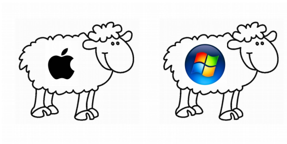
Propagation d'opinion chez les moutons

Outils de MAP : Probabilités conditionnelles, Modélisation, Lois discrètes et continues, Suites récurrentes et matrices, ...
Mots-clés : Renforcement, Sensibilité aux conditions initiales, ...

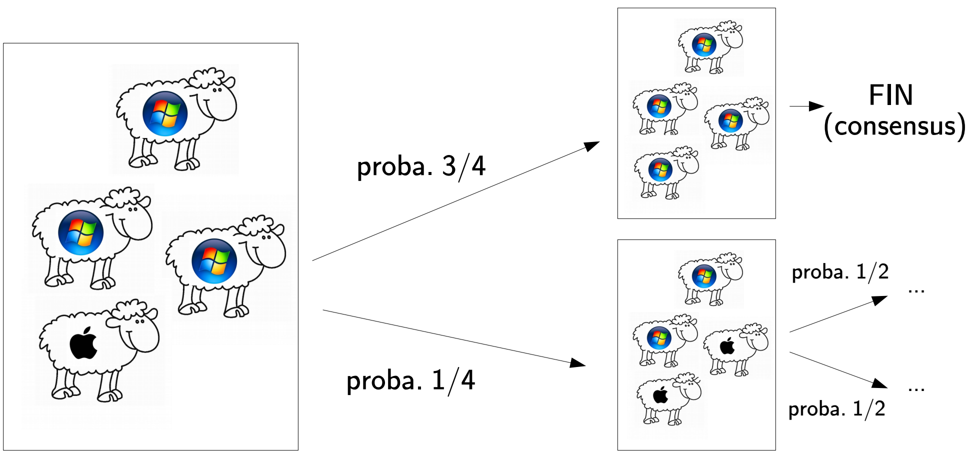
Le but du projet est d'étudier de façon théorique et expérimentale deux modèles simples de propagation d'opinions dans une population de moutons (on considère qu'il y a deux opinions concurrentes : les pro-Mac et pro-Windows). Le but est d'étudier différents aspects du phénomène de renforcement : l'opinion majoritaire a tendance à être renforcée au cours du temps [1].



Partie A : Modèle à population fixe

Soit $N > 1$ fixé, on considère une population de N moutons, chacun est soit pro-Mac soit pro-Windows.

Initialement $0 \leq m \leq N$ moutons sont pro-Mac. A chaque instant $t = 0, 1, 2, ...$, un mouton est choisi uniformément dans la population (indépendamment du passé) et il bêle son opinion. Instantanément, un mouton de l'autre camp (s'il en reste un) change d'opinion. Le processus continue jusqu'à ce qu'un consensus pro-Mac ou pro-Windows soit atteint. Voici un exemple pour $N = 4, m = 1$:



1. Le 2ème modèle a réellement été utilisé en Économie pour décrire une bataille entre deux technologies concurrentes, voir : W.B. Arthur. Self-reinforcing mechanisms in economics. *The economy as an evolving complex system*, vol.5, p.9-31 (1988).

Formellement, le processus $(M(t))_{t\geq 0}$ est donc défini de la façon suivante : $M(0)=m$ et pour $t\geq 0$ si $M(t)=k\notin\{0,N\}$ alors

$$M(t+1)=\begin{cases} k+1 \text{ avec proba. } \frac{k}{N} \\ k-1 \text{ avec proba. } \frac{N-k}{N} \end{cases}.$$

Si jamais $M(t)=0$ (resp. $M(t)=N$) alors $M(t+1)=0$ (resp. $M(t+1)=N$).

