RÉGION ACADÉMIQUE

MINISTÈRE
DE L'ÉDUCATION NATIONALE
ET DE LA JEUNESSE

MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION

INFERENCE BAYESIENNE

Inspection Pédagogique Régionale de Mathématiques

Ce zoom sur « inférence bayésienne » a pour objectifs :

- d’explorer plus avant la partie « probabilités et statistique » du programme de mathématiques complémentaires ;
- de mettre en évidence des liens entre les différents thèmes du programme ;
- de mettre en évidence des liens entre les thèmes du programme et les contenus ;
- de mettre en évidence des liens entre les thèmes et contenus du programme avec d’autres disciplines ;
- de proposer quelques exemples d’activités.

SOMMAIRE

OBJECTIF

DÉCRIRE ET METTRE EN OEUVRE LES PRINCIPES DU CALCUL UTILISANT DES PROBABILITÉS CONDITIONNELLES ET NOTAMMENT LA **FORMULE DE BAYES** POUR **L'INVERSION DES CONDITIONNEMENTS**.

(EXTRAIT DU BO)

OBJECTIF

La formule de Bayes $P_B(A) = \frac{P_A(B)P(A)}{P(B)}$ [...] montre la distinction essentielle entre $P_B(A)$ et $P_A(B)$. **Bien comprendre cette distinction est un objectif majeur.**

(EXTRAIT DU BO)

ENJEUX

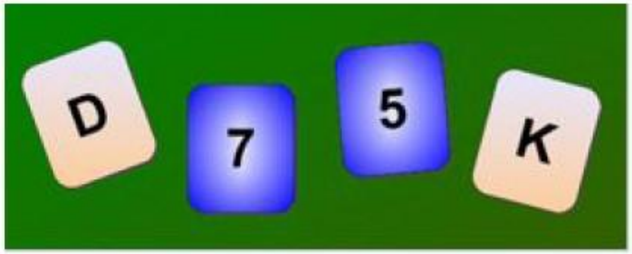
Inférence bayésienne et raisonnement

ENJEUX

Quelle(s) cartes devez-vous retourner (et seulement celles-là) pour déterminer la véracité de la règle suivante :

« Si une carte a un D sur une face,

alors elle porte un 5 sur l'autre face. »



ENJEUX

En notant :
A : « une carte a un D sur une face »
B : « la carte porte 5 sur l'autre face »,
On doit vérifier : A $\Rightarrow$ B.

Pour cela on doit retourner la carte D (vérifier A $\Rightarrow$ B) et la carte 7 (vérifier la contraposée, non B $\Rightarrow$ non A).
Attention : la notion de contraposée figure au programme de la spécialité mathématiques de terminale mais ni en spécialité de première, ni en enseignement optionnel de mathématiques complémentaire.

Erreurs fréquentes :

Ne retourner que la carte D : notion de contraposée négligée.

Retourner les cartes D et 5. Arrêtons nous un instant sur cette erreur.

ENJEUX

Retourner les cartes D et 5.

Explication 1 : on vérifie $B \Rightarrow A$ au lieu de $A \Rightarrow B$.
Explication 2 : on considère que si $A \Rightarrow B$ alors $A \Leftrightarrow B$.

En fait dans les deux cas on peut considérer que le raisonnement suivant est tenu :
Si on a $A \Rightarrow B$, et que B est vrai alors il est très probable que A soit vrai.

Ce raisonnement est celui que l'on tient dans le **raisonnement inductif** lors par exemple d'une étape de recherche dans la résolution d'u problème.

Par exemple, recherche d'une formule, d'une propriété sur quelques cas particuliers avant de la démontrer par raisonnement déductif dans le cas général.

C'est le raisonnement le plus naturel pour un grand nombre d'élèves et celui dont ils se suffisent parfois (voir formation sur Raisonnement – académie de la Guyane).
Bien évidemment c'est un raisonnement qui peut conduire à des erreurs manifestes, comme l'exemple suivant et très classique le montre.

ENJEUX

Raisonnement déductif :
« Tous les hommes sont mortels.
Socrate est un homme.
Donc Socrate est mortel. »

Raisonnement inductif :
« Tous les chats sont mortels.
Socrate est mortel.
Donc Socrate est un chat. »

Pour autant est-ce si simple de traiter ce type d’erreur d’un point de vue pédagogique ?

