Multivariate Correlated Sampling Using Extended Alias Techniques

S. Mohanty, S. Banerjee, D. Goyal,
A. K. Mohanty and F. Carminati

One Dimensional Sampling

- Equal Probability Bin:

  Most Commonly used method in Monte-Carlo Sampling. It works as follows. Consider the probability density p(x) with cumulative density

$$Q(x) = \int_a^x p(x')\,dx'$$

can be sampled using the inverse function method which gives $x=Q^{-1}(u)$ where u is a random variable between (0,1). Numerically, Q^{-1} can be divided into N equal bins. Sampling is carried out choosing a bin randomly between 1 and N with bin number $i=Nu_1+1$. Then x is chosen within a bin by $x=(1-u_2)x_{i+1}+u_2x_i$ where u_1 and u_2 are uniform random numbers between (0,1).

It is quite fast but inaccurate as much of the original distribution is lost by the necessary assumption of uniform likely hood within each bin.

Alias Sampling

Another popular method is alias sampling which is as fast as the equal probable bin and can be made as accurate as table look up. Initially, it was proposed by A. J. Walker in 1977 to sample from a discrete data set. Consider a data set having N discrete tabulated probability values p_i such that $\Sigma p_i = 1$. The alias method recasts this distributions into N equal probable events each with likely hood 1/N. Each event i consists of a non-alias outcome i and an alias outcome λ_i and the probability of non-alias outcome is π_i . As an example consider a Poisson Discrete Distribution with mean value of 2. The first two column show 10 discrete P_i values.

Alias Table for Poisson Distribution

i	P_i	λ	Π
0	0.135	0	0
1	0.270	0	0.835
2	0.270	0	0.972
3	0.180	0	0.839
4	0.090	2	0.902
5	0.030	2	0.361
6	0.012	1	0.120
7	0.002	3	0.034
8	0.0008	1	0.008
9	0.0001	2	0.002

Discrete Alias Sampling

Alias algorithm generates alias outcome λ and alias probability π. The outcome is sampled by first randomly selecting an equal probable non-alias/alias event $i = Nu_1 + 1$. The method then compares a second random number u_2 against the non-alias probability to select either the non-alias or alias outcome as follows. If $u_2 \leq \pi_i$, then non-alias event i is chosen. Otherwise an alias event λ_i is selected.

Linear Alias Method

The above method can be extended to generate continuous distribution. Assuming that the tabulated distribution is piece-wise linear, two random numbers are thrown to estimate x_1 and x_2,

$$x_1 = (1 - u_3)x_i + u_3 x_{i+1}$$

$$x_2 = u_3 x_i + (1 - u_3) x_{i+1}$$

Using another random number u_4, $x = x_1$ is chosen if

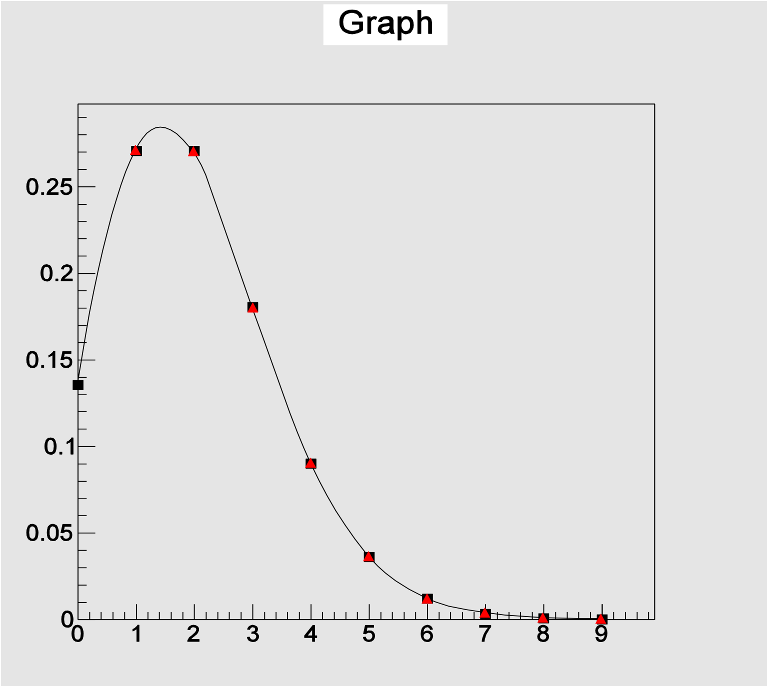
$$u_4 (y_i + y_{i+1}) \leq (1 - u_3) y_i + u_3 y_{i+1}$$

