

# Indice Probabiliste d'Ecart à l'Equilibre pour l'évaluation de la qualité des règles

*Julien BLANCHARD, Fabrice GUILLET, Henri BRIAND, Régis GRAS*

Ecole Polytechnique de l'Université de Nantes  
LINA, Laboratoire d'Informatique de Nantes Atlantique

- Extraction de Connaissances dans les bases de Données
- Découverte de règles d'association  
(huîtres, citron) → (muscadet, pain de seigle)
- Post-traitement des règles:

**Evaluation des règles par des mesures de qualité**

## Deux catégories de mesures de qualité de règles:

- **Indices subjectifs**

→ **orientées décideur**

(Padmanabhan et al. 1998, Liu et al. 1999)

- **Indices objectifs**

→ **orientées données**

(Tan et al. 2004, Guillet 2004,  
Lallich et al. 2004, Lenca et al. 2004)

# Introduction

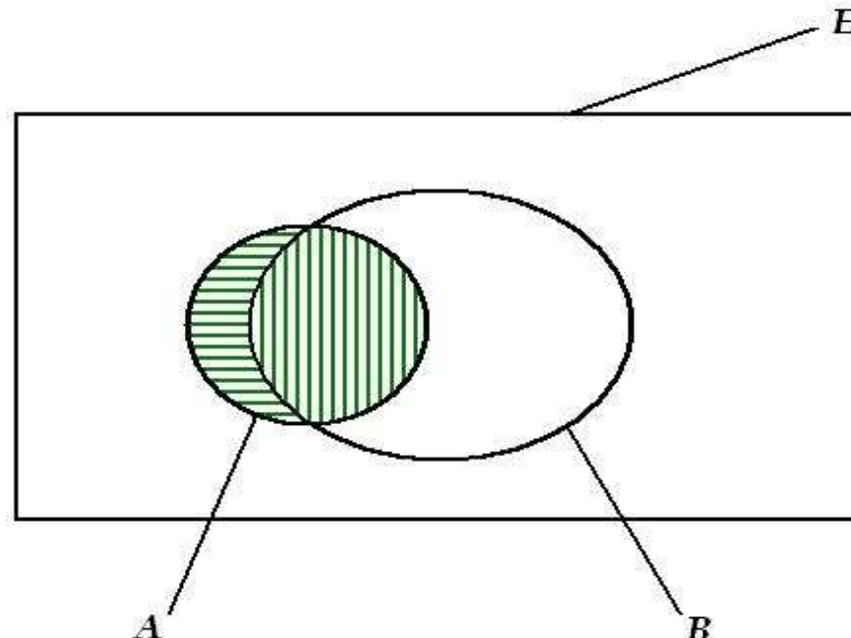
IPEE

Propriétés

Conclusion

**a** 

**b**

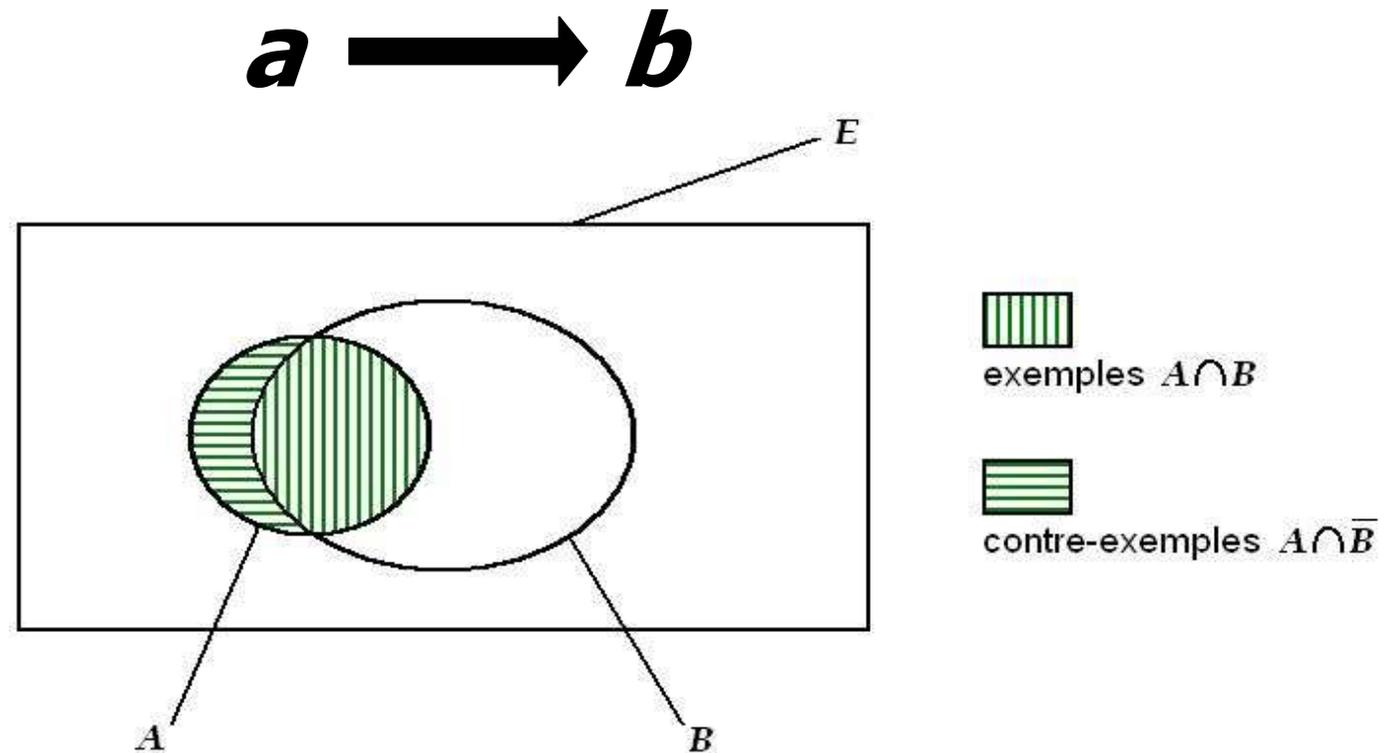


exemples  $A \cap B$



contre-exemples  $A \cap \bar{B}$

	a	non a	
b	$n_{A \cap B}$	$n_{\bar{A} \cap B}$	$n_B$
non b	$n_{A \cap \bar{B}}$	$n_{\bar{A} \cap \bar{B}}$	$n_{\bar{B}}$
	$n_A$	$n_{\bar{A}}$	$n$



