Journal of Radiation Industry 14(4) : 343~351(2020)

A Study on Sampling Methods and Algorithms for Variable Inferences of Diffusion Model

Yunjong Lee[1,*], Hoje Kwon[1] and Shin Ae Kim[1,2]

[1]*Korea Atomic Energy Research Institute (KAERI), 1266, Sinjeoung-dong, Jeongeup-si, Jeollabuk-do 56212, Republic of Korea*
[2]*Departement of Nuclear Engineering, Hanyang University, 222, Wansimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea*

Abstract - Atmospheric transport and dispersion is used to predict the diffusion of radioactive materials to the atmosphere. Several variables are required as inputs to the model for accurate predictions, but in the case of a sudden accident, the concentration, location and time of radioactive emissions must be inferred from monitor measurements. In the inverse modeling method of determining a parameter from a monitor measurement value, sampling is used to estimate a population by extracting some values. In this study, to enable sampling, a function in which the output value x becomes data, was defined, and algorithms for this were compared. Sampling techniques were applied to exponential functions and normal distribution functions. In addition, algorithms for the inverse CDF method, reject sampling, compression reject sampling, adaptive reject sampling, and Importance Sampling were applied through simulation. The inverse CDF method is the most basic method and has the disadvantage of having to find the inverse function. For continuous functions, sampling was difficult near $-\infty$ and ∞ during the sampling process. Rejection Sampling can be easily sampled using the approximation function g(x) to find the target function, but when the c value increases, it takes much time for sampling. Adaptive rejection sampling (ARS) was economical because the area to be rejected was minimized. However, the complexity of the equation increases the computational complexity. In addition, when the dimensions increase, it is very difficult to obtain a sample. Importance sampling must also have a cover function as close as possible to the objective function. If the objective function is large and has a value only in a local area of the data space, poor results may be obtained.

Key words : Inference, Algorithm, Sampling, Rejection sampling, Squeezing rejection sampling, Importance sampling, Atmospheric transport and dispersion

INTRODUCTION

The atmospheric transport and dispersion model is designed to assist emergency responders in effective response and post-emergency assessments in the event of the spread of large quantities of radioactive materials into the atmosphere, such as radioactive material terrorism or the Fukushima nuclear accident. It is used to predict the diffusion of pollutants to provide important information (Connan *et al*. 2013). For accurate predictions, a number of variables besides meteorological data, emission concentration and emission location are required as inputs to the model (Hutchinson *et al*. 2017).

In general, local weather data can be accurately obtained by a monitor installed at a local weather station. However, in case of a sudden accident, there is no concentration, location or time of the release of radioactive material. In this case, concentration, location, etc. must be inferred from the

* Corresponding author: Yunjong Lee, Tel. +82-63-570-3270, Fax. +82-63-570-3259, E-mail. yjlee@kaeri.re.kr

measured values of monitors installed in the area (Christopher *et al*. 2007).

Determining the source parameters in such monitor measurements uses an inverse modeling method, and the method for estimating the population by extracting some values from the model is called sampling. Population indicates all the data you want to know. In general, the population is very large, and it is not easy to know all the values. However, when you want to generate a random number of random variables X that follows all probability distributions f(x), the cumulative distribution function F(X) of all random variables follows a uniform distribution. In the generation of basic random numbers, sampling is possible using the fact that the random number generation range is between 0 and 1. Sampling is establishing a probability model and estimating parameters by taking samples, which are subsets of the population.

The probability density function (PDF) is a function that informs the probability density p(x) according to each input value x. However, it is not possible to know p(x) for a specific x and sampling for data(x). In other words, if you know the PDF, sampling will likely be possible, but only the PDF is not possible.

Therefore, in this paper, in order to enable sampling, a function in which the output value x becomes data, is defined and algorithms are compared. For application to the atmospheric diffusion model, sampling was applied to several exponential functions and normal distribution functions. In addition, with simulations, we created algorithms for the Inverse CDF Method, reject sampling, compression reject sampling, adaptive reject sampling, and importance sampling.

