1 Introduction

The purpose of these notes is examples in undergraduate mathematics that the author considers to be interesting; this could be from applications or pure mathematical interest.

2 One Variable Differential Calculus

2.1 Gauss and the Gauss Distribution

History

A good reference on the history of the gaussian distribution is [2].

The history of how to deal with errors is intimately tied to astronomy; astronomical predictions involve quantities that need to be measured to high precision. Practical limits force astronomers to deal with the errors of predictions or measurements never being in complete agreement.

In the 18th century and early 19th century, there was some confusion as to how to deal with these errors in measurement. As an example, there was some dispute as to whether to use the average or the median of measurements. One of the problems was a theoretical foundation for understanding error was in its infancy. For example, Laplace created a model of typical error that is far from the typical gaussian distribution considered today.

So how did Gauss arrive at his distribution? First it should be noted that he worked on modeling error while solving a problem in astronomy. On January 1, 1801, Giuseppe Piazzi observed the Ceres asteriod. He was interested in whether Ceres was a new planet, but he could only take a small number of observations of its position before it disappered behind the sun. Ceres was estimated to be visible again after about a year, which left many astronomers with the question of where to find it in the sky.

Gauss greatly increased his reputation by correctly solving this problem; in fact, his correct answer was actually in disagreement with most reputable astronomers. Aside from his masterful use of geometry, part of his solution is how to deal with the errors in measurements that were made. It is this problem that lead him to the gaussian distribution as a model for the error.

His approach to modeling the error is the following.

He considers the errors to be random described by a differentiable probability density $p(x)$. The distribution of the errors should satisfy the following:

1. Smaller errors are more probable, i.e. the density $p(x)$ should have a maximum at $x = 0$.

2. The distribution of errors should symmetric, i.e. $p(-x) = p(x)$.

3. Consider any observed quanitity X with true value X_0 and errors modeled by our distribution, i.e. $X = X_0 + G$ where $P(G = x) = p(x)$.

Given any set of observations $\{x_1, x_2, ..., x_n\}$, then the likelihood $P(x_1, x_2, ..., x_n|X_0)$ (i.e. the probability of observing x_1, x_2, ... x_n given the true value is X_0) is maximized by X_0 being the average of $\{x_1, x_2, ..., x_n\}$ (i.e. $X_0 = \frac{x_1+x_2+...+x_n}{n}$). Let us explain this in a little more detail.

We are assuming that the errors $\{x_1, x_2, ..., x_n\}$ are independent. So

$$P(x_1, x_2, ..., x_n|X_0) = p(x_1 - X_0)p(x_2 - X_0)...p(x_n - X_0). \tag{1}$$

When we speak of maximizing the likelihood, we think of all of the observations x_i being fixed. So the above is considered to a function of only the one variable X_0. That is, we are considering the likelihood functions

$$L(X_0) = p(x_1 - X_0)p(x_2 - X_0)...p(x_n - X_0). \tag{2}$$

Then our assumption is that the maximum of $L(X_0)$ occurs at the average of our observations $X_0 = \frac{x_1+x_2+...+x_n}{n}$.

This amounts to Gauss's justification of using averages over median. He is purposefully choosing a model of error where the average of the observations is the most likely explanation of the true value.

The Problem

Show that Gauss' requirements on $p(x)$ force $p(x)$ to be a Gaussian distribution.

The Solution

First, let us consider condition (3) and the consequences of maximizing the likelihood. First note that $L(X_0) \geq 0$, so maximizing $L(X_0)$ is equivalent to maximizing $f(X_0) = \log(L(X_0))$. Using the logarithm will be more convenient as it will turn the product of the $p(x_i - X_0)$ into a sum of logarithms; so we have

$$h(X_0) = \log(p(x_1 - X_0)) + \log(p(x_2 - X_0)) + ... + \log(p(x_n - X_0)). \tag{3}$$

