CAMBRIDGE UNIVERSITY ENGINEERING DEPARTMENT

Part IIA Full Technical Report

3F3 Random Number Generation

Name: Yongqing Jiang

CRSid: yj375

College: Peterhouse

Date of Experiment: Nov. 2025

1 Introduction

This lab activity investigates statistical methods by generating random numbers. Uniform and normal random variables are generated, and they are visualized by histogram and kernel smoothing density (KSD) function. Functions of random variables are also discussed. Inverse CDF method is used to generate random variables from any arbitrary distributions. At last, stable distribution is discussed.

2 Methods, Results, and Discussion

2.1 Uniform and normal random variables

In this section, uniform and normal random variables are generated and visualized by using histogram and kernel density function. The results are compared and discussed.

2.1.1 Histogram and kernel density function

In Figure 1, 1000 random Gaussian numbers and 1000 random uniform numbers are generated. The histograms are generated with 50 bins and the KSD functions has a width of 0.4 with a

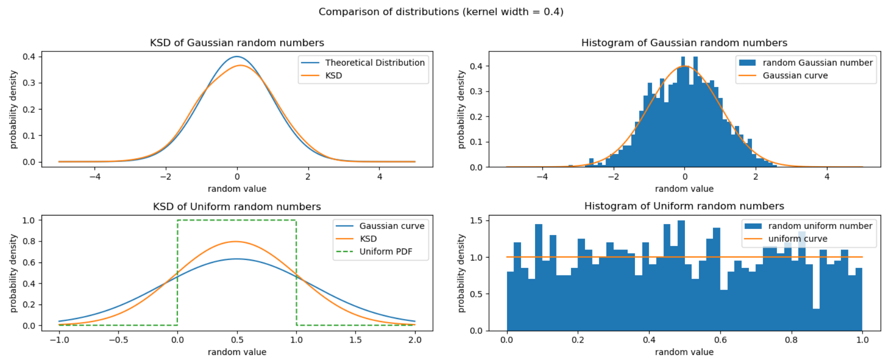

Figure 1: Histogram and KSD of uniform and normal random variables

Gaussian kernel, in the form of[1]:

$$\pi_{KS}(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma}\mathcal{K}\left(\frac{x-x^{(i)}}{\sigma}\right)$$

Where $\mathcal{K}(\cdot) \sim \mathcal{N}(\cdot|0,1)$, σ is the width of the kernel.
We can see that histogram can show more detail of the sample set. However, bin size is a key parameter to determine the underlying distribution. In this illustration, the Gaussian random numbers show a histogram that is really close to a bell-shape, while the uniform numbers do not show a clear flat shape.
On the other hand, KSD smooths out the histogram by averaging kernel functions. The Gaussian KSD is very close to the exact curve. However, the uniform KSD is way off the expected shape. This is due to the choice of kernel width.
To discuss more about the effect of kernel width, several KSDs with distinct kernel widths are plotted.

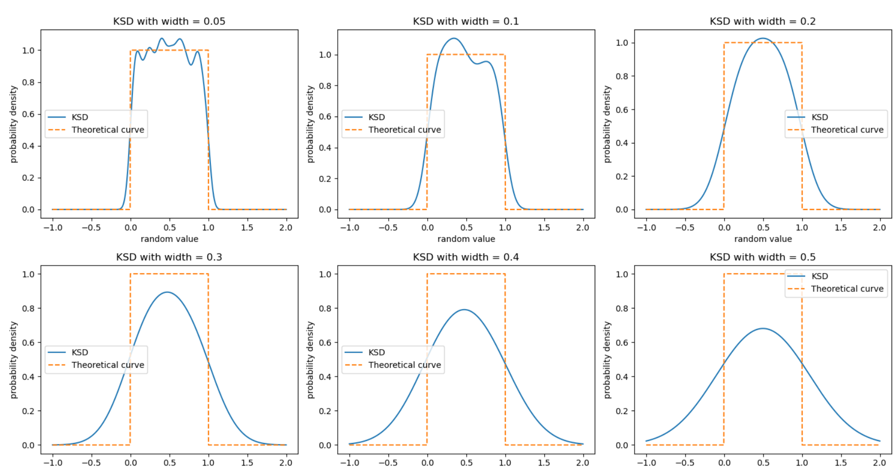

Figure 2: KSD of uniform random variables with different kernel widths

From Figure 2, we can see that with a small kernel width (0.05 and 0.1), the KSD captures more details of the uniform distribution. It is reasonable to state that this is a uniform distribution only by looking at the KSD. However, with a larger kernel width that is comparable to the range of the uniform distribution , the KSD smooths out the details and deviates from the ideal uniform shape. To conclude, there is a trade-off when choosing the kernel width between smoothing and shape deviation.

