De la longueur des chemins de recherche des skip graphs aléatoires binaires

Philippe Duchon, Hubert Larchevêque

Université de Bordeaux, INRIA Bordeaux Sud-Ouest

1. Skip lists
2. Skip graphs
3. Longueur de chemin de recherche dans un skip graph
4. Conclusion

Plan

Exemple ($p = 1/2$)

Une skip-list est basée sur une liste ordonnée d'éléments. Chaque élément, à chaque niveau atteint, tire un booléen avec probabilité p d'obtenir un 1. Dans ce cas, il apparait au niveau supérieur.

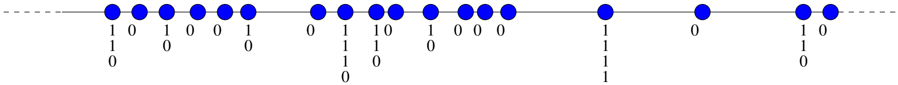


Fig.: Une skip list de paramètre $p = \frac{1}{2}$

Exemple ($p = 1/2$)

Une skip-list est basée sur une liste ordonnée d'éléments. Chaque élément, à chaque niveau atteint, tire un booléen avec probabilité p d'obtenir un 1. Dans ce cas, il apparait au niveau supérieur.

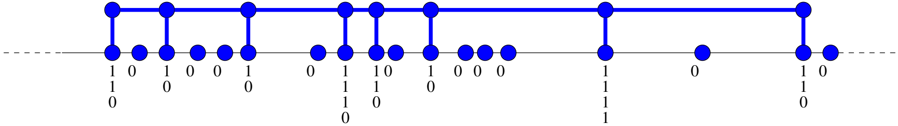


Fig.: Une skip list de paramètre $p = \frac{1}{2}$

Exemple ($p = 1/2$)

Une skip-list est basée sur une liste ordonnée d'éléments. Chaque élément, à chaque niveau atteint, tire un booléen avec probabilité p d'obtenir un 1. Dans ce cas, il apparait au niveau supérieur.

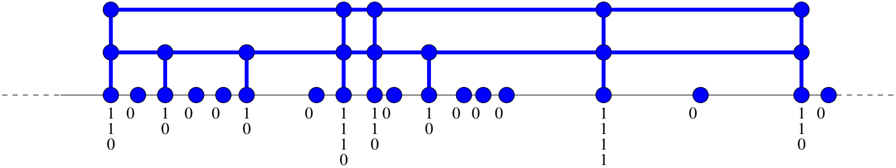


Fig.: Une skip list de paramètre $p = \frac{1}{2}$

Exemple ($p = 1/2$)

Une skip-list est basée sur une liste ordonnée d'éléments. Chaque élément, à chaque niveau atteint, tire un booléen avec probabilité p d'obtenir un 1. Dans ce cas, il apparait au niveau supérieur.

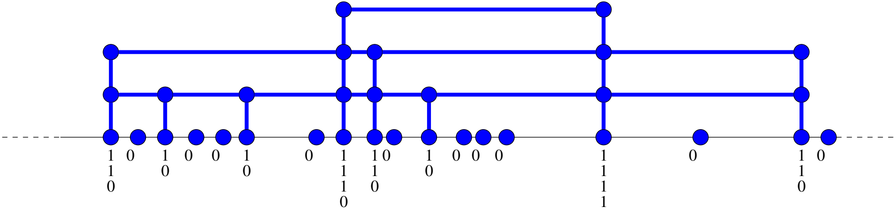


Fig.: Une skip list de paramètre $p = \frac{1}{2}$

Exemple ($p = 1/2$)

Une skip-list est basée sur une liste ordonnée d'éléments. Chaque élément, à chaque niveau atteint, tire un booléen avec probabilité p d'obtenir un 1. Dans ce cas, il apparait au niveau supérieur.