To find the maximum, let's set the derivative to be zero:

$$0 = h'(X_0) = -\left(\frac{p'(x_1 - X_0)}{p(x_1 - X_0)} + \frac{p'(x_2 - X_0)}{p(x_2 - X_0)} + ... + \frac{p'(x_n - X_0)}{p(x_n - X_0)}\right). \tag{4}$$

Now the key is that condition (3) applies to any possible set of observations, no matter how unprobable. Since $p(x)$ is continuous with maximum at $x = 0$, we know that there exists an interval $[-\delta, \delta]$ around $x = 0$ such that $p(x) > 0$ for all $x \in [-\delta, \delta]$. In particular, we know that observations in $[X_0 - \delta, X_0 + \delta]$ are all possible. So now consider any real number $r \in [-\delta, \delta]$ and the observations $\{x_1 = X_0\}$ and $\{x_2 = ... = x_n = X_0 + r\}$.

Since we have already fixed X_0 to represent our true value, let us now use y as the independent variable for our likelihood. So we seek to maximize

$$L(y) = p(x_1 - y)p(x_2 - y)...p(x_n - y). \tag{5}$$

Condition (3) says this maximum is at $y = \frac{x_1+x_2+\dots x_n}{n} = X_0 + \frac{n-1}{n}r$. To simplify notation, let $f(x) = \frac{p'(x)}{p(x)}$. So we get

$$0 = f\left(-\frac{n-1}{n}r\right) + (n-1)f\left(\frac{1}{n}r\right). \tag{6}$$

Now, note that $p(x)$ symmetric implies that $p'(x)$ is anti-symmetric. Therefore $f(x)$ is anti-symmetric. So we get that

$$f\left(\frac{n-1}{n}r\right) = (n-1)f\left(\frac{1}{n}r\right). \tag{7}$$

What are the consequences of this equation? Fix any r_0 small enough such that $2r_0$ is in the interval $[-\delta, \delta]$. Now note that $\frac{n}{n-1}r_0$ is also in the interval for any $n > 1$, and consider $r = \frac{n}{n-1}r_0$. Then we have that

$$\frac{1}{n-1}f(r_0) = f\left(\frac{r_0}{n-1}\right). \tag{8}$$

Now consider $0 < k \leq n+1$. Note that $\frac{k}{n}r$ is in the interval too, and now apply 6 for $n- > k$ and $r- > \frac{k}{n}r$, we get

$$f\left(\frac{k-1}{n}r\right) = (k-1)f\left(\frac{1}{n}r\right). \tag{9}$$

So for any fraction of the form $fracmn$ where $0 < m \leq n$, we have that

$$f\left(\frac{m}{n}r_0\right) = \frac{m}{n}f(r_0). \tag{10}$$

Consider the function $g(x) = f(r_0)x$. We have that $g(x) - f(x) = 0$ for any x that is a rational multiple of r_0 and $0 < |x| \leq r_0$. Hence, by the continuity of $f(x)$, we have that $f(x) = g(x) = f(r_0)x$ for all $|x| \leq r_0$.

Therefore, we may write that $f(x) = kx$ for some constant k and x on some interval around zero. This gives us the differential equation

$$\frac{p'(x)}{p(x)} = k, \tag{11}$$

locally around $x = 0$. Integrating we get

$$\log(p(x)) = \frac{k}{2}x^2 + C. \tag{12}$$

This can be written in the form $p(x) = Ae^{-Bx^2}$. So we see the probability distribution extends to be non-zero on all x. Furthermore, the constants A and B can be related by the fact that $p(x)$ is a probability density so that $\int_{-\infty}^{\infty} Ae^{-Bx^2}\,dx = 1$. It is standard to solve for A in terms of B.

