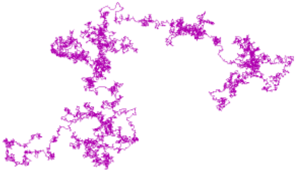
Modèles probabilistes et théorie de l'information

Christophe Lalanne, Emmanuel Chemla



Sommaire

Outline

Qu'est-ce qu'un modèle ?

Une définition opérationnelle

Un modèle est une théorie orientée vers l'action qu'elle doit servir.

$\Rightarrow$ connotation pragmatique

Qu'est-ce qu'un modèle ?

Une définition mathématique

« La modélisation mathématique est le processus par lequel un problème du monde réel est interprété et représenté en termes de symboles abstraits. la description abstraite faisant intervenir une formulation mathématique est appelée *Modèle Mathématique* associé au problème de départ. Le dit problème, issu du réel, peut être alors traité uniquement en termes mathématiques. » [Y. Cherruault, *Modèles et méthodes mathématiques pour les sciences du vivant*, PUF, 1998]

$\Rightarrow$ idée de formalisation

Principes de la modélisation

Élaboration d'un modèle

L'activité de modélisation consiste à dégager une version réduite de la réalité à l'aide d'un **modèle abstrait** (formel), qui permet

- la simulation (informatique) du comportement du système,

Principes de la modélisation

Élaboration d'un modèle

L'activité de modélisation consiste à dégager une version réduite de la réalité à l'aide d'un **modèle abstrait** (formel), qui permet

- la simulation (informatique) du comportement du système,
- la prédiction de son évolution (états possibles),

Principes de la modélisation

Élaboration d'un modèle

L'activité de modélisation consiste à dégager une version réduite de la réalité à l'aide d'un **modèle abstrait** (formel), qui permet

- la simulation (informatique) du comportement du système,
- la prédiction de son évolution (états possibles),
- la comparaison avec des données issues de l'expérimentation.

Principes de la modélisation

Outils

On a besoin de

- concepts permettant de décrire le phénomène ou système considéré,
- relations fonctionnelles liant ces concepts.

Démarche ascendante : complexification croissante du modèle initial

Pourquoi modéliser ?

- La modélisation est utile pour mieux comprendre les systèmes complexes
- Cas des Sciences cognitives : étude et compréhension des mécanismes de la pensée
  - raisonnement
  - langage
  - habiletés sensorimotrices
  - etc.

De la réalité au modèle

- La description d'un phénomène à l'aide d'un modèle implique une réduction de la réalité, donc une **perte d'information** (approximations du réel).
- La modélisation permet également l'extrapolation de certaines propriétés ou la prédiction de certains comportements (non nécessairement observables).
- Dans tous les cas, il y a perte d'information et présence d'incertitude (ou bruit) qu'il faut modéliser (pour la précision de l'estimation)

Exemples classiques

- Modèle proie-prédateur

Exemples classiques

- Modèle proie-prédateur
- Modèle de diffusion

Exemples classiques

- Modèle proie-prédateur
- Modèle de diffusion
- Modèle de reconnaissance de formes

Exemples classiques

- Modèle proie-prédateur
- Modèle de diffusion
- Modèle de reconnaissance de formes
- Modèle de détection du signal (observateur idéal)

Exemples classiques

- Modèle proie-prédateur
- Modèle de diffusion
- Modèle de reconnaissance de formes
- Modèle de détection du signal (observateur idéal)
- Modèle d'intégration de sources multiples d'informations

Exemples classiques

- Modèle proie-prédateur
- Modèle de diffusion
- Modèle de reconnaissance de formes
- Modèle de détection du signal (observateur idéal)
- Modèle d'intégration de sources multiples d'informations
- Modèle d'évolution du marché (macro/micro-économie)

Lien avec la démarche statistique

- élaborer un modèle cohérent et parcimonieux qui rende compte des données observées
- minimiser incertitude et erreur dans le modèle
- Compromis variance minimale *vs.* erreur minimale[1] lors de l'estimation des paramètres du modèle
- sélectionner le « meilleur » modèle (prédictif ou explicatif)

[1] On préférera, quand on a le choix minimiser, la variance

Intégration multisensorielle

Contexte théorique

- Traitement multimodal de l'information sensorielle
- Intégration adaptative des informations issues de 2 canaux de traitement
  - vision + audition
  - vision + modalité haptique (proprioception et mouvements actifs exploratoires de la main)
- Littérature assez vaste. . .

Intégration multisensorielle

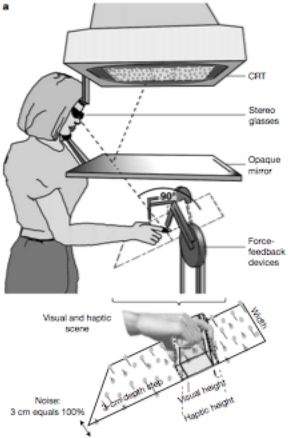
Référence

M. O. Ernst & M. B. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. **Nature**, **415** : 429–433, 2002.

When a person looks at an object while exploring it with their hand, vision and touch both provide information for estimating the properties of the object. Vision frequently dominates the integrated visual-haptic percept, for example when judging size, shape or position, but in some circumstances the percept is clearly affected by haptics. Here we propose that a general principle, which minimizes variance in the final estimate, determines the degree to which vision or haptics dominates. This principle is realized by using maximum-likelihood estimation to combine the inputs. To investigate cue combination quantitatively, we first measured the variances associated with visual and haptic estimation of height. We then used these measurements to construct a maximum-likelihood integrator. This model behaved very similarly to humans in a visual-haptic task. Thus, the nervous system seems to combine visual and haptic information in a fashion that is similar to a maximum-likelihood integrator. Visual dominance occurs when the variance associated with visual estimation is lower than that associated with haptic estimation.

