Démystifier la statistique : de la magie mathématique au sens commun

JOHAN SEGERS

Institut de statistique, biostatistique et sciences actuarielles

Université catholique de Louvain

johan.segers@uclouvain.be

CFIES'2010, Bruxelles, du 8 au 10 septembre 2010

Q Dans une famille, il y a deux enfants. L'ainé est garçon. Quelle est la probabilité que le deuxième soit garçon ?

R 1/2

Q Dans une famille, il y a deux enfants. Au moins un des deux est garçon. Quelle est la probabilité qu'il y ait deux garçons ?

R 1/3

Q Dans une famille, il y a deux enfants. Un des deux s'appelle Marcel. Quelle est la probabilité qu'il y ait deux garçons ?

R **?**

Q Dans une famille, il y a deux enfants. L'ainé est garçon. Quelle est la probabilité que le deuxième soit garçon ?

R 1/2

Q Dans une famille, il y a deux enfants. Au moins un des deux est garçon. Quelle est la probabilité qu'il y ait deux garçons ?

R 1/3

Q Dans une famille, il y a deux enfants. Un des deux s'appelle Marcel. Quelle est la probabilité qu'il y ait deux garçons ?

R **?**

Q Dans une famille, il y a deux enfants. L'ainé est garçon. Quelle est la probabilité que le deuxième soit garçon ?

R 1/2

Q Dans une famille, il y a deux enfants. Au moins un des deux est garçon. Quelle est la probabilité qu'il y ait deux garçons ?

R 1/3

Q Dans une famille, il y a deux enfants. Un des deux s'appelle Marcel. Quelle est la probabilité qu'il y ait deux garçons ?

R **?**

Q Dans une famille, il y a deux enfants. L'ainé est garçon. Quelle est la probabilité que le deuxième soit garçon ?

R 1/2

Q Dans une famille, il y a deux enfants. Au moins un des deux est garçon. Quelle est la probabilité qu'il y ait deux garçons ?

R 1/3

Q Dans une famille, il y a deux enfants. Un des deux s'appelle Marcel. Quelle est la probabilité qu'il y ait deux garçons ?

R **?**

Q Dans une famille, il y a deux enfants. L'ainé est garçon. Quelle est la probabilité que le deuxième soit garçon ?

R 1/2

Q Dans une famille, il y a deux enfants. Au moins un des deux est garçon. Quelle est la probabilité qu'il y ait deux garçons ?

R 1/3

Q Dans une famille, il y a deux enfants. Un des deux s'appelle Marcel. Quelle est la probabilité qu'il y ait deux garçons ?

R ?

Q Dans une famille, il y a deux enfants. L'ainé est garçon. Quelle est la probabilité que le deuxième soit garçon ?

R 1/2

Q Dans une famille, il y a deux enfants. Au moins un des deux est garçon. Quelle est la probabilité qu'il y ait deux garçons ?

R 1/3

Q Dans une famille, il y a deux enfants. Un des deux s'appelle Marcel. Quelle est la probabilité qu'il y ait deux garçons ?

R **?**

- Notre intuition probabiliste semble plutôt faible
- Les mathématiques sont apperçues comme une *magie occulte*

☞ Il faut des outils pour aider les étudiants à intérioriser les concepts

Plan

Outline

Obstacles à la compréhension

- Pauvre intuition de phénomènes aléatoires
- Base faible en mathématiques

En comparaison avec l'arithmétique et la géométrie, notre intuition en probabilité est faible

"Children count and calculate with quantities in their everyday life. Furthermore, a check of success is easy with natural numbers where the calculation of 2+2 yields the immediately testable result of 4.

Such feedback is complicated with probabilities : the extremely small probability of winning in a state lottery is counterveiled by the fact that people win every week." — BOROVCNIK & BENTZ (1991)

En comparaison avec l'arithmétique et la géométrie, notre intuition en probabilité est faible

> *"Children count and calculate with quantities in their everyday life. Furthermore, a check of success is easy with natural numbers where the calculation of 2+2 yields the immediately testable result of 4.*
>
> *Such feedback is complicated with probabilities : the extremely small probability of winning in a state lottery is counterveiled by the fact that people win every week."* — BOROVCNIK & BENTZ (1991)

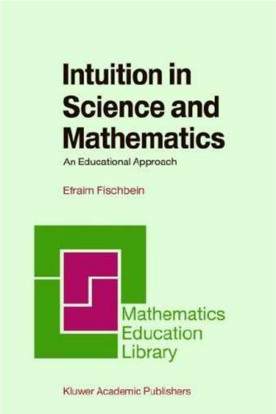
En comparaison avec l'arithmétique et la géométrie, notre intuition en probabilité est faible

"Why has the theory of probability failed to influence mathematics to the same extent that geometry has ?

