Random Variables and Random Number Generation

Full Technical Report

Shanzi (Monica) Ran
Magdalene College
sr2021@cam.ac.uk

December 7, 2024

1 Introduction

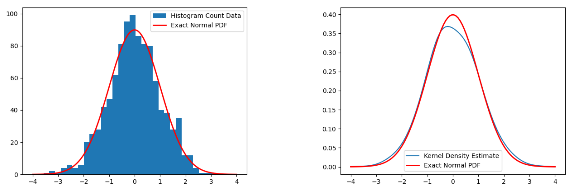
Random vairables(RV) is a key concept in statistical signal and information processing, playing a significant role in various engineering applications, It is therefore important to develop efficient and reliable methods generating and presenting desired non-uniform RV distributions. In this report, investigations and observations of the following objectives are discussed and summarized:

- Generation and visualization of RV using histograms and kernel density estimation(KDE)
- Transformation of RV distributions using Jacobian formula
- Generation of non-uniform RV distributions using cumulative distribution function(CDF) and inverse CDF methods
- Investigation of complex RV distributions and α-stable distribution

2 Results and Discussion

2.1 Uniform and normal RV

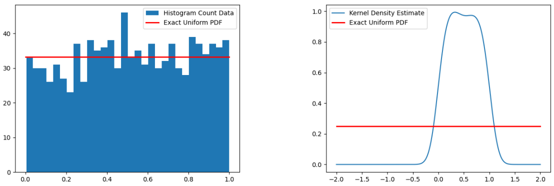
2.1.1 Histogram and KDE



(a) Histogram (b) KDE

Figure 1: *Histogram and KDE plot of 1000 sampled Gaussian RV* $X \sim \mathcal{N}(0,1)$

As shown in Figure 1, both histogram and KDE (with the $\mathcal{N}(0,1)$ Gaussian kernel) produces a close approximate to the density distribution of the normal RV. When comparing between Fig 1(a) and Fig 1(b), it is observed that KDE results in a more accurate approximation, which results from the smoothing of KDE, as discontinuity in histogram resulting from its discrete feature is removed.



(a) Histogram (b) KDE (Overlaid on standard Normal)

Figure 2: *Histogram and KDE plot of 1000 sampled Uniform RV* $X \sim \mathcal{U}(0,1)$

In comparison, the histogram and KDE approximates for a uniform RV in Figure 2 demonstrates the opposite as the histogram describes more accurately the density function of a uniform distribution because of its discontinuous property, which is less correctly depicted by KDE since the KDE smoothing removes the step changes, making the uniform distribution in Fig 2(b) appear to almost coincide a normal distribution. This is also likely related to the specific choice of kernel , as a uniform kernel $\mathcal{U}(0,1)$ may suit the uniform distribution better.

2.1.2 Multinomial distribution

When N samples of a RV with density $p(x)$ is drawn, histogramming the samples is useful in estimating the density. For histogram bin j with edges $[a,b]$, the probability of a sample $x^{(i)}$ falling in this bin is:

$$p_j = \int_a^b p(x)dx$$

Since each of the N samples is independent and identical distributed (iid), the joint probability of a certain outcome $(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$, where X_j describes the number of samples falling in bin j, is the product of each sample falling into the corresponding bins. Therefore, for a specific outcome of bin counts for a histogram with n bins in total (when order does not matter):

$$\prod_{j=1}^{n} {p_j}^{x_j} \tag{1}$$

Since order does not matter, the total number of ways of reaching this outcome is given by combinactorics:

$$\binom{N}{x_1,\ldots,x_k} = \frac{N!}{x_1!x_2!\ldots x_n!} \tag{2}$$

and the $\pm 3\sigma$ lines are growing closer to the mean, which agrees with theoretical calculations as mean increases at N and standard deviation increases slower at a rate of $\sqrt{N}$ with respect to the sample size. Thus, the Python generated uniform distribution is accuarate, with its reliability improving for larger sample sizes which is consistent with the multinomial distribution theory.

So probability of this distribution, equivalently the multinomial distribution probability mass function (pmf), equals *equation* $1 \times$ *equation* 2:

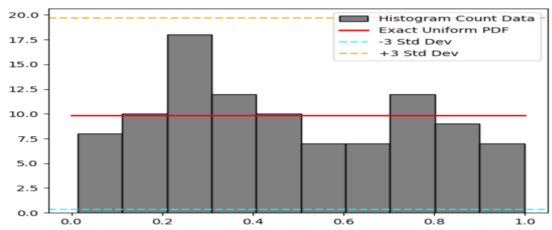
$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{N!}{x_1! \ldots x_n!} {p_1}^{x_1} \ldots {p_n}^{x_n}$$

Uniform RV with varied sample size For a uniform distribution, the theoretical histogram count data of each bin is identical, which can be calculated using the following equations:

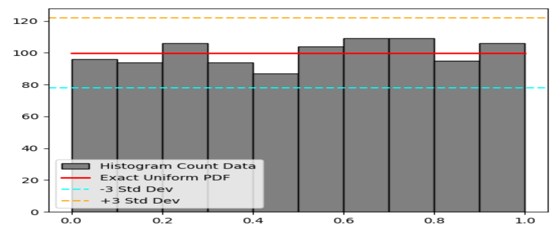
$$\mu = \frac{N}{j}$$

$$\sigma = \sqrt{\frac{N}{j}(1 - \frac{1}{j})}$$

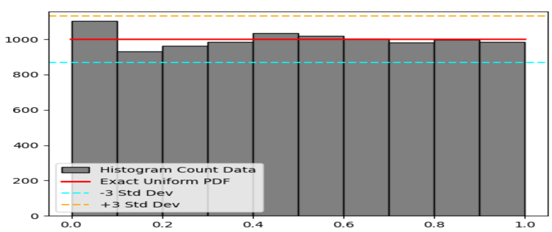
where N is total number of samples, j is total number of bins.



