PROBABILITÉS ET STATISTIQUE
4. Probabilités

Carmelo Vaccaro

Université de Paris-Saclay

M1 - Informatique Science des Données (ISD) - Apprentissage
2022/23: premier semestre

Contenu du chapitre

1. Modélisation déterministe et modélisation probabiliste

Modélisation déterministe 1/5

On parlera de la différence entre des modèles mathématiques déterministes et des modèles probabilistes en analysant un exemple tiré de la mécanique.

En mécanique (la branche de la physique qui étudie le mouvement des objets) on utilise souvent des **modèles déterministes**, c'est-à-dire des modèles qui donnent des prédictions exactes.

Modélisation déterministe 2/5

Un exemple est la deuxième loi de Newton pour un point matériel

$$F = ma,$$

où F est la force agissant sur le point, m la masse du point et a son accélération.

Si l'on connait F, m, la position et la vitesse initiale du point on peut trouver a, puis en intégrant deux fois on trouve la position du point matériel instant par instant. Le mouvement du point est complètement déterminé (d'où le terme **déterminisme**).

Modélisation déterministe 3/5

Les seuls inconvénients à la complète détermination du mouvement du point matériel peuvent dériver :

1. du fait qu'on peut ne pas trouver une solution exacte du problème de l'intégration (d'où des erreurs d'approximation) ;
2. du fait que les mesures de la position et de la vitesse initiale du point sont nécessairement approximées (car une mesure ne peut pas avoir une précision infinie). L'erreur dû à différence entre les vraies valeurs de la vitesse et de la position et leurs mesures peut s'amplifier avec le temps et donc avec le temps la prévision de la position du point matériel peut devenir de moins en moins bonne.

Cet dernier aspect s'appelle **sensibilité aux conditions initiales** (ou **effet papillon**) et c'est la raison pour laquelle par exemple les prévisions météorologiques sont de moins en moins fiables plus l'échéance est lointaine.

Modélisation déterministe 4/5

Étudions maintenant cet exemple. Supposons d'avoir une pièce de monnaie qu'on lance dans l'air et dont on veut prédire le côté (pile ou celui face) qui va apparaitre.

Si l'on connait la force qui agit sur la pièce de monnaie et le moment de cette force par rapport à un axe de symétrie de la pièce, on peut utiliser des équations (dues à Euler) pour déterminer le mouvement de la monnaie. En particulier on peut déterminer l'instant où elle va toucher le sol et sa position dans l'espace. Puis calculer la force de choc que la monnaie subit en tombant par terre et entrer cette valeur de la force à nouveau dans les mêmes équations auparavant.

Après un nombre fini de rebonds (qui peut être de zéro) la monnaie va finalement s'arrêter et cette approche aura prédit le côté qui va apparaitre.

Modélisation déterministe 5/5

Or on voit bien que la procédure décrite auparavant présente des problèmes. Premièrement elle est très complexe, en demandant la résolution de plusieurs systèmes d'équations qui peuvent être très difficiles et dont très probablement on n'aura pas une solution exacte : dans ce cas une solution approximée pourra porter à un faux résultat, pile au lieu de face ou vice-versa.

Deuxièmement, ce problème du lancer de la monnaie est sensible aux conditions initiales, donc une petite erreur de mesure portera encore à un faux résultat.

La question se pose : existe-t-il une approche différente au problème de prédire si c'est le côte pile ou celui face qui va apparaitre ?

Modélisation probabiliste 1/2

Un **modèle probabiliste** (dit aussi **stochastique** ou **aléatoire**) est plus adapté à ce problème.

Dans ce modèle on assume que la probabilité que pile ou face apparaisse soit de 0,5 et que les lancers de la monnaie soient deux à deux indépendants, c'est-à-dire qu'un lancer antérieur n'influence pas les lancers postérieurs.

Avec ce modèle on peut faire des prévisions. On sait que pile et face ont chacun 50% de chance d'apparaitre. La probabilité d'avoir deux fois pile dans deux lancers consécutifs est de 25% comme aussi pour deux fois face ; la probabilité d'avoir une fois pile, une fois face est 50%.

Modélisation probabiliste 2/2

L'exemple précédent nous montre que dans certains situations un modèle probabiliste est préférable à un modèle déterministe.

La modélisation probabiliste s'utilise dans des domaines très variés, comme les jeux de chance, la météorologie, la finance, les systèmes avec un très grand nombre de particules (mécanique statistique), etc...

Le but de ce chapitre est de donner les bases de la théorie des probabilités.