On se pose les question suivantes (en fonction de N,m) :

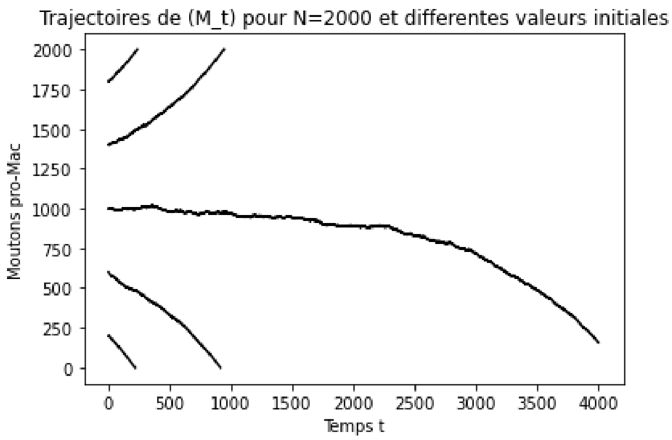
— Que peut-on dire de la proportion de moutons pro-Mac à un instant donné ?

— Quelle est la probabilité d'atteindre le consensus pro-Mac ?

A.1. Quelques simulations

1. **(Code)** Écrire un script python qui trace des simulations de trajectoires $(M(t))_{0\leq t\leq T}$ pour T donné. On choisira les paramètres suivants : $N=2000, T=4000$. Pour l'initialisation m, il est demandé d'afficher (si possible sur le même graphique) une trajectoire partant de chacune des valeurs suivantes : $m=200,600,1000,1400,1800$.

Solution:



A.2. Loi exacte à t fixé

On observe sur les simulations précédentes qu'il y a une forte dépendance en la condition initiale : l'opinion initialement majoritaire a une grande tendance à se répandre dans toute la population. L'objectif des questions qui suivent est de quantifier cette dépendance à l'aide de calculs explicites.

On fixe N pour toute la suite. Pour $k\in\{0,1,\dots,N\}$ et $t\geq 0$ on note

$$p(t,k)=\mathbb{P}(M(t)=k)=\mathbb{P}\left(\text{Au temps } t\text{, exactement } k \text{ moutons sont pro-Mac}\right).$$

On a ainsi

$$p(0,k)=\begin{cases}1 & \text{si } k=m,\\ 0 & \text{sinon}\end{cases}.$$

2. **(Théorie)** Démontrer que pour tout $2\leq k\leq N-2$ et pour $t\geq 1$ on a

$$p(t,k)=\frac{k-1}{N}p(t-1,k-1)+\frac{N-(k+1)}{N}p(t-1,k+1).$$

3. **(Théorie)** Trouver les formules analogues pour $k \in \{0, 1, N-1, N\}$: écrire $p(t,k)$ en fonction de $p(t-1,0), p(t-1,1), \dots, p(t-1,N)$. (Pour cette question il n'est pas besoin de justifier.)
4. **(Théorie)** En déduire qu'il existe une matrice $Q = (q_{k,k'})_{0 \le k,k' \le N}$ de taille $(N+1) \times (N+1)$ telle que pour tout $t \geq 1$:

$$\Big(p(t,0), p(t,1), \dots, p(t,N)\Big) = \Big(p(t-1,0), p(t-1,1), \dots, p(t-1,N)\Big) \times Q.$$

Solution:

2. L'idée est de conditionner par rapport à ce qu'il s'est passé au dernier instant. Si $M(t) = k$ alors nécessairement on a $M(t-1) \in \{k-1, k+1\}$ et on peut donc écrire

$$\begin{aligned}
\mathbb{P}(M(t) = k) &= \mathbb{P}(M(t) = k | M(t-1) = k-1) \times \mathbb{P}(M(t-1) = k-1)) \\
&\quad + \mathbb{P}(M(t) = k | M(t-1) = k+1) \times \mathbb{P}(M(t-1) = k+1)) \\
&= \mathbb{P}(\text{On tire un pro-Mac } | M(t-1) = k-1) \mathbb{P}(M(t-1) = k-1)) \\
&\quad + \mathbb{P}(\text{On tire un pro-Windows } | M(t-1) = k+1) \mathbb{P}(M(t-1) = k+1)) \\
&= \frac{k-1}{N} p(t-1, k-1) + \frac{N-(k+1)}{N} p(t-1, k+1).
\end{aligned}$$

