Étude et conception d'algorithmes pour la génération aléatoire entropique

April 30, 2024

1 Contexte scientifique

La génération aléatoire de structures combinatoires revêt une importance cruciale dans de nombreux domaines, tant du point de vue théorique que pratique. Sur le plan théorique, elle constitue un outil puissant pour étudier et comprendre la distribution des structures combinatoires, par exemple pour l'étude de certains phénomènes biologiques comme l'étude des arbres phylogénétiques. En permettant la création aléatoire de ces structures, on peut explorer l'espace des combinaisons possibles, offrant ainsi des perspectives précieuses sur les propriétés asymptotiques et les comportements statistiques des objets combinatoires étudiés. Sur le plan pratique, la génération aléatoire joue un rôle clé dans le développement d'algorithmes efficaces et rapides pour résoudre des problèmes concrets, on peut citer par exemple le domaine de la physique statistique ou l'étude de certaines structure comme le modèle d'Ising est fondamental pour comprendre la magnétisation des aimants sur une échelle macroscopique. Elle trouve également des applications dans d'autres domaines variés tels que l'optimisation des ressources informatiques, la modélisation de phénomènes complexes, la simulation de scénarios sociaux, et même dans des domaines artistiques tels que le "art générative". En générant des instances aléatoires de structures combinatoires, on peut simuler des situations réelles de manière réaliste, faciliter la prévision, l'analyse de données, et contribuer à une meilleure compréhension des systèmes complexes.

Les générateurs de Boltzmann [DFLS04, FFP07, BRS12] sont des algorithmes permettant de tirer aléatoirement des objets dans une classe combinatoire spécifiable donnée. Plus précisément, c'est une procédure qui permet à partir de la spécification (dans le langage de la méthode symbolique [FS09]) d'une classe combinatoire de produire automatiquement un échantillonneur pour les objets de cette classe. C'est donc une sorte de méta-algorithme qui renvoie un algorithme de tirage aléatoire. Ce cadre a été un bouleversement dans le monde de la génération aléatoire car en terme d'efficacité, il a permis d'atteindre la génération uniforme d'objet de taille inespérée avant. Pour autant, les générateurs de Boltzmann ont quelques défauts. Le premier est consubstantiel, un générateur ne renvoie pas un objet de taille donnée, mais un objet dont la taille suit une certaine distribution appelée distribution de Boltzmann (d'ailleurs cette distribution dépend d'un paramètre). Dès lors, si l'on veut espérer un objet d'un certaine taille, il faut d'une part calibrer le paramètre de Boltzmann pour "viser" la dite taille et d'autre part faire du rejet pour atteindre un objet de la taille voulue (ou presque). Le rejet, ce qui veut dire tirer des objets et les éliminer jusque'à ce que l'on tombe sur un objet dont la taille nous satisfait, a un coût qui a été bien étudié et qui est dans le cas d'une fenêtre acceptation linéaire, linaire en la taille de l'objet. Cela signifie que le générateur gaspille intrinsèquement de l'aléa. Pourtant, les objets rejetés sont des objets de la même classe mais pas de la bonne taille, cela semble donc contre productif de ne pas tenter d'en exploiter un peu la structure.

Plusieurs papiers dans la littérature montrent des algorithmes (mais qui ne sont pas des algorithmes de Boltzmann) où l'on essaie de ne pas rejeter trop vite des structures qui ne serait pas complètement ce que l'on veut, mais plutôt de réutiliser astucieusement l'aléa consommé dans leur génération pour construire l'objet souhaité.

1.1 Travail déjà réalisé en stage

Génération en temps et entropie optimale des structures de partition d'ensembles ainsi que certaines variantes de cette structure (Publication en cours : basé sur [Sta83]).

2 Sujet de la thèse

Travailler sur la génération d'objets combinatoire de manière efficace. Une attention particulière est portée au concept d'entropie (ou entropie de Shannon). C'est à dire au nombre de bits aléatoires utilisé pendant la génération. Cette notion permet d'avoir une borne inférieure quand aux quantités de temps et d'entropie nécessaires à la génération, on peut ainsi parler d'algorithmes optimaux si la borne inférieure est atteinte.

2.1 Objectif

Améliorer la méthode de Boltzmann et la méthode récursive pour générer des structures de manière efficace en temps et (quasi)-optimale en entropie. Les stratégies ne seront pas les mêmes. Pour la méthode récursive, il sera question de ne plus travailler avec les distributions exactes qui necessite la connaissance parfaite du nombre d'objet de toutes tailles, mais de travailler sur des approximations asymptotiques controlées. Pour la méthode de Boltzmann, le travail se concentrera sur la re-exploitation de l'aléa consumé lors du rejet d'objets hors de la fenetre d'acceptation.

3 Encadrement

La thèse se fera au sein de l'équipe CALIN du LIPN à l'Université Sorbonne Paris Nord sous la direction d'Olivier Bodini (Professeurs des Universités en informatique).

4 Calendrier prévisionnel

- Renforcement des bases en combinatoire.
- Étude de certains cas "faciles" : Les arbres croissants n-aires [BFS92]. (Publication en cours sur le cas des arbres croissants binaires).
- S'inspirer d'algorithmes existant et de nos autres travaux pour concevoir une version des générateurs de Boltzmann qui ré-exploitent les objets rejetés pour la construction de l'objet escompté. Le but serait d'être quasi-entropique tout en restant efficace en temps de génération.

References

[BFS92] François Bergeron, Philippe Flajolet, and Bruno Salvy. Varieties of increasing trees. In *CAAP'92: 17th Colloquium on Trees in Algebra and Programming Rennes, France, February 26–28, 1992 Proceedings 17*, pages 24–48. Springer, 1992.

[BRS12] Olivier Bodini, Olivier Roussel, and Michele Soria. Boltzmann samplers for first-order differential specifications. *Discrete Applied Mathematics*, 160(18):2563–2572, 2012.

[DFLS04] Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combinatorics, Probability and Computing*, 13(4-5):577–625, 2004.

[FFP07] Philippe Flajolet, Éric Fusy, and Carine Pivoteau. Boltzmann sampling of unlabelled structures. In *2007 Proceedings of the Fourth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 201–211. SIAM, 2007.

[FS09] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. cambridge University press, 2009.

[Sta83] AJ Stam. Generation of a random partition of a finite set by an urn model. *Journal of Combinatorial Theory, Series A*, 35(2):231–240, 1983.