BERNSTEIN POLYNOMIALS: A BRIDGE BETWEEN THE WEIERSTRASS THEOREM AND A LANDAU TEMPERATURE FLUCTUATION PROBLEM

AIDEN ROHRBACH*
PROJECT ADVISOR: XUWEN CHEN†

October 21, 2025

Abstract

We address a Landau temperature fluctuation problem by applying Bernstein polynomials and the Weierstrass approximation theorem. We consider the problem detailed in [X. Wang and Q.H. Liu, *Temperature fluctuations for a finite system of classical spin-1/2 particles*, Annals of Physics 322 (2007), 2168-2178] in which temperature fluctuations diverge as the temperature of a particle system tends to 0. An application of Bernstein polynomials allows us to solve this issue by proving that the m'th moment of the configurational temperature of a finite particle system tends to the temperature of the surroundings, implying the fluctuations in the temperature must tend to 0.

Contents

*Department of Mathematics, University of Rochester (arohrbachj@gmail.com)
†Department of Mathematics, University of Rochester (xuwenmath@gmail.com), partially supported by NSF DMS-2406620

1 Introduction

Throughout history, pure math and theoretical physics have gone hand-in-hand. Frequently, pure math will discover an important principle that seems disconnected from reality, but upon further inspection, there is a profound application in physics. For example, group theory was developed hundreds of years ago and, upon the invention of quantum mechanics, found an application in describing the spin states of elementary particles through the special unitary group. This illustrates a deep connection between physics and pure math; even the most abstract idea in math can find an incredible application in describing reality.

Another aspect of theoretical physics that can see an application of pure math is the Landau thermodynamic problem. In the standard examination of particle systems, theory predicts a strange occurrence: as the temperature of the surrounding system drops to 0 K, the fluctuations in the temperature grow without bound. This seems to be dissonant with a physical interpretation of a spin system, for temperature fluctuations growing infinitely large has no real meaning.

Temperature, and the fluctuations in temperature, are measurable properties of a physical system. As such, it seems counterintuitive that one of these values could diverge while the other drops to 0. Due to this, there is a necessitiy to create a theory of finite particle systems that reflects what we see experimentally. By utilizing Bernstein polynomials, a topic rooted firmly in analysis, we may be able to solve this problem, justifying the much more reasonable idea that the temperature fluctuations tend to 0 as the temperature tends to 0. X. Wang and Q.H. Liu [5] were the first to document this method of solving the problem. This paper provides the mathematical rigor behind their solution.

2 Background and Main Theorem

2.1 The Limit Definition

In order to understand much of the physics and our main theorem, we need some conception of what it means for a function to "get closer" to a value as it's input varies. The limit helps us with this:

Definition 1 (Limit)**.** *A function $f(x)$ approaches y as x approaches t if for every $\epsilon > 0$, there exists some δ such that $|f(x) - y| < \epsilon$ when $|x - t| < \delta$.*

In this case, we say f approaches y as x approaches t, or $\lim_{x\to t} f(x) = y$. This gives us a mathematical description of a function approaching a number, and will be instrumental in analyzing the behavior of the temperature fluctuations as the temperature of the situation approaches zero.

We may also achieve more definition in the behavior of a function at a specific point by analyzing one-sided limits. There may be a case where a function approaches y as x tends to t *from the left.* In this case, we say $\lim_{x\to t^-} f(x) = y$. An identical definition can be made for x tending to t from the right. In these cases, the function may have differing behavior on each side of t. The formal definition of a one-sided limit follows:

Definition 2 (One-sided limit)**.** *A function $f(x)$ approaches y as x approaches t from the left if for every $\epsilon > 0$, there exists a δ such that $|f(x) - y| < \epsilon$ when $0 < t - x < \delta$.*

In the following section, we may make the statement of a limit more abstract. For example, we may say something like $f(x) \to \infty$ as $x \to 0$. While there is no clear limit notation here, this implicity translates to $\lim_{x\to\infty} f(x) = 0$. This will be important to keep in mind while reviewing the current physics.

2.2 A Review of the Current Physics

In order to understand the physics of the problem we are trying to solve, we must first understand what temperature is. As defined by Landau and Lifshitz in [3], the temperature of a system is defined as the reciprocal of the derivative of entropy S with respect to energy E. Temperature is only meaningful for a system of particles; it is not a quantity that has meaning for individual particles, due to the purely statistical nature of the quantity. The entropy S of a system, used to define temperature, is defines as a constant factor k times the natural log of Σ, the number of accessible microstates of a physical system.

The chief issue that we are solving is described by Landau and Lifschitz [3] in their 12'th chapter. They note how these physical quantities, such as temperature and entropy, are almost always nearly equal to their mean values, but fluctuate. However, in calculating these fluctuations, we must impose the restriction that the fluctuations are small, or the temperature is not near zero. In the cases where these restrictions don't apply, we have to treat the fluctuations are a purely quantum system. The goal of X. Wang and Q.H. Liu [5] was to remedy the issue these restrictions pose, in showing that the fluctuations do tend to zero as the temperature of the system tends to zero. We review their work now.

If we are to assume a system of practically infinite ($N > 10^{23}$), spin-1/2 particles, the temperature fluctuation is defined as

$$\overline{(\Delta T)^2} = \frac{kT_{\text{can}}^2}{C_B} \tag{1}$$

where k is the Boltzmann constant, T_{can} is the fixed temperature of the surrounding system, and C_B is the heat capacity for the system. This heat capacity is

$$C_B = Nk\left(\frac{\epsilon}{kT_{\text{can}}}\right)^2 \text{sech}^2\left(\frac{\epsilon}{kT_{\text{can}}}\right) \tag{2}$$

where ϵ is the energy quantum of independent, spin-1/2 nuclei in a magnetic field B in thermal equilibrium. From this, it can be seen that

$$\overline{(\Delta T)^2} = \frac{k^2T_{\text{can}}^4}{\epsilon^2 N}\cosh^2\left(\frac{\epsilon}{kT_{\text{can}}}\right) \tag{3}$$

This is where the Landau temperature fluctuation problem arises. Equation (3) shows that $\overline{(\Delta T)^2} \to \infty$ as the temperature, T_{can}, tends to 0; a result of the hyperbolic cosine function diverging as the argument grows infintely large. As seen in Landau and Lifshitz's statisitcal mechanics textbook [3], deriving these equations implicitly presumes the temperature fluctuations are small, and so the result requires statistical methods when the fluctuations diverge.
To attempt to solve this problem, let's introduce a system A of a finite number N particles. Consider this system as a small fraction of a much larger closed one A_0, where the rest of the closed system, A_b, makes up the surrounding environment with M particles. So the full system $A_0 = A \oplus A_b$. Using this definition, when $T_{\text{can}} = 0$ K, all of the particles are spin-up. When $T_{\text{can}} \neq 0$, there are n_0 spin-down particles among the $M + N$ total particles. In such a system, we can define the temperature T_{can} as

$$T_{\text{can}} = \frac{2\epsilon}{k}\left[\ln\left(\frac{M+N}{n_0} - 1\right)\right]^{-1} \tag{4}$$

