# Loi de Zipf

On note $\zeta$ la fonction zeta de Riemann, définie par $\zeta(x) = \sum_{n=1}^{+\infty} \frac{1}{n^x}$

## Partie I: Loi de Zipf

Pour $s > 1$ et $\lambda \in \mathbb{R}$, on pose : $P_s(\{n\}) = \frac{\lambda}{n^s}$ pour tout $n \in \mathbb{N}^\star$

1. Pour quelle(s) valeur(s), l'application $P_s$ détermine-t-elle une probabilité sur $(\mathbb{N}^\star, \mathcal{P}(\mathbb{N}^\star))$ ?

On prend $\lambda = \frac{1}{\zeta(s)}$ et on considère désormais l'espace probabilisé $(\mathbb{N}^\star, \mathcal{P}(\mathbb{N}^\star), P_s)$

2. Pour $m \in \mathbb{N}^\star$, on introduit l'événement : $A_m = m\mathbb{N}^* = \{km\ ,\ k \in \mathbb{N}^\star\}$
   Exprimer simplement la probabilité de l'événement $A_m$.
3. Soit $m, n \in \mathbb{N}^*$. Montrer que $A_n$ et $A_m$ sont indépendants si, et seulement, si $m \wedge n = 1$
4. **Application :** On note $p_i$ le $i$-ème nombre premier (avec $i \in \mathbb{N}^*$, $p_1 = 2$, $p_2 = 3$...) et $C_n$ l'ensemble des entiers divisibles par aucun des nombres premiers $p_i$, pour $1 \leqslant i \leqslant n$
   (a) Calculer $P_s(C_n)$
   (b) Déterminer $\bigcap_{n\in\mathbb{N}^*} C_n$
   (c) En déduire le développement eulérien de $\zeta$
   $$\lim_{n\to+\infty} \prod_{i=1}^{n} \left(1 - \frac{1}{p_i^s}\right)^{-1} = \zeta(s)$$
   On écrit alors
   $$\prod_{i=1}^{+\infty} \left(1 - \frac{1}{p_i^s}\right)^{-1} = \zeta(s)$$

## Partie II: Variable aléatoire

Soit $s \in ]1, +\infty[$. On se donne une variable aléatoire $X_s$ à valeurs dans $\mathbb{N}^*$ de loi

$$\forall n \in \mathbb{N}^*, \quad P(X_s = n) = \frac{n^{-s}}{\zeta(s)}$$

5. Pour quelles valeurs de $s$, la variable $X_s$ admet une espérance et la calculer
6. Pour quelles valeurs de $s$, la variable $X_s$ admet une variance et la calculer
7. Soit $k \in \mathbb{N}^*$. On note $B_k$ l'événement « $k$ divise $X_s$ ».
   (a) Calculer $P(B_k)$
   (b) Montrer que si $m$ et $n$ sont premiers entre eux, alors $B_n$ et $B_m$ sont indépendants
8. Montrer que la probabilité qu'aucun carré, autre que 1, ne divise $X_s$ est $\frac{1}{\zeta(2s)}$

<u>*Indication*</u> **:** *On pourra utiliser le développement eulérien de* $\zeta$

**LOI DE ZIPF**

## Partie I: Loi de Zipf

1. La condition $P_s(\mathbb{N}^*) = \sum_{n=1}^{+\infty} P_s(\{n\}) = 1$ détermine $\lambda = \frac{1}{\zeta(s)}$ avec $\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}$. Inversement, cette valeur $\lambda = \frac{1}{\zeta(s)}$ convient car on a alors pour tout $n \in \mathbb{N}^*$ et $\sum_{n=1}^{+\infty} P_s(\{n\}) = 1$
2. On peut exprimer $A_m = \{mk \mid k \in \mathbb{N}^*\}$ et donc
$$P_s(A_m) = \sum_{k=1}^{+\infty} P_s(\{mk\}) = \sum_{k=1}^{+\infty} \frac{\lambda}{m^s k^s} = \frac{1}{m^s}$$
3. Soit $m, n \in \mathbb{N}^*$. D'une part, d'après le calcul précédent, $P(A_m)P(A_n) = \frac{1}{(mn)^s}$.
D'autre part $A_n \cap A_m = A_{\mathbf{ppcm}(m,n)}$. Comme $P(A_n \cap A_m) = \frac{1}{\mathbf{ppcm}(m,n)^s}$.
Les événements $A_m$ et $A_n$ sont indépendants si, et seulement, si $\mathbf{ppcm}(m,n) = mn$, c'est-à-dire si $m$ et $n$ sont premiers entre eux
4. **Application :**
   (a) Soit $p_1, \cdots, p_n$ des nombres premiers deux à deux distincts.
$$\bigcap_{k=1}^{n} A_{p_k} = \{m \in \mathbb{N}^* \mid \forall k \in [\![1,n]\!],\ p_k | m\}$$
   Or les $p_k$ étant des nombres premiers deux à deux distincts
$$(\forall k \in [\![1,n]\!],\ p_k | m) \Longleftrightarrow \prod_{k=1}^{n} p_k \mid m$$
   et donc
$$\bigcap_{k=1}^{n} A_{p_k} = A_{\prod_{k=1}^{n} p_k}$$
   On vérifie alors immédiatement
$$P_s\left(\bigcap_{k=1}^{n} A_{p_k}\right) = P_s\left(A_{\prod_{k=1}^{n} p_k}\right) = \frac{1}{\prod_{k=1}^{n} p_k^s} = \prod_{k=1}^{n} \frac{1}{p_k^s} = \prod_{k=1}^{n} P(A_{p_k})$$
   Donc les $(A_{p_k})_{k\in\mathbb{N}^*}$ sont mutuellement indépendants. Il en va donc de même de leurs complémentaires et avec l'écriture $C_n = \bigcap_{i=1}^{n} \overline{A_{p_i}}$, on obtient
$$P_s(C_n) = \prod_{i=1}^{n} (1 - P_s(A_{p_i})) = \prod_{i=1}^{n} \left(1 - \frac{1}{p_i^s}\right)$$
   (b) L'ensemble $\bigcap_{n\in\mathbb{N}^*} C_n$ est l'ensemble des entiers non nuls qui ne sont divisibles par aucun des nombres premiers.
   Évidemment il y'a que 1 dans ce cas, et donc $\bigcap_{n\in\mathbb{N}^*} C_n = \{1\}$
   (c) Comme la suite $(C_n)_{n\in\mathbb{N}^*}$ est décroissante, alors, en utilisant la limite monotone, on a :
$$\lim_{n\to+\infty} P_s(C_n) = P_s\left(\bigcap_{n\in\mathbb{N}^*} C_n\right) = P_s(\{1\}) = \frac{1}{\zeta(s)}$$
   Donc
$$\zeta(s) = \lim_{n\to+\infty} \prod_{i=1}^{n} \left(1 - \frac{1}{p_i^s}\right)^{-1}$$

