The Log-Gamma Distribution and Non-Normal Error

by Leigh J. Halliwell

ABSTRACT

Because insured losses are positive, loss distributions start from zero and are right-tailed. However, residuals, or errors, are centered about a mean of zero and have both right and left tails. Seldom do error terms from models of insured losses seem normal. Usually they are positively skewed, rather than symmetric. And their right tails, as measured by their asymptotic failure rates, are heavier than that of the normal. As an error distribution suited to actuarial modeling this paper presents and recommends the log-gamma distribution and its linear combinations, especially the combination known as the generalized logistic distribution. To serve as an example, a generalized logistic distribution is fitted by maximum likelihood to the standardized residuals of a loss-triangle model. Much theory is required for, and occasioned by, this presentation, most of which appears in three appendices along with some related mathematical history.

KEYWORDS

Log-gamma, digamma, logistic, Euler-Mascheroni, cumulant, maximum likelihood, robust, bootstrap

1. Introduction

The errors resultant in most actuarial models are not normally distributed.

Since the late 1980s actuarial science has been moving gradually from methods to models. The movement was made possible by personal computers; it was made necessary by insurance competition. Actuaries, who used to be known for fitting curves and extrapolating from data, are now likely to be fitting models and explaining data.

Statistical modeling seeks to explain a block of observed data as a function of known variables. The model makes it possible to predict what will be observed from new instances of those variables. However, rarely, if ever, does the function perfectly explain the data. A successful model, as well as seeming reasonable, should explain the observations tolerably well. So the deterministic model $\mathrm{y} = f(\mathrm{X})$ gives way to the approximation $\mathrm{y} \approx f(\mathrm{X})$, which is restored to equality with the addition of a random error term: $\mathbf{y} = f(\mathrm{X}) + \mathbf{e}$. The simplest model is the homoskedastic linear statistical model (LSM) with vector form $\mathbf{y} = \mathrm{X}\beta + \mathbf{e}$, in which $E[\mathbf{e}] = 0$ and $Var[\mathbf{e}] = \sigma^2\mathrm{I}$. According to LSM theory, $\hat{\boldsymbol{\beta}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathbf{y}$ is the best linear unbiased estimator (BLUE) of β, even if the elements of error vector $\mathbf{e}$ are not normally distributed.[1]

The fact that the errors resultant in most actuarial models are not normally distributed raises two questions. First, does non-normality in the randomness affect the accuracy of a model's predictions? The answer is "Yes, sometimes seriously." Second, can models be made "robust,"[2] i.e., able to deal properly with non-normal error? Again, "Yes." Three attempts to do so are 1} to relax parts of BLUE, especially linearity and unbiasedness (robust estimation), 2} to incorporate explicit distributions into models (GLM), and 3} to bootstrap.[3] Bootstrapping a linear model begins with solving it conventionally. One obtains $\hat{\boldsymbol{\beta}}$ and the expected observation vector $\hat{\mathbf{y}} = \mathrm{X}\hat{\boldsymbol{\beta}}$. Gleaning information from the residual vector $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$, one can simulate proper, or more realistic, "pseudo-error" vectors $\mathbf{e}_i^*$ and pseudo-observations $\mathbf{y}_i = \hat{\mathbf{y}} + \mathbf{e}_i^*$. Iterating the model over the $\mathbf{y}_i$ will produce pseudo-estimates $\hat{\boldsymbol{\beta}}_i$ and pseudo-predictions in keeping with the apparent distribution of error. Of the three attempts, bootstrapping is the most commonsensical.

Our purpose herein is to introduce a distribution for non-normal error that is suited to bootstrapping in general, but especially as regards the asymmetric and skewed data that actuaries regularly need to model.

[1]For explanations of the linear statistical model and BLUE see Judge [1988, §§5.5 and 5.7] and Halliwell [2015]. Although normality is not assumed in the model, it is required for the usual tests of significance, the t-test and the f-test. It would divert us here to argue whether randomness arises from reality or from our ignorance thereof. Between the world wars physicists Niels Bohr and Albert Einstein argued this at length–technically to a stalemate, although most physicists give the decision to Bohr. Either way, actuaries earn their keep by dealing with what the insurance industry legitimately perceives as randomness ("risk"). One reviewer commented, "From a Bayesian perspective, there is no real concept of randomness in the sense of an outcome that is not the result of a cause-effect relationship." This pushes conditional probability too far, to the tautology that $\mathrm{Prob}[X = a|X = b] = \delta_{ab}$. Information about some variables from a group of jointly distributed random variables can affect the probability distribution of the others, as in the classic case of the positive-definite multivariate normal distribution. Such information tightens or narrows the "wave function," without altogether collapsing it. The reviewer's comment presumes the notion of Einstein's hidden variables. Bayes is neither here nor there, for a variable that remains hidden forever, or ontologically, really is random. The truth intended by the comment is that modelers should not be lazy, that subject to practical constraints they should incorporate all relevant information into their models, even the so-called "collateral" information. Judge [1988, Chapter 7] devotes fifty pages to the Bayesian version of the LSM. Of course, whether randomness be ontological or epistemological, the goal of science is to mitigate it, if not to dispel it. This is especially the goal of actuarial science with regard to risk.

