Assignment 1 Report

Arpon Basu
Shashwat Garg

Autumn 2021

Contents

1 Introduction

2 Problem 1

2.1 What Theory Says

In this question we have been asked to plot the PDF and CDF of several different functions, along with calculating the variance whenever possible. The functions are-

- Laplace Distribution

  As per wikipedia, the Laplace distribution is the **distribution of the difference of two independent random variables with identical exponential distributions**.
  Also called the double exponential distribution, this distribution is often used to **model phenomena with heavy tails or when data has a higher peak than the normal distribution.**

- Gumbel Distribution

  The Gumbel distribution is used to model the distribution of the maximum (or the minimum) of a number of samples of various distributions. For modelling the minimum value, we use the negative of the original values.
  The PDF of the Gumbel distribution also has an almost constant shape and shifts depending on the location parameter.

- Cauchy Distribution

  Cauchy Distribution is a very interesting example of Statistical Distributions, It is an example of a "pathological" distribution since both its expected value and its variance are undefined.

2.2 Code Flow

The code does 3 things,

We first define some generic variables, specifying the location and scaling parameters for the various distributions.

Then we define a generic pdf_plot function which given a function and some parameters, plots the function.

This is followed by defining the three distributions in terms of their pdf. Then we repeatedly pass the functions along with the parameters into the plotting functions to obtain the required graphs.

Next, we define a generic function for plotting the CDF of a distribution which uses PDF alongwith calculating Riemann sums for plotting the CDF graph. We have used the ranges from $\mu - 100b$ to $\mu + 100b$, to define $\pm\infty$.

Finally for calculating variance, we have again used Riemann sum method on the random variable defined by $(x - \mu)^2$.

This finishes this question's code of the assignment.

2.3 Any derivations/explanations (If asked for)

None

2.4 Observations and their Rationalization

We observed how increasing the number of sample points makes the graph smoother. By adjusting various values of the location and scaling parameters, we observed and understood the behaviour of parameters on the functions.

We also learnt basic implementation strategies to plot different functions and their PDF, CDF and other corresponding values easily

3 Problem 2

3.1 What Theory Says

This problem deals with operating on the Poisson distribution variable and understanding the behaviour of sum of random variables and Poisson thinning.

The process of summing the random variables, say $Z = X + Y$ is carried out analytically by combining the different possibilities in which the combination of x and y could lead to the required value of z.

Implementing this in code is simpler since we can just take a large amount of samples and then combine the outcomes of the corresponding individual outcomes, to get the overall random variable. No need of the double Σ.

Poisson thinning is basically a filtering process.Suppose Y is the thinned R.V. obtained from X, with probability p. So we require to keep each value of Y as a value from the Binomial Distribution with parameters as values taken by X and the probability factor of the binomial distribution as p.

3.2 Code Flow

The code begins with specifying the parameters for the poisson random variables, X and Y.

We then obtain a distribution of $Z = \ X + Y$ by adding the distributions of the random variables X and Y.

We then construct a plot of the obtained (experimental) distribution of Z and compare it with the actual distribution (analytical) of Z, ie. the poisson R.V. with λ_z as $\lambda_x + \lambda_y$

Both plots seem to match near perfectly.

Next, we implement the Poisson thinning process. We begin with specifying the original Poisson random variable Y and obtain a distribution containing 100,000 instances of the same.

We then use the binomial distribution with probability p on Y to obtain the distribution of the variable Z.

Comparing the obtained values with the actual distribution of Z which is actually the Poisson distribution with $\lambda_z = \lambda_y * p$, gives a near perfect match, confirming the correctness of the procedure.

3.3 Any derivations/explanations (If asked for)

3.4 Observations and their Rationalization

We observe how Poisson thinning starts with applying the binomial distribution on the random variable Y to obtain Z, which finally is a Poisson distribution only. Apart from the mathematical proof of the same, this behaviour can be understood by realising that the λ parameter of the Poisson distribtuion is actually the average number of hits per unit time, so to speak. Thus, is we allow only p fraction to come through the effective distribution would stay the same, just get scaled down in magnitude of values by p.

