Statistiques d'ordre

RAHMANI Naceur

11/11/2025

Introduction aux Statistiques d'Ordre

Une **statistique d'ordre** est la k-ième plus petite valeur dans un échantillon $X_1, X_2, \ldots, X_n$ trié. Par exemple :

- $X_{(1)}$ est le minimum.
- $X_{(n)}$ est le maximum.
- $X_{(k)}$ est la k-ième plus petite valeur de l'échantillon.

La fonction de répartition empirique est la proportion des éléments inférieurs ou égaux à une valeur x.

Densité du Minimum $X_{(1)}$

Le **minimum** d'un échantillon de taille n, $X_{(1)}$, est la plus petite valeur parmi les X_i. La fonction de répartition de $X_{(1)}$ est donnée par :

$$P(X_{(1)} \leq x) = 1 - (1 - F(x))^n$$

Où $F(x)$ est la fonction de répartition de chaque X_i.
La densité de $X_{(1)}$, obtenue par dérivation de la fonction de répartition, est :

$$f_{X_{(1)}}(x) = n \cdot (1 - F(x))^{n-1} \cdot f(x)$$

Où $f(x)$ est la densité associée à la fonction de répartition $F(x)$.

Densité du Maximum $X_{(n)}$

Le **maximum** d'un échantillon de taille n, $X_{(n)}$, est la plus grande valeur parmi les X_i. La fonction de répartition de $X_{(n)}$ est donnée par :

$$P(X_{(n)} \leq x) = F(x)^n$$

Où $F(x)$ est la fonction de répartition de chaque X_i.
La densité de $X_{(n)}$, obtenue par dérivation de la fonction de répartition, est :

$$f_{X_{(n)}}(x) = n \cdot F(x)^{n-1} \cdot f(x)$$

Où $f(x)$ est la densité associée à la fonction de répartition $F(x)$.

Densité Associée à $X_{(k)}$

La densité de la **k-ième statistique d'ordre** $X_{(k)}$ peut être obtenue en différentiant la fonction de répartition de $X_{(k)}$.
Soit un échantillon $X_1, X_2, \ldots, X_n$ de taille n, avec une fonction de répartition $F(x)$. La fonction de répartition de $X_{(k)}$ est donnée par :

$$P(X_{(k)} \leq x) = \sum_{i=k}^{n} \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i}$$

La densité de $X_{(k)}$ est la dérivée de la fonction de répartition :

$$f_{X_{(k)}}(x) = \frac{d}{dx} \left(\sum_{i=k}^{n} \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i} \right)$$

Après différentiation, nous obtenons la densité suivante :

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} f(x) [F(x)]^{k-1} [1 - F(x)]^{n-k}$$

Où $f(x)$ est la densité associée à la fonction de répartition $F(x)$.

Exemple : Distribution Uniforme $U(0,1)$

Supposons que $X_1, X_2, \ldots, X_n$ suivent une distribution uniforme $U(0,1)$, donc $F(x) = x$ pour $x \in [0,1]$ et $f(x) = 1$.
Les densités associées aux statistiques d'ordre sont :

- Densité du minimum $X_{(1)}$:

$$f_{X_{(1)}}(x) = n(1-x)^{n-1} \quad \text{pour} \quad x \in [0,1]$$

- Densité du maximum $X_{(n)}$:

$$f_{X_{(n)}}(x) = nx^{n-1} \quad \text{pour} \quad x \in [0,1]$$

Ces densités sont caractéristiques d'un échantillon de taille n tiré d'une distribution uniforme.

Distribution Asymptotique du Maximum et du Minimum

Dans l'**asymptotique**, pour des valeurs très grandes de n, le maximum $X_{(n)}$ et le minimum $X_{(1)}$ suivent des lois limites bien connues :

- **Maximum** : La loi asymptotique du maximum suit une **loi de Gumbel** :

$$P\left(\frac{X_{(n)} - \mu_n}{\sigma_n} \leq z\right) \to e^{-e^{-z}} \quad \text{(Gumbel pour le maximum).}$$

- **Minimum** : La loi asymptotique du minimum suit également une loi de Gumbel :

$$P\left(\frac{X_{(1)} - \mu_1}{\sigma_1} \leq z\right) \to e^{-e^{z}} \quad \text{(Gumbel pour le minimum).}$$

Où μ_n et σ_n sont les paramètres de localisation et d'échelle respectivement.

Propriétés Générales des Statistiques d'Ordre

Les **statistiques d'ordre** ont plusieurs propriétés fondamentales :

- **Monotonie** : Les statistiques d'ordre sont monotones croissantes : $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$.
- **Indépendance** : Les statistiques d'ordre **ne sont pas indépendantes**. Par exemple, le minimum et le maximum d'un échantillon sont liés entre eux.
- **Symétrie** : La fonction de répartition de $X_{(k)}$ dépend de la fonction de répartition $F(x)$, mais l'ordre des statistiques n'affecte que la position dans l'échantillon trié.