ENJEUX

Un cerveau bayésien ?

Sources et pour aller plus loin :

Le cerveau statisticien : la révolution Bayésienne en sciences cognitives, Stanislas Dehaene.
https://www.college-de-france.fr/site/stanislas-dehaene/p1346267510661_content.htm

Café des sciences
https://webinet.cafe-sciences.org/articles/les-bebes-genies-de-la-statistique/

ENJEUX

Je vous invite à ce moment du diaporama à regarder la vidéo intitulée « Charlie Chaplin » (lien ci-dessous) puis à revenir à cette présentation.

https://youtu.be/QbKw0_v2clo v

ENJEUX

Aux mêmes causes les mêmes effets, l'anamorphose devant l'hôtel de ville de Paris ci-dessous :

(Crédit photos Pascal Berger)

ENJEUX

Si notre cerveau voit une sphère sur la photo de droite même s'il sait que c'est une illusion d'optique, c'est qu'il se dit par expériences et répétitions de ces expériences (observations habituelles) qu'il est bien plus plausible que cette forme soit une sphère que plutôt la résultante de la construction étrange de gauche.

De même pour Charlie Chaplin, la probabilité que ce soit un visage en relief étant bien supérieure que celle d'un masque en creux, notre cerveau « préfère » voir un visage en relief qui tourne dans l'autre sens que le creux du masque.

Dans les deux cas notre cerveau semble avoir des hypothèses à priori sur le monde qui l'entoure. L'expérience suivante tend à le démontrer.

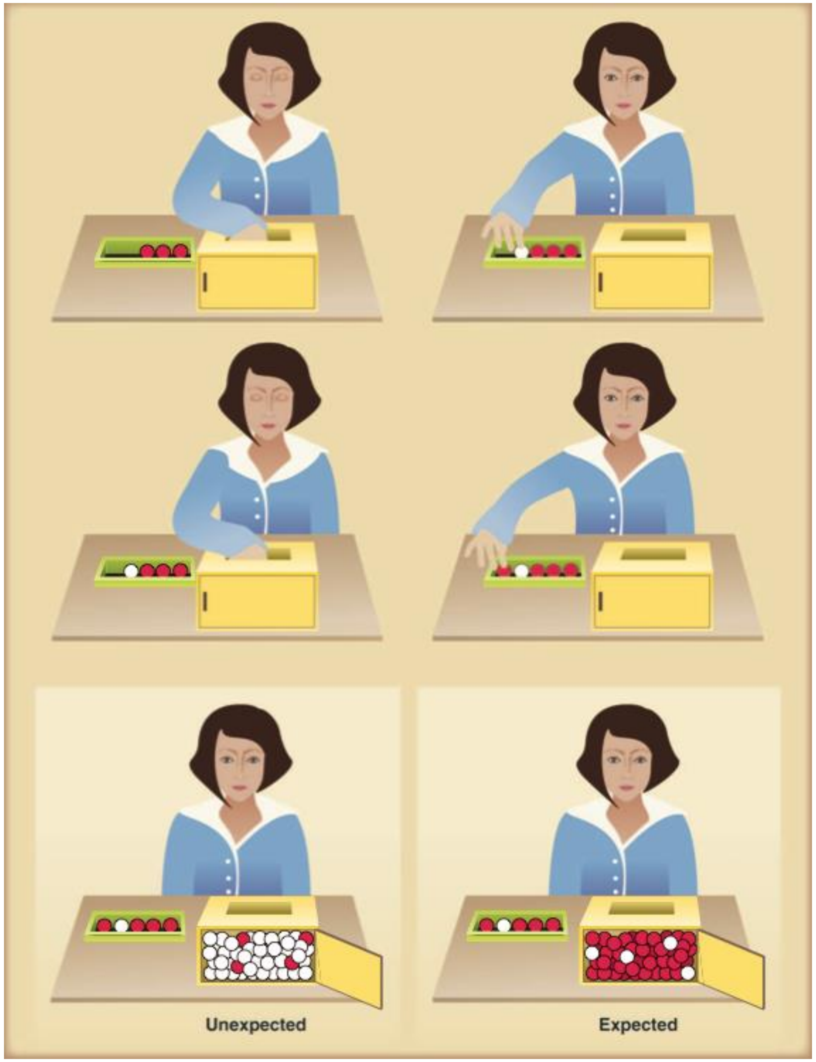
ENJEUX

L'expérience [1] porte sur des enfants de 8 mois.

