Comparison of Different Techniques to Generate Normal Random Variables*

Ritabrata Roy

November 4, 2002

Abstract

This exercise aims at exploring different techniques for creating a random variable X according to a normal distribution with zero mean and unit variance. The methods include the use of an inverse cumulative distribution function, the Box–Muller method, the polar technique and the application of the Central Limit Theorems to uniform random variables. The normal random variables generated by these methods are then compared according to different performance metrics, including their mean, variance and kurtosis, and conclusions are drawn about the performance of each of the techniques.

1 Introduction

The Gaussian distribution, or normal distribution, is a continuous distribution function that is defined completely by the mean and variance of the distribution. The probability density function (pdf) of a Gaussian random variable X is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

*Submitted as partial fulfillment of course requirement in Stochastic Signals and Systems (ECE 330:541) to Dr. Wade Trappe, Assistant Professor, Rutgers, The State University of New Jersey.

where, μ: mean of the distribution
σ^2: variance of the distribution

The Gaussian distribution is real-valued and symmetric about the mean. A standard normal distribution is obtained by putting $\mu = 0$ and $\sigma^2 = 1$ in (1) and is typically represented as $N(0, 1)$.

The Gaussian distribution, in the form of a bell-shaped curve, appears in several man-made and natural phenomena and has been christened *normal* as a tribute to its ubiquity. The frequent ocurrence of the distribution follows from the Central Limit Theorems, which state that the mean of a set of variates with any distribution having a finite mean and variance tends to the Gaussian distribution. As pointed out in [1], in nature, many macroscopic phenomena result from the addition of numerous independent, microscopic processes; this gives rise to the Gaussian random variable.

Historically, the normal distribution was first introduced by de Moivre in the second edition of his *Doctrine of Chances* (1718), in the context of approximations of large binomial coefficients. His result was extended by Laplace in his book *Analytical Theory of Probabilities* (1812), and is now called Theorem of de Moivre-Laplace. Around that time the analysis of errors of experiments was pioneered by Laplace, Legendre and Gauss. The distribution appearing in the theorem of de Moivre–Laplace was called Gaussian as a result of Gauss' work on the method of least squares, introduced by Legendre in the context of the theory of errors. Towards the end of the 19th century, Pearson established the priority of de Moivre, and Poincar coined the name "normal" [2].

2 Generation of Normal Random Variables

This exercise calls for the generation of a standard normal distribution using the following four techniques:

1. Inverse Distribution
2. Box–Muller Method
3. Polar Technique

4. Sums of Uniform Random Variables

For all the methods, 1000 Gaussian random variables were generated and their frequency distribution plotted in the form of a histogram. The mean and variance of the distributions were also calculated, since a normal distribution is completely characterised by these parameters. According to the requirements of the problem these values should be as follows:

- Mean, $\mu = 0$
- Variance, $\sigma^2 = 1$ (*i.e.*, standard deviation, $\sigma = 1$)

The histogram technique obviously necessitates a certain amount of discretisation of the data, but by selecting a sufficently small interval (0.01), a fairly accurate distribution was obtained. It may be pointed out that although the distribution has the classic bell-shaped form, it does not represent the probability density function (pdf) of the Gaussian random variable.

2.1 Inverse Distribution

A random variable X can be created by computing the inverse cumulative distribution function (cdf) of a set of uniformly distributed random numbers. Thus, a Gaussian random variable may be generated by chossing a random (uniformly distributed) real number between 0 and 1, since they are the lower and upper bounds respectively of a cdf, and applying F_X^{-1}, the inverse cdf function, to obtain the random variable X. The inverse cdf function for a Gaussian variable does not have a closed form expression, although there are several approximations available in the literature. One of them is to obtain it in terms of the $Q(x)$ function, and which has been provided as a guideline in the problem statement. However, taking advantage of the freedom to choose an equivalent approximation and by exploiting the powerful built-in functions in MATLAB, a more direct approach was used to obtain the inverse cdf, as has been described below.