Uniformité des Statistiques d'Ordre

Lorsque l'échantillon est tiré d'une **distribution uniforme** $U(0,1)$, les statistiques d'ordre $X_{(k)}$ suivent une distribution uniforme. Plus précisément :

$$P(X_{(k)} \leq x) = F(x)^k \quad \text{et} \quad f_{X_{(k)}}(x) = k(1-x)^{k-1}.$$

Cela signifie que les statistiques d'ordre dans une distribution uniforme suivent une loi uniforme sur l'intervalle $[0,1]$.

Limites des Distributions du Maximum et du Minimum

Les **limites des distributions** des statistiques d'ordre maximum et minimum se trouvent dans le cadre des distributions asymptotiques des extrêmes. Par exemple :

- **Maximum** : La loi de Gumbel pour le maximum devient applicable lorsque $n \to \infty$, et la fonction de répartition asymptotique est :

$$P\left(\frac{X_{(n)} - \mu_n}{\sigma_n} \leq z\right) \to e^{-e^{-z}}.$$

- **Minimum** : La distribution du minimum converge également vers une loi de Gumbel lorsque $n \to \infty$.

Ces lois sont utilisées dans l'analyse des valeurs extrêmes et des phénomènes rares mais significatifs.

Convergence Presque Sûre des Statistiques d'Ordre

Les statistiques d'ordre convergent presque sûrement vers la fonction de répartition réelle $F(x)$. Le **théorème de Glivenko-Cantelli** nous dit que :

$$F_n(x) \xrightarrow{a.s.} F(x) \quad \text{lorsque} \quad n \to \infty.$$

Cela signifie que la fonction de répartition empirique $F_n(x)$ converge vers la vraie fonction de répartition $F(x)$, avec une convergence uniforme, presque sûrement.

Fonctions des Statistiques d'Ordre

Les **fonctions des statistiques d'ordre**, comme les **moyennes** ou les **médianes**, peuvent être étudiées pour en extraire des propriétés intéressantes sur l'échantillon.

- La **médiane** correspond à $X_{(n/2)}$ dans un échantillon de taille n.
- La **moyenne** de $X_{(k)}$ peut être utilisée dans des méthodes de régression robuste.

Les fonctions des statistiques d'ordre sont particulièrement utiles en analyse des données, en estimation de quantiles, et en calcul de mesure d'assimilation.

Définition de la Loi Conditionnelle des Statistiques d'Ordre

Soit $X_1, X_2, \ldots, X_n$ un échantillon de n variables aléatoires indépendantes et identiquement distribuées (i.i.d.), triées en ordre croissant :

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

La statistique d'ordre $X_{(k)}$ est la k-ème plus petite valeur de l'échantillon, et la statistique $X_{(l)}$ est la l-ième plus petite valeur de l'échantillon, avec $k < l$.
La **loi conditionnelle de $X_{(k)}$** étant donné $X_{(l)} = x$ décrit la distribution de $X_{(k)}$, sachant que $X_{(l)}$ a pris la valeur x.

Formule de la Loi Conditionnelle des Statistiques d'Ordre

La fonction de répartition conditionnelle de $X_{(k)}$, sachant $X_{(l)} = x$, est donnée par :

$$P(X_{(k)} \leq x \mid X_{(l)} = x) = \sum_{i=k}^{l} \binom{n}{i} [F(x)]^{i} [1 - F(x)]^{n-i} \cdot [F(x)]^{l-i} \cdot [1 - F(x$$

Où :

- $F(x)$ est la fonction de répartition de chaque variable X_i,
- $\binom{n}{i}$ est le coefficient binomial,
- La somme est effectuée pour i allant de k à l.

Cette formule permet de déterminer la probabilité que $X_{(k)}$ prenne une valeur inférieure ou égale à x, conditionnée par la valeur de $X_{(l)}$.

Interprétation de la Loi Conditionnelle

La loi conditionnelle des statistiques d'ordre nous donne des informations sur la **distribution des quantiles conditionnels**.
Elle permet de comprendre comment une statistique d'ordre comme le **k-ième quantile** se comporte lorsqu'on connaît déjà la valeur d'une autre statistique d'ordre dans l'échantillon.
Par exemple, si nous connaissons la valeur du **maximum** $X_{(n)}$ ou du **minimum** $X_{(1)}$, la loi conditionnelle permet de déterminer la distribution des autres quantiles, comme la médiane ou les quartiles.
Cette approche est particulièrement utile pour l'estimation robuste et dans des modèles de **régression quantile conditionnelle**.