Fonction de Répartition Empirique (FRE) : Théorie et Applications

RAHMANI Naceur

Octobre 2025

Introduction à la Fonction de Répartition Empirique

La **fonction de répartition empirique (FRE)** est une estimation non paramétrique de la fonction de répartition $F(x)$ d'une variable aléatoire basée sur un échantillon de données. Elle est définie par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}},$$

où $\mathbb{1}_{\{X_i \leq x\}}$ est la fonction indicatrice qui vaut 1 si $X_i \leq x$, et 0 sinon. La FRE donne la proportion des observations $X_1, X_2, \ldots, X_n$ inférieures ou égales à x.

Propriétés de la Fonction de Répartition Empirique

Les propriétés suivantes sont importantes pour comprendre le comportement de la FRE :

- **Monotonie** : La FRE est croissante, i.e. $F_n(x_1) \leq F_n(x_2)$ si $x_1 \leq x_2$.
- **Borne inférieure et supérieure** : $0 \leq F_n(x) \leq 1$ pour tout x.
- **Sauts** : La FRE est une fonction en escalier. Elle fait un saut de $\frac{1}{n}$ à chaque valeur observée.
- **Continuité par morceaux** : $F_n(x)$ est continue par morceaux.

Ces propriétés sont liées à la nature discrète de l'échantillon.

Calcul de la Fonction de Répartition Empirique

Le calcul de la fonction de répartition empirique se fait en deux étapes :

1. **Trier l'échantillon** $X_1, X_2, \ldots, X_n$ en ordre croissant : $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$.
2. **Calculer la FRE** : La fonction de répartition empirique est la proportion des éléments de l'échantillon inférieurs ou égaux à x :

$$F_n(x) = \frac{\#\{X_i \leq x\}}{n}$$

Par exemple, si $X = \{3, 7, 2, 5, 6\}$, en triant l'échantillon on obtient $X_{(1)} = 2, X_{(2)} = 3, X_{(3)} = 5, X_{(4)} = 6, X_{(5)} = 7$. Si $x = 5$, alors $F_n(5) = \frac{3}{5} = 0.6$.

Convergence en Probabilité de la FRE

La **loi des grands nombres** garantit que la FRE converge en probabilité vers la fonction de répartition vraie $F(x)$. Formellement :

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

Cela signifie que, à mesure que n devient grand, $F_n(x)$ devient une bonne approximation de $F(x)$.

Convergence Presque Sûre (Théorème de Glivenko-Cantelli)

Le **théorème de Glivenko-Cantelli** stipule que la fonction de répartition empirique $F_n(x)$ converge presque sûrement vers la fonction de répartition vraie $F(x)$:

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

Cela implique que la probabilité que $F_n(x)$ s'écarte de $F(x)$ devient nulle à mesure que n augmente. C'est une convergence plus forte que la convergence en probabilité.

Distance entre $F_n(x)$ et $F(x)$

La **distance** entre la fonction de répartition empirique $F_n(x)$ et la fonction de répartition vraie $F(x)$ peut être mesurée par la **distance de Kolmogorov**, définie comme :

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$

Le **théorème de Glivenko-Cantelli** assure que $D_n \to 0$ presque sûrement à mesure que $n \to \infty$. Cela signifie que la différence maximale entre $F_n(x)$ et $F(x)$ devient négligeable pour des grands n.

Utilisations de la Fonction de Répartition Empirique

La fonction de répartition empirique est utilisée dans plusieurs contextes :

- **Tests de goodness-of-fit** : Comme le test de Kolmogorov-Smirnov, qui compare la fonction de répartition empirique d'un échantillon avec une fonction théorique.
- **Estimation des quantiles** : Comme la médiane, les quartiles, et d'autres percentiles.
- **Méthodes non paramétriques** : Estimation de la fonction de répartition sans faire d'hypothèses sur la distribution sous-jacente.
- **Visualisation** : La **courbe de FRE** permet de visualiser l'adéquation entre l'échantillon et la fonction théorique.

Exemple 1 : Calcul de $F_n(x)$ pour un échantillon

Prenons un échantillon $X = \{3, 7, 2, 5, 6\}$.

1. **Trier l'échantillon** :
   $X_{(1)} = 2, X_{(2)} = 3, X_{(3)} = 5, X_{(4)} = 6, X_{(5)} = 7$.
2. **Calculer $F_n(5)$** : Le nombre d'éléments inférieurs ou égaux à 5 est 3. Donc, $F_n(5) = \frac{3}{5} = 0.6$.

Exemple 2 : Test de Kolmogorov-Smirnov

Le test de **Kolmogorov-Smirnov** compare la fonction de répartition empirique $F_n(x)$ à une fonction de répartition théorique $F(x)$. Si la **distance de Kolmogorov** D_n est suffisamment petite, on accepte l'hypothèse que l'échantillon provient de la distribution théorique.

Conclusion

La fonction de répartition empirique est un outil fondamental en statistiques, notamment pour l'estimation de la fonction de répartition d'un échantillon. Elle est utilisée dans les tests de goodness-of-fit, l'estimation des quantiles, et dans les méthodes non paramétriques. Le théorème de Glivenko-Cantelli assure que la FRE converge presque sûrement vers la fonction de répartition réelle à mesure que la taille de l'échantillon augmente.