The MATLAB function `normcdf` can calculate the normal cdf with a particular mean μ and standard deviation σ corresponding to a particular x. Thus, taking advantage of the fact that the cdf is a monotonically increasing function, the inverse cdf corresponding to a random number y can be obtained by calculating `normcdf(x)` starting at -3 (where the cdf is 0.0013)

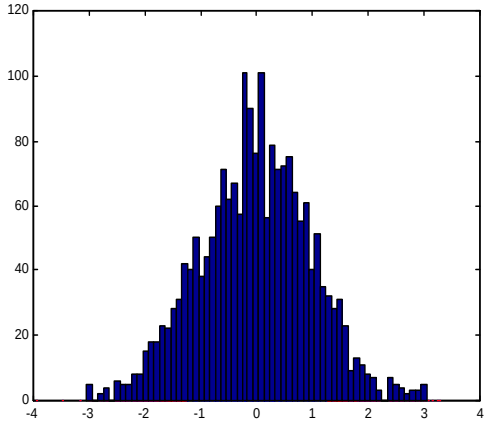

Figure 1: MATLAB plot of frequency distribution (histogram) of random variables generated by the inverse cdf method. The abscissa represents the value of the random variable and the ordinate represents the number of corresponding occurrences.

and progressively increasing the value of x until the following condition is satisfied within a very small error:

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{2}$$

This gives $x = \text{cdf}^{-1}(y)$ and forms the basis for the method used to generate the Gaussian random variable. It may be pointed out that while this technique is computationally inefficent, it can obtain the inverse cdf with a high degree of accuracy since it is equivalent to a look-up table method whose precision depends on the increment value of the variable.

In the execution of the algorithm (MATLAB source code in Appendix A.1), the variable was incremented by 0.01, which is quite a conservative value, since standard cdf tables use the same increment, and a mean interpolation was performed for the final value. Fig. 1 displays a typical frequency distribution obtained for a generated set of 1000 Gaussian random variable. The average parameters of the distribution obtained from 10 successive simulations (with different random number seeds) were found to be:

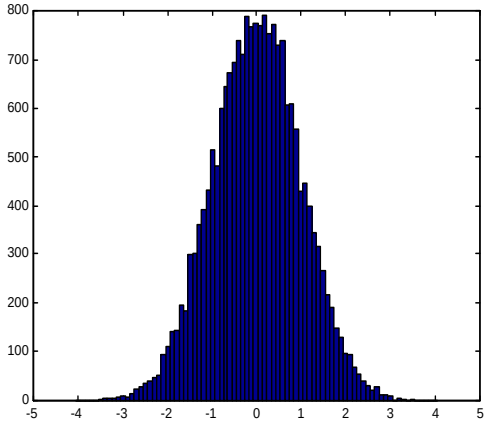

Figure 2: MATLAB plot of frequency distribution (histogram) of random variables generated by the Box–Muller method. The abscissa represents the value of the random variable and the ordinate represents the number of corresponding occurrences.

- Mean, $\mu = 0.0232$
- Variance, $\sigma^2 = 1.03001$

2.2 Box–Muller Method

The Box–Muller method makes use of the fundamental transformation law of probabilites to convert from a two-dimensional continuous uniform distribution (uniformly and independently distributed between 0 and 1) to a two-dimensional Gaussian bivariate distribution. Thus the method gives a set of random numbers which have a Gaussian distribution with zero mean and unit standard deviation. The result may be proved by solving for the uniform random variables in terms of the normal random variables and taking the Jacobian which yields the desired result [3].

$$y_1 = \sqrt{-2\ln(x_1)}\cos(2\pi x_2) \tag{3}$$

$$y_2 = \sqrt{-2\ln(x_2)}\cos(2\pi x_1) \tag{4}$$

Thus,

$$x_1 = e^{-\frac{(y_1^2+y_2^2)}{2}} \tag{5}$$

$$x_2 = \frac{1}{2\pi}\tan^{-1}(\frac{y_2}{y_1}) \tag{6}$$

Taking the Jacobian yields,

$$\frac{\delta(x_1, x_2)}{\delta(y_1, y_2} = -[\frac{1}{2\pi}e^{-\frac{y_1^2}{2}}][\frac{1}{2\pi}e^{-\frac{y_2^2}{2}}] \tag{7}$$

In the execution of the algorithm (MATLAB source code in Appendix A.2), the frequency distribution that was obtained for a typical set of 1000 generated Gaussian random variable is shown in Fig. 2. The average parameters of the distribution obtained from 10 successive simulations (with different random number seeds) were as follows:

- Mean, $\mu = 0.01233$
- Variance, $\sigma^2 = 0.99126$

2.3 Polar Technique

Another form of the Box–Muller method is called the polar technique. This improves over the previous technique in being quicker (since it makes fewer calls to the mathematical library and uses only one transcendental function, instead of three) as well as numerically more robust (the previous method has numerical stability problems when x_1 or x_2 is very close to zero) [2].

