

Probabilités, inférence statistique, et recherche opérationnelle (STAT-S202)

Partie I: Probabilités

Davy Paindaveine

2024–2025

Université libre de Bruxelles
Solvay Brussels School of Economics and Management

Trois parties

Q1 **Probabilités** : Davy Paindaveine

Théorie : 24h (=12×2h), TP : 18h (=9×2h)

Q2 **Inférence statistique** : Catherine Dehon

Théorie : 24h (=12×2h), TP : 18h (=9×2h)

Q2 **Recherche opérationnelle** : Yves De Smet

Théorie : 12h (=6×2h), pas de TP

En janvier, un examen sur le Q1

En juin, un examen commun sur le Q2

Note unique en juin (et, le cas échéant, en septembre)

Le détail des règles de calcul de notes, de report de session et de report d'année sont disponibles sur l'UV.

STATISTIQUE & MATHÉMATIQUES APPLIQUÉES

Probabilités et inférence statistique

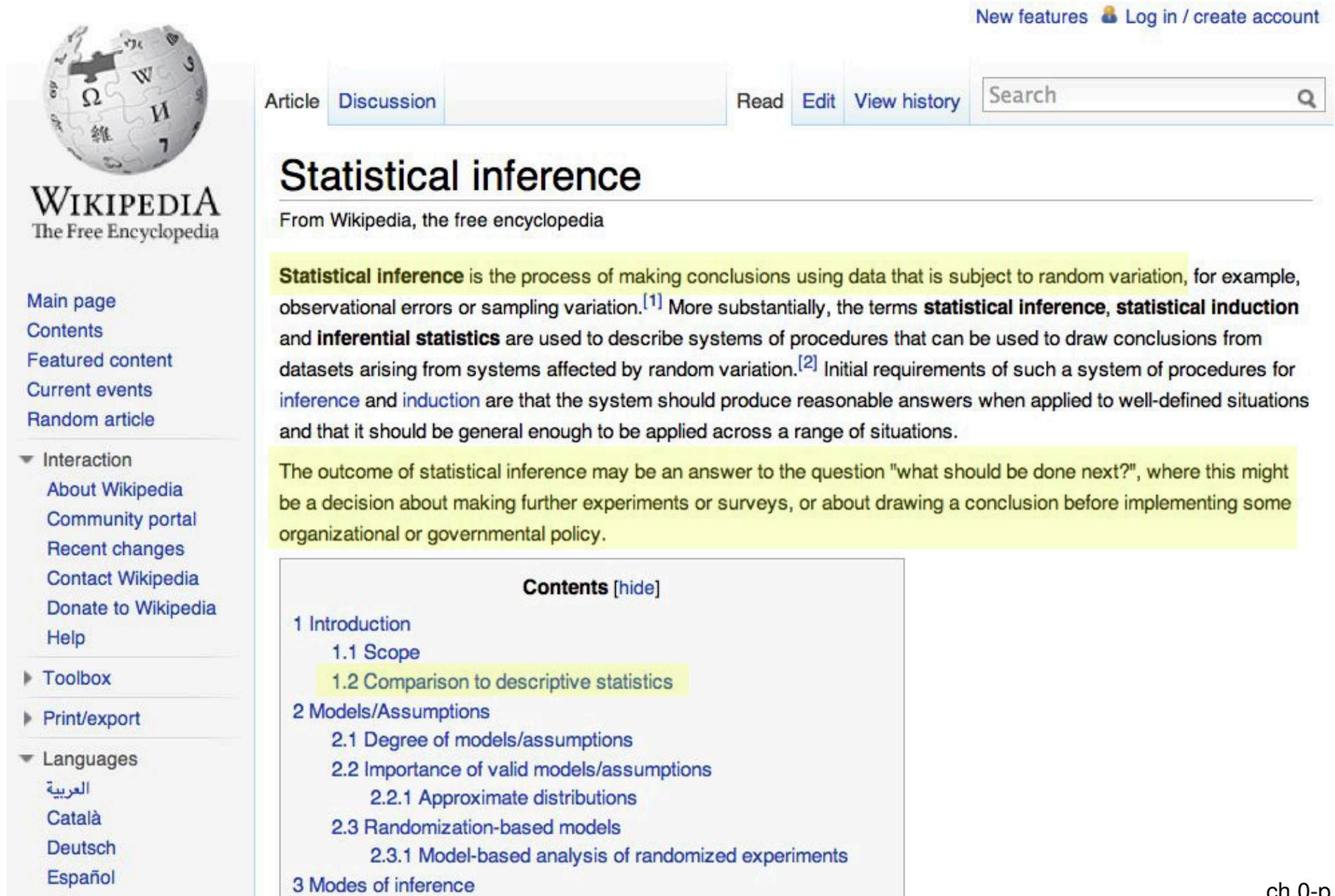
Catherine Dehon
Marc Hallin
Davy Paindaveine
Christine Thomas-Agnan
Catherine Vermandele



Éditions de l'Université de Bruxelles

Éditions Ellipses

L'inférence statistique, quoi et pour quoi ?