2. Introduction aux probabilités

Procédures et résultats des procédures

En considérant les probabilités, nous traitons des *procédures* qui produisent des *résultats*.

Exemples :

- **Procédure** : lancer un dé à six faces. **Résultats** : numéro 4 ; numéro 2 ; numéro 3 ; etc...
- **Procédure** : répondre à une question de test à choix multiple. **Résultats** : BCAD... ; CCAB... ; ABCD... ; etc...
- **Procédure** : se soumettre à un test de dépistage de la consommation de drogues. **Résultats** : Amphétamine, 500 NG/ML ; MDMA, 400 NG/ML ; Opiacés, 2000 NG/ML ; etc...

Quelques concepts

- Un **événement** est un ensemble de résultats ou d'issues d'une procédure.
- Un **événement simple** est un résultat ou un événement qui ne peut pas être décomposé en éléments plus simples.
- Un **événement composé** est un événement qui n'est pas simple, c'est-à-dire qu'il combine deux ou plusieurs événements simples.
- L'**espace d'échantillonnage** d'une procédure est constitué de tous les événements simples possibles. Autrement dit, l'espace d'échantillonnage est constitué de tous les résultats qui ne peuvent pas être décomposés davantage.

Exemple (1/3)

Procédure	**Exemple d'événement**	**Espace d'échantillonnage complet**
Une seule naissance	1 femelle (événement simple)	{F, M}
3 naissances	2 femelles et 1 mâle (FFM, FMF, MFF sont tous des événements simples résultant en 2 femelles et un mâle)	{FFF, FFM, FMF, FMM, MFF, MFM, MMF, MMM}

Exemple (2/3)

- Avec une naissance, le résultat d'une femelle est un événement simple car il ne peut pas être décomposé davantage.
- Avec trois naissances, l'événement "2 femelles et 1 mâle" est un événement composé car il peut être décomposé en événements plus simples, tels que FFM, FMF ou MFF.
- Avec trois naissances, l'espace d'échantillonnage se compose des 8 événements simples (FFF, FFM, FMF, FMM, MFF, MFM, MMF et MMM).

Exemple (3/3)

- Avec trois naissances, le résultat FFM est considéré comme un événement simple, car il s'agit d'un résultat qui ne peut pas être décomposé davantage.

- Nous pouvons penser à tort que FFM puisse être décomposée en résultats individuels de F, F et M, mais F, F et M ne sont pas des résultats de la procédure en question, qui est celle de trois naissances.

Notation et définitions

- P désigne une probabilité.
- A, B, C, ... désignent des événements spécifiques.
- $P(A)$ désigne la probabilité que l'événement A se produise.

Comment déterminer les probabilités des événements

Trois méthodes différentes :

1. Approximation de la probabilité par la fréquence relative.
2. Approche classique de la probabilité (exige des résultats de même probabilité).
3. Probabilités subjectives.

Approximation de la probabilité par la fréquence relative

Effectuez (ou observez) une procédure, et comptez le nombre de fois où l'événement A se produit réellement. Sur la base de ces résultats réels, $P(A)$ est approximé comme suit :

$P(A) =$ (nombre de fois A s'est produit) / (nombre de fois que la procédure a été répétée)

Exemple : probabilité d'un accident de voiture.

Trouver la probabilité qu'une voiture choisie au hasard aux États-Unis ait un accident cette année.

Pour une année récente, 6.511.100 voitures ont eu un accident parmi les 135.670.000 voitures immatriculées aux États-Unis.

Nous pouvons utiliser l'approche de la fréquence relative comme suit :

P(accident) =
(nombre de voitures qui ont eu un accident)/(nombre total de voitures) =
6.511.100/135.670.000 = 0,0480

soit 4,8%.

Approche classique des probabilités

Supposons qu'une procédure donnée comporte n événements simples différents et que *chacun de ces événements simples a une chance égale de se produire*. Si l'événement A peut se produire dans s de ces n façons, alors

P(A) = (nombre de façons que A peut se produire) / (nombre d'événements simples différents) = s / n.

Attention : Lorsque on utilise l'approche classique, il faut vérifier que les résultats sont de probabilité égale.

Exemple : génotypes

Étant donnés des génotypes AA, Aa, aA et aa, quelle est la probabilité que nous sélectionnions un génotype dans lequel les deux composantes sont différentes ?

Dans ce cas, l'espace d'échantillonnage $\{AA, Aa, aA, aa\}$ comprend des résultats de probabilité égale. Parmi les 4 résultats, il y en a exactement 2 dans lesquels les deux composants sont différentes : Aa et aA. Nous pouvons utiliser l'approche classique pour obtenir

P(résultat avec composantes différentes) $= 2 \;/\; 4 = 0{,}5$,

soit 50%.

