RANDOM VARIABLES and RANDOM NUMBER GENERATION
Full Technical Report

Name: Abhishek Shenoy (ams289)

College: Pembroke College

1 Introduction

Random variate analysis is an important part of many engineering applications, this investigation focuses on simulating and analysing the behaviour of different random variables with various different properties. We consider approximation by using the multinomial distribution, functions of random variables, convergence of the Monte-Carlo simulation from the inverse CDF method as well as the derivation of Student's T distribution from a mixture of gaussians and its behaviour.

2 Multinomial Distribution

Explain the form of the multinomial distribution; in particular why do we get the terms $p_j^{n_j}$, and why the factorial terms?

If there are N items that need to be placed in k groups. If you place n_1 items in group 1, n_2 items in group 2, and so on till you place n_k items in the last group, the number of possible combinations Z is given by the multinomial coefficient:

$$Z = \frac{N!}{n_1!n_2!...n_J!}$$

The power terms of probability $p_i^{n_i}$ comes from the multiplication of probabilities for each outcome. So consider a set of symbols $\{x_1, x_2, ..., x_J\}$, and each symbol x_i has a probability p_i of occurring. So if a symbol x_i occurs n_i times then the joint probability (of that specific sequence) is given by $p_i^{n_i}$. However we need to consider this for every symbol (ie. for every i), hence giving the following expression:

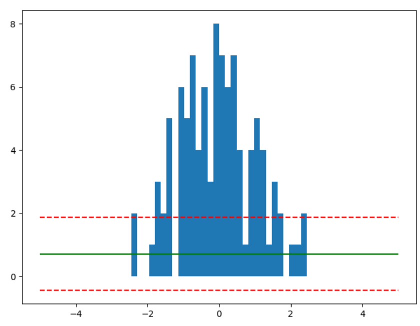
$$\prod_{i=1}^{J} p_i^{n_i}$$

We can now bring the idea of being able to reorder the sequence of outcomes where the number of possible combinations is the multinomial coefficient Z, therefore giving the full multinomial distribution.

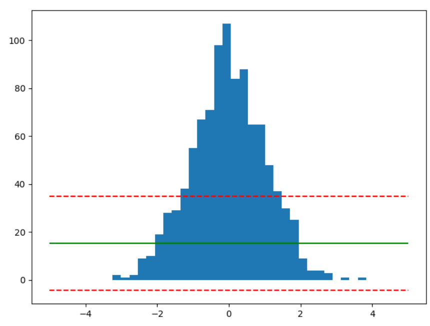
$$P(n_1, n_2, ..., n_J) = Z\prod_{i=1}^{J} p_i^{n_i} = \frac{N!}{n_1!n_2!...n_J!}\prod_{i=1}^{J} p_i^{n_i} = N!\prod_{i=1}^{J}\frac{p_i^{n_i}}{n_i!}$$

where $n_1, n_2, ..., n_J$ are the counts of all the outcomes for their respective symbols $x_1, x_2, ..., x_J$.

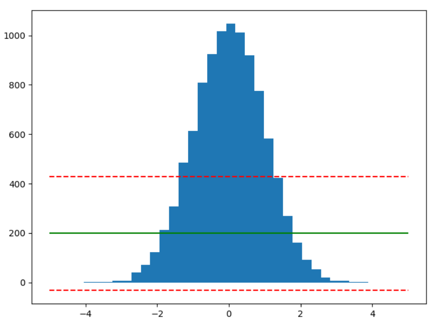
Now, by generating samples from the normal distribution, we can plot the histogram along with the mean and variance calculated using the CDF from the bin probabilities p_j. From the multinomial distribution theory, the mean of the count data in bin j is Np_j and the variance is $Np_j(1-p_j)$.



