Southern Illinois University Carbondale

OpenSIUC

Publications

2017

Derivation of the Exponential Distribution through an Infinite Sine Series

Todd C. Headrick
Southern Illinois University Carbondale, headrick@siu.edu

Michael E. May
Southern Illinois University Carbondale

Follow this and additional works at: http://opensiuc.lib.siu.edu/cqmse_pubs

Recommended Citation

Derivation of the Exponential Distribution

through an Infinite Sine Series

Todd Christopher Headrick

Departments of CQMSE and Psychology (Quantitative Methods-Statistics)
Wham Bldg., Mail Code 4618
Southern Illinois University Carbondale
Carbondale, IL, USA 62901-4618

Michael E. May

Department of CQMSE and the Rehabilitation Institute (Behavior Analysis and Therapy)
Wham Bldg., Mail Code 4618
Southern Illinois University Carbondale
Carbondale, IL, USA 62901-4618

Abstract

This communication provides a derivation of the often-used one-parameter exponential family of distributions based on an infinite sine series. The main results associated with this derivation are the finite approximations of the probability density function (pdf) and the cumulative distribution function (cdf) of the exact one-parameter exponential family of distributions. The limit of these finite functions are the exact pdf and cdf. Numerical examples are provided to compare and contrast the finite approximations of distributions in terms of error using the finite-based cdf and the exact standard exponential cdf. We would also note that the finite approximations of the pdf and cdf vary in terms of both shape and percentage points. In view of this, the finite pdf and cdf offer a user the flexibility to potentially provide more accurate approximations in the context of fitting distributions to data rather than approximations that are based solely on the exact exponential pdf or cdf – most notably when distributions are heavy-(right)tailed.

1 Introduction

The continuous one-parameter exponential distribution is one of the most often used distributions to make inferential or probability statements with respect to events that occur – and reoccur – at random over time e.g. life time events and survival analysis. Specifically, Epstein and Sobel [5] discussed the use of the one-parameter exponential distribution in terms of life testing, fatigue testing, and in other contexts of destructive test situations such as the amount of (i) current needed to blow out a fuse, (ii) voltage needed to break down a condenser, or (iii) force to rupture some physical material.

Further, the exponential distribution is closely related to both the Poisson and geometric distributions. For example, the exponential distribution has been demonstrated to be related to the Poisson distribution in terms of the number of successive changes of discrete observations that occur during a given continuous time interval e.g. the waiting times between successive changes; which are also random variables. Furthermore, the exponential distribution can also be derived based on the Poisson or geometric distributions see e.g. Johnson, Kotz, and Kemp [13]; Johnson, Kotz, and Balakrishnan [12]; Marsaglia [15]; Leemis and McQuestion [14].

The exponential or approximate exponential distributions have also been demonstrated to be useful in many Monte Carlo studies in terms of comparing and contrasting the Type I error rates and (or) power properties associated with parametric and (or) nonparametric statistics. See, for example, Algina, Oshima and Line [1]; Boneau [2]; Conover and Iman [3]; Donaldson [4]; Harwell and Serlin [6]; Headrick and Sawilowsky [9,10]; Headrick [7]; Headrick ([8], p. 106).

In view of the importance of the one-parameter exponential distribution, the purpose of this communication is to derive this statistical distribution through an infinite sine series; which is, as far as we are aware, wholly new. It is demonstrated that the finite derivations of the pdf and cdf provided below can yield good approximations to the exact pdf and cdf. Further, the derivation provided herein of the (smaller values of the natural number index set) finite approximations to the one-parameter exponential distribution may vary considerably in terms of shape and percentage points. As such, this may provide a user with more accurate approximations in the context of fitting distributions to real-world data.

2   Mathematical Development

The exact forms of the one-parameter exponential probability density function (pdf) and cumulative distribution function (cdf) considered herein are respectively as follows (e.g. Hogg and Tanis, 2001, pp. 178-179):

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \tag{1}$$

$$F(x) = 1 - e^{-x/\theta} \tag{2}$$

where $0 \le x < \infty$ and the parameter $\theta > 0$.

