International Journal of Mathematics Research.
ISSN 0976-5840 Volume 7, Number 1 (2015), pp. 35-39

http://www.irphouse.com

Selection Rejection Methodology For N-Dimensional Continuous Random Variables And Its Application To N-Dimensional Normal Distribution

Sachinandan Chanda

Department of Mathematics, Shillong Polytechnic, Mawlai, Shillong –793022

ABSTRACT

In this paper we have generalized the Selection-Rejection Methodology for one dimensional continuous random variables to n-dimensional continuous random variables and applied it to the n-dimensional normal distribution.

Keywords: - random variable, iterations, target probability distribution and proposal probability distribution.

1. INTRODUCTION.

The Selection-Rejection Methodology for one dimensional continuous random variables was developed based on the idea of Acceptance-Rejection Method by the renowned mathematician, Von Neumann, from the University of Berlin. Von Neumann[1] came forward with his method during 1950's but later on Karl Sigman[2] from Columbia University gave the similar methodology in 2007.Again in 1989,Bernard D.Flury[3] from Indiana University came forward with the theory "Acceptance-Rejection Sampling Made Easy".D.P.Kroese[4] from University of Queensland put forward his theory of Acceptance-Rejection in 2011.Selection-Rejection Methodology can be applied to almost all statistical distributions and hence it has got immense physical significance.

2. SELECTION-REJECTION METHODOLOGY FOR N-DIMENSIONAL CONTINUOUS RANDOM VARIABLES.

Let $X_1, X_2, ..., X_n$ be a n dimensional continuous random variable with probability distribution function $f\left(x_1, x_2, ..., x_n\right) \; \forall x_i \in R, i = 1, 2, ..., n$,where R =set of all real numbers. Let $g\left(x_1, x_2, ..., x_n\right) \; \forall x_i \in R, i = 1, 2, ..., n$ where R =set of all real numbers be another probability density function such that $\frac{f(x_1, x_2, ..., x_n)}{g(x_1, x_2, ..., x_n)} \le k \;\; \forall x_i \in R, i = 1, 2, ..., n$, where $k \ge 1$ is a real number. By successively selecting different values of $X_1, X_2, ..., X_n$ we will try to make the ratio $\frac{f(x_1, x_2, ..., x_n)}{kg(x_1, x_2, ..., x_n)}$ as close to 1 as possible. The probability density function $f(x_1, x_2, ..., x_n)$ is called target distribution and the probability density function $g(x_1, x_2, ..., x_n)$ is called proposal distribution.

The step by step procedure for the Selection-Rejection Methodology is as follows.

Step (1):- Let $X_1, X_2, ..., X_n$ be a n dimensional continuous random variable with probability distribution function $f(x_1, x_2, ..., x_n) \;\; \forall x_i \in R, i = 1, 2, ..., n$, where R =set of all real numbers.

Step (2):- Let $X_1', X_2', ..., X_n'$ be another n dimensional continuous random variable (which is independent of $X_1, X_2, ..., X_n$) with probability distribution function $g(x_1, x_2, ..., x_n) \;\; \forall x_i \in R, i = 1, 2, ..., n$, where R =set of all real numbers.

Step (3):- Let $\frac{f(X_1', X_2', ..., X_n')}{g(X_1', X_2', ..., X_n')} \le k \;\; \forall\, X_i' \in R, i = 1, 2, ..., n$, where $k \ge 1$ a real number.

Step (4):- Let $0 < R_i < 1, i = 1, 2, ..., n$, be n random numbers.

Step (5):- Set X_i' in terms of R_i', $i = 1, 2, ..., n$ depending on the expression obtained for the ratio $\frac{f(X_1', X_2', ..., X_n')}{kg(X_1', X_2', ..., X_n')}$.

Step (6):- If $\prod_{i=1}^{n} R_i \le \frac{f(X_1', X_2', ..., X_n')}{kg(X_1', X_2', ..., X_n')}$, then set $(X_1, X_2, ..., X_n) = (X_1', X_2', ..., X_n')$ select the continuous random variable $(X_1', X_2', ..., X_n')$; otherwise reject the variable $(X_1', X_2', ..., X_n')$ and repeat the process from step (1).