(a) N = 100



(b) N = 1000



(c) N = 10000

Figure 1: Plot of normally distributed histograms for $N = 100$, $N = 1000$ and $N = 10000$ with theoretical mean (green) and ± 3 standard deviation lines (red)

As the number of bins with probabilities closer to 1 increase, the mean increases. This can be easily noticed from the fact that the bin count mean Np_j is maximised when $p_j = 1$. In general as N increases, the histogram variance increases however much slower than that of the uniform distribution. Bins with probabilities close to 0 or 1 have the lowest standard deviation. The standard deviation of the bins is maximised for $p_j = 1/2$ and hence intermediate probability bins have higher standard deviation. Therefore as the number of bins with probabilities closer to 0 or 1 increase, the histogram variance decreases but for intermediate probabilities, the variance increases. We can also consider the ratio $R = \frac{\sigma_j^2}{\mu_j} = \frac{\sqrt{Np_j(1-p_j)}}{Np_j}$ for each bin, we can see that $R \to 0$ as $p_j \to 1$ (and $R \to 1$ for $p_j \to 0$ since $\mu = \sigma = 0$). This provides a perspective on how the relative increase in the standard deviation compares to the relative increase in the mean. So for intermediate probabilities, this holds more significance since $R \propto \frac{1}{\sqrt{N}}$ where N is very large.

3 Functions of Random Variables

Take $p(x) = U(x|0, 2\pi)$ and $f(x) = \sin(x)$ corresponding to measuring a carrier signal in a communications system at a random phase offset. Determine the theoretical probability density for $f(x)$ and verify once again by comparison with transformed random samples.

For a given random variable y which is a function $f(x)$ of another random variable x with the density function $p_x(x)$, then the following is true where $f_k^{-1}(y)$ is one of K inverse functions of f:

$$p_y(y) = \sum_{k=1}^{K} \left. \frac{p_x(x)}{\left|\frac{dy}{dx}\right|} \right|_{x=f_k^{-1}(y)} \tag{1}$$

$$y = \sin(x) \text{ and } p_x(x) = \frac{1}{2\pi} \qquad 0 \leq x \leq 2\pi$$
$$\frac{dx}{dy} = \frac{d}{dy}\arcsin(y) = \frac{1}{\sqrt{1-y^2}}$$

For the range $[0, 2\pi]$, there are two inverse functions that are exactly the same just shifted by 2π.

$$\begin{aligned} p_y(y) &= \sum \frac{p_x(x)}{|\frac{dy}{dx}|} \\ &= p_x(\arcsin(y))\ |\frac{d}{dy}\arcsin(y)| + p_x(\pi - \arcsin(y))\ |\frac{d}{dy}(\pi - \arcsin(y))| \\ &= 2(\frac{1}{2\pi})\ |\frac{d}{dy}\arcsin(y)| = 2(\frac{1}{2\pi})\ |\frac{dx}{dy}| \\ &= \frac{1}{\pi\sqrt{1-y^2}} = \frac{1}{\pi\cos(\arcsin y)} \qquad -1 \leq y \leq 1 \end{aligned}$$

Fig. 2 verifies that the analytical solution of the probability density function matches that of the simulation.

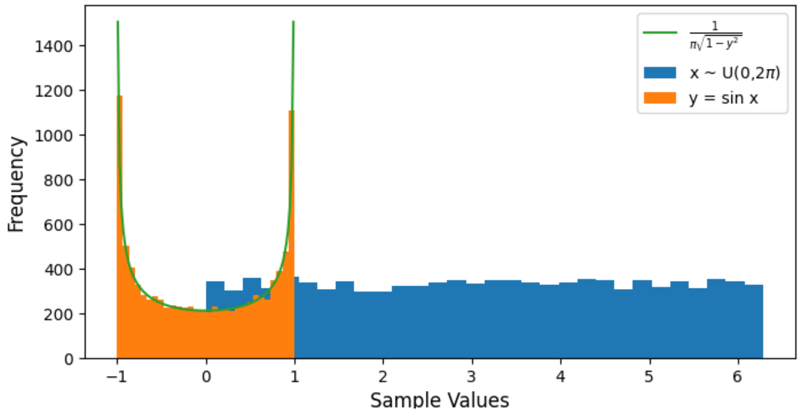


Figure 2: Plot of histograms for x and $y = sin(x)$ with an analytical solution for $P_Y(y)$.