Since this temperature is the temperature of the surrounding environment A_b, every particle has the same probability p to be spin-down, with p given by

$$p \equiv \frac{n_0}{M+N} = \frac{\exp\left(-\frac{\epsilon}{kT_{\text{can}}}\right)}{\exp\left(-\frac{\epsilon}{kT_{\text{can}}}\right) + \exp\left(\frac{\epsilon}{kT_{\text{can}}}\right)} \tag{5}$$

Since the temperature of the system A is not T_{can} if N is relatively small, we can assume $n_0 > N$, implying the number M is so large that $n_0 = p(M+N) > N$ can be satisfied.
If we assume that at an instant t, there is a configuration of n spin-down particles and $N - n$ spin up particles, then the configuration appears with probability

$$\binom{N}{n} p^n (1-p)^{N-n} \tag{6}$$

by using the binomial probability distribution. Then, as specified in the standard treatment [3], the configurational finite N temperature $T(n,N)$ is defined as

$$T(n,N) = \begin{cases} 0 & n = 0, N \\ 0 & n = \frac{N}{2} \text{when } N \text{ is even} \\ \frac{2\epsilon}{k}(\ln(\frac{N-n}{n}))^{-1} & \text{otherwise} \end{cases} \tag{7}$$

This definition suffers from a discontinuity when N is even. A similar situation happens when the number of particles N is odd. When N is odd, we get

$$T(n,N) = \begin{cases} 0 & n = 0, N \\ \frac{2\epsilon}{k}(\ln(\frac{N-n}{n}))^{-1} & \text{otherwise} \end{cases} \tag{8}$$

In this ensemble, the mean value of a quantity $f(n,N)$ is defined as

$$\overline{f}(p,N) = \sum_{n=0}^{N} f(n,N)\binom{N}{n}p^n(1-p)^{N-n} \tag{9}$$

and so the mth moment of the configurational temperature is

$$\overline{T^m}(p,N) = \sum_{n=0}^{N}(T(n,N))^m\binom{N}{n}p^n(1-p)^{N-n} := T \tag{10}$$

We now arrive at our main theorem.

Theorem 1 (Main Theorem)**.** *If $\overline{T^m}(p,N)$ is given by (10), then*

$$\lim_{N\to\infty} \overline{T^m}(p,N) = \left(\frac{2\epsilon}{k}\frac{1}{\ln(\frac{1}{p}-1)}\right)^m = (T_{can})^m \tag{11}$$

When $m = 1$, we recover the probability given by the Maxwell-Boltzmann distribution through algebraic manipulations. This shows that, in the thermodynamic limit, temperature fluctuations do not diverge. In other words,

$$\lim_{N\to\infty} \overline{(\Delta T)^2} = 0 \quad (0 < T_{\text{can}} < \infty) \tag{12}$$

The key behind proving equation (12) is modifying a proof of the Weierstrass approximation theorem. The proof in question hinges on the idea of a Bernstein polynomial.

2.3 Bernstein Polynomials and the Central Limit Theorem

The Bernstein polynomial was first described and used by S. Bernstein in [1] for a constructive proof of the Weierstrass Approximation Theorem (a theorem we will examine and prove in the following section).

Now, these polynomials are pervasive through mathematics. They are critically important in approximating continuous functions and have found a use in computer graphics in the form of Bézier curves. Our use of these polynomials will be in approximating the finite configurational temperature of our system.

We define the Bernstein polynomial now [4].

Definition 3 (Bernstein Polynomial)**.** *A polynomial of the form*

$$B_{n,k}(x) = \binom{n}{k}x^k(1-x)^{n-k},\ \ k \in \{0,...,n\} \tag{13}$$

This polynomial is very reminiscent of the probability mass function of a binomial random variable X with parameters n trials, k successes, and a probability of success x, which is defined as

Definition 4 (Binomial Distribution)**.**

$$f(x) = \binom{n}{k} p^x q^{n-x} \tag{14}$$

where p is the probability of success for each trial and $q = 1 - p$, or the probability of failure for each trial. This implies the polynomial is always nonnegative (as probability distributions are always nonnegative) and implies that the sum over the entire polynomial is 1, for a fixed x.

$$\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = 1 \tag{15}$$

Some special values related to the binomial random variable are the expectation and variance. These can come in handy when proving things using these polynomials. The expectation is defined as

$$E(B_{n,k}(x)) = \sum_{k=0}^{n} k B_{n,k}(x) = nx \tag{16}$$

while the variance is

$$\sigma^2 = nx(1-x) \tag{17}$$

These values come from the standard definition of the expectation and variance of a random variable. There's an interesting point, see [4] for example. By manipulating equations (16) and (17), we can see

$$\sum_{k=0}^{n} \frac{k}{n} B_{n,k}(x) = x \tag{18}$$

and

$$\sum_{k=0}^{n} \left(\frac{k}{n} - x\right)^2 B_{n,k}(x) = \frac{x(1-x)}{n} \tag{19}$$

This is akin to assigning $B_{n,k}$ as the probability to the 'events' $\frac{k}{n}$ and taking the expectation and variance. These equations will be more useful to us.

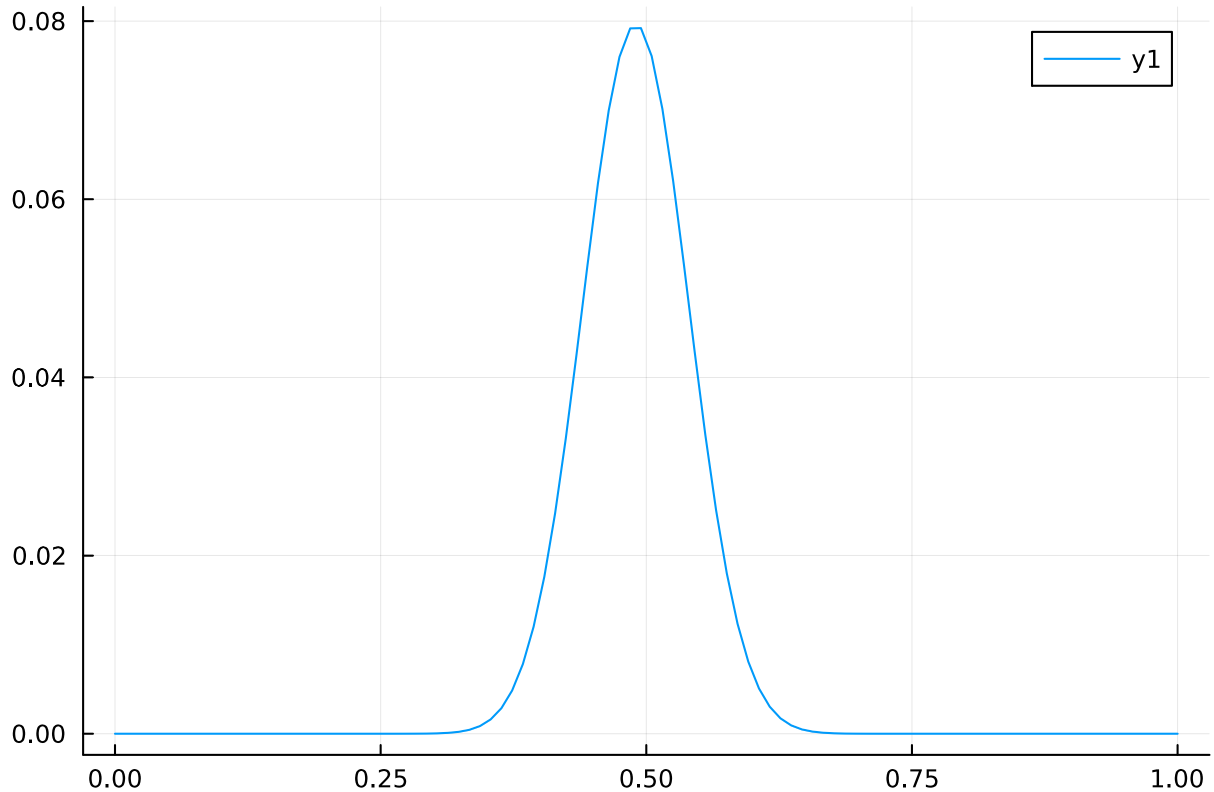
There is one last topic regarding Bernstein polynomials that will be instrumental in providing a proof for the main theorem. We need to find a limiting approximation for a Bernstein polynomial when n is large and, more specifically, when $k \approx \frac{n}{2}$. Let us examine an example of such a case by letting $n = 100$ and $k = 49$, using some code to plot the function.