Because we possess a natural geometric intuition but no probability intuition." — A. ENGEL (1970)

En comparaison avec l'arithmétique et la géométrie, notre intuition en probabilité est faible

> *"Why has the theory of probability failed to influence mathematics to the same extent that geometry has ?*
> *Because we possess a natural geometric intuition but no probability intuition."* — A. ENGEL (1970)



EBRAIM FISCHBEIN (1975)
The Intuitive Sources of Probabilistic Thinking in Children.

Quelques idées fausses typiques

Effet de la taille de l'échantillon

La probabilité d'avoir au moins 2 fois « pile » après 3 lancés d'une pièce est :

- ☐ inférieure à
- ☐ égale à
- ☐ supérieure à

à la probabilité d'avoir au moins 200 fois « pile » après 300 lancés.

Quelques idées fausses typiques

Effet de la taille de l'échantillon

La probabilité d'avoir au moins 2 fois « pile » après 3 lancés d'une pièce est :

- ☐ inférieure à
- ☠ égale à
- ☐ supérieure à

à la probabilité d'avoir au moins 200 fois « pile » après 300 lancés.

Quelques idées fausses typiques

Effet de la taille de l'échantillon

La probabilité d'avoir au moins 2 fois « pile » après 3 lancés d'une pièce est :

- ☐ inférieure à
- ☠ égale à
- ☑ supérieure à

à la probabilité d'avoir au moins 200 fois « pile » après 300 lancés.

Quelques idées fausses typiques

L'effet de l'axe du temps (le phénomène de FALK)

Une boîte contient 3 billes, 2 blanches et 1 noire. Jean en tire une et il la mise de côté sans la regard. Après il tire une 2^{e} et il voit qu'elle est blanche.
La probabilité que la 1re bille soit blanche, est-elle

- ☐ inférieure à
- ☐ égale à
- ☐ supérieure à

à la probabilité qu'elle soit noire ?

Quelques idées fausses typiques

L'effet de l'axe du temps (le phénomène de FALK)

Une boîte contient 3 billes, 2 blanches et 1 noire. Jean en tire une et il la mise de côté sans la regard. Après il tire une 2^{e} et il voit qu'elle est blanche.
La probabilité que la 1^{re} bille soit blanche, est-elle

- ☐ inférieure à
- ☐ égale à
- ☠ supérieure à

à la probabilité qu'elle soit noire ?

Quelques idées fausses typiques

L'effet de l'axe du temps (le phénomène de FALK)

Une boîte contient 3 billes, 2 blanches et 1 noire. Jean en tire une et il la mise de côté sans la regard. Après il tire une 2^{e} et il voit qu'elle est blanche.
La probabilité que la 1re bille soit blanche, est-elle

- ☐ inférieure à
- ☑ égale à
- ☠ supérieure à

à la probabilité qu'elle soit noire ?

Techniques pour construire l'intuition probabiliste

- Présentation : arbres, diagrammes
- Arithmétique : fréquences
- Géométrie : aire, volume
- Analogies
- Expériences : Simulations

  ↯ Par contre, enterrer les concepts dans un formalisme mathématique n'aide pas à la compréhension !

Techniques pour construire l'intuition probabiliste

- Présentation : arbres, diagrammes
- Arithmétique : fréquences
- Géométrie : aire, volume
- Analogies
- Expériences : Simulations

  ϟ Par contre, enterrer les concepts dans un formalisme mathématique n'aide pas à la compréhension !

Outline

L'illusion de la linéarité (1)

GIROLAMO CARDANO
(1501–1576)

P(⚅⚅) $= 1/36$

P(au moins 1 fois ⚅⚅
après 18 lancés de deux dés)

☠
$\neq 18/36 = 1/2$

VAN DOOREN ET AL. (2003)

L'illusion de la linéarité (2)

BLAISE PASCAL
(1623–1662)

ANTOINE DE GOMBAUD, le Chevalier de Méré (1607–1684) :

P(au moins 1 ⚅ après 4 lancés)

☠
$\neq P$(au moins 1 ⚅⚅ après 24 lancés)

Non-linéarité des probabilités : loi additive

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Démonstration :

$$\begin{aligned} P(A \cup B) &= P(A \setminus B) + P(A \cap B) + P(B \setminus A) \\ &= \underbrace{P(A \setminus B) + P(A \cap B)}_{=P(A)} + \underbrace{P(B \setminus A) + P(A \cap B)}_{=P(B)} - P(A \cap B) \end{aligned}$$

Jongler avec des identités, peu convainquant. . .