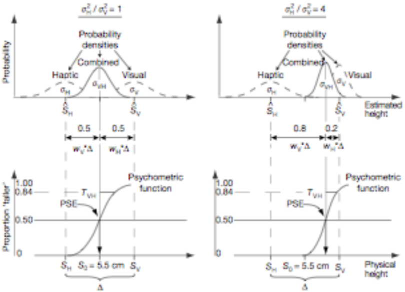
Intégration multisensorielle

Schéma expérimental



Intégration multisensorielle

Modèle proposé



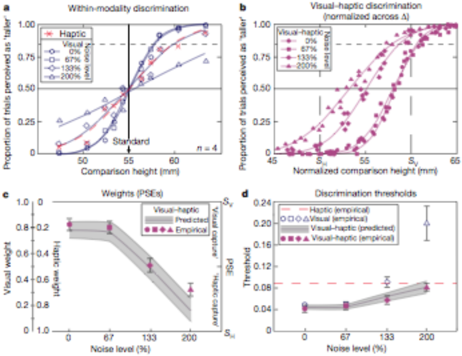
Intégration multisensorielle

Principe général

- prendre en compte les 2 estimations unimodales
- pondérer ces estimations par leur précision relative $(1/\sigma^2)$
- combiner ces estimations pondérées

Intégration multisensorielle

Résultats



Intégration multisensorielle

Conclusion

- bonne adéquation des données au modèle
- filtrage adaptatif de l'information afférente : utilisation optimale de la spécificité/sélectivité de chacune des modalités sensorielles lorsqu'elles sont recrutées simultanément

Contrôle moteur

Contexte théorique

- utilisation optimale des informations sensorielles pour le contrôle de l'action (e.g. calibration d'un geste d'atteinte ou de saisie)
- boucles sensorimotrices
- modèle de contrôle moteur (différent d'un simple servo-contrôle) : modèle direct (*forward model*) + modèle inverse (*inverse model*)

Contrôle moteur

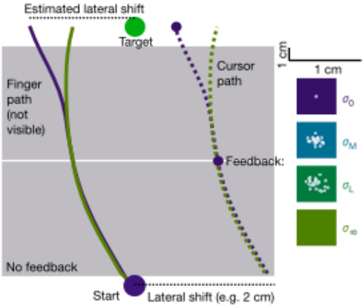
Référence

Körding & Wolpert. Bayesian integration in force estimation. **Journal of Neurophysiology**, **92** : 3161–3165, 2004.

When we interact with objects in the world, the forces we exert are finely tuned to the dynamics of the situation. As our sensors do not provide perfect knowledge about the environment, a key problem is how to estimate the appropriate forces. Two sources of information can be used to generate such an estimate : sensory inputs about the object and knowledge about previously experienced objects, termed prior information. Bayesian integration defines the way in which these two sources of information should be combined to produce an optimal estimate. To investigate whether subjects use such a strategy in force estimation, we designed a novel sensorimotor estimation task. We controlled the distribution of forces experienced over the course of an experiment thereby defining the prior. We show that subjects integrate sensory information with their prior experience to generate an estimate. Moreover, subjects could learn different prior distributions. These results suggest that the CNS uses Bayesian models when estimating force requirements.

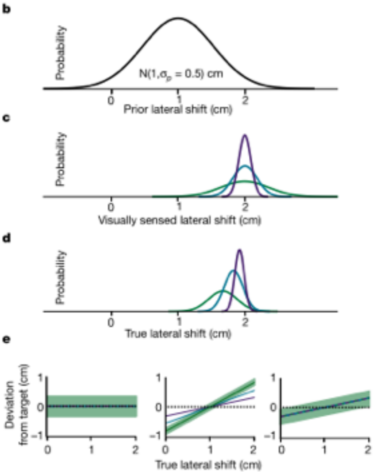
Contrôle moteur

Schéma expérimental



Contrôle moteur

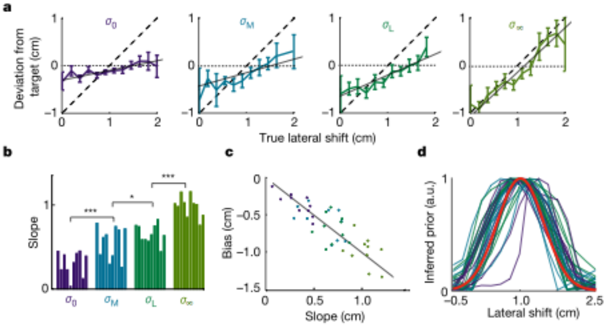
Modèle proposé



Contrôle moteur

Résultats

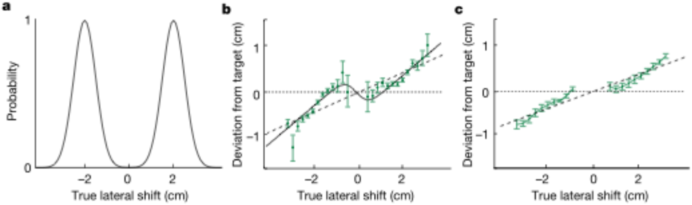
en assumant une distribution gaussienne



Contrôle moteur

Résultats (suite)

en assumant un mélange de distributions



Contrôle moteur

Conclusion

- combinaison bayésienne des informations afférentes et efférentes
- utilisation progressivement optimale des informations sensorielles
- apprentissage sensorimoteur modélisé comme un processus adaptatif tenant compte de l'expérience passée

Outline

Modèles probabilistes

Un modèle probabiliste permet de prendre en compte

- les erreurs de mesure (cf. loi de Laplace)
- l'incertitude liée à la connaissance d'un phénomène à la suite d'une observation unique ou répétée dans le temps
- des connaissances préalables dans l'estimation des caractéristiques d'un système (cadre bayésien)

L'élaboration d'un tel modèle est nécessaire à la réalisation de simulations de systèmes complexes et à la prise en compte du concept d'expériences aléatoires.