2.3.1 Numerical Simulations

Consider the following Julia program:

```
using Plots
bernstein_poly(x) = binomial(BigInt(100), BigInt(49)) * x^49 * (1-x)^(100-49)
x = range(0, 1, length=100)
y = bernstein_poly.(x)
p = plot(x,y)
savefig(p, "graph.pdf")
```

This will output the following plot:



A visual examination of this figure suggests the Bernstein polynomial may be approximated well (for large n and $k \approx \frac{n}{2}$) with a Gaussian function, perhaps something of the form of the normal distribution:

Definition 5 (Normal Distribution)**.**

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{20}$$

For this probability distribution, the mean is μ and the variance is σ^2. We may also write this as $N(\mu, \sigma^2)$.

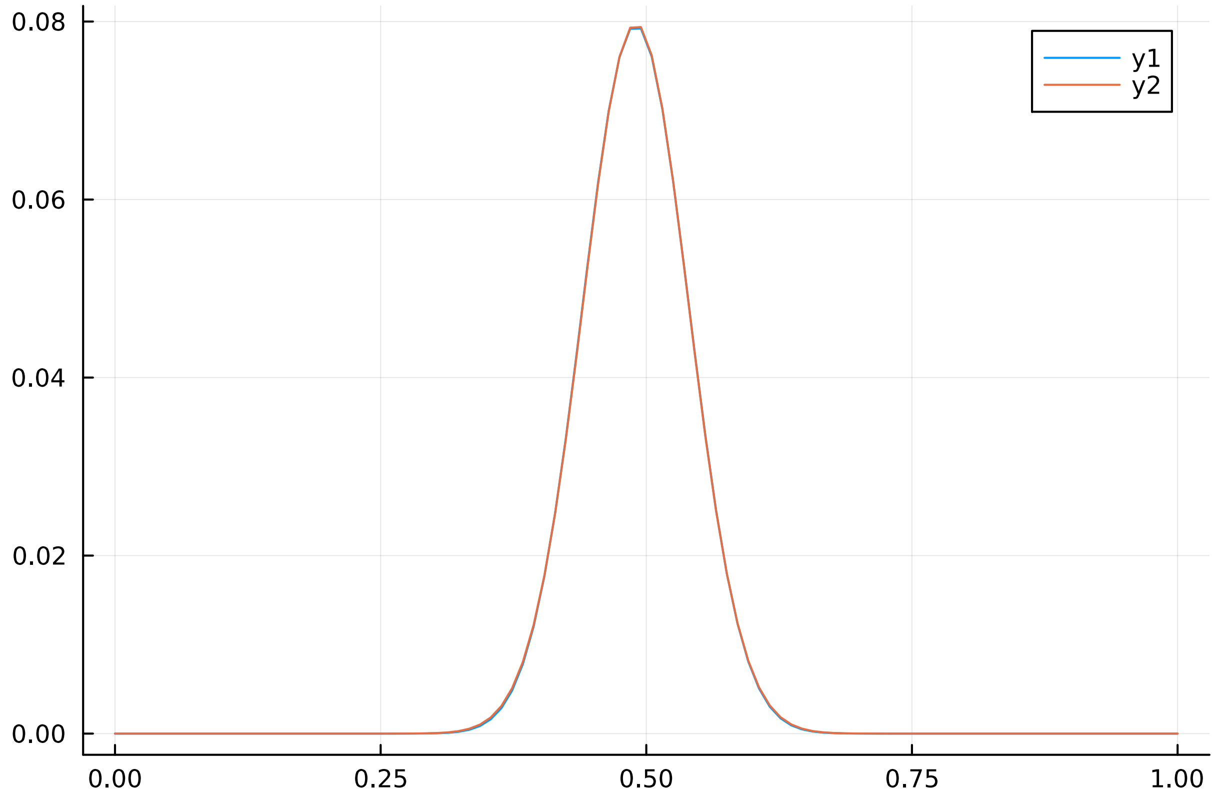
We did not choose this at random. Recall how the definition of the Bernstein polynomial given in (13) resembles the probability mass function of a binomial random variable X with parameters n trials and k successes, with a probability of success x. A very useful approximation of this binomial random variable is with a normal distribution with mean nx and variance $nx(1-x)$. We may be able to approximate our Bernstein polynomials using this theorem. There is a small caveat, however. The Bernstein polynomial is not a function of the number of successes k. It instead describes the probability of achieving k successes out of n trials as a function of the probability of one success. This means the x in the normal distribution function above will be fixed at k and μ and σ will both be functions of p in our Bernstein polynomials.

To provide some evidence to our proposed approximation, we may fit a curve to our data in the program above:

```
using Plots
using LsqFit
bernstein_poly(x) = binomial(BigInt(100), BigInt(49)) * x^49 * (1-x)^(100-49)
x = range(0, 1, length=100)
y = bernstein_poly.(x)
p = plot(x,y)
@. model(x,p) = p[1] * exp(-(x-p[2])^2 / (2*p[3]^2))
```

```
p0 = [0.5, 0.5, 0.5]
fit = curve_fit(model, x, y, p0)
yfit = model(x, fit.param)
plot!(x, yfit)
savefig(p, "graph.pdf")
```

This will produce the following plot:



The model we chose clearly fits the data (nearly) perfectly, so we likely have the right function. We may see the coefficients the curve fit function generated with

```
print(fit.param)
```

This will output $[0.07975379835767381, 0.49009950111114275, 0.0496174134291032 45]$. Interpreting this, we see the curve height is 0.0798, the mean is 0.49, and the standard deviation is 0.0496. We can notice (with some prior knowledge of the mean and standard deviation of the binomial distribution) that $0.0798 \approx \frac{1}{\sqrt{2\pi k(1-\frac{k}{n})}}$, $0.49 \approx \frac{k}{n}$, and $0.0496 \approx \frac{\sqrt{k(1-\frac{k}{n})}}{n}$.