Now, consider a limited sine function (idealised saturation): $f(x) = min(sin(x), 0.7)$

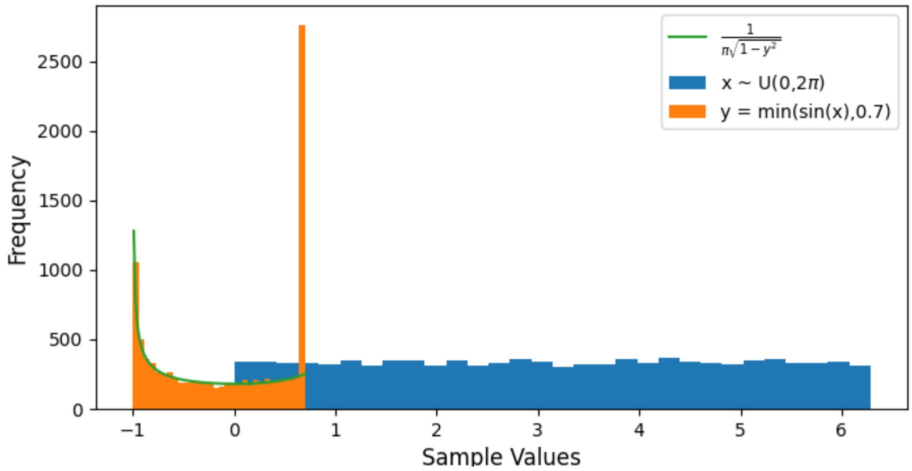


Figure 3: Plot of histograms for x and the limited sine function y

The pdf of this new distribution essentially saturates any value above 0.7 to 0.7 as shown in Fig. 3. This is essentially equivalent to the addition of a delta function $(F_y(1) - F_y(0.7))\delta(y - 0.7) = (1 - F_y(0.7))\delta(y - 0.7)$ to the pdf so that the density of any value ≥ 0.7 is concentrated at 0.7.

This could be predicted from the Jacobian formula by the fact that the saturation above 0.7 acts as a step function and since the delta function is the derivative of the step function, we get the delta function with a value of the PDF area for $x \geq 0.7$.

$$p_y(y) = \begin{cases} \dfrac{1}{\pi\sqrt{1-y^2}} & -1 \leq y < 0.7 \\ \dfrac{1}{\pi}(\arcsin(1) - \arcsin(0.7))\delta(y - 0.7) & y = 0.7 \end{cases}$$

This simplifies to the following:

$$p_y(y) = \begin{cases} \dfrac{1}{\pi\sqrt{1-y^2}} + (\dfrac{1}{2} - \dfrac{1}{\pi}\arcsin(0.7))\delta(y - 0.7) & -1 \leq y \leq 0.7 \end{cases}$$

From this it is possible to see that the delta function is present as the histogram has a sharp rise in frequency at a value of 0.7.

4 Inverse CDF method

Given the exponential distribution $p(x) = e^{-x}$:

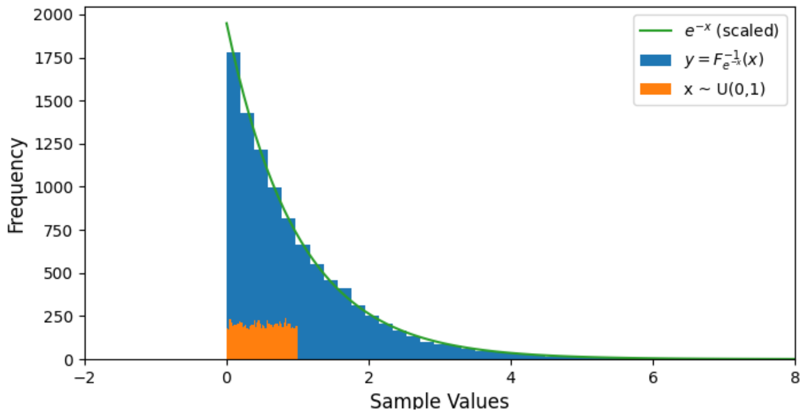


Figure 4: Histogram for the inverse CDF method on the exponential distribution with the desired analytical p.d.f

By generating random samples to estimate the mean and variance of the exponential distribution, we verify that they are close to the expected theoretical values for this exponential distribution.

$$\mu = E(y) = \int_{y=0}^{\infty} yp(y)dy \approx \frac{1}{N}\sum_{i=0}^{N} y_i = \hat{\mu}$$

$$\sigma^2 = E(Y^2) - \mu^2 = \int_{y=0}^{\infty} y^2 p(y)dy - \mu^2 \approx \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \hat{\mu}^2 = \hat{\sigma}^2$$

From the Python simulation using the inverse CDF method to generate the values, we get the following estimates:
Mean $\quad \hat{\mu} = 0.9984917373711222$
Variance $\quad \hat{\sigma}^2 = 0.9931753347028838$

This is very close to being as expected where the true values are $\mu = 1, \sigma^2 = 1$.