2.1.2 Multinomial distribution

The Multinomial distribution is a generalization of the binomial distribution [2], and it can be used to discribe a distribution within a histogram.
Suppose there are some fixed finite number of bins, J, and there are N samples to be placed into these bins. Each bin has a probability p_j of receiving a sample, where $j = 1, 2, \ldots, J$ and $\sum_{j=1}^{J} p_j = 1$.
p_j can be calculated by integrating the underlying probability density function $p(x)$.

$$p_j = \int_{c_j-\delta/2}^{c_j+\delta/2} p(x)dx$$

Where c_j is the center of bin j, and δ is the width of each bin.
Let random variable $\vec{X} = (X_1, X_2, \ldots, X_J)$ represents the number of samples in each bin.
Then the PMF of the Multinomial distribution is given by:

$$P(\vec{X} = \vec{n}) = \frac{N!}{n_1!n_2!\ldots n_J!} p_1^{n_1} p_2^{n_2} \ldots p_J^{n_J}$$

Where $\vec{n} = (n_1, n_2, \ldots, n_J)$. When $\dim \vec{X}$ is 2, it reduced to a binomial distribution.
The term $p_j^{n_j}$ represents the probability of n_j samples falling into bin j, as each sample has a probability p_j of falling into that bin.
The factorial terms account for the different arrangements of samples across the bins, considering that the order of samples within each bin does not matter.
Expectation of X_j is given by:

$$\mathbb{E}[X_j] = Np_j$$

And Variance of X_j is given by:

$$\mathrm{Var}(X_j) = Np_j(1 - p_j)$$

For uniform distributed random variables, $p(x) = \frac{1}{N}$, so $p_j = \frac{\delta}{N}$.
$\mathbb{E}(X_j) = \frac{N\delta}{N} = \delta$, $\mathrm{Var}(X_j) = N\frac{\delta}{N}(1 - \frac{\delta}{N}) = \delta(1 - \frac{\delta}{N})$.
Three sets of uniform random variables are generated with N = 100, 1000, 10000 respectively.
The histograms are plotted with 30 bins, and the results are shown in Figure 3.
We can see that the normalized mean of each bin (height of each bar) is fixed by the bin width, which is $\frac{1}{30} \approx 0.0333$.

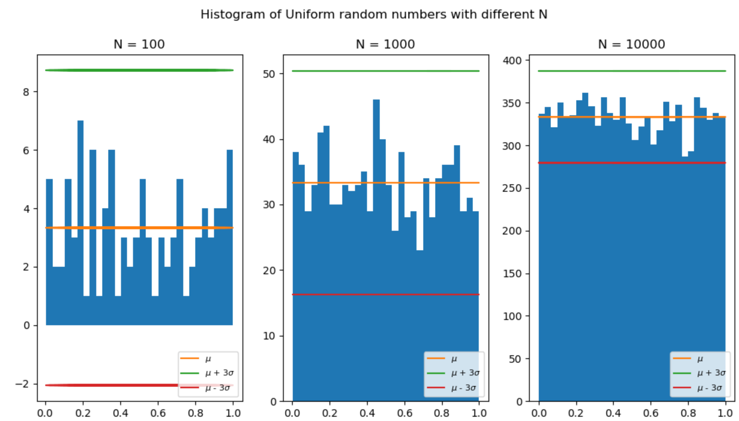

Figure 3: Histograms of uniform random variables with different sample sizes

For N = 1000 and 10000, the histogram bars fluctuate around this value and fits in the interval $[\mu - 3\sigma, \mu + 3\sigma]$, as expected.
However, for N = 100, the histogram bars show significant deviation from the expected range. This is because with a small sample size, the variance is relatively large compared to the mean. It is implied that a small sample size may not accurately represent the underlying distribution by using a histogram.
This analysis can also be applied to Gaussian random variables.
To start with,

$$p_j = \int_{c_j - \delta/2}^{c_j + \delta/2} p(x)dx = F(c_j + \delta/2) - F(c_j - \delta/2)$$