On présente une boite opaque à un enfant, et on tire devant lui l'une après l'autre des balles dans cette boite. Les balles peuvent être blanches ou rouges. Un trucage permet de faire en sorte que la plupart des balles sorties sont rouges, et seulement quelques unes sont blanches.

Une fois ceci réalisé, on ouvre la boite, on présente son contenu à l'enfant, et on observe son degré de surprise (il faut savoir que « le degré de surprise » est quantifié à partir du temps de fixation du regard de l'enfant : plus il est élevé, plus on considère que l'enfant est surpris).

Si le contenu est conforme à l'échantillon (beaucoup de rouge, peu de blanc), on constate que l'enfant n'est pas surpris.
Mais si le contenu est en contradiction avec l'échantillon (beaucoup de blanc, peu de rouge), l'enfant manifeste un long temps de fixation (il est surpris, donc).



[1] *: Revue Science, Alison Gopnik, Scientific Thinking in Young Children: Theoretical Advances, Empirical Research, and Policy Implications, Science 337, 1623 (2012);*

ENJEUX

L'expérience démontre qu'il avait réussi à émettre une hypothèse sur le contenu de la boite à partir de l'échantillon qu'on lui avait présenté.
D'autres expériences de ce genre ont permis de conforter l'idée que dès le plus jeune âge, les enfants sont capables de réaliser des **inductions bayésiennes**.

Et les adultes ?

ENJEUX

Un problème classique
et malheureusement d'actualité

ENJEUX

Une maladie affecte une personne sur 10000.
Un test de dépistage est fiable à 99 % (nous ne distinguerons pas dans cette exemple les notions de sensibilité et de spécificité – voir plus loin exemples d'activité) c'est-à-dire que :

- 99 % des malades testés sont positifs ;
- 99 % des non malades testés sont négatifs.

Quelle est la probabilité d'être malade si le test est positif?

Voir aussi vidéo dans lien diapo 36 *(argument frappant)*

ENJEUX

Raisonnement statistique :
Sur une population de un million d'individus on obtient le tableau suivant :
M : « l'individu est malade »
T : « le test est positif »

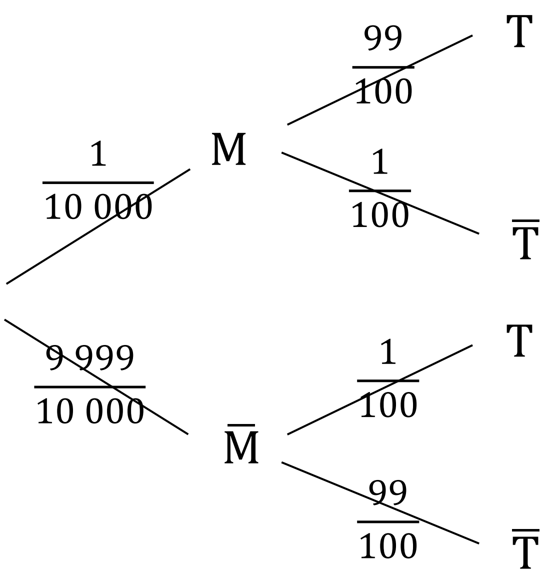
	M	$\overline{M}$	tot
T	99	9 999	10 098
$\overline{T}$	1	989 901	989 902
tot	100	999 900	1 000 000

$$P_T(M) = 0{,}0098 \approx 0{,}001$$

Ne pas oublier l'importance des statistiques et du traitement statistique dans les programmes de lycée : de la seconde à la terminale …

ENJEUX

Raisonnement probabilistique :



Avec la formule de Bayes :

$$P_T(M) = \frac{P_M(T)P(M)}{P(T)}$$

$$P_T(M) = \frac{\frac{1}{10000} \times \frac{99}{100}}{\frac{1}{10000} \times \frac{99}{100} + \frac{9999}{10000} \times \frac{1}{100}}$$

$$P_T(M) = \frac{99 \times 10^{-6}}{0{,}10098}$$

$$P_T(M) = 0{,}0098 \approx 0{,}001$$

… en parallèle des probabilités.