The algorithm can be summarised as follows:

1. Let U_1 and U_2 be independent and identically distributed (iid) uniform random variables.
2. $V_i = 2U_i - 1$ and define $W = V_1^2 + V_2^2$.
3. If $W > 1$ then go back to step 1, else let $y = \sqrt{\frac{-2\ln(W)}{W}}$.
4. Then, $X_1 = V_1 Y$ and $X_2 = V_2 Y$ are iid $N(0, 1)$.

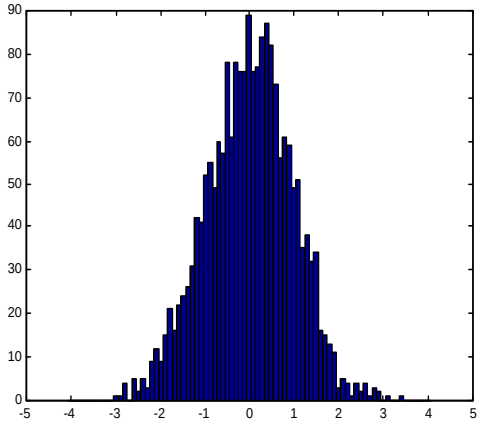

Figure 3: MATLAB plot of frequency distribution (histogram) of random variables generated by the polar technique. The abscissa represents the value of the random variable and the ordinate represents the number of corresponding occurrences.

This method is advantageous of this method over the first form of Box–Muller's method in spite of the fact that the algorithm discards 21% of the values of W in step 3.

The algorithm (MATLAB source code in Appendix A.3) was executed for 1000 iterations in each of its runs and a typical frequency distribution of the Gaussian random variables that was obtained is shown in Fig. 3. The average parameters of the distribution obtained from 10 successive simulations (with different random number seeds) were found to be:

- Mean, $\mu = 0.0034$
- Variance, $\sigma^2 = 0.99908$

2.4 Sum of Uniform Random Variables

As was briefly touched upon in the Introduction, the Central Limit Theorems are a set of weak convergence results that express the fact that any sum of many small independent random variables is approximately normally

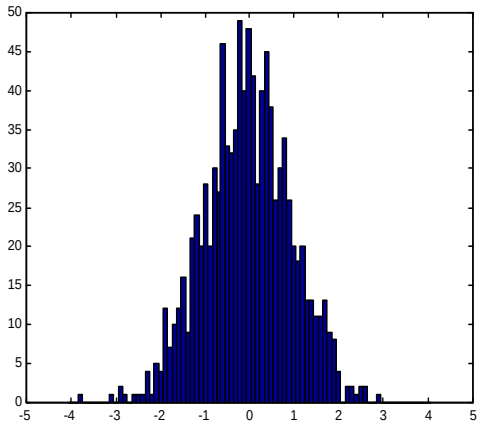

Figure 4: MATLAB plot of frequency distribution (histogram) of random variables generated by the sum of random variables method. The abscissa represents the value of the random variable and the ordinate represents the number of corresponding occurrences.

distributed.

More formally, this result can be expressed as follows [4]: Let $W_n = X_1 + \cdots + X_n$ be an iid random sum with $E[X] = \mu_X$ and $\mathrm{Var}[X] = \sigma_X^2$. The cdf of W_n may be approximated by

$$F_{W_n}(w) \approx \Phi(\frac{w - n\mu_X}{\sqrt{n\sigma_x^2}}) \tag{8}$$

where $\Phi(\cdot)$ represents the cdf of a Gaussian distribution.

For a uniform random variable in the range $(0, 1)$, the mean and variance are given by 1 and $\frac{1}{12}$ respectively. Hence, the normal random variable may be approximated as:

$$Y = \frac{\Sigma_{i=1}^{n} X_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \tag{9}$$

for a sufficiently high value of n.

