

CORRECTION DU DEVOIR MAISON N° 8

```

1. a) def NbDiff(L):
    X=[ ]
    for l in L:
        if l not in X:
            X.append(l)
    return len(X)

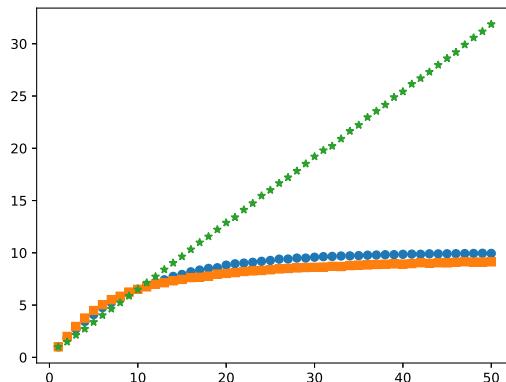
b) from random import randint

def Z(N,k):
    return NbDiff([randint(1,N) for _ in range(k)])

c) def espZ(N,k):
    L=[Z(N,k) for _ in range(1000)]
    return sum(L)/1000

import matplotlib.pyplot as mp
X=[i for i in range(1,51)]
Y1=[espZ(10,x) for x in X]
Y2=[espZ(x,10) for x in X]
Y3=[espZ(x,x) for x in X]
mp.plot(X,Y1,'o')
mp.plot(X,Y2,'s')
mp.plot(X,Y3,'*')
mp.show()

```



On remarque que dans les deux premiers cas on a convergence vers une valeur proche de 10 et dans le dernière cas l'espérance semble diverger vers $+\infty$ et être proportionnelle à N .

2. Z_1 est la variable certaine égale à 1 (avec 1 tirage on a 1 numéro « distinct »). Donc $E(Z_1) = 1$.

$Z_2(\Omega) = \{1; 2\}$ et $P(Z_2 = 1) = P(X_1 = X_2) = \sum_{k=1}^N P([X_1 = k] \cap [X_2 = k]) = N \times \frac{1}{N^2}$ car les tirages sont indépendants.

Donc $P(Z_2 = 1) = \frac{1}{N}$ et $P(Z_2 = 2) = 1 - \frac{1}{N}$.

On en déduit que $E(Z_2) = 1 \times \frac{1}{N} + 2 \times \left(1 - \frac{1}{N}\right) = 2 - \frac{1}{N}$.

3. a) De même que pour Z_2 , l'événement $[Z_k = 1]$ signifie que tous les numéros tirés étaient les mêmes. On a donc, par indépendance des tirages :

$$P(Z_k = 1) = \sum_{i=1}^N P([X_1 = i] \cap \dots \cap [X_k = i]) = N \times \frac{1}{N^k} = \frac{1}{N^{k-1}}.$$

L'événement $[Z_k = k]$ signifie qu'on a obtenu que des numéros différents au cours des k premiers tirages. Ceci est évidemment impossible si $k > N$.

$$\text{Si } k \leq N, \text{ on a } P(Z_k = k) = 1 \times \frac{N-1}{N} \times \frac{N-2}{N} \times \dots \times \frac{N-(k-1)}{N} = \frac{N!}{(N-k)!N^k}.$$

- b) Comme $Z_k(\Omega) = \llbracket 1; k \rrbracket$, les événements $([Z_k = i])_{i \in \llbracket 1; k \rrbracket}$ forment un système complet d'événements. D'après la formule des probabilités totales :

$$P(Z_{k+1} = \ell) = \sum_{i=1}^k P(Z_k = i) P_{[Z_k=i]}(Z_{k+1} = \ell).$$

On peut alors remarquer que à l'étape $k+1$ on a soit le même nombre de numéros distincts qu'à l'étape précédente soit 1 numéro distinct de plus.