Where $F(x)$ is the CDF of the Gaussian distribution. This expression helps to evaluate p_j in *python* using *scipy.stats.norm.cdf* function.
$\mathbb{E}(X_j) = Np_j = N(F(c_j + \delta/2) - F(c_j - \delta/2))$
$\mathrm{Var}(X_j) = Np_j(1 - p_j) = N(F(c_j + \delta/2) - F(c_j - \delta/2))(1 - F(c_j + \delta/2) - F(c_j - \delta/2))$.

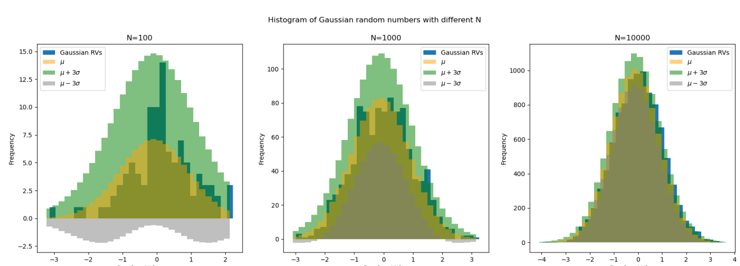

Figure 4: Histograms of Gaussian random variables with different sample sizes

As shown in Figure 4, the histograms of Gaussian random variables are plotted with corresponding expected mean and variance for each bin height. Similar to the uniform distribution case, it is expected that the bar heights lie within the interval $[\mu - 3\sigma, \mu + 3\sigma]$.

However, in this case, it can be observed that σ is no longer a constant throughout the range of random variables, since p_j varies for different bins. $\mathrm{Var}(X_j) = Np_j(1-p_j)$. The variance is larger at the center of the distribution where p_j is larger, and smaller at the tails where p_j approaches to 0 and 1.

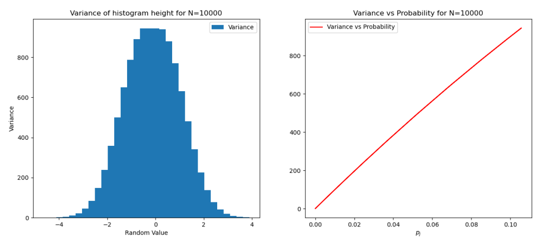

Figure 5: Comparing variance of histogram height and p_j

This argument is visualized in Figure 5, where the variance of histogram height and p_j are plotted together. It is clear that the variance is small when p_j is small and at the tails.
Due to the limited bin size, $\max(p_j)$ is around 0.1. In this range, $\mathrm{Var}(X_j) \approx Np_j$, which agrees with the linear trend shown.

3 Function of Random Variables

The Jacobian formula for change of variables in probability density functions states that if $y = f(x)$ is a differentiable and invertible function, in which x and y are random variables with PDF $p(x)$ and $p(y)$:

$$p(y) = \left.\frac{p(x)}{|dy/dx|}\right|_{x=f^{-1}(y)}$$

For a linear transformation, $y = f(x) = ax + b \Rightarrow x = f^{-1}(y) = \frac{y-b}{a}$, and $dy/dx = a$. With $f(\cdot)$ being the standard Gaussian distribution:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$p(y) = \frac{1}{|a|} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(\frac{y-b}{a})^2}{2}} = \frac{1}{\sqrt{2\pi a^2}} e^{-\frac{(y-b)^2}{2a^2}}$$

We can see that the transformed random variable is $\mathcal{N}(y|b, a^2)$, the linear transformation of a normal distribution is still a normal distribution, with a shift of mean from μ to $\mu + b$ and a scaling of variance from σ^2 to $a^2\sigma^2$.
Now, consider a non-linear transformation, $y = f(x) = x^2$.