MATERIALS AND METHODS

1. Sampling

Most complex probability models are intractable to inferring exact probability values. In practice, since it is impossible to accurately infer a probability value with a probability model, we have to rely on approximation techniques. One of these approximation techniques is the sampling technique. It is important to obtain an expectation value in predicting a certain value through an actual model. How can we obtain the expected value through sampling? First of all, assuming that there is a distribution p(x) and there is a function f(x) about x, what we are interested in is how we can get E[f]. This can be obtained simply by following the equation below.

$$E[f] = \int f(x)p(x)dx \quad \text{(eq. 1)}$$

However, if multiple samples can be extracted from the probability distribution p(x), the approximate value for E[f] can be obtained through the following calculation.

$$\hat{f} = \frac{1}{L}\sum_{l=1}^{L} f(z^l) \quad \text{(eq. 2)}$$

In equation 2, z^l is each sample obtained through sampling. In this way, the expected value for function f(x) can be calculated simply through sampling. However, there are many cases where it is impossible to obtain the actual p(x), and for this purpose, the sampling techniques are very diverse.

2. Types of sampling

2.1. Inverse CDF method

The inverse CDF method is one in which the CDF of all probability distributions follows a uniform distribution. If we generate a random number of random variable X that follows a specific probability distribution f(x), the cumulative distribution function F(x) of all random variables will follow a uniform distribution, and when a basic random number is generated, the uniform distribution ranges from 0 to 1. Therefore, it is possible to easily implement a random number generator using inverse cumulative distribution and the CDF (Maschio and Schiozer 2019).

If the inverse function of F(x), which is the CDF of the random variable X, can be calculated, then sampling for the random variable X is possible using equation 3.

$$\begin{aligned} & X \sim f(x) \\ \Rightarrow & F(X) \equiv U \sim Unif(0,1) \\ & F(X) \equiv U \\ \Leftrightarrow & F(X) \cdot F^{-1}(X) = F^{-1}(U) \\ \Leftrightarrow & X = F^{-1}(U) \end{aligned} \quad \text{(eq. 3)}$$

2.2. Rejection sampling

Assuming that the objective is to obtain a sample from a

specific target function f(x), sampling becomes difficult if the distribution of the function f(x) to be sampled is not a commonly known distribution. In this case, another function e(x) that can cover all f(x) is created and used with g(x), an approximation function of the function f(x) (Xia *et al*. 2018).

2.3. Squeezing rejection sampling

Squeezing rejection sampling is simply a method of sampling using the proposal density, which is similar to that in rejection sampling (Cheng *et al*. 2015). However, in this case, two proposed distributions are used instead of one. Rejection sampling is an improved sampling method to compensate for the shortcomings of squeezing rejection sampling.

2.4. Adaptive rejection sampling, ARS

Adaptive rejection sampling (ARS) is a method of generating a better probability sample by generating a sample through a proposed distribution created by connecting a tangent line to a log-transformed target density and improving the proposed candidate using this (Gilks *et al*. 1995).

2.5. Importance sampling

Rejection sampling has a disadvantage in that it takes a long time to obtain a sample of a desired size because the rejection rate is very large. To compensate for these shortcomings, the importance sampling method enables sampling without rejection, that is, no discarded samples (Merle *et al*. 2016). It is a method designed to efficiently estimate the expected value. It is used in various applications such as probability density estimation and reinforcement learning.

If we assume that we want to get the expected value for a function, there will be a random variable that goes into that function, and there will be a probability distribution that creates that random variable. Here, the random variable is set to X, and the PDF is set to f(x).

3. Simulation

R has been used since 1993 as a free version of S-PLUS by Ross Ihaka and Robert Gentleman of the University of Auckland, New Zealand. R almost absorbed users of the existing S-PLUS (S language), and despite being open source, it has high-performance computing speed and data processing capability and various software and APIs with Google and Amazon cloud services, etc. (Grunsky 2002). R is an open source program and has a language for statistics/data mining and graphing. R is mainly used in research and industry-specific applications, and recently, companies have started to use it. In particular, it is attracting attention for the purpose of big data analysis, and more than 5000 packages support various functions and are being updated from time to time (Braun and Murdoch 2016).

