Random variables and random number generation

3F3 laboratory experiment
Theo A. Brown
Selwyn College, University of Cambridge

December 6, 2021

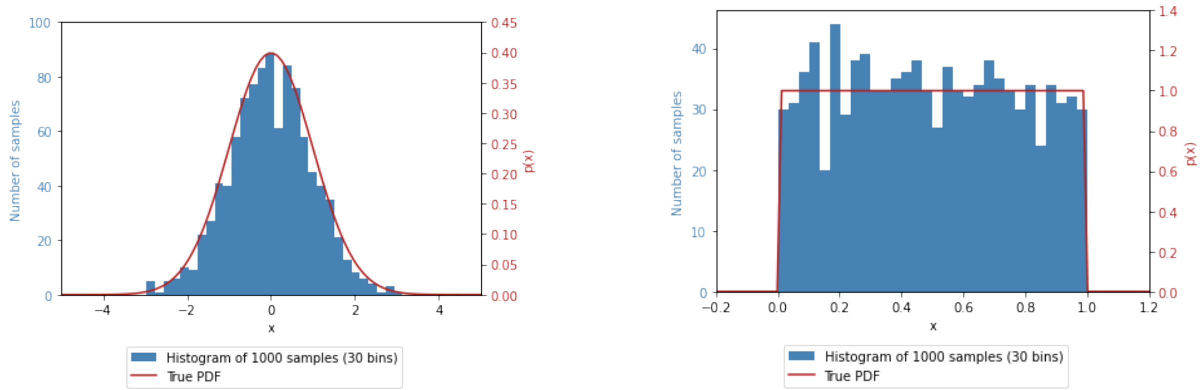
Contents

1 Generating random numbers from the Uniform and Normal distributions

A vector of 10000 samples from a unit Gaussian distribution was generated using `np.random.randn`, and a vector of 10000 samples from a unit Uniform distribution was generated using `np.random.rand`. These sample vectors are used throughout the experiment to investigate different methods of random number generation.

1.1 Comparison of histogram with true probability density function

The sample vectors were binned and plotted as histograms. The true analytical probability density function (PDF) was calculated for the unit Gaussian and unit Uniform distributions, and was overlaid on the histogram plots (Figure 1). It can be seen in Figure 1 that the histograms closely follow the shape of the analytical PDF, showing that the generation of the random numbers for these distributions is accurate.



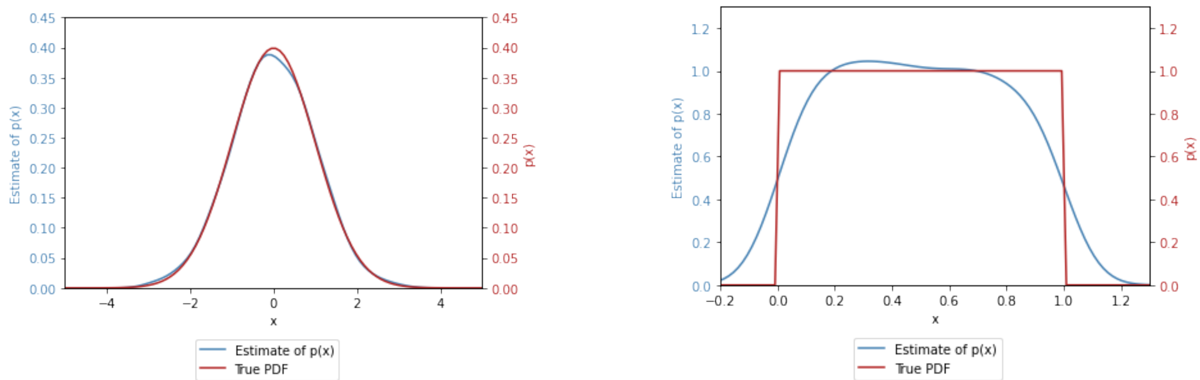
(a) Gaussian distribution

(b) Uniform distribution

Figure 1: Histogram of samples drawn from a distribution, overlaid with the true PDF of the distribution

1.2 Kernel density smoothing

A unit Gaussian kernel $\mathcal{N}(0, 1)$ is applied to the data to provide a continuous estimate for the PDF. The result is plotted for the Gaussian and Uniform distributions in Figure 2.