[2]"Models that perform well even when the population does not conform precisely to the parametric family are said to be robust" (Klugman [1998, §2.4.3]). "A robust estimator . . . produces estimates that are 'good' (in some sense) under a wide variety of possible data-generating processes" (Judge [1988, Introduction to Chapter 22]). Chapter 22 of Judge contains thirty pages on robust estimation.

[3]For descriptions of bootstrapping see Klugman [1998, Example 2.19 in §2.4.3] and Judge [1988, Chapter 9, §9.A.1]. However, they both describe bootstrapping, or resampling, from the empirical distribution of residuals. Especially when observations are few (as in the case of our example in Section 6 with eighty-five observations) might the modeler want to bootstrap/resample from a parametric error distribution. Not to give away "free cover" would require bootstrapping from a parametric distribution, unless predictive errors were felt to be well represented by the residuals. A recent CAS monograph that combines GLM and bootstrapping is Shapland [2016].

As useful candidates for non-normal error, Sections 2 and 3 will introduce the log-gamma random variable and its linear combinations. Section 4 will settle on a linear combination that arguably maximizes the ratio of versatility to complexity, the generalized logistic random variable. Section 5 will examine its special cases. Finally, Section 6 will estimate by maximum likelihood the parameters of one such distribution from actual data. The most mathematical and theoretical subjects are relegated to Appendices A-C.

2. The log-gamma random variable

If $X \sim Gamma(\alpha, \theta)$, then $Y = \ln X$ is a random variable whose support is the entire real line.[4] Hence, the logarithm converts a one-tailed distribution into a two-tailed. Although a leftward shift of X would move probability onto the negative real line, such a left tail would be finite. The logarithm is a natural way, even *the* natural way, to transform one infinite tail into two infinite tails.[5] Because the logarithm function strictly increases, the probability density function of $Y \sim Log\text{-}Gamma(\alpha, \theta)$ is:[6]

$$f_Y(u) = f_{X=e^Y}\left(e^u\right)\frac{de^u}{du} = \frac{1}{\Gamma(\alpha)}e^{-\frac{e^u}{\theta}}\left(\frac{e^u}{\theta}\right)^{\alpha-1}\frac{1}{\theta}e^u$$

$$= \frac{1}{\Gamma(\alpha)}e^{-\frac{e^u}{\theta}}\left(\frac{e^u}{\theta}\right)^{\alpha}$$

Figure 1 contains a graph of the probability density functions of both X and $Y = \ln X$ for $X \sim Gamma(1, 1) \sim Exponential(1)$. The log-gamma tails are obviously infinite, and the curve itself is skewed to the left (negative skewness).

The log-gamma moments can be derived from its moment generating function:

$$M_Y(t) = E\left[e^{tY}\right] = E\left[e^{t\ln X}\right] = E\left[X^t\right] = \frac{\Gamma(\alpha+t)}{\Gamma(\alpha)}\theta^t$$

Even better is to switch from moments to cumulants by way of the cumulant generating function:[7]

$$\psi_Y(t) = \ln M_Y(t) = \ln\Gamma(\alpha+t) - \ln\Gamma(\alpha) + t\ln\theta$$

$$\psi_Y(0) = \ln M_Y(0) = 0$$

The cumulants (κ_n) are the derivatives of this function evaluated at zero, the first four of which are:

$$\psi_Y'(t) = \psi(\alpha+t) + \ln\theta$$

$$\kappa_1 = E[Y] = \psi_Y'(0) = \psi(\alpha) + \ln\theta$$

$$\psi_Y''(t) = \psi'(\alpha+t)$$

$$\kappa_2 = Var[Y] = \psi_Y''(0) = \psi'(\alpha) > 0$$

$$\psi_Y'''(t) = \psi''(\alpha+t)$$

$$\kappa_3 = Skew[Y] = \psi_Y'''(0) = \psi''(\alpha) < 0$$

$$\psi_Y^{iv}(t) = \psi'''(\alpha+t)$$

$$\kappa_4 = XsKurt[Y] = \psi_Y^{iv}(0) = \psi'''(\alpha) > 0$$

The scale factor θ affects only the mean.[8] The alternating inequalities of κ_2, κ_3, and κ_4 derive from the polygamma formulas of Appendix A.3. The variance, of course, must be positive; the negative skewness confirms the appearance of the log-gamma density in Figure 1. The positive excess kurtosis means that the log-gamma distribution is "platykurtic;" its kurtosis is more positive than that of the normal distribution.

[4]Appendix A treats the gamma function and its family of related functions.

