Generating Probability Distributions

Colin Gordon

August 19, 2023

Abstract

In this paper I will discuss a methodology on how to generate random numbers from the number line with a given probability distribution function. The methodology presented is a way to use pseudo-random number generators that are conventionally found in standard programming libraries to produce a random number that has a background frequency of a defined distribution.

1 Introduction

When we collect data, that data can usually be represented as a histogram that follows a semi-martingale pattern produced by Levy-T distributions such Gaussian, Lorentzian, Gamma, etc. It would be computationally useful to produce distributions to be used in simulated models. The methodology that I will present can be used for any differentiable probability distribution function that is defined over the real number line.

2 Theory

2.1 Probability distribution functions

A probability distribution function over the real number line is a distribution where the function the probability that a number lies in an interval is equal to the integral of the probability distribution over that interval.

$$Prob(x \in [a, b]) = \int_a^b \rho(x)dx \tag{1}$$

These probability distribution functions are integrable over the real number line with a total value of 1. they are usually designed to be differentiable and always be positive (no negative probability).

The probability function maps a subset of the real number line to a number in the interval $[0, 1]$. where there are three conditions the first conditions makes probability always positive. if $A \subset R$, $P(A) \geq 0$. The second condition makes the probability of the entire space 1. $P(R) = 1$ The third condition is that the union of disjoint/mutually exclusive sets are equivelant to addition in the probability space. if $E_1, E_2, ..., E_n$ are disjoint sets in R

then $P(E_1) + P(E_2) + ... + P(E_n) = P(\cup_{i=1}^{n} E_i)$ this is also true if the number of disjoint sets are infinite.

In the case of the real number line, we can always partition the real number line into a set of disjoint intervals which I will use later.

2.2 Pseudo Random number generators

Pseudo random number generators in computers are can be created by many different methodologies. One example is to solve the differential equation of a logistic curve with an Euler method, the sink equilibrium point can cause the value to vary in a near random fashion near the equilibrium. By design random number generators should produce a uniform distribution of numbers within some interval $[a, b]$. for this paper we will choose the interval to be $[0, 1]$. Note that we can always transform a number that is in an interval of $[a, b]$ to an interval $[0, 1]$ by taking the number $x \in [a, b]$ to make it the number $y = \frac{x-a}{b-a}$ which will be a random number in $y \in [0, 1]$ or invertably go from $x = a + (b - a)y$. We will utilize the pseudo random number generators in two ways. One will be to partition the interval [0,1] into sub intervals where the length of the intervals is the probability of a sub interval of the real number line. if the pseudo random number lands in that interval then we map that to an associate sub interval on the real number line. from that sub interval we randomly choose a number in that interval to be the randomly generated number that has a background probability density function.

2.3 Partitioning $[0, 1]$

You can constructively partition the number line from $[0, 1]$ with probabilities of intervals in a way where the length of an interval equates to a probability. We can order subintervals of the real number line in a way where we constructively build intervals in $[0, 1]$.

First we define our probability distribution in a way where the maximum occurs at $x = \mu$ which is usually an axis of symmetry or mean of the distribution. there is also another factor that represents the width and flatness of such a distribution to we can call γ for Lorentzian or σ for Gaussian distributions. This is like the standard deviation of the distribution. with these type of descriptive variables we can adjust the distribution to a unitless version where $\Delta = \frac{1}{\gamma}(x - \mu)$. This allows to computationally create a bunch of points for one type of the distribution and save it to memory. when we need a random point we just randomly pick a generated point of Δ and use γ and μ to calculate the associate $x = \gamma\Delta + \mu$. doing this type of transformation we can define our probability density function using change of variable. hence $\rho(\Delta)d\Delta = \rho(x)dx$ this means our distribution is now centered at the axis of symmetry, mean or weakest case maximum by construction and $\frac{1}{\gamma}dx = d\Delta$ making Δ unitless.