(a) Gaussian distribution

(b) Uniform distribution

Figure 2: Estimate of the PDF of a distribution, generated using Gaussian kernel smoothing of random samples from the distribution, overlaid with the true PDF of the distribution

At each point, the kernel method takes the mean of N neighbouring values, weighted using a Gaussian distribution. This is advantageous as it can smooth random irregularities in the samples: for example, there may be histogram bins with a sample count that is greater than the expected sample count, which would give an incorrectly high estimate of the probability of a sample falling in this bin; the kernel smooths out this local peak by taking neighbouring values into account. However, this also means that the kernel smoothing method becomes inaccurate when there is a sudden change or discontinuity in the probability density function. In the Uniform distribution there is zero probability for values outside the range $0 < x < 1$, but as the kernel averages over a window of values it does not capture the step change at $x = 0$ and $x = 1$ and instead decays smoothly.

1.3 Multinomial distribution theory

1.3.1 Derivation of the multinomial distribution

For N samples and M bins, let N_i be the random variable representing the number of samples in bin i and p_i be the probability that a sample falls in bin i. Samples are drawn independently, so the probability of getting n_i samples in bin i is given by:

$$Pr(N_i = n_i) = \prod_{j=1}^{n_i} p_i = p_i^{n_i}$$

The joint probability of obtaining a given distribution of N samples across M bins is given by:

$$Pr(N_1 = n_1, N_2 = n_2, \ldots, N_M = n_M) = \prod_{i=1}^{M} \binom{N_{\text{available},i}}{n_i} p_i^{n_i} \tag{1}$$

where $N_{\text{available},i} = N - \sum_{j=1}^{i-1} n_j$.

The binomial coefficient in Equation 1 comes from considering all possible combinations of getting n_i samples in bin i:

- For the first bin, n_1 samples are chosen from a set of N
- For the second bin, n_2 samples are chosen from a set of $N - n_1$
- For the ith bin, n_i samples are chosen from a set of $N - n_1 \cdots - n_{i-1}$

Simplifying the coefficient in Equation 1 gives:

$$\begin{aligned}
\binom{N}{n_1}\binom{N-n_1}{n_2}\cdots\binom{n_M}{n_M} &= \frac{N!}{n_1!(N-n_1)!}\frac{(N-n_1)!}{n_2!(N-n_1-n_2)!}\cdots\frac{n_M!}{n_M!} \\
&= \frac{N!}{n_1!n_2!\ldots n_M!}
\end{aligned}$$

The last step above is obtained by diagonal cancellation.

Consequently, Equation 1 simplifies to:

$$Pr(N_1 = n_1, N_2 = n_2, \ldots, N_M = n_M) = \frac{N!}{n_1!n_2!\ldots n_M!} p_1^{n_1} p_2^{n_2} \ldots p_M^{n_M}$$

From considering the probability $p_i = Pr(N_i = n_i)$ for N samples, the mean and standard deviation of N_i can be found:

$$\mu_i = \mathbb{E}[N_i] = Np_i \tag{2}$$

$$\sigma_i = \sqrt{Var[N_i]} = \sqrt{Np_i(1-p_i)} \tag{3}$$

1.3.2 Application to the Uniform distribution

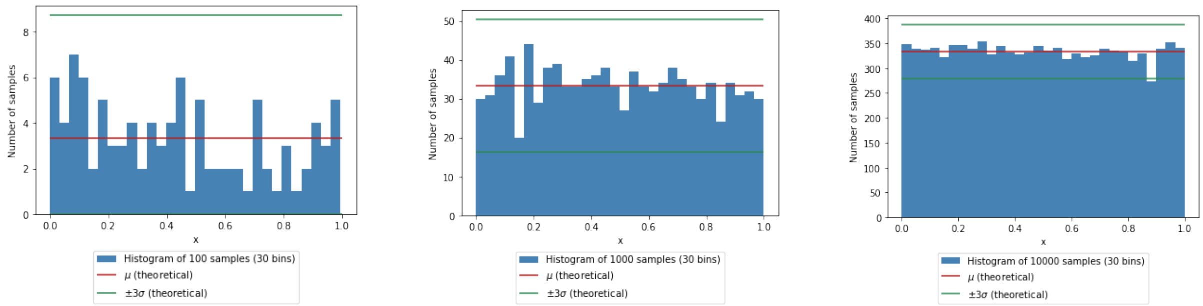
For the Uniform distribution between 0 and 1, the pdf $p(x)$ is defined as:

$$\begin{aligned}
p(x) &= \frac{1}{x_{max} - x_{min}} \\
&= 1
\end{aligned}$$

Hence, using Equation 2 and Equation 3, the mean and standard deviation of the number of samples from the Uniform distribution in a histogram with bin width δ and bin centers c_i:

$$\begin{aligned}
p_i &= \int_{c_i - \frac{\delta}{2}}^{c_i + \frac{\delta}{2}} 1 \, dx \\
&= \delta \\
\mu_i &= N\delta \\
\sigma_i &= \sqrt{N\delta(1-\delta)}
\end{aligned}$$

Figure 3 shows that the sample count lies within the 3σ interval for almost all bins, which is consistent with the multinomial distribution theory.