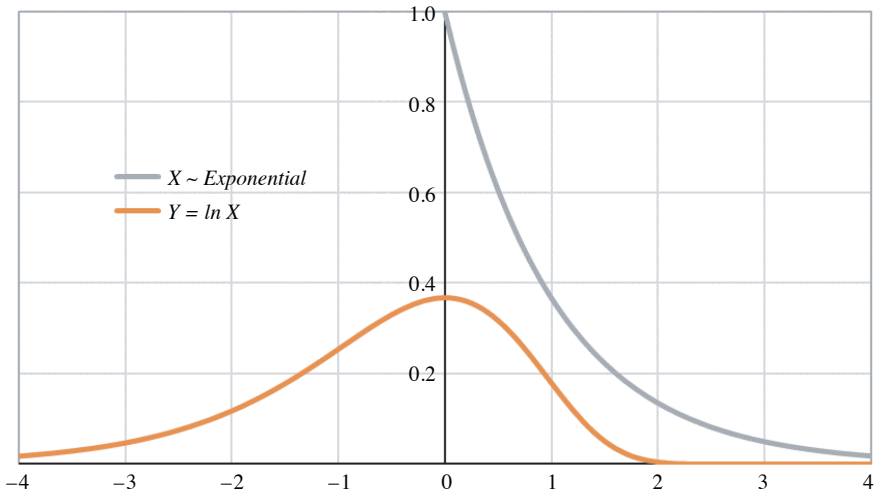
[5]Meyers [2013], also having seen the need for two-tailed loss distributions, is intrigued with skew normal and "mixed lognormal-normal" distributions, which in our opinion are not as intelligible and versatile as log-gamma distributions.

[6]With Leemis ["Log-gamma distribution"], we use the 'log' prefix for $Y = \ln(Gamma)$, rather than for $\ln(Y) = Gamma$. Hogg [1984, j2.6] defines his "loggamma" as $Y = e^{Gamma}$, after analogy with the lognormal.

[7]Moments, cumulants, and their generating functions are reviewed in Appendix B. The customary use of 'ψ' for both the cumulant generating function and the digamma function is unfortunate; however, the presence or absence of a subscripted random variable resolves the ambiguity.

[8]See footnote 16 in Appendix A.1.

Figure 1. Probability densities



Since the logarithm is a concave downward function, it follows from Jensen's inequality:

$$E[Y = \ln X] \le \ln E[X]$$

$$\psi(\alpha) + \ln\theta \le \ln(\alpha\theta) = \ln(\alpha) + \ln(\theta)$$

$$\psi(\alpha) \le \ln(\alpha)$$

Because the probability is not amassed, the inequality is strict: $\psi(\alpha) < \ln(\alpha)$ for $\alpha > 0$. However, when $E[X] = \alpha\theta$ is fixed at unity and as $\alpha \to \infty$, the variance of X approaches zero. Hence, $\lim_{\alpha\to\infty} (\ln(\alpha) - \psi(\alpha)) = \ln 1 = 0$. It is not difficult to prove that $\lim_{\alpha\to 0^+} (\ln(\alpha) - \psi(\alpha)) = \infty$, as well as that $\ln(\alpha) - \psi(\alpha)$ strictly decreases. Therefore, for every $y > 0$ there exists exactly one $\alpha > 0$ for which $y = \ln(\alpha) - \psi(\alpha)$.

The log-gamma random variable becomes an error term when its expectation equals zero. This requires the parameters to satisfy the equation $E[Y] = \psi(\alpha) + \ln\theta = 0$, or $\theta = e^{-\psi(\alpha)}$. Hence, the simplest of all log-gamma error distributions is $Y \sim Log\text{-}Gamma(\alpha, e^{-\psi(\alpha)}) = \ln(X \sim Gamma(\alpha, e^{-\psi(\alpha)}))$.

3. Weighted sums of log-gamma random variables

Multiplying the log-gamma random variable by negative one reflects its distribution about the y-axis. This does not affect the even moments or cumulants; but it reverses the signs of the odd ones. For example, the skewness of $-Y = -\ln X = \ln X^{-1}$ is positive.

Now let Y be a γ-weighted sum of independent log-gamma random variables Y_k, which resolves into a product of powers of independent gamma random variables $X_k \sim Gamma(\alpha_k, \theta_k)$:

$$Y = \sum_k \gamma_k Y_k = \sum_k \gamma_k \ln X_k = \sum_k \ln X_k^{\gamma_k} = \ln \prod_k X_k^{\gamma_k}$$

Although the X_k must be independent of one another, their parameters need not be identical. Because of the independence:

$$\psi_Y(t) = \ln E\left[e^{tY}\right] = \ln E\left[e^{t\sum_k \gamma_k Y}\right] = \ln E\left[\prod_k e^{t\gamma_k Y_k}\right]$$

$$= \ln \prod_k E\left[e^{t\gamma_k Y_k}\right] = \sum_k \ln E\left[e^{t\gamma_k Y_k}\right] = \sum_k \psi_{Y_k}(t\gamma_k)$$

The nth cumulant of the weighted sum is:

$$\kappa_n(Y) = \frac{d^n \psi_Y(t)}{dt^n}\bigg|_{t=0} = \frac{d^n}{dt^n} \sum_k \psi_{Y_k}(t\gamma_k)\bigg|_{t=0}$$

$$= \sum_k \frac{d^n \psi_{Y_k}(t\gamma_k)}{dt^n}\bigg|_{t=0} = \sum_k \gamma_k^n \psi_{Y_k}^{[n]}(0)$$

$$= \sum_k \gamma_k^n \kappa_n(Y_k)$$

So the nth cumulant of a weighted sum of independent random variables is the weighted sum of the

cumulants of the random variables, the weights being raised to the nth power.[9]

Using the cumulant formulas from the previous section, we have:

The not overly complicated "generalized logistic" distribution is versatile enough for modeling non-normal error.