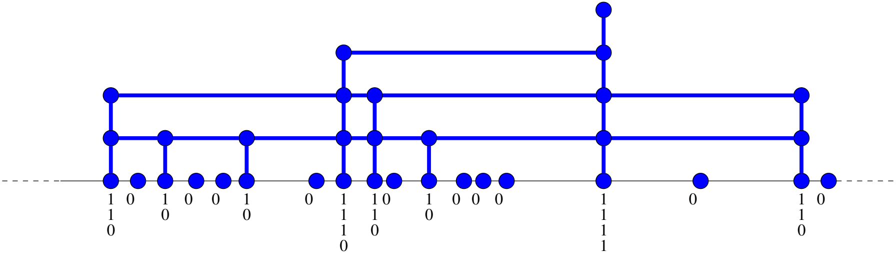
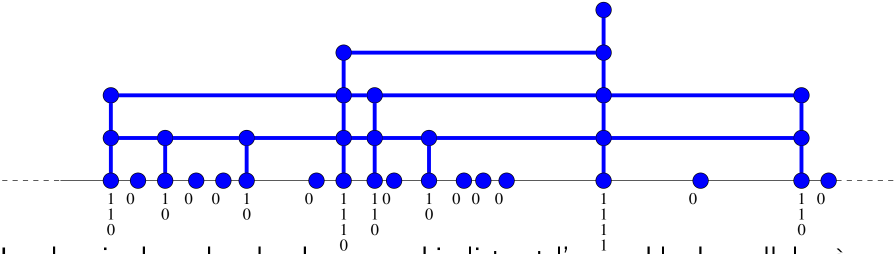


Fig.: Une skip list de paramètre $p = \frac{1}{2}$

Quelques propriétés connues des skip lists

- Nombre de niveaux : $O(\log n)$ w.h.p..
- Hauteur H_n d'une skip list : longueur maximale d'un chemin de recherche pour n'importe quelle clé, depuis le sommet sentinelle. Devroye[1] a prouvé que H_n est d'ordre $\log n$.

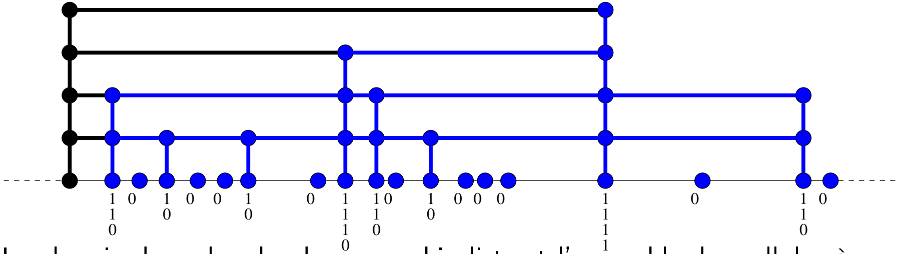


Le chemin de recherche dans une skip list est l'ensemble des cellules à parcourir lors de la recherche d'une clé, depuis le sommet sentinelle.

[1] Luc Devroye. A limit theory for random skip lists.

Quelques propriétés connues des skip lists

- Nombre de niveaux : $O(\log n)$ w.h.p..
- Hauteur H_n d'une skip list : longueur maximale d'un chemin de recherche pour n'importe quelle clé, depuis le sommet sentinelle. Devroye[1] a prouvé que H_n est d'ordre $\log n$.

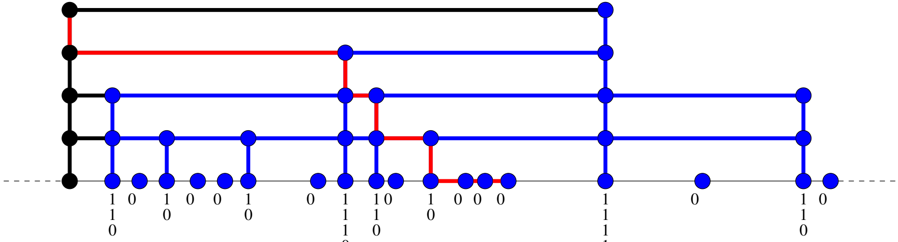


Le chemin de recherche dans une skip list est l'ensemble des cellules à parcourir lors de la recherche d'une clé, depuis le sommet sentinelle.