To solve $\int_{-\infty}^{\infty} e^{-Bx^2}\,dx$, we square the integral and switch to polar coordinates to get

$$\int_{-\infty}^{\infty} e^{-Bx^2}\,dx = \sqrt{\frac{\pi}{B}}. \tag{13}$$

So we get that $A = \sqrt{\frac{B}{\pi}}$, and then

$$p(x) = \sqrt{\frac{B}{\pi}} e^{-Bx^2}. \tag{14}$$

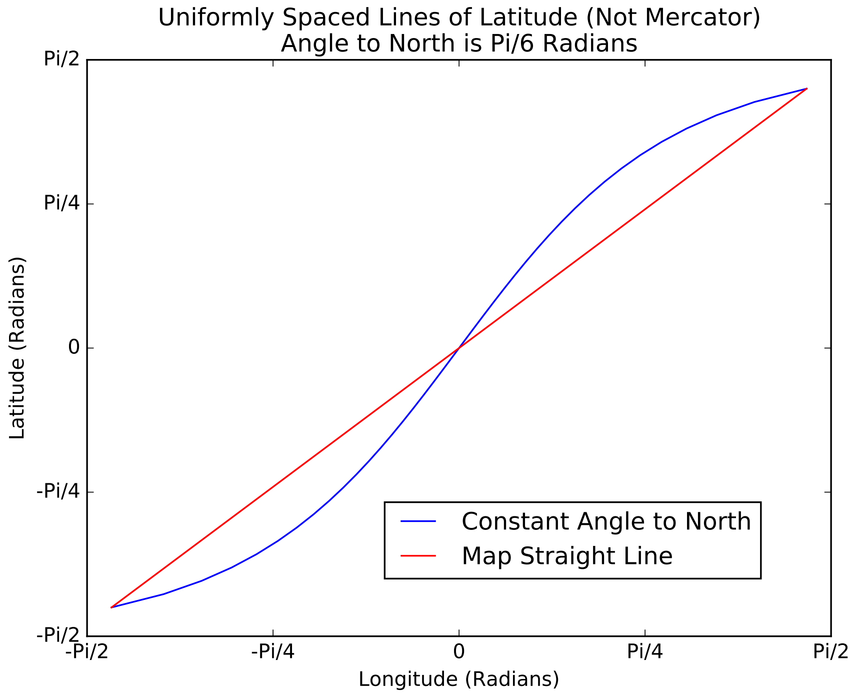
3 One Variable Integral Calculus

3.1 The Mercator Map and the Integral of Secant

Historical Motivation

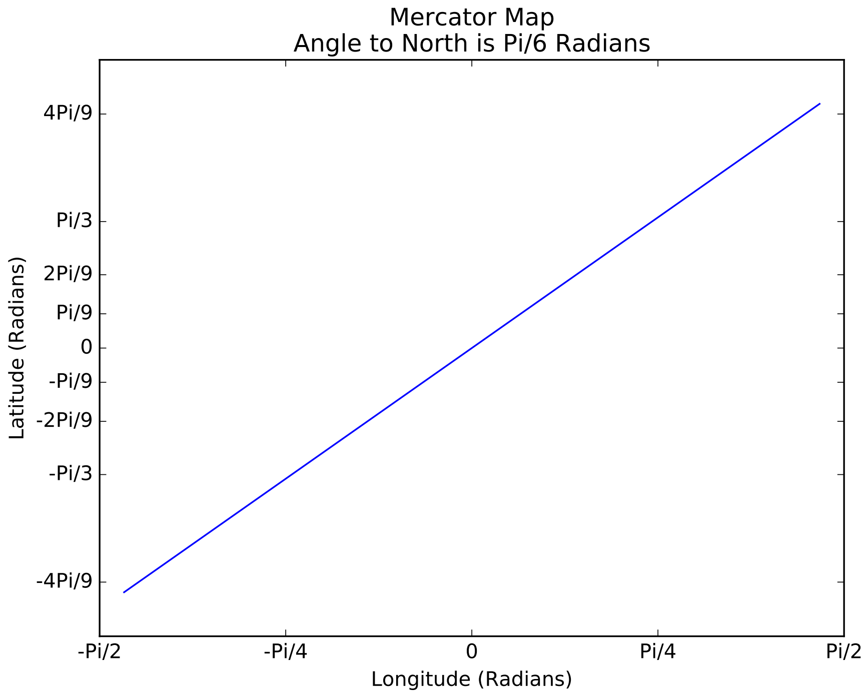
A great reference for the history of the integral of the secant function and its relation to the Mercator map is [1]. We will give a brief overview here.