(a) N=100



(b) N=1000



(c) N=10000

Figure 3: *Histograms for different values of uniform RV N with theoretical μ and 3σ lines*

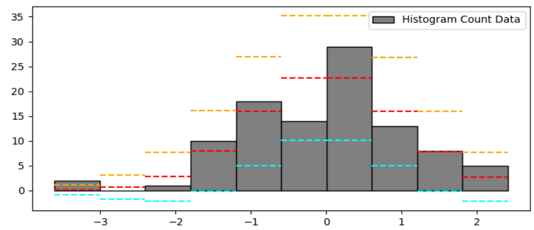
As shown in Fig 3, when sample size N enlarges, the histogram count approaches to the straight line mean,

Normal RV with varied sample size Similarly, theoretical histogram count mean and standard deviation for normal RV is calculated as below according to the multinomial distribution:

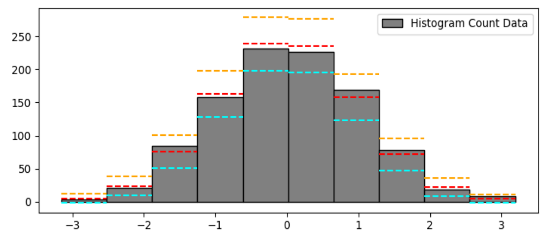
$$\mu = N p_j$$

$$\sigma = \sqrt{N p_j (1 - p_j)}$$

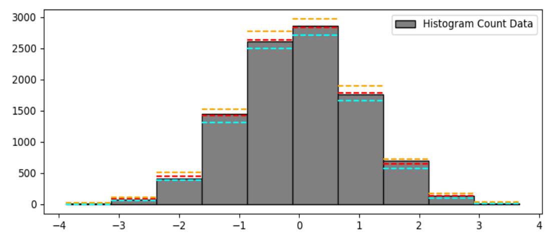
It should be noted that this μ and σ differs for each of the 10 bins, with $p_j = \phi(b) - \phi(a)$ for bin j with edges $[a, b]$.



(a) N=100



(b) N=1000



(c) N=10000

Figure 4: *Histograms for different values of normal RV N with theoretical μ and 3σ lines*

According to Fig 7, a similar trend of mean and $\pm 3\sigma$ lines approaching to be closer is observed as N increases. Meanwhile, it is observed that when approaching the tail of the distribution where $p_j \approx 0$, since $\sigma^2 \approx Np_j = \mu$, $\sigma \approx \sqrt{\mu}$, resulting in -3σ lines more likely to be negative as $\mu - 3\sqrt{\mu} < 0$.

When p_j is closer to 1, since $\sum_{j=1}^{n} p_j = 1$ as probability theory requires, there is only 1 bin left on the histogram, variance is hence 0 regardless of value of N.

2.2 Functions and transformations of RV

2.2.1 Linear transformation $y = ax + b$

Using the Jacobian method, density function of y can be transformed from $X ~ \mathcal{N}(0,1)$ by function $f(x) = ax + b$ as follows:

$$y = f(x) = ax + b, \text{ so } x = f^{-1}(y) = \frac{y-b}{a}$$
$$\frac{dy}{dx} = a$$
$$p(y) = \frac{p(x)}{\left|\frac{dy}{dx}\right|}\Big|_{x=f^{-1}(y)}$$
$$= \frac{1}{a} p\left(\frac{y-b}{a}\right)$$
$$= \frac{1}{\sqrt{2\pi}a} exp(-\frac{1}{2}(\frac{y-b}{a})^2)$$

which indicates that $p(y)$ is a normal density function with $\mu = b, \quad \sigma = a$. The above results is also supported by Fig 5 where the theoretical calculated density function of y is overlaid upon the sampled results with sample size 10000.

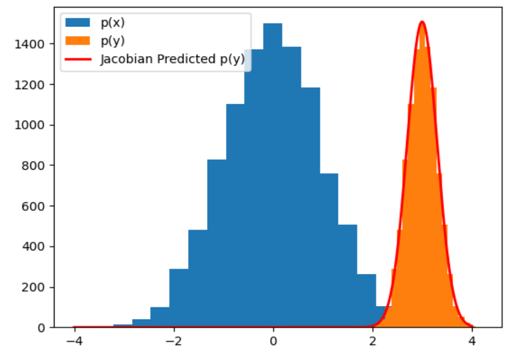


Figure 5: *Sampled transformation for y=ax+b*

2.2.2 Quadratic transformation $y = x^2$

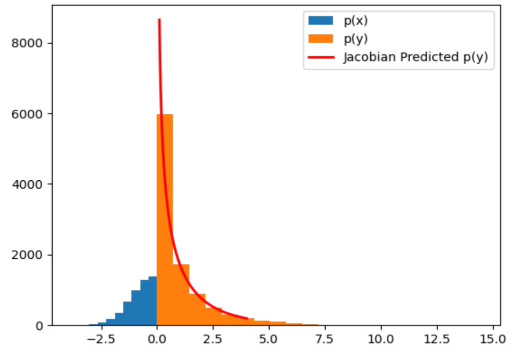


Figure 6: *Sampled transformation for $y = x^2$*

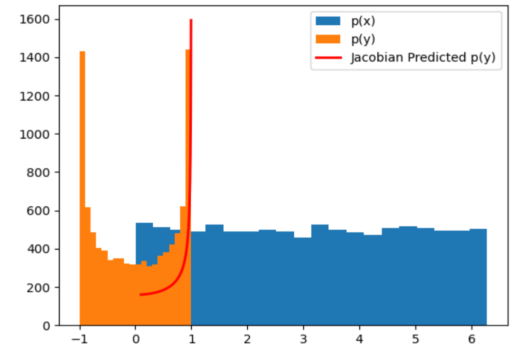
When $y = f(x) = x^2$, the Jacobian method involves summing all possible values of $f^{-1}(y)$ since this inverse function is a one-many mapping. Fig 6 depicts the result gathered over 10000 samples, which aligns with the calculated tranformations.

$$y = f(x) = x^2, \text{ so } x = f^{-1}(y) = \pm\sqrt{y}\text{ , where } x_1 = \sqrt{y}, x_2 = -\sqrt{y}$$
$$\frac{dy}{dx} = 2x$$
$$p(y) = \sum_{k=1}^{2} \frac{p(x)}{|\frac{dy}{dx}|}\Big|_{x=x_k(y)}$$
$$= \frac{1}{\sqrt{2\pi}} exp(-\frac{y}{2})(\frac{1}{2\sqrt{y}} + |\frac{1}{2\sqrt{y}}|)$$
$$= \frac{1}{\sqrt{2\pi y}} exp(-\frac{y}{2})$$