4 Problem 3

4.1 Introduction

This problem deals with the classical topic of random walks (for our case, one-dimensional), the importance of whom can't be underestimated. So we'll analyse random walks to answer all the questions posed in the assignment. Also, a significant portion of the assignment involves proving results associated with the Law of Large Numbers. Thus, that too will be presented in a different section. Finally, **we have carried out an extra computation on our own, ie:- to theoretically calculate and then show the envelope of the random walk plots that are obtained**. Since the calculations are a bit involved, they too are put in a different subsection. Finally, we present all our results in the results subsection.

4.2 Random Walk Analysis

So basically, what is a random walk in one dimensions? A person starts at the origin, and and for each unit interval of time δt, takes either a unit step (δz) forwards, or backwards, with probabilities p and $q = (1 - p)$ respectively. In the context of this problem, we shall assume that $p = \frac{1}{2}$.
Even without any mathematical justification, from the above formulation itself it's intuitively clear that over a large number of steps, the person will, on an average, take as many steps forward as backward, and thus the **expected** position of the person after a long time t ($= n\delta t$, where n is the number of steps taken) will be at $z = 0$.
Now, suppose we want to mathematically analyze the probability distribution of a random walk. How do we do so? We first observe that if x is the number of forward steps taken, then:
$z = x\delta z - (n - x)\delta z = (2x - n)\delta z$
Also, the probability that exactly x forward steps will be taken can be modeled via the

binomial distribution, ie:-
$P(x) = \binom{n}{x} p^x q^{n-x}$
Considering n, x to be much much greater than 1 (so as to allow us to apply the Stirling's approximation, as in $n! \sim \sqrt{2\pi n}(n/e)^n \mathcal{O}(1+\frac{1}{n})$), and after some mathematical jugglery, we obtain:
$P(x) = \frac{1}{(2\pi npq)^{1/2}} e^{-\frac{(x-np)^2}{2npq}}$
Applying the transformations $t = n\delta t$, $z = (2x-n)\delta z$, $P(z;t)\cdot 2\delta z = P(x)$, and most importantly, defining the **diffusion coefficient** D as $\frac{(\delta z)^2}{2\delta t}$, we obtain
$P(z;t) = \frac{1}{(4\pi Dt)^{1/2}} e^{-\frac{z^2}{4Dt}}$
Thus, as asked for in the report,
The true mean of the terminal positions for a random walk with probability $p = \frac{1}{2}$ is zero.
The true variance of the terminal positions for a random walk with probability $p = \frac{1}{2}$ is $2Dt = 2\frac{(\delta z)^2}{2\delta t}(n\delta t) = n(\delta z)^2 = 0.001$, since in our case $n = 10^3$ steps and $\delta z = 0.001$.

4.3 Calculations involving averages of Random Variables, the Law of Large Numbers

Coming to the next question raised by the assignment, which digresses a bit and asks us to prove that the random variable $\hat{M} := \frac{1}{n}\sum_{j=1}^{n} X_j$ converges to it's mean $M := \mathbb{E}[X]$ (ie:- becomes a one point random variable or a Dirac delta), where X_1, X_2, ..., X_n are independent draws from a random variable X. Note that $\mathbb{E}[\hat{M}] = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[X_j] = \frac{1}{n}\sum_{j=1}^{n} M = M$.
Thus, we just need to prove that as $n \longrightarrow \infty$, **the probability that $\hat{M}$ takes any other value than M tends to zero**.
For proving this, note that by Chebyshev's inequality,
$P(|\hat{M} - M| \geq \epsilon) \leq \frac{\sigma_n^2}{\epsilon}$,
where σ_n^2 is the variance of $\hat{M}$. But then also note that since X_1, X_2, ..., X_n are all mutually independent (and hence the variance of their sum is the sum of their variances),
$\text{Var}(\frac{1}{n}\sum_{j=1}^{n} X_j) = \frac{1}{n^2}\sum_{j=1}^{n}\text{Var}(X_j) = \frac{\sigma^2}{n} \implies \sigma_n = \frac{\sigma}{\sqrt{n}}$
Thus,
$P(|\hat{M} - M| \geq \epsilon) \leq \frac{\sigma_n^2}{\epsilon} = \frac{\sigma^2}{n\epsilon} \longrightarrow 0$ as $n \longrightarrow \infty$; $\forall \epsilon > 0$, as desired (Note that this effectively was a proof of the **weak version of the Law of Large Numbers**. See Problem 5 for more details).
For the random variable $\hat{V} := \frac{1}{n}\sum_{j=1}^{n}(X_j - M)^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mathbb{E}[X])^2$, we have, after expanding the square
$\hat{V} = \frac{1}{n}\sum_{j=1}^{n}(X_j - M)^2 = M^2 + \frac{1}{n}\sum_{j=1}^{n} X_j^2 - \frac{2M}{n}\sum_{j=1}^{n} X_j$
As $n \longrightarrow \infty$, $\frac{2M}{n}\sum_{j=1}^{n} X_j \longrightarrow 2M^2$, borrowing from the above proof.