RESULTS

1. Inverse CDF method

The inverse CDF method is a method that uses a feature in which the cumulative distribution function (CDF) of all probability distributions follows a uniform distribution. As shown in equation 4, an arbitrary exponential function was defined and the sampling method was performed using it.

$$X \sim \mathrm{Exp}(\lambda)\ (\lambda > 0)$$
$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}\ (0 \le x \le \infty) \qquad \text{(eq. 4)}$$

$$F(X) \equiv U$$
$$\Leftrightarrow F(x) = \int_0^x f(t)dt = \int_0^x \frac{1}{\lambda} e^{-\frac{t}{\lambda}} = 1 - e^{-\frac{x}{\lambda}} = u \qquad \text{(eq. 5)}$$

$$X \equiv F^{-1}(U)$$
$$\Leftrightarrow e^{-\frac{x}{\lambda}} = 1 - u$$
$$\Leftrightarrow -\frac{x}{\lambda} = \log(1-u)$$
$$\Leftrightarrow x = -\lambda \log(1-u)$$
$$\Leftrightarrow x = -\lambda \log(u) \qquad \text{(eq. 6)}$$

The result is the same as substituting U for the inverse function of the cumulative distribution function as in equation (6). That is, a random variable that follows an exponential distribution can be created by substituting U into the inverse function of F(x).

Through the R uniformly distributed random number generation function runif, a total of 100000 uniformly distributed random numbers following unif (0,1) were generated. The resulting U-value was substituted into the inverse function of the cumulative distribution function of the exponential distribution. In Table (2), the distribution of X created by the random number generator was drawn with the hist ()

Table 1. R was used for simulation. Through the equal distribution random number generation function runif, a total of 100,000 equal distribution random numbers were generated following unif (0,1)

```
## Let's Create a random number generator using the Inverse Cumulative Distribution Function.
## Suppose the random variable X is from Exponential dist w/ parameters lamda = 1.[X~Exp(1)]

U = runif (100000, min = 0, max = 1)
X = − log (U)
```

Table 2. Coding to check table random number generation

```
#Check the distribution of the random number generator.
#"freq = FALSE" is based on probability density, not frequency.

hist (X, nclass = 100, freq = FALSE)

# Let's compare with the built-in random number generator "rexp ()"
Y = rexp (100000, 1)
lines (density (Y), lwd = "2")
```

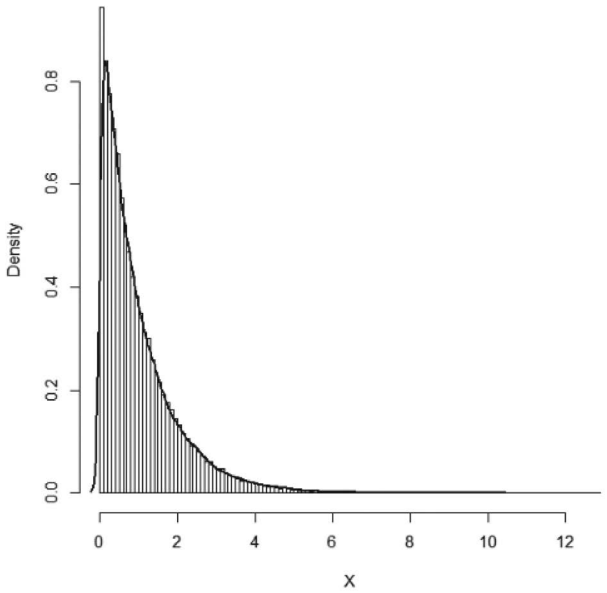

Fig. 1. The result of computational simulation using R: coincides with the random number generator of exponential distribution and the graph.

function to check whether the exponential random number generator works well. In addition, in order to compare with the random number generator of the exponential distribution built in R, the PDF was drawn as a thick line with the lines () function.

The results of computational simulation using R are shown in Fig. 1. As the two graphs match, it can be confirmed that the random number generator of the generated exponential distribution is working well. In other words, we confirmed that samples following the exponential distribution were well produced.

