could cache the samples from the reference distributions at each vector p. assumes that the same data set dimensions are being used. test statistic is being used.

tried sampling from the preimage of theta. intersection of simplex with a hyperplane.

Do we need exchangeability? The data are n IID vectors of length m, e.g., $(X_{11}, X_{12}, \dots, X_{1m}), (X_{21}, X_{22}, \dots, X_{2m}), \dots, (X_{n1}, X_{n2}, \dots, X_{nm})$. We don't make any assumptions about the dependence structure within each vector $(X_{i1}, X_{i2}, \dots, X_{im})$, for any given i. Still the sums $\sum_{j=1}^m X_{ij}, i = 1, \dots, n$ are IID. Since each X_{ij} is 0 or 1, the sums are IID random variables each taking a value in $0, 1, \dots, m$. Viewing $0, 1, \dots, m$ as $m+1$ categories, we can identify each vector sum $\sum_{j=1}^m X_{ij}$ as a (weighted) choice of one of thse $m+1$ categories. These choices are IID, so their sum is multinomial.

The hyperplane $c^t x = 1$ in $\{x \geq 0\} \subset \mathbb{R}^m$, for constant $c \geq 0$, is contained in $[0, c_1^{-1}] \times \dots \times [0, c_m^{-1}]$. So can use rejection sampling to sample points uniformly on its intersection with the solid simplex. Acceptance probability $O(1/m!)$ (volume of solid simplex in $\mathbb{R}^m$.

1 Method

Let x be an observaiton from the multinomial distribution with sample size n and parameter $p = (p_1, \dots, p_m)$. Let $c = (0, \dots, m-1)/(m-1)$. The goal is a confidence interval for $\theta = c^t p$.

One CI is given by maximum likelihood. The MLE ... is asymptotically normal with variance ..., which may be approximated by ... . For a given finite sample size, the coverage of this CI deteriorates as the multinomial parameter p approaches the boundary of the parameter space, the simplex in R^m. We therefore look for a more efficient CI.

We may obtain an exact CI by inverting a test statistic. describe inversion. describe sampling procedure.

This CI is exact, i.e., its coverage equals the nominal coverage, subject to provisos:

1. monte carlo error, which may be reduced arbitrarily by increasing the tuning parameters ...

2. discreteness. e.g., when n=2, ... This may be removed by introducing randomness ... , though we don't do so here.

3. the null hypothesis $\theta = \theta_0 = \{p : c^t p = \theta_0\}$ is a composite null hypothesis, which leads to a conservative CI. That is, the null consists of multiple distributions, the p-value is corresponds to the least favorable.

2 Simulation