Comment définir le hasard ?

- multiplicité des définitions selon le domaine d'étude
- définition classique (statistique) : « la rencontre de deux séries causales indépendantes »
- concepts connexes : chance, incertitude, aléa numérique

▸ Paradoxe de Bertrand

Le concept de probabilité

Historique

Pascal (1654)

Étude de la répartition (équitable) de la mise entre 2 joueurs après interruption d'un jeu de pari en 3 parties

- le gain attribué à un des 2 joueurs = moyenne pondérée des gains possibles si le jeu se termine
- à partir de cette espérance mathématique, on en déduit la probabilité de gain

Le concept de probabilité

Historique

Kolmogorov (1930)

- Axiomatisation des principes du calcul des probabilités
- Démarche inverse : probabilité $\longrightarrow$ espérance
- La probabilité est considérée comme une mesure

Le concept de probabilité

La probabilité vue comme une mesure

Si A, B, C, ... sont des ensembles munis d'une fonction m dont les propriétés sont :

- à valeurs positives
- croissante relativement à l'inclusion

$$A \subset B \Rightarrow \mu(A) \leq \mu(B)$$

- additive pour la réunion de 2 ensembles disjoints

$$\mu(A \cap B) = \mu(A) + \mu(B)$$

alors A, B, C, ... sont dits mesurables par m.

Le concept de probabilité

Vocabulaire

- expérience aléatoire
- espace des possibles (Ω)
- événement (élémentaire ou sous-ensemble de Ω)

Le concept de probabilité

Exemples

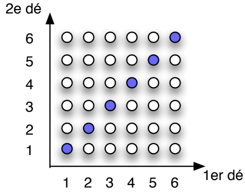
- jeu de pile ou face
  - expérience aléatoire = « jeter une pièce »
  - $\Omega = \{P; F\}$
  - un événement = observer « pile »
- jeu de dés
  - expérience aléatoire = « lancer un dé »
  - $\Omega = \{1; 2; 3; 4; 5; 6\}$
  - un événement = observer « 1 »

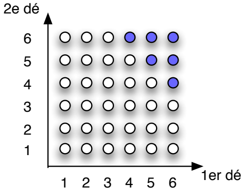
Le concept de probabilité

Exemples (suite)

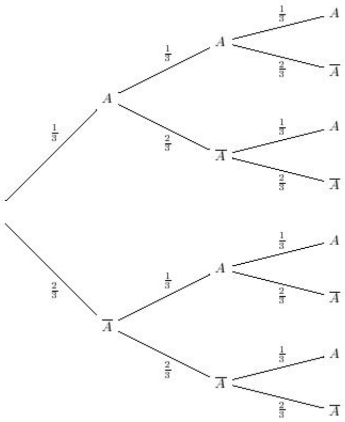
Lancer de 2 dés (réguliers)

- $E =$ « les deux dés sont identiques » : événement réalisé par 6 résultats
- $E =$ « la somme est > 9 » : événement réalisé par 6 résultats





Arbres de Bernoulli



Formalisation du concept de probabilité

Définition classique (idéale)

$$\Pr(A) = \frac{n_A}{n}$$

Si tous les événements observables (Ω) sont également possibles, la probabilité d'observer A est égale au nombre de fois où A a été observé rapporté à l'ensemble des événements observés.

Exemple : jeu infini pile/face ou lancer de dé

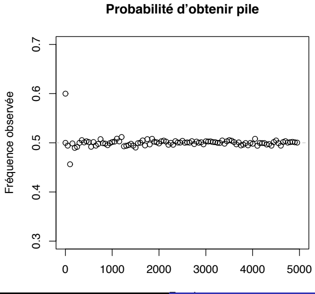
- P(« Face »)=1/2
- P(« {1 ;2} »)=1/3

Formalisation du concept de probabilité

Définition fréquentiste (objective)

$$\Pr(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

Lorsque l'expérience est répétée un (très) grand nombre de fois, la fréquence relative de l'événement approche une valeur constante – loi des grands nombres (faible/forte)



Formalisation du concept de probabilité

Définition bayésienne (subjective)

- La définition fréquentiste implique la possibilité de répéter indéfiniment *et* à l'identique l'expérience : peu réaliste !
- Probabilité d'un événement = mesure du degré de croyance (personnelle) dans l'occurrence de celui-ci
- schéma inductif et causalité (schéma de Bayes)
- Rôle des connaissances a priori (« prior distribution » en inférence bayésienne)

Formalisation du concept de probabilité

Définition bayésienne (subjective) (suite)

- Événement A
  - (1) valeur que j'attribue à A
  - (2) valeur intrinsèque de A, indépendante de (1)
  - $(1) \leq (2)$

  $\mathrm{P}(A) =$ valeur attribuée/valeur intrinsèque
- Analogie avec les paris : « ce que je suis prêt à parier »vs. « ce que je devrais parier si j'étais sûr que A se produit ».
- Cox (1946) : expression de règles quantitatives assurant un raisonnement (causal) logique et consistant