Also, in the above proof (for $\hat{M}$), let X^2 be our random distribution instead of X. Then we also have
$\frac{1}{n}\sum_{j=1}^{n} X_j^2 \longrightarrow \mathbb{E}[X^2]$ as $n \longrightarrow \infty$
Finally, to finish the proof, consider the following argument : If $Q_{1,n}$ and $Q_{2,n}$ are 2 random variables which tend to their means as $n \longrightarrow \infty$ (ie:- become Dirac Deltas), then $P(Q_{1,n}+Q_{2,n} \not\longrightarrow \mathbb{E}[Q_{1,n}]+\mathbb{E}[Q_{2,n}]; n \longrightarrow \infty) \leq P(Q_{1,n} \not\longrightarrow \mathbb{E}[Q_{1,n}]; n \longrightarrow \infty) + P(Q_{2,n} \not\longrightarrow \mathbb{E}[Q_{2,n}]; n \longrightarrow \infty) \longrightarrow 0$ as both $P(Q_{1,n} \not\longrightarrow \mathbb{E}[Q_{1,n}]; n \longrightarrow \infty)$ and $P(Q_{2,n} \not\longrightarrow \mathbb{E}[Q_{2,n}]; n \longrightarrow \infty) \longrightarrow 0$.
Hence $\hat{V}$ equals $M^2 + \mathbb{E}[X^2] - 2M^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 =Var(X)$, for large enough n, with probability 1.

4.4 Envelope Calculation of the Random Walk plots

Having answered all the questions in the problem, we now return to random walks for some concluding remarks.
As shown by the formula of $P(z;t)$, for a fixed time t (ie:- fixed n), we have that z, **which is the terminal position of the walk**, will have a Gaussian distribution for large enough n. Also, even though we saw that the expected **position** of a random walker is zero, the expected **translation**, which is the absolute value of position, ie:- $|z|$, is not zero. In fact, the PDF of z is what is known as the **folded normal distribution** in literature (click here to obtain reference). Directly using the result given in the source, we get
$\mathbb{E}[|Z|] = \sigma\sqrt{\frac{2}{\pi}}e^{-\frac{\mu^2}{2\sigma^2}} + \mu(1-2\Phi(-\frac{\mu}{\sigma}))$
Since $\mu = \mathbb{E}[Z] = 0$, much of the terms vanish, leaving behind
$\mathbb{E}[|Z|] = \sigma\sqrt{\frac{2}{\pi}} = \sqrt{2Dt}\sqrt{\frac{2}{\pi}}$
Putting $2Dt = n(\delta z)^2$, we get
$\mathbb{E}[|Z|] = \sqrt{\frac{2n}{\pi}}\delta z$
Also,
$\mathbb{E}[|Z|^2] = \mathbb{E}[Z^2] = \text{Var}(Z) + \mathbb{E}[Z]^2 = \text{Var}(Z) = 2Dt$
Thus $\text{Var}(|Z|) = \mathbb{E}[|Z|^2] - \mathbb{E}[|Z|]^2 = \sigma^2(1-\frac{2}{\pi})$
Thus, assuming translation varies normally (we can do so due to Lippmann's joke:)), we obtain that the **envelope of the random walk plots would be** (note that we assume that 3 times the standard deviation in a normal distribution encompasses most (99.7%) of the data)
$\text{env}(n) = \pm(\mathbb{E}[|Z|] + 3(\mathbb{E}[|Z|^2]))^{\frac{1}{2}} = \pm\sigma(\sqrt{\frac{2}{\pi}} + 3(1-\frac{2}{\pi})^{\frac{1}{2}}) \approx \pm 2.606(\delta z)\sqrt{n}$
On plotting these curves over the random walk plots, one obtains excellent encompassing by the envelopes.