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

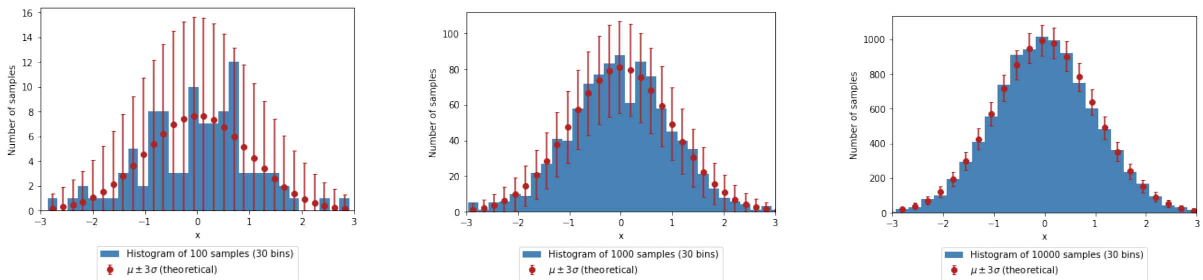
Figure 3: Histogram of N samples from a Uniform distribution, showing mean and ±3 standard deviation

1.3.3 Application to the Gaussian distribution

For the Gaussian distribution:

$$
\begin{aligned}
p_i &= \int_{c_i - \frac{\delta}{2}}^{c_i + \frac{\delta}{2}} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}x^2\right) dx \\
&= \Phi\left(c_i + \frac{\delta}{2}\right) - \Phi\left(c_i - \frac{\delta}{2}\right)
\end{aligned}
$$

The mean and variance are calculated according to Equation 2 and Equation 2.



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

Figure 4: Histogram of N samples from a Gaussian distribution, showing mean and ±3 standard deviation

Figure 4 shows that the sample count lies within the 3σ interval for almost all bins, which is consistent with the multinomial distribution theory.

1.3.4 Effect of increasing N

Multinomial distribution theory suggests that the method of using a histogram of samples to estimate properties of a distribution increases in accuracy as the number of samples N increases. For a given bin probability p_j, the standard deviation increases less rapidly than the mean with increasing N ($\sigma \propto \sqrt{N}$ whereas $\mu \propto N$). Consequently an increase in N will result in the standard deviation becoming smaller compared to the mean ($\frac{\sigma}{\mu} \propto \frac{1}{\sqrt{N}}$). This can be seen in Figure 3 and Figure 4, where the standard deviation reduces as N increases.

1.3.5 Effect of bin probabilities

Figure 4 shows that the variance in the number of samples N_i in bin i increases as p_i increases. The relationship between $Var[N_i]$ and p_i is clear from Equation 3. Figure 5 is a plot of the relationship. For the Gaussian distribution with 30 bins, the bin probabilities remain small, so the variance of N_i increases with increasing bin probability. For another distribution or a histogram with some bin probabilities greater than 0.5, it would be observed that the variance decreases with increasing bin probability.

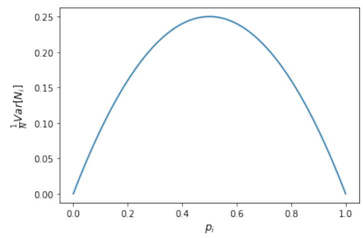


Figure 5: Plot of normalised variance against bin probability

2 Functions of random variables

2.1 $f = ax + b$

For $x \sim \mathcal{N}(0, 1)$, let $y = f(x) = ax + b$. Then:

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

Comparing with the equation for a general Gaussian distribution, it is clear that $p_Y(y)$ is a Gaussian distribution with mean b and variance a^2. The result is compared with a histogram of transformed samples in Figure 6. The close agreement suggests this hypothesis is correct.