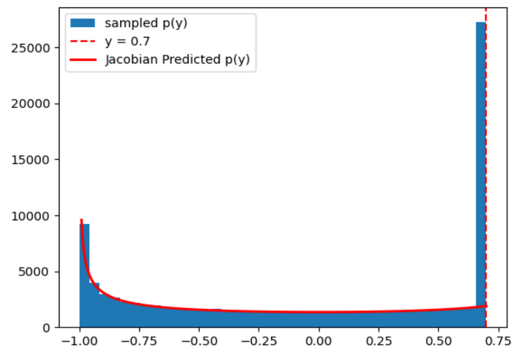
2.2.3 Trignometric transformation $y = sin(x)$

For a transformation from $X ~ \mathcal{U}(0, 2\pi)$ using $f(x) = sin(x)$:

$$y = f(x) = sin(x), \text{ so } x = f^{-1}(y) = sin^{-1}(y)$$
$$\frac{dy}{dx} = cos(y)$$
$$p(y) = \frac{p(x)}{|\frac{dy}{dx}|}\Big|_{x=f^{-1}(y)}$$
$$= \frac{1}{2\pi|cos(sin^{-1}(y))|}$$



(a) Sampled density for $y = sin(x)$



(b) Sampled density for $f(x) = min(sin(x), 0.7)$ (zoomed in)

Figure 7: *Sampled transformation for* $f(x) = sin(x)$ *and* $f(x) = min(sin(x), 0.7)$

where the absolute value is required due to the periodic feature of trigonometric transformations. This response agrees with Fig 7(a).

For a value-limited sine transformation $f(x) = min(sin(x), 0.7)$, the density function is more complicated due to this "clipping":

$$y = f(x) = \min(\sin(x), 0.7)$$

$$x = f^{-1}(y) = \begin{cases} \sin^{-1}(y) & \text{if } -1 \leq y < 0.7, \\ \delta(y - 0.7) & \text{if } y = 0.7, \end{cases}$$

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

$$= \begin{cases} \frac{1}{2\pi\cos(\sin^{-1}(y))} & \text{if } -1 \leq y < 0.7, \\ \frac{\pi - 2\sin^{-1}(0.7)}{2\pi}\,\delta(y - 0.7) & \text{if } y = 0.7. \\ 0 & \text{if otherwise} \end{cases}$$

The experimentally sampled data agrees with this Jacobian calcualtion and is shown in Fig 7(b), overlaid upon the $y = 0.7$ boundary line.

2.3 Inverse CDF method

2.3.1 Generation of exponential distribution using CDF and inverse CDF

For an exponential distribution $p(y) = exp(-y)$ for $y \geq 0$, its CDF and inverse CDF can be calculated as:

$$p(y) = \frac{df^{-1}(y)}{dy} = \exp(-y)$$

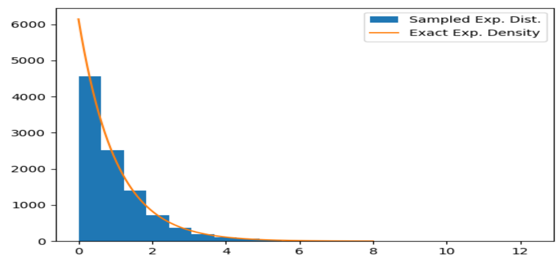
CDF: By definition of CDF, $F(y) = \int_0^y p(y)\,dy$

$$= 1 - \exp(-y)$$

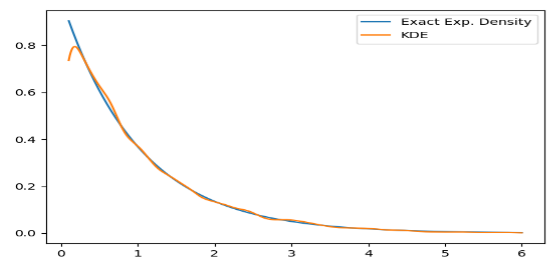
Inverse CDF: $y = f(x)\ x = f^{-1}(y)$

From CDF, $x = 1 - e^{-y}$

$$F^{-1}(x) = f(x) = y = -\ln(1 - x)$$



(a) Histogram



(b) KDE

Figure 8: *Sampled exponential distribution*

As demonstrated in Fig 8, it can be observed that the inverse CDF method produces reliable result generating the exponential distribution that matches with both histogram and KDE graphically.

2.3.2 Monte Carlo estimation of mean and variance of exponential distribution.

Theoretically, an exponential distribution of $\lambda = 1$ has

$$\mu = \frac{1}{lambda} = 1$$
$$\sigma^2 = \frac{1}{\lambda^2} = 1$$

Using Monte Carlo estimates, the mean and variance of the samples of exponential distribution can be calculated as:

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

	mean	variance
Theoretical	1	1
Sampled	1.05	0.96

Table 1: *Mean and variance of theoretical and sampled distributions*

Comparing the calculated mean and variance from the sampled data to that of the theoretical values, it can be observed that the values are approximately close, suggesting the reliability of the Monte Carlo estimation.

2.3.3 Proof of validity of Monte Carlo estimate

Unbiased mean estimate

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

where $y_1, ...y_n$are iid samples from $Y\ Exp(1)$

$$E[\hat{\mu}] = E[\frac{1}{n}\sum_{i=1}^{N} y_i]$$
$$= \frac{1}{N}E[\sum_{i=1}^{N} y_i]$$
$$= \frac{1}{N}\sum_{i=1}^{N} E[y_i]$$
$$= \frac{1}{N}\sum_{i=1}^{N} \mu = \mu$$

hence $E[\hat{\mu}] = \mu$, the Monte Carlo mean is unbiased.