The derivation of equations (1) and (2) begins by making use of the following lemma. **LEMMA**: If n is the index set, such that $n \in \mathbb{N}$, then the following infinite sine series

$$\sum_{n=1}^{\infty} \frac{\sin nx}{n} = \frac{1}{2}(\pi - x), \quad 0 < x < 2\pi. \tag{3}$$

PROOF: The proof begins by first setting

$$z = \cos x + i \sin x. \tag{4}$$

As such, we will let

$$\sum_{n=1}^{m} y^{n-1} z^n = \frac{z\{1 - (yz)^m\}}{1 - yz}. \tag{5}$$

It follows that for $|y| < 1$ yields

$$\sum_{n=1}^{\infty} y^{n-1}(\cos nx + i \sin nx) = \frac{\cos x + i \sin x}{1 - y \cos x - yi \sin x} = \frac{(\cos x - y) + i \sin x}{1 - 2y \cos x + y^2}. \tag{6}$$

Consequently, we have

$$\sum_{n=1}^{\infty} y^{n-1}(\sin nx) = \frac{\sin x}{1 - 2y \cos x + y^2}. \tag{7}$$

The infinite series in equation (7) is uniformly convergent for all values of y and for $|y| \le p < 1$. It follows that integrating with respect to y, where $0 < y < 1$, yields

$$\sum_{n=1}^{\infty} y^n \left(\frac{\sin nx}{n} \right) = \sin x \int_0^y \frac{du}{1 - 2u \cos x + u^2}$$

$$= \sin x \int_0^y \frac{du}{(u - \cos x)^2 + (\sin x)^2}$$

$$= \left[\tan^{-1} \left(\frac{u - \cos x}{\sin x} \right) \right]_{u=0}^{u=y}$$

$$= \tan^{-1} \left(\frac{y - \cos x}{\sin x} \right) - \tan^{-1}(- \cot x)$$

and thus we have

$$\sum_{n=1}^{\infty} y^n \left(\frac{\sin nx}{n} \right) = \begin{cases} \tan^{-1} \left(\dfrac{y - \cos x}{\sin x} \right) + \left(\dfrac{\pi}{2} - x \right), & 0 < x < \pi \\[3mm] \tan^{-1} \left(\dfrac{y - \cos x}{\sin x} \right) + \left(\dfrac{3\pi}{2} - x \right), & \pi < x < 2\pi \end{cases} . \quad (8)$$

Let us suppose that x is neither zero nor a multiple of 2π. As such, it necessarily follows that the infinite sine series is convergent – for $0 \le y \le 1$ and y^n – and is positive, monotonic, decreasing, and bounded.

Hence, if we let $y \to 1$ in equation (8), and (again) if x is neither zero nor a multiple of 2π, yields

$$\sum_{n=1}^{\infty} \frac{\sin nx}{n} = \begin{cases} \tan^{-1} \left(\tan \dfrac{1}{2} x \right) + \left(\dfrac{\pi}{2} - x \right), & 0 < x < \pi \\[3mm] \tan^{-1} \left(\tan \dfrac{1}{2} x \right) + \left(\dfrac{3\pi}{2} - x \right), & \pi < x < 2\pi \end{cases} . \quad (9)$$

Whence, based on equation (9), we then have equation (3) as given above i.e.

$$\sum_{n=1}^{\infty} \frac{\sin nx}{n} = \frac{1}{2} (\pi - x), \quad 0 < x < 2\pi. \blacksquare$$

REMARK: We would note that the final step, to obtain equation (3), is justified by the fact that if the series $\sum a_n(x)$ is uniformly convergent for $a \le x \le b$ to the sum $a(x)$, and if, for each value of n, $a_n(x)$ tends to the limit (s_n) as $x \to x_0$, where x_0 is some point in the range of (a, b), then, as $x \to x_0$, $a(x)$ tends to the limit δ, where δ is the sum of the infinite series $\sum_{n=1}^{\infty} s_n$.