$$\kappa_1(Y) = \sum_k \gamma_k \kappa_1(Y_k) = \sum_k \gamma_k \psi(\alpha_k) + \sum_k \gamma_k \ln(\theta_k)$$
$$= \sum_k \gamma_k \psi(\alpha_k) + C$$

$$\kappa_2(Y) = \sum_k \gamma_k^2 \kappa_2(Y_k) = \sum_k \gamma_k^2 \psi'(\alpha_k)$$

$$\kappa_3(Y) = \sum_k \gamma_k^3 \kappa_3(Y_k) = \sum_k \gamma_k^3 \psi''(\alpha_k)$$

$$\kappa_4(Y) = \sum_k \gamma_k^4 \kappa_4(Y_k) = \sum_k \gamma_k^4 \psi'''(\alpha_k)$$

In general, $\kappa_{m+1}(Y) = \sum_k \gamma_k^{m+1} \psi^{[m]}(\alpha_k)$ + IF($m = 0$, C, 0). A weighted sum of n independent log-gamma random variables would provide $2n + 1$ degrees of freedom for a method-of-cumulants fitting. All the scale parameters θ_k would be unity. As the parameter of an error distribution, C would lose its freedom, since the mean must then equal zero. Therefore, with no loss of generality, we may write $Y = \ln \prod_k X_k^{\gamma_k} + C$ for $X_k \sim Gamma(\alpha_k, 1)$.

4. The generalized logistic random variable

Although any finite weighted sum is tractable, four cumulants should suffice in most practice. So let $Y = \gamma_1 \ln X_1 + \gamma_2 \ln X_2 + C = \ln(X_1^{\gamma_1} X_2^{\gamma_2}) + C$. Even then, one gamma should be positive and the other negative; in fact, letting one be the opposite of the other will allow Y to be symmetric in special cases. Therefore, $Y = \gamma \ln X_1 - \gamma \ln X_2 + C = \gamma \ln(X_1/X_2) + C$ for $\gamma > 0$ should be a useful form of intermediate complexity. Let the parameterization for this purpose be $X_1 \sim Gamma(\alpha, 1)$ and $X_2 \sim Gamma(\beta, 1)$. Contributing to the usefulness of this form is the fact that X_1/X_2 is a generalized Pareto random variable, whose probability density function is:[10]

$$f_{X_1/X_2}(u) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{u}{u+1}\right)^{\alpha-1} \left(\frac{1}{u+1}\right)^{\beta-1} \frac{1}{(u+1)^2}$$

Since $e^Y = e^C (X_1/X_2)^\gamma$, for $-\alpha < \gamma t < \beta$:

$$M_Y(t) = E\left[e^{tY}\right] = e^{Ct} E\left[(X_1/X_2)^{\gamma t}\right]$$

$$= e^{Ct} \int_{u=0}^{\infty} u^{\gamma t} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{u}{u+1}\right)^{\alpha-1} \left(\frac{1}{u+1}\right)^{\beta-1} \frac{du}{(u+1)^2}$$

$$= e^{Ct} \frac{\Gamma(\alpha+\gamma t)\Gamma(\beta-\gamma t)}{\Gamma(\alpha)\Gamma(\beta)} \int_{u=0}^{\infty} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\gamma t)\Gamma(\beta-\gamma t)} \left(\frac{u}{u+1}\right)^{\alpha+\gamma t-1} \left(\frac{1}{u+1}\right)^{\beta-\gamma t-1} \frac{du}{(u+1)^2}$$

$$= e^{Ct} \frac{\Gamma(\alpha+\gamma t)\Gamma(\beta-\gamma t)}{\Gamma(\alpha)\Gamma(\beta)}$$

Hence, the cumulant generating function and its derivatives are:

$$\psi_Y(t) = \ln M_Y(t) = Ct + \ln\Gamma(\alpha+\gamma t) - \ln\Gamma(\alpha) + \ln\Gamma(\beta-\gamma t) - \ln\Gamma(\beta)$$

$$\psi_Y'(t) = \gamma\left(\psi(\alpha+\gamma t) - \psi(\beta-\gamma t)\right) + C$$

$$\psi_Y''(t) = \gamma^2\left(\psi'(\alpha+\gamma t) + \psi'(\beta-\gamma t)\right)$$

[9] The same is true of all the ordinary, or non-central, moments. It is true also of the first three central moments, but only because they are identical to the first three cumulants. Halliwell [2011, §4] explains why the fourth and higher central moments of independent random variables are not additive.