In the execution of the algorithm (MATLAB code in the Appendix), 1000 uniformly distributed random variables were used to generate each Gaussian random variable. The frequency distribution for 1000 of these Gaussian random variables in a typical run is shown in Fig. 4. The average parameters of the distribution obtained from 10 successive simulations (with different random number seeds) were calculated to be:

- Mean, $\mu = -0.00377$
- Variance, $\sigma^2 = 1.00722$

3 Results

As has been mentioned earlier, a Gaussian distribution is completely defined by its mean and variance. It follows, therefore, that a natural choice of performance metrics would involve the mean and variance obtained for each of the simulations. Since the exact values of the mean and variance for a standard normal distribution are known (namely, $\eta = 0$ and $\sigma^2 = 1$ respectively), it is possible to treat the deviation of the obtained values from the expected values as an *error* and consequently calculate the following error estimates:

- mean square error,
- mean absolute error, and
- maximum absolute error.

However, the use of error estimates to compute the discrepancy in the mean and variance does not address the more central issue about the *gaussianity* of the distribution. To take an extreme example, a standard Laplacian distribution with zero mean and unit variance would return a zero error estimate, and yet would be a far cry from the expected Gaussian curve. To overcome this shortcoming, use is also made of other standard measures for the nongaussianity of a distribution. These performance metrics, like *kurtosis* and *negentropy*, are discussed in the following subsection.

The table above summarises the results obtained for mean, variance and kurtosis for ten runs of each of the four algorithms discussed in Section II.

	1	2	3	4	5	6	7	8	9	10	Average
μ_1	.037	.048	.069	.055	.026	.032	.003	-.014	.021	-.045	0.02320
σ_1^2	1.101	1.044	.987	.998	.978	1.057	.997	1.011	1.043	1.085	1.03001
K_1	-.045	-.074	-.027	-.018	.101	-.218	-.034	-.163	.073	-.180	-0.05841
μ_2	.008	.034	.032	-.006	.045	.042	-.004	.036	-.033	-.032	0.01233
σ_2^2	.985	1.011	.970	.955	1.026	.935	.978	1.031	1.010	1.011	0.99126
K_2	-.029	.079	.009	-.033	-.032	.010	.096	.059	-.078	.029	0.01096
μ_3	-.007	.039	.024	-.004	-.004	.014	.024	-.016	-.051	.015	0.00340
σ_3^2	.993	1.032	.991	1.006	.970	1.002	.997	.996	.999	1.003	0.99908
K_3	-.032	-.074	-.027	-.097	-.028	-.053	-.137	.174	.035	-.161	-0.04009
μ_4	-.043	.037	-.038	-.064	.027	.028	.027	-.020	-.013	.022	-0.00377
σ_4^2	.946	1.015	.973	.953	1.041	.988	1.073	1.037	1.001	1.045	1.00722
K_4	.018	.073	.012	-.097	-.025	-.292	-.266	-.079	-.107	-.027	-0.07912

3.1 Measures of Nongaussianity

Certain standard performance metrics are used in Independent Component Analysis to reveal hidden patterns underlying sets of random variables. Common metrics for measuring the degree of nongaussianity in a distribution include kurtosis and negentropy [5]

The classical measure of nongaussianity is the kurtosis or the fourth order cumulant, and is defined by [6]:

$$K(x) = \frac{E[(x-\mu)^4]}{\sigma^4} \tag{10}$$

where, x: random variable
μ: mean (first moment) of the distribution
σ^4: fourth moment of the distribution

The kurtosis for a standard normal distribution is 3. For that reason, some authors prefer to represent kurtosis as the *excess* value over 3, and define it as follows:

$$K(x) = \frac{E[(x-\mu)^4]}{\sigma^4} - 3 \tag{11}$$

This second definition is used in the analysis that follows since it leads to a more natural interpretation of nongaussianity.

Random variables that have a negative kurtosis are called subgaussian, and those with positive kurtosis are called supergaussian. In statistical literature, the corresponding expressions *platykurtic* and *leptokurtic* are also used. Supergaussian random variables typically have a "spiky" pdf with heavy tails, *i.e.*, the pdf is relatively large at zero and at large values of the variable, while being small for intermediate values. Subgaussian random variables, on the other hand, have typically a "flat" pdf, which is rather constant near zero, and very small for larger values of the variable.

A second very important measure of nongaussianity is given by negentropy, and is defined by $J(y)$:

$$J(y) = H(y_{gauss}) - H(y) \tag{12}$$

$$H(y) = -\int f(y) \log(f(y)) dy \tag{13}$$

where, y_{gauss}: a Gaussian random variable of the same covariance matrix as y

Negentropy is always non-negative, and is zero if and only if y has a Gaussian distribution.