$$x = f^{-1}(y) = \begin{cases} \sqrt{y}, & x \geq 0 \\ -\sqrt{y}, & x \leq 0 \end{cases}$$

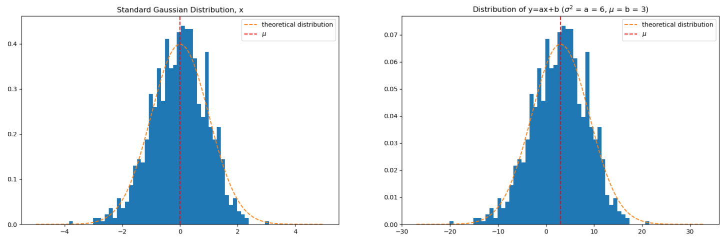

Figure 6: Histogram of linearly transformed Gaussian random variables and corresponding PDF

We can see that $f^{-1}(\cdot)$ is one-to-many, so we need to consider both branches of the inverse function. Also, $|dy/dx| = 2|x| = 2\sqrt{y}$.

$$
\begin{aligned}
p(y) &= \frac{p(x)}{|dy/dx|}\bigg|_{x=f^{-1}(y)} + \frac{p(x)}{|dy/dx|}\bigg|_{x=-f^{-1}(y)} \\
&= \frac{1}{2\sqrt{y}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} + \frac{1}{2\sqrt{y}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{y})^2}{2}} \\
&= \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}
\end{aligned}
$$

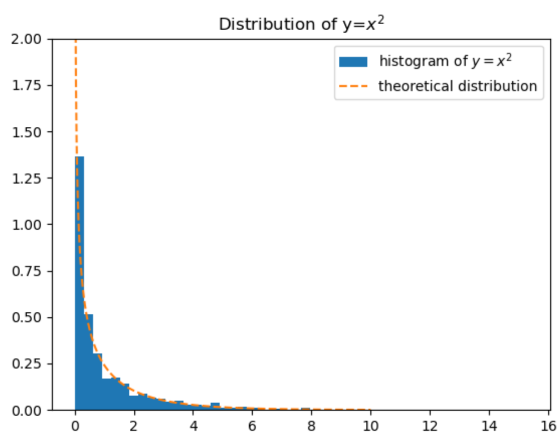

Figure 7: Histogram of squared Gaussian random variables and corresponding PDF

As shown by Figure 7, the theory agrees well to the histogram of squared Gaussian random variables.
Lastly, consider $p(x) = \mathcal{U}(x|0, 2\pi)$, and transformation $y = f(x) = \sin(x)$. The inverse of $f(\cdot)$ is also not one-to-one within the range of x.
There are four branches of the inverse function within $[0, 2\pi]$, and it is better to defind $\mathrm{Arcsin}(\cdot)$

function that maps $[0,1] \rightarrow [0, \frac{\pi}{2}]$.

$$x = f^{-1}(y) = \begin{cases} \text{Arcsin}(y), & 0 \le x \le \frac{\pi}{2}, 0 \le y \le 1 \\ \pi - \text{Arcsin}(y), & \frac{\pi}{2} \le x \le \pi, 0 \le y \le 1 \\ \pi + \text{Arcsin}(-y), & \pi \le x \le \frac{3\pi}{2}, -1 \le y \le 0 \\ 2\pi - \text{Arcsin}(-y), & \frac{3\pi}{2} \le x \le 2\pi, -1 \le y \le 0 \end{cases}$$

Also, $|dy/dx| = |\cos(x)|$.
Now, the transformed PDF can be calculated as:

$$p(y) = \sum_{i=1}^{2} \frac{p(x)}{|dy/dx|}\Bigg|_{x=f_i^{-1}(y)} = \sum_{i=1}^{2} \frac{1}{2\pi|\cos(x)|}\Bigg|_{x=f_i^{-1}(y)}, \quad 0 \le y \le 1$$

$|\cos(x)| = \sqrt{1-y^2}$ for all four branches of $f^{-1}(\cdot)$, so

$$p(y) = \frac{2}{2\pi\sqrt{1-y^2}} = \frac{1}{\pi\sqrt{1-y^2}}, \quad 0 \le y \le 1$$