3. On a

$$\begin{aligned}
p(t,0) &= p(t-1,0) + \frac{N-1}{N} p(t-1,1), \\
p(t,1) &= \frac{N-2}{N} p(t-1,2), \\
p(t,N-1) &= \frac{2}{N} p(t-1,N-2), \\
p(t,N) &= p(t-1,N) + \frac{1}{N} p(t-1,N-1)
\end{aligned}$$

4. Si l'on combine tout on obtient

$$\Big(p(t,0), p(t,1), \dots, p(t,N)\Big) = \Big(p(t-1,0), p(t-1,1), \dots, p(t-1,N)\Big) \times Q.$$

avec

$$Q = \begin{array}{c} 0 \\ 1 \\ \vdots \\ k-1 \\ k \\ k+1 \\ \vdots \\ N-1 \\ N \end{array}
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{N-1}{N} & 0 & \frac{1}{N} & 0 & 0 & 0 & 0 \\
 & & \ddots & & & & \\
0 & \frac{N-(k-1)}{N} & 0 & \frac{k-1}{N} & 0 & 0 & 0 \\
0 & 0 & \frac{N-k}{N} & 0 & \frac{k}{N} & 0 & 0 \\
0 & 0 & 0 & \frac{N-(k+1)}{N} & 0 & \frac{k+1}{N} & 0 \\
 & & & & \ddots & & \\
0 & 0 & 0 & 0 & \frac{1}{N} & 0 & \frac{N-1}{N} \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}.$$

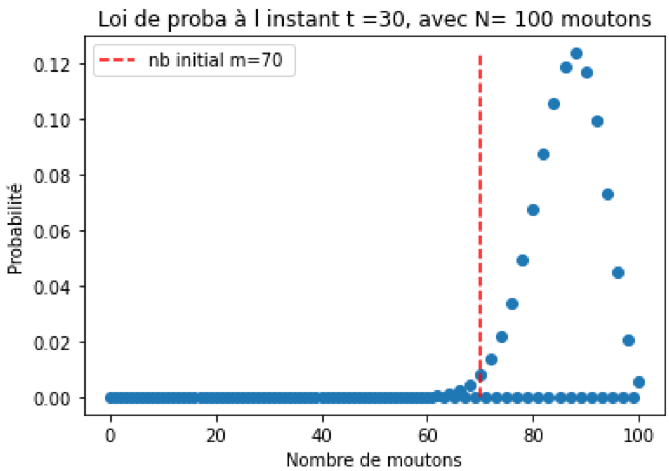
5. **(Code)** Ecrire une fonction python qui prend comme paramètres N et génère la matrice Q. Afficher Q pour $N = 6$.
6. **(Code)** Utiliser les questions précédentes pour écrire une fonction python qui calcule :
$$(t, k, N, m) \mapsto p(t, k).$$
Pour les paramètres $N = 100, t = 30, m = 70$, tracer la courbe $k \mapsto p(t, K)$.
(Pour vérifier cotre code, je trouve $p(30, 86) = 0.1190564\ldots$ *avec* $N = 100, m = 70$*.)*

Solution:

5. On trouve

```
[[1.    0.    0.    0.    0.    0.    0.   ]
 [0.833 0.    0.167 0.    0.    0.    0.   ]
 [0.    0.667 0.    0.333 0.    0.    0.   ]
 [0.    0.    0.5   0.    0.5   0.    0.   ]
 [0.    0.    0.    0.333 0.    0.667 0.   ]
 [0.    0.    0.    0.    0.167 0.    0.833]
 [0.    0.    0.    0.    0.    0.    1.   ]]
```

6.



