

Introduction aux statistiques

Maxime Forriez^{1,2, a} and Maxime Forriez^{1,2,a}

¹ Sorbonne université, 2, rue Francis de Croisset, 75 018 Paris

² Institut de géographie, 191, rue Saint-Jacques, Bureau 105,
75 005 Paris,

^amaxime.forriez@sorbonne-universite.fr

24 septembre 2025

Dans le cadre des statistiques, il est difficile de ne pas évoquer la notion de probabilité. Toutefois, il est fondamental, pour bien comprendre ce chapitre, qu'il ne s'agit que de récapituler les notions utiles dans le cadre des statistiques des variables quantitatives. Dit autrement, ce chapitre n'est qu'une vaste introduction à la notion mathématique de probabilité.

Le calcul de probabilité, « reposant sur des définitions et conventions purement logiques, constitue, en soi, un corps de doctrine indépendant de toute réalité ; mais qu'à l'expérience il s'est révélé applicable à certaines réalités bien concrètes » [Baulig, 1959, p. 399]. H. Baulig¹, un géographe, exprimait ainsi sa méfiance vis-à-vis des probabilités. Sont-elles pertinentes dans un contexte géographique ? Ce chapitre n'a pas la prétention de répondre à cette question. Il expose les bases nécessaires pour vous forger votre propre opinion.

Le calcul de probabilité s'occupe des phénomènes aléatoires (ou processus stochastiques). De tels phénomènes ont pour caractéristique de ne pas arriver tout le temps à la même issue. Les probabilités permettent ainsi de décrire des situations dans lesquelles le futur est incertain.

D'un point de vue mathématique, les probabilités ont longtemps été formulées intuitivement. Elles trouvent leur origine dans le **problème du chevalier de Méré**² au XVII^e siècle. La formalisation définitive et l'usage dans la modélisation des probabilités est assez récente. Elle fut inventée par A. N. Kolmogorov³ en 1933. Toutefois, certaines notions clés furent inventées bien plutôt, ce

1. Henri Baulig (1877-1962)

2. Antoine Gombaud (1607-1684) dit le chevalier de Méré

3. Andrei Nikolaïevitch Kolmogorov (1903-1987)

qui confèrent aux probabilités une histoire ancienne. Par exemple, T. Bayes⁴ inventa la notion de **probabilités conditionnelles**, ainsi qu'un théorème sur lequel certains ont longtemps pensé fonder la statistique. Malheureusement, s'il est parfait d'un point de vue théorique, il s'est avéré quasiment inutilisable en pratique jusqu'à l'arrivée des ordinateurs.

Les usages des probabilités sont nombreuses. On les rencontre principalement :

1. dans les jeux de hasard ;
2. dans la gestion du risque (assurance, industrie) ;
3. dans la gestion des processus stochastiques ;
4. dans la finance (portefeuilles).

Le géographe est-il confronté à la gestion du hasard ? La réponse est **affirmative**. La géographie du risque regroupe aussi bien des problèmes de géographie physique (crues, avalanches, précipitations, *etc.*) que de géographie humaine (circulation routière, *etc.*) En réalité, le hasard est présent dans toutes les branches de la géographie. Il faut savoir l'ordonner, l'étudier. Pour ce, il convient d'utiliser : 1. les raisonnements probabilistes ; 2. les modèles statistiques associés.

Toutefois, l'information géographique est souvent partielle, non exhaustive. C'est là qu'intervient l'**inférence statistique** et l'**échantillonnage**, qui seront au cœur des séances. Deux théories répondent au problème d'une information limitée. La principale question de la théorie de l'estimation est « Comment utiliser un échantillon pour décrire la population dans son entier ? ». La théorie des tests est « Comment la pertinence d'une hypothèse à partir d'une information limitée à un échantillon ? ». Elle permet de jalonner une démonstration, et de valider ou d'invalidier les résultats obtenus, d'en fixer la limite conditionnelle. Quelle est la probabilité que la relation établie sur une statistique univariée, bivariée, ou multivariée se produise au hasard ? Plus elle est basse, plus la relation établie est forte. Si elle est forte, au niveau de quel seuil peut-on parler d'une loi ?

Les statistiques utilisent les probabilités. Il convient de fait de connaître un minimum celles-ci. Il faut bien comprendre la présentation d'un ensemble fini, les dénombrements, les règles de calcul des probabilités, *etc.* Les probabilités permettent la définition de lois concernant des variables aléatoires servant de référence aux modèles statistiques. Tout l'art d'une analyse est de comparer des fréquences observées avec des lois de probabilités théoriques.

4. Thomas Bayes (1702-1761)

Références

[Baulig, 1959] BAULIG, H. (1959). Morphométrie. Annales de Géographie, 68(369):385–408.