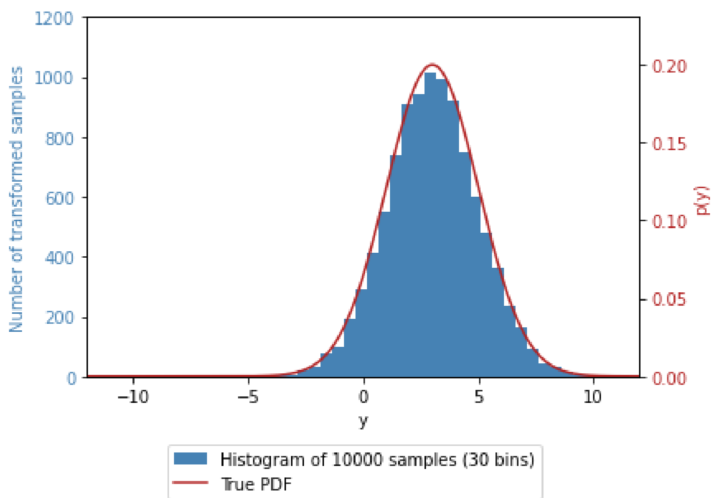


Figure 6: Histogram of linear function ($f(x^{(i)}) = 2x^{(i)} + 3$) of Gaussian samples overlaid with calculated PDF

2.2 $f = x^2$

For $x \sim \mathcal{N}(0,1)$, let $y = f(x) = x^2$. Then:

$$\begin{aligned}
f^{-1}(y) &= \pm\sqrt{y} \\
\left|\frac{dy}{dx}\right| &= |2x| \\
p_Y(y) &= \sum_{i=1}^{2} \frac{1}{|2\sqrt{y}|} p_X\left(\sqrt{y}\right) \\
&= \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{1}{2}y\right)
\end{aligned}$$

The result is compared with a histogram of transformed samples in Figure 7. The close agreement suggests this hypothesis is correct.

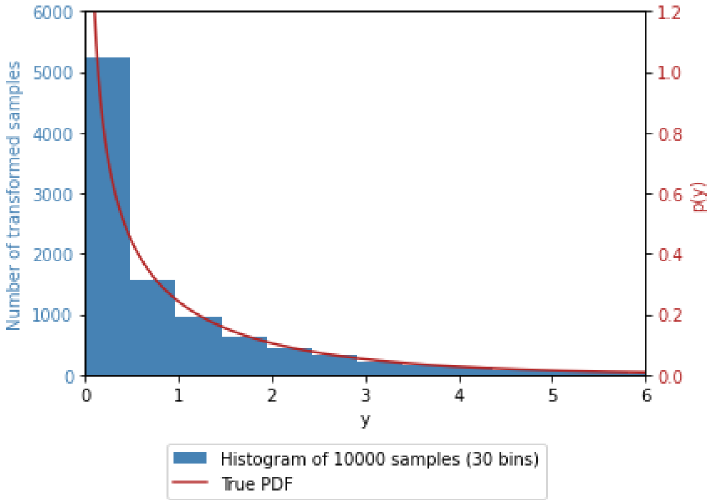


Figure 7: Histogram of quadratic function $\left(f(x^{(i)}) = \left(x^{(i)}\right)^2\right)$ of Gaussian samples overlaid with calculated PDF

2.3 $f = \sin(x)$

For $x \sim \mathcal{U}(0, 2\pi)$, let $y = f(x) = \sin(x)$. Then:

$$\begin{aligned}
p(x) &= \frac{1}{2\pi} \\
f^{-1}(y) &= \sin^{-1}(y) \\
\left|\frac{dy}{dx}\right| &= |\cos(x)| \\
p_Y(y) &= \sum_{i=1}^{2} \frac{\frac{1}{2\pi}}{\left|\cos\left(\sin^{-1}(y)\right)\right|}
\end{aligned}$$

By drawing a right-angled triangle with hypotenuse of length 1 and a side of length y, and applying trigonometry and Pythagoras' theorem, it can be shown that:

$$\cos\left(\sin^{-1}(y)\right) = \sqrt{(1-y^2)}$$

Hence:

$$p_Y(y) = \frac{1}{\pi\sqrt{(1-y^2)}}$$

The result is compared with a histogram of transformed samples in Figure 8. The close agreement suggests this hypothesis is correct.