Donc pour $j \notin \{\ell; \ell-1\}$, $P_{[Z_k=j]}(Z_{k+1} = \ell) = 0$.

$$\text{Et } P_{[Z_k=\ell]}(Z_{k+1} = \ell) = \frac{\ell}{N} \text{ et } P_{[Z_k=\ell-1]}(Z_{k+1} = \ell) = \frac{N-(\ell-1)}{N} \text{ (pour } \ell \geq 2).$$

Donc, pour $\ell \in \llbracket 2; k \rrbracket$:

$$\begin{aligned} P(Z_{k+1} = \ell) &= P(Z_k = \ell) P_{[Z_k=\ell]}(Z_{k+1} = \ell) + P(Z_k = \ell-1) P_{[Z_k=\ell-1]}(Z_{k+1} = \ell) \\ &= \frac{\ell}{N} P(Z_k = \ell) + \frac{N-\ell+1}{N} P(Z_k = \ell-1). \end{aligned}$$

On peut remarquer que la formule fonctionne encore pour $\ell = 1$ car $P(Z_k = 0) = 0$ et $P(Z_{k+1} = 1) = \frac{1}{N^k} = \frac{1}{N} P(Z_k = 1)$.

Pour $\ell = k+1$, la formule fonctionne encore car $P(Z_k = k+1) = 0$ et on connaît les valeurs de $P(Z_{k+1} = k+1)$ et $P(Z_k = k)$ avec la question précédente.

Enfin, pour $\ell > k+1$, la formule fonctionne aussi car

$$P(Z_{k+1} = \ell) = 0 = \frac{\ell}{N} \times 0 + \frac{N-\ell+1}{N} \times 0.$$

La formule donnée est bien vérifiée pour tout $\ell \in \llbracket 1; N \rrbracket$.

- c) On a, par définition de l'espérance :

$$\begin{aligned} E(Z_{k+1}) &= \sum_{\ell=1}^{k+1} \ell P(Z_{k+1} = \ell) \\ &= \sum_{\ell=1}^N \ell P(Z_{k+1} = \ell) && \text{on ajoute des 0} \\ &= \sum_{\ell=1}^N \frac{\ell^2}{N} P(Z_k = \ell) + \sum_{\ell=1}^N \ell \times \frac{N-\ell+1}{N} P(Z_k = \ell-1) \\ &= \sum_{\ell=1}^N \frac{\ell^2}{N} P(Z_k = \ell) + \sum_{\ell=2}^N \ell \times \frac{N-\ell+1}{N} P(Z_k = \ell-1) \\ &= \sum_{\ell=1}^N \frac{\ell^2}{N} P(Z_k = \ell) + \sum_{\ell=1}^{N-1} (\ell+1) \times \frac{N-\ell}{N} P(Z_k = \ell) \\ &= \sum_{\ell=1}^N \frac{\ell^2}{N} P(Z_k = \ell) + \sum_{\ell=1}^N \times \frac{N+(N-1)\ell - \ell^2}{N} P(Z_k = \ell) \end{aligned}$$

on a ajouté 0 dans la deuxième somme

$$\begin{aligned} &= \sum_{\ell=1}^N P(Z_k = \ell) + \frac{N-1}{N} \sum_{\ell=1}^N \ell P(Z_k = \ell) \\ &= 1 + \frac{N-1}{N} E(Z_k). \end{aligned}$$

4. Montrons par récurrence que la propriété $\mathcal{P}(k) : E(Z_k) = N \left(1 - \left(\frac{N-1}{N} \right)^k \right)$ est vraie pour tout $k \geq 1$.

Pour $k=1$, on a $N \left(1 - \left(\frac{N-1}{N} \right)^k \right) = N \times \frac{1}{N} = 1 = E(Z_1)$.

$\mathcal{P}(1)$ est bien vérifiée.

Soit $k \geq 1$ fixé. Supposons $\mathcal{P}(k)$ vraie.

D'après la question précédente :

$$\begin{aligned} E(Z_{k+1}) &= \frac{N-1}{N} \times N \left(1 - \left(\frac{N-1}{N} \right)^k \right) + 1 \\ &= N - \frac{(N-1)^{k+1}}{N^k} + 1 \\ &= N - \frac{(N-1)^{k+1}}{N^k} = N \left(1 - \left(\frac{N-1}{N} \right)^{k+1} \right). \end{aligned}$$

Ainsi $\mathcal{P}(k+1)$ est bien vérifiée.

D'après le principe de récurrence, la propriété est vérifiée pour tout $k \geq 1$.

On aurait aussi pu utiliser la méthode des suites arithmético-géométriques.