Below we show that the Monte Carlo mean estimate $\hat{\mu}$ is unbiased, ie. that $E(\hat{\mu}) = \mu$, where the first expectation is taken wrt the distribution of the random samples y(i), ie. p(y).

$$\begin{aligned}
E(\hat{\mu}) &= E(\frac{1}{N}\sum_{i=1}^{N} y_i) = \frac{1}{N}\sum_{i=1}^{N} E(y_i) \\
&= \frac{1}{N}\sum_{i=1}^{N} E(Y) = \frac{1}{N}\sum_{i=1}^{N} \mu \\
&= \frac{1}{N} N\mu \\
E(\hat{\mu}) &= \mu
\end{aligned}$$

Below we show from expectation formulae that the Monte Carlo mean estimator has variance equal to σ^2/N, ie. $E(\hat{\mu}^2 - \mu^2) \propto 1/N$

$$\begin{aligned} E((\hat{\mu}^2 - \mu^2)) &= E(\hat{\mu}^2) - E(\mu^2) \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E(y_i y_j) - \mu^2 \end{aligned}$$

Note however that this is essentially the sum of correlation functions:

$$\sum_i \sum_j E(y_i y_j) = \sum_i \sum_j r_{yy}[j-i]$$

$$r_{yy}[k] = \begin{cases} E(Y_n^2) & k = 0 \\ E(Y_n Y_{n+k}) = E(Y_n)E(Y_{n+k}) & k \neq 0 \end{cases}$$

$$\begin{aligned} E((\hat{\mu}^2 - \mu^2)) &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E(y_i y_j) - \mu^2 \\ &= \frac{1}{N^2}(NE(y^2) + (N^2 - N)E(y)E(y)) - \mu^2 \\ &= \frac{1}{N^2}(NE(y^2) + (N^2 - N)\mu^2) - \mu^2 = \frac{1}{N^2}(NE(y^2) - N\mu^2) \\ &= \frac{1}{N}(E(y^2) - \mu^2) = \frac{\sigma^2}{N} \end{aligned}$$

We illustrate the above property by plotting the squared error $(\hat{\mu} - \mu)^2$ **as the number of Monte Carlo samples increases.**

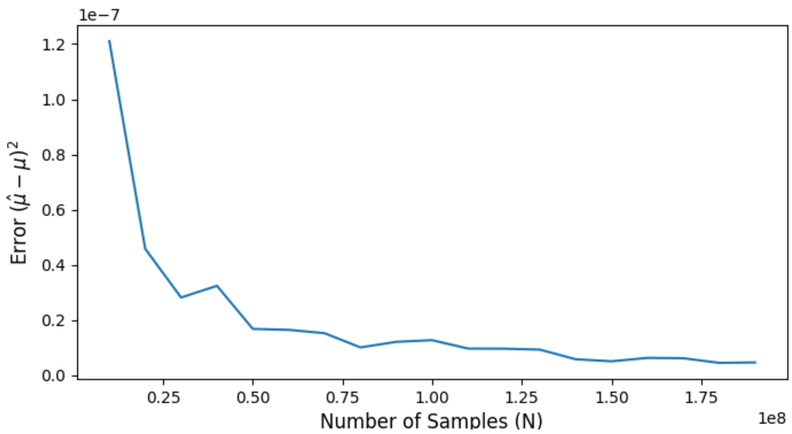


Figure 5: Plot of squared error against the number of Monte Carlo samples

In Fig.5, we have shown that the squared error is proportional to $1/N$ as the number of Monte Carlo samples increases. We plotted the squared error against N for different values of N. This was accomplished by averaging the error over lots of Monte Carlo estimates of $\hat{\mu}$ for each sample set $\underline{y}$ containing N number of samples.

Here we have shown that the Monte Carlo estimator is consistent since it delivers the true value of μ as the number of samples N goes to infinity.

This is considered quite a slow convergence rate as the values of the estimate only converge as $1/\sqrt{N}$ and hence Monte Carlo should only be used when other approximations fail.

5 Simulation from a ‘non-standard’ density

Now we investigate a distribution formed by a scaled mixture of Gaussians using the gamma distribution $G(x|a, b)$ where a and b are positive valued parameters. Here, $\Gamma(.)$ represents the gamma function, which is the normalising constant of the density (it ensures that the pdf integrates to 1).

$$p(x) = \int_0^\infty \mathcal{N}(x|0, u)p(u)du \tag{2}$$

We will consider the Gamma distribution

$$\mathcal{G}(v|a, b) = \frac{1}{b^a\Gamma(a)}v^{a-1}e^{-\dfrac{v}{b}}$$

along with the use of the transformation $u = 1/v$.

