DEVOIR MAISON V

ECG2 MATHS APPLIQUÉES

HEC/ESSEC 2023 Sujet 0 Énoncé 2

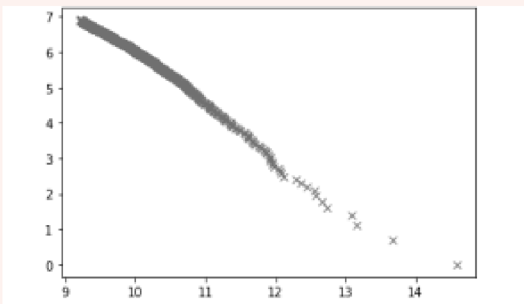
Loi de Pareto-Zipf. On souhaite modéliser la loi de probabilité de la variable aléatoire T qui, à une ville, choisie au hasard parmi les villes françaises, associe l'effectif de sa population. On note n le nombre de villes.

1. a. Pour 2018, on dispose d'un fichier Data1.csv de données provenant de l'INSEE et on utilise la bibliothèque pandas. On exécute les instructions suivantes :

```
import pandas as pd
import numpy.random as rd
import numpy as np
import matplotlib.pyplot as plt

dataset = pd.read_csv('Data1.csv', sep =";")
donnees = dataset [["Libelle","Population"]]
donnees = donnees.sort_values(by = ["Population","Libelle"], ascending =
    False, ignore_index =True)

logPopu = [np.log(t) for t in donnees["Population"]]
logRang = [np.log(k) for k in range (1,lenPopu)+1)]

plt.plot(logPopu, logRang, '*')
```

ce qui produit le graphique suivant :

Graphique



Expliquer pourquoi le programme et le graphique précédent justifient qu'il existe deux réels a et b tels que, pour toute ville, le réel $\frac{a}{t^b}$ où t est l'effectif de celle-ci, est une approximation raisonnable du rang de celle ci dans la liste des villes classées par ordre décroissant de population ?

Date: 26 Février 2024.
http://louismerlin.fr.

Quelle grandeur peut-on calculer pour confirmer ce que l'on constate graphiquement ? Quelle méthode peut-on utiliser pur obtenir le meilleur couple (a, b) en un certain sens ?

b. On suppose que l'on a pour toutes les villes $r = \frac{a}{t^b}$, r étant le rang de cette ville, t son effectif, a et b deux réels identiques pour toutes les villes. Si x est l'effectif d'une des villes françaises, quelle est, en fonction de a, b, x et n, la proportion de villes dont la population urbaine est supérieure ou égale à x ?

• On suppose désormais que T suit la loi de Pareto de paramètre $\theta > 1$ et $x_0 > 0$, c'est-à-dire qu'elle admet pour densité la fonction f définie par

$$f(x) = \begin{cases} \theta \frac{x_0^\theta}{x^{\theta+1}} & \text{si } x \geqslant x_0 \\ 0 & \text{sinon} \end{cases} .$$

On note $\mathrm{Par}(\theta, x_0)$ cette loi.

2. **a.** Déterminer la fonction de répartition F de T.

b. En calculant $\mathbb{P}([T > x])$, montrer que ce résultat est cohérent avec le résultat de la question **1** et exprimer x_0 et θ en fonction de a, b et n.

c. Montrer que $\mathbb{E}(T)$ existe et vaut $\frac{\theta}{\theta-1} x_0$.

3. Soit $x \geqslant x_0$, on note $\mathbb{P}_x$ la probabilité conditionnelle sachant $[T > x]$.

a. Montrer que pour tout $t \geqslant x \geqslant x_0$, $\mathbb{P}_x(T > t) = \left(\frac{x}{t}\right)^\theta$.

b. On pose pour tout $t \in \mathbb{R}$, $F_x(t) = \mathbb{P}_x([T \leqslant t])$. De quelle loi F_x est-elle la fonction de répartition ? Quelle est alors l'espérance de cette loi ?

4. Soit $\delta \in]1, +\infty[$

• On suppose que Y est une variable aléatoire à valeurs dans $[x_0, +\infty[$ dont f_Y est une densité, continue sur $[x_0, +\infty[$, F_Y la fonction de répartition. On suppose que pour tout $x \geqslant x_0$, $\mathbb{P}([Y \geqslant x]) > 0$.

a. Soit $x \geqslant x_0$. Montrer que la fonction G_x définie sur $\mathbb{R}$ par $G_x : t \mapsto \mathbb{P}_{[Y \geqslant x]}(Y \leqslant t)$ est une fonction de répartition d'une variable aléatoire à densité dont on déterminera une densité.

• On suppose que pour tout $x \geqslant x_0$, l'espérance d'une variable aléatoire de fonction de répartition G_x existe et vaut δx.

b. En déduire que pour tout $x \geqslant x_0$, $(\delta - 1) x f_Y(x) = \delta(1 - F_Y(x))$.

c. Résoudre l'équation différentielle $(1 - \delta) x y' - \delta y = 0$ sur $[x_0, +\infty[$.

d. En déduire que Y suit une loi de Pareto dont on précisera les paramètres.