By construction, we can take a finite set of mutually exclusive intervals $\{I_k\}$ that cover R for the variable Δ. Since the intervals are mutually exclusive and cover the universal set then $\sum_k P(I_k) = P(\cup_k I_k) = P(R) = 1$. hence we can constructively partition $[0, 1]$ by defining.

$$P_0 := 0, P_k := P(I_k) + P_{k-1} \tag{2}$$

In essence $P_r = \sum_{k=0}^{r} P(I_k)$. If r is the number of the last interval then we know $P_r = 1$. hence $[0, 1]$ can be partitioned with a set of mutually exclusive intervals $\{\tilde{P}_k\}$ of the forms $[P_{k-1}, P_k]$, $[P_{k-1}, P_k)$, $(P_{k-1}, P_k]$, or (P_{k-1}, P_k) depending on how we define our I_k intervals and topological continuity we will impose. (closed intervals are closed, open are open, half and half are half and half). The nature of these intervals also have the length of $P(I_k)$. Constructively we pick $[P_{k-1}, P_k)$ usually unless we are dealing with edge cases.

The imposed relationship is that there is a function $\tilde{g} : \{\tilde{P}_k\} \to \{I_k\}$ that is one to one and onto. hence we can define a well-defined surjective relationship such that $g : [0, 1] \to I_k$ if $x \in \tilde{P}_k$ then $g(x) = I_k$

2.4 combining partitioning with pseudo random number generators

The construction of the function g allows us to take any number from the interval $[0, 1]$ and project that number to a unique interval in the real number line. If a number in the interval $[0, 1]$ was choosen randomly in a fair matter, then the probability of that number being in the interval $\tilde{P}_k$ is $P(I_k)$. This is due to the fact that the length of the interval is $P(I_k)$ by construction and the length of the universe $[0, 1]$ is 1. Since pseudorandom number generators are designed to have a near uniform distribution of generated numbers in an interval, this function g will can take a generated number y and turn that number into an interval $g(y)$. If we choose our intervals of the real number line to be small enough in a matter where $\rho(\Delta)$ has an extremely small variance across the interval, then randomily choosing another variable Δ from the interval $g(y)$ will approximate the likelyhood of that variable Δ being choosen from the background PDF $\rho(\Delta)$.

2.5 Generating subintervals of the real number line

there are two methods of how to create a subintevals $\{I_k\}$ on the real number line that have a small variance in $\rho(\Delta)$ for $\Delta \in I_k$. One method requires $\rho(\Delta)$ to be differentiable and non-zero but is the faster of the two. the other is to stretch the interval in a way where the max and min differ within a certain limit. The main idea is that we want to pick an interval where

$$sup[\rho(I_k)] - inf[(\rho(I_k))] \leq h < \rho(I_k) > \tag{3}$$

where $< \rho(I_k) >$ is the average over the interval, h is some scale factor, sup is the supremum/max and inf is the inferium/min of the interval. The equation (3) defines a constraint of the interval that relates the to an idea colloquially called percentage/relative difference h. This idea means that the highest mountain and the lowest trough has an elevation difference proportional to h. A smaller choice of h will result in a smaller relative elevation difference. Choosing a value of $h = 0.01$ would mean that the probability density will vary at most 1 percent within the interval making the density function have a near uniform probability distribution across the interval. This constraint allows us to use the random number generator once again where our Δ can be randomly chosen from I_k in an fair fashion with loosing a controlled tiny sliver of error from non-uniformity.

2.5.1 Differential method

This method approximates the relative difference by using the linear approximation methods from calculus. let I_k be an interval where the boundary points are Δ_{k-1} and $\Delta_k = \Delta_{k-1} + \delta$ which $\delta > 0$. if ρ is differentiable over the interval I_k then $\rho(\Delta_0 + \delta) - \rho(\Delta_{k-1}) = \rho'(\Delta_{k-1})\delta$. in general the sign of the derivative signifies if the function is locally decreasing or increasing. for sufficiently small enough intervals the function is strictly increasing or decreasing. hence

$$sup[\rho(I_k)] - inf[(\rho(I_k))] = |\rho(\Delta_{k-1} + \delta) - \rho(\Delta_{k-1})| \approx |\rho'(\Delta_{k-1})|\delta \tag{4}$$

if δ is sufficiently small then we can approximate $< \rho(I_k) > \approx \rho(\Delta_{k-1})$ and setting the $\leq$ to an equals sign we get.