Similarly, $p(y) = 1/(\pi\sqrt{1-y^2})$, for $-1 \le y \le 0$. Combining both parts, we have:

$$p(y) = \frac{1}{\pi\sqrt{1-y^2}}, \quad -1 \le y \le 1$$

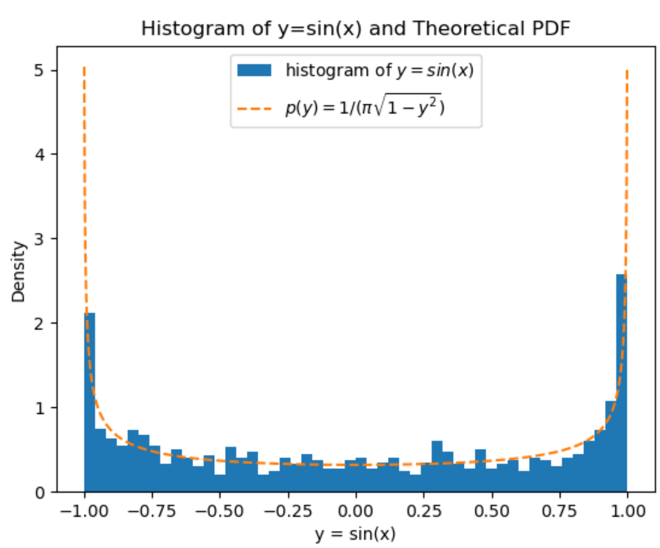

Figure 8: Histogram of $y = \sin(x)$ and corresponding PDF

As shown by Figure 8, the histogram of $y = \sin(x)$ agrees well with the derived PDF.

Now consider a clipping function:

$$y = f(x) = \min(\sin x, 0.7) = \begin{cases} \sin(x), & 0 < x < \arcsin(0.7), \pi - \arcsin(0.7) < x < 2\pi \\ 0.7, & \arcsin(0.7) < x < \pi - \arcsin(0.7) \end{cases}$$

The derivative can be evaluated piecewisely:

$$\left|\frac{dy}{dx}\right| = \begin{cases} |\cos(x)|, & -1 < y < 0.7 \\ 0, & 0.7 < y < 1 \end{cases}$$

We can see that, since the derivative is at the denominator, $p(y)$ doesn't exist in the interval $0.7 < y < 1$.
However, $x \in (0.7, 1)$ is all mapped to $y = 0.7$. Now $p(y)$ is expected to be a combination of a continuous PDF that is derived before for $y \in [-1, 0.7]$, and a finite spike at $y = 0.7$.
The magnitude of the spike is expected to be $\int_{0.7}^{1} \frac{1}{\pi\sqrt{1-y^2}} dy \approx 0.184$.

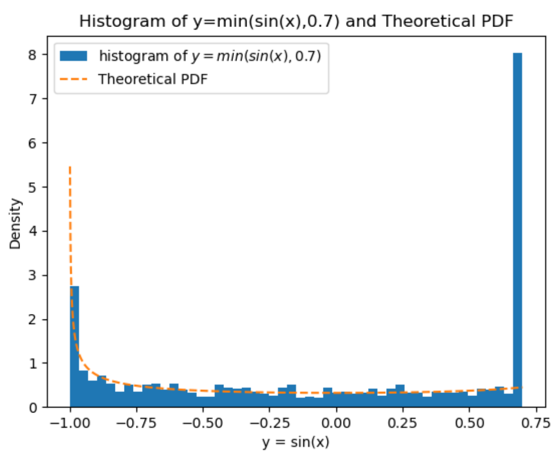

Figure 9: Histogram of $y = \min(\sin(x), 0.7)$ and corresponding PDF

4 Inverse CDF Method

To generate random variables from an arbitrary distribution $p(y)$, the inverse CDF method is used.
Take a uniform distribution $p(x) = \mathcal{U}(x|0, 1)$, and transform it by $y = f(x)$. By the Jacobian formula,

$$p(y) = \begin{cases} \left.\dfrac{1}{|dy/dx|}\right|_{x=f^{-1}(y)}, & f^{-1}(y) \in (0, 1) \\ 0, & \text{otherwise} \end{cases}$$