## Indices objectifs de qualité de règle :

(Blanchard et al. 2004)

- Écart à l'indépendance

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

- Écart à l'équilibre

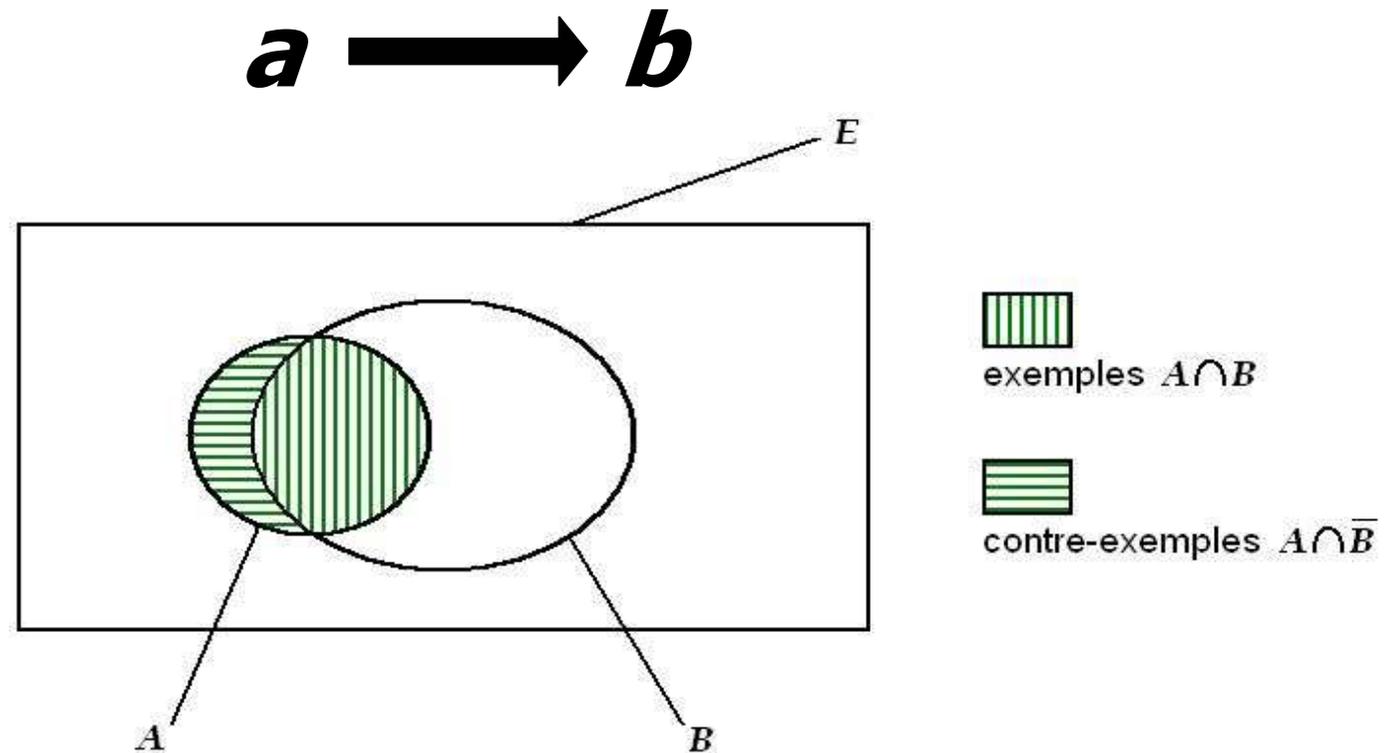
$$P(A \cap B) = P(A) / 2$$

## Introduction

IPEE

Propriétés

Conclusion



## Indices objectifs de qualité de règle :

(Lallich et al. 2004, Gras et al. 2004)

- Indices statistiques
- Indices descriptifs

## Introduction

IPEE

Propriétés

Conclusion

# Indices objectifs de qualité de règle :

	Indice d'écart à l'équilibre	Indice d'écart à l'indépendance
Indice descriptif	<ul style="list-style-type: none"><li>- confiance,</li><li>- indice de Sebag et Schoenauer,</li><li>- taux des exemples et contre-exemples,</li><li>- indice de Ganascia,</li><li>- moindre-contradiction,</li><li>- indice d'inclusion...</li></ul>	<ul style="list-style-type: none"><li>- coefficient de corrélation,</li><li>- indice de Loevinger,</li><li>- lift,</li><li>- conviction,</li><li>- TIC,</li><li>- rapport de cote,</li><li>- multiplicateur de cote...</li></ul>
Indice statistique		<ul style="list-style-type: none"><li>- intensité d'implication,</li><li>- indice d'implication,</li><li>- indice de vraisemblance du lien,</li><li>- contribution orientée au <math>\chi^2</math>,</li><li>- rule-interest...</li></ul>