ENJEUX

Soit une probabilité avant test de $\frac{1}{10000}$, **probabilité à priori,** qui passe après test à $\frac{1}{1000}$, **probabilité à postériori**.

Et qui contredit généralement toute intuition.

ENJEUX

Point de vue pédagogique

ENJEUX

Enjeu double pour la compétence « raisonner » :

- Éléments de logique : implication, équivalence, réciproque …

- Éléments de statistique et probabilités :

Quel degré de confiance accordé à telles ou telles observations, propositions ? Problème de la vraisemblance.

Question cruciale du programme puisqu’elle est revient à plusieurs reprises :

- influence bayésienne mais aussi :
- tests statistiques,
- interprétation des notions d’espérance, de variance et d’écart-type,
- déterminer un intervalle I pour lequel la probabilité $P(X \in I)$ est inférieure à une valeur donnée α, ou supérieure à $1 - \alpha$.

ENJEUX

Comme dans la partie « analyse » (voir zoom sur progression), dans la partie « Probabilités et statistique » le programme tend à :

- créer des liens autour d’enjeux, de questions génératrices (approche thématique) ;

- apporter une diversité de réponses, dans des cadres et registres divers (approche par contenu).

QUELQUES EXEMPLES D'ACTIVITE

Questions flash

QUELQUES EXEMPLES D'ACTIVITE

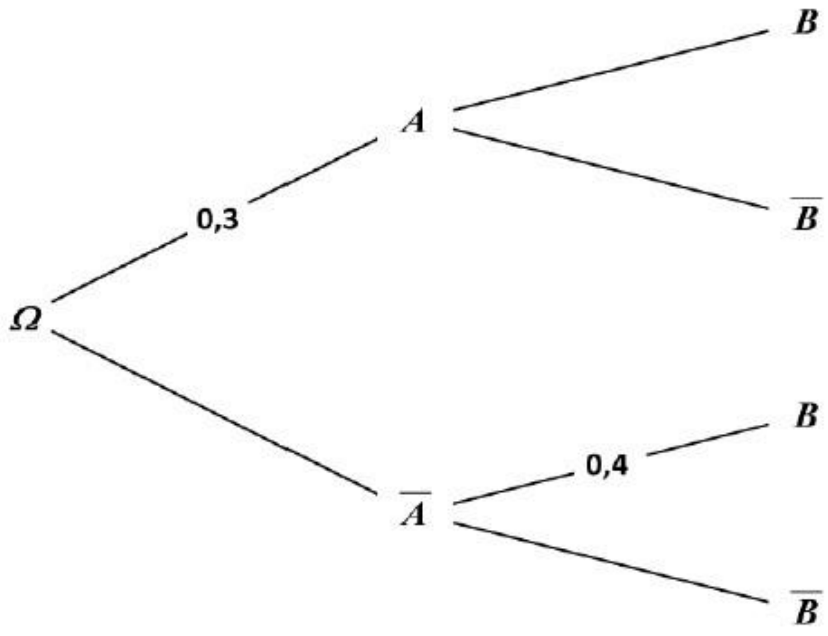
Exemple 1

En s'appuyant sur l'arbre pondéré ci-dessous et sachant que :

$$P(A \cap B) = 0{,}18$$

quelle est la valeur de $P_A(B)$?

a) 0,82 b) 0,6 c) 0,4



QUELQUES EXEMPLES D'ACTIVITE

Exemple 2

Lors d'un examen de deuxième année, 25% des étudiants échouent en biochimie, 15% en anatomie et 10% à la fois en anatomie et en biochimie. Si un élève a échoué en anatomie, quelle est la probabilité qu'il n'échoue pas en biochimie ?