Probabilités subjectives

La probabilité de l'événement A est estimée en utilisant la connaissance des circonstances pertinentes.

Exemple. Lorsqu'ils tentent d'estimer la probabilité qu'un astronaute survive à une mission à bord d'une navette spatiale, les experts tiennent compte des événements passés ainsi que de l'évolution des technologies et des conditions afin d'élaborer une estimation de la probabilité. Actuellement cette probabilité a été estimée par les scientifiques de la NASA à 0,99.

Notes

1. L'approche classique exige des résultats de probabilité égale. Si les résultats ne sont pas également probables, nous devons utiliser l'approximation de la fréquence relative ou les probabilités subjectives.

2. Lorsque nous trouvons des probabilités avec l'approche de la fréquence relative, nous obtenons une approximation au lieu d'une valeur exacte. Lorsque le nombre total d'observations augmente, les approximations correspondantes ont tendance à se rapprocher de la probabilité réelle. Cette propriété est énoncée sous la forme d'un théorème communément appelé *loi des grands nombres*.

Loi des grands nombres

Lorsqu'une procédure est répétée à plusieurs reprises, la probabilité de fréquence relative d'un événement tend à se rapprocher de la probabilité réelle.

La loi des grands nombres nous dit que les approximations de la fréquence relative ont tendance à s'améliorer avec le nombre d'observations. Cette loi reflète une notion simple soutenue par le bon sens :

Une estimation de la probabilité basée sur quelques essais seulement peut être erronée de manière substantielle, mais avec un très grand nombre d'essais, l'estimation tend à être beaucoup plus précise.

Probabilités et résultats qui ne sont pas également probables

Une erreur fréquente consiste à supposer à tort que des résultats ont la même probabilité, simplement parce que nous ne savons rien de la probabilité de chaque résultat.

Lorsque nous ne savons rien de la probabilité de différentes issues possibles, nous ne pouvons pas nécessairement supposer qu'elles ont la même probabilité.

Par exemple, nous ne devons pas conclure que la probabilité de passer un examen est de 1/2 ou 0,5 (parce que soit nous passons l'examen, soit nous ne le passons pas). La probabilité réelle dépend de facteurs tels que le degré de préparation et la difficulté de l'examen.

Exemple

Dans une enquête menée auprès de consommateurs âgés de 12 ans et plus, on a demandé aux répondants combien de téléphones cellulaires étaient utilisés par le ménage.

Parmi les répondants, 211 ont répondu “aucun”, 288 ont répondu “un”, 366 ont répondu “deux”, 144 ont répondu “trois” et 89 ont répondu “quatre” ou plus.

Trouvez la probabilité qu'un ménage choisi au hasard utilise deux téléphones cellulaires ou plus.

Solution

Le nombre total de répondants est de $211 + 288 + 366 + 144 + 89 = 1098$. Les personnes interrogées ayant deux téléphones portables ou plus sont $366 + 144 + 89 = 599$. La probabilité que deux téléphones portables ou plus soient utilisés est la suivante

$$\frac{599}{1098} \approx 0{,}55.$$

Exercice.

Considérons le tableau ci-dessous.

	Le sujet a-t-il réellement menti ?	
	Non (N'a pas menti)	**Oui (A menti)**
Résultat du test positif (Le test polygraphique a indiqué que le sujet a *menti*.)	15 (**faux positif**)	42 (**vrai positif**)
Résultat du test négatif (Le test polygraphique a indiqué que le sujet n'a *pas menti*.)	32 (**vrai négatif**)	9 (**faux négatif**)

En supposant que l'un des résultats de test résumés dans le tableau est choisi au hasard, trouvez la probabilité qu'il s'agisse d'un résultat de test positif.

Solution.

Le nombre d'événements dans l'espace d'échantillonnage est de $15 + 42 + 32 + 9 = 98$.

Parmi les 98 résultats, 57 d'entre eux sont des résultats positifs (trouvés à partir de $42 + 15$).

Comme chaque résultat de test a la même probabilité d'être sélectionné, nous pouvons appliquer l'approche classique comme suit :

$$P(\textit{Résultat du test positif}) =$$
$$\text{(nombre de résultats de test positifs) / (nombre total de résultats)} =$$
$$57/98 = 0{,}582$$

Valeurs de probabilité

- La probabilité d'un événement impossible est de 0.
- La probabilité d'un événement qui est certain de se produire est de 1.
- Pour tout événement A, la probabilité de A est comprise entre 0 et 1 inclus. C'est-à-dire ,
$$0 \leq P(A) \leq 1.$$
- La somme des probabilités de tous les événements simples est 1. Cette somme est une série si le nombre d'événements simples est infini discret ou un intégral si le nombre d'événements simples est infini continu.