4.5 Results

Based on the discussion above, we present the histogram, random walk plots and the errors between the empirical and theoretical means and variances below for the graders perusal. Wherever curve fitting (A normal distribution fit to the terminal position histogram) or enveloping (a parabolic envelope to the random walk plots) has been done, one can see the curves as thick black lines. The equations of those have not been included in the image for aesthetic reasons. The grader can check for their implementation though, in the code.

Figure 1: Problem 3 : The Random Walk Plot

5 Problem 4

5.1 What Theory Says

5.2 Code Flow

5.3 Any derivations/explanations (If asked for)

5.4 Observations and their Rationalization

6 Problem 5

6.1 Introduction

The motivation behind this problem was to actually demonstrate through the proper choice (ie:- independent and identically distributed) of random variables that the conclusions of the Law of Large Numbers holds. Thus, we'll first have a discussion regarding the various nuances of the Law of large numbers, amply supported by insights borrowed from here.
The grader may also wish to refer the contents of Problem 3 for some more discussion regarding the Law of Large numbers.

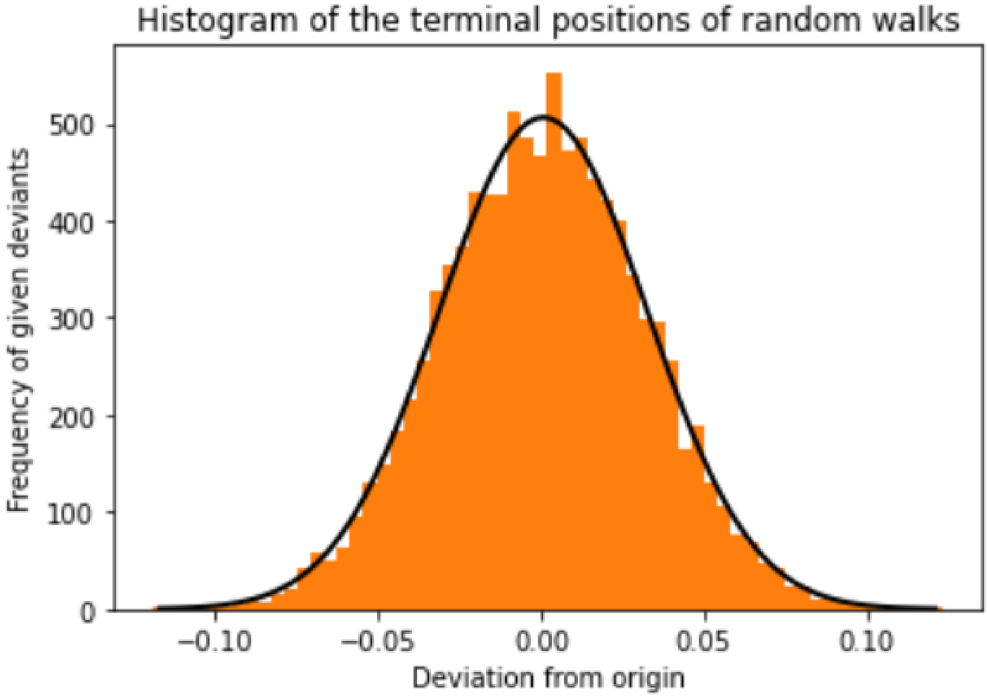

Figure 2: Problem 3 : The Terminal Position Histogram

```
The empirical mean of the last locations of the random walkers is 0.0007304000000000004
The error between true mean and empirical mean is 0.07304000000000004 %
The empirical variance of the last locations of the random walkers is 0.0009708553158400014
The error between true variance and empirical variance is -0.0029144684159998664 %
```

Figure 3: Problem 3 : Empirical Errors

6.2 A discussion regarding the Nuances of the Law of Large Numbers, and it's Implications for this problem

So, what does this law say?