2.3.2 The Central Limit Theorem

As mentioned before, we are able to approximate the binomial distribution with a normal distribution. This is proven in the Central Limit Theorem, which we will state now (as provided in [2]):

Theorem 2 (Central Limit Theorem). *Let $X_1, ..., X_n$ be n identical and independently distributed random variables such that the mean of X_i is μ and the variance is σ^2. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then*

$$\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \quad \textit{as } n \to \infty$$

We may apply this to a binomial random variable X by noting that X is the sum of n identical Bernoulli trials with probability of success p. We can write $X = X_1 + ... + X_n$, where each X_i is the Bernoulli trial mentioned before. Then since each X_i has mean $\mu = p$ and variance $\sigma^2 = pq$, the central limit theorem tells us that $X \sim \mathcal{N}(np, npq)$ when n is sufficiently large.

This approximation of the Bernstein polynomial will be critical in understanding how to prove that the temperature fluctuations tend to 0 as the temperature tends to 0. If we are to apply the normal approximation to the general Bernstein polynomial outright, we get

$$b_{\text{approx}}(p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}} \tag{21}$$

We can note that due to the extremely rapid decay of the exponential term away from $x \approx \frac{k}{n}$, we can very accurately approximate this function with the following:

$$b_{\text{approx}}(p) \approx \frac{1}{\sqrt{2\pi k(1-\frac{k}{n})}} e^{-\frac{(k-np)^2}{2k\left(1-\frac{k}{n}\right)}} \tag{22}$$

Note that now the scaling factor is exactly what we calculated numerically above. And by simply factoring an n^2 out of the numerator of the fraction in the exponential, we see the linear shift of this function is $\frac{k}{n}$ and the denominator is $2\frac{\sqrt{k(1-\frac{k}{n})}}{n}$. Since all of these constants are exactly what we calculated with our curve fit, both the theory and the code agree with our suspicion. This means we can approximate the Bernstein polynomials with the b_{approx} function. We will redefine the b_{approx} function. We get

$$B_{n,k}(p) \sim b_{\text{approx}}(p) := \frac{1}{\sqrt{2\pi k(1-\frac{k}{n})}} e^{-\frac{(k-np)^2}{2k\left(1-\frac{k}{n}\right)}} \tag{23}$$

With this approximation, we have one last theorem to discuss before we can prove the main theorem.

2.4 Continuity and a Proof of the Weierstrass Approximation Theorem

Before we can arrive at the incredibly interesting Weierstrass approximation theorem, we have to understand one of the most fundamental topics in mathematics: continuity. From a high-level point of view, a function is described as ”continuous” if the outputs of the function are somewhat close when the inputs are somewhat close. A more precise definition is:

Definition 6 (Pointwise Continuity)**.** *A function $f(x)$ is continuous at a point x_0 if for every $\epsilon > 0$, there exists a $\delta(\epsilon)$ such that $|f(x) - f(x_0)| < \epsilon$ when $|x - x_0| < \delta(\epsilon)$.*

However, this definition only defines continuity at a specific point. Many functions, including the functions we will examine, are continuous on entire sets of points. So we have continuity on a set:

Definition 7 (Continuity on a set)**.** *A function $f(x)$ is continuous on a set S if for all $s \in S$, $f(x)$ is pointwise continuous at s.*

Now that we understand these concepts, we may arrive at the Weierstrass theorem.

The Weierstrass Approximation theorem is an incredibly important and theoretically rich theorem. It was originally developed and proven by K. Weierstrass in the journal *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*. Later in the paper published by Bernstein [1], this theorem was proven using a probabilistic lens with the Bernstein polynomial.

This theorem is not only theoretically interesting, but has countless applications in real-world problems. One of the most important and wide-spread uses of this theorem is in polynomial interpolation. Many continuous functions that may have complicated analytic forms can be approximated by polynomials to a high degree of accuracy. Due to the fact that polynomials are easy for computers to directly evaluate, approximating these continuous functions using the Weierstrass theorem speeds up computations. As discussed previously, because Bernstein provided a constructive proof of the theorem [1], we have a direct way to find the polynomials we can use to approximate such functions.

2.4.1 Proof of the Weierstrass Approximation Theorem

With the information we now have, we are sufficiently equipped to prove the Weierstrass approxmation theorem. The following proof is given in a note by M. Loss [4]:

Theorem 3 (Weierstrass Approximation Theorem)**.** *Let* $f : [0, 1] \to \mathbb{R}$ *be a continuous function. Define the polynomial*

$$B_n(f)(x) := \sum f\left(\frac{k}{n}\right) B_{n,k}(x). \tag{24}$$

Then for any $\epsilon > 0$*, there exists* N *such that* $\forall n > N$ *and all* $x \in [0, 1]$*,* $|B_n(f)(x) - f(x)| < \epsilon$*.*

This theorem claims this polynomial will approximate a function f to arbitrary accuracy on an interval where f is continuous.

Proof. Since f is continuous on the closed interval, it is uniformly continuous. This means for every $\epsilon > 0$, there exists some $\delta > 0$, depending only on ϵ, such that $|f(x) - f(y)| < \frac{\epsilon}{2}$ when $|x - y| < \delta$. Then consider the following quantity:

$$B_n(f)(x) - f(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) B_{n,k}(x) - f(x) \sum_{k=0}^{n} B_{n,k}(x) \tag{25}$$

$$= \sum_{k=0}^{n} \left[f\left(\frac{k}{n}\right) - f(x)\right] B_{n,k}(x) \tag{26}$$

where the first step follows from the fact that a sum over the entirety of a probability distribution is 1. Then consider splitting the sum at $|\frac{k}{n} - x|$, depending on when it's above and below δ. Our goal is to bound the following distance by some ϵ.

$$|B_n(f)(x) - f(x)| \le \sum_{|\frac{k}{n}-x|<\delta} \left| f\left(\frac{k}{n}\right) - f(x)\right| B_{n,k}(x) + \sum_{|\frac{k}{n}-x|\ge\delta} \left| f\left(\frac{k}{n}\right) - f(x)\right| B_{n,k}(x) \tag{27}$$

This step uses the fact that the polynomial is nonnegative. Now we use the fact that f is uniformly continuous on the interval and the fact that in the second term of equation (27), the distance between any two values of f can't be greater than twice the maximum value of f, $2\max|f(x)|$, to obtain

$$|B_n(f)(x) - f(x)| \le \frac{\epsilon}{2} \sum_{|\frac{k}{n}-x|<\delta} B_{n,k}(x) + 2\max|f(x)| \sum_{|\frac{k}{n}-x|\ge\delta} B_{n,k}(x) \tag{28}$$