The image shows a screenshot of the Wikipedia article titled "Statistical inference". The page layout includes a sidebar on the left with navigation links, a main content area with a title and introductory text, and a table of contents. The text in the main content area is highlighted in yellow.

Statistical inference is the process of making conclusions using data that is subject to random variation, for example, observational errors or sampling variation.^[1] More substantially, the terms **statistical inference**, **statistical induction** and **inferential statistics** are used to describe systems of procedures that can be used to draw conclusions from datasets arising from systems affected by random variation.^[2] Initial requirements of such a system of procedures for **inference** and **induction** are that the system should produce reasonable answers when applied to well-defined situations and that it should be general enough to be applied across a range of situations.

The outcome of statistical inference may be an answer to the question "what should be done next?", where this might be a decision about making further experiments or surveys, or about drawing a conclusion before implementing some organizational or governmental policy.

Contents [hide]

- 1 Introduction
 - 1.1 Scope
 - 1.2 Comparison to descriptive statistics
- 2 Models/Assumptions
 - 2.1 Degree of models/assumptions
 - 2.2 Importance of valid models/assumptions
 - 2.2.1 Approximate distributions
 - 2.3 Randomization-based models
 - 2.3.1 Model-based analysis of randomized experiments
- 3 Modes of inference

Le 18 avril 2022, le bureau de campagne d'Emmanuel Macron juge que

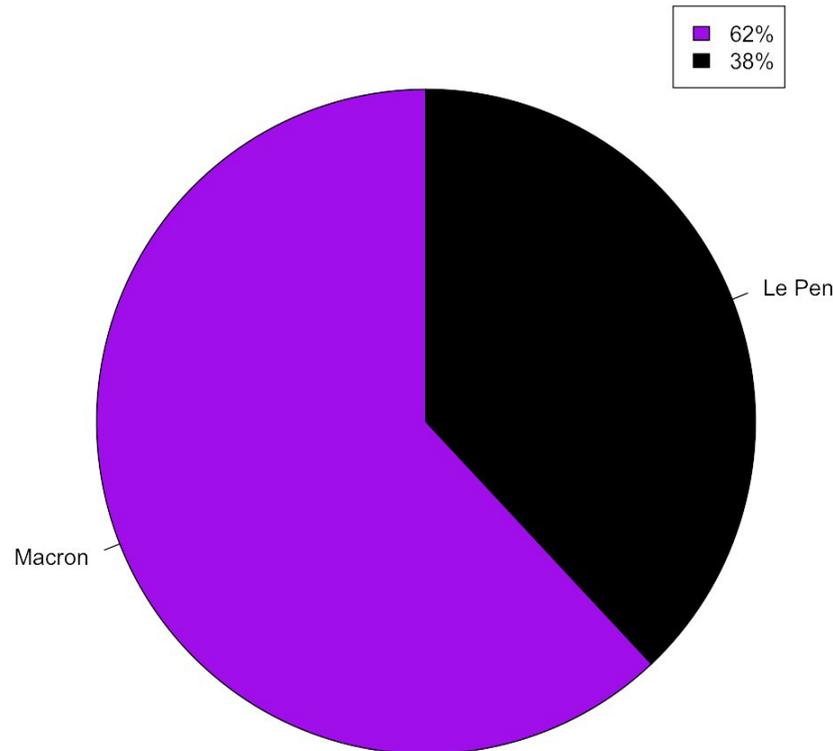
- si la proportion p des Français en faveur d'Emmanuel Macron est $\geq 52\%$, il faut opter pour une fin de campagne prudente, et que
- si $p < 52\%$, il faut au contraire opter pour une fin de campagne agressive.

Comment **décider** de ce qu'il faut faire ?

La décision dépend de la valeur de p , qui est malheureusement inconnue.

Puisqu'il est exclu d'interroger tous les français pour évaluer p , il faut réaliser un sondage : interroger 100 (ou 1000 ou...) futurs votants sur leurs intentions de vote.

La **statistique descriptive** s'arrête à la description des résultats de ce sondage.



Ce résultat de 62% peut se réaliser pour presque toutes les valeurs de $p \in (0, 1)$,
en raison des "variations aléatoires" auxquelles le résultat du sondage est soumis.