The recipe for generating X is now:

(a)Pick a value for parameter $\theta > 0$

(b)Generate a sample from the Gamma distribution $p(v) = \mathcal{G}(v|\theta, 1/\theta)$

(c)Calculate the variance as $u = 1/v$

(d)Sample from a Normal distribution $p(x|u) = N(x|0, u)$

$(e)x$ is the random variable of interest.

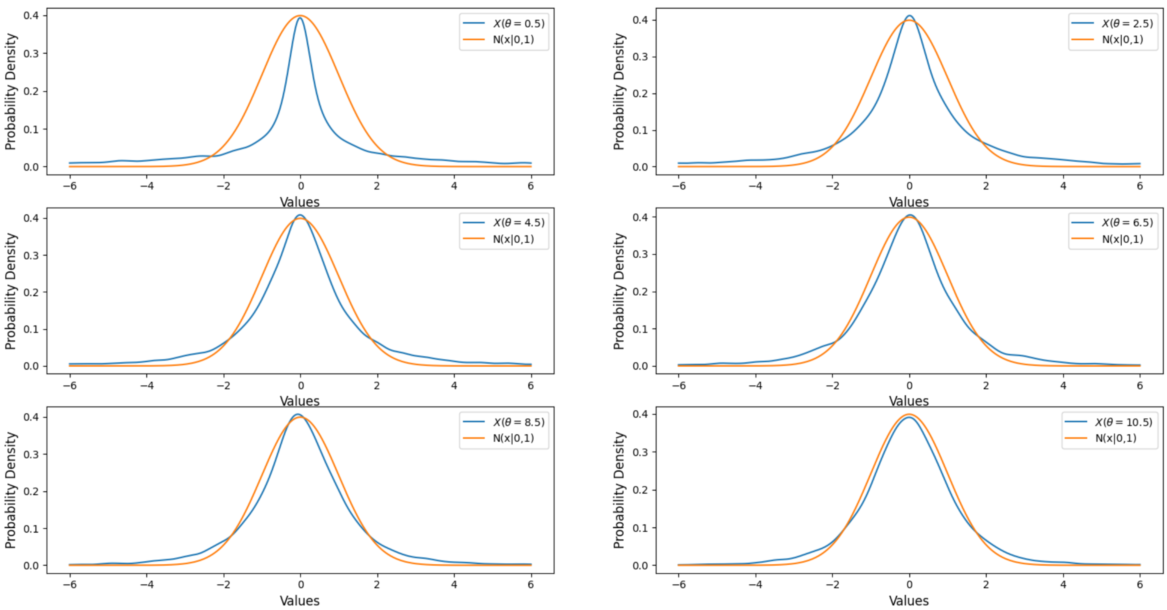


Figure 6: Plot of resulting distribution $p_X(x)$ parametrised by θ

How does θ affect the final distribution? What happens as θ gets large? How does the resulting distribution differ from the previous one with an exponential mixing distribution?

As theta decreases, the distribution becomes more concentrated at 0. As theta tends to infinity, $p_X(x)$ tends to $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ (ie. a standard normal distribution $\mathcal{N}(x|0,1)$). This behaviour is very similar to that of varying the standard deviation of the normal distribution or the scale of the Laplace distribution.

The exponential mixing distribution (Laplace distribution) is sharper at 0 than the resulting distribution due to the exponentials causing the maximum to occur at 0. Also unlike the exponential mixing distribution for which the tails tend towards 0, the tails of the resulting distribution is much thicker for large values of θ.

Investigate the tail behaviour for large values of $|x|$, and suggest an approximate formula for the tails in the form $p(x) \propto |x|^{-\beta}$, where β will be related to θ, using suitably devised logarithmic kernel density plots.

By plotting the logarithmic kernel density plots and by plotting the $\ln(|x|^{-\beta})$, we can determine and verify that the tail approximation is correct if the gradients of the two plots are the same for large values of x.