A.3. Probabilité de consensus

On peut démontrer (et on admet) que si l'on attend suffisamment longtemps alors presque-sûrement on atteint le consensus, soit pour Mac soit pour Windows. On définit donc

$$\alpha(m) = \mathbb{P}(\exists t, M(t) = N) = \mathbb{P}(\text{Le consensus est finalement atteint pour Mac})$$
$$\beta(m) = \mathbb{P}(\exists t, M(t) = 0) = \mathbb{P}(\text{Le consensus est finalement atteint pour Windows}).$$

Grâce à la remarque précédente on a $\alpha(m) + \beta(m) = 1$.

7. **(Théorie)** Démontrer que
$$\alpha(m) = \lim_{t \to +\infty} p(t, N) \tag{1}$$
$$\beta(m) = \lim_{t \to +\infty} p(t, 0).$$

(Indication : Utiliser la Proposition 2.12 dans le Polycopié de MAP361.)

8. **(Théorie)** Votre fonction python de la question 6 permet donc de calculer une approximation de $\alpha(m)$, en prenant t assez grand. Donner une méthode numérique permettant de calculer $\alpha(m)$ à 10^{-5} près.
   (Indication : que peut-on dire de la suite $(p(t,N)+p(t,0))_{t\geq 0}$?)
9. **(Code)** Tracer avec python les probas $m \mapsto \alpha(m)$ pour $N = 30$ (si vous n'avez pas résolu la question 8 vous pouvez considérer que $t = N^2$ est suffisant dans l'équation (1) pour que l'approximation soit correcte).
   (Pour vérifier votre code : je trouve $\alpha(12) = 0.1324654....$)

Solution:

7. On a donc

$$\mathbb{P}(M_(t) = N) = \mathbb{P}(\text{ Le consensus est atteint pour Mac avant l'instant } t).$$

La suite d'événements $t \mapsto \{M(t) = N\}$ est une suite croissante :

$$\big\{M(t) = N\big\} \subset \big\{M(t+1) = N\big\}.$$

(En effet, le processus n'évolue plus lorsqu'il y a consensus.) Ainsi en utilisant la Proposition 2.12 dans le Polycopié de MAP361 on obtient

$$\begin{aligned}\lim_{t\to+\infty} \mathbb{P}(M(t) = N) &= \mathbb{P}(\cup_t\{M(t) = N\})\\ &= \mathbb{P}(\exists t,\ \text{Le consensus est atteint pour Mac avant l'instant } t)\\ &= \alpha(m).\end{aligned}$$

8. On a (par le même argument que la question précédente) que la suite $t \mapsto p(t,N) + p(t,0)$ est croissante en t, et tend vers $\alpha(m) + \beta(m) = 1$. Pour $\varepsilon > 0$, soit t_0 tel que si $t \geq t_0$ alors