[10] Hogg [1884, §2.6] derives this by the change-of-variable technique. Venter [1983, Appendix D] uses the mixing technique, since *Gamma* (α, 1)/*Gamma* (β, 1) ~ *Gamma* (α, θ = 1/*Gamma* (β, 1)). See also Klugman [1998, §2.7.3.4 and Appendix A.2.2.1]. Our formulation assumes a scale factor of unity in the generalized Pareto.

$$\psi_Y'''(t) = \gamma^3(\psi''(\alpha+\gamma t) - \psi''(\beta-\gamma t))$$

$$\psi_Y^{iv}(t) = \gamma^4(\psi'''(\alpha+\gamma t) + \psi'''(\beta-\gamma t))$$

And so, the cumulants are:

$$\kappa_1 = E[Y] = \psi_Y'(0) = C + \gamma(\psi(\alpha) - \gamma(\beta))$$

$$\kappa_2 = Var[Y] = \psi_Y''(0) = \gamma^2(\psi'(\alpha) + \psi'(\beta)) > 0$$

$$\kappa_3 = Skew[Y] = \psi_Y'''(0) = \gamma^3(\psi''(\alpha) - \psi''(\beta))$$

$$\kappa_4 = XsKurt[Y] = \psi_Y^{iv}(0) = \gamma^4(\psi'''(\alpha) + \psi'''(\beta)) > 0$$

The three parameters α, β, γ could be fitted to empirical cumulants κ_2, κ_3, and κ_4. For an error distribution C would equal $\gamma(\psi(\beta) - \psi(\alpha))$. Since $\kappa_4 > 0$, the random variable Y is platykurtic.

Since $Y = \gamma \ln(X_1/X_2) + C$, $Z = (Y - C)/\gamma = \ln(X_1/X_2)$ may be considered a reduced form. From the generalized Pareto density above, we can derive the density of Z:

$$f_Z(u) = f_{X_1/X_2=e^Z}(e^u)\frac{de^u}{du}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{e^u}{e^u+1}\right)^{\alpha-1}\left(\frac{1}{e^u+1}\right)^{\beta-1}\frac{1}{(e^u+1)^2}\cdot e^u$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{e^u}{e^u+1}\right)^{\alpha}\left(\frac{1}{e^u+1}\right)^{\beta}$$

$Z \sim \ln(Gamma(\alpha, 1)/Gamma(\beta, 1))$ is a "generalized logistic" random variable (Wikipedia [Generalized logistic distribution]).

The probability density functions of generalized logistic random variables are skew-symmetric:

$$f_{\frac{1}{Z}=\ln\frac{X_2}{X_1}}(-u) = \frac{\Gamma(\beta+\alpha)}{\Gamma(\beta)\Gamma(\alpha)}\left(\frac{e^{-u}}{e^{-u}+1}\right)^{\beta}\left(\frac{1}{e^{-u}+1}\right)^{\alpha}$$

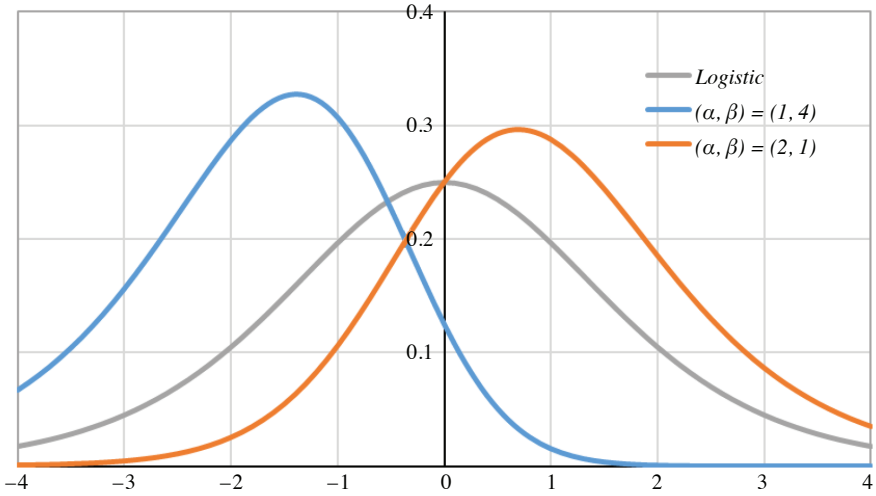
$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{e^u}{e^u+1}\right)^{\alpha}\left(\frac{1}{e^u+1}\right)^{\beta}$$

$$= f_{Z=\ln\frac{X_1}{X_2}}(u)$$

In Figure 2 are graphed three generalized logistic probability density functions.

The density is symmetric if and only if $\alpha = \beta$; the gray curve is that of the logistic density, for which $\alpha = \beta = 1$. The mode of the generalized-logistic (α, β) density is $u_{mode} = \ln \alpha/\beta$. Therefore, the mode is positive [or negative] if and only if $\alpha > \beta$ [or $\alpha < \beta$]. Since the digamma and tetragamma functions ψ, ψ'' strictly increase over the positive reals, the signs of $E[Z] = \psi(\alpha) - \psi(\beta)$ and $Skew[Z] = \psi''(\alpha) - \psi''(\beta)$ are the same as the sign of $\alpha - \beta$. The positive mode of the orange curve (2, 1) implies positive mean and skewness, whereas for the blue curve (1, 4) they are negative.