The Mercator Map of the world spaces out the lines of latitude in a particular way inorder to solve a problem in naval navigation. The problem is that ships would navigate by sailing with a fixed angle to due north (e.g. as seen on a compass). This creates an issue for making map. Consider a map where the lines of latitude are spaced out evenly in the vertical direction (so NOT the Mercator map); for such a map, a course with fixed angle to magnetic north is NOT a straight line on the map. The figure below shows the path of a course with constant angle to magnetic north on a map with uniformly spaced lines of latitude.



The problem is that the lines of latitude get represent shorter and shorter distances as you move from the equator towards either of the poles. This means that there is a complicated relationship between the angle measured on this map and the true angle to magnetic north it represents.

In 1569, Mercator had the idea that he could create a map where the lines of latitude are NOT spaced evenly; if you choose the variation in spacing in the correct manner, then a course with fixed angle to magnetic north will be a straight line on this new map. Furthermore, the angle measured on the map will match the true angle to magnetic north. Consider the following figure that shows a course with constant angle to magnetic north on a Mercator map, note that the lines of latitude are not evenly spaced (the values marked are multiples of $\pi/9$ radians).



Unfortunately, Mercator didn't give a clear formula to precisely describe how to space out the lines of latitude. However, in 1599, Edward Wright found a precise mathematical description of how to space out the lines; he found that the spacing depended on the area under the secant function. He didn't know how to precisely compute this area, but he was able to approximate it.

Later in the 1640's, Henry Bond looked at a table of these approximate areas and a table of logarithms of trigonometic functions. He noticed a similarity in the two tables, and he was able to conjecture a precise formula for the area under the secant function. We now know that his conjecture was correct, but at the time there was no proof beyond numerical tables.

A proof was later given by Isaac Barrow; this proof is the earliest known publication of the use of integration by partial fractions.

The Problem

Compute the integral

$$\int_0^x \sec(u)\,du. \tag{15}$$

The Solution

Recall that $\sec(u) = \frac{1}{\cos(u)}$. First, let's use algebraic manipulation combined with the trigonometric formula $\cos^2(u) + \sin^2(u) = 1$.

$$\int_0^x \frac{1}{\cos(u)}\,du = \int_0^x \frac{\cos(u)}{\cos^2(u)}\,du, \tag{16}$$

$$= \int_0^x \frac{\cos(u)}{1-\sin^2(u)}\,du. \tag{17}$$

Now, we do a u-substitution. However, we are already use the variable u, so let's make it a "w-substitution". We use $w = \sin(u)$, and so $dw = \cos(u)\,du$. Then we have that our integral is:

$$\int_0^{\sin(x)} \frac{1}{1-w^2}\,dw. \tag{18}$$

Now, we use partial fractions:

$$\frac{1}{1-w^2} = \frac{1}{(1-w)(1+w)}, \tag{19}$$

$$= \frac{A}{1-w} + \frac{B}{1+w}. \tag{20}$$

Combining terms and comparing numerators, we get $A+B+(A-B)w = 1$. So we have

$$\begin{cases} A+B=1, \\ A-B=0. \end{cases} \tag{21}$$

Solving we get $A = B = \frac{1}{2}$.