3   Main Result

Exponentiation of equation (3) and using a general scaler of $1/\theta$ (in lieu of $1/2$ in equation 3) for values of $\theta > 0$ and subsequently integrating with respect to x yields:

$$\int_0^\infty e^{(1/\theta)(\pi - x)} dx = \theta e^{\pi/\theta}. \tag{10}$$

Solving equation (10) for a constant to ensure that this equation integrates to one yields a constant of $(e^{-\pi/\theta})/\theta$ for $\theta > 0$. As such, it necessarily follows that the functions provided below in equations (11) and (12) are the finite approximations of the exponential pdf and cdf in equations (1) and (2). Specifically, for finite values of k we have

$$f(x) \cong \frac{1}{\theta} \left\{ (e^{-\pi/\theta}) \left(1 + \frac{\pi - x}{k\theta} \right)^k \right\}, \tag{11}$$

and

$$F(x) \cong \frac{1}{\theta(1 + k)} \left\{ (e^{-\pi/\theta}) \left(\frac{\pi + k\theta - x}{k\theta} \right)^k (-\pi - k\theta + (\pi + k\theta)^{1+k}(\pi + k\theta + x)^{-k}) + x \right\}, \tag{12}$$

where $x > 0$, $0 < k < \infty$, $\theta > 0$, and $\pi + k\theta > x$. Taking the limit as $k \to \infty$ in equations (11) and (12) will result in these approximating functions to converge to the exact one-parameter exponential pdf and cdf given in equations (1) and (2), respectively.

4   Numerical Examples and Conclusion

Figure 1 below gives the graphs of the approximate pdfs associated with equation (11) for an exponential distribution with $\theta = 1$, and $k = 10, 10^2, 10^3, 10^5$. Further, presented below in Table 1 are approximate cumulative proportions, based on equation (12), and are associated with the exact 90-th percentile value for the standard exponential pdf ($x = 2.302585\ldots$), as well as the error in these proportions. Inspection of Table 1 indicates that as the value of k increases then the approximation becomes closer to the exact proportion (0.90) of the standard exponential distribution. If we were to increase the value of k to $k = 10^6$ (or $k = 10^7$) in Table 1, then the error would be less than 2.7×10^{-6} (or 2.3×10^{-7}).

In conclusion, it is worthy to point out that there is no real-world set of data that will *exactly* follow a theoretical statistical distribution. In view of this, the derivation presented in this paper offers a variety of approximate exponential

distributions with varying shapes and percentage points. Thus, this allows a user to alter the shape of an approximate exponential distribution in such a manner that may create a more suitable approximation to a set of data in terms of distribution fitting. Specifically, when empirical distributions are contaminated with outliers in the right tail of the distribution.

Conflict of Interests

The authors declare that there are no conflict of interests regarding the publication of this paper.

References

[1] J. Algina, T.C. Oshima and W. Lin, Type I error for Welch's test and James's second-order test under non-normality and inequality of variance when there are two groups, *Journal of Educational and Behavioral Statistics*, **19** (1994), 275-291. https://doi.org/10.3102/10769986019003275

[2] C.A. Boneau, The effects of violations of assumptions underlying the t-test, *Psychological Bulletin*, **57** (1960), 49-64. https://doi.org/10.1037/h0041412

[3] W.J. Conover and R.L. Iman, Analysis of covariance using the rank transformation, *Biometrics*, **38** (1982), 715-724.
https://doi.org/10.2307/2530051

[4] T.S. Donaldson, Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio, *Journal of the American Statistical Association*, **63** (1968), 660-676.
https://doi.org/10.2307/2284037

[5] B. Epstein and M. Sobel, Life testing, *Journal of the American Statistical Association*, **48** (1953), 486-502.
https://doi.org/10.1080/01621459.1953.10483488

[6] M.R. Harwell and R.C. Serlin, An empirical study of a proposed test of nonparmetric analysis of covariance, *Psychological Bulletin*, **104** (1988), 268-281. https://doi.org/10.1037/0033-2909.104.2.268

[7] T.C. Headrick, Fast fifth-order polynomial transforms for generating univariate and multivariate non-normal distributions, *Computational Statistics & Data Analysis*, **40** (2002), 685-711. https://doi.org/10.1016/s0167-9473(02)00072-5

[8] T.C. Headrick, *Statistical Simulation: Power Method Polynomials and other Transformations*, Chapman & Hall/CRC, 2010.