## Introduction

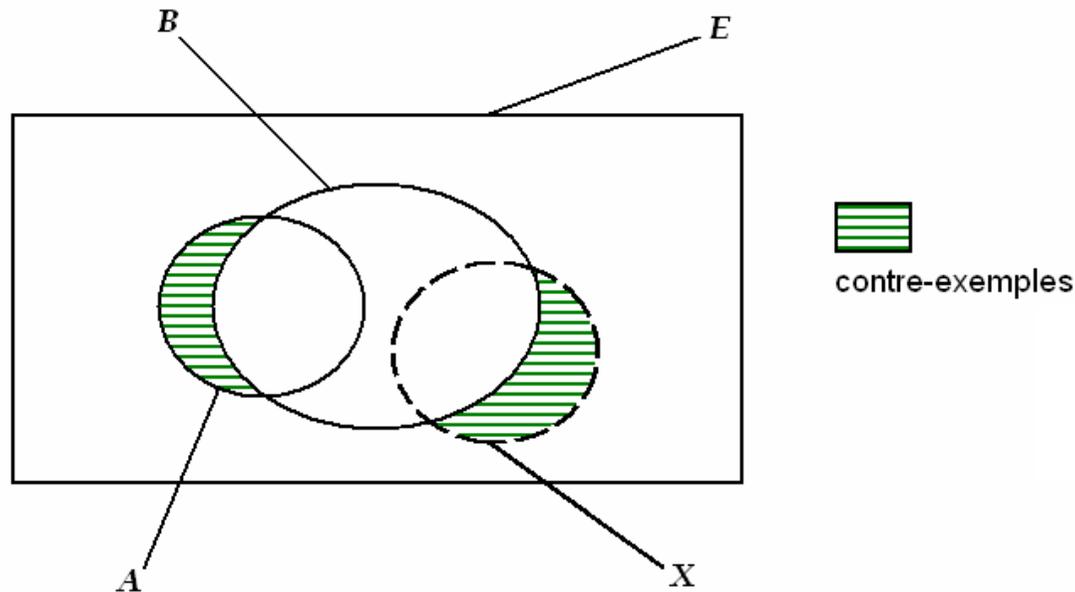
## IPEE

### ➤ Modèle aléatoire

### ➤ Expression analytique

## Propriétés

## Conclusion



Hypothèse  $H_0$  :

équiprobabilité des exemples et des contre-exemples

$$P(A \cap B) = P(A \cap \bar{B})$$

Tirage aléatoire de  $X$  dans  $E$  :

$$P(X \cap B) = P(X \cap \bar{B})$$

## Introduction

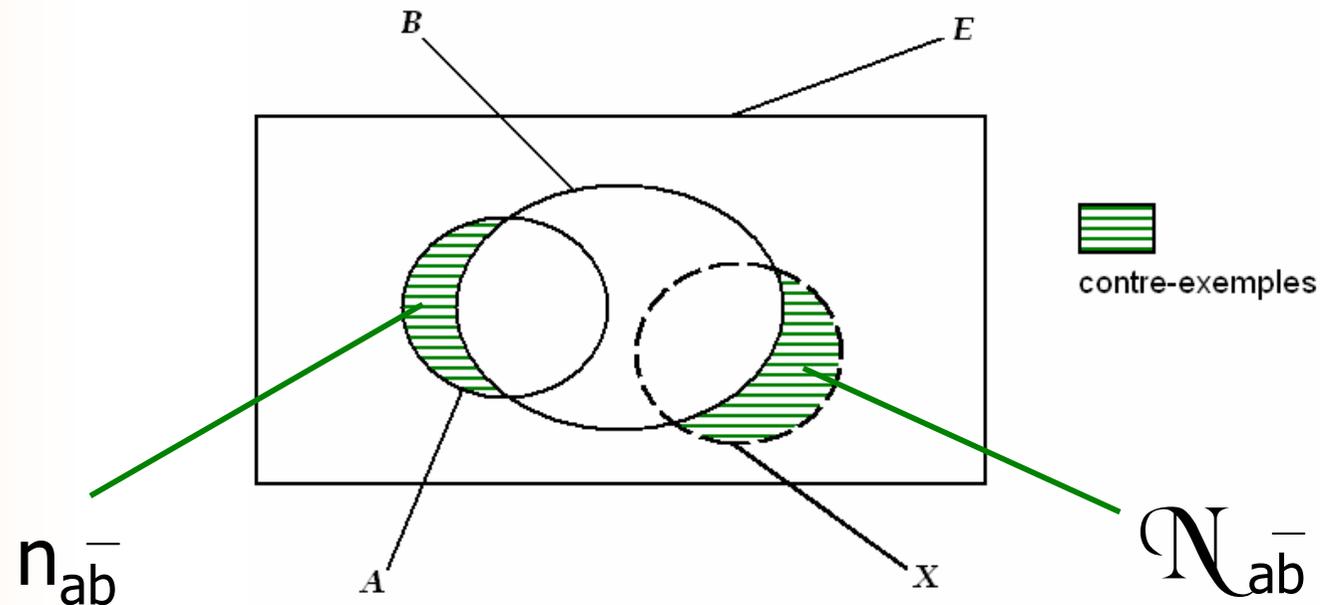
## IPEE

### ➤ Modèle aléatoire

### ➤ Expression analytique

## Propriétés

## Conclusion



Nombre de contre-exemples attendus sous  $H_0$  :

variable aléatoire  $\mathcal{N}_{ab}$

Indice Probabiliste d'Ecart à l'Equilibre :

