Proceedings of the 3rd International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2025.DL30179

The Stirling Formula and the Law of Large Numbers Analysis and Application in Probability Analysis

Miluo Wang[1], Peizhe Li[2*]

[1]The Stony Brook School, New York, USA
[2]Chengdu No.7 High School International Department, Chengdu, China
**Corresponding Author. Email: Lipeizhe1108@163.com*

Abstract. We derive the Stirling formula and the Law of Large Numbers (LLN) and provide step-by-step proofs and analysis for both formula and law, along with their applications in probability and statistics. Stirling's formula provides an asymptotic approximation for factorials, which is crucial for handling large-number computations in combinatorics and probability theory. Additionally, we provide a proof of Stirling's formula using the Gamma function and Laplace's method, discuss its convergence properties, and introduce the Law of Large Numbers with proofs for both the Weak and Strong versions. We also demonstrate the synergy between these two mathematical tools through applications in binomial concentration and asymptotic analysis of central binomial coefficients. Finally, we illustrate how these methods can be applied to real-world statistical problems, especially referring to the probability calculation and analysis.

Keywords: Stirling Formula, Law of Large Numbers, Asymptotic Analysis, Binomial Distribution, Probability Convergence

1. Introduction

Stirling formula aims to solve relatively large number calculations. In any realm of study, estimation of the extremely big numbers or their factorials can cause inaccuracies which affect the research's results. For example, the Binomial calculation, which can be applied in the calculation of probability. Using this formula:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Also, the combinatorics formula:

$$\binom{n}{k} = \frac{n!}{(n-k)!}$$

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.DL30179

While calculating the distribution, the combinatorics can be extremely enormous depending on different research scenarios. When n grows very big, the two factorials in this formula can be very big which leads to uncertainty and inaccuracy.

So, in order to get precise results during research, Stirling Formula can be effective in calculating these enormous factorials. This is the formula:

$$n! \approx \sqrt{2\pi n\left(\frac{n}{e}\right)^n}$$

We will later talk about how to derive this equation. But for now, we will discuss some basic features of this equation. When estimating greater numbers, this equation is useful because the percentage difference is small. However, if using it to estimate small numbers of factorials the percentage difference will be very large.

The Stirling formula can also combine with other tools or equations of estimating large numbers. For example, the Large Number Theory. Because that Stirling formula is limited to estimating only the factorials to some extent, other tools may be necessary for further and broader steps of estimation.

2. Proof of Stirling's formula

2.1. Introduction

Stirling's formula provides an asymptotic approximation for factorials and is widely used in probability, statistics, and combinatorics. This paper represents detailed proof of Stirling's formula:

As

$$n \to \infty, n! \ \frac{n^n}{e^n}\sqrt{2\pi n} \tag{1.1}$$

That is,

$$\frac{n!}{\left(n^n/e^n\right)\sqrt{2n\pi}} = 1 \tag{1.2}$$

The proof relies on the Gamma function, Laplace's method, and asymptotic analysis [1].

The factorial function n! grows rapidly with n, making exact computation impractical for large n. Stirling's approximation offers an efficient estimate that is crucial in many mathematical applications. For example, in probability theory, it is used to approximate binomial coefficients for large n, and in statistical mechanics, it helps in deriving the entropy of ideal gases.

This paper presents a rigorous proof using [2]:

1.Gamma function representation of n!.

2.Laplace's method for approximating integrals.

3.Asymptotic expansions to refine the approximation.

DOI: 10.54254/2753-8818/2025.DL30179

2.2. 2 preliminaries

2.2.1. Gamma function representation

The factorial can be expressed via the Gamma function:

$$n! = \Gamma(n+1) = \int_0^\infty x^n e^{-x} dx \tag{2.1}$$

2.2.2. Laplace's method

Laplace's method approximates integrals of the form:

$$\int_a\hat{}\,b\,e\hat{}\,\left(Mf(x)\right)\,dx\ for\ large\ M \tag{2.2}$$

Where f(x) has a unique maximum at x_0 . The method involves expanding f(x) around x=0 and evaluating a Gaussian integral.