$$p(t,N) + p(t,0) \geq 1 - \varepsilon.$$

On a $p(t,0) \leq \beta(m) = 1 - \alpha(m)$. Pour $t \geq t_0$ on a donc

$$0 \leq \alpha(m) - p(t,N) \leq 1 - p(t,0) - p(t,N) \leq \varepsilon$$

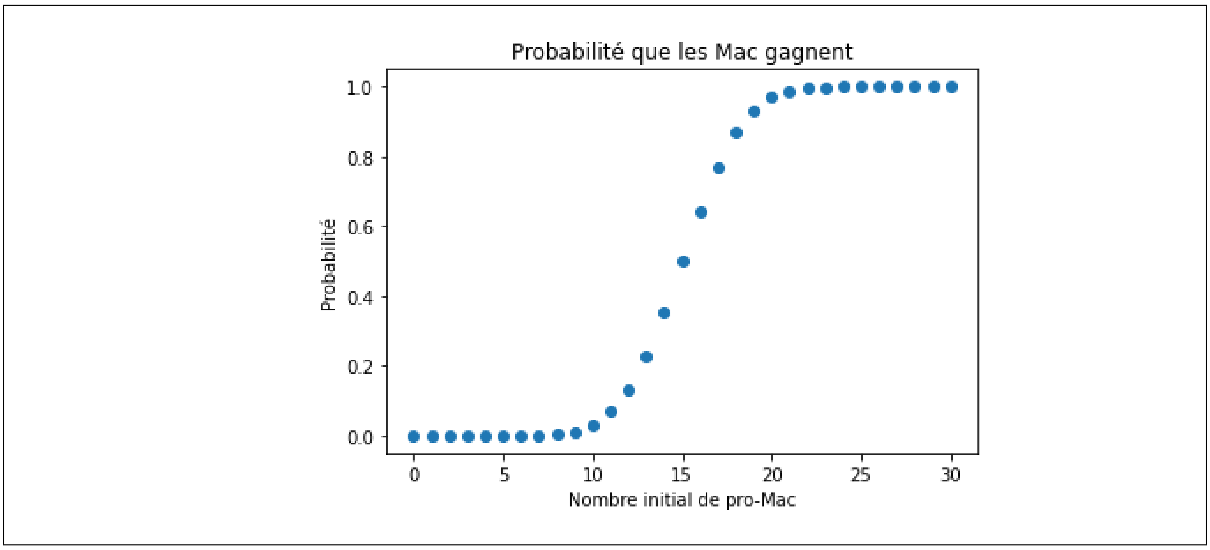
et donc

$$|\alpha(m) - p(t,N)| \leq \varepsilon.$$

Donc l'algorithme suivant renvoie une approximation de $\alpha(m)$ à ε près :

$t = 1$
TantQue $p(t,N) + p(t,0) < 1 - \varepsilon$:
  Augmenter t (par exemple, $t \leftarrow 2t$)
Renvoyer $p(t,N)$.

9.



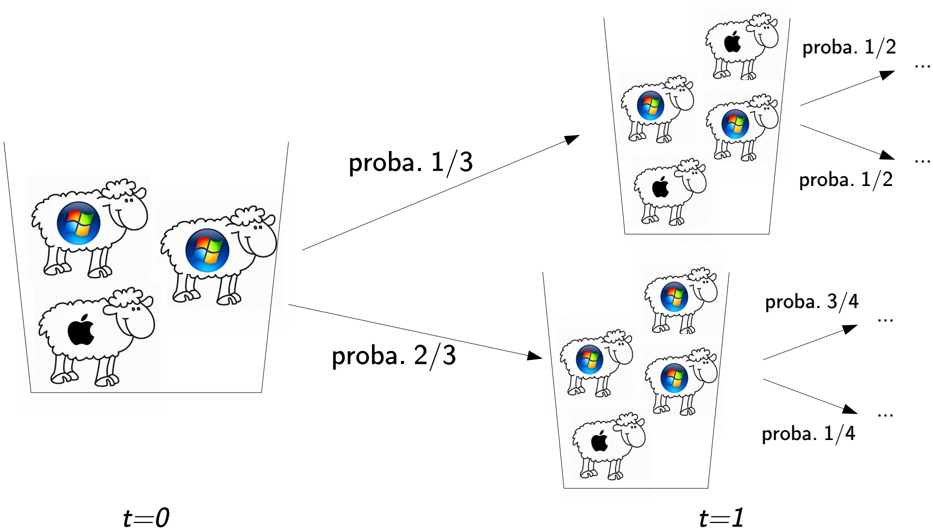
Partie B : Modèle avec population qui augmente

On considère cette fois une population constituée initialement de m moutons pro-Mac et w moutons pro-Windows. À chaque instant $t = 1, 2, ...$, un nouveau mouton arrive dans la population. Pour déterminer son opinion il choisit uniformément au hasard un mouton déjà présent et choisit la même opinion [2].