$$\text{IPEE}(a \rightarrow b) = P(\mathcal{N}_{ab} > n_{ab} \mid H_0)$$

# Introduction

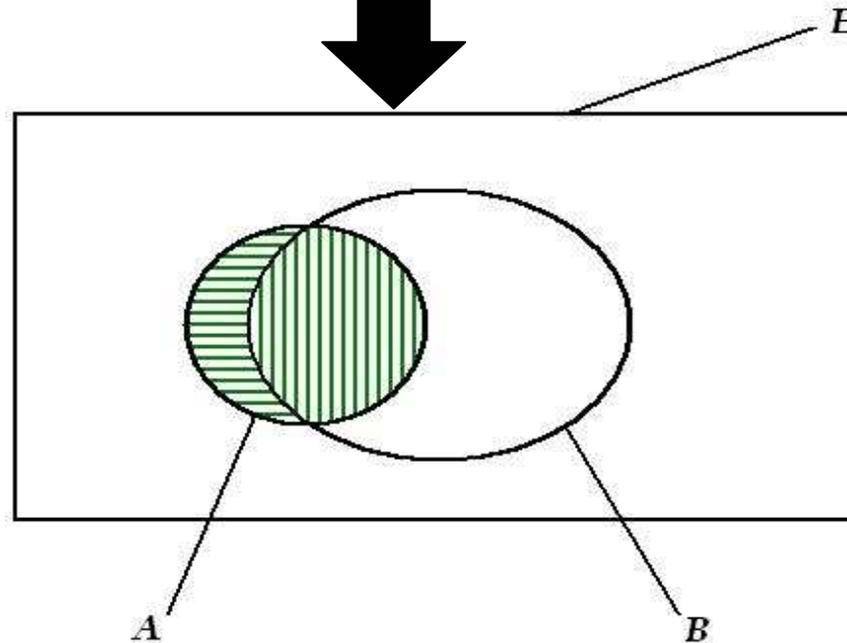
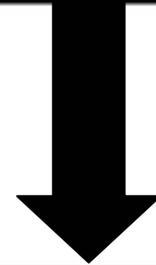
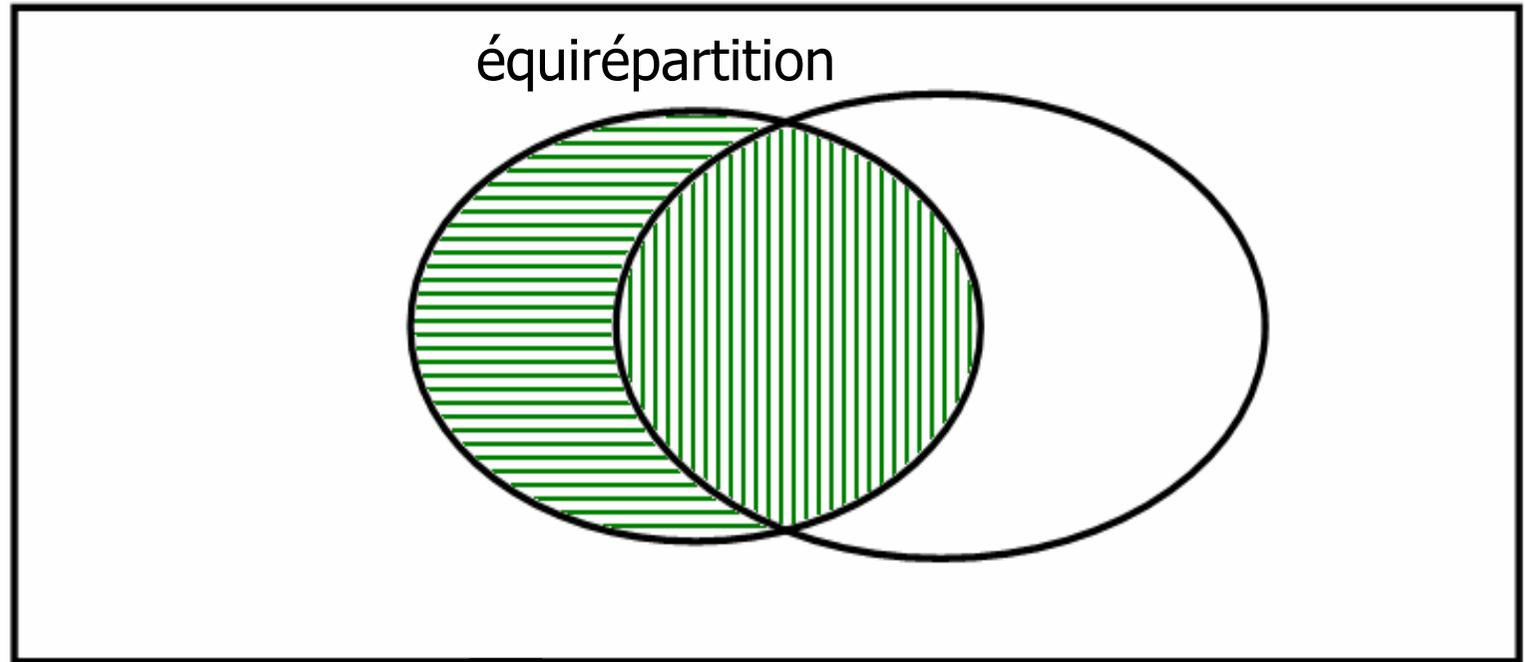
## IPEE