2.3. Proof of Stirling's formula

2.3.1. Rewriting the integral

Let x=n(1+t), then:

$$n! \ = \ n^{n+1}e^{-n}\int_{-1}^{\infty}(1+t)^n e^{-nt}dt \tag{3.1}$$

2.3.2. Taylor expansion and substitution

Taking logarithms and expanding:

$$ln\left(1+t\right) = t - \frac{t^2}{2} + O\left(t^3\right)$$

so:

$$(1+t)^n e^{-nt} = e^{nln(1+t)-nt} \ \approx \ e^{-nt^2/2}$$

Substitute $t = s/\sqrt{n}$, leading to:

$$n! \approx n^{n+1}e^{-n}\int_{-\infty}^{\infty} e^{-s^2/2}ds \ \cdot \ \frac{1}{\sqrt{n}} \tag{3.2}$$

2.3.3. Gaussian integral evaluation

Proof of Gaussian Integral:

Square the Integral:

DOI: 10.54254/2753-8818/2025.DL30179

Let I = $\int_{-\infty}^{\infty} e^{-x^2} dx$ .Then,

$I^2 = \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)} dxdy$.

Switch to Polar Coordinates:

Let

$x = rcoscos\,\theta\,, y = rsinsin\,\theta\,,\ so\ dxdy = rdrd\theta,\ and\ x^2 + y^2 = r^2. The\ limits\ become\ r \in [0, \infty), \theta \in [0,2\pi]$.

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2} r dr d\theta.$$

Evaluate the Radial Integral:

Let $u = r^2,\ so\ du = 2rdr,\ and\ rdr\ =\ du/2.\ Then,$

$$\int_0^{\infty} e^{-r^2} r dr\ =\ \int_0^{\infty} e^{-u} \frac{u}{2} = \frac{1}{2} \int_0^{\infty} e^{-u} du = \frac{1}{2}$$

Evaluate the Angular Integral:

$$I^2 = \int_0^{2\pi} (\tfrac{1}{2}) d\theta = \tfrac{1}{2} \cdot 2\pi = \pi$$

So,

I= $\sqrt{\pi}$

The integral evaluates to:

$\int_{-\infty}^{\infty} e^{-s^2/2} ds = \sqrt{2\pi}$,

yielding:

$$n!\sim\sqrt{2\pi n}\left(\tfrac{n}{e}\right)^n \tag{3.3}$$

2.3.4. Refinement with higher-order terms (optional)

A more precise expansion includes correction terms:

$$n! = \sqrt{2\pi n}\left(\tfrac{n}{e}\right)^n \left(1 + \tfrac{1}{12n} + O\left(n^{-2}\right)\right) \tag{3.4}$$

This refined form can be derived using techniques such as the Euler–Maclaurin formula or saddle-point methods.

Example: For n=10, the simple Stirling approximation gives 10!≈3598695.6, while the refined version gives 10!≈3628509.7, compared to the exact value 3628800. The relative error decreases significantly with the refined approximation.

DOI: 10.54254/2753-8818/2025.DL30179

2.4. Conclusion

Stirling's formula is proven using the Gamma function and Laplace's method, confirming its asymptotic behavior. The approximation is fundamental in many fields, such as:

Probability: Approximating binomial coefficients in large trials.

Statistical Mechanics: Deriving the entropy of an ideal gas.

Combinatorics: Estimating the number of permutations for large sets.

The method demonstrated here highlights the power of asymptotic analysis and integral approximation techniques in deriving widely used mathematical formulas.

3. Large number law

3.1. Introduction

The Law of Large Numbers, a pivotal formula in the realm of Statistics, which studies probability. It describes in repeated independent trials, the sample means of a random event will approach its theoretical expected value as the number of trials increases. A phenomenon described by the (LLN) basic theorem in statistics and probability theory: the average of the results of repeated experiments or observations will settle down and converge toward the predicted value. In other words, over time, randomness will "average out" and the predictable pattern will be clearer and clearer.