LOI DE ZIPF

## Partie II: Variable aléatoire

5. $X_s$ admet une espérance si, et seulement, si la SATP $\sum_{n\geqslant 1} nP(X_s = n)$ converge, c'est-à-dire si $s > 2$. Auquel cas

$$\mathbb{E}(X_s) = \sum_{n=1}^{+\infty} \frac{1}{\zeta(s)n^{s-1}} = \frac{\zeta(s-1)}{\zeta(s)}$$

6. $X_s$ admet une variance si, et seulement, si la SATP $\sum_{n\geqslant 1} n^2 P(X_s = n)$ converge, c'est-à-dire si $s > 3$. Auquel cas

$$\mathbb{E}\left(X_s^2\right) = \sum_{n=1}^{+\infty} \frac{1}{\zeta(s)n^{s-2}} = \frac{\zeta(s-2)}{\zeta(s)}$$

Par la formule de Huygens kœnig

$$\mathbb{V}(X) = \mathbb{E}\left(X_s^2\right) - \mathbb{E}(X_s)^2 = \frac{\zeta(s-2)\zeta(s) - \zeta(s-1)^2}{\zeta(s)^2}$$

7. On note $B_k$ l'événement « $k$ divise $X_s$ ».
   (a) On écrit $B_k = \bigcup_{n\in\mathbb{N}^*} [X_s = kn]$, avec les événements $([X_s = kn])_{n\in\mathbb{N}^*}$ sont deux à deux incompatibles, alors d'après la $\sigma$-additivité

$$\mathbb{P}(B_k) = \sum_{n=1}^{+\infty} \mathbb{P}(X = nk) = \sum_{n=1}^{+\infty} \frac{n^{-s}k^{-s}}{\zeta(s)} = \frac{1}{k^s}$$

   (b) Supposons que $m$ et $n$ sont premiers entre eux, alors $B_n \cap B_m$ est l'événement « $m$ et $n$ divisent $X_s$ ». Comme $m$ et $n$ sont premiers entre eux, cela signifie exactement « $mn$ divise $X_s$ ». Donc

$$\mathbb{P}(B_m \cap B_n) = \frac{1}{(mn)^s} = \frac{1}{m^s}\frac{1}{n^s} = \mathbb{P}(B_m)\,\mathbb{P}(B_n)$$

   Ainsi l'indépendance de $B_m$ et $B_n$

8. Soit $C$ l'événement « aucun carré ne divise $X_s$ ». L'événement peut aussi s'écrire « aucun carré de nombre premier ne divise $X_s$» et son contraire « au moins un carré de nombre premier divise $X_s$ ». Alors

$$\overline{C} = \bigcup_{n\in\mathbb{N}^*} B_{p_n^2}, \text{ soit } C = \bigcap_{n\in\mathbb{N}^*} \overline{B_{p_n^2}}$$

Puisque les entiers $p_n^2$ sont premiers entre eux, les événements $B_{p_n^2}$ sont mutuellement indépendants. En utilisant le théorème de la limite monotone et le développement eulérien de la fonction $\zeta$ on arrive à

$$\mathbb{P}(C) = \lim_{n\to+\infty} \mathbb{P}\left(\bigcap_{k=1}^{n} \overline{B_{p_k^2}}\right) = \lim_{n\to+\infty} \prod_{k=1}^{n} \mathbb{P}\left(\overline{B_{p_k^2}}\right) = \lim_{n\to+\infty} \prod_{k=1}^{n} \left(1 - \frac{1}{p_k^{2s}}\right) = \frac{1}{\zeta(2s)}$$