Axiomes des probabilités

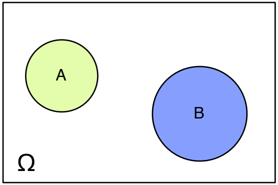
Propriétés fondamentales

Ax. 1 $P_i \geq 0$

Ax. 2 $P(\Omega) = 1$

Ax. 3 si A et B disjoints,
$P(A \cup B) = P(A) + P(B)$
généralisation : si
$A_1, A_2, \ldots, A_k$ disjoints,

$$P(\bigcup A_i) = \sum_{i=1}^{k} P(A_i)$$



Axiomes des probabilités

Conséquences

- $$A \subset B \Rightarrow P(A) \leq P(B) \tag{1}$$
- $P(A) \leq 1$ car $A \subset \Omega$
- $P(A) + P(\bar{A}) = 1$
- $$P(A \cup B) = P(A) + P(B) - P(A) \cap P(B) \tag{2}$$
- $P(\emptyset) = 1 - P(\Omega) = 0$

Conditionnement et indépendance

La probabilité de l'événement A peut être influencée par l'information apportée par la connaissance de l'événement B :

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

A et B sont dits indépendants si l'observation d'un événement n'influence pas la probabilité d'observer l'autre :

$$P(A \mid B) = P(A)$$

D'où l'on déduit

$$P(A \cap B) = P(A) \cdot P(B)$$

Exemples

Probabilité conditionnelle

On jette 2 dés et on note les n° sortis. Soit X la somme de ces 2 nombres (e.g. $X((3,2)) = 5$). On considère également l'événement Y : les 2 nombres sont pairs. On a donc

$$P(Y) = \frac{1}{4} \quad P(\{X = 6\} \cap Y) = \frac{1}{18}$$

Quelle est la probabilité d'observer $X = 6$ sachant Y ?

$$P(X = 6 \mid Y) = \frac{1/18}{1/4} = \frac{2}{9}$$

Exemples

Indépendance

- Obtenir un « 6 » et un « 2 » en lançant 2 dés : $P = 1/36$
- $A =$ « voter Chirac », $P(A) = 1/3$ (pop. totale)
  $B =$ « famille Dupont » contient 3 votants et 1 personne a voté Chirac : $P(A \mid B) = 1/3 \Rightarrow A$ et B indépendants.

Loi des probabilités totales

Dans certains cas, la probabilité d'un événement ne peut être calculée directement, mais seulement à travers un ensemble de probabilités conditionnelles connues.

Soit $B_1, B_2, \ldots, B_k$ une partition de Ω ($B_i \cap B_j = \emptyset \; \forall i \neq j$ et $\bigcup_{i=1}^{k} B_i = \Omega$),

$$P(A) = \sum_{i=1}^{k} P(A \mid B_i)P(B_i)$$

Exemple

3 boules blanches + 1 boule rouge dans une urne. On choisit une boule au hasard, puis une seconde (tirage sans remise). Quelle est la probabilité que la seconde boule soit rouge ?

$$\begin{aligned} P(R^{(2)}) &= P(R^{(2)} \mid B^{(1)})P(B^{(1)}) + P(R^{(2)} \mid R^{(1)})P(R^{(1)}) \\ &= 1/3 \cdot 3/4 + 0 \cdot 1/4 = 1/4 \end{aligned}$$

Le cadre bayésien

Principe général

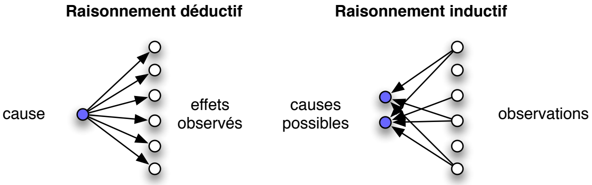
- la probabilité peut être modifiée par l'observation ou la connaissance préalable de l'événement (raisonnement causal inductif)
- formule de Bayes :

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_j P(A \mid B_j)P(B_j)}$$

- lorsqu'on connaît les probabilités a priori d'intervention des différentes causes, la formule de Bayes est une simple conséquence mathématique des théorèmes des probabilités totales et composées
- l'évaluation de ces probabilités a priori est parfois délicate (recours à une distribution uniforme)

Le cadre bayésien

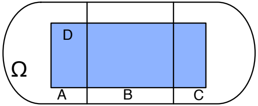
Principe général (suite)



Le cadre bayésien

Principe général (suite)

- partition de Ω
- $P(B_i)$ et $P(A \mid B_i)$ connus



Problème : ayant observé A, quelle est la nouvelle probabilité associée à B_i ?

$$P(A \mid D) = \frac{P(A \cap D)}{P(D)} = \frac{P(A)P(D \mid A)}{P(D)} \qquad (3)$$

Le cadre bayésien

Principe général (suite)

- $P(A \mid D)$, probabilité a posteriori
- $P(A)$, probabilité a priori
- $P(D \mid A)$, fonction de vraisemblance (coefficient de réévaluation)

On peut réexprimer (3) comme

$$P(\text{hypothèse} \mid \text{données}) \propto P(\text{données} \mid \text{hypothèse}) \times P(\text{hypothèse})$$

En d'autres termes, on peut changer de point de vue par rapport au symbole de conditionnement. . .