a) $\frac{1}{3}$ b) $\frac{1}{4}$ c) 0,75 d) $\frac{2}{3}$ e) 0,15

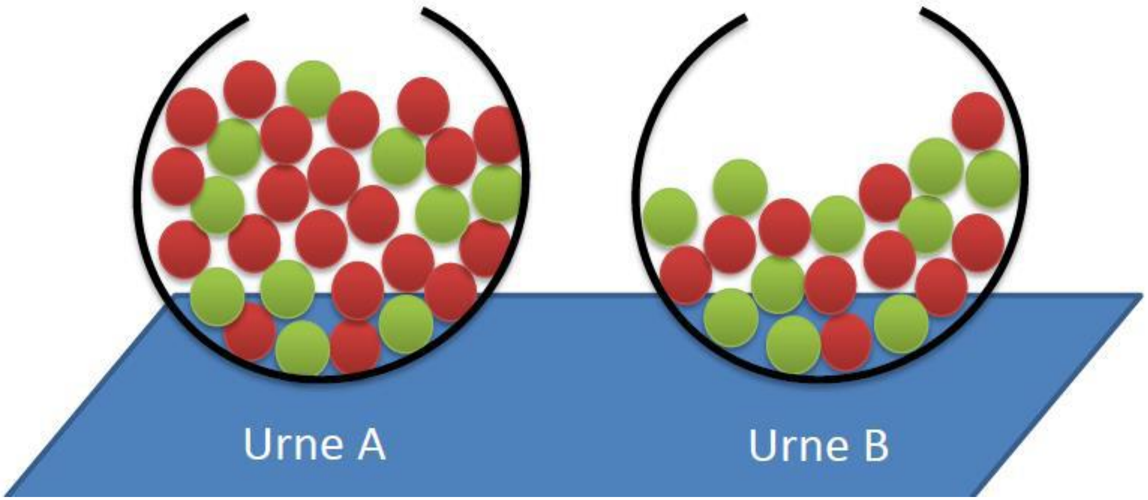
D'après médecine Lille 2007

QUELQUES EXEMPLES D'ACTIVITE

Des classiques

QUELQUES EXEMPLES D'ACTIVITE

« DE QUELLE URNE VIENT LA BOULE ? »



DEUX URNES A ET B SONT REMPLIES DE BOULES INDISCERNABLES AU TOUCHER.
L'URNE A CONTIENT DIX BOULES VERTES ET VINGT BOULES ROUGES.
L'URNE B CONTIENT DIX BOULES VERTES ET DIX BOULES ROUGES.
ALEX CHOISIT AU HASARD UNE DES DEUX URNES.
DANS L'URNE CHOISIE, IL TIRE ALORS UNE BOULE AU HASARD. ELLE EST ROUGE.
QUELLE EST LA PROBABILITÉ QU'ALEX AIT TIRÉ CETTE BOULE DANS L'URNE A SACHANT QU'ELLE EST ROUGE ?

QUELQUES EXEMPLES D'ACTIVITE

Retour sur le test en médecine : précision de vocabulaire.

L'efficacité d'un test peut se mesurer à l'aide de deux grandeurs : **la sensibilité et la spécificité.**

Considérons une étude de validité d'un nouveau test de dépistage d'une maladie, effectuée auprès de 200 personnes : 115 personnes malades et 85 personnes qui ne sont pas malade.
Les résultats de cette étude sont consignés dans le tableau ci-dessous :

M : « l'individu est malade »
T : « le test est positif »

	M	$\overline{M}$	tot
T	90	10	100
$\overline{T}$	25	75	100
tot	115	85	200

« faux positifs »

QUELQUES EXEMPLES D'ACTIVITE

La sensibilité est la probabilité qu'un test réalisé sur une personne malade soit positif, soit $P_M(T)$.

Ici $P_M(T) = \frac{90}{115} \approx 0,78$

La spécificité est la probabilité qu'un test réalisé sur une personne qui n'est pas malade soit négatif, soit $P_{\overline{M}}(\overline{T})$.

Ici $P_{\overline{M}}(\overline{T}) = \frac{75}{85} \approx 0,88$

QUELQUES EXEMPLES D'ACTIVITE

La valeur prédictive positive est la probabilité qu'un patient dont le test réalisé est positif soit réellement malade, soit $P_T(M)$.