[1] Luc Devroye. A limit theory for random skip lists.

Quelques propriétés connues des skip lists

- Nombre de niveaux : $O(\log n)$ w.h.p..
- Hauteur H_n d'une skip list : longueur maximale d'un chemin de recherche pour n'importe quelle clé, depuis le sommet sentinelle. Devroye[1] a prouvé que H_n est d'ordre $\log n$.

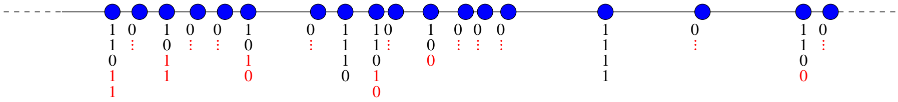


Le chemin de recherche dans une skip list est l'ensemble des cellules à parcourir lors de la recherche d'une clé, depuis le sommet sentinelle.

[1] Luc Devroye. A limit theory for random skip lists.

Plan

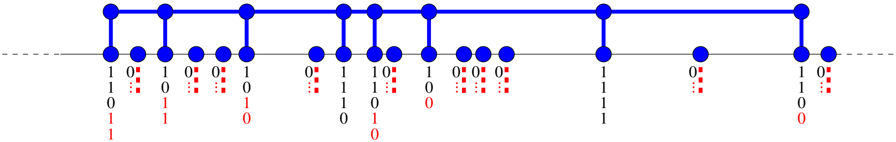
Exemple



Un skip graph basé sur la même liste ordonnée.[2]

[2] James Aspnes and Gauri Shah. Skip graphs.

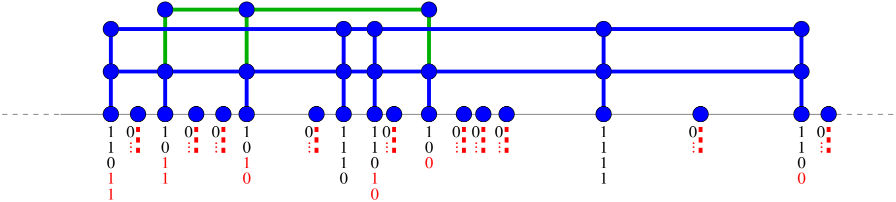
Exemple



Un skip graph basé sur la même liste ordonnée.[2]

[2] James Aspnes and Gauri Shah. Skip graphs.

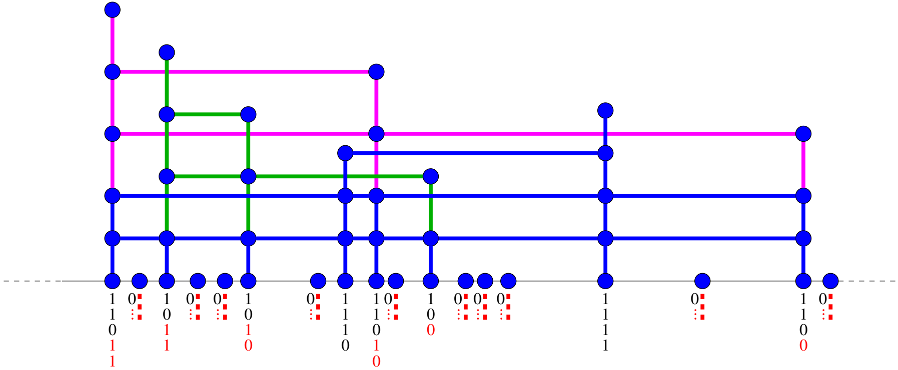
Exemple



Un skip graph basé sur la même liste ordonnée.[2]

[2] James Aspnes and Gauri Shah. Skip graphs.

