
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 (bandwidth 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.

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

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

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

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

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

Explain behaviour as N becomes large:

From above,

$$\text{overall mean} = \frac{N}{j} \propto N$$

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

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

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 ??, 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 ?? 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:

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

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

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

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

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:

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

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

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

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

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:

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

$$\begin{aligned}\text{CDF: By definition of CDF, } F(y) &= \int_0^y p(y) dy \\ &= \int_0^y \exp(-y) dy \\ &= 1 - \exp(-y)\end{aligned}$$

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

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

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

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)(b)(c)(d)(e)
 $\alpha = 0.5, \beta = -1, -0.5, 0, 0.5, 1$

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

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

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.