⇒ Ceci ne dit rien de définitif sur p

Introduction

Ceci dit, la valeur $p = 1\%$, par exemple, rend ce résultat de 62% **très peu probable** et est donc à écarter

Les probabilités

= un processus **déductif** :

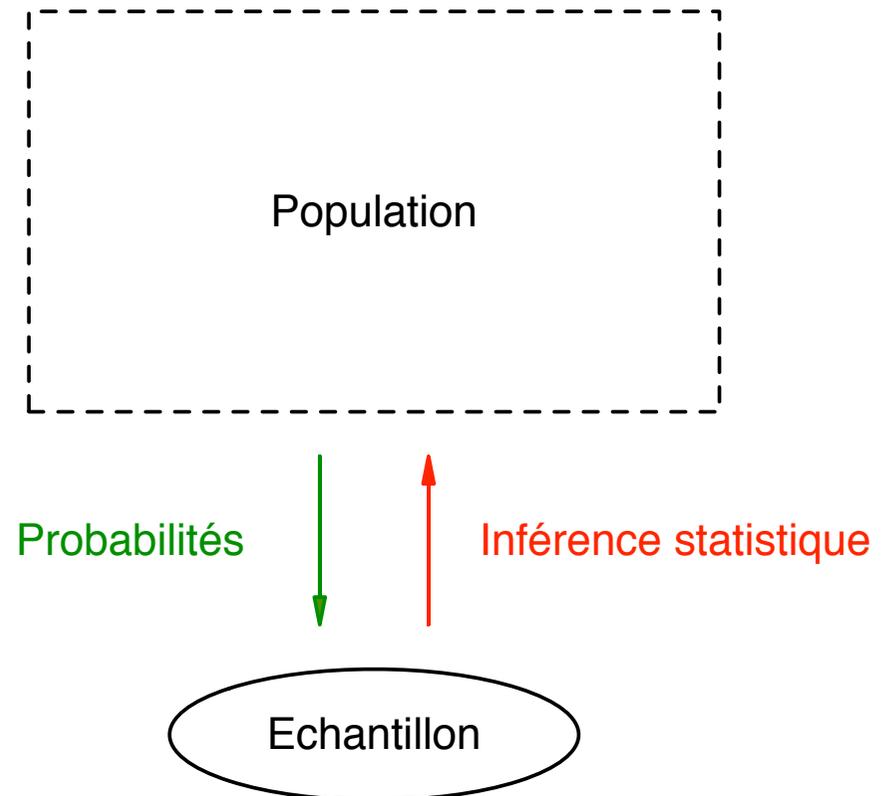
Une connaissance parfaite de la population permet de "prédire" les caractéristiques de l'échantillon qui sera obtenu aléatoirement



La statistique inférentielle

= un processus **inductif** :

L'échantillon observé permet d'obtenir de l'information sur la population qui n'est que très partiellement connue



Ce sont ainsi **les probabilités** qui engendrent la **statistique inférentielle**, laquelle va plus loin que la statistique descriptive : elle permet de tirer des conclusions (et donc de prendre des décisions). Comme le cours le montrera, elle établira ici

- que si on tolère une probabilité de 5% qu'on opte à tort pour une fin de campagne prudente, il convient d'opter en effet pour la prudence (alors qu'un résultat de sondage de 58% ne mènerait pas à cette conclusion),
- qu'une "fourchette" pour p , associée à un "taux d'erreur de 5%", est donnée par [52.5%, 71.5%].

Clairement, **toute "preuve statistique" comportera un risque d'erreur.**

- Comment définir cette erreur ?
- Comment la contrôler ? (p.ex., comment choisir une taille de sondage assurant une erreur inférieure à un seuil fixé par le cabinet Macron ?)
- Comment interpréter les résultats des procédures d'inférence statistique ?

Les applications des probabilités et de l'inférence statistique sont innombrables :

- **En économie** : quel est le lien entre les dépenses et les revenus des ménages ?
Comment prévoir le PNB en fonction d'autres grandeurs macroéconomiques ?
- **En finance** : comment apprécier les risques associés aux actifs financiers ?
Comment construire un portefeuille optimisant les profits en minimisant le risque ?
- **En assurances** : comment fixer les primes pour pouvoir faire face (avec une probabilité suffisante) à l'ensemble des sinistres qui se produiront à l'avenir ?
- **En politique de l'éducation** : quel est l'impact d'une augmentation de la taille des classes sur l'efficacité de l'enseignement ?
- **En santé publique** : quelle est l'importance du tabagisme passif ? Comment valider un médicament avant de l'introduire sur le marché ?
- ...

- 1 Mesures de probabilité
- 2 Variables aléatoires
- 3 Vecteurs aléatoires
- 4 Théorèmes limites et lemme de Fisher