The probability that the continuous random variable $(X_1', X_2', ..., X_n')$ is selected is $\frac{1}{k}$.

The number of iterations required to select $(X_1', X_2', ..., X_n')$ is k.

It may be noted that $0 \le \dfrac{f\left(X_1', X_2', \ldots\ldots\ldots, X_n'\right)}{kg\left(X_1', X_2', \ldots\ldots\ldots, X_n'\right)} \le 1$

To prove that the probability for the selection of $X_1', X_2', \ldots\ldots\ldots, X_n'$ **is** $\dfrac{1}{k}$**,**

Proof: -

$$P\left(Select \middle| X_1', X_2', \ldots.., X_n'\right) = P\left(\prod_{i=1}^{n} R_i \le \frac{f\left(X_1', X_2', \ldots\ldots, X_n'\right)}{kg\left(X_1', X_2', \ldots\ldots, X_n'\right)}\right) = \frac{f\left(X_1', X_2', \ldots\ldots, X_n'\right)}{kg\left(X_1', X_2', \ldots\ldots, X_n'\right)} \,..$$

$$P\left(X_1', X_2', \ldots.., X_n' \ \text{is selected}\right) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\int_{-\infty}^{x_3}\ldots\ldots\int_{-\infty}^{x_n} \frac{f\left(w_1, w_2, \ldots\ldots, w_n\right)}{kg\left(w_1, w_2, \ldots\ldots, w_n\right)} g\left(w_1, w_2, \ldots\ldots, w_n\right) dw_1 dw_2 \ldots.. dw_n$$

$$= \frac{1}{k}\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\int_{-\infty}^{x_3}\ldots\ldots\int_{-\infty}^{x_n} f\left(w_1, w_2, \ldots\ldots, w_n\right) dw_1 dw_2 \ldots\ldots. dw_n = \frac{1}{k}$$

$$\left[\because \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\int_{-\infty}^{x_3}\ldots\ldots.\int_{-\infty}^{x_n} f\left(w_1, w_2, \ldots\ldots, w_n\right) dw_1 dw_2 \ldots\ldots dw_n = 1\right]$$

Hence the proof.

Since the probability of selection (i.e. success) is $\dfrac{1}{k}$,the number of iterations needed will follow a geometric distribution with $p = \dfrac{1}{k}$.So, on average it will take k iterations to generate a number.

III. APPLICATION TO N-DIMENSIONAL NORMAL DISTRIBUTION.

Two dimensional normal distribution is given by

$$f\left(x_1, x_2, \ldots\ldots, x_n\right) = \frac{1}{\sqrt{2\pi}} e^{-\left(\sum_{i=1}^{n}\frac{x_i^2}{2}\right)}, x_i \ge 0; x_i \in R, i = 1, 2, \ldots., n. \qquad (1)$$

Here $f\left(x_1, x_2, \ldots\ldots\ldots, x_n\right)$ is the target distribution.

Let $g(x_1, x_2, \ldots\ldots\ldots, x_n) = e^{-\left(\sum_{i=1}^{n} x_i\right)}, x_i \ge 0, i = 1, 2, \ldots.n.$ be the proposal distribution. (2)

$$\text{Let} \quad h(x_1, x_2, \ldots\ldots., x_n) = \frac{f(x_1, x_2, \ldots\ldots., x_n)}{g(x_1, x_2, \ldots\ldots., x_n)} = \frac{1}{\sqrt{2\pi}} \exp\left(-\left(\frac{\sum_{i=1}^{n} x_i^2 - 2x_i}{2}\right)\right) \qquad (3)$$

With the help of differential calculus we can show that $h(x_1, x_2, \ldots\ldots, x_n)$ attains maximum at $\underbrace{1,1,\ldots\ldots,1}_{n\ times}$ and the maximum value of $h(x_1, x_2, \ldots\ldots, x_n)$ is $\dfrac{e^{\frac{n}{2}}}{\sqrt{2\pi}}$