5. a) Soit N fixé. On a alors $\lim_{k \rightarrow +\infty} \left(\frac{N-1}{N} \right)^k = 0$ car $-1 < \frac{N-1}{N} < 1$. Donc $E(Z_k) \underset{k \rightarrow +\infty}{\sim} N$.
- b) Soit k fixé. On a alors $\lim_{N \rightarrow +\infty} \left(\frac{N-1}{N} \right)^k = \lim_{N \rightarrow +\infty} e^{k \ln(1-1/N)} = 1$. On est face à une forme indéterminée.
Mais on peut remarquer que :

$$1 - \left(\frac{N-1}{N} \right)^k = 1 - e^{k \ln(1-1/N)} \underset{N \rightarrow +\infty}{\sim} -k \ln \left(1 - \frac{1}{N} \right) \underset{N \rightarrow +\infty}{\sim} \frac{k}{N}$$

$$\text{Donc } N \left(1 - \left(\frac{N-1}{N} \right)^k \right) \underset{N \rightarrow +\infty}{\sim} k.$$

- c) En reprenant l'écriture exponentielle comme dans la question précédente on a

$$1 - \left(\frac{N-1}{N} \right)^N = 1 - e^{N \ln(1-1/N)} \rightarrow 1 - e^{-1},$$

$$\text{donc } E(Z_N) \underset{N \rightarrow +\infty}{\sim} N(1 - e^{-1}).$$

Facultatif : la métaphore de la cantine (ENS).

1. a) L'événement $[K_{n+1} = 1]$ signifie que les $n + 1$ convives ont choisi la même table.

Notons Y_i la VAR égale au numéro de la table choisie par le $i^{\text{ème}}$ convive. On a :

$$\begin{aligned}
 q_{n+1,1} &= P(K_{n+1} = 1) = P\left(\bigcup_{j=1}^{+\infty} [Y_1 = j] \cap [Y_2 = j] \cap \dots \cap [Y_{n+1} = j]\right) \\
 &= \sum_{j=1}^{+\infty} P([Y_1 = j] \cap [Y_2 = j] \cap \dots \cap [Y_{n+1} = j]) \\
 &\quad \text{union d'événements disjoints} \\
 &= \sum_{j=1}^{+\infty} P(Y_1 = j) \times P_{[Y_1=j]}(Y_2 = j) \times \dots \times P_{[Y_1=j] \cap \dots \cap [Y_n=j]}(Y_{n+1} = j) \\
 &\quad \text{formule des probabilités composées} \\
 &= \sum_{j=1}^{+\infty} P(Y_1 = j) \times \frac{1}{1+\theta} \times \frac{2}{2+\theta} \dots \times \frac{n}{n+\theta} \\
 &= \frac{n!}{(1+\theta)(2+\theta)\dots(n+\theta)} \sum_{j=1}^{+\infty} P(Y_1 = j) \\
 &= \frac{n!}{(1+\theta)(2+\theta)\dots(n+\theta)}. \\
 &\quad ([Y_1 = j])_{j \in \mathbb{N}^*} \text{ système complet d'événements}
 \end{aligned}$$

On a donc bien
$$q_{n+1,1} = \frac{n!}{(1+\theta)(2+\theta)\dots(n+\theta)}.$$

- b) Soit $i \in \llbracket 2; n \rrbracket$. Comme $K_n(\Omega) = \llbracket 1; n \rrbracket$, $([K_n = j])_{j \in \llbracket 1; n \rrbracket}$ est un système complet d'événements. D'après la formule des probabilités totales

$$P(K_{n+1} = i) = \sum_{j=1}^n P(K_n = j) P_{[K_n=j]}(K_{n+1} = i).$$

Or si $j \notin \{i-1; i\}$, $P_{[K_n=j]}(K_{n+1} = i) = 0$ car il n'y a qu'un seul convive qui arrive en plus entre l'étape n et $n+1$.

De plus, $P_{[K_n=i]}(K_{n+1} = i) = \frac{n}{n+\theta}$ car ici le $(n+1)^{\text{ème}}$ convive a choisi l'une des tables déjà occupé et

$P_{[K_n=i-1]}(K_{n+1} = i) = \frac{\theta}{n+\theta}$ car ici le $(n+1)^{\text{ème}}$ convive a choisi une nouvelle table.