$1/|dy/dx| = |df^{-1}/dx|$. Thus, the CDF of y, $F(y) = \int p(y)dy = f^{-1}(y)$. Equivalently, $f(y) = F^{-1}(y)$.
This method is a general way to generate random numbers that has a CDF $F(\cdot)$. A set of uniform random numbers is transferred by $F^{-1}(\cdot)$ to give the required random number set.
This method is verified by generating exponential random variables.
The PDF of an exponential distribution Y with mean 1 is:

$$p(y) = e^{-y}$$

The corresponding cdf is found by integration:

$$F(y) = \int_0^y p(t)dt = \int_0^y e^{-t}dt = 1 - e^{-y}$$

The inverse of this function is: $F^{-1}(x) = -\ln(1-x)$. Random numbers $y^{(i)}$ can be generated by $y^{(i)} = F^{-1}(x^{(i)})$, where $x^{(i)}$ are uniform random numbers between 0 and 1.

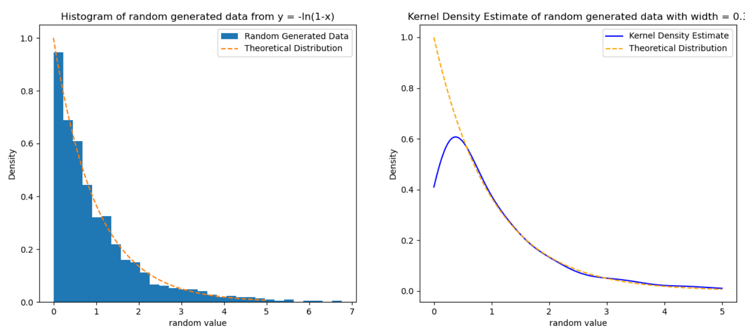

Figure 10: Histogram of exponential random variables generated by inverse CDF method and corresponding PDF

As shown in Figure 10, the histogram of generated exponential random variables fits well with the theoretical PDF, as well as the KSD, which verifies the correctness of the inverse CDF method for large random values.
A Monte Carlo simulation is also performed to estimate the mean and variance of the exponential distribution.

$$\mu = \mathbb{E}[Y] \approx \frac{1}{N}\sum_{i=1}^{N} y^{(i)} = \hat{\mu}$$

$$\sigma^2 = \mathrm{Var}(Y) \approx \frac{1}{N}\sum_{i=1}^{N} (y^{(i)})^2 - (\hat{\mu})^2 = \hat{\sigma}^2$$

$\mathbb{E}[\hat{\mu}] = \mathbb{E}[\frac{1}{N}\sum_{i=1}^{N} y^{(i)}] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[y^{(i)}]$ by linearality of expectation. $\mathbb{E}[y^{(i)}] = \mu$ by definition, so $\mathbb{E}[\hat{\mu}] = \mu$. Which implies that the Monte Carlo estimator of mean is unbiased.

The variance of the estimator of mean can be evaluated from the definition of variance: $\mathrm{Var}(\hat{\mu}) = \mathbb{E}[\hat{\mu}^2] - (\mathbb{E}[\hat{\mu}])^2 = \mathbb{E}[\hat{\mu}^2] - \mu^2$.

$$\hat{\mu}^2 = \frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} y^{(i)}y^{(j)}\right)$$

The expecation of $\hat{\mu}^2$ can be evaluated as:

$$\begin{aligned}
\mathbb{E}[\hat{\mu}^2] &= \frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{E}[y^{(i)}y^{(j)}]\right) \\
&= \frac{1}{N^2}\left(\sum_{i=1}^{N} \mathbb{E}[{y^{(i)}}^2] + 2\sum_{1\leq i\leq j\leq N}^{N} \mathbb{E}[y^{(i)}y^{(j)}]\right) \\
&= \frac{1}{N^2}\left(\sum_{i=1}^{N} \mathbb{E}[{y^{(i)}}^2] + 2\sum_{1\leq i\leq j\leq N}^{N} \mathbb{E}[y^{(i)}]\mathbb{E}[y^{(j)}]\right)
\end{aligned}$$