[9] T.C. Headrick and S.S. Sawilowsky, Properties of the rank transformation in factorial analysis of covariance, *Communications in Statistics: Simulation and Computation*, **29** (2000), 1059-1087.
https://doi.org/10.1080/03610910008813654

[10] T.C. Headrick and S.S. Sawilowsky, Weighted simplex procedures for determining boundary points and constants for the univariate and multivariate power methods, *Journal of Educational and Behavioral Statistics*, **25** (2000), 417-436. https://doi.org/10.3102/10769986025004417

[11] R.V. Hogg and E.A. Tanis, *Probability and Statistical Inference*, Prentice Hall, 2001.

[12] N.L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 1, John Wiley and Sons, 2001.

[13] N.L. Johnson, S. Kotz and A.W. Kemp, *Univariate Discrete Distributions*, Vol. 1, John Wiley and Sons, 1993.

[14] L.M. Leemis and J.T. McQuestion, Univariate distribution relationships, *American Statistician*, **62** (2008), 45-53.
https://doi.org/10.1198/000313008x270448

[15] G. Marsaglia, Generating exponential random variables, *Annals of Mathematical Statistics*, **32** (1961), 899-900.
https://doi.org/10.1214/aoms/1177704984

Value of k	Cumulative Proportion	Error
$k = 10$	$0.697852\ldots$	$0.202148\ldots$
$k = 10^2$	$0.873523\ldots$	$0.026476\ldots$
$k = 10^3$	$0.897268\ldots$	$0.002732\ldots$
$k = 10^5$	$0.899973\ldots$	$0.000027\ldots$

Table 1: Cumulative proportions using the exact value ($x = 2.302585\ldots$) associated with the standard exponential distribution for the 90-th percentile. The proportions are based on equation (12) and the values of $k = 10, 10^2, 10^3, 10^5$ that are provided in Figure 1.

(a) $k = 10$
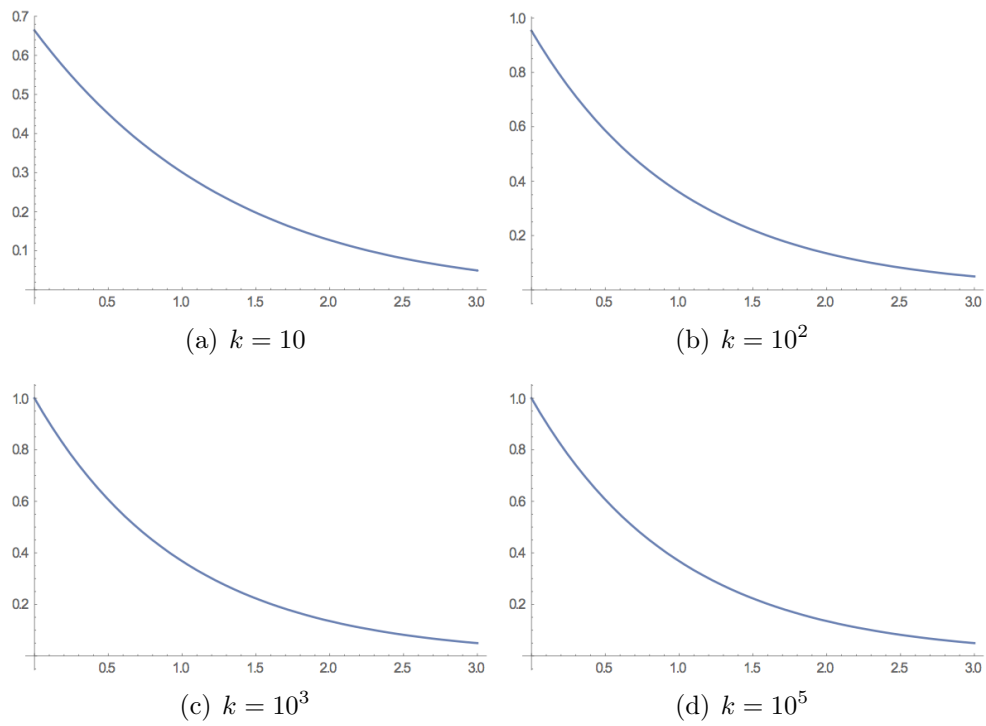
(b) $k = 10^2$

(c) $k = 10^3$

(d) $k = 10^5$

Figure 1: Approximations of the standard exponential distribution based on equation (11)