### ➤ Modèle aléatoire

### ➤ Expression analytique

## Propriétés

## Conclusion



exemples  $A \cap B$



contre-exemples  $A \cap \bar{B}$

## Introduction

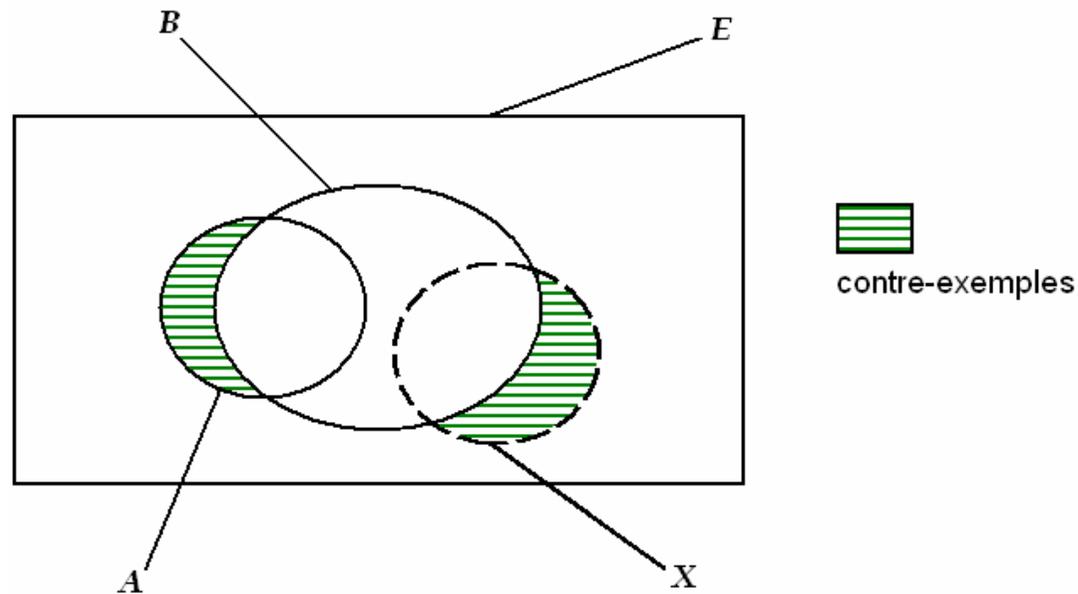
### IPEE

➤ Modèle aléatoire

➤ Expression analytique

Propriétés

Conclusion



$$\text{IPEE}(a \rightarrow b) = P(\mathcal{N}_{ab}^- > n_{ab}^- \mid H_0)$$

Loi binomiale de paramètres  $1/2$  et  $n_a$  :

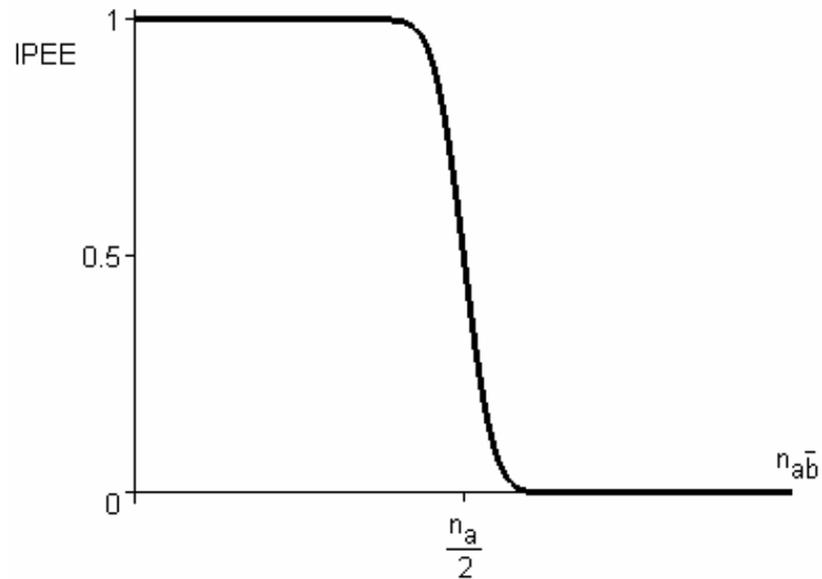
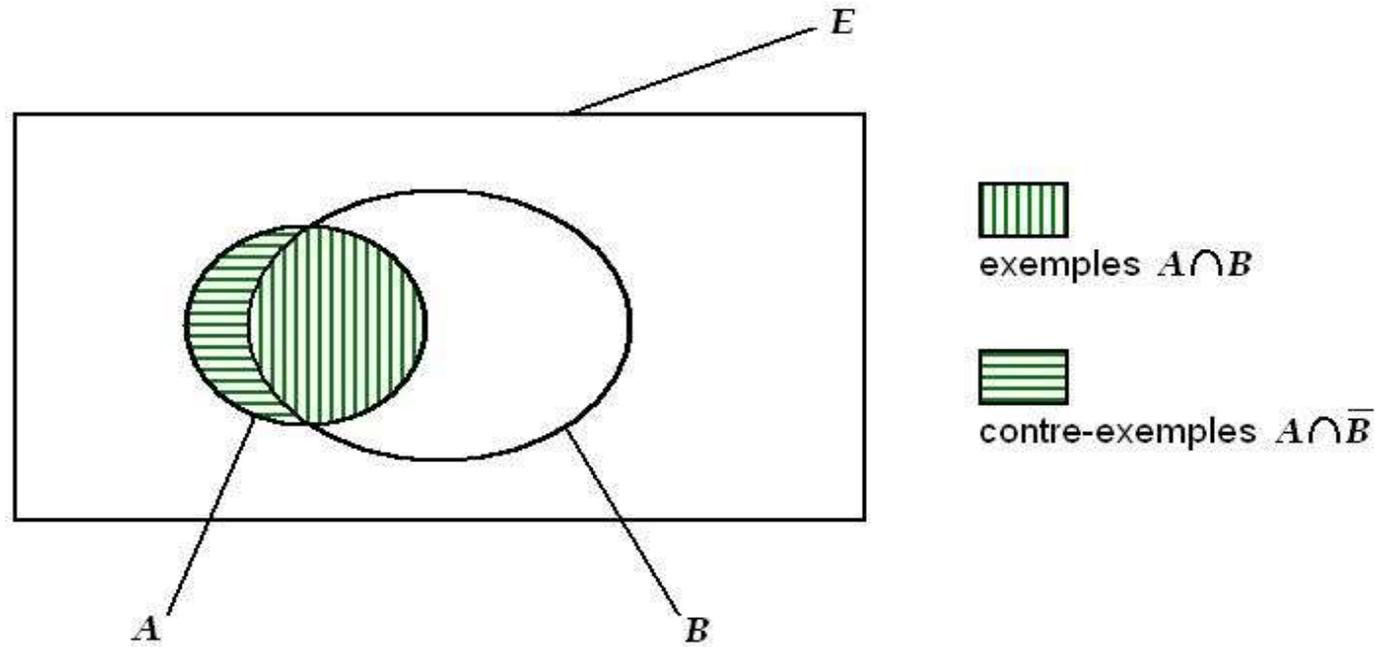
$$\text{IPEE}(a \rightarrow b) = 1 - \frac{1}{2^{n_a}} \sum_{k=0}^{n_{a\bar{b}}} C_{n_a}^k$$