Therefore, our integral becomes

$$\int_0^{\sin(x)} \frac{1}{2(1-w)} + \frac{1}{2(1+w)}\,dw = \frac{1}{2}\log\left(\frac{1+w}{1-w}\right)\Bigg|_0^{\sin(x)}, \tag{22}$$

$$= \frac{1}{2}\log\left(\frac{1+\sin(x)}{1-\sin(x)}\right). \tag{23}$$

To simplify things, we can now use some trigonometric identities.

$$\frac{1}{2}\log\left(\frac{1+\sin(x)}{1-\sin(x)}\right) = \log\sqrt{\frac{1+\sin(x)}{1-\sin(x)}}, \tag{24}$$

$$= \log\sqrt{\frac{(1+\sin(x))^2}{1-\sin^2(x)}}, \tag{25}$$

$$= \log\left(\frac{1+\sin(x)}{\cos(x)}\right), \tag{26}$$

$$= \log(\sec(x)+\tan(x)). \tag{27}$$

4 Multivariable Differential Calculus

Here are examples related to differential calculus in more than one variable.

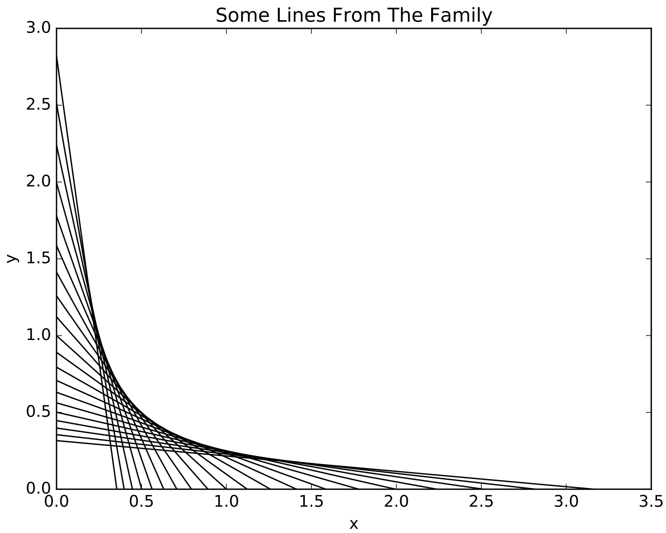
4.1 Envelopes

In the simplist cases, the **envelope** of a family $\mathfrak{F}$ of curves is a curve γ that is in some sense extremal to the entire family of curves. What is often the case, is that every point of the curve γ touches exaclty one curve from the the family $\mathfrak{F}$, and furthermore this touching is only tangential (i.e. they cross at an angle of zero). This is best illustrated with examples.

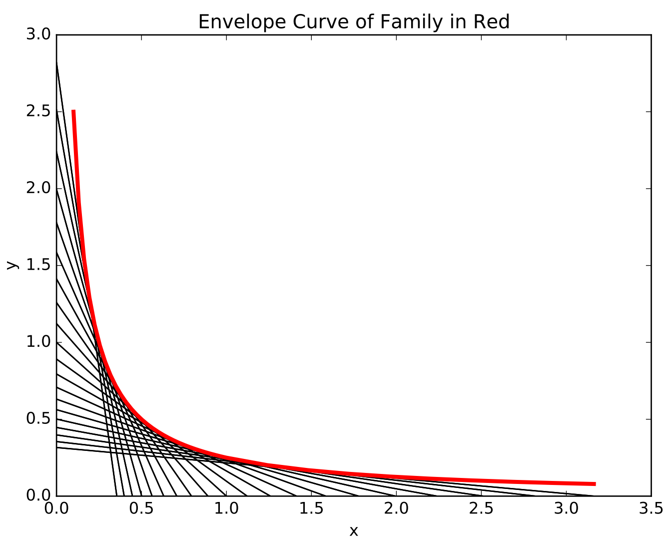
4.1.1 The Hyperbola as an Envelope

The Set Up

Consider the family $\mathfrak{F}$ of straight lines in $\mathbb{R}^2$, where each line crosses the x-axis and y-axis at pairs of points of the form $(s,0)$ and $(0,1/s)$ for some $s>0$. So we see that each line is of the form $\frac{1}{s}x + sy = 1$ for some $s>0$. Some of the lines from the family are pictured in the following figure.



We can see that extemal to the family of lines is a curve concave up in the first quadrant $\{x, y > 0\}$. In the following figure, you can see the curve superimposed with some of the lines from the family.



