MCMC procedure for contingency tbales

Ruriko Yoshida

1 Preliminaries

This is based on the paper "Algebraic algorithms for sampling from conditional distributions" by Persi Diaconis and Bernd Sturmfels in Ann. Statist. Volume 26, Number 1 (1998), 363-397.

Also please see the paper "The Diaconis-Sturmfels algorithm and rules of succession" by Ian H. Dinwoodie in Bernoulli Volume 4, Number 3 (1998), 401-410.

I put them in Dropbox as well as week09.pdf file for MCMC introduction.

2 Metropolis Algorithm

Please see page 288 on Gelman et al "Bayesian data analysis". This is one algorithm to run a discrete time Markov chain on the state space. If we want to sample tables according to the hypergeometric distribution we have to put weight on the ratio of hypergeoemtric probabilities of the present step and the proposed step (it is the ratio of products of factorials, and you just get the ratio of changed cell frequencies).

Algorithm 2.1 (Metropolis Algorithm on the set of tables). *Input The observed table x_0 and the sample size N. Model F.*

Output Sampled tables according to the hypergeometric distribution.

Algorithm

1. *Compute a Markov basis M via* `4ti2` *under the model F.*
2. *Set $S = \emptyset$.*
3. *for $i = 1, \cdots, N$ DO*
   (a) *Puck a move $z \in M$ uniformly.*
   (b) *Sample a proposal $x^* = z + x_{i-1}$.*

(c) *Compute the ratio*

$$r = \frac{p(x^*|m)}{p(x_{i-1}|m)}$$

where m *is the set of given marginals. In our case that is*

$$r = \frac{\prod_{all\ cell\ counts\ j\ in\ x_{i-1}} j!}{\prod_{all\ cell\ counts\ k\ in\ x^*} k!}.$$

(d) *Set*

$$x_i = \begin{cases} x^* & \textit{with probability } \min(r, 1) \textit{ and if } x^* \geq 0 \\ x_{i-1} & \textit{else.} \end{cases}$$

(e) *Add* x_i *to* S.

4. *Return* S.

3 MCMC for goodness-of-fits

Suppose we have the null hypothesis H_0 vs the alternative H_1. We apply MCMC for goodness-of-fit test if the sample size (some cell counts are small). The distribution of LRTs would converges to the chi square distribution od d.f. (the number of parameter in the allternative minus the number of parameters in the null hypothesis).

Algorithm 3.1 (MCMC for goodness-of-fits)**.** *Input The observed two way table* x_0 *(Just for now since it is easy to write) and the sample size* N*. Number of burn-in* B*. Number of skip* S*. Model for* H_0*,* F_0 *and model for* H_1*,* F_1*.*

Output A list of log ratios computed from sampled tables according to the hypergeometric distribution.

Algorithm

1. *Sample* B *many tables (let* x_0^* *be the last, i.e.,* B*th sampled table) under* F_0 *with initial table* x_0 *using Algorithm 2.1.*
2. *Set* $L = \emptyset$.
3. *For* $i = 1, \cdots, N$ *do:*
   (a) *Sample a table* x *under* F_0 *with initial* x_{i-1}^* *using Algorithm 2.1.*
   (b) *Compute MLE* μ_0 *under* H_0 *via IPF.*

(c) Compute the pearson's test statistics

$$l = \sum_{j,k} \frac{(x_i[j][k] - \mu_0[j][k])^2}{\mu_0[j][k]}$$

(d) Add l to L.

(e) Sample S many tables (let x_i^* be the last sampled table, i.e., Sth sampled table) under F_0 with the initial x using Algorithm 2.1.

4. Return L.