A direct sampler from A-hypergeometric distributions

Shuhei Mano, The Institute of Statistical Mathematics
Nobuki Takayama, Kobe University

The A-hypergeometric distributions defined by [1] are conditional distributions for an exchangeable sample from log-affine models and appear in typical count data analyses. Consider m cells and $t_i \in \{1, ..., m\}$ is the i-th observation of a sample of size n. A sample is specified by the count vector $(c_1, ..., c_m)$, where $c_i := \#\{j; t_j = i\}$. If we assume $(c_1, ..., c_m) \sim Multinom(n; p_1(\xi), ..., p_m(\xi))$ with log-affine models $\log p_i(\xi) = \sum_{j=1}^{d+g} a_{ji}\xi_j - \phi_i(\xi)$, we have

$$\mathbb{P}(C = c) = \frac{n!}{c!} \exp\left\{\sum_{j=1}^{d+g} \xi_j b_j - \sum_{i=1}^{m} \phi_i(\xi)\right\}, \qquad c_! := \prod_{i=1}^{m} c_i!.$$

Here, $(b_1, ..., b_{d+g})$ are sufficient statistics satisfying $b_j = \sum_{i=1}^{m} a_{ji}c_i$, $j \in \{1, ..., d+g\}$. Let the conditions for $j \in \{1, , ..., d\}$ be denoted by $Ac = b$, where A is a non-negative integer-valued $d \times m$ matrix of rank d and $b \in \mathbb{N}_0^d$. The conditional distribution

$$\mathbb{P}(C = c|AC = b) \propto \frac{1}{c!} \exp\left\{\sum_{j=d+1}^{d+g} \xi_j b_j\right\} = \frac{x^c}{c!}, \qquad \log x_i := \sum_{j=d+1}^{d+g} a_{ji}\xi_j$$

is called the A-hypergeometric distribution, because the normalization constant is the A-hypergeometric polynomial defined by Gel'fand et al. A typical example is the distribution of the two-by-two contingency table with fixed marginal sums.

In this talk, a direct sampler that enables i.i.d. sampling from general A-hypergeometric distributions is discussed. The algorithm is based on a holonomic gradient method, which computes the A-hypergeometric polynomials without enumerating possible counts.

Algorithm 1 ([2]) *Sampling from the A-hypergeometric distribution of matrix A and vector b.*

1. *Pick $t_1 = j$ with probability $p_A(b; j)$.*
2. *For $i = 2, ..., n$, pick $t_i = j$ with probability $p_A(b - (a_{t_1} + \cdots + a_{t_{i-1}}); j)$.*

Here, $p_A(b; i)$ is the expected proportion of the i-th cell count (c_i over the degree) when the sufficient statistics are b.

References

[1] Takayama N., Kuriki, S., Takemura, A.: A-hypergeometric distributions and Newton polytope. Adv. in Appl. Math. **99**, 109–133 (2018)

[2] Mano, S.: Partition structure and the A-hypergeometric distribution associated with the rational normal curve. Electron. J. Stat. **11**, 4452–4487 (2017)