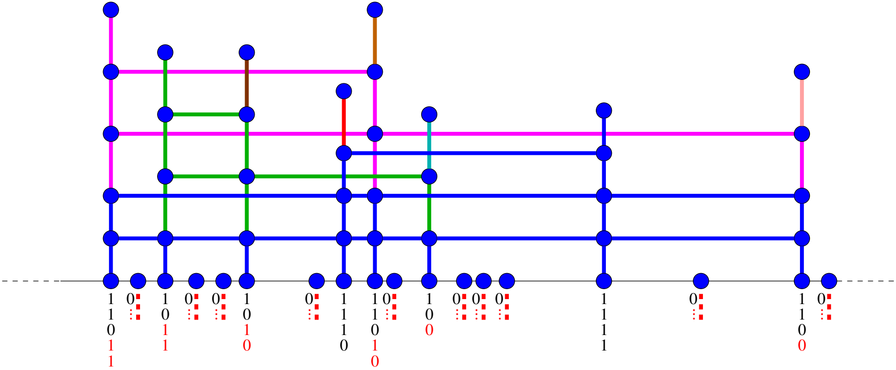
Exemple



Un skip graph basé sur la même liste ordonnée.[2]

[2] James Aspnes and Gauri Shah. Skip graphs.

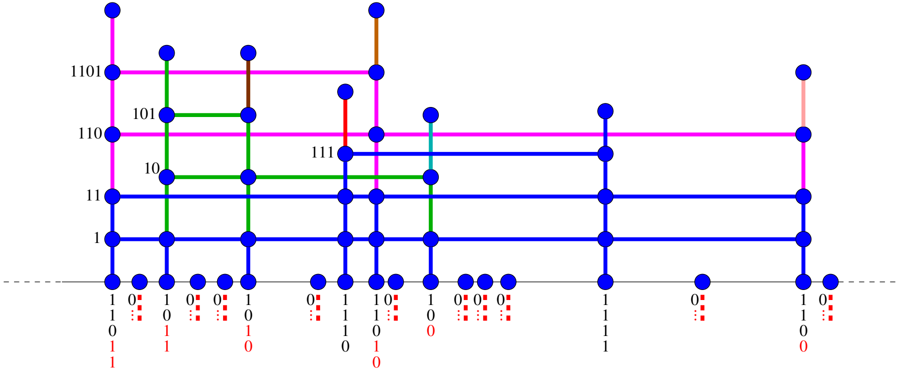
Exemple



Un skip graph basé sur la même liste ordonnée.[2]

[2] James Aspnes and Gauri Shah. Skip graphs.

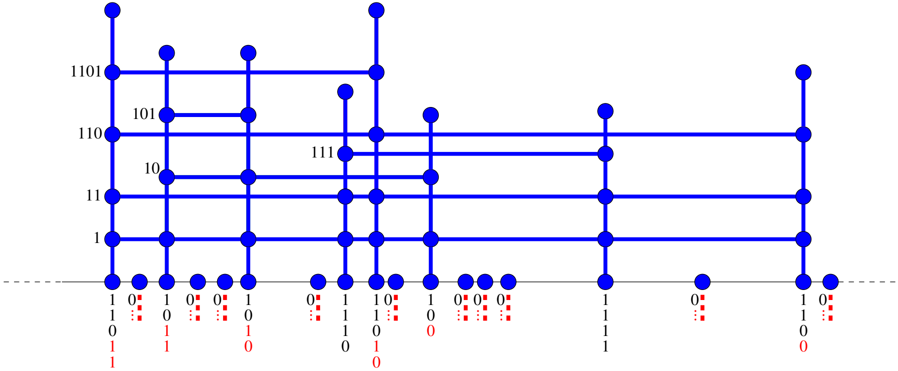
Exemple



Un skip graph basé sur la même liste ordonnée.[2]

[2] James Aspnes and Gauri Shah. Skip graphs.