Flipping a coin is a proper example: a single run of flipping may result in a skewed heads-to-tails ratio. Some special and unpredictable results may occur. However, if we flip the coin thousands or millions of times, the probability of heads, for example, will gradually converge to 0.5. According to this rule, this phenomenon can also be applied to other realms of study. For instance, it is the basis for everything from financial modeling and insurance actuarial calculations to factory quality control and machine learning algorithm efficacy.

There are two types of LLN: Weak LLN and Strong LLN. Their functions are mostly similar, but some distinctions exist, and their conditions vary to some extent. Weak LLN, focusing on the sample average, converges in probability toward the expected value as the sample size increases. Strong LLN is a stronger version of Weak LLN, which ensures exact convergence. However, using Strong LLN means that the mean must be finite, and variables must be independent, identically distributed.

3.2. Analysis

Before going deep toward what Weak and Strong LLN look like, there is a general agreement between these two versions that the sample mean calculation is

$$\underline{X}_n = \frac{1}{n}\left(X_1 + X_2 + X_3 + \ldots + X_n\right)$$

And both of them converge to exact value:

$$(\underline{X}_n) = \mu$$

DOI: 10.54254/2753-8818/2025.DL30179

Another important concept is variance. In Statistics, it indicates how far the data with the mean or the median. During probability calculations, we assume identical finite variance has no connection to random variables in the data set, which $Var(X_i) = \sigma^2$. If the data set has the mean n, the variance will look like this:

$$Var(\underline{X}_n) = Var\left(\frac{1}{n}(X_1 + X_2 + X_3 + \ldots + X_n)\right)$$

This is equal to

$$\frac{1}{n^2} Var(X_1 + X_2 + X_3 + \ldots + X_n) = \frac{\sigma^2}{n}$$

3.3. Weak law of large numbers

The Weak LLN, also known as Khinchin's Law, describes a set of identically distributed data with finite mean converges to the expected value in probability [3]. Therefore, for any positive number ε , there is:

$$P(|\underline{X}_n - \mu| > \varepsilon) = 0$$

Still taking flipping coins for example, if we flip n times and n tends to infinity, there is almost no probability of seeing more heads than 0.50001 times of n.

3.3.1. Convergence target

As for the proof of the Weak LLN, we set the expectation value as

$$E(\underline{X}_n) = \mu$$

As mentioned before, the variance would be

$$Var(\underline{X}_n) = \frac{\sigma^2}{n}$$

3.3.2. Apply Chebyshev's inequality

By using Chebyshev's Inequalities, which provides conservative data about how much data lies within a certain range of the mean. It has a form $P\left\{\left|X - \mu\right| \geq k\right\} \leq \frac{\sigma^2}{k^2}$, if X has a mean μ , variance σ^2 , and $k > 0$. There will be [4]

DOI: 10.54254/2753-8818/2025.DL30179

$$P\{\geq \varepsilon\} \leq \frac{Var\left(\underline{X}_n\right)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

So, no matter how small ε is, the RHS has the tendency to 0 when n, times of trials increase.

3.4. Strong law of large numbers

The Strong LLN, also known as Kolmogorov's Law, if a data set X_n is a finite mean μ, there will be [5]:

$X_n \to \mu$ a.s.(a.s. is denoted as "almost sure")

In the identically and independent distributed random variables converge almost sure to the expected value, in this case is μ. If more samples are taken, it is going to be closer to the mean almost surely.

