

## Séance 4 questions - Analyse de données

1. Quels critères mettriez-vous en avant pour choisir entre une distribution statistique avec des variables discrètes et une distribution statistique avec des variables continues?

Le choix d'une loi statistique, donc d'une distribution, dépend en premier lieu de la nature du phénomène étudié, ce qui permet de choisir entre « loi discrète et loi continue ». Viennent ensuite la forme de la distribution empirique (visuellement ou statistiquement testable), les caractéristiques de la série : espérance, médiane, variance, asymétrie etc. et le nombre de paramètres de la loi, certaines lois s'adaptant davantage selon leur flexibilité .

On choisit donc une loi/distribution discrète lorsque :

- les valeurs possibles sont dénombrables, souvent limitées à des entiers
- il s'agit de comptages: nombre d'événements, de succès/échecs, d'individus ( cf lois discrètes : Binomiale, Bernoulli, Poisson, Hypergéométrique... ) .

Parallèlement, on choisit une loi/distribution continue lorsque :

- la variable peut prendre toutes les valeurs d'un intervalle, non dénombrables
- il s'agit de mesures continues : temps, distance, altitude, température... (cf les lois continues : normale, log-normale, exponentielle, uniforme continue...) .

Le critère majeur est donc la nature du phénomène et la structure des valeurs observée, le tout appuyé par la forme empirique de la distribution et les paramètres statistiques.

2. Expliquez selon vous quelles sont les lois les plus utilisées en géographie?

Certaines lois sont particulièrement importantes la géographie, en premier lieu la loi de Poisson qui est décrite comme « indispensable pour les événements rares » et apparaît lorsque l'on compte des occurrences dans une surface ou un intervalle. Son utilité en géographie est de modéliser des événements ponctuels dans l'espace ou le temps comme par exemple le nombre d'accidents, de séismes, d'occurrences d'un phénomène localisé.

En outre, la loi normale (Gauss) est elle décrite comme « la plus fréquente » et constitue souvent la distribution limite de nombreux phénomènes Elle permet de nombreuses variables naturelles et sociales approximées par une normale par exemple la distribution des hauteurs, les températures, les revenus ou les rendements.

La loi log-normale est mentionnée comme essentielle pour des variables multiplicatives et asymétriques (loi de Galton-Gibrat), elle est utile pour tout ce qui est taille des villes, surface des parcelles, revenus, intensité de certains flux.

Les lois rang-taille (Zipf et Zipf-Mandelbrot) sont utilisées en géographie pour les distributions rang-taille notamment pour les tailles de villes . Elle permet de modéliser la hiérarchie urbaine et analyser la structure polarisée d'un territoire.

Enfin, La loi exponentielle est utilisée pour les processus liés au temps d'attente ou aux risques, décrite comme adaptée aux phénomènes de fiabilité et de survie .Donc utile pour les durées

d'événements naturels ou techniques ou la modélisation de probabilité de défaillance, de temps entre deux occurrences.