$$|\rho'(\Delta_{k-1})|\delta \approx h\rho(\Delta_{k-1}) \tag{5}$$

the equation (5) can be used to generate δ in turn generate Δ_k by the following equation.

$$\Delta_k = \Delta_{k-1} \pm h\frac{\rho(\Delta_{k-1})}{|\rho'(\Delta_{k-1})|} \tag{6}$$

where we have $\pm$ depending if we want to increase or decrease our delta values. Note that the function divided by the derivative is the same as the reciprocal of $\frac{dLog(\rho)}{d\Delta}$

2.5.2 Direct methods

A direct approach maybe needed if the derivative is near zero. You can do this by directly substituting the absolute difference in (4) to (5) to get

$$|\rho(\Delta_k) - \rho(\Delta_{k-1})| \approx h\rho(\Delta_{k-1}) \tag{7}$$

the solution could be done analytically or utilize a standard root finding method, like Secant/midpoint/Newton methods. Another methodology that can be used is

$$|Log(\frac{\rho(\Delta_k)}{\rho(\Delta_{k-1})})| \approx h \tag{8}$$

Depending on the type of distribution one methodology maybe easier than the other. to give an example of each. For the center of the Lorentzian $\rho(\Delta) = \frac{1}{\pi(1+\Delta^2)}$ such that $\gamma = 1$ and $\mu = 0$ where $\Delta_{k-1} = 0$ then

$$|\frac{1}{\pi(1+\Delta_k^2)} - \frac{1}{\pi})| \approx h\frac{1}{\pi} \tag{9}$$

which Δ_k can be solved analytically directly as $\Delta_k^2 = \frac{h}{1-h}$ which if we want to be suave we can make it a bit simpler $\Delta_k = h^{\frac{1}{2}}$ since this number is slightly smaller than the fraction.

for a Gaussian $\rho(\Delta) \sim e^{-\frac{\Delta^2}{2}}$ the Log method would be beneficial at $\mu = 0$, $\Delta_{k-1} = 0$ and $\sigma = 1$

$$|Log(\frac{\rho(\Delta_k)}{\rho(0)})| = |\frac{-\Delta_k^2}{2} - 0| = \frac{\Delta_k^2}{2} \approx h \tag{10}$$

Direct method implementation is more accurate at employing intervals that maintain the relative difference being a maximum of h. however the smaller h is the difference accuracy diminishes. to save on computation time it is in general to employ the method in (8) than in the direct methods unless the derivative is absolutely zero or the change of Delta is so great that you would like to restrain it. On a computer you can catch the zero condition by assuming it is non-zero and try calculating. the zero in a denominator will result in an error which will activate a catch that would implore one of the two methods listed above associated with a root finding method like the secant method. on the other hand, you can use an if condition to change the values if the change of Delta is too big.

2.6 Dealing with Tail ends of the distribution

Since the real number line is not compact, there will not be a finite number of bounded intervals that will cover the entire real number line. Thus we need a methodology of how one should deal with the intervals with infinite length. To explore this lets look at the effects of the interval $[\Delta_0, \infty)$, the opposing direction $(-\infty, -\Delta_0]$ is analogues.

First lets define an equation to analyze the Tail of our cumulative function

$$F_+(\Delta) := P(x \geq \Delta) = \int_{\Delta}^{\infty} \rho(x)dx \tag{11}$$

$$F_-(\Delta) := P(x \leq \Delta) = \int_{-\infty}^{\Delta} \rho(x)dx \tag{12}$$

Properties of the probability density being normalized means that as $\Delta \to \infty$ that $F(\Delta) \to 0$ and $\rho(\Delta) \to 0$, hence we can expect that our function $\frac{\rho(\Delta)}{F(\Delta)} \to 0, L, \pm\infty$ for $\Delta \to \infty$. For the conditions where the limit is infinite, we can analyze the reciprocal of this function instead.