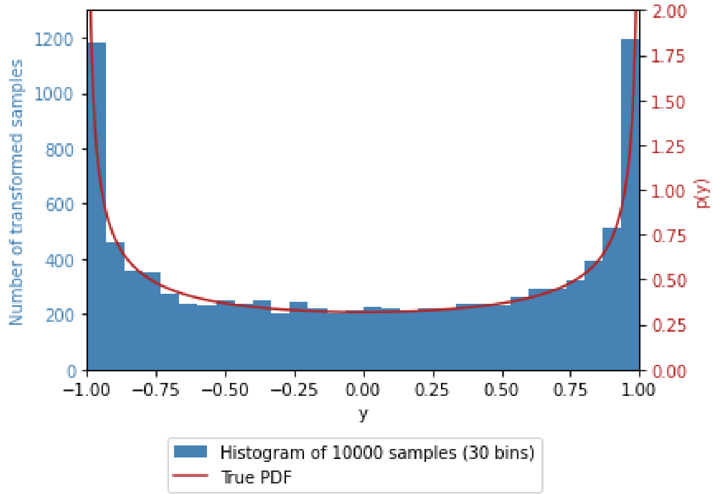


Figure 8: Histogram of sinusoidal function $\left(f(x^{(i)}) = \sin\left(x^{(i)}\right)\right)$ of Uniform samples overlaid with calculated PDF

2.4 $f = \mathbf{clip}(\sin(x), 0.7)$

For $x \sim \mathcal{U}(0, 2\pi)$, let:

$$y = f(x) = \begin{cases} -0.7 & \sin(x) < -0.7 \\ \sin(x) & |\sin(x)| < 0.7 \\ 0.7 & \sin(x) > 0.7 \end{cases}$$

Note that this function differs from the one in the handout by being symmetric, which is a better model of realistic signal clipping. This produces a probability density with the following properties:

- $|y|$ will never be exceed 0.7. Hence $p_Y(y)$ will be zero for $|y| > 0.7$.
- $f(x)$ collapses all values less than -0.7 to -0.7, and all values greater than 0.7 to 0.7. Hence $p_Y(y)$ will have delta functions at $y = 0.7$ and $y = -0.7$, containing all of the probability mass from $y > 0.7$ and $y < -0.7$ respectively.
- For $|y| < 0.7$, $p_Y(y) = \frac{1}{\pi\sqrt{(1-y^2)}}$ as before.

The area A of each delta function can be calculated using the fact that the PDF must integrate to 1:

$$\begin{aligned} A &= \frac{1}{2}\left(1 - \int_{-0.7}^{0.7} \frac{1}{\pi\sqrt{1-y^2}}\, dy\right) \\ &= \frac{1}{2} - \frac{1}{\pi}\sin^{-1}(0.7) \end{aligned}$$

Which gives the PDF as:

$$p_Y(y) = \begin{cases} A\delta(y+0.7) & y = -0.7 \\ \frac{1}{\pi\sqrt{(1-y^2)}} & |y| < 0.7 \\ A\delta(y-0.7) & y = 0.7 \\ 0 & \text{otherwise} \end{cases}$$

The presence of the delta function in the PDF can also be explained using the Jacobian formula. Consider the derivative of y:

$$\frac{dy}{dx} = \begin{cases} \cos(x) & 0 < x < \sin^{-1}(0.7) \\ 0 & \sin^{-1}(0.7) < x < \pi - \sin^{-1}(0.7) \\ \cos(x) & \pi - \sin^{-1}(0.7) < x < \pi + \sin^{-1}(0.7) \\ 0 & \pi + \sin^{-1}(0.7) < x < 2\pi - \sin^{-1}(0.7) \end{cases}$$

As the Jacobian formula involves dividing by $\frac{dy}{dx}$, $p_Y(y)$ is undefined for the range $|y| > 0.7$.

The calculated PDF is compared with a histogram of transformed samples in Figure 8. The close agreement suggests this hypothesis is correct.

