Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/040104

Simple Statistical Derivations of the Stirling's Formula Using Normal Approximations and Laplace's Method

Jose Guardiola[1,*] *and Hassan Elsalloukh*[2]

[1] Department of Mathematics and Statistics, Texas A&M University- Corpus Christi, 6300 Ocean Drive, CI-109, Corpus Christi, TX 78412, USA

[2] Department of Mathematics and Statistics, University of Arkansas at Little Rock, 2801 South University Avenue, Little Rock, Arkansas, 72204, USA

Received: 16 Sep. 2016, Revised: 15 Dec. 2016, Accepted: 21 Dec. 2016
Published online: 1 Jan. 2017

Abstract: In this article we show several methods that are useful to introduce students to the derivation of Stirling's formula. The methods shown here use the normal density as the limiting distribution to other distributions. The last procedure uses Laplace's method by completing a definite integral to the kernel of a normal distribution in order to solve it. The advantages of these procedures over many others found in the literature are their simplicity and their relationship with a normal limiting distribution. Under this approach, Laplace's method can also be considered as an approximation to a normal distribution to obtain the derivation of Stirling's formula. The derivations shown here are simple enough and sufficiently short such that they can be used for educational purposes on some undergraduate courses in statistics.

Keywords: Stirling's formula, convergence in distribution, Asymptotic theory, Convergence to a normal distribution, Taylor series

1 Introduction

Stirling's approximation or Stirling's formula is an approximation for large factorials. We can trace the practical origins of Stirling's formula back to De Moivre [1] , whose method was later modified to a more elegant solution by James Stirling as it is mentioned by Stigler [10]. In this article we discuss some statistical derivations of Stirling's formula by using convergence in distributions that have a limiting normal distribution. Convergence in distribution is studied in mathematical statistics courses and is commonly referred to as asymptotic theory. We say that the distribution of X is the asymptotic distribution or the limiting distribution of the sequence $\{Xn\}$. We may write this convergence as $X_n \xrightarrow{D} X$, see Hogg and Craig [5]. In this paper, X will always be a normal distribution, and we will refer to this limiting distribution as a "normal approximation".

Most elementary statistics students are already familiar with this convergence in distribution to a normal density, even if they have not been formally introduced to this concept, through their exposure to the central limit theorem. Another concept that elementary statistics students commonly see is that the binomial distribution has a limiting normal distribution for a large number of trials, so that binomial probabilities can be approximated by a normal distribution when the number of trials is large. These approximations are often mentioned in elementary statistics courses without explicit reference to the formal asymptotic theory. In this article we use a similar approach and provide some examples of convergence to a normal distribution that lead to simple derivations of Stirling's formula using only the formulas of distributions and basic algebra or, in one case, calculus. The convergence to a normal approximation provides a useful framework in which students can apply previously learned statistics concepts that allow them to derive Stirling's formula. These examples provide the students with a clear understanding of the convergence to a limiting normal distribution, the Laplace's method, and Stirling's formula. These procedures will empower the students to reinforce the learned concepts and to link them together in a cohesive manner.

* Corresponding author e-mail: jose.guardiola@tamucc.edu

2 Deriving Stirling’s Formula

2.1 Deriving Stirling’s Formula Using Convergence to a Normal Approximation

The first derivation of Stirling’s formula shown here uses a probability distribution function that approaches a normal distribution for large values of one of the parameters. It is a heuristic proof that is adequate for the purposes in this article. Consider for example a gamma probability distribution with shape parameter α and scale parameter β written as $\Gamma(\alpha,\beta)$. This distribution approaches the shape of a normal distribution for large values of the shape parameter α, see for example Grice [4]. This gamma distribution is then compared with a similarly shaped normal distribution with mean μ and variance σ^2, written as $N(\mu,\sigma^2)$, that exhibits the same mean and variance as the gamma distribution. In order to do that we set up the parameters as follows; $\alpha\cdot\beta=\mu$, and $\alpha\cdot\beta^2=\sigma^2$. We substitute α by n and for the sake of simplicity we use the parameter $\beta=1$, then each of the two distributions has both its mean and variance equal to n. The modes of these distributions are evaluated and equated at the common mean at $x=n$. By equating a gamma distribution $\Gamma(n,1)$ to a Normal distribution $N(n,n)$ we get,

$$\frac{1}{\Gamma(n)}x^{n-1}e^{-x}\approx\frac{1}{\sqrt{2\pi n}}\exp\left[-\frac{(x-n)^2}{2n}\right]. \tag{1}$$

Evaluating (1) at $x=n$ yields

$$n\cdot\Gamma(n)=n!\approx n^n e^{-n}\sqrt{2\pi n},$$

which is the well known Stirling’s approximation.