Variance of Monte Carlo mean estimator

$$Var[\hat{\mu}] = Var[\frac{1}{N}\sum_{i=1}^{N} y_i]$$
$$= \frac{1}{N^2}\sum_{i=1}^{N} Var[y_i]$$
$$= \frac{1}{N^2}(N\sigma^2)$$
$$= \frac{\sigma^2}{N}$$
$$= E[\hat{\mu}^2] - E[\hat{\mu}]^2$$

$$From above proof, E[\hat{\mu}]^2 = \mu^2 = E[\mu]^2$$
$$Var[\hat{\mu}] = E[\hat{\mu}^2] - E[\mu]^2$$
$$= E[\hat{\mu}^2 - \mu^2] \propto \frac{1}{N}$$

2.3.4 Squared mean error $(\hat{\mu} - \mu)^2$ and Monte Carlo sample size

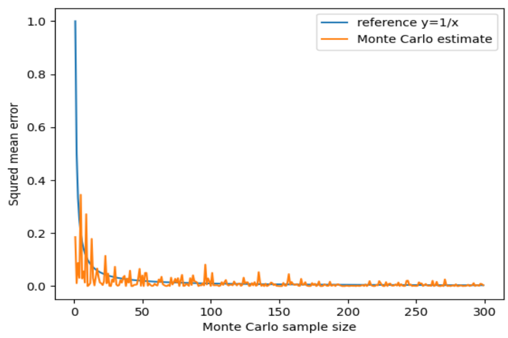


Figure 9: *Squared mean error with reference to* $y = \sqrt{(x)}$

Fig 9 shows that the squared error between a Monte Carlo mean estimate and the theoretical mean decays with increasing Monte Carlo sample size. This corresponds to a decay rate of $\frac{1}{N}$, where N is the sample size taken by Monte Carlo estimator, as the trend follows the $y = \sqrt{x}$ line in the figure.

As N grows, the squared mean error quickly approaches 0, further supporting the above proof that the Monte Carlo mean is unbiased.

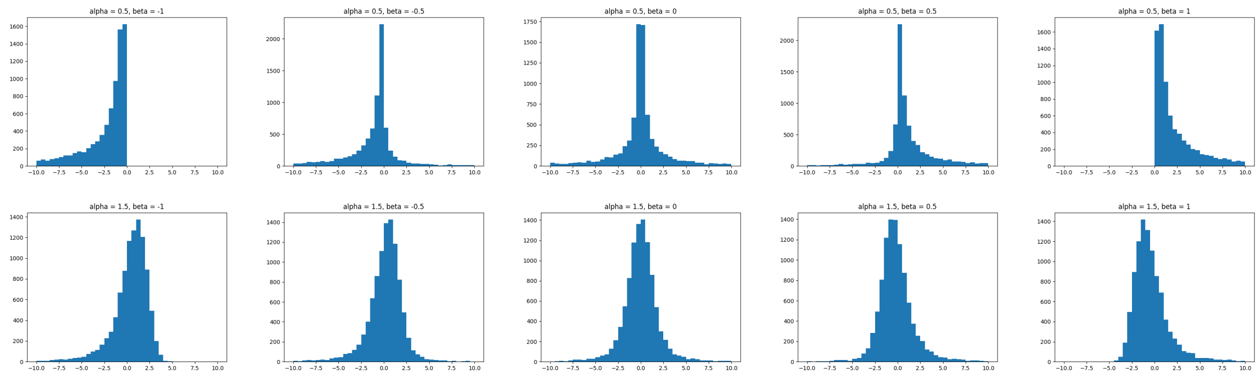


Figure 10: *Histogram of sampled RV X for different values of* α *and* β

2.4 Simulation for complex RV distributions

2.4.1 Generation of complex RV by sampling

One strength of using the sampling method to generate RV is that it allows construction of RV with complex density distributions, where a general density function would be hard to calculate using methods including the Jacobian formula.

For a certain RV X that is defined as:

$$\alpha \in (0,2)(a \neq 1), \beta \in [-1,+1]$$
$$b = \frac{1}{a} tan^{-1}(\beta tan(\pi\alpha/2))$$
$$s = (1+\beta^2 tan^2(\pi\alpha/a))^{\frac{1}{2\alpha}}$$
$$\text{Generate RVs: } U \sim \mathscr{U}(-\pi/2, +\pi/2)$$
$$V \sim \mathscr{E}(1)$$
$$X = s\frac{sin(\alpha(U+b))}{(cos(U))^{1/a}}(\frac{cos(U-\alpha(U+b))}{V})^{\frac{1-\alpha}{\alpha}}$$

Fig 10 shows histogram plots of the RV X when different values of the parameters α and β are chosen. Each figure is produced using sample size $N = 10000$ and 40 bins, each of width 0.5cm, which is tested to give an optimal demonstration of the density distribution.

Comparing the plots obtained for $\alpha = 0.5$(first row) and $\alpha = 1.5$(second row), it is observed that when α increase, the distribution becomes fuller and less steep, while a lower α value like 0.5 results in a sharp density peak centered around 0. Therefore α is likely to contribute to the spread of the distribution, and a higher α value gives a lower kurtosis.

Comparing the plots obtained for different values of β that spread across the given permitted range of it, with α fixed, it is observed that β controls the skewness of the sampled distribution.

A negative β value results in a negatively-skewed distribution where the median is left of the mode(peak of histogram), and a positive β value results in a positive skew where the median is on the right of the mode. At $\beta = 0$, the sampled distribution of x appears to be almost symmetrical around 0.0.