So we have

$$\sum_{|\frac{k}{n}-x|\ge\delta} B_{n,k}(x) = \sum_{|\frac{k}{n}-x|\ge\delta} \frac{|\frac{k}{n} - x|^2}{|\frac{k}{n} - x|^2} B_{n,k}(x) \le \frac{1}{\delta^2} \sum_{|\frac{k}{n}-x|\ge\delta} \left|\frac{k}{n} - x\right|^2 B_{n,k}(x) \le \frac{x(1-x)}{n\delta^2} \tag{29}$$

The second to last step uses the uniform continuity of f, while the last step uses the formula for the variance given in equation (19). Then in the distance inequality we are examining, we obtain

$$|B_n(f)(x) - f(x)| \le \frac{\epsilon}{2} \sum_{|\frac{k}{n}-x|<\delta} B_{n,k}(x) + 2\max|f(x)| \frac{x(1-x)}{n\delta^2} \le \frac{\epsilon}{2} + \frac{\max|f(x)|}{2n\delta^2} \tag{30}$$

The last step comes from the fact that the $x(1 - x) \le \frac{1}{4}$ on the interval $[0, 1]$. So by choosing $n > \frac{\max|f(x)|}{\epsilon\delta^2}$, we arrive at

$$|B_n(f)(x) - f(x)| \le \frac{\epsilon}{2} + \frac{\max|f(x)|}{2n\delta^2} \le \epsilon \tag{31}$$

yielding the theorem. □

This theorem is quite interesting. The concept of approximating any continuous function with a polynomial is very enticing, especially in physics. To apply this to our problem, we need to make some modifications, due to the apparent discontinuity in the definition of the configuational temperature (8).

3 Landau's Temperature Fluctuation Law

We can now mathematically show the configurational temperature of a finite N particle system will tend to T_{can} as $N \to \infty$. We will do this by slightly modifying the proof we used above for the Weierstrass Approximation Theorem. This provides a solution to the problem described by Landau.

Before we provide a proof of the main theorem, let us first provide some motivation for the method that is used to prove it. The method we use is by defining a function $\gamma(x) = \left(\frac{2\epsilon}{k}\left(\ln\left(\frac{1}{x}-1\right)\right)^{-1}\right)^m$, and then proving the following statement:

$$\forall \epsilon > 0,\ \exists n_0 \text{ s.t. } |\overline{T^m}(p,N) - \gamma(p)| < \epsilon,\ \forall N > n_0 \tag{32}$$

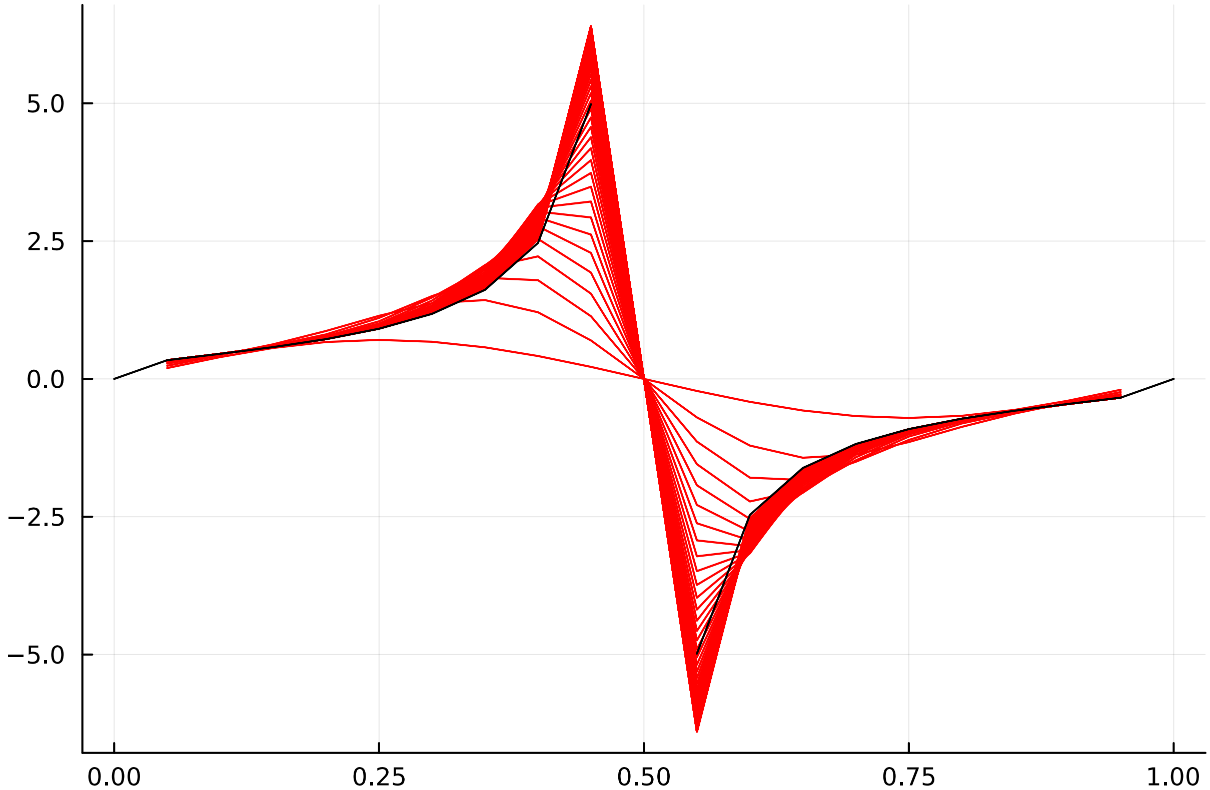
Using the definition of the limit we have provided in the introduction, it's clear that this will prove our theorem, as $\gamma(p) = (T_{\text{can}})^m$. To provide some evidence to show this method is reasonable, let us create some Julia code to examine the difference between $\overline{T^m}(p,N)$ and $\gamma(p)$ for various N.

