



HAL
open science

**Sean Gailmard, Statistical Modeling and Inference for
Social Science, New York, Cambridge University Press,
2014**

Cyril Benoît

► **To cite this version:**

Cyril Benoît. Sean Gailmard, Statistical Modeling and Inference for Social Science, New York, Cambridge University Press, 2014. *Revue Française de Science Politique*, 2016, 66 (3-4), pp.619-620. 10.3917/rfsp.663.0566 . halshs-01380804

HAL Id: halshs-01380804

<https://shs.hal.science/halshs-01380804>

Submitted on 25 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gailmard Sean, *Statistical Modeling and Inference for Social Science*, New York, Cambridge University Press, 2014, (Analytical Methods for Social Research)

Connu pour ses contributions dans le domaine de la théorie des jeux et pour ses analyses du comportement des organisations bureaucratiques, Sean Gailmard propose dans cet ouvrage de dresser un panorama des principes de la modélisation et de l'inférence statistique en sciences sociales. S'adressant prioritairement à un public d'étudiant.e.s de niveau master ou doctorat, l'auteur y adopte une démarche résolument pédagogique. Les chapitres introductifs sont ainsi l'occasion d'un retour sur les fondements de la régression, depuis les notions de variable et de distribution jusqu'aux problèmes de multicollinéarité. Des paliers supplémentaires en termes de complexité sont ensuite franchis de façon progressive. Une liste de problèmes portant sur des aspects clés des thèmes discutés est également proposée à la fin de chacun des chapitres.

Si le format du livre est celui du manuel, et que ses premières parties peuvent utilement servir à la préparation d'un cours d'introduction aux méthodes statistiques, le/la lecteur.trice souhaitant approfondir ses connaissances en mathématiques appliquées trouvera dans les pages suivantes une présentation formelle de méthodologies plus élaborées. Ainsi que le rappelle S. Gailmard, il est en effet rare que l'étude de relations entre plusieurs variables ne soit pas associée à l'examen de la portée d'une théorie dépassant le seul contexte des données observées (chapitre 3). Au contraire, son ambition est plus régulièrement de mettre à jour le principe gouvernant la manière dont des facteurs sont reliés les uns aux autres, comme lorsque Theda Skocpol s'interroge sur les déterminants structurels des révolutions sociales en comparant différentes « situations révolutionnaires » (p. 70). Dans ce contexte, le raisonnement statistique peut être envisagé comme une tentative d'identification du « processus générateur des données observées » (*DGP*, pour *Data Generating Process*). Toutefois, cette opération ne saurait se réduire à la mise au jour des conditions « nécessaires » ou « suffisantes » à la réalisation d'un processus : outre les simplifications charriées par les théories éprouvées et les contraintes inhérentes au travail d'échantillonnage, une telle définition du *DGP* amènerait le chercheur à ignorer la part d'incertitude qui caractérise la réalisation de tout phénomène historique (et qui le désigne comme tel). Si un.e statisticien.ne parvenait à construire un modèle extrêmement complexe pour capturer la totalité des événements ayant conduit à la Révolution française, la Révolution n'apparaîtrait certainement pas sous la même forme, voire n'apparaîtrait pas du tout, si le modèle était « relancé » (p. 77). Représenter cette part d'aléatoire implique de construire des modèles à même de refléter la dimension stochastique de la plupart des *DGP*. Pour comprendre comment des phénomènes s'y combinent et opérer des montées en généralité sur de plus solides assises, l'analyse ne doit pas se cantonner à l'étude de la manière dont des variables sont liées les unes aux autres dans des bases de données spécifiques. Au contraire, il faut également tenir compte des événements qui auraient pu mais ne sont pas advenus, dans toutes les bases de données qui auraient pu être construites pour ce *DGP* (p. 73 et 85). Pour formaliser ce postulat, la théorie des probabilités peut constituer un recours. C'est l'objet du chapitre 4, qui en introduit les concepts fondamentaux.

Si les chapitres 3 et 4 justifient le recours à la modélisation pour déterminer la pertinence explicative d'une théorie, les chapitres 5 (comportement des variables aléatoires, notion de covariance) et 6 (familles de distributions de probabilité, moyenne conditionnelle) abordent plus frontalement les enjeux de l'analyse de données en elle-même, ainsi que les modalités d'identification du *DGP* qui les a générées. Les chapitres 7 à 9, qui présentent notamment les notions d'échantillon aléatoire, de tests d'hypothèse et d'estimation (avec une section bienvenue sur l'approche Bayésienne), discutent des conditions de mise en rapport d'une théorie avec les propriétés d'un jeu de données observables. À vocation conclusive, le chapitre 10 est consacré au problème de la causalité. Par contraste avec le reste du livre, et en dépit de sa dimension transversale, la présentation de ce concept est assez décevante. Si les solutions apportées sont d'une grande qualité sur le plan technique, les enjeux soulevés par leur utilisation (comment passer d'une théorie libellée en termes de relations causales à des méthodes focalisées sur des associations entre variables ?) sont trop vite relégués au rang des « considérations métaphysiques ». Paradoxalement, il s'agit de l'une des seules parties de l'ouvrage où l'auteur pêche par excès de positivisme. Ailleurs, les méthodes sont discutées de façon réflexive et critique. Les nuances exprimées dans le texte comme dans l'écriture des formules mathématiques permettent ainsi d'apprécier la souplesse autorisée par les procédures présentées. La promesse pédagogique faite en introduction est par conséquent tenue : le/la statisticien.ne débutant.e ou confirmé.e trouvera dans *Statistical Modeling and Inference for Social Science* des ressources pour l'enseignement, ainsi qu'une exposition rigoureuse des approches permettant une modélisation « réaliste » des phénomènes sociaux.

Cyril Benoît

Sciences Po Bordeaux, Centre Emile Durkheim