lets suppose that $\frac{\rho(\Delta)}{F_+(\Delta)} \to L$ then $-\frac{\rho'(\Delta)}{\rho(\Delta)} \to L$ by L'hospital's theorem since $F_+(\Delta) \to 0$ and $\rho(\Delta) \to 0$. Thus for every $\epsilon > 0$ there exists δ such that

$$|\frac{\rho(\Delta)}{F_+(\Delta)} + \frac{\rho'(\Delta)}{\rho(\Delta)}| \leq |\frac{\rho(\Delta)}{F_+(\Delta)} - L| + |L + \frac{\rho'(\Delta)}{\rho(\Delta)}| < 2\epsilon \tag{13}$$

for all $\Delta > \delta$. This allows us to approximate

$$F_+(\Delta) \approx -\frac{\rho^2(\Delta)}{\rho'(\Delta)} \tag{14}$$

for large Δ. and analogously we can show that

$$F_-(\Delta) \approx \frac{\rho^2(\Delta)}{\rho'(\Delta)} \tag{15}$$

2.6.1 Capping Constructed intervals

Since the intervals $\{I_k\}$ are constructed to be bounded and there are a finite number possible for a computer, it is impossible to cover the real numbers R due to its lack of compactness. hence we can set an arbitrary cap for our construction that will cover a large portion of the probability space such as 99.999 percent. we will call the cap number we choose as P_{max}. From here we can define our interval maximum for the intervals $\{I_k\}$ We will call this number Δ_{max} where $\Delta_{max} \geq sup[\cup_k I_k]$ Since the functions $F_\pm$ define the tails of our cumulative function then, we can set

$$F_-(-\Delta_{max}) + F_+(\Delta_{max}) = 1 - P_{max} \tag{16}$$

substituting the left-hand side with the equations (14) and (15) we can approximate Δ_{max}.

$$\frac{\rho^2(-\Delta_{max})}{\rho'(-\Delta_{max})} - \frac{\rho^2(\Delta_{max})}{\rho'(\Delta_{max})} \approx 1 - P_{max} \tag{17}$$

using a standard root solving this equation with a root finding method, will allow us to partition a large fraction of the probability space in the interval $[0, 1]$ making intervals $\{I_k\}$ cover the range from $[-\Delta_{max}, \Delta_{max}]$

2.6.2 Dealing with the rest of the tail

After covering the majority of the probability space with the intervals $\{\tilde{P}_k\}$ corresponding the $\{I_k\}$ there is going to still be a small sliver of probability space that needs computer logic to deal with. First you have the divide up the space in a way that represents the possibility of a large negative number and a large positive number. let $k = \kappa$ be the largest number P_k available of the constructed intervals $\tilde{P}_k$. note that $P_\kappa \approx P_{max}$. Since this space is approximated then we will have to normalize each space. We may switch the order which space should be defined first, either being the negative space or the positive space. In this instance we will define the positive space first.

we will define

$$P_{\kappa+1} = P_\kappa + \frac{F_+(\Delta_{max})}{F_-(-\Delta_{max}) + F_+(\Delta_{max})}(1 - P_\kappa) \tag{18}$$

this makes the interval $\tilde{P}_{\kappa+1} = [P_\kappa, P_{\kappa+1})$ and the last interval for the negative space is $\tilde{P}_{\kappa+2} = [P_{\kappa+1}, 1]$ which has a length of $\frac{F_-(-\Delta_{max})}{F_-(-\Delta_{max})+F_+(\Delta_{max})}(1 - P_\kappa)$ if a number is choosen within those intervals, then we have another selection mechanism to determine what the number ought to be. The region of large $|\Delta|$ can actually be broken down in a geometric fashion. To demonstrate this we will start with the fact that (14) and (15) can be manipulated to show this relationship.