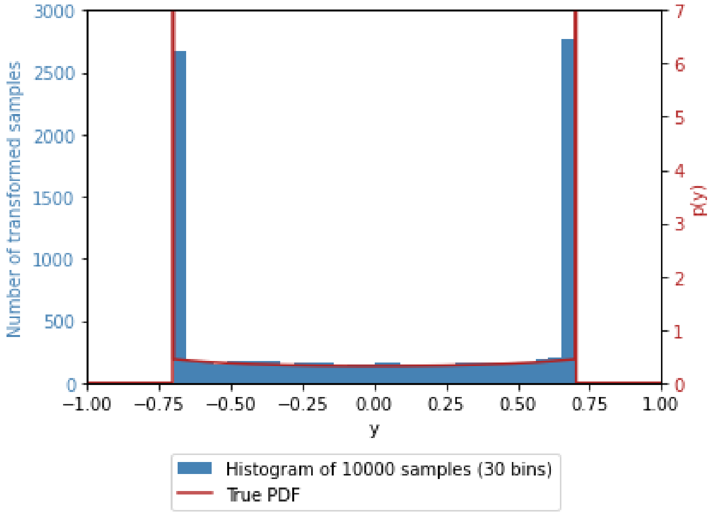


Figure 9: Histogram of limited sinusoidal function $\left(f(x^{(i)}) = \text{clip}\left(\sin\left(x^{(i)}\right), 0.7\right)\right)$ of Uniform samples overlaid with calculated PDF

3 Generating random variables using the inverse CDF method

3.1 Exponential distribution

For the exponential distribution with mean one:

$$\begin{aligned}
&\text{PDF: } p(y) = \exp(-y), \;\; y \geq 0 \\
&\text{CDF: } F(y) = \int_0^y p(y')dy' = 1 - \exp(-y) \\
&\text{iCDF: } F^{-1}(x) = -\ln(1-x)
\end{aligned}$$

Figure 10 shows that samples from the exponential distribution generated using the iCDF follow the true PDF.

3.2 Estimating properties of the exponential distribution

Properties of the distribution can be estimated using Monte Carlo methods:

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

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

3.2.1 Properties of Monte Carlo mean estimator

The Monte Carlo mean estimate $\hat{\mu}$ can be shown to unbiased:

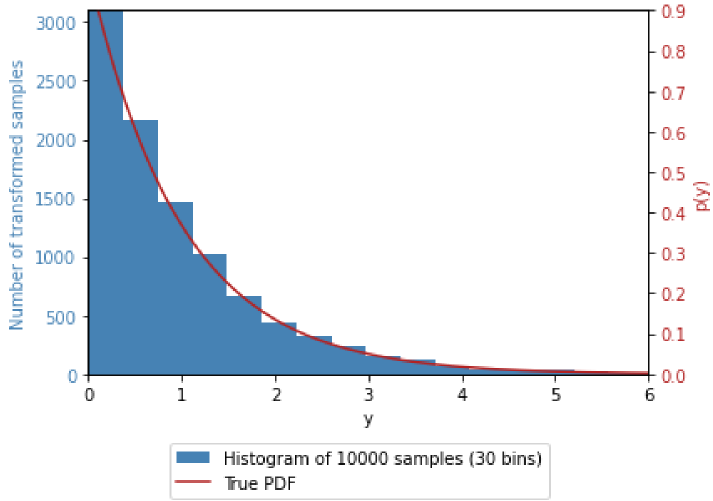


Figure 10: Histogram of samples drawn from the uniform distribution and transformed using the iCDF method to follow the exponential distribution, overlaid with the true exponential PDF

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

To derive the variance of the Monte Carlo mean estimator, consider the variance of the sum of random variables X and Y:

$$
\begin{aligned}
Var[X+Y] &= \mathbb{E}\left[(X+Y-\mu_X-\mu_Y)^2\right] \\
&= \mathbb{E}\left[((X-\mu_X)+(Y-\mu_Y))^2\right] \\
&= \mathbb{E}\left[(X-\mu_X)^2\right]] + \mathbb{E}\left[(Y-\mu_Y)^2\right] + 2\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]
\end{aligned}
$$

Using the fact that X and Y are independent gives $\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] = 0$, hence:

$$Var[X+Y] = Var[X] + Var[Y]$$

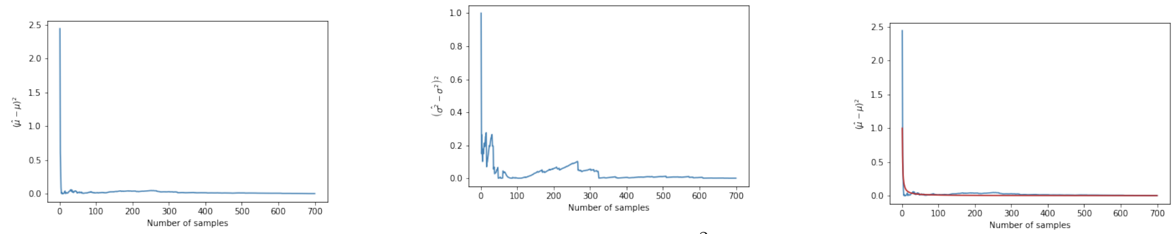
When multiplying a random variable by a scalar, the variance is multiplied by the square of the scalar:

$$
\begin{aligned}
Var[kX] &= \mathbb{E}\left[(kX)^2\right] - \mathbb{E}[kX]^2 \\
&= k^2\mathbb{E}\left[X^2\right] - k^2\mathbb{E}[X]^2 \\
&= k^2 Var[kX]
\end{aligned}
$$