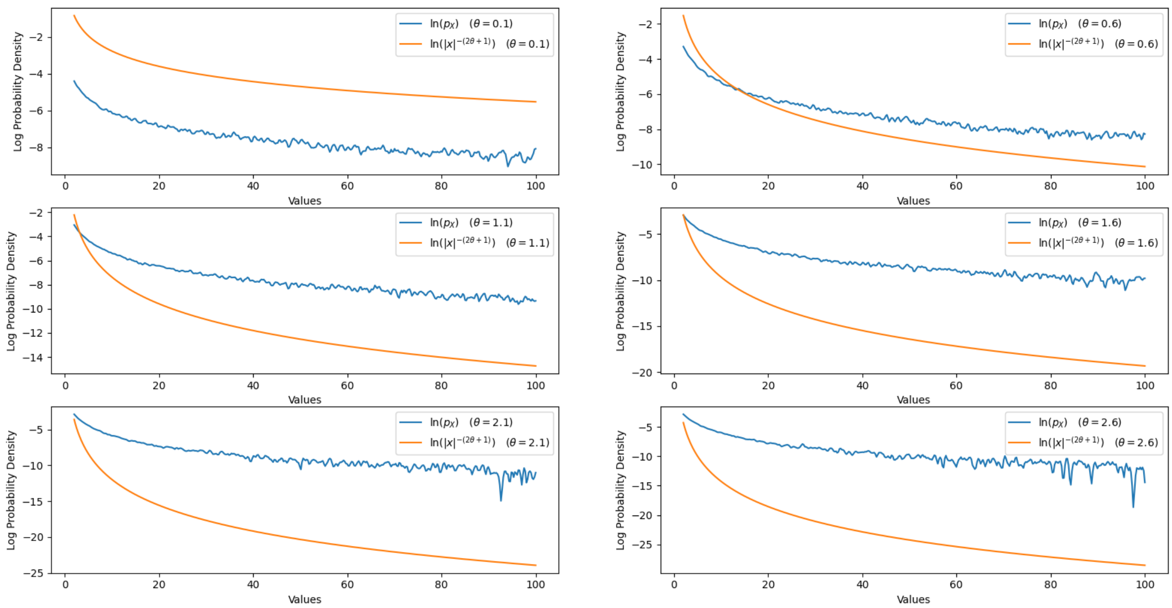


Figure 7: Plot of logarithmic KDE of $p_X(x)$ and the approximation parametrised by θ

By considering and testing β as functions of different orders of magnitude of θ, we find that $\beta \propto \theta$. By testing values, we were able to determine an approximation for $\beta = 2\theta + 1$. We can see that when theta is very small $\theta = 0.1$, the approximation for the tails is more visibly correct for the range of values that we have used.

Determine the probability density function for $U = 1/V$ by the Jacobian method.
Using the transformation $v = 1/u$, we can find the pdf $p_U(u)$:

$$
\begin{aligned}
p_U(u) &= p_V(v)|\frac{dv}{du}| \\
&= p_V\Big(\frac{1}{u}\Big)| - \frac{1}{u^2}| \\
&= \frac{1}{b^a\Gamma(a)}\Big(\frac{1}{u}\Big)^{a-1} e^{-\frac{\left(\frac{1}{u}\right)}{b}} | - \frac{1}{u^2}| \\
&= \frac{1}{b^a\Gamma(a)} u^{1-a}\ e^{-\frac{1}{bu}}\ u^{-2} \\
&= \frac{1}{b^a\Gamma(a)} u^{-1-a} e^{-\frac{1}{bu}}
\end{aligned}
$$

Now derive the formula for p(x) using Eq. 2 and your expression for $p(u)$. The result should be expressed in terms of the gamma function, $\Gamma()$, x and θ. Is this formula consistent with your analysis of the tail behaviour above?