$$\frac{\rho}{F_\pm} \approx |\frac{\rho'}{\rho}| \tag{19}$$

which is mathematically equivalent to

$$\frac{dLog(F_\pm)}{d\Delta} \approx |\frac{dLog(\rho)}{d\Delta}| \tag{20}$$

utilizing the fact that ρ is monotonic at the tail ends then by integration

$$\frac{F_{\pm}(\Delta_1)}{F_{\pm}(\Delta_2)} \approx \frac{\rho(\pm\Delta_1)}{\rho(\pm\Delta_2)} \tag{21}$$

this relationship is rather interesting since we generate Δ_k with the relationship in (8) we can use this method and combine it with (21) to define $\Delta_\kappa = \Delta_{max}$ then $\Delta_{\kappa+n}$ where

$$\frac{F_{\pm}(\Delta_{\kappa+n})}{F_{\pm}(\Delta_\kappa)} \approx \frac{\rho(\pm\Delta_{\kappa+n})}{\rho(\pm\Delta_\kappa)} \tag{22}$$

If we are looking at interval where $\frac{F_{\pm}(\Delta_{\kappa+n})-F_{\pm}(\Delta_{\kappa+n-1})}{F_{\pm}(\Delta_{\kappa+n-1})} = -h$ then we also know that $\frac{\rho(\pm\Delta_{\kappa+n})-\rho(\pm\Delta_{\kappa+n-1})}{\rho(\pm\Delta_{\kappa+n-1})} \approx -h$ which holds our constraint of the interval differences being a way where the density does not change significantly. what is beautiful about this is that

$$\frac{F_{\pm}(\Delta_{\kappa+n})}{F_{\pm}(\Delta_{\kappa+n-1})} \approx \frac{\rho(\pm\Delta_{\kappa+n})}{\rho(\pm\Delta_{\kappa+n-1})} \approx 1-h \tag{23}$$

which through recursive multiplication it breaks down into

$$\frac{F_{\pm}(\Delta_{\kappa+n})}{F_{\pm}(\Delta_\kappa)} \approx \frac{\rho(\pm\Delta_{\kappa+n})}{\rho(\pm\Delta_\kappa)} \approx (1-h)^n \tag{24}$$

Just subtracting two equations of the same form we can get the equation below

$$P([\Delta_{\kappa+n}, \Delta_{\kappa+n+1})|[\Delta_\kappa, \infty)) = \frac{F_{\pm}(\Delta_{\kappa+n}) - F_{\pm}(\Delta_{\kappa+n+1})}{F_{\pm}(\Delta_\kappa)} \approx h(1-h)^n \tag{25}$$

This conditional probability can be used to dertermine the random numbers that we can get if our first random variable generated is in $\tilde{P}_{\kappa+1}$ or $\tilde{P}_{\kappa+2}$.

let r be a random number from the interval $[0, 1]$, then we can calculate

$$n = Floor[\frac{Log(r/h)}{Log(1-h)}] \tag{26}$$

This n points to an interval $[\Delta_{\kappa+n}, \Delta_{\kappa+n+1})$ much like how $\tilde{P}_k \to I_k$. Using equation (24) we can determine $\Delta_{\kappa+n}$ and $\Delta_{\kappa+n+1}$ with root finding methods. Since that ρ over the interval $[\Delta_{\kappa+n}, \Delta_{\kappa+n+1})$ has a relative difference of h then we can randomly choose a number from this interval to be our randomly generated point from our probability density function.

Further examination eludes to an idea that instead of choosing from an interval again, we can look at the limit of $h \to 0^+$. let us define a new variable $\chi/h = n$, we will call $\Delta_\chi := \Delta_{\kappa+n}$ making $\Delta_{\chi+h} = \Delta_{\kappa+n+1}$ this transformation rewrites equations (24) and (25). in (24) the righthand side of the equation becomes $(1-h)^{\chi/h}$ which the limit of $h \to 0^+$ is $e^{-\chi}$ hence we get

$$\frac{F_{\pm}(\Delta_\chi)}{F_{\pm}(\Delta_\kappa)} \approx \frac{\rho(\pm\Delta_\chi)}{\rho(\pm\Delta_\kappa)} \approx e^{-\chi} \tag{27}$$