3. Formalisation des probabilités

Définition mathématique d'espace de probabilités 1/2

Soit Ω un ensemble ; une σ-**algèbre** sur Ω est un ensemble $\mathcal{A}$ de sous-ensembles de Ω qui contient Ω et qui est stable par complémentaire et par union dénombrable.

Cela veut dire que

1. $\mathcal{A} \ni \Omega$;
2. pour tout $E \in \mathcal{A}$, le complémentaire de E (dénoté $\overline{E}$ et défini par $\Omega \setminus E$) est une élément de $\mathcal{A}$;
3. pour tout $E_n \in \mathcal{A}$ pour $n \in \mathbb{N}$, l'union $\bigcup_{n\in\mathbb{N}} E_n$ est un élément de $\mathcal{A}$.

Dans la définition d'espace de probabilités Ω représente l'ensemble des événements simples et $\mathcal{A}$ celui des événements composés.

Soit $\mathcal{A}$ une σ-algèbre sur un ensemble Ω. Une *fonction de probabilités* sur $\mathcal{A}$ est une fonction $P : \mathcal{A} \to [0, 1]$ telle que

1. $P(\Omega) = 1$,
2. pour tous $E_n \in \mathcal{A}$ pour $n \in \mathbb{N}$ tels que E_m et E_n soient disjoints pour $m \neq n$, alors $P\big(\bigcup_{n\in\mathbb{N}} E_n\big) = \sum_{n\in\mathbb{N}} P(E_n)$.

Un **espace de probabilités** est une terne $(\Omega, \mathcal{A}, P)$, où Ω est un ensemble, $\mathcal{A}$ une σ-algèbre sur Ω et P une fonction de probabilité sur $\mathcal{A}$.

Probabilité de la différence

Pour tous $E, F \in \mathcal{A}$ on a que $P(E \setminus F) = P(E) - P(E \cap F)$.

Preuve. Comme E est l'union disjointe de $E \cap F$ et $E \setminus F$, alors d'après la 2 de la slide 34 on a que

$$P(E) = P(E \cap F) + P(E \setminus F)$$

qui implique le résultat.

Probabilité du complémentaire

Pour tout $F \in \mathcal{A}$ on a que $P(\overline{F}) = 1 - P(F)$.

Preuve. Il suffit d'utiliser le résultat de la slide précédente en prenant $E = \Omega$ et en utilisant le fait que $P(\Omega) = 1$ d'après la 1 de la slide 34.

Probabilité de l'union

Pour tous $E, F \in \mathcal{A}$ on a que $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Preuve. On a que $E \cup F$ est union à deux à deux disjointe des ensembles $E \cap F$, $E \setminus F$ et $F \setminus E$. Donc d'après la propriété 2 de la slide 34 et d'après la slide 35 on a que

$$P(E \cup F) = P(E \cap F) + P(E \setminus F) + P(F \setminus E) =$$

$$P(E \cap F) + P(E) - P(E \cap F) + P(F) - P(E \cap F) = P(E) + P(F) - P(E \cap F).$$

Probabilité de l'union disjointe

Pour tous $E, F \in \mathcal{A}$ tels que $E \cap F = \emptyset$ on a que $P(E \cup F) = P(E) + P(F)$.

Preuve. C'est une conséquence immédiate de la slide précédente.

Exercice

1. Calculer $P(E \cup F)$ avec $P(E) = 0{,}5$, $P(F) = 0{,}2$, $P(E \cap F) = 0{,}1$.
2. Calculer $P(E \cap F)$ avec $P(E) = 0{,}3$, $P(F) = 0{,}4$, $P(E \cup F) = 0{,}5$.
3. Calculer $P(\overline{E})$ avec $P(E) = 0{,}65$.
4. Calculer $P(E)$ avec $P(\overline{E}) = 0$.

Solution 1/2

1. Calculer $P(E \cup F)$ avec $P(E) = 0{,}5$, $P(F) = 0{,}2$, $P(E \cap F) = 0{,}1$.
   $P(E \cup F) = P(E) + P(F) - P(E \cap F) = 0{,}5 + 0{,}2 - 0{,}1 = 0{,}6.$

2. Calculer $P(E \cap F)$ avec $P(E) = 0{,}3$, $P(F) = 0{,}4$, $P(E \cup F) = 0{,}5$.
   On a que
   $$P(E \cap F) = P(E) + P(F) - P(E \cup F),$$
   donc
   $$0{,}5 = 0{,}3 + 0{,}4 - P(E \cup F) = 0{,}7 - P(E \cup F),$$
   qui implique que $P(E \cup F) = 0{,}2$.