Donc

$$P(K_{n+1} = i) = \frac{n}{n+\theta} P(K_n = i) + \frac{\theta}{n+\theta} P(K_n = i-1).$$

En résumé,
$$q_{n+1,i} = \frac{n}{n+\theta} q_{n,i} + \frac{\theta}{n+\theta} q_{n,i-1}.$$

2. a) Soit $n \in \mathbb{N}^*$.

$$\begin{aligned}
P_{n+1} &= \sum_{i=1}^{n+1} q_{n+1,i} X^i \\
&= q_{n+1,1} X + \frac{n}{n+\theta} \sum_{i=2}^{n+1} q_{n,i} X^i + \frac{\theta}{n+\theta} \sum_{i=2}^{n+1} q_{n,i-1} X^i \\
&= \frac{n}{n+\theta} q_{n,1} X + \frac{n}{n+\theta} \sum_{i=2}^{n+1} q_{n,i} X^i + \frac{\theta}{n+\theta} \sum_{j=1}^n q_{n,j} X^{j+1} \quad \text{question 1.a) et } j = i - 1 \\
&= \frac{n}{n+\theta} P_n + \frac{n}{n+\theta} q_{n,n+1} X^{n+1} + \frac{\theta}{n+\theta} X P_n \\
&= \frac{n+\theta}{n+\theta} P_n \quad \text{car } q_{n,n+1} = 0.
\end{aligned}$$

Ainsi $\boxed{P_{n+1} = \frac{n+\theta}{n+\theta} P_n}$

b) À l'aide d'une récurrence rapide, on peut montrer que

$$P_n = \frac{n-1+\theta X}{n-1+\theta} \times \frac{n-2+\theta X}{n-2+\theta} \times \dots \times \frac{1+\theta X}{1+\theta} P_1.$$

Comme $P_1 = q_{1,1} X = X$ on a donc

$$P_n = \frac{\prod_{i=1}^{n-1} (i+\theta X)}{\prod_{i=1}^{n-1} (i+\theta)} \times X = \frac{\prod_{i=0}^{n-1} (i+\theta X)}{\prod_{i=0}^{n-1} (i+\theta)} = \frac{L_n(\theta X)}{L_n(\theta)}.$$

On a bien $\boxed{P_n = \frac{L_n(\theta X)}{L_n(\theta)}}$.

3. a) K_n est une VAR finie car $K_n(\Omega) = [\![1; n]\!]$. Par définition $\mathbb{E}(K_n) = \sum_{i=1}^n iP(K_n = i)$.

De plus, $P'_n = \sum_{i=1}^n P(K_n = i) \times iX^{i-1}$ donc $P'_n(1) = \sum_{i=1}^n iP(K_n = i)$.

Ainsi, $\boxed{\mathbb{E}(K_n) = P'_n(1)}$.

Notons h la fonction $x \mapsto \ln(P_n(x))$. Comme les $q_{n,i}$ sont des probabilités ce sont des réels positifs (au moins l'un d'entre eux n'est pas nul), la fonction est donc définie au moins sur \mathbb{R}^{+*} .

D'après la question 2.b), pour tout $x \in \mathbb{R}^{+*}$, $h(x) = \sum_{i=0}^{n-1} \ln(\theta x + i) - \ln(L_n(\theta))$.

On a donc, en dérivant, pour tout $x \in \mathbb{R}^{+*}$, $h'(x) = \frac{P'_n(x)}{P_n(x)} = \sum_{i=0}^{n-1} \frac{\theta}{\theta x + i}$.

On en déduit que $P'_n(1) = P_n(1) \times \left(\sum_{i=0}^{n-1} \frac{\theta}{\theta + i} \right) = \sum_{i=0}^{n-1} \frac{\theta}{\theta + i}$, car $P_n(1) = 1$.

En conclusion $\boxed{\mathbb{E}(K_n) = \sum_{i=0}^{n-1} \frac{\theta}{\theta + i}}$.

b) On a tout d'abord $P''_n = \sum_{i=1}^n i(i-1)q_{n,i} X^{i-2}$ donc

$$P''_n(1) = \sum_{i=1}^n i(i-1)q_{n,i} = \sum_{i=1}^n i^2 P(K_n = i) - \sum_{i=1}^n iP(K_n = i) = \mathbb{E}(K_n^2) - \mathbb{E}(K_n).$$

D'après la formule de Koenig-Huygens,

$$\mathbb{V}(K_n) = \mathbb{E}(K_n^2) - \mathbb{E}(K_n)^2 = P_n''(1) + P_n'(1) - (P_n'(1))^2.$$

On a bien $\boxed{\mathbb{V}(K_n) = P_n''(1) + P_n'(1) - (P_n'(1))^2}.$

On reprend notre fonction h de la question précédente et on a

$$h''(x) = \frac{P_n''(x)P_n(x) - (P_n'(x))^2}{(P_n(x))^2} = \sum_{i=0}^{n-1} -\frac{\theta^2}{(\theta x + i)^2}.$$

Donc, en évaluant en $x = 1$, $P_n''(1) - (P_n'(1))^2 = -\sum_{i=0}^{n-1} \frac{\theta^2}{(\theta + i)^2}$.