2.2 Deriving Stirling’s Formula Using the Central Limit Theorem

Students that already took a course in elementary statistics know that the sum of a large number of random variables, under certain conditions, has a normal limiting distribution, this result is known as the central limit theorem (CLT), see Lehmann [8]. The CLT can also be used to provide a derivation of Stirling’s Formula. First, we define S as a sum of n gamma random variables $X_i\sim\Gamma(1,\beta)$, each with mean β and variance β^2, then

$$S=X_1+X_2+\cdots+X_n. \tag{2}$$

Methods from mathematical statistics show that $S\sim\Gamma(n,\beta)$, see Hogg and Craig [5]. On the other hand, applying the central limit theorem for a sum of large enough n independent, identically distributed variables shows that the sum approaches a normal distribution with a mean equal to n times the mean of each individual variable and a variance equal to n times the variance of each individual variable, see Lehmann [8], that is

$$S\xrightarrow{D}N(n\beta,n\beta^2). \tag{3}$$

Then, from (2) and (3) we have

$$\Gamma(n,\beta)\xrightarrow{D}N(n\beta,n\beta^2). \tag{4}$$

Thus, by equating these expressions for large n and applying the convergence in distribution shown in (3) and (4), we can write

$$\frac{1}{\Gamma(n)\beta^n}x^{n-1}e^{-x/\beta}\approx\frac{1}{\sqrt{2\pi}\sqrt{n\beta^2}}\exp\left[-\frac{(x-n\beta)^2}{2n\beta^2}\right].$$

Considering equal probability density function values of these distributions at the common mean, that is at $x=n\beta$, we get

$$\frac{1}{\Gamma(n)\beta^n}(n\beta)^{n-1}e^{-n}\approx\frac{1}{\sqrt{2\pi}\sqrt{n\beta^2}}, \tag{5}$$

evaluating (5) for $n!$ yields

$$n!\approx n^n e^{-n}\sqrt{2\pi n}.$$

3 Laplace's Method

3.1 Derivation of Laplace's Method Using the Kernel of a Normal Distribution

First, we introduce the idea that the Laplace's method can also be considered as an approximation to the definite integral of a normal distribution. We use a truncated Taylor series approximation and rearrange terms in order to solve a definite integral that can be expressed in terms of a normal distribution whose solution is known. The result from this integral, widely known as Laplace's method [9], can also be used to derive Stirling's formula. First consider a non-normalized probability density function $f(x)$ with continuous second derivatives and positive values of the function between the constants a and b. First consider rewriting the function $f(x)$ as

$$f(x) = \exp\left[Lnf(x)\right]. \tag{6}$$

Using Taylor's approximation we have

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 + \varepsilon, \tag{7}$$

Where ε includes all higher order terms, which in this approximation will be considered negligible. Then from (6) we have

$$\int_a^b f(x)dx = \int_a^b \exp\left[Lnf(x)\right]dx, \tag{8}$$

and applying Taylor series approximation to $Lnf(x)$ in (8) we have

$$\int_a^b f(x)dx = \int_a^b \exp\left\{[Lnf(x_0)] + \frac{d[Lnf(x)]_{x=x_0}}{dx}(x-x_0) + \frac{1}{2}\frac{d^2[Lnf(x)]_{x=x_0}}{dx^2}(x-x_0)^2 + \varepsilon\right\}dx, \tag{9}$$

After dropping the higher order terms ε, and computing the first derivative on (9), we have

$$\int_a^b f(x)dx \approx \int_a^b \exp\left\{[Lnf(x_0)] + \frac{f'(x_0)}{f(x_0)}(x-x_0) + \frac{1}{2}\frac{d^2[Lnf(x)]_{x=x_0}}{dx^2}(x-x_0)^2\right\}dx, \tag{10}$$

Assuming that $f(x)$ has a global maximum at x_0 and thus $f'(x_0) = 0$, then from (8) and (10) we have

$$\int_a^b f(x)dx \approx \int_a^b \exp\left\{[Lnf(x_0)] + \frac{1}{2}\frac{d^2[Lnf(x)]_{x=x_0}}{dx^2}(x-x_0)^2\right\}dx, \tag{11}$$

Rearranging terms in (11) to set up the kernel of a normal probability density function, whose integral is known, we have the following expression

$$\int_a^b f(x)dx \approx \exp[Lnf(x_0)]\int_a^b \exp\left\{-\frac{1}{2}\frac{(x-x_0)^2}{\dfrac{1}{-\dfrac{d^2[Lnf(x)]}{dx^2}|_{x=x_0}}}\right\}dx, \tag{12}$$

Expression (12) can be evaluated by recognizing the well known kernel of the normal distribution, then, the solution of this integral when $a \to -\infty$ and $b \to \infty$ is the normalizing constant of the Gaussian distribution, that is $\sqrt{2\pi\sigma^2}$. Heuristically it is not hard to imagine that as n gets large, and assuming that the variance term decreases, the most important contribution to the value of the integral will be between a fixed and b. We can evaluate this definite integral by identifying and matching the corresponding terms to these constants. This method of evaluating a definite integral is the well known Laplace's method. As can be seen here, this method is related to the previous methods by again approximating a normal distribution, this time with a Taylor polynomial. Then we have

$$\int_a^b f(x)dx \approx f(x_0)\cdot\sqrt{\frac{2\pi}{-\dfrac{d^2[Lnf(x)]_{x=x_0}}{dx^2}}}$$

which is the well known Laplace's method.