Solution 2/2

3. Calculer $P(\overline{E})$ avec $P(E) = 0{,}65$.
   On a que
   $$P(\overline{E}) = 1 - P(E) = 1 - 0{,}65 = 0{,}35.$$

4. Calculer $P(E)$ avec $P(\overline{E}) = 0$.
   On a que
   $$P(E) = 1 - P(\overline{E}) = 1 - 0 = 1.$$

4. Probabilité conditionnelle

Probabilité conditionnelle

La probabilité conditionnelle d'un événement est une probabilité obtenue avec l'information supplémentaire qu'un autre événement s'est déjà produit.

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilités et soient $E, F \in \mathcal{A}$ deux événements. On dénote $P(E|F)$ la probabilité conditionnelle que l'événement E se produise, étant donné que l'événement F s'est déjà produit.

Cette probabilité est égale à

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

La probabilité conditionnelle de E étant donné F peut être trouvée en supposant que l'événement F s'est produit, puis en calculant la probabilité que l'événement E se produise.

Remarque

Soient $E, F \in \mathcal{A}$. D'après la slide précédente on a que

$$P(E \cap F) = P(E|F)P(F).$$

Exemple.

	Le sujet a-t-il réellement menti ?	
	Non (N'a pas menti)	**Oui (A menti)**
Résultat du test positif (Le test polygraphique a indiqué que le sujet a *menti*.)	15 (**faux positif**)	42 (**vrai positif**)
Résultat du test négatif (Le test polygraphique a indiqué que le sujet n'a *pas menti*.)	32 (**vrai négatif**)	9 (**faux négatif**)

Si un sujet est choisi au hasard, trouvez la probabilité que le sujet ait un résultat positif au test, étant donné que le sujet a effectivement menti. Autrement dit, trouvez $P(\text{résultat positif} \mid \text{le sujet a menti})$.

Solution.

Il y a deux façons de résoudre cet exercice.

Première méthode, intuitive. Nous voulons $P(\text{résultat positif} \mid \text{le sujet a menti})$, la probabilité d'obtenir une personne avec un résultat positif, étant donné que le sujet sélectionné a menti.

Si nous supposons que le sujet sélectionné a effectivement menti, nous n'avons affaire qu'aux 51 sujets de la deuxième colonne du tableau. Parmi ces 51 sujets, 42 ont eu un résultat positif, nous obtenons donc ce résultat :

$$P(\text{résultat positif} \mid \text{le sujet a menti}) = \frac{42}{51} = 0{,}824.$$

Solution.

Deuxième méthode, avec la formule.

$$P(\text{résultat positif} \mid \text{le sujet a menti}) =$$

$$P(\text{résultat positif } \textit{et} \text{ le sujet a menti}) \ / \ P(\text{le sujet a menti}) =$$

$$\frac{\left(\frac{42}{98}\right)}{\left(\frac{51}{98}\right)} = \frac{42}{51} = 0{,}824.$$

Exercice

Les probabilités pour le sexe des jumeaux sont les suivantes

Sexe des jumeaux				
	gar./gar.	gar./fille	fille/gar.	fille/fille
Jumeaux identiques	1/6	0	0	1/6
Jumeaux fraternels	1/6	1/6	1/6	1/6

Un médecin annonce à une femme enceinte qu'elle va donner naissance à des jumeaux garçons. D'après les données ci-dessus, quelle est la probabilité qu'elle ait des jumeaux identiques ? Autrement dit, trouvez la probabilité de jumeaux identiques étant donné que les jumeaux sont deux garçons.

Solution.

Soit $E =$ "jumeaux identiques", $F =$ "jumeaux garçons" ; nous devons calculer $P(E|F)$.

Nous avons que $P(F) = 1/6 + 1/6 = 1/3$, $P(E \cap F) = 1/6$, donc la formule donne

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{1/6}{1/3} = 0.5.$$

Confusion de l'inverse

La *confusion de l'inverse* est l'erreur consistant à croire que $P(E|F) = P(F|E)$.

En général, $P(E|F) \neq P(F|E)$. D'après la formule des probabilités conditionnelles, $P(E|F) = P(F|E)$ si et seulement si $P(E) = P(F)$.