Choosing $k = \dfrac{e^{\frac{n}{2}}}{\sqrt{2\pi}}$, we get

$$\frac{f\left(x_1, x_2, \ldots\ldots, x_n\right)}{kg\left(x_1, x_2, \ldots\ldots, x_n\right)} = \prod_{i=1}^{n} \exp\left(-\frac{\left(x_i - 1\right)^2}{2}\right) \tag{4}$$

Selection-Rejection Methodology for the n-dimensional distribution is as follows

Step (1):- Let $X_1, X_2, \ldots\ldots, X_n$ be n dimensional continuous random variable with probability distribution function $f\left(x_1, x_2, \ldots\ldots, x_n\right)\ \forall x_i \in R, i = 1, 2, \ldots., n$, where R=set of all real numbers.

Step (2):- Let $X_1', X_2', \ldots\ldots, X_n'$ be another n dimensional continuous random variable with probability distribution function $g\left(x_1, x_2, \ldots\ldots, x_n\right)\ \forall x_i \in R, i = 1, 2, \ldots., n$, where R=set of all real numbers.

Step (3):-Let $0 < R_i < 1, i = 1, 2, \ldots., n$, be n random numbers.

Step (4):- Set $X_i' = 1 + \sqrt{-2\ln(R_i)}, i = 1, 2, \ldots.., n$,

Step (5):-If $\prod_{i=1}^{n} R_i \le \prod_{i=1}^{n} \exp\left(-\frac{\left(X_i' - 1\right)^2}{2}\right)$,then set $\left(X_1, X_2, \ldots\ldots, X_n\right) = \left(X_1', X_2', \ldots\ldots, X_n'\right)$ and select $\left(X_1', X_2', \ldots\ldots, X_n'\right)$;otherwise reject $\left(X_1', X_2', \ldots\ldots, X_n'\right)$ and repeat the process from Step(1).\

Conclusion

Selection-Rejection Methodology is valid for any dimension of continuous random variable.In this method we approximate the target function to proposal function so that after a number of successive iterations the proposal function becomes almost equal to target function and proposal function is selected.

References

(1) JOHN VON NEUMANN, "Various techniques used in connection with random digits, in *Monte Carlo Method,* Appl. Math. Series, vol, 12, U. S. Nat. Bureau of Standards, 1951,pp. 36-38 (Summary written by George E. Forsythe);reprinted in John von Neumann, *Collected Works.* Vol. 5, Pergamon Press; Macmillan, New York, 1963, pp. 768-770. MR 28#1104.

(2) Acceptance-Rejection Method by Karl Sigman, 2007, Columbia University.

(3) BERNARD D. FLURY, "ACCEPTANCE-REJECTION SAMPLING MADE EASY",SIAM Review, Vol. No. 3. Pp 474-476, September 1990.

(4) D.P.Kroese, "Acceptance-Rejection Method", 2011, University of Queensland.

(5) Loernzo Pareschi,"Part III: Monte Carlo methods",2003,University of Ferrara,Italy.

(6) WILLIAM FELLER, "An Introduction to Probability Theory and its Applications", Vol I, Wiley, New York, 1950, Lemma 2, P-131(P-166 of 2nd ed).MR 12,424.

(7) Richard Saucier, "Computer Generation of Statistical Distributions", March 2000, *ARMY RESEARCH LABORATORY.*

(8) J.H. Ahrens and U. Dieter, "Computer Methods for Sampling from Exponential and Normal Distributions", Comm. A.C.M 15 (1972), 873-882.

(9) Fill, J. A. (1998) "An interruptible algorithm for perfect sampling via Markov chains". *Annals of Applied Probability,* 8(1) 131-162. MR1620346.

(10) Fill, J. A. Machide, M., Murdoch, D.J. and Rosenthal, J. S. (1999). "Extension of Fill's perfect rejection sampling algorithm to general chains. Random *Structures and Algorithms"* **17** 219-316. MR1801136.

(11) Gilks, W. R. and Wild, P. (1992). "Adaptive rejection sampling for Gibbs sampling". *Appl. Statist.* **41** 337-348.

(12) Propp, J.G. and Wilson, D. B. (1996). "Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and* Algorithms". **9**(1 & 2), 223-252. MR1611693.