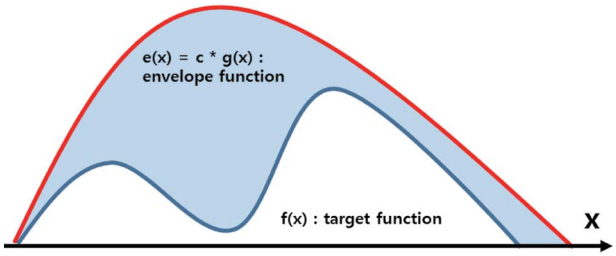

Fig. 2. Relationship chart between envelope function and target function.

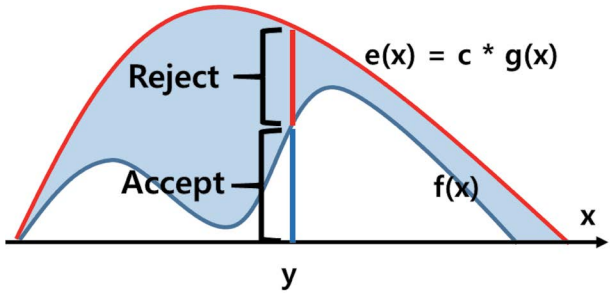

Fig. 3. Schematic explaining Rejection Sampling.

2. Rejection sampling

The sampling probability was corrected by sampling candidates from a distribution e(x) easier than the target density, f(x), and then randomly rejecting some candidates. e(x) is the proposal density. For the proposal density, uniform distribution and normal distribution were available. Whenever possible, it was better to use a distribution similar to f(x).

g(x) is a distribution in which the sample knows how to calculate easily, and the constant c is a constant that allows g(x) to cover all of f(x). By multiplying g(x) and c, we made a cover function as shown in equation 7.

$$e(x) = c^* g(x) \geq f(x), \;\; c = \max_x \frac{f(x)}{g(x)} \geq 1 \qquad \text{(eq. 7)}$$

Table 3. Rejection Sampling Algorithm Description

Step 1. Create Sample Y from g(x).
Step 2. To evaluate the sample, sample U is created from a uniform distribution.
Step 3. Calculate the rejection ratio.
Step 4. If U ≥ Rejection Ratio, sample Y is rejected. And it goes back to step1.
Step 5. If U ≤ Rejection Ratio, accept sample Y. And by specifying Y = X, it is saved as a sample of target density f(x). Then go back to step 1 and repeat the algorithm until the desired number of samples is obtained.

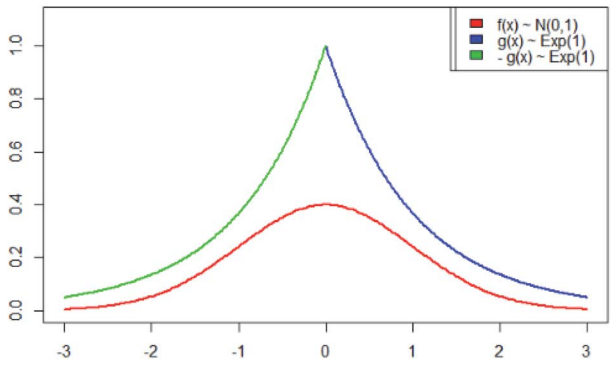

Fig. 4. Target function and proposed function: f(x) is the target function, and g(x) is the proposed function.

$$\text{rejection ratio} = \frac{f(y)}{e(y)} = \frac{f(y)}{c * g(y)} \quad \text{(eq. 8)}$$

Rejection Sampling can be used to efficiently sample if you know the PDF of the target function, but it is very difficult or impossible to sample directly from that function.

When extracting a sample from a certain probability distribution, this probability distribution is called the target density and you need to know the probability density function of the target probability distribution. If at least the proportionality constant of the PDF can be calculated, a sample can be accurately obtained from the target density using rejection sampling g.

Rejection conditions were when the rejection ratio was greater than that of the uniformly distributed sample U.

Reject sampling was implemented under the assumption that samples that follow a normal distribution are extracted using an exponential distribution.