Figure 2. Generalized logistic densities



5. Special cases

Although the probability density function of the generalized logistic random variable is of closed form, the form of its cumulative distribution function is not closed, save for the special cases of $\alpha = 1$ and $\beta = 1$. The special case of $\alpha = 1$ reduces X_1/X_2 to an ordinary Pareto. In this case, the cumulative distribution is $F_Z(u) = 1 - \left(\frac{1}{e^u+1}\right)^{\beta}$. Likewise, the special case of $\beta = 1$ reduces X_1/X_2 to an inverse Pareto. In that case, $F_Z(u) = \left(\frac{e^u}{e^u+1}\right)^{\alpha}$. It would be easy to simulate values of Z in both cases by the inversion method (Klugman [1998, Appendix H.2]).[11]

Quite interesting is the special case $\alpha = \beta$. For Z is symmetric about its mean if and only if $\alpha = \beta$, in which case all the odd cumulants higher than the mean equal zero. Therefore, in this case:

$$f_Z(u) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}\frac{(e^u)^{\alpha}}{(e^u+1)^{2\alpha}}$$

Symmetry is confirmed in as much as $f_Z(-u) = f_Z(u)$. When $\alpha = \beta = 1$, $f_Z(u) = \frac{e^u}{(e^u+1)^2}$, whose cumulative distribution is $F_Z(u) = \frac{e^u}{e^u+1} = \left(\frac{e^u}{e^u+1}\right)^{\alpha=1} = 1 - \left(\frac{1}{e^u+1}\right)^{\beta=1}$. In this case, Z is a logistic distribution. Its mean and skewness are zero. As for its even cumulants:[12]

$$\kappa_2 = Var[Z] = \psi'(\alpha) + \psi'(\beta) = 2\psi'(1)$$

$$= 2\sum_{k=1}^{\infty}\frac{1}{k^2} = 2\cdot\zeta(2) = 2\frac{\pi^2}{6} = \frac{\pi^2}{3} \approx 3.290$$

$$\kappa_4 = XsKurt[Z] = \psi'''(\alpha) + \psi'''(\beta) = 2\psi'''(1)$$

$$= 2\cdot 6\sum_{k=1}^{\infty}\frac{1}{k^4} = 12\cdot\zeta(4) = 12\cdot\frac{\pi^4}{90} = \frac{2\pi^4}{15} \approx 12.988$$

Instructive also is the special case $\alpha = \beta = ½$. Since $\Gamma(½) = \sqrt{\pi}$, the probability density function in this case is $f_Z(x) = \frac{1}{\pi}\frac{e^{x/2}}{e^x+1}$. The constant $\frac{1}{\pi}$ suggests a connection with the Cauchy density $\frac{1}{\pi}\frac{1}{u^2+1}$ (Wikepedia [Cauchy distribution]). Indeed, the density function of the random variable $W = e^{Z/2}$ is:

$$f_W(u) = f_{Z=2\ln W}(2\ln u)\frac{d(2\ln u)}{du}$$

$$= \frac{1}{\pi}\frac{u}{u^2+1}\frac{2}{u} = \frac{2}{\pi}\frac{1}{u^2+1}$$

This is the density function of the absolute value of the standard Cauchy random variable.[13]

6. A maximum-likelihood example

Table 1 shows eighty-five standardized[14] error terms from an additive-incurred model. A triangle of incremental losses by accident-year row and

[11]Workable Excel formulas are $Z = LN(-1 + (1 - RAND())^{(-1/\beta)})$ and $Z = -LN(RAND())^{(-1/\alpha)} - 1)$. To simulate the generalized Pareto and logistic random variables requires gamma or beta inverse functions. Two equivalent forms are $Z1 = LN(GAMMA.INV(RAND(), \alpha, 1)) - LN(GAMMA.INV(RAND(), \beta, 1))$ and $Z2 = -LN(-1 + 1/BETA.INV(RAND(), \alpha, \beta))$. Mathematically, Z1 ~ Z2; but the Z2 formula overflows when $RAND() \approx 1$.

[12]See Wikipedia [Logistic distribution] and Wikipedia [Riemann zeta function]. Often cited is the coefficient of excess kurtosis: $\gamma_2 = \kappa_4/\kappa_2^2 = (2\pi^4/15)/(\pi^2/3)^2 = 1.2$. Havil [2003, Chapter 4] shows how Euler in 1735 proved that $\sum_{k=1}^{\infty} 1/k^2$ converges to the value $\pi^2/6$. Since the mid-1600s, determining the value of $\sum_{k=1}^{\infty} 1/k^2$ had eluded Wallis, Leibniz, and the Bernoullis. Because the latter lived in Basel, it became known as the Basel problem.

[13]Another derivation is:

$$W = e^{Z/2} = \sqrt{e^Z} = \sqrt{e^{\ln\left(\frac{Gamma(½,1)}{Gamma(½,1)}\right)}} = \sqrt{\frac{Gamma(½,1)}{Gamma(½,1)}}$$

$$= \sqrt{\frac{Gamma(½,2)}{Gamma(½,2)}} = \sqrt{\frac{N(0,1)^2}{N(0,1)^2}} = \left|\frac{N(0,1)}{N(0,1)}\right|$$

The *Gamma* (½, 2) ~ χ_1^2 random variable is the square of the standard normal. Wikipedia [Cauchy distribution] and Hogg [1984, pp. 47, 49] show that the standard Cauchy random variable is the quotient of two independent standard normals.