On note $Z(0) = m, W(0) = w$ et pour $t \geq 1$ on note $Z(t)$ (resp. $W(t)$) le nombre de moutons pro-Mac (resp. pro-Windows) immédiatement après l'instant t, c'est-à-dire lorsque le t-ème nouveau mouton est arrivé. On a pour tout t

$$Z(t) + W(t) = m + w + t.$$

Voici un schéma du début du processus, lorsque $m = 1, w = 2$:



Le processus $(Z(t))_{t\geq 0}$ est donc défini de la façon suivante : $Z(0) = m$ et pour $t \geq 0$

2. Dans ce modèle les moutons ne changent jamais d'opinion.

si $Z(t) = k$ et $W(t) = k'$ alors

$$Z(t+1) = \begin{cases} k+1 \text{ avec proba. } \frac{k}{k+k'} \\ k \text{ avec proba. } \frac{k'}{k+k'} \end{cases}.$$

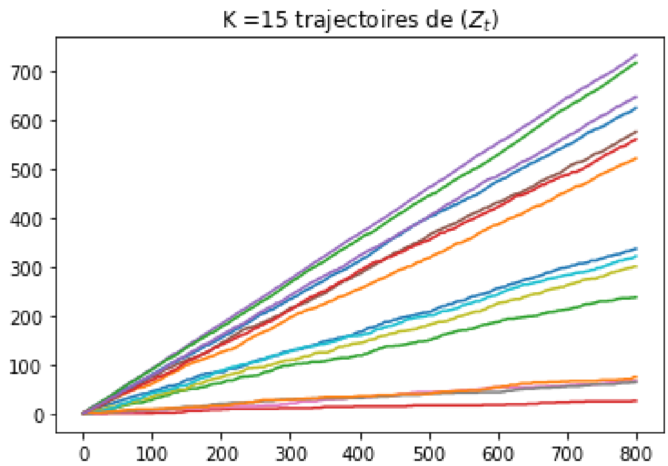
B.1. Simulations de trajectoires

On fixe pour l'instant $m = w = 1$. On va s'intéresser au processus $(Z(t))_{t\geq 0}$.

1. **(Code)** Écrire un script python qui trace (sur le même graphique) 15 trajectoires $(Z(t))_{0\leq t\leq T}$, pour $T = 800$.

Solution:

1.



B.2. Cas $m = w = 1$: uniformité

Sur vos simulations on doit normalement observer que sur chaque trajectoire $(Z(t))_t$ augmente à peu près linéairement avec t. On va donc étudier le comportement de la variable aléatoire $Z(t)/t$.

2. **(Code)** Toujours dans le cas $m = w = 1$, écrire un script python qui réalise K tirages de la variable aléatoire $\frac{Z(T)}{T}$ et affiche les résultats dans un histogramme. On prendra $K = 10000$, $T = 500$ et on représentera des histogrammes à 20 bâtons. Avec `matplotlib` un histogramme des valeurs `Donnees` avec 20 bâtons s'affiche avec

```
plt.hist(Donnees, bins=20,ec='black')
```

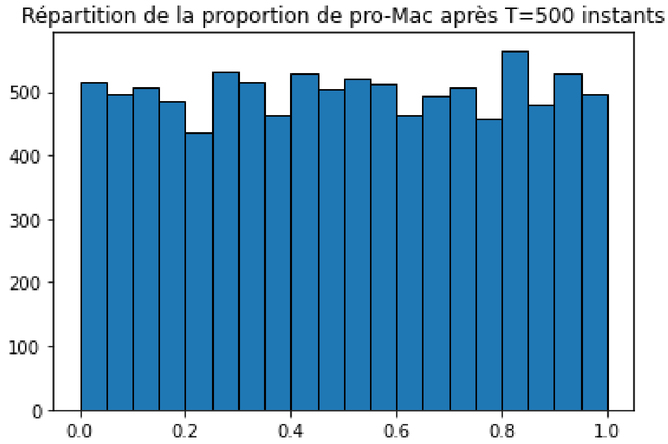
(le `ec='black'` dessine les bords des bâtons).

