 0.48$

  $P(bilingue|femme) = 1500/4800 = 0.31$

Populations de locuteurs bilingues

Quelle est la probabilité qu'un individu soit bilingue étant donné qu'il (ou qu'elle) est homme (ou femme) ? Il faut appliquer la notion de probabilité conditionnelle :

- Pour la population A :

  $P(bilingue|homme) = 2080/5200 = 0.4$

  $P(bilingue|femme) = 1920/4800 = 0.4$

- Pour la population B

  $P(bilingue|homme) = 2500/5200 = 0.48$

  $P(bilingue|femme) = 1500/4800 = 0.31$

Populations de locuteurs bilingues

Quelle est la probabilité qu'un individu soit bilingue étant donné qu'il (ou qu'elle) est homme (ou femme) ? Il faut appliquer la notion de probabilité conditionnelle :

- Pour la population A :

  $P(bilingue|homme) = 2080/5200 = 0.4$

  $P(bilingue|femme) = 1920/4800 = 0.4$

- Pour la population B

  $P(bilingue|homme) = 2500/5200 = 0.48$

  $P(bilingue|femme) = 1500/4800 = 0.31$

Populations de locuteurs bilingues

Quelle est la probabilité qu'un individu soit bilingue étant donné qu'il (ou qu'elle) est homme (ou femme) ? Il faut appliquer la notion de probabilité conditionnelle :

- Pour la population A :

  $P(bilingue|homme) = 2080/5200 = 0.4$

  $P(bilingue|femme) = 1920/4800 = 0.4$

- Pour la population B

  $P(bilingue|homme) = 2500/5200 = 0.48$

  $P(bilingue|femme) = 1500/4800 = 0.31$

Populations de locuteurs bilingues

Quelle est la probabilité qu'un individu soit bilingue étant donné qu'il (ou qu'elle) est homme (ou femme) ? Il faut appliquer la notion de probabilité conditionnelle :

- Pour la population A :

  $P(bilingue|homme) = 2080/5200 = 0.4$

  $P(bilingue|femme) = 1920/4800 = 0.4$

- Pour la population B

  $P(bilingue|homme) = 2500/5200 = 0.48$

  $P(bilingue|femme) = 1500/4800 = 0.31$

En général, on applique la formule de calcul de la probabilité conditionnelle

$$P(\mathit{bilingue}|\mathit{homme}) = \frac{P(\mathit{bilingue} \wedge \mathit{homme})}{P(\mathit{homme})} = \frac{0.25}{0.52} = 0.48$$

Pour la population A, est-ce que la probabilité d'être bilingue dépend du sexe ou pas ?

Théorème de Bayes

Dans l'exemple précédent (populations A et B), on a vu qu'il y a deux probabilités conditionnelles : la probabilité du sexe étant donné le statut linguistique et la probabilité du statut linguistique étant donné le sexe.
Peut-on définir une relation entre les deux ? Oui, cette relation est ce qu'on appelle la règle ou le théorème de Bayes :

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\,\mathrm{P}(A)}{\mathrm{P}(B)} \tag{5}$$

On peut dériver cette règle à partir du théorème de la multiplication :

$$\mathrm{P}(A|B)\,\mathrm{P}(B) = \mathrm{P}(A \cap B) = \mathrm{P}(B|A)\,\mathrm{P}(A) \tag{6}$$

qu'on divise des deux cotés par $\mathrm{P}(B)$ pour obtenir (5).

Pourquoi les tests médicaux doivent être très précis

Un patient a une certaine forme de cancer ou non. Une biopsie retourne deux résultats : + si le test dit que le patient est malade, $-$ si le test indique que le patient n'est pas malade. La biopsie a une précision de 98% d'identification des cancéreux et 97% d'identification des non-cancéreux. Nous savons aussi que la probabilité a priori qu'une personne quelconque ait un cancer est de 0.008.
Quelle est alors la probabilité a posteriori qu'une personne pour qui le test a donné + soit vraiment malade d'un cancer ?
Quels sont le rappel et la précision de ce test ?

La règle de l'élimination

Comment peut-on calculer la probabilité de B si on connaît $\mathrm{P}(B|A)$ et $\mathrm{P}(A)$? Si on connaît aussi $\mathrm{P}(B|\overline{A})$, on peut utiliser la règle de l'élimination :

$$\mathrm{P}(B) = \mathrm{P}(B|A)\,\mathrm{P}(A) + \mathrm{P}(B|\overline{A})\,\mathrm{P}(\overline{A})$$

Elle est souvent utile avec le théorème de Bayes :

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\,\mathrm{P}(A)}{\mathrm{P}(B)} = \frac{\mathrm{P}(B|A)\,\mathrm{P}(A)}{\mathrm{P}(B|A)\,\mathrm{P}(A) + \mathrm{P}(B|\overline{A})\,\mathrm{P}(\overline{A})}$$

Exercice

Rappel : Le concept de probabilité conditionnelle formalise l'idée que la probabilité d'un événement puisse être changée selon qu'un autre événement se soit réalisé ou pas.
Par exemple, supposons qu'on cherche à déterminer la langue maternelle X d'un citoyen helvétique tiré au hasard. On dispose pour cela d'un modèle, disons $\mathrm{P}(a) = 0.45$, $\mathrm{P}(f) = 0.35$, $\mathrm{P}(i) = 0.15$ et $\mathrm{P}(r) = 0.05$ ($a =$ allemand, $f =$ français, $i =$ italien, $r =$ romanche). Sans autres connaissances, le pari le plus sûr serait donc d'affirmer que notre locuteur parle allemand. Si on possède une information complémentaire, comme par exemple la région d'origine Y (à choisir parmi les modalités suivantes : $A =$ Suisse alémanique, $R =$ Suisse romande, $T =$ Tessin, $G =$ Grisons), la situation peut être très différente.