On en déduit que $\boxed{\mathbb{V}(K_n) = \sum_{i=0}^{n-1} \frac{\theta}{\theta + i} - \sum_{i=0}^{n-1} \frac{\theta^2}{(\theta + i)^2}}.$

4. a) Soit $i \in \llbracket 1; n \rrbracket$. On note ε_i la VAR égale à 1 si le $i^{\text{ème}}$ convive occupe une nouvelle table et 0 sinon.

La variable ε_i suit une loi de Bernoulli de paramètre $\frac{\theta}{i-1+\theta}$ d'après l'énoncé. Cela fonctionne bien pour $i = 1$ car ε_1 est la VAR certaine égale à 1 ce qui correspond bien à une loi de Bernoulli de paramètre 1.

Les variables $(\varepsilon_i)_{1 \leq i \leq n}$ sont indépendantes car l'énoncé précise que les choix successifs des convives se font au hasard et ne dépendent donc pas des choix précédents.

Et enfin on a bien $K_n = \sum_{i=1}^n \varepsilon_i$ car $\sum_{i=1}^n \varepsilon_i$ correspond au nombres de convives qui ont choisi une nouvelle table et donc au final on obtient bien le nombre de tables occupées.

- b) D'après la propriété de linéarité de l'espérance $\mathbb{E}(K_n) = \sum_{i=1}^n \mathbb{E}(\varepsilon_i)$.

Or on sait que $\mathbb{E}(\varepsilon_i) = \frac{\theta}{i-1+\theta}$.

Donc $\boxed{\mathbb{E}(K_n) = \sum_{i=1}^n \frac{\theta}{i-1+\theta} = \sum_{i=0}^{n-1} \frac{\theta}{\theta+i}}.$

Les variables $(\varepsilon_i)_{1 \leq i \leq n}$ étant indépendantes, par propriété de la variance

$$\mathbb{V}(K_n) = \sum_{i=1}^n \mathbb{V}(\varepsilon_i) = \sum_{i=1}^n \frac{\theta}{i-1+\theta} \left(1 - \frac{\theta}{i-1+\theta}\right).$$

$\boxed{\mathbb{V}(K_n) = \sum_{i=0}^{n-1} \frac{\theta}{\theta+i} \left(1 - \frac{\theta}{\theta+i}\right)}.$

5. a) Soit $i \in \llbracket 1; n \rrbracket$.

Pour tout $x \in [i-1; i]$,

$$\frac{\theta}{\theta+i} \leq \frac{\theta}{\theta+x} \leq \frac{\theta}{\theta+i-1}$$

car $\theta > 0$ donc la fonction $x \mapsto \frac{\theta}{\theta+x}$ est décroissante sur $[i-1; i]$.

Par croissance de l'intégrale on obtient que

$$\int_{i-1}^i \frac{\theta}{\theta+i} dx \leq \int_{i-1}^i \frac{\theta}{\theta+x} dx \leq \int_{i-1}^i \frac{\theta}{\theta+i-1} dx \Leftrightarrow \frac{\theta}{\theta+i} \leq \int_{i-1}^i \frac{\theta}{\theta+x} dx \leq \frac{\theta}{\theta+i-1}.$$

On peut réécrire cet encadrement, pour $i \in \llbracket 1; n-1 \rrbracket$ sous la forme :

$$\int_i^{i+1} \frac{\theta}{\theta+x} dx \leq \frac{\theta}{\theta+i} \leq \int_{i-1}^i \frac{\theta}{\theta+x} dx$$