3. **(Théorie)** Sur votre histogramme on doit normalement observer que la variable aléatoire $\frac{Z(T)}{T}$ est répartie à peu près uniformément dans $[0, 1]$. Démontrer par récurrence que, en effet, on a pour tout $t \geq 0$ que la variable aléatoire $Z(t)$ est uniforme dans l'ensemble

$$\{1, 2, 3, \ldots, t+1\}.$$

Solution:

2.



3. Pour $t = 0$ c'est vrai si l'on part de $m = w = 1$ puisque $Z(0)$ est uniforme dans $\{1\}$. Supposons que cela est vrai à l'instant t : on a donc

$$\mathbb{P}(Z(t) = 1) = \mathbb{P}(Z(t) = 2) = \cdots = \mathbb{P}(Z(t) = t+1) = \frac{1}{t+1}.$$

Montrons l'hérédité : pour $k \in \{2, \ldots, t+2\}$,

$$\begin{aligned}
\mathbb{P}(Z(t+1) = k) &= \mathbb{P}(Z(t+1) = k \cap Z(t) = k-1) + \mathbb{P}(Z(t+1) = k \cap Z(t) = k) \\
&= \mathbb{P}(\text{On tire un pro-Mac } |Z(t) = k-1)\mathbb{P}(Z(t) = k-1)) \\
&+ \mathbb{P}(\text{On tire un pro-Windows } |Z(t) = k)\mathbb{P}(Z(t) = k)) \\
&= \frac{k-1}{t+2}\mathbb{P}(Z(t) = k-1)) + \frac{t+2-k}{t+2}\mathbb{P}(Z(t) = k)) \\
&= \frac{k-1}{t+2}\frac{1}{t+1} + \frac{t+2-k}{t+2}\frac{1}{t+1} \\
&= \frac{t+1}{(t+2)(t+1)} = \frac{1}{t+2}.
\end{aligned}$$

Il reste le cas $k = 1$ que l'on traite facilement puisque $Z(t+1) \in \{1, 2, 3, \ldots, t+2\}$:

$$\mathbb{P}(Z(t+1) = 1) = 1 - \sum_{k \geq 1} \mathbb{P}(Z(t+1) = k) = 1 - \frac{t+1}{t+2} = \frac{1}{t+2}.$$

Finalement nous avons bien prouvé que $Z(t+1)$ est uniforme dans

$$\{1, 2, 3, \ldots, t+2\}.$$

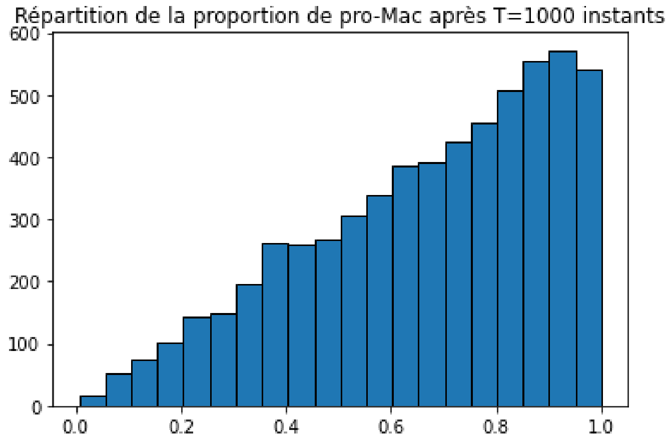
B.3. Cas $m, w \neq 1$: non-uniformité

4. **(Code)** On prend maintenant $m = 2$, $w = 1$, écrire un script python qui réalise K tirages de la variable aléatoire $\frac{Z(T)}{T}$ et affiche les résultats dans un histogramme.