The distribution of the target function and the proposed function was as follows.

- Target density : f(x)~N(0,1) $(-\infty \le x \le +\infty)$
- Proposal density : g(x)~Exp(1) $(0 \le x \le +\infty)$

Even if the x ranges of the two distributions were different, this was not a big problem because it could be adjusted to fit the range of the target density during the sampling process.

As shown in Fig. 4, since the exponential distribution wraps around the target density, which is a normal distribution, in the interval where x is positive, it can be confirmed that the exponential distribution is valid as the proposal density.

The procedure for extracting a sample that follows a normal distribution using the Rejection Sampling method is as follows. First, to find the envelope density e(x), was calculated a constant c suitable for g(x) and then defined e(x). Then, the rejection ratio was calculated.

In equation 9, the process of defining the constant c value and the covering function for the normal distribution function is explained.

$$f(x) \sim N(0,1),\ f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \ (-\infty \le x \le \infty)$$

$$g(x) \sim \mathrm{Exp}(1),\ g(x) = e^{-x} \ (0 \le x \le \infty)$$

$$c = \max \frac{f(x)}{g(x)} = \log \frac{f(x)}{g(x)} = \log f(x) - \log(g)$$

$$= -\frac{x^2}{2} + constant$$

$$\frac{\partial}{\partial x}\left(\log \frac{f(x)}{g(x)}\right) = -x + 1 = 0 \quad \therefore x = 1$$

$$c = \frac{f(x=1)}{g(x=1)} = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}}$$

$$e(x) = c \times g(y) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}} \times e^{-y} = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}-y}$$

Rejection ratio :

$$r = \frac{f(y)}{e(y)} = \frac{f(y)}{c \times g(y)} \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}}{\frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}-y}} = e^{-\frac{1}{2}(y-1)^2} \quad \text{(eq. 9)}$$

To check the sampling result, x (sample data) values extracted by rejection sampling were plotted as a histogram. The normal distribution function built into R was used, and the normal distribution graph was drawn with a thick solid line. As shown in Fig. 5, the histogram and the solid line match, which confirms that samples that follow the normal distribution were well produced.

Table 4. Computational simulation syntax implementing rejection sampling

```
## Rejection Sampling.
## Consider generating a Normal (0,1) by using Exponential (1).
n = 10000          # Sample size
x = rep (NA, n)    # Sample of target dist (normal).

index = 1
while (index < = n) {
    #---Step1. Generate Proposal density.----------#
    y = rexp (1)
    #---Step2. Calculate acceptance ratio.----------#
    r = exp (- (y-1)^2/2)
    #---Step3. Rejection Sampling.----------#
    u = runif (1)          # For evaluate samples.
    if (u < r){            # Condition of acceptance.
      u2 = runif (1)       # For adjustment range of samples.
      if (u2 < 0.5) x[index] = - y
                else x[index] = y

    index = index + 1
    }
  }

## Check the distribution of generating samples.
hist (x, nclass = 100, freq = FALSE)

## Compare with the built-in random number generator "rnorm ()"
x1 < - seq (-4, 4, length = 200)
lines (x1, dnorm (x1, mean = 0, sd = 1),type = 'l', col = 1, lwd = "2")
```

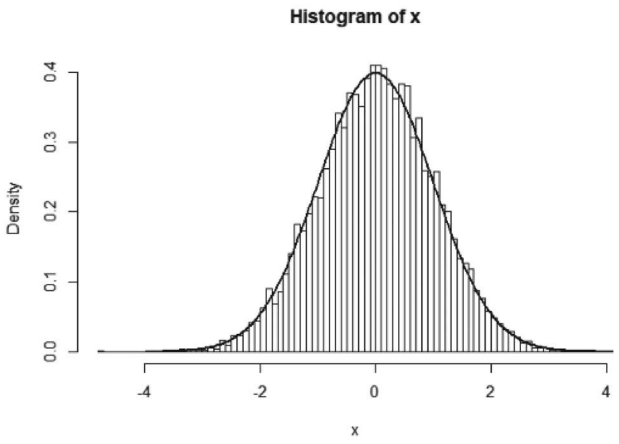

Fig. 5. Sample extraction result of function following normal distribution using exponential distribution with R.