# Introduction

## IPEE

### Propriétés

## Conclusion

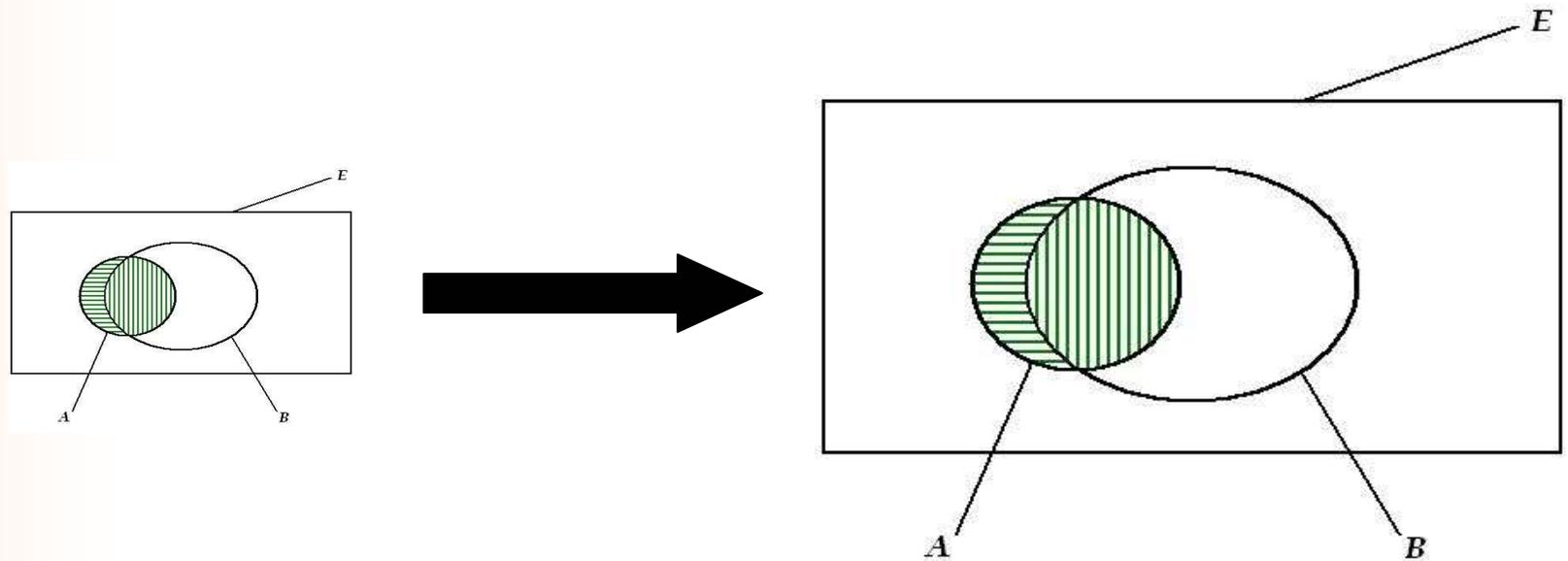
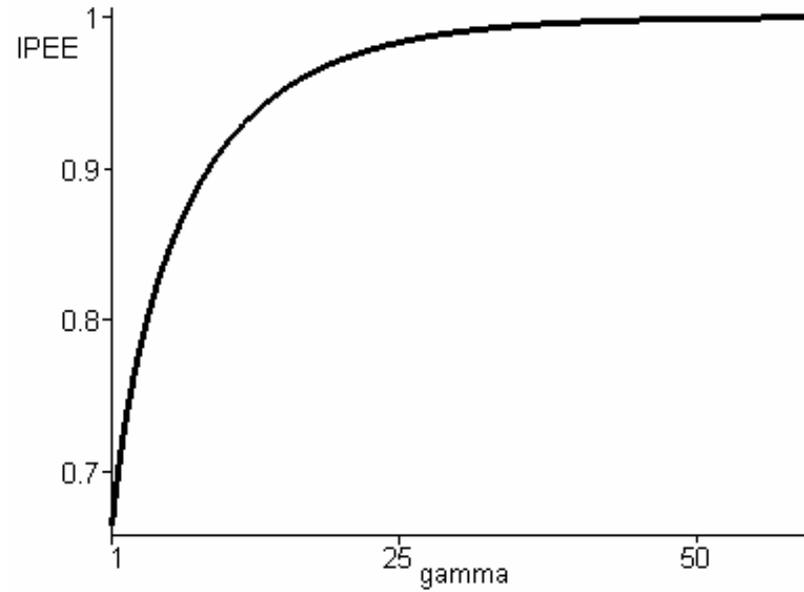