if you divide both sides of equation (25) and take the limit $h \to 0^+$ you will get

$$\rho(\Delta_\chi | [\Delta_\kappa, \infty)) = \frac{\rho(\pm\Delta_\chi)}{F_\pm(\Delta_\kappa)} \frac{d\Delta_\chi}{d\chi} \approx e^{-\chi} \tag{28}$$

plugging in (27) on the right handside of (28) we get

$$\frac{\rho(\pm\Delta_\chi)}{F_\pm(\Delta_\kappa)} \frac{d\Delta_\chi}{d\chi} \approx \frac{\rho(\pm\Delta_\chi)}{\rho(\pm\Delta_\kappa)} \tag{29}$$

which in turn shows that Δ_χ can be approximated in a near linear fashion like so

$$\frac{d\Delta_\chi}{d\chi} \approx \frac{F_\pm(\Delta_\kappa)}{\rho(\pm\Delta_\kappa)} \approx \frac{\rho(\pm\Delta_\kappa)}{\rho'(\pm\Delta_\kappa)} \tag{30}$$

The equation (30) comes from a bit of an over use of limits to calculate this approximation. hence it is the most accurate for $\chi \to \kappa$. I don't believe this (30) would be accurate enough to truly capture the nature of the tail end unlike (25) or (27) however it illustrates the geometric nature of how the curve changes for large values of Δ.

3 Procedures

Now that we have all of the pieces necessary to calculate random variables from the distribution I will outline the logic of how this will be done in code.

3.1 General Algorithm Development

3.1.1 optimizing by making unitless

This was discussed in the previous section where for a given distribution type, there are usually variables that can be set to 0 or 1. For a distribution like the Gaussian distribution the variable known as the mean can be set to 0, and the standard deviation can be set to 1. This transformation to make the probability distribution function to a unitless version allows the programmer to only have to do one run of generating variables and store those into long term memory. the unitless version should always have an invertable transformation to transform the generated variable into a variable with the correct units. To illustrate with our Gaussian example, the general curve has the form $\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ the unitless version would have $\sigma = 1$ and $\mu = 0$ making the function $\rho(\Delta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\Delta^2}{2}}$ where $\rho(\Delta)d\Delta = \rho(x)dx$ and $\Delta = \frac{1}{\sigma}(x - \mu)$. Thus if we generated a large set of random variables from the Δ distribution, we can translate that back into the unit based x definition with $x = \sigma\Delta + \mu$. randomly choosing from an already generated set will save on computation time significantly. This will be useful for standard distributions. This method however has to be manually done per each type of distribution hence is more of a user optimization rather than anything else.

3.1.2 All reals or not

if the non-zero probability space does not require the whole set of real numbers to analyze it may be of use to define the min/max Δ value possible for the curve at $\Delta = 0$ if only one max/min Δ exists, or make $\Delta = 0$ represent the midpoint for the case of both a max and min Δ exist. the existence of at least one min/max of Δ will make an adjustment to how we will deal with tails. If both exist, then the space is compact and can have a finite cover of intervals covering the distribution. Hence the cap is implied to be $P_{max} = 1$.

3.1.3 Defining Center

If there is no standard means of making the equation unitless or the equation does not have any max/min for Δ then the best choice would to make the center of the graph occur at an axis of symmetry to break up the tail piece a bit. if there is no axis of symmetry than a global maximum would suffice.