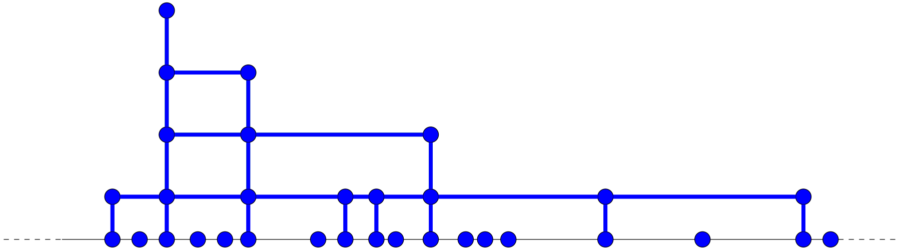
Exemple



Un skip graph basé sur la même liste ordonnée.[2]

[2] James Aspnes and Gauri Shah. Skip graphs.

Exemple



Un skip graph basé sur la même liste ordonnée.[2]

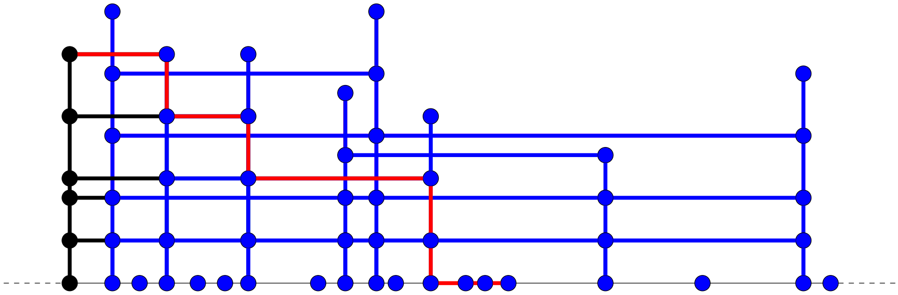
[2] James Aspnes and Gauri Shah. Skip graphs.

Quelques propriétés connues des skip graphs

- Meilleures performances dans un environnement distribué.
- La structure de données peut être distribuée sur un grand nombre de noeuds.
- Bon équilibrage de charge, bonne tolérance aux pannes.
- Nombre de niveaux : $O(\log n)$.

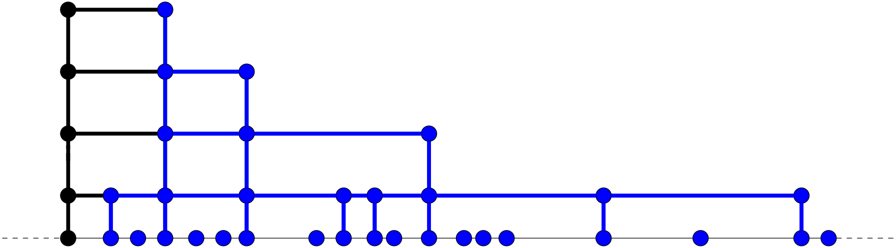
Plan

Définition et lien avec les chemins de recherche dans une skip list



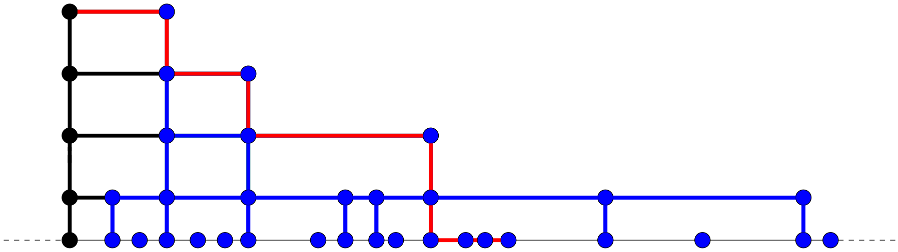
Le chemin de recherche dans un skip graph peut être vu comme le chemin de recherche dans l'une des n différentes skip lists.