Because of the complexities involved in the evaluation of the negentropy, only the kurtosis was used in measuring the nongaussianity of the generated random variables in this study. It may also be pointed out that since the Statistical Toolbox of MATLAB 6 defines the kurtosis of a standard normal distribution as 3, a modified version of the function `kurtosis` was used in the simulations.

Applying the definitions of kurtosis to the data summarised in the table, it becomes obvious that the kurtosis for the four methods are within 8% of the expected values and may be considered Gaussian to quite a high degree of accuracy. Moreover, while the random variables obtained from the first Box–Muller method were supergaussian, the random variables generated by the other three methods were subgaussian. At the same time, the Box–Muller distribution is closest to having an average kurtosis of zero, and is therefore nearest to a standard normal distribution. This also agrees with the fact that visually, the Box–Muller histogram has the "most pleasing" bell-shape, with the least number of spikes.

The use of kurtosis in the comparison of the different normal random variable generation schemes is motivated by the fact that although a Gaussian random variable is completely characterised by its mean and variance, the converse is not true– namely, other distributions can have the same values for these parameters as the standard normal distribution. Thus, given that a distribution is normal, the mean and variance may be used to check whether it is a standard normal distribution with excellent results. But when the distribution itself may not be Gaussian with a high degree of accuracy, the kurtosis becomes more effective as it contains information about the *shape* of the curve in a way that is not provided be merely the mean and variance.

3.2 Mean Square Error

The mean square error is one of the most commonly used performance metrics for the estimation of error or population variance. It is the average of the square of the difference between the desired response and the actual system output. One of the reasons for its popularity is that it lends itself easily to mathematical evaluation and derivations. It is defined as:

$$\epsilon = E[(X - \hat{x})^2] \tag{14}$$

where, X: true value
$\hat{x}$: estimated value

Once the kurtosis values have established that the random variables obtained are very close to a Gaussian random distribution, the mean square error metric may be used to obtain a better idea of how close the generated distributions are to the requirements of being standard normal distributed ($\eta(0, 1)$). To that end, the mean square error in the mean, variance and kurtosis (assuming the deviation of the values of the mean, variance and kurtosis from 0, 1 and 0 to be their respective "errors") were evaluated for the four methods based on a sample space of size 10 (assuming each of them to be equiprobable). The results that were obtained are listed below:

1. Inverse Cumulative Distribution
    - $\epsilon_{\mu} = 0.001582$
    - $\epsilon_{\sigma^2} = 0.002531$

- $\epsilon_K = 0.01317$

2. Box–Muller Method
   - $\epsilon_\mu = 0.00095$
   - $\epsilon_{\sigma^2} = 0.000984$
   - $\epsilon_K = 0.002875$
3. Polar Technique
   - $\epsilon_\mu = 0.000612$
   - $\epsilon_{\sigma^2} = 0.000215$
   - $\epsilon_K = 0.009639$
4. Sums of Uniform Random Variables
   - $\epsilon_\mu = 0.001199$
   - $\epsilon_{\sigma^2} = 0.001676$
   - $\epsilon_K = 0.019034$

3.3 Mean Absolute Error

The mean absolute error is a performance metric that is easy to understand, but sometimes difficult to implement mathematically. It is the average of the absolute difference between the desired response and the actual system output. It is often used in image processing applications in lieu of the mean square error estimator because it requires much less memory to calculate. The mean absolute error is defined as:

$$\epsilon = E[|X - \hat{x}|] \tag{15}$$

where, X: true value
$\hat{x}$: estimated value

Once the kurtosis values have established that the random variables obtained are very close to a Gaussian random distribution, the mean absolute error metric may be used to obtain a better idea of how close the generated distributions are to the requirements of being standard normal distributed ($\eta(0, 1)$). To that end, the mean absolute error in the mean, variance and

kurtosis (assuming the deviation of the values of the mean, variance and kurtosis from 0, 1 and 0 to be their respective "errors") were evaluated for the four methods based on a sample space of size 10 (assuming each of them to be equiprobable). The results that were obtained are listed below:

1. Inverse Cumulative Distribution
    - $\epsilon_{\mu} = 0.03498$
    - $\epsilon_{\sigma^2} = 0.03813$
    - $\epsilon_{K} = 0.09325$
2. Box–Muller Method
    - $\epsilon_{\mu} = 0.02723$
    - $\epsilon_{\sigma^2} = 0.02668$
    - $\epsilon_{K} = 0.04522$
3. Polar Technique
    - $\epsilon_{\mu} = 0.02$
    - $\epsilon_{\sigma^2} = 0.00968$
    - $\epsilon_{K} = 0.08185$
4. Sums of Uniform Random Variables
    - $\epsilon_{\mu} = 0.03181$
    - $\epsilon_{\sigma^2} = 0.0353$
    - $\epsilon_{K} = 0.09962$