Typiquement, la probabilité qu'un Suisse parle français n'est pas la même selon sa région d'origine—elle ne le serait que si les deux variables (langue et région d'origine) étaient indépendantes et nous savons que ce n'est pas le cas. Considérons un cas particulier : supposons que 80% des romands soient de langue maternelle française, 15%, allemande, 4.8%, italienne et 0.2%, romanche. Supposons aussi que les Suisses soient 70% de Suisse allemande, 25% romands, 4.5% tessinois et 0.5% des Grisons. Comment évaluer la probabilité conditionnelle $P(R|a)$ qu'un Suisse qui parle allemand soit originaire de Suisse romande ?

Variables aléatoires et fonctions de probabilité

VARIABLE ALÉATOIRE : une variable aléatoire est une fonction $X : \Omega \rightarrow \mathbb{R}$ associant un nombre réel à chaque point dans l'ensemble fondamental.

FONCTION DE PROBABILITÉ : une fonction associant une probabilité aux nombres réels qui constituent les résultats d'une variable aléatoire. (Elle doit bien sûr respecter les 3 axiomes des probabilités.)

Variables aléatoires et fonctions de probabilité

VARIABLE ALÉATOIRE : une variable aléatoire est une fonction $X : \Omega \rightarrow \mathbb{R}$ associant un nombre réel à chaque point dans l'ensemble fondamental.

FONCTION DE PROBABILITÉ : une fonction associant une probabilité aux nombres réels qui constituent les résultats d'une variable aléatoire. (Elle doit bien sûr respecter les 3 axiomes des probabilités.)

exemple

Expérience : le jet de trois pièces équilibrées.
Variable aléatoire : le nombre de pièces tombées sur le côté face.

Ensemble fondamental	
• (FFF)	• (PPF)
• (FPF)	• (PFP)
• (FFP)	• (FPP)
• (PFF)	• (PPP)

Distribution de la V.A.

Nombre de « face »	Prob
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

Ceci est un exemple d'une distribution binomiale.

Distribution binomiale

Expérience : le jet de trois pièces pipées de sorte à ce que P(F) $= p$.

Issue	Probabilité
F F F	p^3
F F P	$p^2(1-p)$
F P F	$p^2(1-p)$
P F F	$p^2(1-p)$
F P P	$p(1-p)^2$
P F P	$p(1-p)^2$
P P F	$p(1-p)^2$
P P P	$(1-p)^3$

Y	$\mathrm{P}(Y=k)$
0	$(1-p)^3$
1	$3p(1-p)^2$
2	$3p^2(1-p)$
3	p^3

Y : le nombre de pièces tombées sur face.

Forme générale de la distribution binomiale

Soit une séquence de n expériences indépendantes, dont le résultat est soit « succès » soit « échec » (ou pile et face ou 0 et 1). Soit p = P(succès d'une expérience quelconque). Supposons aussi que p soit constante d'une expérience à l'autre. Soit Y la variable correspondant au nombre total des succès pendant n expériences. Alors on dit que Y a une distribution binomiale et

$$\mathrm{P}(Y = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1 - p)^{n-k}, for\ k = 0, 1, \ldots, n.$$

Rappel

$$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$$

Une application bayesienne

Le raisonnement bayesien utilise une estimation a priori d'une probabilité pour éclairer une décision, mais ensuite peut modifier cette probabilité selon les événements qui se sont passés. L'exemple des assurances utilisait cette idée. L'exemple qui suit l'utilise aussi, ainsi que la distribution binomiale.

Supposons que M. Wehrli et Mme Merlo veuillent ouvrir un restaurant ensemble. M. Wehrli croit en son personnel et croit que les chances sont de 4 contre 1 qu'un employé embauché restera au moins un an. Mme Merlo est un peu plus pessimiste et pense que ces chances sont seulement de 3 contre 2.

Question 1 : quelle probabilité donnent-ils chacun à l'événement A : « 9 des 10 employés embauchés la première année restent au moins un an ? »

Maintenant, le banquier à qui ils ont demandé un prêt, ayant déjà fait affaire avec M. Wehrli et Mme Merlo, croit ce premier trois fois plus fiable que cette dernière pour ce genre de prédictions. Si B est l'événement « M. Wehrli a raison, » il suppose donc a priori que $\mathrm{P}(B) = .75$.
Question 2 : a priori, quelle probabilité donne le banquier à l'événement A ?
Question 3 : si, à la fin de la première année, l'événement A s'est avéré, quelle probabilité donnera le banquier a posteriori à l'événement B ?

Référence : John E. Freund, 1973, *Introduction to Probability*, pp. 182–183.

Probabilités conditionnelles et argmax

La notation $\operatorname{argmax}_x f(x)$ est souvent utilisée en TALN. Elle veut tout simplement dire de choisir la valeur de x qui maximise $f(x)$. Parfois on « triche » en utilisant argmax au lieu d'une probabilité conditionnelle. En effet, supposons qu'on ait dévelopé et estimé un modèle probabiliste $\mathrm{P}(X, A, B, C)$. Quand on veut trouver la valeur la plus probable de X étant donné des valeurs fixées de A, B et C, on utilise souvent :

$$\operatorname*{argmax}_x \mathrm{P}(X = x, A = a, B = b, C = c)$$

alors qu'il faudrait en fait utiliser :

$$\operatorname*{argmax}_x \mathrm{P}(X = x | A = a, B = b, C = c)$$

Est-ce équivalent ?