3. Squeezing rejection sampling

Rejection sampling generally has a high rejection rate, so it takes much time to sample from the target function f(x). Squeezing rejection sampling is a method of sampling using the proposal density as in Rejection Sampling. However, two proposed distributions were used instead of one. The envelope function was defined as a distribution with a value greater than the target density f(x) in all regions of x as in rejection sampling. Squeezing function s(x) cannot exceed target density f(x) in all areas of x. It consists of a distribution smaller than the f(x) value. By setting the proposal density to two in this way, the rejection area is minimized, so it is possible to sample at a faster rate. However, because the equation is complex, the computational complexity increases.

The calculation of the envelope function e(x) and the squeezing function s(x) is as follows. It is defined as in equation (10) in the same way as the envelope function used in rejection sampling.

$$e(x) = c * g(x) \geq f(x), \quad c = \max_x \frac{f(x)}{g(x)} \geq 1 \quad \text{(eq. 10)}$$

The squeezing function s(x) is a function related to the Taylor series and could be easily obtained by using the Taylor expansion of the target function f(x).

Assuming that the Taylor series is expressed by the following equation, let s(x) be the equation excluding the remaining term (R_n).

Then, since s (x) is always the lower bound to f(x), it can be used as a squeezing function.

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_{n+1}$$

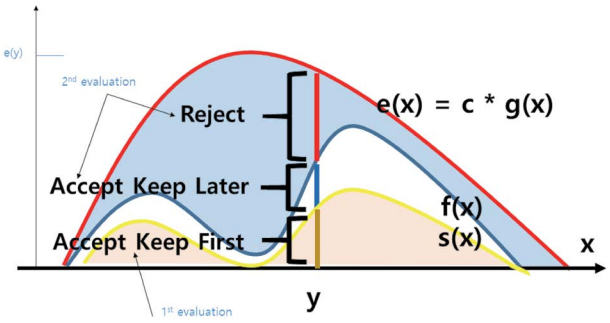

Fig. 6. Graph for explaining Squeezing Rejection Sampling.

Table 5. The process of squeezing rejection sampling algorithm

Step 1. Create sample Y in g(x).
Step 2. To evaluate Sample Y, sample U is created from a uniform distribution.
Step 3. Calculate the 1st Rejection Ratio (r1) and the 2nd Rejection Ratio (r2).
Step 4. If U ≤ r1 = s(y)/e(y), accept sample Y.
And by specifying Y = X, it is saved as a sample of target density f(x).
And go to Step 7.
Step 5. Otherwise, if U ≥ r1, we evaluate whether U ≤ r2 = f(y)/e(y).
If the second inequality is satisfied, sample Y is accepted.
And by specifying Y = X, it is saved as a sample of target density f(x).
And go to Step 7.
Step 6. If U ≥ r2 = f(y)/e(y), sample Y is rejected.
And go to Step 7.
Step 7. Go back to Step 1 and repeat the algorithm until you get the desired number of samples.

Table 6. Adaptive rejection sampling ARSgorithm

Step 1. Define the initial number of supports k and set. Ex) s1, s2, s3
Step 2. Create a function for each section for the squeezing function and envelope function using k support points. Ex) (s1, s2) and (s2, s3)
Step 3. Perform Squeezing Rejection Sampling.
Step 4. Update the support points.
Step 5. Go back to step1.

$$\text{then } s(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k \le f(x) \qquad \text{(eq. 11)}$$

At this time, the rejection ratio is as shown in equation 12.

$$1^{st} \text{ Rejection Ratio : } r_1 = \frac{s(y)}{e(y)}$$

$$2^{nd} \text{ Rejection Ratio : } r_2 = \frac{f(y)}{e(y)} = \frac{f(y)}{c * g(y)} \qquad \text{(eq. 12)}$$

4. Adaptive rejection sampling

In rejection sampling it is often not easy to find the envelope function e(x) itself. Adaptive Rejection Sampling is a sampling method that has a strategy of automatically generating an envelope function for squeezing rejection sampling. That is, it can be seen as an improved sampling method that compensates for the disadvantages of rejection sampling and squeezing rejection sampling.