3.4.1. Use 4th movement of Chebyshev's inequality for proof

Proof of the Strong LLN relates to the 4th movement of Chebyshev's inequality, which is

$$Pr\left(\left|Z - EZ\right| > \varepsilon\right) \leq \frac{E\left[(Z - EZ)^4\right]}{\varepsilon^4}$$

To make the calculation simpler, we set the mean of the data set equals to 0, and the mean of the data set X_n is μ, which $E(X_n) = \mu = 0$. Our goal is to calculate $E[X4\,n\,]$. We expand it to the 4th power:

$$X4\,n\, = \left(\frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}\right)^4 = \frac{1}{n^4}\left(\sum_i X_i\right)^4$$

When we write $\left(\sum_i X_i\right)^4$, there will be different types of terms. Consider all of them, concluding that

$$E\left[X4\,n\,\right] = \frac{1}{n^4} E\Bigg(\sum_i {X_i}^4 + 4\sum_{i\neq j} {X_i}^3 X_j + 3\sum_{i\neq j} {X_i}^2 {X_j}^2 + 6\sum_{i,j,k} X_i X_j {X_k}^2 + \sum_{i,j,k,l} X_i X_j X_k X_l$$

In this case, X_i to the first power vanishes because X_i are independent and centered at $E(X_i) = 0$. Then any product with unpaired X has expectation value of 0. If this term is eliminated, what else remains?

DOI: 10.54254/2753-8818/2025.DL30179

$$E\left(X_n^4\right)=\frac{1}{n^4}\left[nE\left({X_i}^4\right)+3n\left(n-1\right)E\left({X_j}^2{X_i}^2\right)\right]$$

Simplify the terms, since both X_j and X_i are independent and identically distributed:

$$E\left({X_j}^2{X_i}^2\right)=E\left({X_j}^2\right)E\left({X_i}^2\right)=\sigma^4$$

As mentioned before, $Var\left(X_n\right)=\sigma^2$, so

$$E\left(X_n^4\right)=\frac{1}{n^4}\left[nm_4+3n\left(n-1\right)\sigma^4\right]=\frac{m_4}{n^3}+\frac{3n\left(n-1\right)}{n^3}\sigma^4$$

Because $\frac{n-1}{n^3}\leq\frac{1}{n^2}$, we can get

$$E\left(X_n^4\right)\leq\frac{m_4}{n^3}+\frac{3\sigma^4}{n^2}$$

Then apply the 4th moment Markov's Inequality:

$$Pr\left[X_n-EX_n>\epsilon\right]\leq\frac{\frac{m_4}{n^3}+\frac{3\sigma^4}{n^2}}{\epsilon^4}$$

Now apply the Borel-Cantelli, with fix ϵ value, and let $X_n>\epsilon$, we then have:

$$\sum_{n=1}^{\infty}PrPr\ \left(X_n\right)\ <\infty$$

So, only finite amounts of X_n will appear as the probability is 1. This explains precisely why $X_n\rightarrow 0\ a.s.$ and since we previously assumed $E(X_n)=0$, so X_n converges to its expected value almost surely.

3.5. Synergy between Stirling formula and LLN

The relationship between the law of large numbers and the Stirling formula is that both can concentrate a combinatorial structure with large numbers to an expected value. Using Law of Large Numbers to find the bound and use Stirling Formula to estimate precise asymptotic. For example the Binomial Probabilities, $PrPr\ \left(Binomial\left(n,k\right)=k\right)\ =\left(n\ k\right)p^k(1-p)^{n-k}$. Using Stirling

DOI: 10.54254/2753-8818/2025.DL30179

Formula can calculate big number binomials, and for large n , binomial concentration is strongly concentrated around np . According to the LLN, $\frac{k}{n} \to p\ with\ high\ probability$.

4. Application

4.1. Binomial Concentration

Binomial Concentration, which describes how the outcome behaves when making multiple trials, the outcome of successes. If let X_n~$Bin(n,p)$ and a fix number ε which $0 < \varepsilon < 1$. Prove

$$PrPr\left(\left|\frac{X_n}{n} - p\right| > \varepsilon\right) \to 0\left(n \to \infty\right)$$

Use Weak LLN or Strong LLN, there will be $\frac{X_n}{n} \to p$, which proves the probability formula converges to 0 in probability or almost surely accordingly.