The Problem

Let us consider computing the envelope curve $\gamma(x)$ of the family $\mathfrak{F}$.

The Solution

To compute the envelope curve $\gamma(x)$, let us consider the auxilliary function $g(x, y, s) = \frac{1}{s}x + sy - 1$. Let us see how the Implicit Function Theorem of vector calculus let's us use $g(x, y, s)$ to find the exremal envelope curve $\gamma(x)$. As we discuss this, please consider the similarities to the ordinary first derivative test.

First, consider any point (x_0, y_0) NOT on the extremal envelope curve $\gamma(x)$, but is touched by some line in $\mathfrak{F}$. So there is some $s_0 > 0$ such that $\frac{1}{s_0}x_0 + s_0 y_0 = 1$; note that this is equivalent to $g(x_0, y_0, s_0) = 0$. Since (x_0, y_0) isn't on the boundary of the region of points touched by lines in $\mathfrak{F}$, we know that for any other points (x_1, y_1) close to (x_0, y_0) we may find another line in $\mathfrak{F}$ touching (x_1, y_1). That is, for every (x_1, y_1) close to (x_0, y_0), we may find $s_1 > 0$ such that $g(x_1, y_1, s_1) = 0$.

This can be summarized as saying that for all points (x_0, y_0) that are touched by a line in $\mathfrak{F}$ and also isn't on the envelope γ, we can locally solve $s = S(x, y)$ such that $g(x, y, S(x, y)) = 0$. Now, you may begin to see the connection to the Implicit Function Theorem.

Recall that the Implicit Function Theorem can only confirm that we CAN locally solve $s = S(x, y)$ such that $g(x, y, S(x, y)) = 0$. However, we seek for the extremal points where we CAN'T locally solve. This is similar to the first derivative test of ordinary calculus. Technically, the first derivate test only says when a point is NOT an extemum of a function; then the candidate points for extrema are reduced to some finite list by solving for the vanishing of the derivative.

Here, we are in a similar situation. We solve for a set of candidate points that must contain our extremal curve γ. It will happen to be the case that our candidate set will allow only one curve and so this must be the envelope. However, we are being a little reckless here as we haven't proven the curve must exist; we will consider the picture to be very convincing and ignore this technical detail.

So we seek for when we can't locally solve $s = S(x, y)$ such that $g(x, y, S(x, y)) = 0$. The Implicit Function Theorem tells us this will only be possible for those (x, y, s) with $g(x, y, s) = 0$ and $\frac{\partial g}{\partial s}(x, y, s) = 0$.

So we look for

$$0 = \frac{\partial g}{\partial s}, \tag{28}$$

$$= -\frac{x}{s^2} + y \tag{29}$$

We wish to find an equation restricting x and y; so it is most efficient to solve the above for s. Also, from the picture it is clear that we should restrict to $x, y > 0$. Therefore, for $x, y > 0$, we have $s = \sqrt{\frac{x}{y}}$. Plugging this into the equation for $g(x, y, s) = 0$, we get

$$\sqrt{xy} + \sqrt{xy} - 1 = 0. \tag{30}$$

Therefore, we find that the envelope curve must lie inside the set $S = \{xy = \frac{1}{4}\}$. However, one will recognize that for each point $x > 0$, there is only one y such that $y \in S$. Therefore, the envelope must be this curve.

So the envelope $\gamma(x)$ is the curve $y = \frac{1}{4x}$ for $x > 0$.

Final Remark

Note that the set S is actually a hyperbola. Therefore, the hyperbola can be realized as the envelope of a simple family of straight lines. For this reason, hyperbolas are (approximately) reproducible in ”string art”: art formed from straight line segments where each segment is made by tightened string.

References

[1] V. Frederick Rickey and Philip M. Tuchinsky. An application of geography to mathematics: History of the integral of the secant. *Mathematics Magazine*, 1980.

[2] Saul Stahl. The evolution of the normal distribution. *Mathematics Magazine*, 2006.