Accepted Manuscript

Discretization-based direct random sample generation

Liqun Wang, Chel Hee Lee

PII: S0167-9473(13)00223-5
DOI: http://dx.doi.org/10.1016/j.csda.2013.06.011
Reference: COMSTA 5618

To appear in: *Computational Statistics and Data Analysis*

Received date: 1 July 2012
Revised date: 5 June 2013
Accepted date: 6 June 2013

Please cite this article as: Wang, L., Lee, C.H., Discretization-based direct random sample generation. *Computational Statistics and Data Analysis* (2013), http://dx.doi.org/10.1016/j.csda.2013.06.011

Discretization-based Direct Random Sample Generation

Liqun Wang[a,b,*], Chel Hee Lee[c,]

[a] *Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2*
[b] *School of Science, Beijing Jiaotong University, Beijing, China*
[c] *The Collaborative Biostatistics Program, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*

Abstract

An efficient Monte Carlo method for random sample generation from high dimensional distributions of complex structures is developed. The method is based on random discretization of the sample space and direct inversion of the discretized cumulative distribution function. It requires only the knowledge of the target density function up to a multiplicative constant and applies to standard distributions as well as high-dimensional distributions arising from real data applications. Numerical examples and real data applications are used for illustration. The algorithms are implemented in statistical software R and a package `dsample` has been developed and is available online.

Key words: Direct sampling; discretization; Monte Carlo sampling; multivariate random variates generation; R package; visualization.

1. Introduction

Multivariate random sample generation plays an important role in computational statistics and data analysis. It is a central part of solutions to many high-dimensional integration and optimization problems as well as Bayesian data analysis (e.g., Sotto et al. 2011). Many numerical methods and algorithms have been developed in the literature, most notably the Markov chain Monte Carlo (MCMC) methods. However, in real applications the MCMC have practical difficulties when dealing with high-dimensional, multi-modal distributions, as well as other technical problems such as convergence (e.g., O'Hagan, Murphy and Gormley 2012).

*Corresponding author: Tel: 204-474-6270, Fax: 204-474-7621.
Email addresses: `liqun.wang@umanitoba.ca` (Liqun Wang), `chl948@mail.usask.ca` (Chel Hee Lee)