Treating the samples $y^{(i)}$ as independent random variables and combining the above properties gives:

$$Var\left[\frac{1}{N}\sum_{i=1}^{N} y^{(i)}\right] = \frac{1}{N^2}\sum_{i=1}^{N} Var\left[y^{(i)}\right] = \frac{\sigma}{N}$$

The convergence of the Monte Carlo estimators for the mean and variance is shown in Figure 11. It is apparent that the mean estimator does indeed converge proportional to $\frac{1}{N}$ (see Figure 11c).



(a) $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y^{(i)}$ (b) $\hat{\sigma^2} = \frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)}\right)^2 - \hat{\mu}^2$ (c) $\hat{\mu}$ with $\frac{1}{N}$ overlay

Figure 11: Behaviour of Monte Carlo estimators for increasing number of samples

3.2.2 Application to the exponential distribution

The exponential distribution $p(y) = \exp(-y),\ y \geq 0$ has $\mu = 1$ and $\sigma^2 = 1$. Applying Monte Carlo estimation to 10000 samples gives $\hat{\mu} = 1.0035$ and $\hat{\sigma^2} = 1.0128$, which are close to the true values.

4 Generating random variables using a scaled mixture of Gaussians

4.1 Using the Exponential distribution

Using the iCDF method, we can generate samples that are distributed according to:

$$p(u) = \frac{\alpha^2}{2}\exp\left(-\frac{\alpha^2}{2}u\right)$$

This distribution has:

$$\text{CDF: } F(u) = \int_0^u p(u')du' = 1 - \exp\left(-\frac{\alpha^2}{2}u\right)$$
$$\text{iCDF: } F^{-1}(v) = -\frac{2}{\alpha^2}\ln(1-v)$$

Samples $u^{(i)}$ can therefore be generated from uniformly distributed samples $v^{(i)}$. The samples $u^{(i)}$ are then used to set the variance of the random variable X:

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

The limit of the integral is from zero to infinity because the standard deviation must be positive. A histogram of samples of $x^{(i)}$ is plotted in Figure 12 for different values of α. It appears that α sets the standard deviation of the distribution.

The kernel method was applied to the data, and the smoothed density was plotted on linear and log axes at different scales (Figure 13). From the shape of the distribution seems to be a sharper version of a Gaussian distribution, with significantly higher probability density in the centre and tails that last for longer. Closer examination of the logarithmic plots suggests that the distribution behaves more like a Gaussian close to $x = 0$, where the logarithmic graph resembles an inverted parabola, and more like an exponential further from $x = 0$, where the logarithmic graph becomes linear.

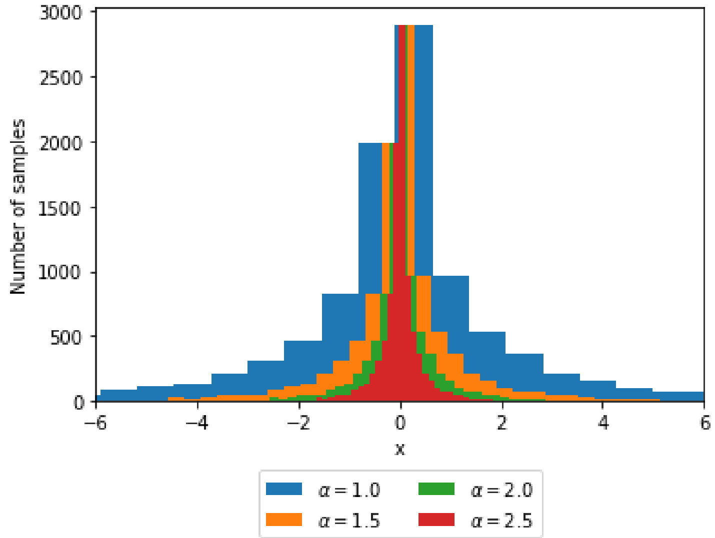
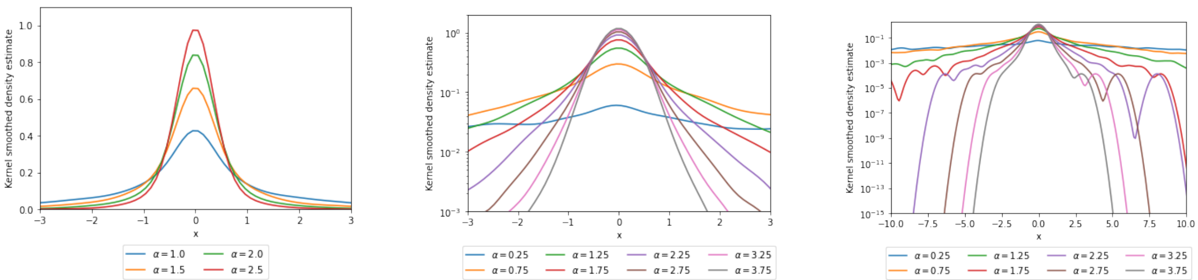