$$
\begin{aligned}
p_X(x) &= \int_{u=0}^{u=\infty} \mathcal{N}(x|0,u)p_U(u)du \\
&= \int_{v=0}^{v=\infty} \mathcal{N}(x|0,\frac{1}{v})p_V(v)dv \quad \text{(By substituting } u = \frac{1}{v} \text{ giving } p_U(u)du = -p_V(v)dv) \\
&= \int_{v=0}^{v=\infty} \frac{\sqrt{v}}{\sqrt{2\pi}} e^{-\frac{1}{2}(vx^2)} \cdot \frac{1}{b^a\Gamma(a)} v^{a-1} e^{-\frac{v}{b}} dv \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{b^a\Gamma(a)}\int_{v=0}^{v=\infty} v^{a-\frac{1}{2}} e^{-\frac{1}{2}v(x^2+\frac{2}{b})} dv \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{b^a\Gamma(a)}\int_{v=0}^{v=\infty} v^{(a+\frac{1}{2})-1} e^{-\frac{v}{(\frac{x^2}{2}+\frac{1}{b})^{-1}}} dv \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{b^a\Gamma(a)}\int_{v=0}^{v=\infty} v^{A-1} e^{-\frac{v}{B}} dv \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{b^a\Gamma(a)} B^A\ \Gamma(A) \qquad \text{where } A = a + \frac{1}{2} \text{ and } B = \Big(\frac{x^2}{2} + \frac{1}{b}\Big)^{-1} \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{b^a\ \Gamma(a)}\ \Big(\frac{x^2}{2} + \frac{1}{b}\Big)^{-(a+\frac{1}{2})}\ \Gamma(a + \frac{1}{2}) \\
p_X(x) &= \frac{1}{\sqrt{2\pi}}\frac{\Gamma(a+\frac{1}{2})}{b^a\ \Gamma(a)}\ \Big(\frac{x^2}{2} + \frac{1}{b}\Big)^{-(a+\frac{1}{2})}
\end{aligned}
$$

In terms of θ, ie. when $a = \theta$ and $b = \frac{1}{\theta}$.

$$
p_X(x) = \frac{1}{\sqrt{2\pi}}\frac{\Gamma(\theta+\frac{1}{2})}{\Gamma(\theta)}\theta^\theta\ (\frac{x^2}{2} + \theta)^{-(\theta+\frac{1}{2})}
$$

The resulting distribution is Student's t distribution with 2θ degrees of freedom while the exponential mixing distribution is a form of the Laplace distribution. This is consistent with the analysis of the tail behaviour since the the t distribution is a fat-tailed distribution unlike a normal distribution which decays exponentially.

By considering approximations, we can get the form of the tail distribution:

$$
\begin{aligned}
p_X(x) &= \frac{1}{\sqrt{2\pi}} \underbrace{\frac{\Gamma(\theta+\frac{1}{2})}{\Gamma(\theta)}}_{(\theta-\frac{1}{2})^{\frac{1}{2}}} \theta^{\theta} \underbrace{(\frac{x^2}{2}+\theta)^{-(\theta+\frac{1}{2})}}_{(\frac{\sqrt{2}}{x}+\ldots)^{2\theta+1}} \\
&= \frac{1}{\sqrt{2\pi}} (\theta-\frac{1}{2})^{\frac{1}{2}} \theta^{\theta} 2^{\theta+\frac{1}{2}} x^{-(2\theta+1)} \\
p_X(x) &\propto |x|^{-(2\theta+1)} \qquad \text{for } x \to \infty
\end{aligned}
$$

The approximation for $(\frac{x^2}{2}+\theta)^{-(\theta+\frac{1}{2})}$ is generated by calculating taking the first order approximation from the expansion of $(\frac{x^2}{2}+\theta)^{-\frac{1}{2}}$ as $x \to \infty$ using the Laurent series and then raising to the power of $2\theta+1$.

To get a more accurate approximation, we can use the more accurate approximation for $\frac{\Gamma(\theta+\frac{1}{2})}{\Gamma(\theta)}$ as $\left((\theta-\frac{1}{2})^2+\frac{\theta-\frac{1}{2}}{2}+\frac{1}{8}\right)^{\frac{1}{4}}$ instead of $(\theta-\frac{1}{2})^{\frac{1}{2}}$.

By considering the approximations of $p_X(x)$, we plotted this at https://www.desmos.com/calculator/sy8hldhrsj to verify that the proportionality holds true.

6 Conclusion

In this report, we covered several different topics in random variable simulation. We nalysed the terms that arose in the probability mass function of the multinomial distribution and applied the theory to consider the effects of generating histograms by sampling. Our main analysis consisted of simulating different random variables. We investigated functions of a random variable and the properties of the new density function as well as considering the convergence of Monte-Carlo sampling of a distribution from its inverse CDF. Then we considered a more advanced aspect of random variate simulation where we generated and studied the T distribution using a scaled mixture of Gaussians. Overall this experiment consisted of insightful mathematical and modelling techniques that are used to simulate and analyse the behaviour of random variables.

7 Appendix

Full Code can be viewed here: https://gitlab.com/ams289/3F3-code/