On prendra également $K = 10000$, $T = 500$ et on représentera des histogrammes à 20 bâtons.

5. **(Théorie)** Décrivez et commentez l'histogramme. En particulier, est-ce que le résultat est intuitif ?

Solution:

4. 

5. L'histogramme est clairement non-uniforme. Les moutons pro-Mac sont très fortement favorisés, ce qui est cohérent avec la condition initiale en faveur des pro-Mac.

B.4. Bonus : Une question théorique

Ces dernières questions bonus sont bien plus difficiles et ne seront pas notées très généreusement, elles ne sont destinées qu'aux fans de moutons et de produits infinis.

Les valeurs initiales m, w sont maintenant quelconques. Le but est de démontrer le résultat suivant (apparemment évident) : presque-sûrement la population de moutons pro-Mac ne cesse jamais d'augmenter, autrement dit :

$$\mathbb{P}\left(\exists t \text{ tel que } Z(t) = Z(t+1) = Z(t+2) = Z(t+3) = \dots\right) = 0.$$

6. **(Théorie)** Calculer pour tout $t < t'$ et tout $\mu \in \{1, 2, \dots\}$ la probabilité

$$\mathbb{P}\left(Z(t) = Z(t+1) = \cdots = Z(t') \mid Z(t) = \mu\right).$$

7. **(Théorie)** En déduire une majoration de

$$\mathbb{P}\left(Z(t) = Z(t+1) = \cdots = Z(t')\right).$$

et démontrer que pour tout t

$$\mathbb{P}\left(Z(t) = Z(t+1) = Z(t+2) = Z(t+3) = \dots\right) = 0.$$

8. **(Théorie)** Conclure.
(Indication : Utiliser la Proposition 2.13 du Polycopié de MAP361.)

Solution:

6. On a

$$\mathbb{P}\left(Z(t) = Z(t+1) = \cdots = Z(t') \mid Z(t) = \mu\right) = \left(1 - \tfrac{\mu}{t+2}\right)\times\left(1 - \tfrac{\mu}{t+3}\right)\times \dots \left(1 - \tfrac{\mu}{t'+1}\right).$$

   En effet si l'événement de gauche est réalisé alors on ne tire jamais l'un des μ moutons pro-Mac.

7. On a donc

$$\begin{aligned}
\mathbb{P}\left(Z(t) = Z(t+1) = \cdots = Z(t')\right) &= \sum_{\mu} \mathbb{P}\left(Z(t) = Z(t+1) = \cdots = Z(t') \mid Z(t) = \mu\right) \\
&\quad \times \mathbb{P}(Z(t) = \mu) \\
&\leq \left(1 - \tfrac{1}{t+2}\right) \times \left(1 - \tfrac{1}{t+3}\right) \times \dots \left(1 - \tfrac{1}{t'+1}\right) \\
&\leq \exp\left(\sum_{s=t+2}^{t'+1} \log(1 - \tfrac{1}{s})\right) \\
&\leq \exp\left(\sum_{s=t+2}^{t'+1} -\frac{1}{s}\right) \qquad \text{(car } \log(1-u) \leq u\text{)}
\end{aligned}$$

   qui tend vers zéro quand $t' \to 0$. Et puisque pour tout t'

$$\mathbb{P}\left(Z(t) = Z(t+1) = \dots\right) \leq \mathbb{P}\left(Z(t) = Z(t+1) = \cdots = Z(t')\right),$$

   alors le terme de gauche vaut zéro.

8. Finalement, en utilisant la Proposition 2.13 du Polycopié de MAP361 :

$$\mathbb{P}\left(\exists t \text{ tel que } Z(t) = Z(t+1) = \dots\right) \leq \sum_{t} \mathbb{P}\left(Z(t) = Z(t+1) = \dots\right) = 0.$$