————————-IRRELEVANT TEMPLATE STUFF BELOW ————————————-

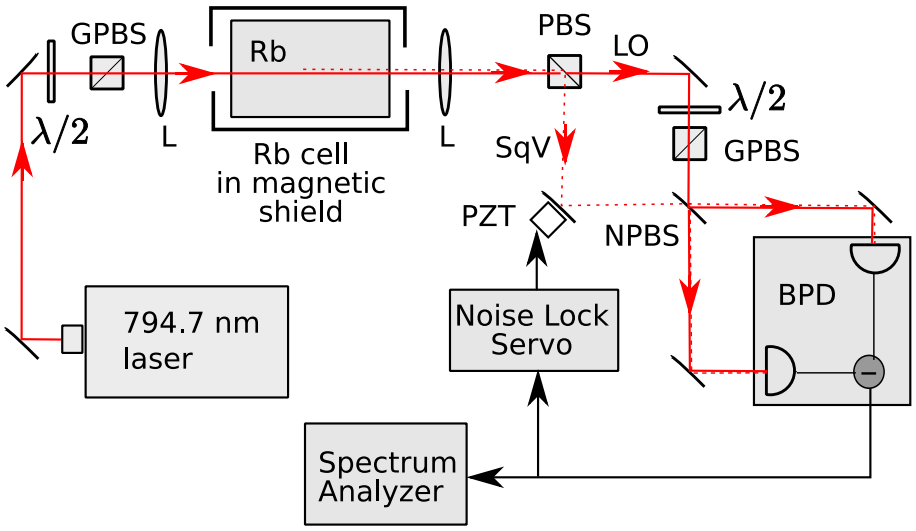

Figure 1: Every figure MUST have a caption.

Don't forget to list all important steps in your experimental procedure!

Use active voice either in past or present through all the report and be consistent with it: The laser light comes from to ... and eventually arrived to the balanced photodiode as seen in the figure 1.

Sentences in the past voice while correct are generally considered hard to read in large numbers. The laser light was directed to ..., wave plates were set to ... etc.

4 Analysis

In this section you will need to show your experimental results. Use tables and graphs when it is possible. Table 1 is an example.

Table 1: Every table needs a caption.

x (m)	V (V)
0.0044151	0.0030871
0.0021633	0.0021343
0.0003600	0.0018642
0.0023831	0.0013287

Analysis of equation **??** shows ...

Note: this section can be integrated with the previous one as long as you address the issue. Here explain how you determine uncertainties for different measured values. Suppose that in the experiment you make a series of measurements of a resistance of the wire R for different applied voltages V, then you calculate the temperature from the resistance using a known equation and make a plot temperature vs. voltage squared. Again suppose that this dependence is expected to be linear [2], and the proportionality coefficient is extracted from the graph. Then what you need to explain is that for the resistance and the voltage the uncertainties are instrumental (since each measurements in done only once), and they are .... Then give an equation for calculating the uncertainty of the temperature from the resistance uncertainty. Finally explain how the uncertainty of the slop of the graph was found (computer fitting, graphical method, *etc.*)

If in the process of data analysis you found any noticeable systematic error(s), you have to explain them in this section of the report.

It is also recommended to plot the data graphically to efficiently illustrate any points of discussion. For example, it is easy to conclude that the experiment and theory match each other rather well if you look at Fig. 1 and Fig. 2.

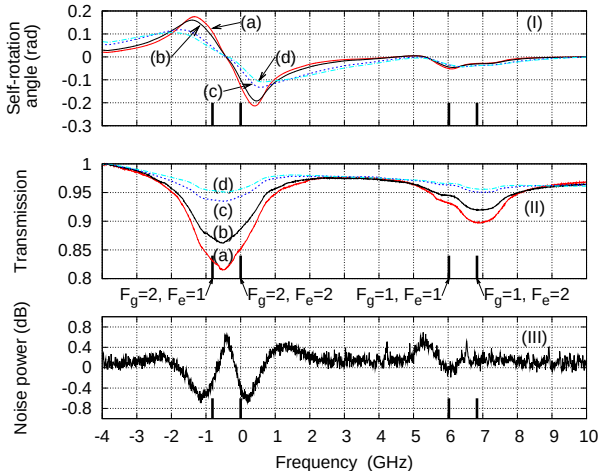

Figure 2: Every plot must have axes labeled.

5 Conclusions

Here you briefly summarize your findings.

References

[1] A. C. Melissinos and J. Napolitano, *Experiments in Modern Physics*, (Academic Press, New York, 2003).

[2] N. Cyr, M. Têtu, and M. Breton, IEEE Trans. Instrum. Meas. **42**, 640 (1993).

[3] *Expected value*, available at `http://en.wikipedia.org/wiki/Expected_value`.