En effet $P(E|F)/P(F|E)$ est égal à $P(E)/P(F)$, donc si $P(E)$ est 2 fois plus grand que $P(F)$, alors $P(E|F)$ est 2 fois plus grand que $P(F|E)$.

Confusion de l'inverse : exemple

Considérons la probabilité qu'à Paris il soit sombre dehors, étant donné qu'il est minuit : cela est certain, car il est toujours sombre à minuit, donc $P(\text{sombre à Paris} \mid \text{minuit à Paris}) = 1$.

Mais $P(\text{minuit à Paris} \mid \text{sombre à Paris})$ est presque nulle, car il fait sombre à beaucoup d'horaires différents de minuit.

Formule des probabilités totales

Pour tous $E, F \in \mathcal{A}$ on a que

$$P(F) = P(F \cap E) + P(F \cap \overline{E}) = P(F|E)P(E) + P(F|\overline{E})P(\overline{E}).$$

Preuve. En effet F est l'union disjointe de $F \cap E$ et de $F \cap \overline{E}$, donc d'après la slide 38 on a que

$$P(F) = P(F \cap E) + P(F \cap \overline{E}).$$

D'après la slide 44 on a l'autre égalité.

Probabilité conditionnelle du complémentaire

Pour tous $E, F \in \mathcal{A}$ on a que $P(\overline{E}|F) = 1 - P(E|F)$.

Preuve. D'après la formule de la probabilité conditionnelle, du fait que $\overline{E} \cap F = F \setminus E$ et du résultat de la slide 35, on a que

$$P(\overline{E}|F) = \frac{P(\overline{E} \cap F)}{P(F)} = \frac{P(F \setminus E)}{P(F)} =$$

$$\frac{P(F) - P(F \cap E)}{P(F)} = 1 - \frac{P(F \cap E)}{P(F)} = 1 - P(E|F).$$

5. Événements indépendants

Événements indépendants

Deux événements E et F sont **indépendants** si l'occurrence de l'un n'affecte pas la probabilité d'occurrence de l'autre.

Si E et F sont indépendants, alors $P(E|F) = P(E)$. Puisque $P(E|F) = \dfrac{P(E \cap F)}{P(F)}$, cela implique que

$$P(E \cap F) = P(E)\,P(F).$$

Si E et F ne sont pas indépendants, on dit qu'ils sont *dépendants*.

Exemple

Il y a cinq boules dans un sac, trois vertes et deux rouges. Vous sélectionnez au hasard deux boules dans le sac. Trouvez la probabilité que la première boule sélectionnée soit verte et que la seconde le soit également.

Utilisez chacune des hypothèses suivantes.

1. Supposons que les deux sélections aléatoires soient faites *avec remplacement*, donc après que la première balle ait été sélectionnée, cette balle est mise dans le sac avant que la deuxième sélection ne soit faite.
2. Supposons que les deux sélections aléatoires soient effectuées *sans remplacement*, de sorte que la première balle sélectionnée ne soit pas remise dans le sac avant que la deuxième sélection ne soit effectuée.

Solution de 1.

Les deux sélections sont indépendantes car le second événement n'est pas affecté par le premier résultat. De plus, les probabilités sont les mêmes dans les deux événements.

Donc

$$P(\text{première verte et deuxième verte}) =$$
$$P(\text{première verte})\ P(\text{deuxième verte}) = \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{25} = 0{,}36.$$

Solution de 2.

Si les deux boules sont choisies au hasard sans remplacement, les deux choix ne sont pas indépendants mais on peut considérer l'événement de tirer deux boules vertes comme deux événement indépendants avec des probabilités différentes : la première a une probabilité de 3/5, la seconde de 2/4.

La probabilité est donc de

$$\frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10} = 0.3.$$

Remarque

Souvent dans les livres et cours de probabilité on dit que comme les sélections aléatoires sans remplacement sont dépendantes on ne peut pas utiliser la formule des probabilités d'événements dépendants. À la place on utilise des méthodes très compliquées pour calculer la probabilité.

On montre maintenant comment utiliser des événements indépendants pour des sélections sans remplacement.

La probabilité de sélectionner sans remplacement deux boules vertes dans un sachet avec trois boules vertes et deux rouges est égale à la probabilité de sélectionner une boule verte dans un sachet avec trois boules vertes et deux rouges et puis de de sélectionner une boule verte dans un autre sachet avec deux boules vertes et deux rouges.

Les deux sélections sont évidemment indépendantes car on a deux sachets différents et donc on peut utiliser la formule des probabilités d'événements dépendants.

Exercice

Calculer la probabilité que deux personnes choisies au hasard soient toutes deux nées de lundi.