Theorem 6.2.1. *Let X_1, X_2, ..., X_n be n independent random variables (of mean μ and variance σ^2) which are identically distributed. Define the empirical mean of the variables $\hat{\mu_n}$ to be $\mathbb{E}[\frac{1}{n}\sum_{j=1}^{n} X_j]$. Then,*

(**Weak Version**)

$$\lim_{n\to\infty} P(|\hat{\mu_n} - \mu| > \epsilon) = 0 \ \forall \epsilon > 0$$

(**Strong Version**)

$$P(\lim_{n\to\infty} \hat{\mu_n} = \mu) = 1$$

Thus, what the weak version says is that for any positive threshold we set, for large

enough n, there will be a very high probability that the empirical mean will be within the threshold of the theoretical mean, while what the strong version says is that for large enough n, the empirical mean $\hat{\mu_n}$, which is a function of n, converges, from the analysis point of view, to μ, ie:- $\forall \epsilon > 0\ \exists N \in \mathbb{N}$ such that $|\hat{\mu_n} - \mu| < \epsilon\ \forall n > N$.
This is in fact, also the point in which the strong law guarantees a stronger condition than the weaker one: The weak law says that $|\hat{\mu_n} - \mu| \leq \epsilon$ with **probability 1** (for large enough n), but one would also do well to remember that many non empty events have zero probabilities, especially in the case of continuous random variables (in fact, if the PDF of X is continuous, then any denumerable subset of $\mathbb{R}$ has a probability measure of zero).
Thus, in fact, the weak law leaves room for the empirical mean to overshoot any threshold a countably infinitely many times ! This is where the strong law of large numbers kicks in : It guarantees, based on the premise that X_1, X_2, ..., X_n are **identically distributed**, that, the above scenario, will never happen since for any $\epsilon > 0$ it provides us a bound N after which the empirical mean doesn't cross the threshold of the theoretical mean.
As with anything in life, the extra guarantee doesn't come without riders, the rider here being that X_1, X_2, ..., X_n are **identically distributed**. What this also tells us is that the weak version of the law, in fact, doesn't require X_1, X_2, ..., X_n to be identically distributed, **the sequence of random variables just needs to have the same mean and variance** (following which a simple application of Chebyshev's inequality **proves the weak law**).
The above insight also helps us segue into the the final discussion, which is about the standard deviation $\hat{\sigma_n}$ of $\frac{1}{n}\sum_{j=1}^{n} X_j$. As shown in Problem 3 too, since X_1, X_2, ..., X_n are all mutually independent,
$\text{Var}(\frac{1}{n}\sum_{j=1}^{n} X_j) = \frac{1}{n^2}\sum_{j=1}^{n} \text{Var}(X_j) = \frac{\sigma^2}{n} \implies \hat{\sigma_n} = \frac{\sigma}{\sqrt{n}}$
Thus, unlike $\hat{\mu_n}$, which is a stochastic quantity only on the asymptotic behaviour of whom we can comment upon, $\hat{\sigma_n}$ is a monotonically decreasing quantity which can be exactly calculated at any stage.

6.3 Results

Thus, we verify all the above predictions by using two sequences of random variables, one with an uniform distribution, and another with a Gaussian distribution, using a **box and whisker plot** to plot the set of deviations $|\hat{\mu_N} - \mu|$ for $M = 100$ repeats across a large set of values of N. Though the box and whisker plot shows us the median of the distribution (as opposed to the mean) and it's spread in terms of the difference between the maxima and minima (as opposed to the standard deviation), the general trend of predictions by the Law of Large Numbers is apparent from these metrics too, ie:- the median goes to zero (though not monotonically, as expected) as N grows larger, and the **distribution of errors becomes more and more concentrated around the central tendency as N grows large**. One may also cursorily note that the median (of deviations) in the Gaussian box plot are larger than the corresponding medians of the uniform box plot.
The copies of the plots are attached below for the grader's perusal :