2.4.2 Tail probability for the complex distribution

The tail probability of X that $P(|X| > t)$ can be calculated as:

$$\begin{aligned} P(|X| > t) &= P(X > t) + P(X < -t) \\ &= (1 - P(X < t)) + P(X < -t) \\ &= 1 - F(t) + F(-t) \end{aligned}$$

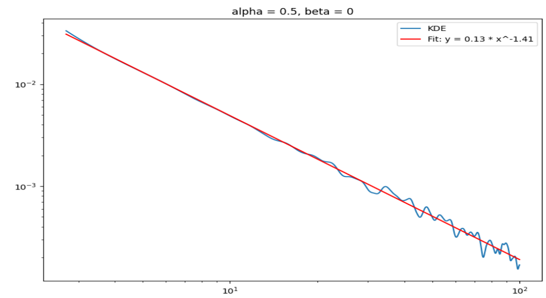
In terms of the distribution of X generated by sampling described in the last section, this tail probability can also be approximated to the sum of histogram count data of samples outside the tail boundary. When β is fixed at 0, this would result in the tail probability as follows for $\alpha = 0.5$ and $\alpha = 1.5$ at tail boundaries $t = 0, 3, 6$ (with reference to tail probabilities of a standard normal distribution):

As shown in Table 2, the tail probability of X for both α settings is much higher than that of a standard normal distribution, which indicates that X converges slower and hence is more spread-out than the standard normal. Additionaly, the $\alpha = 1.5$ distribution has

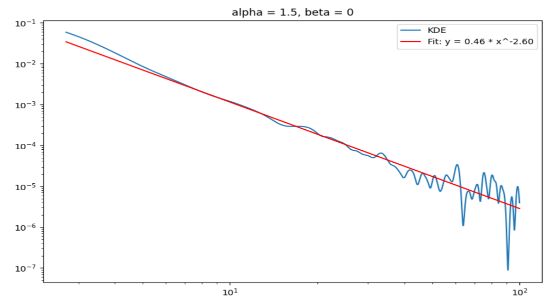
	t=0	t=3	t=6
$\alpha = 0.5$	1	0.3674	0.2768
$\alpha = 1.5$	1	0.1032	0.03129
$\mathcal{N}(0,1)$	1	0.0027	$1.973e^{-9}$

Table 2: *Tail probabilities of X at different parameter combinations and standard normal*

a significantly reduced tail probability than $\alpha = 0.5$, which agrees with the graphical representations as the density distribution becomes wider yet more concentrated around 0 as α increases.



(a) $\alpha = 0.5, \beta = 0$



(b) $\alpha = 1.5, \beta = 0$

Figure 11: *Tail probability of X and linear regression fittings (sample size 80000; log-log)*

From closer observation of the distribution as sample size N get large, it can be suspected that for large tail boundary $|x|$, the pdf can be approximated by a power expression $p(x) \approx cx^{\gamma}$, where c and γ are constants. This relationship is better depicted as a log-log equation $log(p(x)) = log(c) + \gamma log(x)$, which means $log(c)$ and γ can be obtained as y-intercept and slope of the graph in log-log scale.

Experimenting with $\alpha = 0.5$ or $\alpha = 1.5$, at $\beta = 0$, it is shown in Fig 11 that c is relatively stable around 0.15. When repeating the regression fitting for other α values in the permitted range, it is concluded that γ is related to α as $\gamma = -0.128\alpha - 0.077$.

2.4.3 Relating X with Gaussian distribution

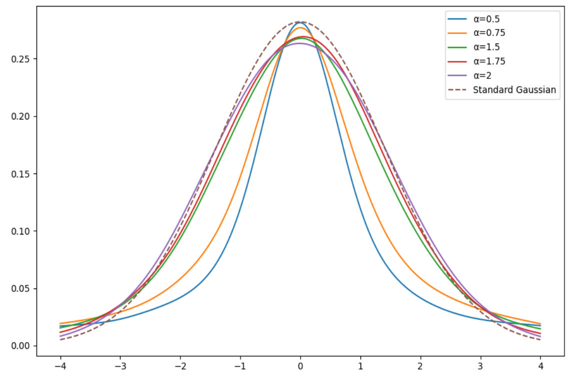


Figure 12: *X at different values of α compared to gaussian*

As shown in Fig 12, the distribution of X approaches that of Gaussina distribution with variance 2 $\mathcal{N}(0,2)$ as α get closer to 2. This result aligns with the $\alpha = 2$ special case of the stable distribution.

3 Conclusion

In this report, the investigations on applying Python programming to statistics of RV is explored, with a focus on data visualization using histograms and KDE, and generation of specific RV distributions using various methods including the Jacobian formula and inverse CDF methods. The capabilities of generating distributions through sampling is also discussed in detail, providing a better understanding of some of the properties of the stable diffusion. Manipulations of sampled data including Monte Carlo estimates of mean and variance and tail probability calculations is investigated in this lab exercise, which prove to be useful tools that can transfer to other statistical applicaitons.

In conclusion, generating and presenting RV distributions using the above discussed methods and transformations contains great potentials and is expected to employ wide usage, as it makes scientific and engineering investigations efficient and accurate. Further explorations on this topic remains open as examining more complex and advanced methods would strengthen understanding of this profound area.

ENGINEERING TRIPOS PART II A

EIETL MODULE EXPERIMENT 3F3

RANDOM VARIABLES and RANDOM NUMBER GENERATION
Short Report Template

Name: Shanzi (Monica) Ran

College: Magdalene College

This is a template suitable for the short report write-up. Simply edit the Latex or Word document to include your calculations/ results/ code.

1. **Uniform and normal random variables.**

   Histogram of Gaussian random numbers overlaid on exact Gaussian curve (scaled):

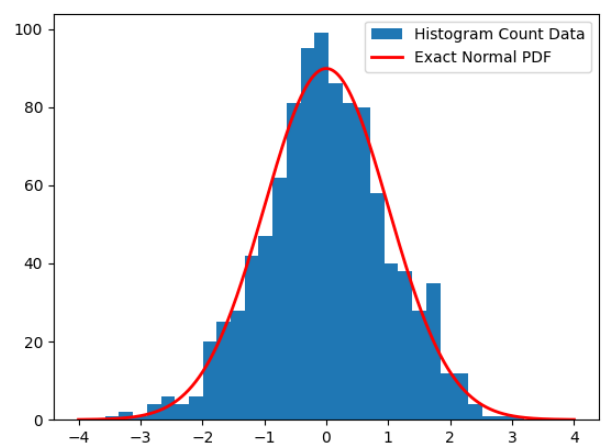


Figure 1: Histogram of Gaussian random numbers overlaid on exact Gaussian curve

Histogram of Uniform random numbers overlaid on exact Uniform curve (scaled):