Solution

La probabilité que la première personne soit née le lundi est de 1/7 et la probabilité que la deuxième personne soit également née le lundi est aussi de 1/7. Comme les deux événements sont indépendants, la probabilité que les deux personnes soient nées le lundi est la suivante

$$\frac{1}{7} \cdot \frac{1}{7} = \frac{1}{49}.$$

6. Probabilité d'au moins un

Probabilité d'au moins un

Supposons que l'on veuille trouver la probabilité que parmi 3 enfants, il y ait "au moins une" fille.

"Au moins une" signifie "une ou plusieurs". Le complément d'obtenir au moins une fille parmi 3 enfants est équivalent à n'obtenir aucune fille (ou 3 garçons).

Pour trouver la probabilité d'obtenir au moins une fille parmi 3 enfants, on calcule la probabilité de son complément et on utilise la formule $P(E) = 1 - P(\overline{E})$.

La probabilité de n'avoir aucune fille parmi 3 enfants est de $1/8$, donc la probabilité d'avoir au moins une fille est de

$$1 - \frac{1}{8} = \frac{7}{8}.$$

Exemple

La probabilité qu'un pneumatique soit défectueux est de 0,0003. Trouvez la probabilité que sur 5 pneus, il y en ait au moins 1 qui soit défectueux.

Solution

La probabilité qu'un pneumatique soit défectueux est de 0,0003. Trouvez la probabilité que sur 5 pneus, il y en ait au moins 1 qui soit défectueux.

Solution. Soit E l'événement "au moins 1 des 5 pneus est défectueux". Le complément de E est $\overline{E}$ = "les 5 pneus sont bons".

La probabilité qu'un pneu soit bon est de 1 - 0,0003 = 0,9997. Les événements indiquant que les pneus sont bons sont indépendants, donc

$$P(\overline{E}) = 0,9997^5 \approx 0,9985.$$

Ainsi

$$P(E) = 1 - P(\overline{E}) \approx 1 - 0.9985 = 0.0015.$$

Exercice

Supposons que la probabilité qu'une voiture choisie au hasard ait un accident en France au cours d'une année est de 0,048. Si une famille possède quatre voitures, trouvez la probabilité qu'au moins une d'entre elles ait un accident de voiture au cours de l'année.

Solution

Soit E l'événement "au moins 1 voiture parmi 4 accidents en un an". Le complément de E est $\overline{E}$ = "aucun accident de voiture parmi 4 en un an". La probabilité qu'une voiture ait un accident est de 1 - 0,048 = 0,952. Les événements où les voitures n'ont pas d'accident sont indépendants, donc

$$P(\overline{E}) = 0.952^4 \approx 0.821.$$

Donc

$$P(E) = 1 - P(\overline{E}) \approx 1 - 0.821 = 0.179.$$

7. Théorème de Bayes

Théorème de Bayes 1/2

Le théorème de Bayes donne la relation entre $P(E|F)$ et $P(F|E)$:

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}$$

En effet $P(E|F) = \dfrac{P(E \cap F)}{P(F)}$ et d'après la remarque à la slide 44, on a $P(E \cap F) = P(F|E)P(E)$, d'où la formule plus haut.

Théorème de Bayes 2/2

La formule des probabilités totales donne

$$P(F) = P(F|E)P(E) + P(F|\overline{E})P(\overline{E})$$

et donc

$$P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|\overline{E})P(\overline{E})}.$$

Théorème de Bayes : exemple (1/7)

Un test de dépistage pour une maladie fournit les résultats suivants: il est positif à 95% pour une personne atteinte de la maladie, et négatif à 95% pour une personne non atteinte de la maladie. Supposons qu'à un moment donné, la maladie touche une personne sur 150, c'est-à-dire qu'en moyenne sur 150 personnes une est atteinte de la maladie et 149 non.

Supposons qu'une personne arbitraire fait le test et le résultat est positif : quelle est la probabilité que la personne soit effectivement atteinte de la maladie ?

La réponse est (justification donnée après) : 11%, c'est-à-dire il y a 89% de chance que la personne ne soit pas malade. Cela montre que c'est complètement inutile de faire des tests de dépistage de façon massive, on ne finirait par trouver que des faux positifs.

Théorème de Bayes : exemple (3/7)

Soit Mal l'événement "la personne est malade", Pos l'événement "la personne a un test positif". Alors $\overline{Mal}$ est l'événement "la personne n'est pas malade", $\overline{Pos}$ l'événement "la personne a un test negatif".