[14]The generic model of the observations is $\mathbf{y} = X\beta + \mathbf{e}$, where $Var[\mathbf{e}] = \sigma^2\Phi$. The variance of the residuals is $Var[\hat{\mathbf{e}}] = Var[\mathbf{y} - X\hat{\beta}] = \sigma^2(\Phi - X(X'\Phi^{-1}X)^{-1}X')$, as derived by Halliwell [1996, Appendix D]. The standardized residual of the ith element is then $\hat{\mathbf{e}}_i/\sqrt{Var[\hat{\mathbf{e}}]_{ii}}$. Usually the estimate $\hat{\sigma}^2$ must be used for σ^2.

Table 1. Ordered error sample (85 = 17 × 5)

–2.287	–0.585	–0.326	–0.009	0.658
–1.575	–0.582	–0.317	0.009	0.663
–1.428	–0.581	–0.311	0.010	0.698
–1.034	–0.545	–0.298	0.012	0.707
–1.022	–0.544	–0.267	0.017	0.821
–1.011	–0.514	–0.260	0.019	0.998
–0.913	–0.503	–0.214	0.038	1.009
–0.879	–0.501	–0.202	0.090	1.061
–0.875	–0.487	–0.172	0.095	1.227
–0.856	–0.486	–0.167	0.123	1.559
–0.794	–0.453	–0.165	0.222	1.966
–0.771	–0.435	–0.162	0.255	1.973
–0.726	–0.429	–0.115	0.362	2.119
–0.708	–0.417	–0.053	0.367	2.390
–0.670	–0.410	–0.050	0.417	2.414
–0.622	–0.384	–0.042	0.417	2.422
–0.612	–0.381	–0.038	0.488	2.618

evaluation-year column was modeled from the exposures of its thirteen accident years. Variance by column was assumed to be proportional to exposure. A complete triangle would have $\sum_{k=1}^{13} k = 91$ observations; but we happened to exclude six.

The sample mean is nearly zero at 0.001. Three of the five columns are negative, and the positive errors are more dispersed than the negative. Therefore, this sample is positively skewed, or skewed to the right. Other sample cumulants are 0.850 (variance), 1.029 (skewness), and 1.444 (excess kurtosis). The coefficients of skewness and of kurtosis are 1.314 and 1.999.

By maximum likelihood we wished to explain the sample as coming from $Y = \gamma Z + C$, where $Z = \ln(X_1/X_2) \sim \ln(Gamma(\alpha, 1)/Gamma(\beta, 1))$, the generalized logistic variable of Section 4 with distribution:

$$f_Z(u) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{e^u}{e^u+1}\right)^{\alpha}\left(\frac{1}{e^u+1}\right)^{\beta}$$

So, defining $z(u;\ C,\ \gamma) = (u - C)/\gamma$, whereby $z(Y) = Z$, we have the distribution of Y:

$$f_Y(u) = f_{Z=z(Y)}(z(u))z'(u)$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{e^{z(u)}}{e^{z(u)}+1}\right)^{\alpha}\left(\frac{1}{e^{z(u)}+1}\right)^{\beta}\frac{1}{\gamma}$$

The logarithm, or the log-likelihood, is:

$$\ln f_Y(u) = \ln\Gamma(\alpha+\beta) - \ln\Gamma(\alpha) - \ln\Gamma(\beta)$$

$$+ \alpha z(u) - (\alpha+\beta)\ln\left(e^{z(u)}+1\right) - \ln\gamma$$

This is a function in four parameters; C and γ are implicit in $z(u)$. With all four parameters free, the likelihood of the sample could be maximized. Yet it is both reasonable and economical to estimate Y as a "standard-error" distribution, i.e., as having zero mean and unit variance. In Excel it sufficed us to let the Solver add-in maximize the log-likelihood with respect to α and β, giving due consideration that C and γ, constrained by zero mean and unit variance, are themselves functions of α and β. As derived in Section 4, $E[Y] = C + \gamma(\psi(\alpha) - \psi(\beta))$ and $Var[Y] = \gamma^2(\psi'(\alpha) + \psi'(\beta))$. Hence, standardization requires that $\gamma(\alpha, \beta) = 1/\sqrt{\psi'(\alpha)+\psi'(\beta)}$ and that $C(\alpha, \beta) = \gamma(\alpha, \beta) \cdot (\psi(\beta) - \psi(\alpha))$. The log-likelihood maximized at $\hat{\alpha} = 0.326$ and $\hat{\beta} = 0.135$. Numerical derivatives of the GAMMALN function approximated the digamma and trigamma functions at these values. The remaining parameters for a standardized distribution must be $\hat{C} = -0.561$ and $\hat{\gamma} = 0.122$. Figure 3 graphs the maximum-likelihood result.

The maximum absolute difference between the empirical and the fitted, $|\max| = 0.066$, occurs at $x = 0.1$. We tested its significance with a "KS-like" (Kolmogorov-Smirnov, cf. Hogg [1984, p. 104]) statistic. We simulated one thousand samples of eighty-five instances from the fitted distribution, i.e., from the Y distribution with its four estimated parameters. Each simulation provided an empirical cumulative distribution, whose maximum absolute deviation we derived from the fitted distribution over the interval [–4, 4] in steps of 0.1. Figure 4 contains the graph of the cumulative density function of $|\max|$.