Définition et lien avec les chemins de recherche dans une skip list



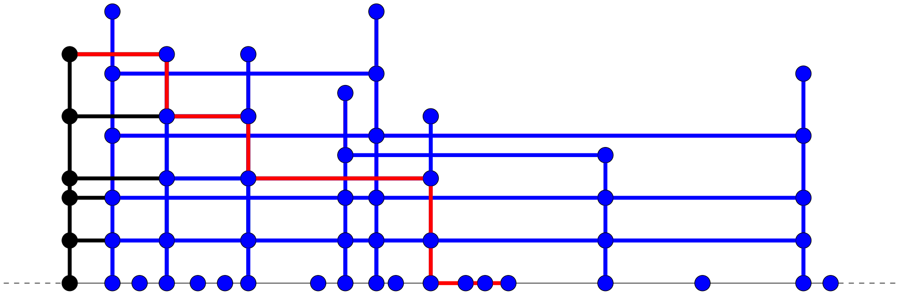
Le chemin de recherche dans un skip graph peut être vu comme le chemin de recherche dans l'une des n différentes skip lists.

Définition et lien avec les chemins de recherche dans une skip list



Le chemin de recherche dans un skip graph peut être vu comme le chemin de recherche dans l'une des n différentes skip lists.

Définition et lien avec les chemins de recherche dans une skip list



Le chemin de recherche dans un skip graph peut être vu comme le chemin de recherche dans l'une des n différentes skip lists.

Résultat principal

Theorem (Duchon, L.)

Pour n'importe quel $\epsilon > 0$, la hauteur H'_n d'un skip graph construit à partir d'une liste de n éléments est telle que $\mathbb{P}(H'_n \leq (2c + \epsilon)\log_2 n) = 1 - o(1)$, où c est l'unique solution de l'équation

$$x - 1 - x\log_2(x) + (x-1)\log_2(x-1) = 0.$$

Schéma de preuve

En étendant le résultat de Devroye, on prouve le Lemme suivant :

Lemma

Quelque soit $\alpha \geq 1$ et quelque soit $c > \alpha c_p$, où c_p est l'unique solution positive de l'équation

$$x - 1 + (x-1)\log_{1/p}(x-1) - x\log_{1/p} x = 0,$$

la hauteur H_n d'une skip list aléatoire (de paramètre p) de n éléments est telle que $\mathbb{P}(H_n > c\log_{1/p} n) = o(n^{1-\alpha})$.

Schéma de preuve du Lemme

- $\mathbb{P}\{H_n > k\} \leq \mathbb{P}\{L(\mathcal{S}) > l\} + n\mathbb{P}\left\{\sum_{j=0}^{l} N_j > k\right\}$ où $L(\mathcal{S})$ est le niveau maximal de $\mathcal{S}$, et N_j est le nombre de noeuds au niveau j dans le chemin de recherche de x_n.
- Or on a $\mathbb{P}(L(\mathcal{S}) > l) \leq n.p^l$ car le niveau maximal d'une skip list est le maximum de n variables aléatoires indépendantes de paramètre p.
- On prend donc : $l = \alpha \log_{1/p} n + \sqrt{\log_{1/p} n}$ et on cherche $k = \lceil c \log_{1/p} n \rceil$, avec c tel que $\mathbb{P}\left\{\sum_{j=0}^{l} N_j > k, L(\mathcal{S}) \leq l\right\} = o(n^{-\alpha})$.
- $\Rightarrow$ en recombinant, on aura bien $\mathbb{P}\{H_n > k\} < o(n^{1-\alpha})$

Schéma de preuve du Lemme (2)

Via l'inégalité de Chernoff, on trouve une borne inférieure sur la valeur de c : $c > \alpha c_p$.