Linearity of expectation and independence of samples are used in the last step.
$\mathbb{E}[{y^{(i)}}^2] = \mathrm{Var}(Y) + (\mathbb{E}[Y])^2 = \sigma^2 + \mu^2$ and $\mathbb{E}[y^{(i)}]\mathbb{E}[y^{(j)}] = \mu^2$ by definition. Thus, $\mathbb{E}[\hat{\mu}^2]$ can be simplified as:

$$\begin{aligned}
\mathbb{E}[\hat{\mu}^2] &= \frac{1}{N^2}\left(N(\sigma^2+\mu^2) + 2\frac{N(N-1)}{2}\mu^2\right) \\
&= \frac{\sigma^2}{N} + \mu^2
\end{aligned}$$

$\mathrm{Var}(\hat{\mu}) = \mathbb{E}[\hat{\mu}^2] - \mu^2 = \frac{\sigma^2}{N}$.

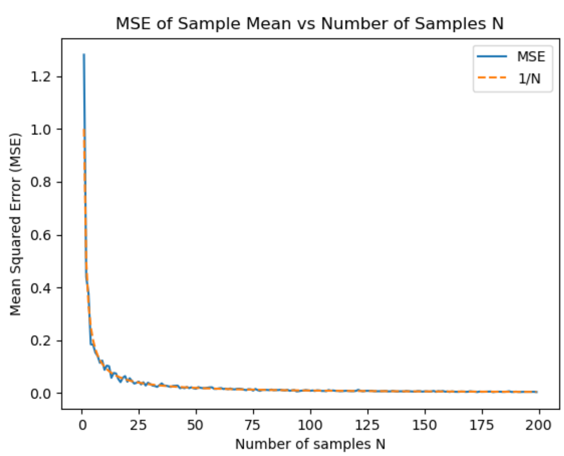

Figure 11: MSE against sample size N

A final simulation is performed to verify the variance of the estimator of mean. As shown in Figure 11, the MSE of the estimator of mean decreases with increasing sample size N. For each N, 100 tests are performed to evaluate the MSE by Monte Carlo simulation.

5 Stable Distribution

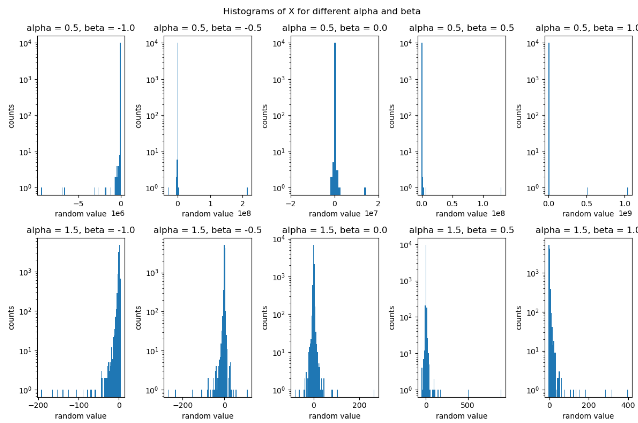

Figure 12: Histogram of α-stable random variables with different α and β

From Figure 12, we can see that α determines the extent of the data with extreme values. A small value of α gives a distribution that has more extreme values, while a large value of α gives a distribution that is more concentrated around the mean. When $\alpha = 2$, we can see that the distribution has become a standard Gaussian distribution.

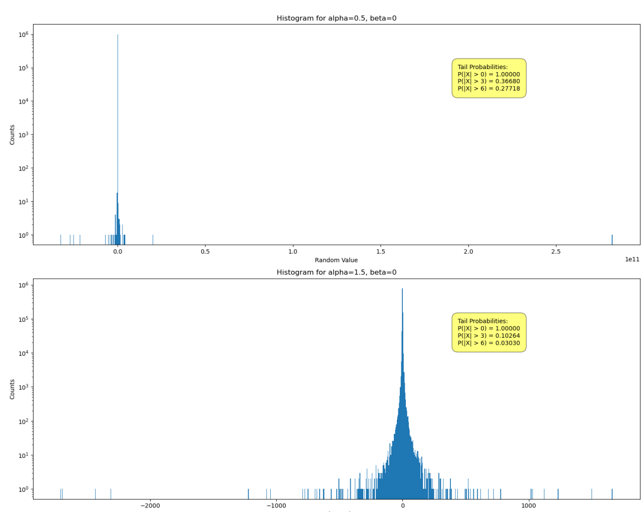

Figure 13: Tail Probability of stable random variables with different α

The value of β determines the skewness of the distribution. A positive β skews the distribution to the right, and a negative β skews the distribution to the left.
For the case that $\beta = 0$, the tail probability of α-stable random variables is plotted in Figure 13. To compare with the standard Gaussian distribution, $\mathcal{N}(0, 1)$, that has a tail probability