3.1 Numerical Simulations

Consider the following code:

```
using Plots
using SpecialFunctions

function T(n,N)
        if n == 0 || n == N
                return 0
        else
                return 1 / log((N-n)/n)
        end
end
function Tm(p,N,m)
        result = 0.0
        for n in 0:N
                Tn = T(n,N)^m
                log_binom = lgamma(N+1) - lgamma(n+1) - lgamma(N-n+1)
                log_prob = n*log(p)+(N-n)*log(1.0-p)
                log_term = log_binom + log_prob

                term = exp(log_term) * Tn
                result += term
        end
        return result
end
gamma(x,m) = (1 / log((1/x)-1))^m

xs = 0:0.05:1
gs = gamma.(xs, Ref(1))

ns = [i for i in 1:1001 if i % 2 != 0 && i % 5 == 0]
lines = [Tm.(xs, Ref(i), Ref(1)) for i in ns]

p = plot(xs, lines[1], legend=false, color=:red)
for line in lines[2:end]
        plot!(xs, line, color=:red)
end
plot!(xs, gs, color=:black)
savefig(p, "tn-plot.pdf")
```

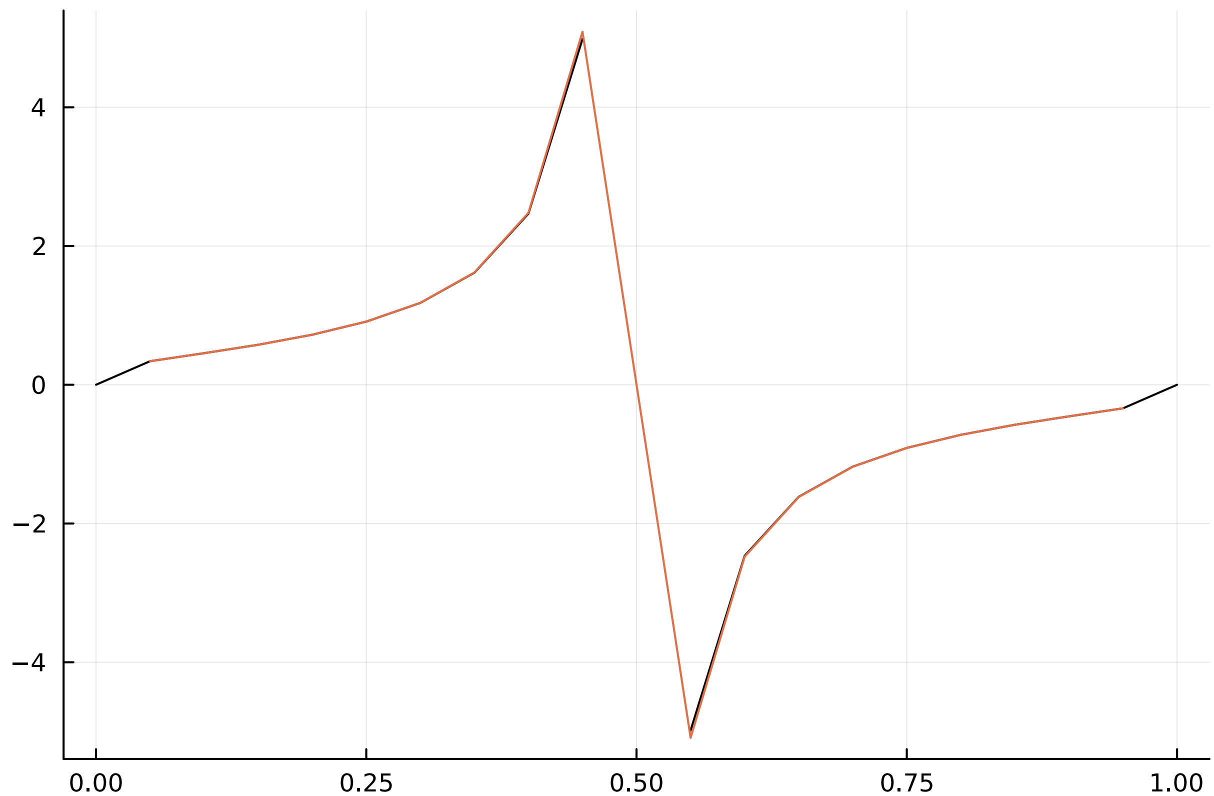
This will produce the following plot:



The black line pictured is our function γ and the numerous red lines are $\overline{T^m}(p, N)$ for various values of N, specifically odd multiples of 5 from 1 to 1001. We can see that these red lines seem to converge to the black line, however our maximum N is still quite small. Let's now produce a plot with a higher N value. Continuing our code from before:

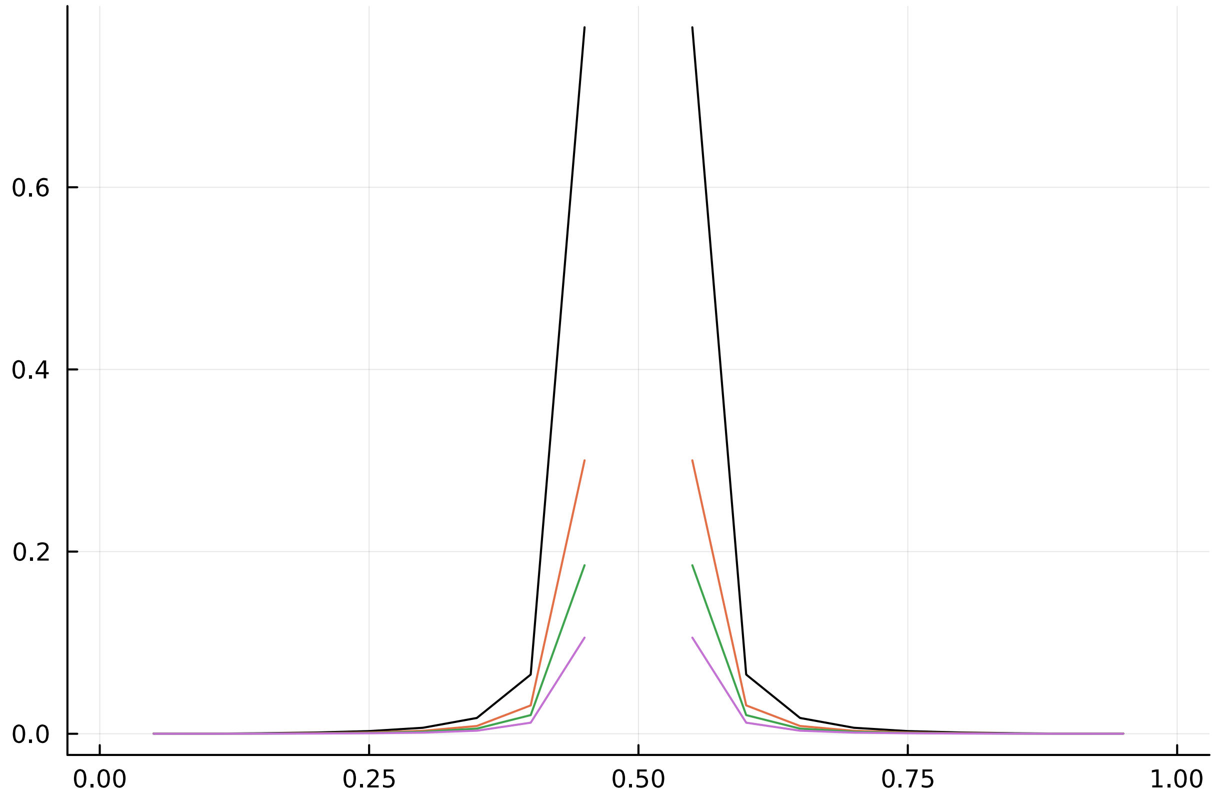
```
ts = Tm.(xs, Ref(5001), Ref(1))
p = plot(xs, gs, legend=false, color=:black)
plot!(xs, ts)
savefig(p, "tn-plot-converged.pdf")
```