Application mécanique d'une formule occulte

$$P(\text{au moins 1 fois « pile » après 2 lancés d'une pièce}) = P(A \cup B)$$

$$A = \{\text{« pile » au 1}^{\text{er}}\text{ lancé}\}$$
$$B = \{\text{« pile » au 2}^{\text{e}}\text{ lancé}\}$$

Grâce à la loi additive :

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 1/2 + 1/2 - 1/4 = 3/4 \end{aligned}$$

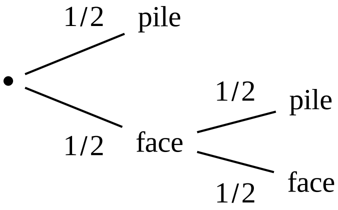
Représentation de l'espace d'échantillonnage au moyen d'un tableau

	B	
A	PP	PF
	FP	FF

Calcul :

$$P(A \cup B) = \frac{2+2-1}{4} = \frac{3}{4}$$

Représentation de l'espace d'échantillonnage au moyen d'un arbre



Calcul :

$$1/2 + (1/2 \times 1/2) = 3/4$$

Outline

Probabilité conditionnelle : changement de classe de référence

THOMAS BAYES (1702–1761)

"Essay Towards Solving a Problem in the Doctrine of Chances"

$$P(\text{malade} \mid \text{test positif})$$

$$\neq P(\text{test positif} \mid \text{malade})$$

La solution offerte par le pasteur BAYES est ingénieuse mais peu intuitive

Théorème de BAYES

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

Démonstration simple. Toutefois :

- Changement de classe de référence (de A à B)
- Facile à se tromper dans la formule
- Conclusion peu intuitive

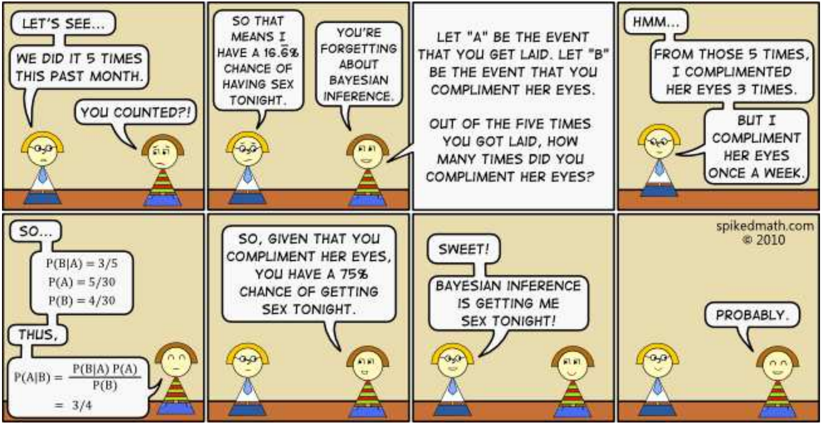
La solution offerte par le pasteur BAYES est ingénieuse mais peu intuitive

Théorème de BAYES

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

Démonstration simple. Toutefois :

- Changement de classe de référence (de A à B)
- Facile à se tromper dans la formule
- Conclusion peu intuitive



`http://spikedmath.com/226.html`

$$A = \{\text{getting laid}\}$$
$$B = \{\text{complimenting her eyes}\}$$

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

Approche classique

$$A = \{\text{getting laid}\}, \qquad P(A) = 5/30$$

$$B = \{\text{complimenting her eyes}\}, \qquad P(B) = 4/30$$

$$P(B \mid A) = 3/5 \qquad \Longrightarrow \qquad P(A \mid B) = \ ?$$

Grâce au théorème de BAYES :

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} = \frac{(3/5) \times (5/30)}{4/30} = 3/4$$

Approche classique

$$A = \{\text{getting laid}\}, \qquad P(A) = 5/30$$
$$B = \{\text{complimenting her eyes}\}, \qquad P(B) = 4/30$$