On sait que $P(Pos|Mal) = 0{,}95$, on veut calculer $P(Mal|Pos)$. Par le théorème de Bayes on a

$$P(Mal|Pos) = \frac{P(Pos|Mal)P(Mal)}{P(Pos)}.$$

Théorème de Bayes : exemple (4/7)

On sait aussi que $P(Pos|Mal) = 0{,}95$, $P(Mal) = 1/150$, $P(\overline{Mal}) = 1 - P(Mal) = 149/150$, $P(Pos|\overline{Mal}) = 1 - P(\overline{Pos}|\overline{Mal}) = 0{,}05$ et par la formule des probabilités totales on a

$$P(Pos) = P(Pos|Mal)P(Mal) + P(Pos|\overline{Mal})P(\overline{Mal}) =$$

$$0{,}95 \times 1/150 + 0{,}05 \times 149/150 = 0{,}056.$$

Théorème de Bayes : exemple (5/7)

Donc

$$P(Mal|Pos) = \frac{P(Pos|Mal)P(Mal)}{P(Pos)} = \frac{0{,}95 \times 1/150}{0{,}056} \approx 0{,}11.$$

Cela veut dire que si le test est positif il y a 89% de chance que la personne ne soit pas malade. Cela montre que c'est complètement inutile de faire des tests de dépistage de façon massive, on finirait par trouver que des faux positifs.

Théorème de Bayes : exemple (6/7)

Quelle est donc la bonne stratégie pour détecter des malades ? C'est celle de tester seulement les symptomatiques ou les personnes à risque, où la probabilité d'avoir la maladie serait beaucoup plus grand que 1/150.

Par exemple supposons que la personne qui obtient un test positif est symptomatique ou à risque et que cela implique que la probabilité d'avoir la maladie est 1/15. Alors $P(Mal) = 1/15$ et $P(\overline{Mal}) = 1 - P(Mal) = 14/15$.

Théorème de Bayes : exemple (7/7)

Alors on aurait

$$P(Pos) = P(Pos|Mal)P(Mal) + P(Pos|\overline{Mal})P(\overline{Mal}) =$$

$$0{,}95 \times 1/15 + 0{,}05 \times 14/15 = 0{,}11$$

et donc

$$P(Mal|Pos) = \frac{P(Pos|Mal)P(Mal)}{P(Pos)} = \frac{0{,}95 \times 1/15}{0{,}11} \approx 0{,}58.$$

Théorème de Bayes : approche simple

On examine l'exemple précédent avec une autre approche.

Sur 150.000 personnes, 149.000 sont saines et 1000 malades. Si je teste 150.000 personnes, sur 149.000 sains j'aurais 5% de faux positifs, donc 7450 faux positifs. Sur les 1000 malades, j'aurais 95% de vrais positifs, donc 950 vrais positifs.

Donc au total j'aurai $7450 + 950 = 8400$ positifs, dont 7450 faux positifs. Donc le taux de faux positifs est $7450/8400 \approx 89\%$.

Exercice

Une population de souris est étudiée dans un laboratoire par rapport à la présence ou non de deux types d'anticorps, appelés anticorps A et anticorps B. On a les faits suivants :

- si une souris porte l'anticorps A, alors 2 fois sur 5 elle porte aussi l'anticorps B ;
- si une souris ne porte pas l'anticorps A, alors 4 fois sur 5 elle ne porte pas l'anticorps B.
- La moitié de la population porte l'anticorps A.

En utilisant le théorème de Bayes, calculez la probabilité que, si une souris porte l'anticorps B, alors elle porte aussi l'anticorps A.

Solution 1/2

Il faut calculer $P(A|B)$. On a les données suivantes :

- $P(B|A) = \frac{2}{5} = 0{,}4$;
- $P(\overline{B}|\overline{A}) = \frac{4}{5} = 0{,}8$, d'où $P(B|\overline{A}) = 1 - P(\overline{B}|\overline{A}) = 1 \text{ - } 0{,}8 = 0{,}2$;
- $P(A) = 0{,}5$.

En utilisant le théorème de Bayes on a que

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0{,}4 \cdot 0{,}5}{P(B)} = \frac{0{,}2}{P(B)}.$$

Solution 2/2

Il reste à calculer $P(B)$ en utilisant la formule des probabilités totales

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

$$0{,}4 \cdot 0{,}5 + 0{,}2 \cdot 0{,}5 = 0{,}2 + 0{,}1 = 0{,}3.$$

Donc

$$P(A|B) = \frac{0{,}2}{0{,}3} = \frac{2}{3} \approx 0{,}67.$$

Fin.