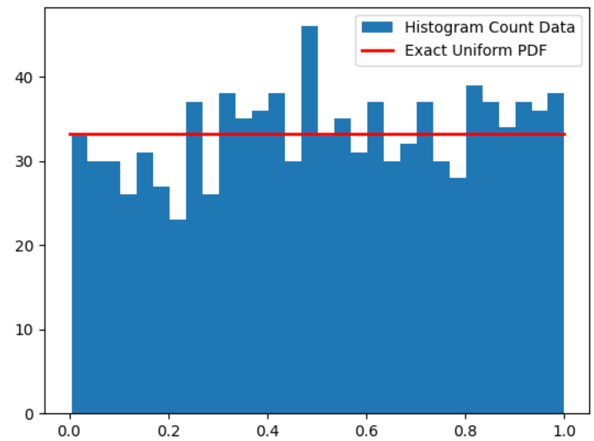


Figure 2: Histogram of Uniform random numbers overlaid on exact Uniform curve

Kernel density estimate for Gaussian random numbers overlaid on exact Gaussian curve:

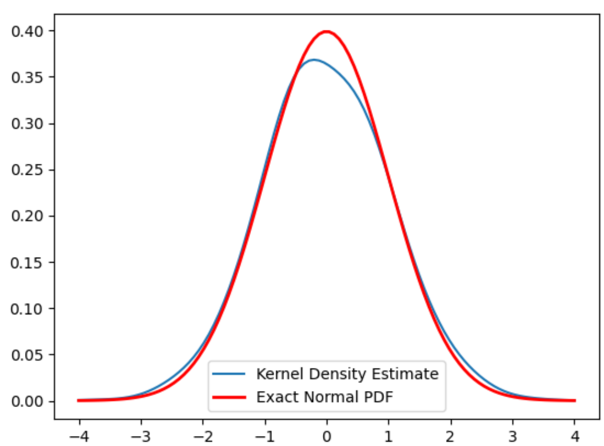


Figure 3: Kernel density estimate for Gaussian random numbers overlaid on exact Gaussian curve

Kernel density estimate for Uniform random numbers overlaid on exact Gaussian curve:

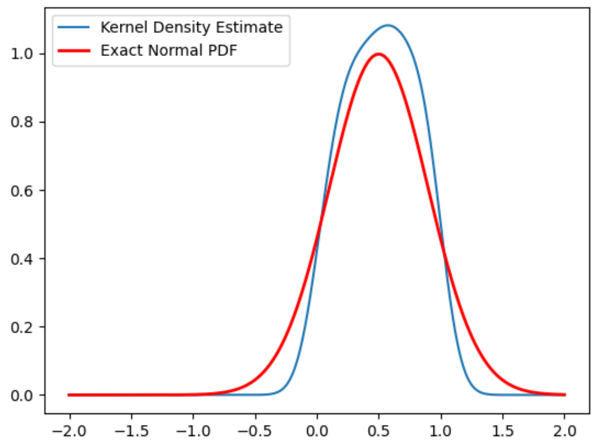


Figure 4: Kernel density estimate for Uniform random numbers overlaid on exact Gaussian curve

Comment on the advantages and disadvantages of the kernel density method compared with the histogram method for estimation of a probability density from random samples:

Kernel Density Method (KDE) gives an estimation with a more continuous presentation due to its smoothing function, while histogram presentations are better for discrete data.

On the other hand, KDE requires a careful choice of the operation bandwidth as overfitting (bandwith too small, estimation not representative of original data) or underfitting (bandwidth too small, estimation very spiky and difficult for statistical analysis) are highly likely. Therefore, histograms are generally more intuitive and robust to bandwidth selection.

Theoretical mean and standard deviation calculation for uniform density as a function of N:

$X \sim \mathcal{U}(0,1)$

Assume histogram has j bins, for N random variables.

Since $\displaystyle\sum_{i=0}^{j} p_i = 1, \quad p_i = \frac{1}{j}$.

Mean per bin: $Np_j = \dfrac{N}{j}$.

Variance per bin: $Np_j(1-p_j) = \dfrac{N}{j}\left(1-\dfrac{1}{j}\right)$.

Uniform distribution implies mean and variance are the same for each bin.

Overall (summed up): Mean $= \dfrac{N}{j} \propto N, \quad$ Standard deviation $= \sqrt{\dfrac{N}{j}\left(1-\dfrac{1}{j}\right)} \propto \sqrt{N}$.

Explain behaviour as N becomes large:

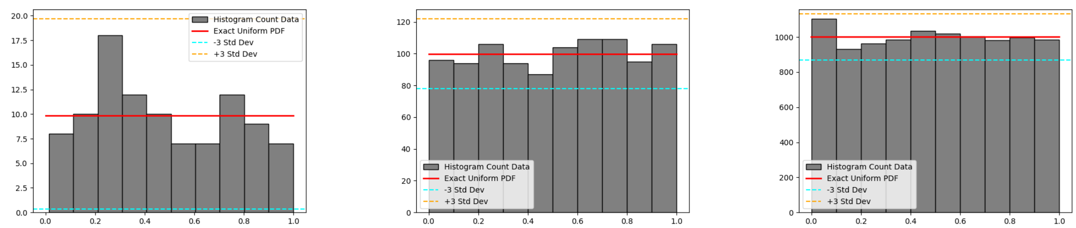
From above,

overall mean $= \dfrac{N}{j} \propto N$

overall standard deviation $= \sqrt{\dfrac{N}{j}(1-\dfrac{1}{j})} \propto \sqrt{N}$

so as N grow large, mean of histogram count grows at$N$, standard deviation grows at$N\sqrt{N}$.

Plot of histograms for $N = 100$, $N = 1000$ and $N = 10000$ with theoretical mean and ± 3 standard deviation lines:



(a) N=100 (b) N=1000 (c) N=10000

Figure 5: Histograms for different values of N with theoretical mean and standard deviation lines

Are your histogram results consistent with the multinomial distribution theory?

As shown in Figure 5, the mean of histogram count data approaches the theoretical value (marked by the red horizontal line) as N grows larger. The standard deviation of the sampled data points is also shrinking, indicating that the distribution of bin counts is more uniform than before. Therefore, the histogram results in Figure 5 is consistent with the multinomial distribution theory from which the mean and standard deviation formula for each bin j is developed from.