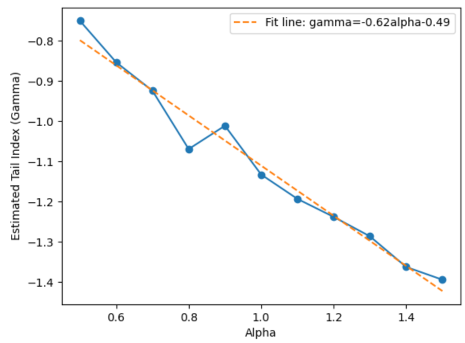

Figure 14: γ against α

of:

$$P(|X| > 0) = 1$$
$$P(|X| > 3) \approx 2.7 \times 10^{-3}$$
$$P(|X| > 6) \approx 9.87 \times 10^{-10}$$

It is clear that stable distributions generally have heavier tails than the Gaussian distribution for $\beta = 0$.
To discuss more about the tail behavior, the tail probabilities are evaluated for different α values with $\beta = 0$.
The tail probabilities are defined as $P(|X| > t)$, where $t = 3, 6$, defined by value of α. Threshold t is chosen to be 3 for small α values, and 6 for large α values. 100000 samples are generated for each α value to evaluate the tail probabilities. And the generation is repeated for 50 times for each value of α to get an average value of tail probability $p(x) = cx^{\gamma}$. For each set of samples, the tail probabilities are linearized by taking the logarithm on both sides: $\ln(p(x)) = \ln(c) + \gamma \ln(x)$. Then, linear regression is performed to estimate the value of γ.
As shown by Figure 14, there is a linear relationship between γ and α. By a linear fit, we have:

$$\gamma \approx -0.61\alpha - 0.49$$

At last, it is worth exploring the behavior of the stable distribution as α approaches to 2. As shown by Figure 15, when α is very close to 2, the stable distribution shows little difference among each value of α. However, when $\alpha = 2$, the distribution becomes a Gaussian distribution $\mathcal{N}(0, 2)$, which is significantly different from the stable distributions with α close to 2.

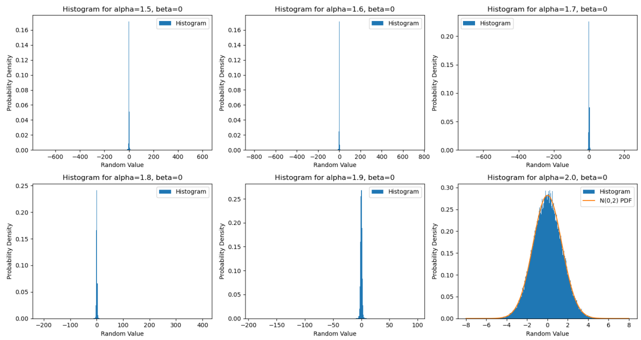

Figure 15: Histogram of stable distribution for different values of α close to 2

6 Conclusion

This report links theoretical statistical concepts ,including random number generation and variable transformation, to empirical validation via simulation and visualization. We confirmed that distribution visualization depends on parameter tuning and sample size, validated the Jacobian formula for variable transformations and the inverse CDF method for arbitrary distribution generation, and clarified how stable distribution parameters control tail heaviness and skewness. These techniques form a foundational toolkit for engineering applications involving randomness, from signal processing to stochastic modeling.

References

[1] Cambridge University Engineering Department. *Random Variables and Random Number Generation Lab Sheet*. 2025.

[2] S Sinharay. *Discrete Probability Distributions*. ETS, Princeton, NJ, USA, 2010

Appendix: Code Listings

The code used to generate the results in this report is provided below: `https://github.com/OliverJiang2025/3F3_lab`