▸ Diagnostic médical

Variable aléatoire

- fonction $X : \Omega \to \mathbb{R}$
- permet de quantifier les événements

Exemple

Expérience aléatoire = lancer deux fois une pièce
$\Omega = \{(P, F); (F, P); (P, P); (F, F)\}$
Le nombre de « pile » est une VA définie comme suit :

$X((F, F))$	$X((P, F))$	$X((F, P))$	$X((P, P))$
0	1	1	2

Lois de probabilité et espérance mathématique

- cas discret (Ω fini, dénombrable) et continu

$$P(X = x) \qquad P(a < X < b) = \int_a^b f(x)dx$$

- espérance = valeur attendue, moyenne (en probabilité)

$$E(X) = \sum_x xp(x) \qquad E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

Espérance mathématique

Exemples

Lancer de 2 pièces

$\Omega = \{(P,F);(F,P);(P,P);(F,F)\}$

$$E(X) = 1/2 \cdot 1 + 1/4 \cdot 2 + 1/4 \cdot 0 = 1$$

Espérance mathématique

Exemples

Lancer de 2 pièces

$\Omega = \{(P,F);(F,P);(P,P);(F,F)\}$

$$E(X) = 1/2 \cdot 1 + 1/4 \cdot 2 + 1/4 \cdot 0 = 1$$

Cas d'une VA réelle

$$f(x) = \begin{cases} \frac{1}{2} & \text{si } 0 \leq x \leq 2, \\ 0 & \text{sinon.} \end{cases}$$

alors

$$E(X) = \int_0^2 \frac{x}{2} dx = 1$$

Entropie

Définition

Il s'agit de la quantité moyenne d'information apportée par l'observation de la valeur prise par une VA.
L'information contenue dans une réalisation est donnée par :

$$H(X = x) = \ln \frac{1}{p(x)} = - \ln p(x)$$

Si on répète l'expérience (n grand) :

$$H(X) = - \sum_i p(x_i) \ln p(x_i)$$

On parle de quantité moyenne d'information générée par les événements observés.

Entropie

Exemple

Soit une pièce « faussée » telle que $P(\text{pile}) = 0.9$. La VA associée est définie comme $X(\text{pile}) = 1$, $X(\text{face}) = 0$.
On a alors :

$$H(X = 1) = -\ln 0.9 = 0.105$$

Observer « pile » apporte peu d'information puisque $P(\text{pile}) = 0.9$ (événement presque certain).
Si on considère un lancer infini, avec $p = 0.5$:

$$H(X) = -(0.5 \ln 0.5 + 0.5 \ln 0.5) = 0.693$$

Outline

Théorie de l'information

- un schéma et une théorie de la communication (Shannon, 1949)
- formalisation et quantification de l'information circulant entre un système (ouvert) et son environnement
- exemples : télécommunications, langage naturel, transmission de l'information neuronale, etc.

Schéma de la communication

- transmission d'un message codé (signal) entre un émetteur et un récepteur, éventuellement en présence de bruit sur le canal de transmission
- signal = phénomène physique discriminable
- code = bijection, optimalité

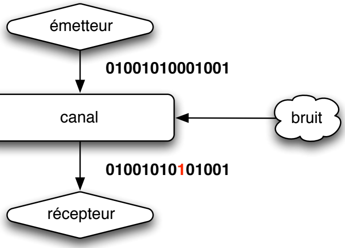


Schéma de la communication

Quantité d'information

- information = sélection de possibilités (e.g. suite de choix binaires, 0/1)
- « être informé » = apprendre quelque chose de nouveau, i.e. diminution de l'incertitude

La quantité d'information est assimilée à la probabilité de survenue d'un état par rapport à tous ses états possibles.

Schéma de la communication

Quantité d'information (suite)

- Bit (binary digit) : unité de mesure élémentaire
- Bit = quantité d'information apportée par un événement dont la probabilité de survenue vaut $1/2$

Exemples

- choix binaire, $Q(I) = 1\,\text{bit}$
- 4 possibilités, $Q(I) = 2\,\text{bits}$
- assigner un n° à chaque être humain : $Q(I) > 32\,\text{bits}$

Schéma de la communication

Intérêts et limites

Intérêts

- réduction de la masse d'information (redondance, compression, correction)
- notion de codage optimal par la maximisation de la quantité d'information

Limites

- suppression de la richesse sémantique du message (langage naturel $\neq$ code)

Exemple : codes correcteurs d'erreur

Principe général

- émetteur et receveur se mettent d'accord pour choisir les mots de code
- envoi du message C, réception du message R
- détection :
  - quand R n'est pas un mot de code, la transmission est fausse
  - quand R est un mot de code, on n'en sait rien
- vecteur d'erreur : $e = C \oplus R$
- $w(e) = d(C, R)$ (d : distance de Hamming)
- correction : calcul de $C = R \oplus e$, i.e. choix du mot de plus petit poids car c'est le vecteur d'erreur le plus probable
- le receveur corrige R en le remplaçant par le mot de code le plus proche

Exemple : codes correcteurs d'erreur

Choix des mots de code

- codage par blocs d'un ensemble de k bits à transmettre ($k =$ dimension des blocs)
- à chaque bloc on ajoute r bits de contrôle
- longueur du mot/code : $n = k + r$
- rendement du code $= k/n$
- objectif : le rendement doit être le plus près possible de 1
- mots binaires de longueur $n =$ messages ; mots de code $=$ certains messages particuliers
- 2^n messages parmi lesquels 2^k sont les mots de code

Exemple : codes correcteurs d'erreur

Test de parité

Soit $k = 3$, $r = 1$: on ajoute 1 bit de parité de façon que le poids du mot obtenu soit pair. On a $n = 4$, donc le rendement vaut 0.75.