Lemma

Quelque soit $\alpha \geq 1$ et quelque soit $c > \alpha c_p$, où c_p est l'unique solution positive de l'équation

$$x - 1 + (x-1)\log_{1/p}(x-1) - x\log_{1/p} x = 0,$$

la hauteur H_n d'une skip list aléatoire (de paramètre p) de n éléments est telle que $\mathbb{P}(H_n > c\log_{1/p} n) = o(n^{1-\alpha})$.

Nous utiliserons le cas $\alpha = 2$ et $p = \frac{1}{2}$:$\mathbb{P}(H_n > 2c_{1/2}\log_2 n) = o(\frac{1}{n})$.

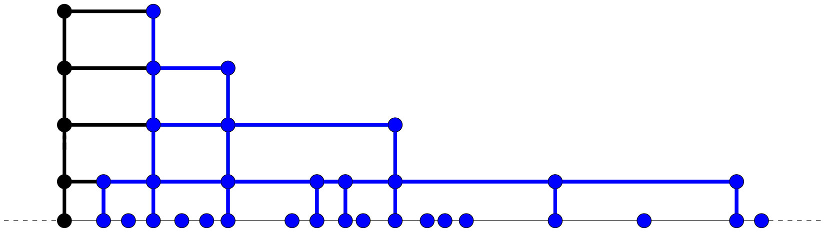
Schéma de preuve (2)

Considérons la skip list “vue” par x_i, $\mathcal{S}(x_i)$:

- La longueur du plus long préfixe commun entre w_{x_i} et w_{x_j} est distribuée suivant une loi géométrique de paramètre $1/2$
- $\Rightarrow$ le nombre de listes de $\mathcal{S}(x_i)$ dans lesquelles x_j apparait est distribué selon une loi géométrique.
- $\Rightarrow$ $\mathcal{S}(x_i)$ peut être considéré comme une skip list aléatoire de paramètre $p = 1/2$ de taille $n - 1$ dans laquelle x_i est inséré au rang i, seul au plus haut niveau.

Schéma de preuve (3)

$\mathcal{S}(x_i)$ peut être considéré comme une skip list aléatoire de paramètre $p = 1/2$ de taille $n-1$ dans laquelle x_i est inséré au rang i, seul au plus haut niveau.



La longueur maximale d'un chemin de recherche dans un skip graph peut donc être considéré comme la longueur maximale des chemins de recherche dans n skip lists pas indépendantes de paramètre $p = 1/2$.

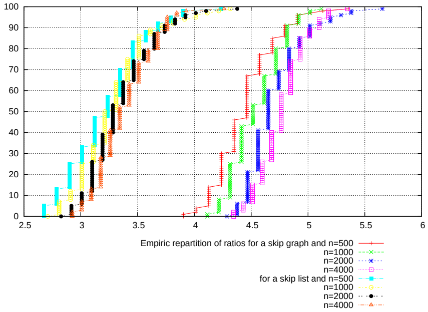
Schéma de preuve (4)

- $\mathbb{P}(H_n > 2c_{1/2} \log_2 n) = o(\frac{1}{n})$,
- H'_n a pour borne supérieure la longueur maximale des chemins de recherche dans n skip lists indépendantes de paramètre $1/2$,
- $\Rightarrow \mathbb{P}(H'_n \leq (2c + \epsilon) \log_2 n) = 1 - o(1)$ où c est la constante de Devroye pour les skip lists, vérifiant l'équation suivante :

$$x - 1 - x \log_2(x) + (x - 1) \log_2(x - 1) = 0.$$

Plan

Un espace à combler



- Pour les skip lists, la constante $c_{1/2} \simeq 4.403$ prédite par Devroye correspond aux sommets des courbes.
- Le rapport $\frac{H'_n}{\log n}$ semble inférieur à notre borne supérieure de $\simeq 8.807$.
- Aucun résultat concernant une possible borne inférieure pour ce rapport.
- Existe-t-il une constante c' telle que $\frac{H'_n}{\log n} \to c'$ en probabilité ?

Merci !