Rejection Sampling from the Quasi-Multinomial Distribution

Nobuaki Hoshino

School of Economics, Kanazawa University

The Quasi-Multinomial (QM) distribution (type 2) is a generalized multinomial distribution, which has a parameter to control overdispersion with marginal expectations fixed. This property is especially useful when random samples from this distribution are used as perturbed data in order to protect information, since the degree of protection can be tuned while the unbiasedness of univariate margins holds. Therefore the present article considers a simple method of sampling from the QM distribution.

Firstly, let us define the probability mass function (pmf) of the QM distribution by

$$p(y_1,\ldots,y_F) = \frac{n!}{y_1!\ldots y_F!}\frac{1}{(1+n\beta)^{n-1}}\prod_{f=1}^{F}\pi_f(\pi_f+y_f\beta)^{y_f-1}, \tag{1}$$

where $y_f, f=1,2,\ldots,F$, is a nonnegative integer, and $\sum_{f=1}^{F} y_f = n$. The parameter π_f of (1) is a nonnegative real number, and $\sum_{f=1}^{F}\pi_f = 1$ the same as a cell probability of the multinomial distribution. The parameter β of (1) is a nonnegative real number in this article, and $\beta = 0$ represents the multinomial distribution.

We denote the QM distribution (1) by $\mathrm{QM}(\pi_1,\ldots,\pi_F;n,\beta)$. Its special case of $F=2$ is called the Quasi-Binomial (QB) distribution (type 2), which is referred to by $\mathrm{QB}(\pi;n,\beta)$.

The conditional distribution method of sampling exploits the following relationship:

$$(Y_1,\ldots,Y_F) \overset{d}{=} (Y_1|Y_2,\ldots,Y_F)\ldots(Y_{F-2}|Y_{F-1},Y_F)(Y_{F-1}|Y_F)Y_F. \tag{2}$$

The conditional probabilities of the right hand side are explicitly given below on the QM distribution:

Proposition 1 *If* $(Y_1,\ldots,Y_F)\sim QM(\pi_1,\ldots,\pi_F;n,\beta)$ *then* $(Y_g|Y_{g+1}=y_{g+1},\ldots,Y_F=y_F)\sim QB(\pi_g/(1-\sum_{f=g+1}^{F}\pi_f);n-\sum_{f=g+1}^{F}y_f,\beta)$ *for* $g=1,\ldots,F$.

Therefore sampling from the QM distribution is accomplished by sequential sampling from QB distributions, and rejection sampling from the QB distribution is rather simplified when we use the Beta Binomial mixture distribution as "proposal", where we can show its acceptance ratio on $y\in\{0,1,2\ldots,n\}$ as

$$\frac{\Gamma(a_1+n)\Gamma(a_2+1)}{\Gamma(a_1+y)\Gamma(a_2+n-y)}\frac{(a_1+y)^{y-1}(a_2+n-y)^{n-y-1}}{(a_1+n)^{n-1}} =: \rho(y). \tag{3}$$

By symmetry, it suffices to present our result for $0<a_1<a_2$.

Algorithm 1 *The following procedure generates a sample from* $QB(a_1/(a_1+a_2);n,1/(a_1+a_2))$ *for a positive integer* n *when* $0<a_1<a_2$:

1. *Generate* $p\sim Beta(a_1,a_2)$
2. *Generate* $y|p\sim Binomial(n,p)$
3. *Generate* $u\sim U(0,1)$
4. *If* $u>\rho(y)$ *then goto 1*
5. *Output* y.