Codage :

```
000 -> 0000
001 -> 0011
010 -> 0101
011 -> 0110
100 -> 1001
101 -> 1010
110 -> 1100
111 -> 1111
```

Exemple : codes correcteurs d'erreur

Test de parité (suite)

0000	q^4
0001	pq^3
0010	pq^3
0011	p^2q^2
0100	pq^3
0101	p^2q^2
0110	p^2q^2
0111	p^3q
1000	pq^3
$\vdots$	$\vdots$

- Total : $q^4 + 4pq^3 + 6p^2q^2 + 4p^3q + p^4 = (p+q)^4 = 1$
- Probabilité de bonne transmission : q^4
- Probabilité de mauvaise transmission : $4pq^3 + 6p^2q^2 + 4p^3q + p^4 = 4p - 6p^2 + 4p^3 - p^4$
- Probabilité de ne pas s'en apercevoir : $6p^2q^2 + p^4 = 6p^2 - 12p^3 + 7p^4$
- Proportion de messages faux non détectés : $\frac{6p^2-12p^3+7p^4}{4p-6p^2+4p^3-p^4} \approx 1.5$

Chaînes de Markov

Principe général

- processus stochastique = fonction aléatoire, i.e. un résultat observable ne peut être défini qu'en termes probabiliste
- processus de Markov = la seule source de modification potentielle de l'état du processus au temps n est constituée par ses états précédents
- si la probabilité de transition d'état P ne dépend pas du temps, alors on parle de chaîne de Markov
- Théorème de Chapman-Kolmogorov : probabilité de transition au temps n obtenue à partir de P^n (cas fini)
- états absorbants = états à partir desquels le processus n'évolue plus

Chaînes de Markov

Exemple

Soit la matrice de transition suivante :

$$P = \begin{array}{c|ccc} & R & N & S \\ \hline R & 1/2 & 1/4 & 1/4 \\ N & 1/2 & 0 & 1/2 \\ S & 1/4 & 1/4 & 1/2 \end{array}$$

où $R =$ « il pleut », $N =$ « il fait beau », $S =$ « il neige », et $p_{ij} =$ probabilité de passer de l'état i à l'état j.

Chaînes de Markov

Exemple

En d'autres termes, $p_{ij} =$ probabilité de l'événement j demain sachant que l'événement i s'est produit aujourd'hui, i.e.

$$P = \begin{array}{c|ccc} & R & N & S \\ \hline R & \Pr(R \mid R) & \Pr(N \mid R) & \Pr(S \mid R) \\ N & \Pr(R \mid N) & \Pr(N \mid N) & \Pr(S \mid N) \\ S & \Pr(R \mid S) & \Pr(N \mid S) & \Pr(S \mid S) \end{array}$$

Chaînes de Markov

Exemple (suite)

Question : sachant qu'il pleut aujourd'hui, quelle est la probabilité qu'il neige dans 2 jours ?

résultat = réunion (disjointe) de 3 événements

- il pleut demain et il neige dans 2 jours
- il fait beau aujourd'hui et il neige dans 2 jours
- il neige aujourd'hui et il neige dans 2 jours

soit

$$\Pr(S^{(2)} \mid R) = \Pr(S \mid R) \cdot \Pr(R \mid R) + \\ + \Pr(S \mid N) \cdot \Pr(N \mid R) + \Pr(S \mid S) \cdot \Pr(S \mid R)$$

soit

$$p_{13}^{(2)} = p_{13} \cdot p_{11} + p_{23} \cdot p_{12} + p_{33} \cdot p_{13}$$

Chaînes de Markov

Exemple (suite)

Or $p_{13} \in P^2$ et il suffit de calculer directement les éléments de $P \times P$.
Si on généralise :

$$P^1 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 1/2 & 1/4 & 1/4 \\ N & 1/2 & 0 & 1/2 \\ S & 1/4 & 1/4 & 1/2 \end{array}, \quad P^2 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.438 & 0.188 & 0.378 \\ N & 0.375 & 0.250 & 0.378 \\ S & 0.375 & 0.188 & 0.438 \end{array}, \quad P^3 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.406 & 0.203 & 0.391 \\ N & 0.406 & 0.188 & 0.406 \\ S & 0.391 & 0.203 & 0.406 \end{array},$$

$$P^4 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.402 & 0.199 & 0.398 \\ N & 0.398 & 0.203 & 0.398 \\ S & 0.398 & 0.199 & 0.402 \end{array}, \quad \ldots \quad P^7 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.400 & 0.200 & 0.400 \\ N & 0.400 & 0.200 & 0.400 \\ S & 0.400 & 0.200 & 0.400 \end{array}$$

Chaînes de Markov

Exemple (suite)

Question : quelle est la probabilité qu'il pleuve demain en se donnant une loi de probabilité R/N/S pour aujourd'hui ?

Soit $u^{(n)}$ un vecteur de probabilité pour l'état de départ. On peut montrer que $u^{(n)} = uP^n$

Par exemple, si $u = \{.9, .05, .05\}$ et $v = \{.05, .05, .9\}$, quelle est la probabilité qu'il neige

- après 3 jours ?
- après 7 jours ?

Chaînes de Markov

Exemple (suite)

$$u^{(3)} = uP^3 = [.9, .05, .05] \begin{bmatrix} 0.406 & 0.203 & 0.391 \\ 0.406 & 0.188 & 0.406 \\ 0.391 & 0.203 & 0.406 \end{bmatrix} = [.406, .202, \mathbf{.392}]$$

$$u^{(7)} = uP^7 = [.9, .05, .05] \begin{bmatrix} 0.400 & 0.200 & 0.400 \\ 0.400 & 0.200 & 0.400 \\ 0.400 & 0.200 & 0.400 \end{bmatrix} = [.400, .200, \mathbf{.400}]$$

$$v^3 = uP^3 = [.392, .202, \mathbf{.406}] \quad v^7 = uP^7 = [.400, .200, \mathbf{.400}]$$

Il y a convergence indépendamment de la distribution initiale : il s'agit d'une chaîne régulière.