From this we get the following:



Here, the black line is still γ and the orange line is $\overline{T^m}(p, 5001)$. We can see with a much higher N value, this $\overline{T^m}(p, N)$ function is nearly exactly our γ. This is providing some significant evidence that our claim is true. Let us finally examine the function $|\overline{T^m}(p, N) - \gamma(p)|$ for various N to see the exact difference between the two functions:

```
a(p,N,m) = abs(Tm(p,N,m) - gamma(p,m))
as = [a.(xs,Ref(i),Ref(1)) for i in [1001, 2001, 3001, 5001]]
p = plot(xs, as[1], legend=false)
for line in as[2:end]
        plot!(xs, line)
end
savefig(p, "tn-plot-converged-distances.pdf")
```



Each N value corresponds to a line in descending order (1001 is the black line and 5001 is the purple line). It's easy to see the distances tend to 0 (except around the singularity, where strange things happen), but we now have sufficient motivation to pursue this method of proof.

3.2 Proof of the Main Theorem

Proof of Main Theorem. Let's define the function γ as follows:

$$\gamma(x) = \left(\frac{2\epsilon}{k}\left(\ln\left(\frac{1}{x} - 1\right)\right)^{-1}\right)^m \tag{33}$$

This function has three points of discontinuity: 0, $\frac{1}{2}$, and 1. Since $\lim_{x\to 0^+}\gamma(x) = 0$ and $\lim_{x\to 1^-}\gamma(x) = 0$, the discontintuities at 0 and 1 are removable. Let's define γ to be 0 at these points. The discontinuity at $\frac{1}{2}$ is not removable, since both the left and right sided limits of γ do not exist. So consider a closed sub-interval $[a, b] \subset [0, 1]$ such that γ is continuous on $[a, b]$. This amounts to considering a closed sub-interval of $[0, 1]$ that does not contain $\frac{1}{2}$.

Let's consider the distance $|\overline{T^m}(p, N) - \gamma(p)|$ where $p \in [a, b]$. We know

$$|\overline{T^m}(p, N) - \gamma(p)| = \left|\sum_{n=0}^{N}\left(T(n, N)^m\binom{N}{n}p^n(1-p)^{N-n}\right) - \gamma(p)\sum_{n=0}^{N}\binom{N}{n}p^n(1-p)^{N-n}\right| \tag{34}$$

$$= \left|\sum_{n=0}^{N}\left(T(n, N)^m B_{N,n}(p)\right) - \gamma(p)\sum_{n=0}^{N}B_{N,n}(p)\right| \tag{35}$$

$$\leq \sum_{n=0}^{N}\left|T(n, N)^m - \gamma(p)\right| B_{N,n}(p) \tag{36}$$

where the third line follows from the fact that Bernstein polynomials are strictly non-negative. Now since γ is continuous at p, we know for every $\epsilon > 0$, there exists some $\delta > 0$ such that $|\gamma(x) - \gamma(p)| < \frac{\epsilon}{2}$ when

$|x-p|<\delta$. Let's split (36) according to when $|\frac{n}{N}-p|$ is less than or greater than δ.

$$|\overline{T^m}(p,N)-\gamma(p)| \leq \sum_{|\frac{n}{N}-p|<\delta} \left|\gamma\left(\frac{n}{N}\right)-\gamma(p)\right| B_{N,n}(p) + \sum_{|\frac{n}{N}-p|\geq\delta} |T(n,N)^m-\gamma(p)|\, B_{N,n}(p) \tag{37}$$

Now, by appealing to the continuity of γ in our interval $[a,b]$, the distance in the first sum of (37) will be less than $\frac{\epsilon}{2}$. Now we need to examine the righthand sum of (37). We will break this up into two cases; one where $|\frac{n}{N}-\frac{1}{2}|<\delta$ and one where $|\frac{n}{N}-\frac{1}{2}|\geq\delta$. Let's first assume $|\frac{n}{N}-\frac{1}{2}|<\delta$. We note the following asymptotic behavior: when $N\to\infty$,

$$B_{N,n}(p) \sim b_{\text{approx}}(p) = \frac{1}{\sqrt{2\pi n(1-\frac{n}{N})}} e^{-\frac{(Np-n)^2}{2n(1-\frac{n}{N})}} \tag{38}$$

as a result of the normal distribution approximation to the binomial distribution using the central limit theorem. Using (38), we can see that $|T(n,N)^m|B_{N,n}(p) \sim |T(n,N)^m|b_{\text{approx}}(p)$. To see another important fact, notice that because δ was chosen with respect to p, the distance between p and $\frac{1}{2}$ will always be greater than δ. So when $|\frac{n}{N}-\frac{1}{2}|<\delta$, $T(n,N)^m$ will be very close to the singularity, and $|T(n,N)^m|>|\gamma(p)|$. This means that $|T(n,N)^m-\gamma(p)|\leq 2|T(n,N)^m|$. Additionally, if $T(n,N)$ is 0, then (38) dominates $\gamma(p)$. Let's split the right hand sum in (37) into the following:

$$\begin{aligned}\sum_{|\frac{n}{N}-p|\geq\delta} |T(n,N)^m-\gamma(p)|\, B_{N,n}(p) = &\sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|<\delta}} |T(n,N)^m-\gamma(p)|\, B_{N,n}(p) \\ &+ \sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|\geq\delta}} |T(n,N)^m-\gamma(p)|\, B_{N,n}(p)\end{aligned} \tag{39}$$

Now utilizing the facts above, we can transform this first portion in (39).

$$\sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|<\delta}} |T(n,N)^m-\gamma(p)|\, B_{N,n}(p) \leq 2 \sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|<\delta}} |T(n,N)^m|\, b_{\text{approx}}(p) \tag{40}$$

So we have this $|T(n,N)^m|b_{\text{approx}}(p)$ term in the sum. In the summands, we have $|\frac{n}{N}-p|\geq\delta$, so $b_{\text{approx}}(p)\leq \frac{1}{\sqrt{N}}e^{-2N\delta^2}$, where this inequality comes from the following and the fact that because $\frac{n}{N}$ is close to $\frac{1}{2}$, $\frac{n}{N}(1-\frac{n}{N})$ is close to $\frac{1}{4}$:

$$e^{-\frac{(Np-n)^2}{2n(1-\frac{n}{N})}} = e^{-\frac{(N(p-\frac{n}{N}))^2}{2n(1-\frac{n}{N})}} \tag{41}$$

$$\leq e^{-\frac{(N\delta)^2}{2n(1-\frac{n}{N})}} \tag{42}$$

$$= e^{-\frac{N\delta^2}{2\frac{n}{N}(1-\frac{n}{N})}} \tag{43}$$

Due to the restriction $|\frac{n}{N}-\frac{1}{2}|<\delta$, we have $T(n,N)^m\leq(\ln(\frac{2N}{N\pm1}-1))^{-m}$ as the worst term. We can also see that (40) has around δN terms. This means we have:

$$2\sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|<\delta}} |T(n,N)^m|\, b_{\text{approx}}(p) \leq 2N\delta\left(\ln\left(\frac{2N}{N\pm1}-1\right)\right)^{-m}\frac{1}{\sqrt{N}}e^{-2N\delta^2} \tag{44}$$

