Séance 4 : les distributions statistiques

Pour choisir entre une distribution statistique avec des variables quantitatives discrètes et une distribution statistique avec des variables quantitatives continues il faut déjà regarder type de données sur lesquelles on appliquera ces variables. Il faut aussi choisir le type de distribution statistique en fonction de ce que l'on souhaite faire ensuite ressortir au moment du traitement de l'analyse de données étudiées. Dit simplement les variables discrètes servent à compter en s'appliquant sur une série statistique précise limitée avec un décompte effectué. Par exemple on se sert d'une variable discrète pour compter le nombre de véhicules défectueux dans un garage automobile comptant 20 voitures. Le nombre de véhicules correspondant à cette variable sera forcément X sur 20 en sachant que X = {0, 1, 2, …, 20}. Les variables continues servent à mesurer donc elles utilisent des outils comme moyenne, la somme, l'écart-type pour mettre en exergue ce que l'on étudie.

Rentrons dans le détail d'une distribution statistique avec des variables (aléatoire) discrètes. Elles sont utilisées dans la vie courante dans le résultats des jeux de hasard, dans les sondages d'opinion, ou encore pour analyser des processus aléatoires tels que les files d'attentes. Ce que j'ai décrit dans le premier paragraphe concernant les variables discrètes correspond à la loi uniforme discrète où X peut prendre chaque valeur de l'ensemble avec la même probabilité. Cette loi est utilisée quotidiennement dans les jeux de pile ou face (la pièce a autant de probabilité de retomber sur pile ou sur face en sachant qu'il n'y a que ces deux scénarios qui sont possibles), dans les jeux de cartes (on aura forcément une carte du jeu de 54 cartes) ou encore à la loterie. La variable de Bernoulli rentre également dans les variables discrètes puisqu'elle permet d'indiquer aléatoirement. La variable binomiale permet quant à elle de mesurer la fréquence du nombre de succès dans plusieurs expériences aléatoires identiques et indépendantes. Elle permet de définir deux paramètres : **n** (le nombre d'expérience) et **p** (la probabilité de succès dans chaque expérience). Elle modélise une situation d'échec ou de succès. Lorsque que la probabilité qu'un événement se produise devient rare on utilise plutôt la loi de Poisson. Cela s'applique quand le nombre d'expérience est élevé, donc que l'événement devient rare. On peut aussi inverser le raisonnement en cherchant à savoir combien de fois faut-il reproduire une expérience pour obtenir un succès avec la loi géométrique. On calcule alors le nombre d'essais qu'il faut faire pour obtenir succès. La loi de Pascal en revanche cherche à connaître le nombre d'expériences nécessaires pour obtenir le nombre précis de réussite.

Passons maintenant aux distributions statistiques des variables continues. Elles servent plutôt à effectuer des mesures sur les distributions statistiques étudiées. La loi de Poisson est aussi utilisable pour des distributions statistiques continues pour se concentrer sur l'observation d'événements aléatoires, indépendants et identiquement distribuées. Mais quand les événements sont très rares, la probabilité est donc faible et peut tendre vers zéro. Avec cette distribution on peut décrire de nombreux processus dont la probabilité est petite et constante. Quand une mesure est opérée, il reste une incertitude sur le résultat unique obtenu. Ensuite la loi normale se retrouve dans de nombreuses situations, caractérisée par une représentation graphique avec une courbe en forme de cloche (l'ogive de Galton). Elles est caractérisée par deux paramètres : la moyenne **μ** et l'écart-type **σ**. Elle est définie par la fonction de densité de probabilité. Plus l'écart-type **σ** est petit plus la distribution est resserrée par rapport à la moyenne **μ**. Sur la représentation graphique, le point le plus de la courbe en forme de cloche correspond à la moyenne de la distribution statistique. Plusieurs propriétés générales sont applicables à la loi Normale comme par exemple le probabilité que X se trouve à un écart-type de la moyenne est d'environ 68% ; la probabilité que $P(X>\mu+1{,}65.\sigma) = P(X<\mu-1{,}65.\sigma) = 5\%$; enfin la probabilité que $P(X>\mu+1{,}96.\sigma) = P(X<\mu-1{,}96.\sigma) = 2{,}5\%$, dans ce dernier cas l'aire complémentaire est à

95%. La loi de Student permet d'étudier un échantillon pour une distribution normale dont on ne connaît pas la variance.

Pour la géographie, on peut se servir souvent de la loi Zipf et Zipf-Mandelbrot. Au départ mise en place pour s'intéresser à la fréquence de l'utilisation d'un mot dans un texte, elle sert pour la géographie en dans les variables discrètes appliquées aux territoires. Les lois rang-taille pour les villes par exemple avec d'un côté la place d'une ville dans la hiérarchie du réseau urbain et de l'autre côté son nombre d'habitants. La loi de Pareto est aussi utile pour mesurer une petite proportion d'unité qui concentre une grande part du phénomène, comme par exemple dans le cas des revenus, de la concentration foncière, de la répartition d'équipement ou encore des flux (20% des communes concentrent 70% des infrastructures sportives – données créées pour l'exemple). La loi normale peut servir pour la géographie climatique et environnementale (températures, altitudes, précipitations ; rendements agricoles…). En revanche on ne peut pas utiliser la loi normale pour les variables aléatoires concernant la fréquences du dénombrement de certains objets géographiques (lacs, montagnes…). La loi log-normale s'utilise quand les valeurs sont positives et asymétriques comme pour les revenues ou les surfaces des exploitations agricoles. On peut également se servir de la loi de Poisson pour étudier des événements rares et discrets dans l'espace comme pour les analyses routières en localisant les accidents, ou encore en géographie physique pour cartographier les séisme et l'apparition de nouveaux cratères. La loi exponentielle sert pour décrire la décroissance avec la distance ou le temps (la diminution de la densité de population autour d'un centre urbain).

Ces quelques lois évoquées sont donc souvent utilisées en géographie pour étudier différents phénomènes. Si on veut résumer brièvement, les lois de Zipf et de Pareto servent surtout en géographie humaine et urbaine ; tandis que les lois normale, log-normale et Poisson sont utilisées davantage en géographie physique et environnementale.