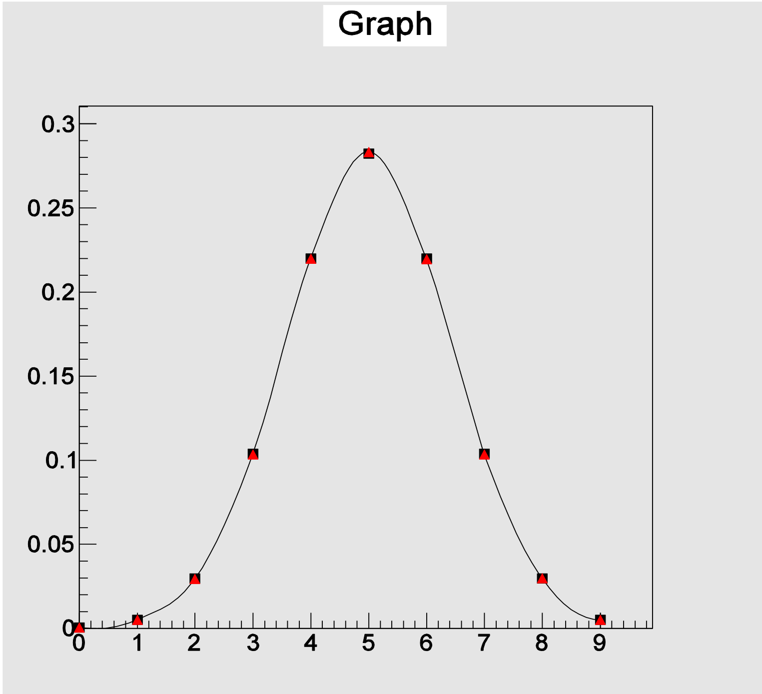
Otherwise, $x = x_2$ is selected.

Advantage and Disadvantage

- The EPB method requires two random numbers where as the alias linear sampling requires four random numbers.
- The alias method stores three tables i , λ and π . Whereas EPB method stores two tables i and P_i .
- The linear alias method is accurate and as fast as EPB method.
- The only major disadvantage of the Alias sampling over the EPB or other sampling methods is that it consumes more memory which is required for storage.

Example :Using Linear Alias Sampling





Solid Squares are input data points for a Poisson distribution and continuous curve is the result of linear alias sampling.

Solid Squares are input data points for a Gaussian distribution and continuous curve is the result of linear alias sampling

Example of Alias and EPB Samplings

We carry out a comparative study using both EPB and linear alias samplings. We have considered ENDF/ B-VI file 5 for LF=1 which gives the secondary energy probability distribution. We select the reaction MT=91 for neutron interacting with ^{238}U via inelastic continuum reaction.