3.2 Using Laplace's Method to Derive Stirling's Formula

Now let's use Laplace's method to approximate the definite integral of the Chi-square distribution. Let the random variable X have a Chi-square distribution with p degrees of freedom, written as $X \sim \chi^2_p$. Considering the definite integral of the Chi-square distribution, we know that for any probability density function

$$1 = \int_0^\infty \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2} dx$$

or equivalently

$$\Gamma(p/2)2^{p/2} = \int_0^\infty x^{(p/2)-1} e^{-x/2} dx, \tag{13}$$

Then we define

$$f(x) = x^{(p/2)-1} e^{-x/2},$$

the mode for a Chi-square distribution [3] is at

$$x_0 = p - 2. \tag{14}$$

The second derivative of $Ln[f(x)]$ is

$$\frac{d^2 Ln(f(x))}{dx^2} = -\frac{(p-2)}{2x^2} \tag{15}$$

Applying Laplace's method to approximate the integral in (13) we have

$$\Gamma(p/2)2^{p/2} \approx f(x_0) \sqrt{\frac{2\pi}{\left(-\frac{d^2[Ln f(x)]_{x=x_0}}{dx^2}\right)}}. \tag{16}$$

Then substituting (14) and (15) into (16), we get

$$\Gamma(p/2)2^{p/2} \approx (p-2)^{p/2-1} e^{-(p-2)/2} \sqrt{4\pi(p-2)}.$$

Considering large degrees of freedom, in a way that the Chi-square distribution reduces skewness and looks relatively symmetric, and allowing $n = \frac{p-2}{2}$ yields

$$n! \approx n^n e^{-n} \sqrt{2\pi n},$$

obtaining Stirling's formula again.

4 Conclusion

The derivations shown above are simple enough and sufficiently short that they can be used for educational purposes in some undergraduate courses in statistics. These proofs are related to other proofs of a probabilistic nature shown in Hu [6], Diaconis and Freedman [2], and Kahn [7]. The relationship between Laplace's method and a normal approximation seems intuitive and easy to remember as applicable in the derivation of Stirling's formula. There is an abundance of methods leading to Stirling's formula, but many of them are not suitable for students with only an elementary statistics background. The approaches developed here are appropriate for students with a basic background in statistics and mathematics. We emphasize that these methods are simple enough to be used in a calculus based introduction to mathematical statistics, and the first two methods may be suitable for courses in statistics without any calculus prerequisites.

References

[1] De Moivre, A., Approximatio ad Summan Terminorum Binomili $\overline{a+b^n}$ in Seriem Expansi (1733).
[2] Diaconis, P. and Freedman D., "An Elementary Proof of Stirling's Formula". American Mathematical Monthly, **93**, 123-125 (1986).
[3] Dowdy, S., Wearden S., Statistics for Research, John Wiley & Sons, 2nd Ed (1991).
[4] Grice, J. and Bain L., "Inferences Concerning the Mean of the Gamma Distribution", Journal of the American Statistical Association, Vol.**75**, No. 372 , pp. 929-933 (1980)

[5] Hogg, R., Craig, A. and McKean, J., Introduction to Mathematical Statistics, Prentice Hall; 6th edition (2004).
[6] Hu T. C. , "A Statistical Method of Approach to Stirling's Formula". The American Statistician, **42**, 204-205 (1988).
[7] Kahn R., "A Probabilistic Proof of Stirling's Formula". American Mathematical Monthly, Volume **81**, Issue 4 (1974).
[8] Lehmann, E., Elements of Large-Sample Theory, Springer-Verlag, New York (1999).
[9] Laplace, P. S., "Memoire sur la Probabilite des Causes par les Evenements", Memoire de Mathematique et de Physique, **6**, 621-656 (1774).
[10] Stigler S., The History of Statistics, Belknap Press, Cambridge Massachusetts (1990).

Jose Guardiola is an Associate Professor, Department of Mathematics and Statistics, Texas A&M University- Corpus Christi, 6300 Ocean Drive, CI-109, Corpus Christi, TX 78412 (email: jose.guardiola@tamucc.edu).

Hassan Elsalloukh is an Associate Professor, Department of Mathematics and Statistics, University of Arkansas at Little Rock, 2801 South University Avenue, 616 DKSN, Little Rock, Arkansas, 72204 (email: hxelsalloukh@ualr.edu). We are grateful to Dean Young for his encouragement and support, to Blair Sterba-Boatwright and Thomas McMillan for their help editing this article, and to Richard Arena for his help on searching related articles.