The ARS method was a method proposed to improve the disadvantages of the existing rejection sampling method. A sample was created through a proposal distribution created by connecting the tangent line to the log-transformed target density. After that, it is a method of generating a better probability sample by improving the proposed candidate using this. However, the use of this sampling method was limited because the target function f(x) must be continuous and differentiable in all ranges of X, and it can only be applied to a kind of bell-shaped distribution in which the function is log-concave.

The most representative advantage of ARS is that the envelope function (cover function) used in reject sampling is used as a piecewise linear envelope function (piecewise linear envelope function) created by connecting a tangent line to a log-transformed probability distribution. It is improved to be close to the target function in the process of sampling. In other words, it has the effect of reducing the expected number of iterations, which is more efficient than rejection sampling.

ARS generated a sample through a proposal distribution created by connecting a tangent line to a log-transformed target density. It is a method for generating better probability samples by improving the proposed candidate using this. However, it was not suitable as the dimensions increased. When the Gaussian distribution p(z) and the cover function are q(z), an equation must be established to satisfy $kq(z) \le p(z)$.

In the end $\sigma_q^2 \le \sigma_q^2$ is required. When this is expanded, it can be seen that the constant k becomes $k = (\sigma_q/\sigma_p)^D$. When the dimension increases in one dimension for the Gaussian distribution-covered function, the difference area is very small, and the rejection rate is expected to be very low. However, in the formula, the adoption rate is inversely proportional to the dimension D, so this problem eventually

becomes a problem in which k increases exponentially with respect to the input dimension. If the dimension of the sample is D = 1000, the adoption rate is 1/20000. Therefore, when the dimension increases, a situation in which it is very difficult to obtain a sample will occur.

5. Importance sampling

In rejection sampling methods, the rejection rate during sample extraction is very large, so it takes a long time to obtain a sample of a desired size. The Importance Sampling method makes sampling possible without wasted samples (Merle *et al*. 2016). The purpose of sampling first is the calculation of the average value of a specific probability density function second is the calculation of a specific probability value.

If it is sampling to calculate a specific value, it does not need to take many samples because you only need to know that value. In other words, it is necessary to efficiently sample so that no samples are discarded during sampling. Most problems in statistics are closely related to finding the expected value of a particular probability distribution. Importance sampling is a method designed to efficiently estimate expected values.

Importance sampling has been used in various applications such as probability density estimation and reinforcement learning. Assuming something want to find an expected value for a function, there will be a random variable that goes into that function, and there will be a probability density that creates that random variable. Here, the random variable is set as X (= sample data), and the probability distribution function (PDF) is f(x) (= target density).

$$E[f] = \int f(x)p(x)dx \qquad \text{(eq. 13)}$$

The average is shown in equation 13. As seen in previous sampling, in reality, it is difficult to sample at f(x), so we tried a sampling method using the approximate distribution g(x). Since we cannot sample directly at f(x), we have to sample instead from the approximate distribution g(x), which is easy to sample. In the formula, if you multiply and divide the approximate distribution g(x), it could be calculated as follows.

$$E[f] = \int f(x)p(x)dx \int p(x)\frac{f(x)}{g(x)}g(x)dx$$

$$= E_g\left[p(x)\frac{f(x)}{g(x)}\right] \qquad \text{(eq. 14)}$$

equation 14 above means adding an infinite number of instances, but in reality, it is impossible to find infinite instances. This is because 100% of the target population is unknown. Instead, we could sample an individual instance from f(x) and calculate the probability of that individual instance occurring. Therefore, the sampling was performed using the formula calculated by changing it to the form of addition to the normalizing constant as follows.