# Introduction

## IPEE

### Propriétés

### Conclusion



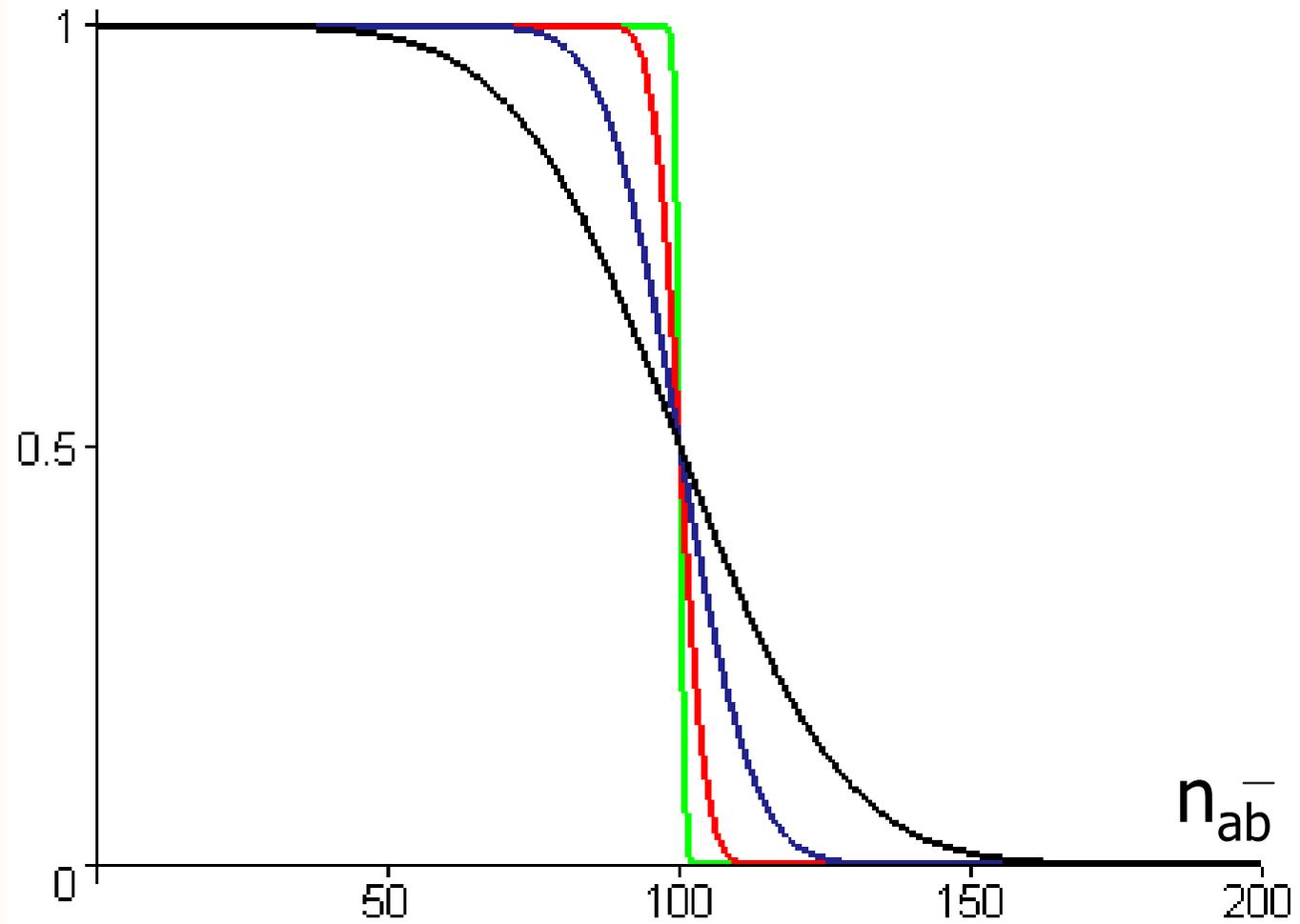
Règles logiques :  $IPEE = 1 - 2^{-n_a}$

Introduction

IPEE

Propriétés

Conclusion

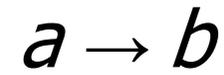


Introduction

IPEE

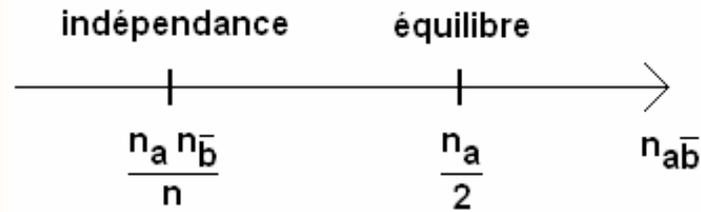
Propriétés

Conclusion

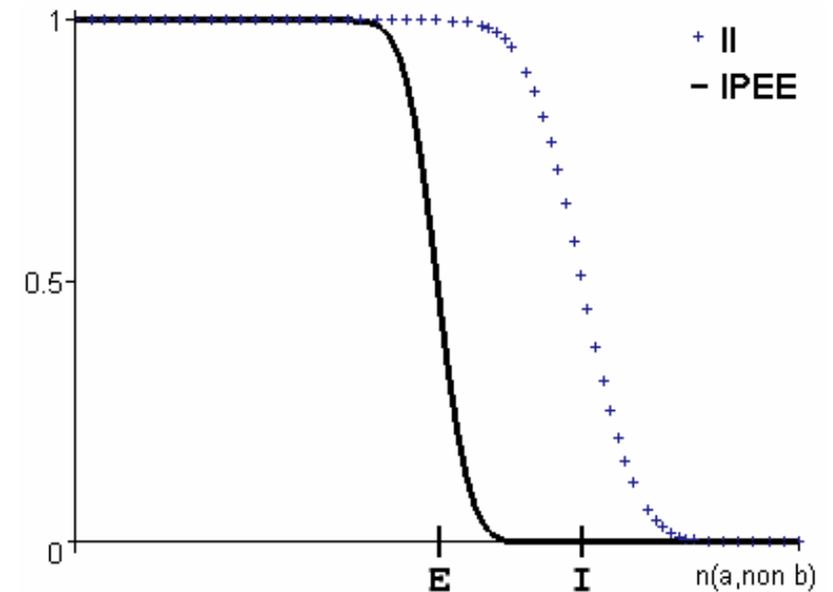
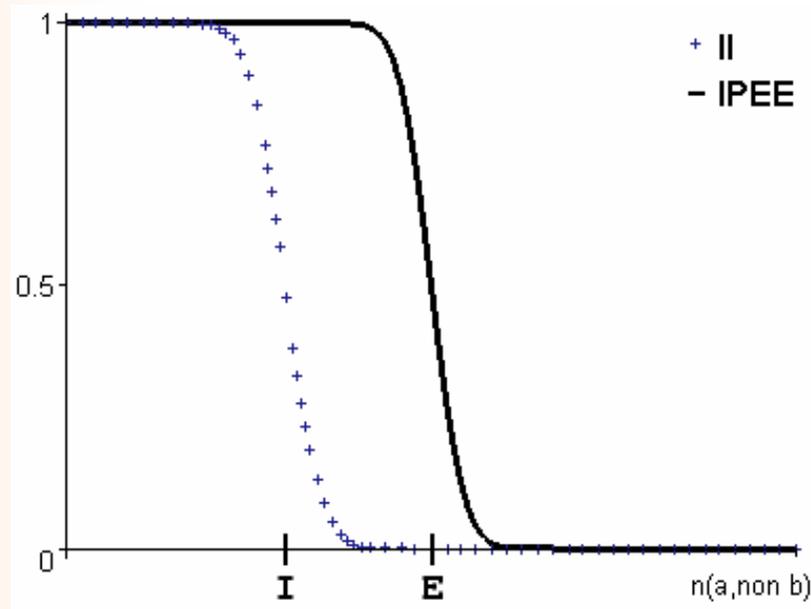
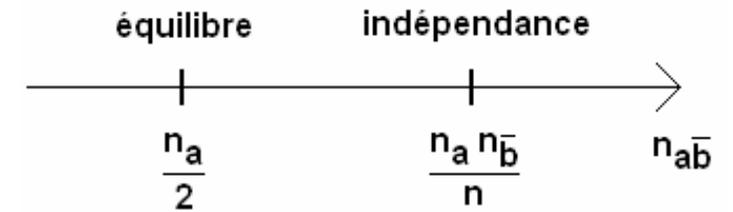