En sommant pour i allant de 1 à $n - 1$ on obtient :

$$\begin{aligned}
 & \sum_{i=1}^{n-1} \int_i^{i+1} \frac{\theta}{\theta + x} dx \leqslant \sum_{i=1}^{n-1} \frac{\theta}{\theta + i} \leqslant \sum_{i=1}^{n-1} \int_{i-1}^i \frac{\theta}{\theta + x} dx \\
 & \Leftrightarrow \int_1^n \frac{\theta}{\theta + x} dx \leqslant \mathbb{E}(K_n) - 1 \leqslant \int_0^{n-1} \frac{\theta}{\theta + x} dx \\
 & \Leftrightarrow \boxed{1 + \int_1^n \frac{\theta}{\theta + x} dx \leqslant \mathbb{E}(K_n) \leqslant 1 + \int_0^{n-1} \frac{\theta}{\theta + x} dx.} \quad \text{relation de Chasles}
 \end{aligned}$$

b) D'après la question précédente

$$1 + \theta(\ln(\theta + n) - \ln(\theta + 1)) \leqslant \mathbb{E}(K_n) \leqslant 1 + \theta(\ln(\theta + n - 1) - \ln(\theta)).$$

Pour $n > 1$, comme $\ln(n) > 0$, on en déduit que

$$\frac{1}{\ln(n)} + \theta \left(\frac{\ln(\theta + n)}{\ln(n)} - \frac{\ln(\theta + 1)}{\ln(n)} \right) \leqslant \frac{\mathbb{E}(K_n)}{\ln(n)} \leqslant \frac{1}{\ln(n)} + \theta \left(\frac{\ln(\theta + n - 1)}{\ln(n)} - \frac{\ln(\theta)}{\ln(n)} \right).$$

On peut remarquer que $\ln(\theta + n) = \ln(n) + \ln\left(1 + \frac{\theta}{n}\right)$ et $\ln(\theta + n - 1) = \ln(n) + \ln\left(1 + \frac{\theta - 1}{n}\right)$.

Donc $\lim_{n \rightarrow +\infty} \frac{\ln(\theta + n)}{\ln(n)} = 1$ et $\lim_{n \rightarrow +\infty} \frac{\ln(\theta + n - 1)}{\ln(n)} = 1$.

On a donc $\lim_{n \rightarrow +\infty} \frac{1}{\ln(n)} + \theta \left(\frac{\ln(\theta + n)}{\ln(n)} - \frac{\ln(\theta + 1)}{\ln(n)} \right) = \theta$.

Et $\lim_{n \rightarrow +\infty} \frac{1}{\ln(n)} + \theta \left(\frac{\ln(\theta + n - 1)}{\ln(n)} - \frac{\ln(\theta)}{\ln(n)} \right) = \theta$.

Par encadrement de limites, on a donc $\lim_{n \rightarrow +\infty} \frac{\mathbb{E}(K_n)}{\ln(n)} = \theta$ et donc, étant donné que $\theta \neq 0$,

$$\boxed{\mathbb{E}(K_n) \underset{n \rightarrow +\infty}{\sim} \theta \ln(n).}$$

6. On a $\mathbb{V}(K_n) - \mathbb{E}(K_n) = - \sum_{i=0}^{n-1} \frac{\theta^2}{(\theta + i)^2}$.

Or $\frac{\theta^2}{(\theta + i)^2} \underset{i \rightarrow +\infty}{\sim} \frac{\theta^2}{i^2}$ et on sait que la série $\sum_{i \geqslant 1} \frac{1}{i^2}$ est convergente.

Donc, d'après le critère d'équivalence pour les séries à termes positifs, $\sum \frac{\theta^2}{(\theta + i)^2}$ est convergente.

On peut donc en déduire que la limite de $\mathbb{V}(K_n) - \mathbb{E}(K_n)$ est un réel, que nous pouvons noter ℓ .

Cela nous permet donc de démontrer que $\lim_{n \rightarrow +\infty} \frac{\mathbb{V}(K_n)}{\mathbb{E}(K_n)} = 1$ car $\frac{\mathbb{V}(K_n)}{\mathbb{E}(K_n)} - 1 = \frac{\mathbb{V}(K_n) - \mathbb{E}(K_n)}{\mathbb{E}(K_n)} \rightarrow 0$.

En conclusion, $\boxed{\mathbb{V}(K_n) \underset{n \rightarrow +\infty}{\sim} \mathbb{E}(K_n) \underset{n \rightarrow +\infty}{\sim} \theta \ln(n)}$