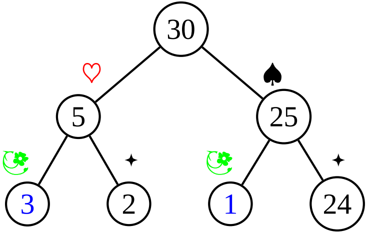
$$P(B \mid A) = 3/5 \qquad \Longrightarrow \qquad P(A \mid B) = \ ?$$

Grâce au théorème de BAYES :

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} = \frac{(3/5) \times (5/30)}{4/30} = 3/4$$

Autre répresentation : fréquences naturelles

$A = \heartsuit$, $B =$ ❦



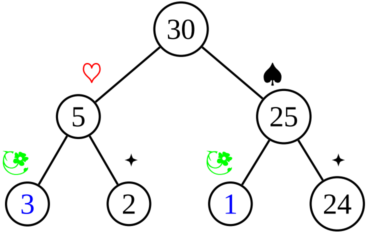
Arithmétique élémentaire :

$$P(\heartsuit \mid \text{❦}) = \frac{3}{3+1} = \frac{3}{4}$$

La classe de référence (30) reste fixée tout au long du calcul.

Autre répresentation : fréquences naturelles

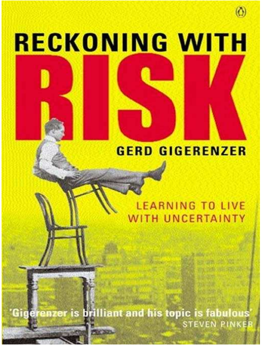
$A = \heartsuit$, $B =$ ✿



Arithmétique élémentaire :

$$P(\heartsuit \mid ✿) = \frac{3}{3+1} = \frac{3}{4}$$

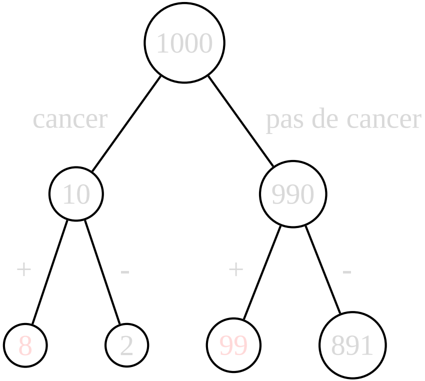
La classe de référence (30) reste fixée tout au long du calcul.



GERD GIGERENZER (1947)

Aider les médecins à comprendre les implications d'un test positif

$$\left.\begin{array}{lcr} P(\text{cancer}) & = & 1\% \\ P(\text{test} + \mid \text{cancer}) & = & 80\% \\ P(\text{test} + \mid \text{pas de cancer}) & = & 10\% \end{array}\right\} \Rightarrow P(\text{cancer} \mid \text{test} +) = \ ?$$

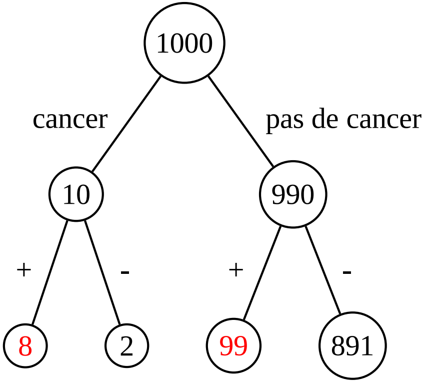


$$P(\text{cancer} \mid \text{test} +) = \frac{8}{8+99} = 7.5\%$$

GIGERENZER & EDWARDS (2003)

Aider les médecins à comprendre les implications d'un test positif

$$\left.\begin{array}{lcr} P(\text{cancer}) & = & 1\% \\ P(\text{test} + \mid \text{cancer}) & = & 80\% \\ P(\text{test} + \mid \text{pas de cancer}) & = & 10\% \end{array}\right\} \Rightarrow P(\text{cancer} \mid \text{test} +) = \ ?$$



$$P(\text{cancer} \mid \text{test} +)$$
$$= \frac{8}{8+99} = 7.5\%$$

GIGERENZER & EDWARDS (2003)

Outline

Approche classique : théorème fondamental de l'analyse

Fonction de répartition F et fonction de densité f :

$$f(x) = \frac{d}{dx} F(x)$$

$$\int_{-\infty}^{x} f(y)\, dy = F(x)$$

- ϟ Calcul de dérivées et d'intégrales :
  apperçue comme une collection de formules à apprendre par cœur

Faute typique

$$f(x) \overset{☠}{\neq} P(X = x) = 0$$

Approche classique : théorème fondamental de l’analyse

Fonction de répartition F et fonction de densité f :