$$(n_a, n_b, n_{a\bar{b}}, n)$$

$$n_b \geq n/2$$



$$n_b \leq n/2$$



Introduction

IPEE

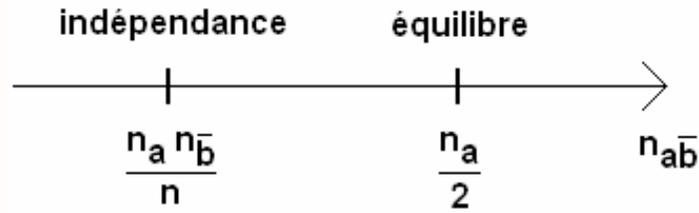
Propriétés

Conclusion

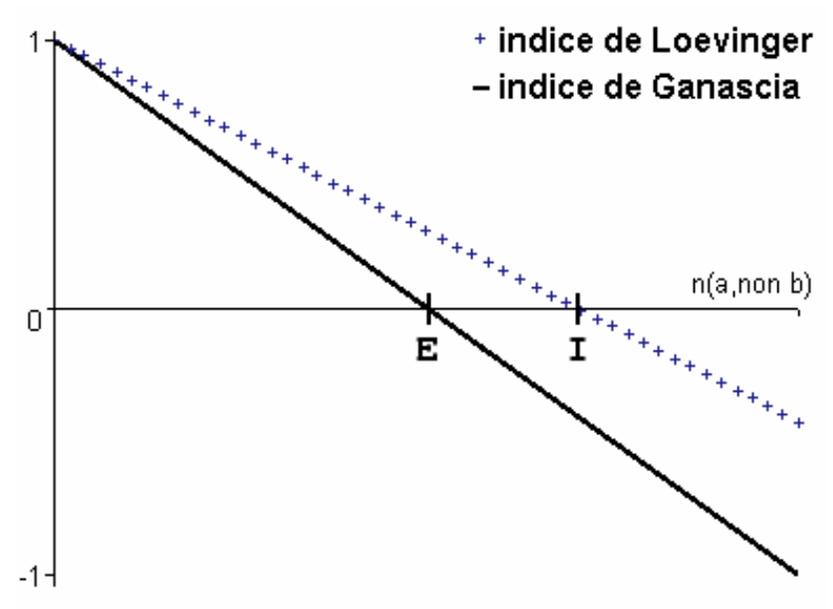
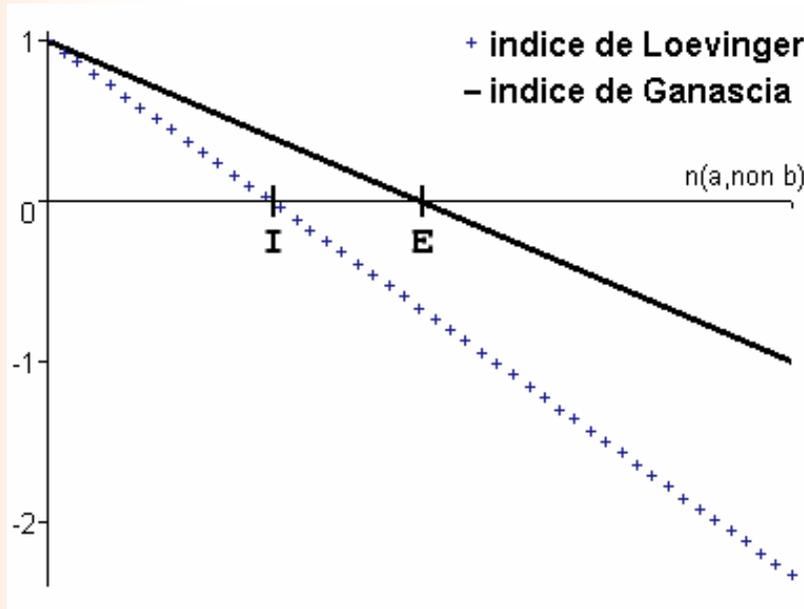
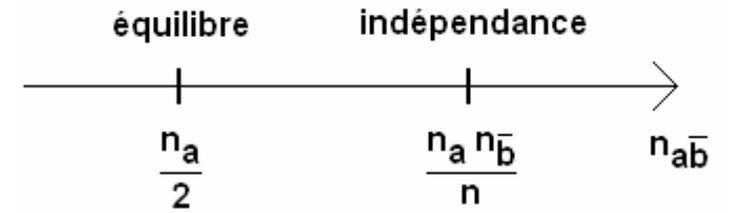
$$a \rightarrow b$$

$$(n_a, n_b, n_{ab}, n)$$

$$n_b \geq n/2$$



$$n_b \leq n/2$$



Introduction

IPEE

Propriétés

**Conclusion**

**IPEE :**

## **Indice Probabiliste d'Ecart à l'Equilibre**

- Mesure la significativité statistique de l'écart à l'équilibre
- Prend en compte la taille des phénomènes étudiés
- Fait référence à une échelle de valeurs intelligible
- Indice analogue à l'intensité d'implication  
➔ évaluation statistique complète des règles