ENDF Data at Energy 8MeV

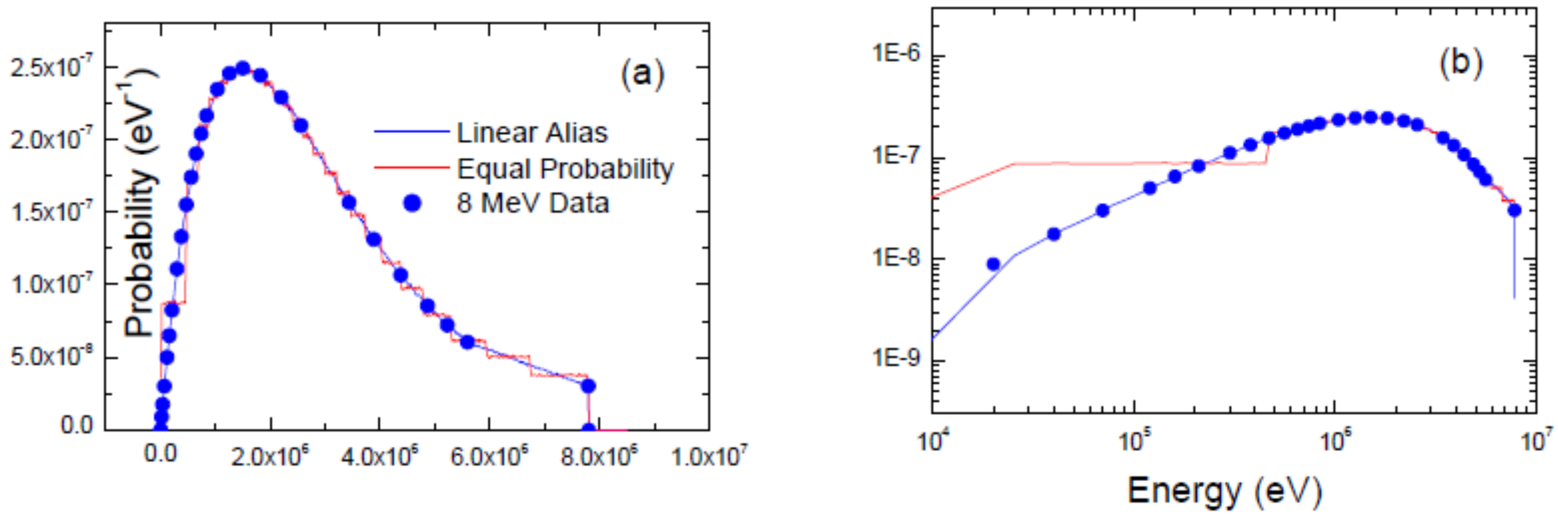


FIG. 1: (a) The probability distribution of the out going neutron of the reaction $n +^{238} U \rightarrow n' +^{238}$ at 8 MeV. The data points (blue filled circles) are taken from ENDF data file. The red curve is obtained using EPB samplings where as the blue curve is due to linear alias sampling. (b) Same as (a), but shown on log scale. The sampling has been carried out using $10E7$ events.

2D Correlated Sampling

The Alias Sampling can be extended for multivariate correlated sampling. As an example, we consider a correlated Gaussian distribution in two dimension. Although the method can be used for any non-Gaussian distribution as well.

- Method:

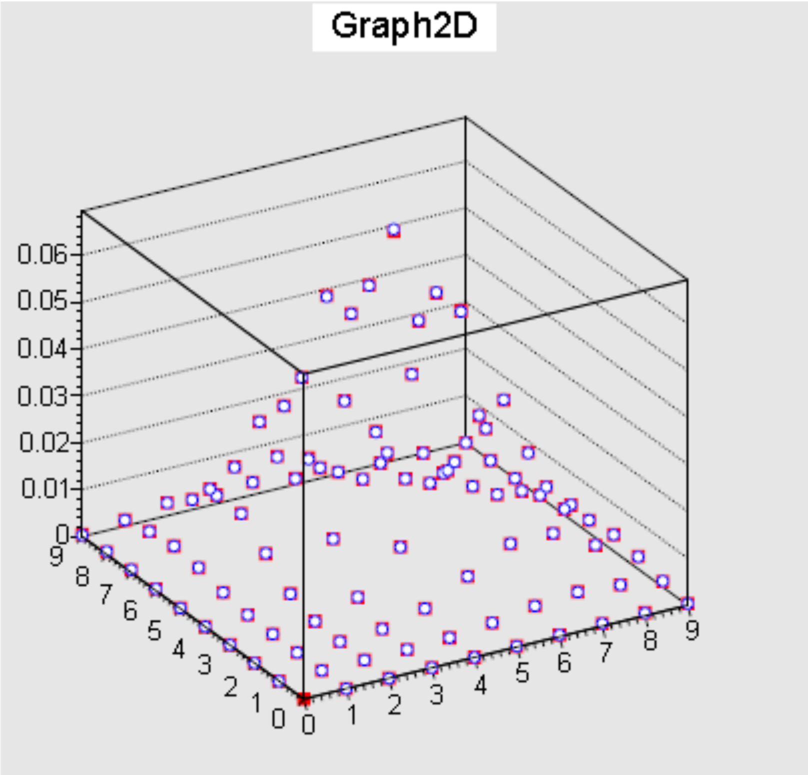
  Consider 10 Gaussian discrete points p[i][j] (σ^2=1) corresponding to i = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and j = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. We choose a row in random using alias sampling based on the distribution px[i] where px[i] = Σ_j p[i][j]. Once the row i is chosen the corresponding j values are chosen depending on the distribution p[i][j].