Ici $P_T(M) = \frac{90}{100} \approx 0{,}90$

La valeur prédictive négative est la probabilité qu'un patient dont le test réalisé est négatif ne soit pas malade**,** , soit $P_{\bar{T}}(\bar{M})$.

Ici $P_{\bar{T}}(\bar{M}) = \frac{75}{100} \approx 0{,}75$

C'est une occasion de vérifier les formules de Bayes:

$$P_T(M) = \frac{P_M(T)P(M)}{P(T)} = \frac{\frac{90}{115} \times \frac{115}{200}}{\frac{100}{200}} \approx 0{,}90$$

$$P_{\bar{T}}(\bar{M}) = \frac{P_{\bar{M}}(\bar{T})P(\bar{M})}{P(\bar{T})} = \frac{\frac{75}{85} \times \frac{85}{200}}{\frac{100}{200}} \approx 0{,}75$$

QUELQUES EXEMPLES D'ACTIVITE

Mais la sensibilité et la spécificité ne sont pas des informations utiles directement au patient qui se demande non pas :
« si je suis malade, quelle est la probabilité que le test soit positif ? »
Mais bien :
« si mon test est positif, quelle est la probabilité que je sois malade ? »

Soit $T \Rightarrow M$ et non $M \Rightarrow T$: lien avec les éléments de logique.

Pour répondre à la question du patient, comme nous l'avons vu précédemment (diapo 17), une donnée est indispensable :
la prévalence de la maladie, c'est-à-dire la fréquence de la maladie dans la population.
Ici le choix du nombre de malades et de non malades a été fait par les chercheurs et ne tient pas forcément compte de cette prévalence, d'autres raisons peuvent intervenir, et surtout ce n'est pas l'objectif recherché.

QUELQUES EXEMPLES D'ACTIVITE

Le problème de la diapo 18 peut donc être reformulé et traité avec ces nouvelles informations.

« Une maladie affecte une personne sur 10 000.
Un test de dépistage de sensibilité 95 % et de spécificité 99 % est mis au point.
Quelle est la probabilité d'être malade si le test est positif ? »

QUELQUES EXEMPLES D’ACTIVITE

Pour diversifier les supports et les thèmes

QUELQUES EXEMPLES D'ACTIVITE

https://padlet.com/mathspcwallon/vjpwhnmsv968

Un lien vers une page où sont explorés entre autres :

- le paradoxe de Monty Hall avec une programmation Python permettant des simulations,
- un cas en justice : l'affaire Sally Clarke,
- deux vidéos de « Monsieur Phi », sur le thème de l'inférence Bayésienne, reprenant l'exemple de la diapo 18 …

QUELQUES EXEMPLES D'ACTIVITE

En informatique : gestion des spams

Principe :
À partir d'observations faites sur le contenu d'un message (expéditeur, mots employés, présence de liens …) un algorithme met à jour son estimation de la probabilité qu'un message soit un spam.
En fonction de la valeur de cette probabilité à postériori, il peut décider de classer ou non le message comme spam.

Exemple :
On considère qu'un message sur deux reçu dans la boîte mail d'un certain utilisateur est une promotion.
On remarque que la fréquence d'apparition du mot « soldes » dans un message classique est de $\frac{1}{100\ 000}$ et dans un spam cette fréquence est dix fois supérieure.

1. Quand un message arrive, quelle est la probabilité qu'il soit une promotion ?
2. Si l'algorithme de la boîte de messagerie détecte alors la présence du mot « soldes » dans le message, quelle est alors la probabilité que ce message soit une promotion ?

QUELQUES EXEMPLES D'ACTIVITE

En informatique : gestion des spams

- **Lien avec l'enseignement scientifique** : intelligence artificielle, apprentissage machine (ou apprentissage automatique).
  Si en enseignement scientifique, conformément au programme, « *il est possible de raisonner sur des tableaux à double entrée sans faire appel explicitement à la théorie des probabilités conditionnelles ni à la formule de Bayes* ». (extrait du BO)
  En enseignement optionnel de mathématiques complémentaires, on pourra s'attacher à explorer, plus particulièrement et en complément, le « raisonnement bayésien ».