Now, use Striling Formula to estimate for an exact k, and the number k equals to $[qn]$. So

$$\left(n\ k\right) \sim \frac{1}{\sqrt{2\pi nq\left(1-q\right)}} exp\left(nH\left(q\right)\right)$$

In the formula above, the $H\left(q\right)$ is a Binary Entropy, which often cooperates with Stirling Formula to calculate the Binary Coefficient. It measures on uncertainty of each trail, quantifies the "spread" of binomial distribution. $H\left(q\right)$ is equal to

$$H\left(q\right) = -qlogq - (1-q)log(1-q)$$

According to this, there is

$$PrPr\ \left(X_n = k\right) = \left(n\ k\right)p^k(1-k)^{n-k} \sim \frac{1}{\sqrt{2\pi nq\left(1-q\right)}} exp\left(n\left(H\left(q\right) + \left(-qlogq - (1-q)log\left(1-q\right)\right)\right)\right)$$

Let $H\left(q\right) + \left(-qlogq - (1-q)log(1-q)\right) = -D(q||p)$, according to the Kullback–Leibler divergence. So

$$PrPr\ \left(X_n = k\right) = \frac{1}{\sqrt{2\pi nq\left(1-q\right)}} e^{-nD(q||p)}$$

DOI: 10.54254/2753-8818/2025.DL30179

This claims if the situation $p \neq q$, then $D(q||p) > 0$ and the probability decays. So, $PrPr\ (X_n = k)$ will be exponentially small in n . It is generally a formula proving that the binomial probability concentrates on the mean value precisely.

4.2. Asymptotic of central binomial coefficients

Another example may be relevant to the asymptotic of the central binomial coefficients and probability at the mean. Using Stirling Formula can show the result of $(2n\ n\ )$, which is

$$(2n\ n\ ) \sim \frac{4^n}{\sqrt{\pi n}}$$

It describes the asymptotic size, $PrBin\left(2n, \frac{1}{2}\right) = n\right)$. For the times of success trials $p = \frac{1}{2}$: $PrPr\ \left(Bin\left(2n, \frac{1}{2}\right) = n\right)\ = (2n\ n\ ) \times 2^{-2n} = \frac{1}{\sqrt{\pi n}}$, The result shows the decay rate of a result's convergence, depending on the n . This gives a range.

5. Conclusion

Both the Stirling Formula and the Large Number Theory deal with enormous data set problems, which gives fewer differences to the actual value. Striling formula has the function of estimating relatively precise value of factorials, which can be applied in factorial and combinatorics calculation. For example, binomial expansion, concentration, etc. Also assists quantifies the asymptotic of central binomial coefficients as previously mentioned, which gives a clear range. Stirling's Formula enables tractable estimation of binomial coefficients, probabilities in discrete distributions, and other large-scale combinatorial quantities. Large Number Theory, on the other hand, is to ensure the stability of convergence and probability occurring estimation. It claims that as the trails increase, the final probability will converge to the $E(X_n) = \mu$ also known as the mean of a data set in probability or almost surely by the Weak Law of Large Numbers and the Strong Law of Large Numbers accordingly. In conclusion, Stirling Formula deals with potential challenges brought by the enormous numbers of calculations and estimations, while the Law of Large Numbers gives a better understanding toward convergence and the probability of data with randomness. A combination of them can deal with many complex data sets and analysis.

References

[1] Conrad, Keith. Stirling's Formula. University of Connecticut, Department of Mathematics. Accessed Sept 15, 2025.
[2] Robert M. De Jong, Weak Laws of Large Numbers for Dependent Random Variables, Jul-Sept 1998.
[3] Scott Sheffield, 18.600: Lecture 30 .1in Weak law of large numbers, MIT, Accessed Sept 16, 2025.
[4] Stout W F. Almost Sure Convergence [M], New York: Academic Press, 1974.
[5] Zvi Artstein and Richard A. Vitale, A Strong Law of Large Numbers for Random Compact Sets, Oct 1975.