$$E[f] \cong \frac{1}{n}\sum_{i=0}^{n} p(x)\frac{f(x)}{g(x)} \qquad \text{(eq. 15)}$$

$$E[f] \cong \frac{1}{n}\sum_{i=0}^{n} p(x)\frac{f(x)}{g(x)} = \frac{1}{n}\sum_{i=0}^{n} p(x_i)w(x_i)$$

$$*\textit{weight}: w(x_i) = \frac{f(x_i)}{g(x_i)} \qquad \text{(eq. 16)}$$

Looking at equation (16), it can be seen that w(x) is acting as a weight for the mean of h(x). In other words, it can be seen that the average is weighted. Therefore, weight is regarded as important, and this is called importance sampling or weight sampling. Moreover, weights are precisely called "unstandardized weights" or "importance ratios". Importance sampling, like the reject-sampling method, needs to have the cover function as close as possible to the objective function in order for this sampling method to have good results.

It is very common, but if the objective function has a variable shape and only has a value in a local area of the data space, the influence of the main weight will be largely influenced by some of the values. Therefore, that method may produce poor results.

CONCLUSION

The Inverse CDF Method was a very basic method and had the disadvantage of having to find the inverse function. During the sampling process, sampling was difficult near $-\infty$ and ∞. The purpose of Rejection Sampling is to obtain a sample from a specific target function f(x). If the distribution of the function f(x) to be sampled is not a commonly known distribution, sampling is performed, and it becomes

difficult. In this case, another cover function e(x) that can cover all f(x) is created, which is an approximate function of function f(x). However, in order to cover all of target function f(x), the value of c must be increased, which may take much time. Therefore, it can be effective only when the distribution most similar to the target function f(x) is used.

The advantage of Adaptive Rejection Sampling (ARS) is that the rejected area is minimized, and in this study, we were able to generate samples at a faster rate. However, because the equation is complex, the computational complexity is greater, and when the dimensions increase, it becomes very difficult to obtain a sample.

Importance sampling, like the reject sampling method, has to have a cover function as close as possible to the objective function in order for this sampling method to produce good results. If the objective function has a large change and has a value only in a local area of the data space, it has a disadvantage in that it may produce poor results.

ACKNOWLEDGEMENTS

This paper is a research conducted with the support of the National Research Foundation of Korea (NRF) with the funding of the government (Ministry of Science and ICT) (NRF-2020M2D2A2062571).

REFERENCES

Braun WJ and Murdoch DJ. 2016. A first course in statistical programming with R, Cambridge University Press.

Cheng L, Lu Z and Zhang L. 2015. Application of Rejection Sampling based methodology to variance based parametric sensitivity analysis. *Reliab. Eng. Syst. Saf.* **142**:9-18.

Christopher T, AllenGeorge S and EllenHaupt YS. 2007. Improving pollutant source characterization by better estimating wind direction with a genetic algorithm. *Atmos. Environ.* **41**:2283-2289.

Connan O, Smith K, Organo C, Solier L, Maro D and Hébert D. 2013. Comparison of RIMPUFF, HYSPLIT, ADMS atmospheric dispersion model outputs, using emergency response procedures, with 85Kr measurements made in the vicinity of nuclear reprocessing plant. *J. Environ. Radioact.* **124**:266-277.

Gilks WR, Best NG and Tan KKC. 1995. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *JR Stat. Soc. Ser C Appl. Stat.* **44**:455-472.

Grunsky EC. 2002. R: a data analysis and statistical programming environment-an emerging tool for the geosciences. *Comput. Geosci.* **28**(10):1219-1222.

Hutchinson M, Oh H and Chen W. 2017. A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors. *Inf. Fusion* **36**:130-148.

Maschio C and Schiozer DJ. 2019. A new parameterization method for data assimilation and uncertainty assessment for complex carbonate reservoir models based on cumulative distribution function. *J. Pet. Sci. Eng.* **183**:106400.

Merle C, Leblois R, Rousset F and Pudlo P. 2016. Resampling: an improvement of Importance Sampling in varying population size models. *Theor. Popul. Biol.* **114**:70-87.

Xia Y, Yang X and Zhang Y. 2018. A rejection inference technique based on contrastive pessimistic likelihood estimation for P2P lending. *Electron. Commer. Res. Appl.* **30**:111-124.

Received: 16 October 2020
Revised: 2 November 2020
Revision accepted: 14 November 2020