We can see that as $N\to\infty$, the right hand side of (44) tends to 0, since the exponential term dominates the logarithmic term. So the first sum in the right hand side of (39) can be ignored. This means we get the following

$$\sum_{|\frac{n}{N}-p|\geq\delta} |T(n,N)^m-\gamma(p)|\, B_{N,n}(p) \to \sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|\geq\delta}} |T(n,N)^m-\gamma(p)|\, B_{N,n}(p), \quad N\to\infty \tag{45}$$

Since we are working in a closed interval where γ is continuous, the extreme value theorem will tell us that γ is bounded above by some M. So the distance in the right-hand side of (45) is guaranteed to be less than $2M$. From this we achieve

$$|\overline{T^m}(p,N) - \gamma(p)| \leq \sum_{|\frac{n}{N}-p|<\delta} |T(n,N)^m - \gamma(p)|\, B_{N,n}(p) + \sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|\geq\delta}} |T(n,N)^m - \gamma(p)|\, B_{N,n}(p) \tag{46}$$

$$\leq \frac{\epsilon}{2} \sum_{|\frac{n}{N}-p|<\delta} B_{N,n}(p) + 2M \sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|\geq\delta}} B_{N,n}(p) \tag{47}$$

Now examining the second sum in (47), we can perform the following manipulation:

$$\sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|\geq\delta}} B_{N,n}(p) = \sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|\geq\delta}} \frac{|\frac{n}{N}-p|^2}{|\frac{n}{N}-p|^2} B_{N,n}(p) \tag{48}$$

$$\leq \frac{1}{\delta^2} \sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|\geq\delta}} \left|\frac{n}{N}-p\right|^2 B_{N,n}(p) \tag{49}$$

$$\leq \frac{p(1-p)}{N\delta^2} \tag{50}$$

The last step follows from substituting the variance given in (19). Now we can see

$$|\overline{T^m}(p,N) - \gamma(p)| \leq \frac{\epsilon}{2} \sum_{|\frac{n}{N}-p|<\delta} B_{N,n}(p) + 2M \sum_{\substack{|\frac{n}{N}-p|\geq\delta \\ |\frac{n}{N}-\frac{1}{2}|\geq\delta}} B_{N,n}(p) \leq \frac{\epsilon}{2} + \frac{2Mp(1-p)}{N\delta^2} \tag{51}$$

Since the quantity $p(1-p)$ is always less than $\frac{1}{4}$ on the interval $[0,1]$, this must also be true for the sub-interval $[a,b]$. So

$$|\overline{T^m}(p,N) - \gamma(p)| \leq \frac{\epsilon}{2} + \frac{2Mp(1-p)}{N\delta^2} \leq \frac{\epsilon}{2} + \frac{M}{2N\delta^2} \tag{52}$$

So by choosing $n_0 > \frac{M}{\epsilon\delta^2}$, we can see that when $N > n_0$,

$$|\overline{T^m}(p,N) - \gamma(p)| \leq \frac{\epsilon}{2} + \frac{M}{2N\delta^2} \leq \epsilon \tag{53}$$

□

This means we arrive at the following statement:

$$\forall\epsilon > 0,\ \exists n_0 \text{ s.t. } |\overline{T^m}(p,N) - \gamma(p)| < \epsilon,\ \forall N > n_0 \tag{54}$$

This precisely means $\lim_{N\to\infty} \overline{T^m}(p,N) = \gamma(p) = (T_{\text{can}})^m$ for p in the sub-interval $[a,b]$. Since this is an arbitrary sub-interval of $[0,1]$ not including $\frac{1}{2}$, we can say $\overline{T^m}(p,N)$ converges to $\gamma(p) = (T_{\text{can}})^m$ on the set $[0,\frac{1}{2}) \cup (\frac{1}{2},1]$. The influence of the proof of the Weierstrass Approximation Theorem is apparent. The main change that is required is examining closed sub-intervals of the interval where γ is defined. This proof also gives insight into why the temperature fluctuations tend to 0 as N grows large. The configuational temperature of the system A, as the number of particles grows large, will tend to the temperature of the surroundings. A rewording of the last statement of the above proof can describe this more clearly:

$$\forall\epsilon > 0,\ \exists n_0 \text{ s.t. } |\overline{T}(p,N) - T_{\text{can}}| < \epsilon,\ \forall N > n_0,\ m = 1 \tag{55}$$

Any fluctuation in temperature will be reflected in the value $\overline{T}(p,N)$. Since we have determined that the distance between this temperature and the temperature of the surroundings T_{can} must be bounded, any fluctuation in temperature must also be bounded. Furthermore, as $\overline{T}(p,N)$ gets closer to T_{can}, the fluctuations will necessarily get smaller, implying the fluctuations tend to 0 as N grows large.

4 Discussions and Conclusions

The main result of this paper is a mathematical proof of (13), as X. Wang and Q.H. Liu [5] presented. We aimed to reconcile the dissonance between the current models of a spin system and the supposed infinite temperature fluctuations that are predicted. Our proof provides a rigorous connection between the Landau thermodynamic problem and analysis, specifically by using Bernstein polynomials to understand why the temperature fluctuations of a finite particle system tend to 0, even as the temperature of the surrounding heat bath tends to 0 K.

This paper provides a rigorous foundation for the newly proposed ensemble, thereby further validating the previous physics-based work on the subject.

Acknowledgements

We would like to thank X.W. Chen for his invaluable assistance in guiding this paper. We are also grateful to the referees for their careful review of this manuscript and their encouraging comments that have significantly improved this paper.

References

[1] S. Bernstein, *Démonstration du théorème de weierstrass fondée sur le calcul des probabilités*, Communications of the Kharkov Mathematical Society **13** (1912), 1–2, Translated by Michael S. Floater, 2017, available at `https://www.mn.uio.no/math/english/people/aca/michaelf/translations/bernstein_english.pdf`.

[2] J. Cain and L. Yan, *Central limit theorem*, `https://web.stanford.edu/class/archive/cs/cs109/cs109.1212/lectureNotes/LN18_clt.pdf`, Accessed: 2025-10-06.

[3] L.D. Landau and E.M. Lifshitz, *Statistical physics*, 3rd ed., Pergamon Press, Oxford, 1980.

[4] M. Loss, *Weierstrass approximation and bernstein polynomials*, `https://loss.math.gatech.edu/16FALLTEA/NOTES/weierstrassbernstein.pdf`, Accessed: 2024-11-02.

[5] X. Wang and Q.H. Liu, *Temperature fluctuations for a finite system of classical spin-1/2 particles*, Annals of Physics **322** (2007), 2168–2178.