Marche aléatoire et mouvement brownien

- système à dynamique discrète utilisé pour modéliser une succession d'états aléatoires (e.g. positions spatiales décorrélées)
- à chaque instant, le futur dépend de son état présent uniquement : le système n'a pas de mémoire
- utilisation dans le parcours des graphes (longueur et direction des pas fixes ou aléatoires) : modélisation de l'exploration « aveugle » par un processus stochastique
- modélisation de phénomènes naturels, e.g. mouvements aléatoires (en apparence) de grains de pollens dans l'eau : mouvement brownien

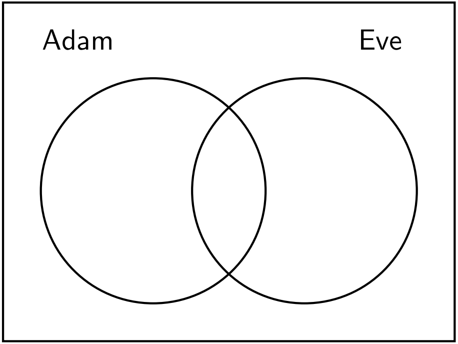
Petits exercices corrigés

Dénombrement I

Quand Adam et Eve se sont mariés, chacun d'eux avait déjà plusieurs enfants de mariages précédents. Au bout de quelques années, il y a 8 enfants dans leur maison : Adam est le pére de 6 d'entre eux, Eve est la mére de 5. Combien d'enfants ont-ils eu ensemble ?

Dénombrement I

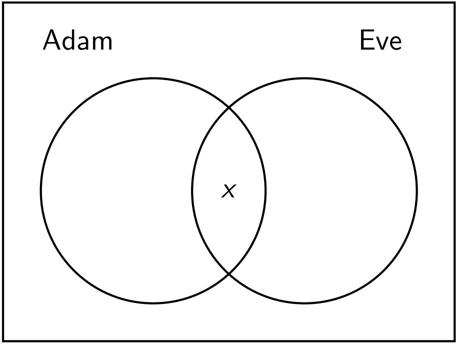
Quand Adam et Eve se sont mariés, chacun d'eux avait déjà plusieurs enfants de mariages précédents. Au bout de quelques années, il y a 8 enfants dans leur maison : Adam est le pére de 6 d'entre eux, Eve est la mére de 5. Combien d'enfants ont-ils eu ensemble ?



Source : Euréka. *Faites vos jeux*, Dunod, 2002.

Dénombrement I

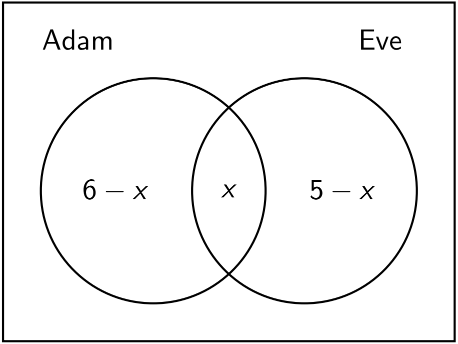
Quand Adam et Eve se sont mariés, chacun d'eux avait déjà plusieurs enfants de mariages précédents. Au bout de quelques années, il y a 8 enfants dans leur maison : Adam est le pére de 6 d'entre eux, Eve est la mére de 5. Combien d'enfants ont-ils eu ensemble ?



Source : Euréka. *Faites vos jeux*, Dunod, 2002.

Dénombrement I

Quand Adam et Eve se sont mariés, chacun d'eux avait déjà plusieurs enfants de mariages précédents. Au bout de quelques années, il y a 8 enfants dans leur maison : Adam est le pére de 6 d'entre eux, Eve est la mére de 5. Combien d'enfants ont-ils eu ensemble ?



Source : Euréka. *Faites vos jeux*, Dunod, 2002.

Dénombrement II

On jette 3 dés. On peut obtenir une somme de 9 ou 10 de 6 façons différentes. L'expérience montre que l'on obtient 10 plus souvent que 9, pourquoi ?

Dénombrement II

On jette 3 dés. On peut obtenir une somme de 9 ou 10 de 6 façons différentes. L'expérience montre que l'on obtient 10 plus souvent que 9, pourquoi ?

Solution

Sommes qui font 9	Sommes qui font 10

Dénombrement II

On jette 3 dés. On peut obtenir une somme de 9 ou 10 de 6 façons différentes. L'expérience montre que l'on obtient 10 plus souvent que 9, pourquoi ?

Solution

Sommes qui font 9

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

Sommes qui font 10

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

Dénombrement II

On jette 3 dés. On peut obtenir une somme de 9 ou 10 de 6 façons différentes. L'expérience montre que l'on obtient 10 plus souvent que 9, pourquoi ?

Solution

Sommes qui font 9

1 2 6	1 3 5	1 4 4	2 2 5	2 3 4	3 3 3
x6	x6	x3	x3	x6	x1

Sommes qui font 10

1 3 6	1 4 5	2 2 6	2 3 5	2 4 4	3 3 4
x6	x6	x3	x6	x3	x3

Régle de Bayes I – niveau : facile

Rappel : $P(A \text{ et } B) = P(A|B) \times P(B)$
Enoncé :
Un dé est pipé :

Il tombe une fois sur 5 sur un nombre pair.