2. **Functions of random variables** For normally distributed $\mathcal{N}(x|0,1)$ random variables, take $y = f(x) = ax + b$. Calculate $p(y)$ using the Jacobian formula:

$X \sim \mathcal{N}(0,1)$

$y = f(x) = ax + b$, so $x = f^{-1}(y) = \dfrac{y-b}{a}$

$\dfrac{dy}{dx} = a$

$$\begin{aligned} p(y) &= \frac{p(x)}{\left|\frac{dy}{dx}\right|}\Bigg|_{x=f^{-1}(y)} \\ &= \frac{1}{a} p\left(\frac{y-b}{a}\right) \\ &= \frac{1}{\sqrt{2\pi}a} exp(-\frac{1}{2}(\frac{y-b}{a})^2) \end{aligned}$$

Explain how this is linked to the general normal density with non-zero mean and non-unity variance:

Compare the pdf of this transformed y with that of a standard normal, it is observed that:

$$\mu = b, \quad \sigma = a$$

This indicate that y corresponds to a normal density distribution with mean $= b$, variance $= a^2$.

Verify this formula by transforming a large collection of random samples $x^{(i)}$ to give $y^{(i)} = f(x^{(i)})$, histogramming the resulting y samples, and overlaying a plot of your formula calculated using the Jacobian:

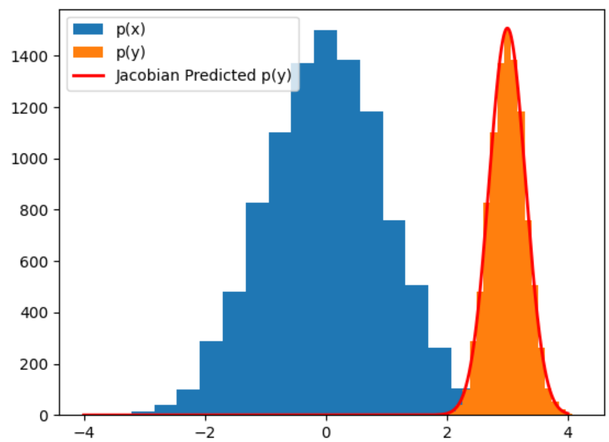


Figure 6: Sampled transformation for y=ax+b

Now take $p(x) = \mathcal{N}(x|0, 1)$ and $f(x) = x^2$. Calculate $p(y)$ using the Jacobian formula:

$X \sim \mathcal{N}(0, 1)$

$y = f(x) = x^2$, so $x = f^{-1}(y) = \pm\sqrt{y}$, where $x_1 = \sqrt{y}$, $x_2 = -\sqrt{y}$

$$\frac{dy}{dx} = 2x$$

$$\begin{aligned}
p(y) &= \sum_{k=1}^{2} \frac{p(x)}{|\frac{dy}{dx}|}|_{x=x_k(y)} \\
&= \frac{1}{\sqrt{2\pi}} exp(-\frac{y}{2})(\frac{1}{2\sqrt{y}} + |\frac{1}{2\sqrt{y}}|) \\
&= \frac{1}{\sqrt{2\pi y}} exp(-\frac{y}{2})
\end{aligned}$$

Verify your result by histogramming of transformed random samples:

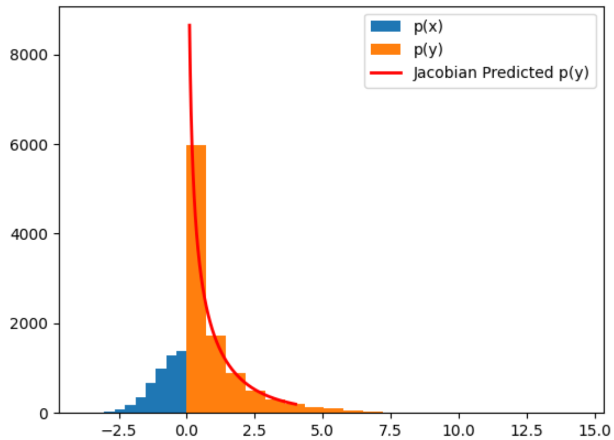


Figure 7: Sampled transformation for $y = x^2$

3. **Inverse CDF method**

   Calculate the CDF and the inverse CDF for the exponential distribution:

   $p(y) = \dfrac{df^{-1}(y)}{dy} = \exp(-y)$

   CDF: By definition of CDF, $F(y) = \int_0^y p(y)\,dy$

   $$= \int_0^y \exp(-y)\,dy$$

   $$= 1 - \exp(-y)$$

   Inverse CDF: $y = f(x)$ $x = f^{-1}(y)$

   From CDF, $x = 1 - e^{-y}$

   $F^{-1}(x) = f(x) = y = -\ln(1-x)$

   Matlab/Python code for inverse CDF method for generating samples from the exponential distribution:

```
import numpy as np

def inverse_cdf(n, exp_lambda):
    x = np.random.rand(n)
    samples = - np.log(1 - x) / exp_lambda
    return samples
```

Plot histograms/ kernel density estimates and overlay them on the desired exponential density:

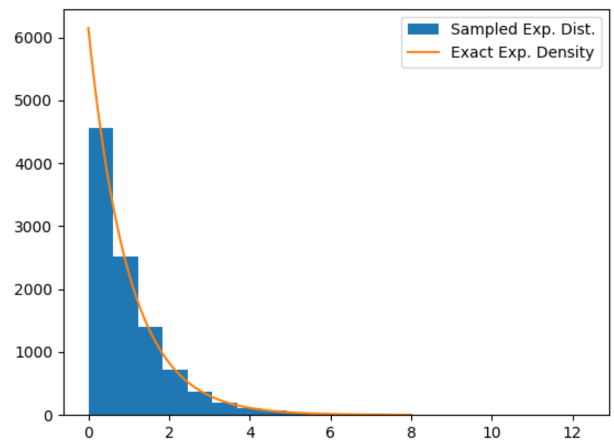


Figure 8: Generating samples from exponential distribution (histogram)