3.4 Maximum Absolute Error

The maximum absolute error is a performance metric that gives the worst-case error, but its practicality is again limited by the difficulty in implementing the absolute function. It is the maximum of the absolute differences between the desired response and the actual system output. Mathematically it is defined as:

$$\epsilon = max\{|X - \hat{x}|\} \tag{16}$$

where, X: true value
$\hat{x}$: estimated value

Once the kurtosis values have established that the random variables obtained are very close to a Gaussian random distribution, the maximum absolute error metric may be used to obtain a better idea of the deviation of the generated distribution from the requirements of being standard normal distributed ($\eta(0, 1)$). To that end, the maximum absolute error in the mean, variance and kurtosis (assuming the deviation of the values of the mean, variance and kurtosis from 0, 1 and 0 to be their respective "errors") were evaluated for the four methods based on a sample space of size 10 (assuming each of them to be equiprobable). The results that were obtained are listed below:

1. Inverse Cumulative Distribution
   - $\epsilon_\mu = 0.0689$
   - $\epsilon_{\sigma^2} = 0.1014$
   - $\epsilon_K = 0.2182$
2. Box–Muller Method
   - $\epsilon_\mu = 0.0452$
   - $\epsilon_{\sigma^2} = 0.0648$
   - $\epsilon_K = 0.0956$
3. Polar Technique
   - $\epsilon_\mu = 0.0514$
   - $\epsilon_{\sigma^2} = 0.0322$
   - $\epsilon_K = 0.1735$
4. Sums of Uniform Random Variables
   - $\epsilon_\mu = 0.0637$
   - $\epsilon_{\sigma^2} = 0.0542$
   - $\epsilon_K = 0.2925$

4 Conclusion

This section attempts to draw a comparison between the different performance metrics employed to evaluate the *gaussianity* of the generated distributions. These measures range from subjective metrics that return qualitative (for instance, visual cues) information about the distribution, to more quantitative measures that summarise the same statistics in the form of a well-defined function.

Although not glorified to the same extent as the other metrics, the humble histogram that was used to plot the distribution of the random variables is itself an important tool in determining how close the distribution is to a Gaussian curve. Gaussian random variables are distributed such that the incidence of the variables is high near the mean of the distribution and falls off when moving away from the mean. Thus, the random variables generated by each of the techniques were plotted according to their frequency of occurrence, and it was seen that within limits of the resolution accuracy, the mean was equal to the mode, which is a property of a normal distribution. Thus, the histogram, or frequency distribution curve, serves an important role in giving a visual qualitative measure of the generated random variables.

The most important metric for determining the inherent gaussianity of the distribution is arguably the kurtosis. As has been explained in Section 3.1, this returns a quantitative measure for how close a distribution satisfies the properties of a normal distribution. In this case, the average kurtosis values for all the random variable generation techniques come within 8% of the desired value ($K = 0$) for a standard normal distribution. This result suggests that all the four methods are acceptable as legitimate means to create Gaussian random variables. It may be pointed out that the first Box–Muller method ($K = 0.01096$) comes closest to meeting the Gaussian requirement of zero kurtosis, despite its tendency to have numerical stability problems near zero.

Method		Mean Sq. Error	Mean Abs. Error	Max. Abs. Error
Inverse Distribution	ϵ_μ	0.001582	0.03498	0.0689
	ϵ_{σ^2}	0.002531	0.03813	0.1014
Box–Muller Method	ϵ_μ	0.000950	0.02723	0.0452
	ϵ_{σ^2}	0.000984	0.02668	0.0648
Polar Technique	ϵ_μ	0.000612	0.02000	0.0514
	ϵ_{σ^2}	0.000215	0.00097	0.0322
Sum of Uniform RVs	ϵ_μ	0.001199	0.03181	0.0637
	ϵ_{σ^2}	0.001676	0.03530	0.0542