En revanche :
– les chiffres pairs sont tous équiprobables entre eux ;
– et les chiffres impairs sont aussi tous équiprobables.

Question : Quelle est la probabilité de faire 2 ?

Régle de Bayes I – niveau : facile

Rappel : $P(A \text{ et } B) = P(A|B) \times P(B)$

Enoncé :

Un dé est pipé :

Il tombe une fois sur 5 sur un nombre pair.

En revanche :

– les chiffres pairs sont tous équiprobables entre eux ;

– et les chiffres impairs sont aussi tous équiprobables.

Question : Quelle est la probabilité de faire 2 ?

$$P(2) = P(2 \text{ et } pair)$$

Régle de Bayes I – niveau : facile

Rappel : $P(A \text{ et } B) = P(A|B) \times P(B)$
Enoncé :
Un dé est pipé :

Il tombe une fois sur 5 sur un nombre pair.

En revanche :
– les chiffres pairs sont tous équiprobables entre eux ;
– et les chiffres impairs sont aussi tous équiprobables.

Question : Quelle est la probabilité de faire 2 ?

$$\begin{aligned} P(2) &= P(2 \text{ et } \mathit{pair}) \\ &= P(2|\mathit{pair}) \times P(\mathit{pair}) \end{aligned}$$

Régle de Bayes I – niveau : facile

Rappel : $P(A \text{ et } B) = P(A|B) \times P(B)$
Enoncé :
Un dé est pipé :

Il tombe une fois sur 5 sur un nombre pair.

En revanche :
– les chiffres pairs sont tous équiprobables entre eux ;
– et les chiffres impairs sont aussi tous équiprobables.

Question : Quelle est la probabilité de faire 2 ?

$$\begin{aligned} P(2) &= P(2 \text{ et } \mathit{pair}) \\ &= P(2|\mathit{pair}) \times P(\mathit{pair}) \\ &= 1/3 \times 1/5 \end{aligned}$$

Régle de Bayes I – niveau : facile

Rappel : $P(A \text{ et } B) = P(A|B) \times P(B)$
Enoncé :
Un dé est pipé :
Il tombe une fois sur 5 sur un nombre pair.
En revanche :
– les chiffres pairs sont tous équiprobables entre eux ;
– et les chiffres impairs sont aussi tous équiprobables.
Question : Quelle est la probabilité de faire 2 ?

$$\begin{aligned} P(2) &= P(2 \text{ et } pair) \\ &= P(2|pair) \times P(pair) \\ &= 1/3 \times 1/5 \\ &= 1/15 \end{aligned}$$

Régle de Bayes II - un peu plus difficile

Rappel : $P(A \text{ et } B) = P(A|B) \times P(B)$
Bon sens : $P(A \text{ et } B) = P(B \text{ et } A)$...

Enoncé :
Pour dépister une maladie, on applique un test.
– Si le patient est effectivement malade, le test donne un résultat positif dans 99% des cas.
– Mais il se peut aussi que le résultat du test soit positif alors que le patient est en bonne santé, et ceci se produit dans 2% des cas.
– En moyenne, un patient sur 1000 est atteint de cette maladie.

Question : Quelle est la probabilité pour que soit atteint quelqu'un dont le test est positif ?

Facultatif : Comment améliorer ce resultat ?

Régle de Bayes II - Solution

$$\begin{aligned}
\text{M} : &\quad \text{“le patient est malade”} \\
\text{non-M} : &\quad \text{“le patient n'est pas malade”} \\
+ : &\quad \text{“le résultat du test est positif”}
\end{aligned}$$

Régle de Bayes II - Solution

$$
\begin{aligned}
\text{M} : &\quad \text{“le patient est malade”} \\
\text{non-M} : &\quad \text{“le patient n'est pas malade”} \\
+ : &\quad \text{“le résultat du test est positif”}
\end{aligned}
$$

On cherche : $P(M|+)$

Régle de Bayes II - Solution

$$\begin{aligned} \text{M} &: \text{“le patient est malade”} \\ \text{non-M} &: \text{“le patient n'est pas malade”} \\ + &: \text{“le résultat du test est positif”} \end{aligned}$$

On cherche : $P(M|+)$

On connaît :

$$\begin{aligned} P(+|M) &= 99/100 \\ P(+|\text{non-M}) &= 2/100 \\ P(M) &= 1/1000 \end{aligned}$$

Régle de Bayes II - Solution

$$\begin{aligned} \text{M :} &\quad \text{“le patient est malade”} \\ \text{non-M :} &\quad \text{“le patient n'est pas malade”} \\ + \text{ :} &\quad \text{“le résultat du test est positif”} \end{aligned}$$

On cherche : $P(M|+)$

On connaît :

$$\begin{aligned} P(+|M) &= 99/100 \\ P(+|\text{non-M}) &= 2/100 \\ P(M) &= 1/1000 \end{aligned}$$

et la régle de Bayes...

Régle de Bayes II - Solution

M : "le patient est malade"
non-M : "le patient n'est pas malade"
$+$: "le résultat du test est positif"

$$
\begin{aligned}
P(M|+) &= \frac{P(M \text{ et } +)}{P(+)} \\
&= \frac{P(+|M) \times P(M)}{P(+)} \\
&= \frac{P(+|M) \times P(M)}{P(+|M) \times P(M) + P(+|\text{non-M}) \times P(\text{non-M})} \\
&= \frac{99/100 \times 1/1000}{99/100 \times 1/1000 + 2/100 \times 999/1000} \\
&= \frac{99}{2097} \approx \frac{100}{2000} \approx 5\%
\end{aligned}
$$