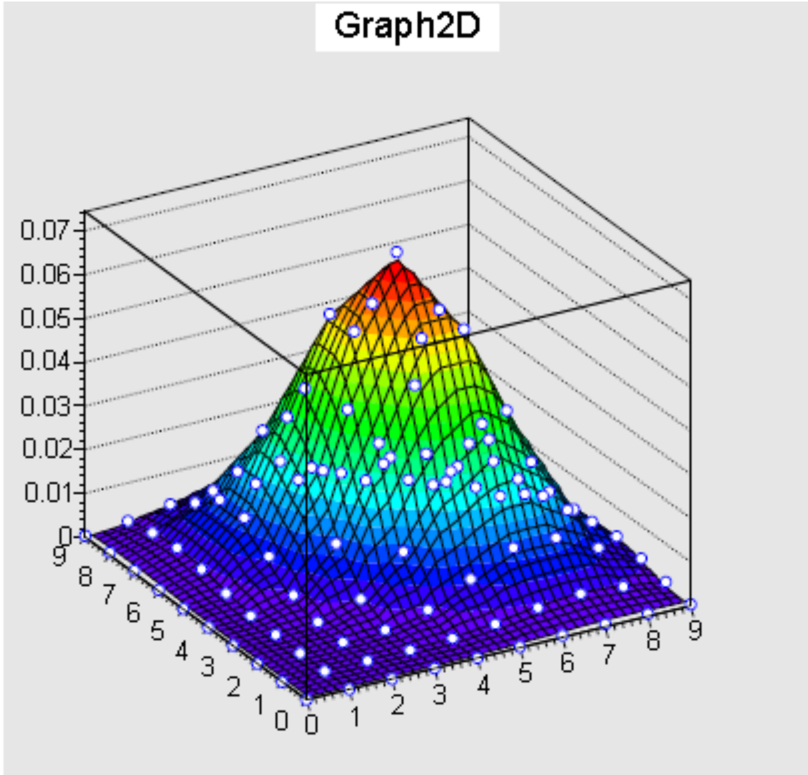
Example :

$$f(x,y) = e^{-\left[(x-x_0)^2+(y-y_0)^2+\alpha(x-x_0)(y-y_0)\right]}$$

We Sample using Walker's alias techniques which requires 4 random numbers per sample. In this $\sigma^2 = 1$ and α is the correlation parameter.

Plots for original and generated distribution





Left: Circles represent original points and squares are the generated points. Right: Circles represent original points and continuous curve is for generated points for α=1.0

Chi-Square Test

$$\chi^2 = \frac{1}{N}\sum\left[\frac{p[i][j]-gx[i][j]}{p[i][j]}\right]^2$$

TABLE II: The table for χ^2 test(σ^2=10.0)

α	χ^2/N $n = 10^2K$	χ^2/N $n = 10^3K$	χ^2/N $n = 10^4K$
0.0	2.56E-3	1.78E-4	2.92E-5
1.0	3.77E-3	1.53E-3	1.45E-4
2.0	2.42E-2	5.02E-3	4.51E-4
3.0	3.25E-1	4.24E-2	2.42E-2

Covariance Test

TABLE IV: The table for co-variance test , N= 100K

α	σ^2	σ^2_{xy}	σ^2	σ^2_{xy}
0.0	1.985	-3.32E-17	1.977	-.0008
1.0	2.579	-1.264	2.569	-1.255
2.0	6.856	-5.968	6.845	-5.978
3.0	16.30	-15.94	16.31	-15.94

Second and Third Columns are the input covariances at different correlation parameter α, fourth and fifth columns are the generated covariances. The agreement is extremely good. Which justifies the accuracy of the alias samplings.

Conclusions

- The alias sampling is fast and efficient.
- As compared to commonly used EPB sampling, Alias Sampling is accurate although it requires slightly higher memory for storage.
- This method can be applied to sample from any arbitrary multivariate correlated distribution.

- Although this method is quite general and can be applied to any dimensions, in this work we have restricted sampling only from a two dimensional correlated distribution. The motivation behind this study has been to develop a ROOT based Monte-Carlo application package for low energy neutron transport down in energy to a few keV using the evaluated nuclear data file (ENDF) which is available in ROOT format. Work is in progress to apply this new method of alias technique to the ENDF data set where the angle and energy distributions are strongly correlated.