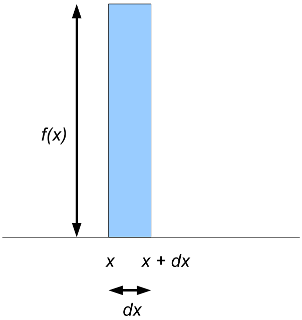
$$f(x) = \frac{d}{dx} F(x)$$
$$\int_{-\infty}^{x} f(y)\, dy = F(x)$$

- ↯ Calcul de dérivées et d’intégrales :
  apperçue comme une collection de formules à apprendre par cœur

Faute typique

$$f(x) \overset{☠}{\neq} P(X = x) = 0$$

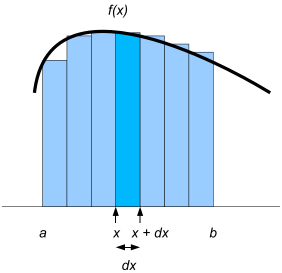
Géométrie et calcul infinitésimal



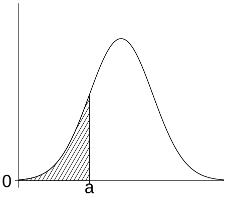
$$P(x \leqslant X \leqslant x + dx) \approx f(x)\,dx$$

$$dx = \text{difference en } x$$

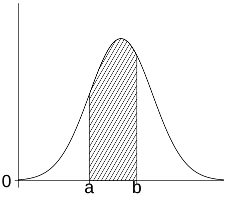
Une probabilité est une aire est une intégrale



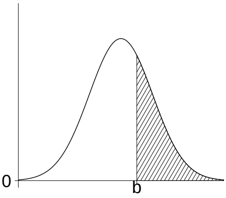
$$P(a \leqslant X \leqslant b) = \text{Somme pour } a \leqslant x \leqslant b \text{ des } f(x)\,dx = \int_a^b f(x)\,dx$$



$$P(X \leqslant a) = \int_{-\infty}^{a} f(x)\, dx$$

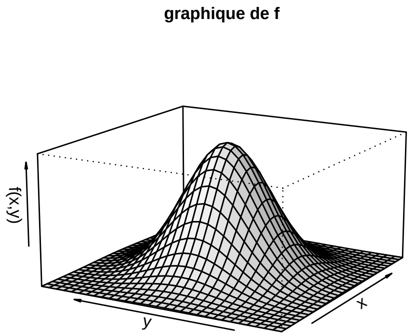


$$P(a \leqslant X \leqslant b) = \int_{a}^{b} f(x)\, dx$$



$$P(X \geqslant b) = \int_{b}^{\infty} f(x)\, dx$$

La même intuition géométrique s'applique au cas multivarié

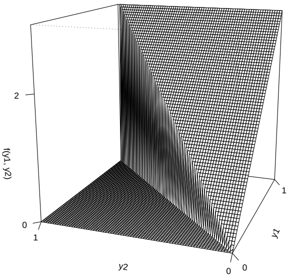


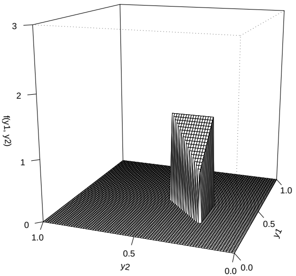
$f(x,y) =$ hauteur d'un pavé droit de base $dx \times dy$ émanant du point (x,y)

La même intuition géométrique s'applique au cas multivarié

univarié	*bivarié*
variable aléatoire X	variables aléatoires X, Y
▸ densité $f : \mathbb{R} \to [0, +\infty[$	▸ densité $f : \mathbb{R}^2 \to [0, +\infty[$
▸ Pour $A \subset \mathbb{R}$, $$\begin{aligned} & P(X \in A) \\ = \; & \int_{x \in A} f(x)\, dx \end{aligned}$$	▸ Pour $A \subset \mathbb{R}^2$, $$\begin{aligned} & P[(X, Y) \in A] \\ = \; & \iint_{(x,y) \in A} f(x, y)\, dx\, dy \end{aligned}$$
▸ l'aire au-dessus de A et en dessous de f	▸ le volume au-dessus de A et en dessous de f