Once the above two measures established that the generated distributions were indeed Gaussian, both qualitatively and quantitatively, error functions were defined to evaluate the deviation of the mean and the variance from the expected values (0 and 1 respectively, in the case of a standard normal distribution). These measures indicate that the parameters were well within an acceptable error margin. The mean absolute error, which is probably the most popular error estimate, was within 1% for the mean and variance, while the mean absolute error was around 3% for all the four techniques. Finally, the maximum absolute error, which gives an estimate about the worst-case deviation, suggests that the errors are of the order of 10% in mean and variance for the four techniques compared to a standard normal distribution. It is worth noting, however, that since mean square error and mean average error are averaging metrics, they tend to balance out variations in a distribution provided they occur sufficiently infrequently. So, depending on the degree of precision required, the maximum absolute error may be a better measure of the *quality* of the generated random variables. It may also be pointed out that, once again, the Box–Muller methods outperform the other techniques for generating Gaussian random variables.

In conclusion, there is no single performance metric that encapsulates information about the gaussianity of the distribution *as well as* the closeness to the standard normal distribution. However, it has been established that it would be foolhardy to design a metric that is based *solely* on the mean and variance of the distribution. A good metric could combine information from the kurtosis with an estimation of the error provided by the maximum absolute error, to assess the quality of a generated Gaussian random variable. Based on this metric, the Box–Muller method for generating Gaussian random variables seems to provide the best performance.

A Appendix

The appendix lists the source codes used for simulating the four algorithms discussed in this report. The programs were written in MATLAB version 5.2.0.3084, and were executed in a *Microsoft Windows 98SE* environment.

A.1 Inverse Distribution

```
% This generates normal random variables
% and calculates their mean and variance
% using inverse CDF of uniform RVs
for i=1:1000
z=rand(1);
j=-3;
while normcdf(j)<z
j=j+0.01;
end
rv(i,1)=j-0.05;
end
x=-4.0:0.01:4.0;
hist(rv,x)
avg=mean(rv)
var=(std(rv))*(std(rv))
kurt=kurtosis(rv)
```

A.2 Box–Muller Method

```
% This plots and calculates the mean and
% variance for the distribution generated
% by the Box-Muller method (proj0102.m)
for i=1:1000
u1=rand(1);
u2=rand(1);
```

```
x1=sqrt(-2*log(u1))*cos(2*pi*u2);
x2=sqrt(-2*log(u2))*cos(2*pi*u1);
rv1(i,1)=x1;
rv2(i,1)=x2;
end
rv=[rv1;rv2];
x=-4.0:0.1:4.0;
hist(rv,x)
avg=mean(rv)
var=(std(rv))*(std(rv))
kurt=kurtosis(rv)
```

A.3 Polar Technique

```
% This plots and calculates the mean and
% variance for the distribution generated
% by the polar technique (proj0103.m)
for i=1:1000
w=2;
while w>1
u1=rand(1);
v1=2*u1-1;
u2=rand(1);
v2=2*u2-1;
w=v18v1+v2*v2;
end
y=sqrt((-2*log(w))/w);
x1=v1*y;
x2=v2*y;
rv1(i,1)=x1;
rv2(i,1)=x2;
end
rv=[rv1;rv2];
x=-4.0:0.1:4.0;
hist(rv,x)
avg=mean(rv)
var=(std(rv))*(std(rv))
```

```
kurt=kurtosis(rv)
```

A.4 Sums of Uniform Random Variables

```
% This generates normal random variables
% and calculates their mean and variance
% using sums of uniform random variables
for i=1:1000
for j=1:1000
x(j,1)=rand(1);
end
y=sum(x(:,1));
z=(y-(j/2))/sqrt(j/12);
rv(i,1)=z;
end
x=-4.0:0.1:4.0;
hist(rv,x)
avg=mean(rv)
var=(std(rv))*(std(rv))
kurt=kurtosis(rv)
```

References

[1] Alberto Leon-Garcia. *Probability and Random Processes for Electrical Engineering*. Addison-Wesley, Reading, MA, 1994.

[2] Wikipedia-The Free Encyclopedia. http://www.wikipedia.org.

[3] Eric W. Weisstein. http://www.treasure-troves.com.

[4] Roy D. Yates and David J. Goodman. *Probability and Stochastic Processes*. John Wiley, New York, NY, 1999.

[5] Aapo Hyvarinen. http://www.cis.hut.fi/aapo/papers.

[6] NIST Sematech. e-Handbook of Statistical Mehods. Technical report, http://www.itl.nist.gov/div898/handbook, 1998.