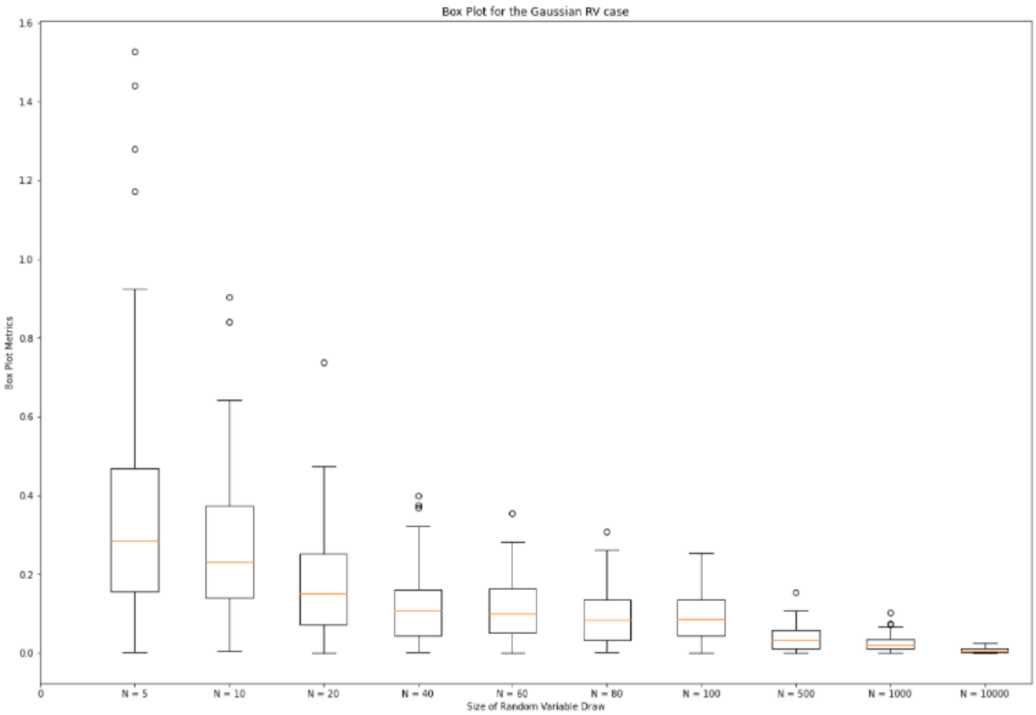

Figure 4: Problem 5 : The Gaussian RV Box Plot

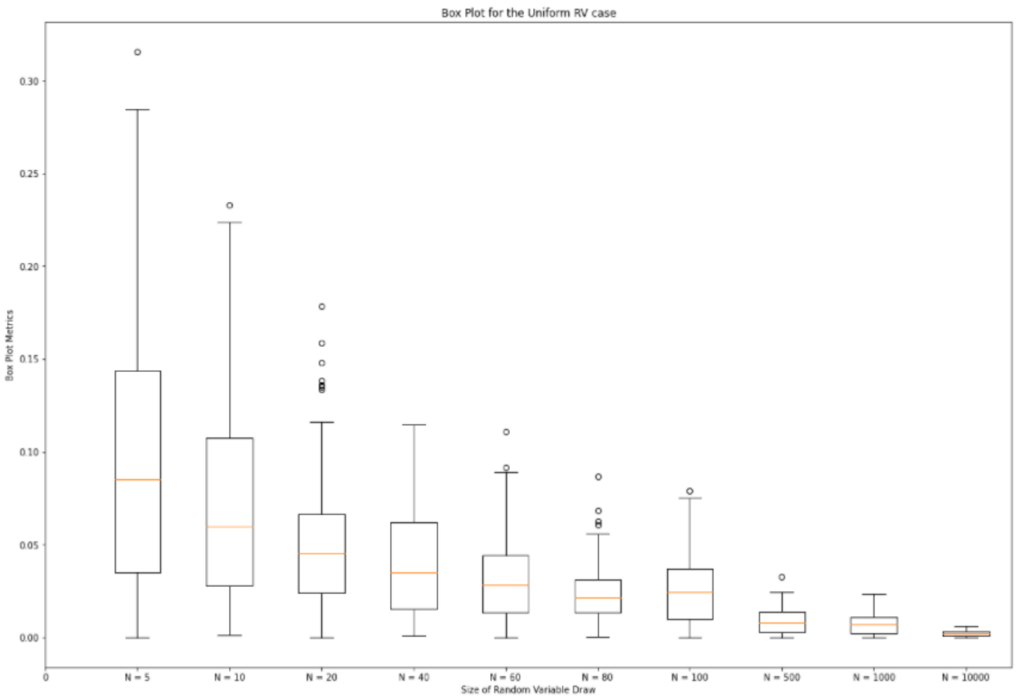

Figure 5: Problem 5 : The Uniform RV Box Plot