Une probabilité est un volume est une double intégrale

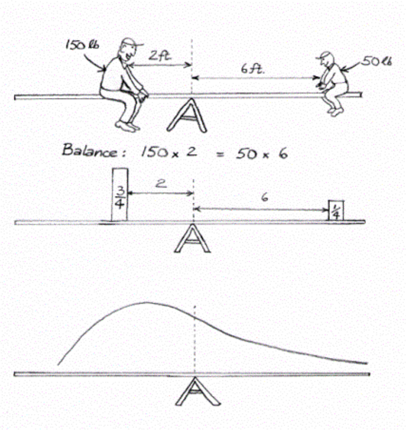




$$f(x,y) = \begin{cases} 3x & \text{si } 0 \leqslant y \leqslant x \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

$$A = \{(x,y) : 1/4 \leqslant y \leqslant x \leqslant 1/2\}$$

Une espérance est comme le point d'équilibre d'une bascule



$$E(X) = \begin{cases} \sum x\, p(x) \\ \int x\, f(x)\, dx \end{cases}$$

$$= \text{somme}(\text{distance} \times \text{poids})$$

MARTIN (2003) JSE

Outline

Quid de Marcel ?

Dans une famille, il y a deux enfants. Un des deux s'appelle Marcel. Quelle est la probabilité qu'il y ait deux garçons ?

Nombre de familles à deux enfants

		2e enfant Marcel	garçon, autre	fille
	Marcel	m	$n-m$	n
1er enfant	garçon, autre	$n-m$	...	...
	fille	n	...	...

$$m \ll n \Rightarrow P(\text{deux garçons} \mid \text{Marcel}) = \frac{2n-m}{4n-m} \approx \frac{2n}{4n} = \frac{1}{2}$$

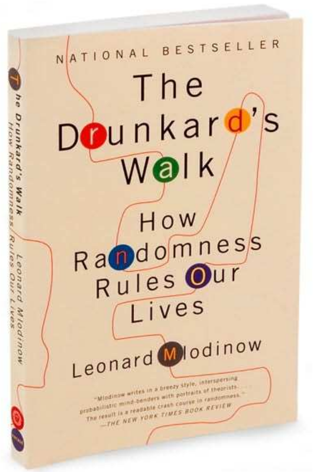
Quid de Marcel ?

Dans une famille, il y a deux enfants. Un des deux s'appelle Marcel. Quelle est la probabilité qu'il y ait deux garçons ?

Nombre de familles à deux enfants

		2e enfant		
		Marcel	garçon, autre	fille
	Marcel	m	$n-m$	n
1er enfant	garçon, autre	$n-m$	$\ldots$	$\ldots$
	fille	n	$\ldots$	$\ldots$

$$m \ll n \Rightarrow P(\text{deux garçons} \mid \text{Marcel}) = \frac{2n-m}{4n-m} \approx \frac{2n}{4n} = \frac{1}{2}$$



LEONARD MLODINOW (2008)

Conclusion

- Notre intuition probabiliste semble plutôt faible
- La magie mathématique est apperçue comme un *deus ex machina*
- Outils pour aider les étudiants à intérioriser les concepts :
  - Présentation : arbres, diagrammes
  - Arithmétique : fréquences
  - Géométrie : aire, volume
  - Analogies
  - Expériences : Simulations

`www.uclouvain.be/isba`

Bibliography (1)

BOROVCNIK, M. & BENTZ, H.-J. (1991)
Empirical research in understanding probability
Chance Encounters : Probability in Education, 73–105.

ENGEL, E. (1970)
Teaching probability in intermediate grades
The Teaching of Probability and Statistics, 97–150.

FALK, R. (1979)
Revision of probabilities and the time axis
Proc. of the Third Int. Conf. for the Psych. of Math. Education, 64–66.

FISCHBEIN, E. & SCHNARCH, D. (1997)
The Evolution with Age of Probabilistic, Intuitively Based Misconceptions
Journal for Research in Mathematics Education 28, 96–105.

GIGERENZER, G. & EDWARDS, A. (2003)
Simple tools for understanding risks : from innumeracy to insight
British Medical Journal 327, 741–744.

Bibliography (2)

GREER, B. (2001)
Understanding probabilisic thinking : The legacy of EFRAIM FISCHBEIN.
Educational Studies in Mathematics 45, 15–33.

MARTIN, M. A. (2003)
"It's like... you know" : The Use of Analogies and Heuristics in Teaching Introductory Statistical Methods
Journal of Statistics Education 11, Number 2.

VAN DOOREN, W., DE BOCK, D., DEPAEPE, F., JANSSENS, D. & VERSCHAFFEL, L. (2003)
The illusion of linearity : Expanding the evidence towards probabilistic reasoning
Educational Studies in Mathematics 53, 113–138.