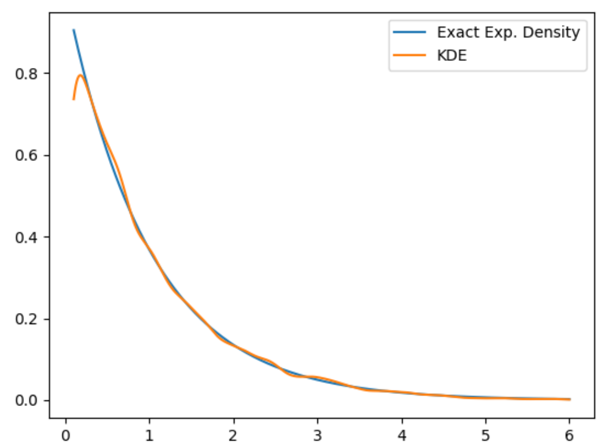


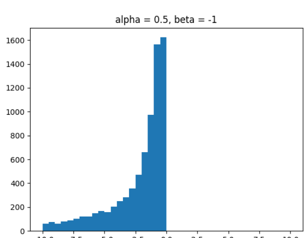
Figure 9: Generating samples from exponential distribution (KDE)

4. **Simulation from a ‘non-standard’ density.**

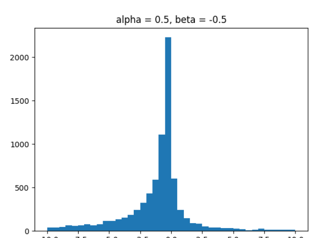
Matlab/Python code to generate N random numbers drawn from the distribution of X:

```
import numpy as np

def gen_x(n, alpha, beta):
    b = np.arctan(beta * np.tan(np.pi * alpha / 2)) /
       alpha
    s = (1 + (beta ** 2) * (np.tan(np.pi * alpha / 2)
       ** 2)) ** (1 / (2 * alpha))

    u = np.random.uniform(low=(-np.pi/2), high=(np.pi
       /2), size=n)
    v = np.random.exponential(scale=1, size=n)

    x = s * np.sin(alpha * (u + b)) * ((np.cos(u -
       alpha * (u + b)) / v) ** ((1 - alpha)/alpha)) /
       (np.cos(u) ** (1 / alpha))
    return x
```

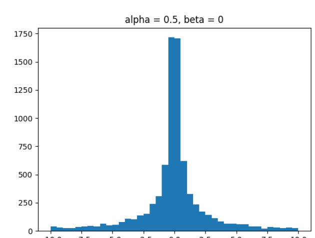
Plot some histogram density estimates with $\alpha = 0.5$, 1.5 and several values of β.



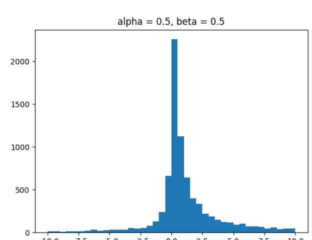
(a) $\alpha = 0.5, \beta = -1$



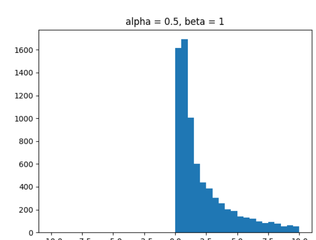
(b) $\alpha = 0.5, \beta = -0.5$



(c) $\alpha = 0.5, \beta = 0$

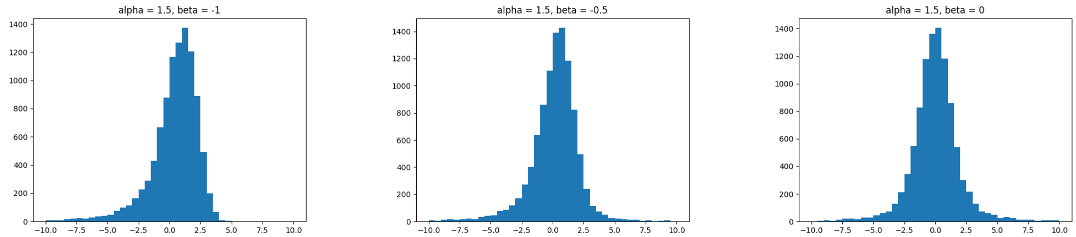


(d) $\alpha = 0.5, \beta = 0.5$

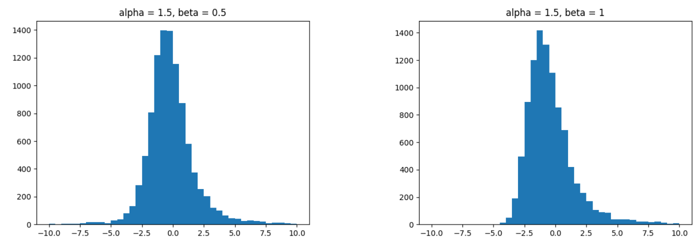


(e) $\alpha = 0.5, \beta = 1$

Figure 10: Histograms density estimates for $\alpha = 0.5$ at $\beta \in \{-1, -0.5, 0, 0.5, 1\}$



(a) $\alpha = 1.5, \beta = -1$ (b) $\alpha = 1.5, \beta = -0.5$ (c) $\alpha = 1.5, \beta = 0$



(d) $\alpha = 1.5, \beta = 0.5$ (e) $\alpha = 1.5, \beta = 1$

Figure 11: Histograms density estimates for $\alpha = 1.5$ at $\beta \in \{-1, -0.5, 0, 0.5, 1\}$

Hence comment on the interpretation of the parameters α and β.

Comparing the plots obtained for $\alpha = 0.5$ and $\alpha = 1.5$, it is observed that when α increase, the distribution becomes fuller and less steep, while a lower α value like 0.5 results in a sharp density peak centered around 0. Therefore α is likely to contribute to the **kurtosis** of the distribution, and a higher α value gives a lower kurtosis.

Comparing the plots obtained for different values of β that spread across the given permitted range of it, and as α is fixed, it is observed that β controls the **skewness** of the sampled distribution. A negative β value results in a negatively-skewed distribution where the median is left of the mode(peak of histogram), and a positive β value results in a positive skew where the median is on the right of the mode. At $\beta = 0$, the sampled distribution of x appears to be almost symmetrical around 0.0.