Figure 12: 100-bin histogram of samples drawn from the example distribution for different values of α



(a) Linear scale

(b) $10^{-3} < \ln(p(x)) < 1$

(c) $10^{-15} < \ln(p(x)) < 1$

Figure 13: Kernel smoothed probability density estimates for the example distribution

4.2 Using the Gamma distribution

4.2.1 PDF of $U = \frac{1}{V}$

The probability density function of the random variable U can be calculated by applying the Jacobian method:

$$p(v) = \mathcal{G}(v|\theta, \theta^{-1}) = \frac{1}{\theta^{-\theta}\Gamma(\theta)} v^{\theta-1} e^{-v\theta}$$

$$\left|\frac{du}{dv}\right| = \left|-v^{-2}\right| = v^{-2}$$

$$f^{-1}(u) = u^{-1}$$

$$\begin{aligned} p(u) &= \left.\frac{p(v)}{v^{-2}}\right|_{v=u^{-1}} \\ &= \frac{\theta^{\theta}}{u^2\Gamma(\theta)} u^{1-\theta} e^{-\frac{\theta}{u}} \end{aligned}$$

$$p(u) = \frac{\theta^{\theta}}{\Gamma(\theta)} u^{-1-\theta} e^{-\frac{\theta}{u}} \tag{4}$$

4.2.2 PDF of $X \sim \mathcal{N}(x|0, u)$

The PDF of $X \sim \mathcal{N}(x|0, u)$, where $u \sim p(u)$ from Equation 4 can be obtained by marginalisation:

$$\begin{aligned}
p(x) &= \int_0^\infty \mathcal{N}(x|0, u)p(u)du \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} u^{-\frac{1}{2}} e^{-\frac{1}{2}x^2 u^{-1}} \frac{\theta^\theta}{\Gamma(\theta)} u^{-1-\theta} e^{-\theta u^{-1}} du \\
&= \frac{\theta^\theta}{\sqrt{2\pi}\Gamma(\theta)} \int_0^\infty \left(u^{-1}\right)^{\frac{3}{2}+\theta} e^{-\left(\theta+\frac{1}{2}x^2\right)u^{-1}} du
\end{aligned}$$

Setting $u = v^{-1}$, and using the fact that $\frac{du}{dv} = -v^{-2}$:

$$p(x) = -\frac{\theta^\theta}{\sqrt{2\pi}\Gamma(\theta)} \int_{-\infty}^0 v^{\frac{3}{2}+\theta} e^{-\left(\theta+\frac{1}{2}x^2\right)v} v^{-2} dv$$

$$p(x) = \frac{\theta^\theta}{\sqrt{2\pi}\Gamma(\theta)} \int_0^\infty v^{-\frac{1}{2}+\theta} e^{-\left(\theta+\frac{1}{2}x^2\right)v} dv \tag{5}$$

From the definition of the Gamma distribution, we know that:

$$\int_0^\infty v^{a-1} e^{-\frac{v}{b}} dv = b^a \Gamma(a)$$

Which is the integral from Equation 5 with

$$\begin{aligned}
a &= \theta + \frac{1}{2} \\
b &= \left(\theta + \frac{1}{2}x^2\right)^{-1}
\end{aligned}$$

Hence Equation 5 solves to become:

$$p(x) = \frac{\theta^\theta \Gamma\left(\theta + \frac{1}{2}\right)}{\sqrt{2\pi}\Gamma(\theta)} \left(\theta + \frac{1}{2}x^2\right)^{-\theta-\frac{1}{2}} \tag{6}$$