Figure 3. Empirical and fitted CDFs

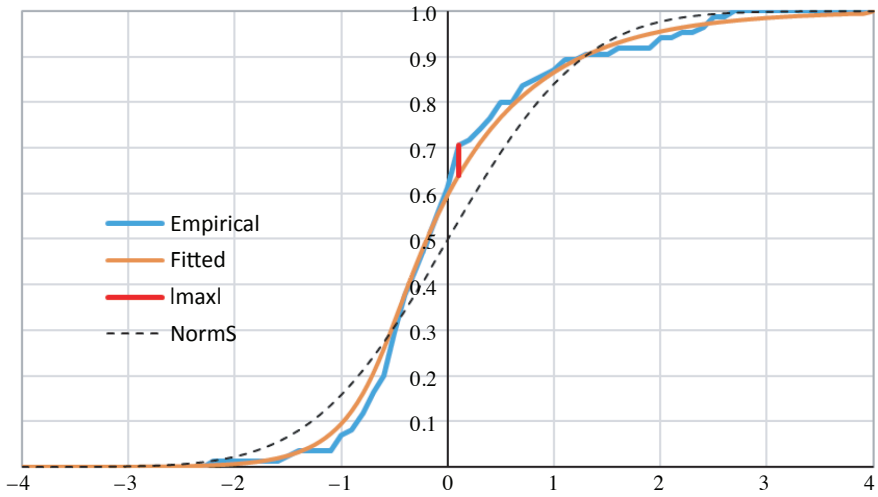
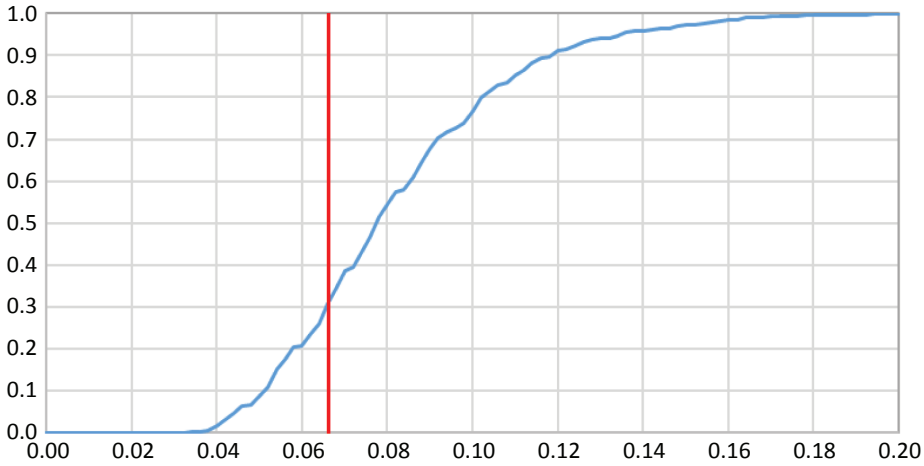


Figure 4. Distributon of |maxI|



The actual deviation of 0.066 coincides with the thirty-first percentile, or about with the lower tercile. The dashed line in Figure 3 (legend "NormS") represents the cumulative standard-normal distribution. The empirical distribution has more probability below zero and a heavier right tail. Simulations with these logistic error terms are bound to be more accurate than simulations defaulted to normal errors.[15]

[15]Taking trigamma and tetragamma values from www.easycalculation.com/statistics/polygamma-function.php, we calculated the higher cumulants: $\kappa_3 = Skew[Y] = 1.496$ and $\kappa_4 = XsKurt[Y] = 4.180$. Since Y is standardized, these values double as coefficients of skewness and kurtosis. That they, especially the kurtosis, are greater than the empirical coefficients, 1.314 and 1.999, is evidence for the inferiority of the method of moments.

7. Conclusion

Actuaries are well schooled in loss distributions, which are non-negative, positively skewed, and right-tailed. The key to a versatile distribution of error is to combine logarithms of loss distributions. Because most loss distributions are transformations of the gamma distribution, the log-gamma distribution covers most of the possible combinations. The generalized logistic distribution strikes a balance between versatility and complexity. It should be a recourse to the actuary seeking to bootstrap a model whose residuals are not normally distributed.

References

Artin, E., *The Gamma Function*, New York: Dover, 2015 (ET "Einführung in die Theorie der Gammafunktion," *Hamburger Mathematische Einzelshriften*, 1931).

Daykin, C. D., T. Pentikäinen, and M. Pesonen, *Practical Risk Theory for Actuaries*, London: Chapman & Hall, 1994.

Halliwell, L. J., "Loss Prediction by Generalized Least Squares," *Proceedings of the Casualty Actuarial Society*, 83 (1996), pp. 156–193, www.casact.org/pubs/proceed/proceed96/96436.pdf.

Halliwell, L. J., "Conditional Probability and the Collective Risk Model," Casualty Actuarial Society *E-Forum*, Spring 2011, www.casact.org/pubs/forum/11spforum/Halliwell.pdf.

Halliwell, L. J., "The Gauss-Markov Theorem: Beyond the BLUE," Casualty Actuarial Society *E-Forum*, Fall 2015, www.casact.org/pubs/forum/15fforum/Halliwell_GM.pdf.

Havil, J., *Gamma: Exploring Euler's Constant*, Princeton University Press, 2003.

Hogg, R. V., and S. A. Klugman, *Loss Distributions*, New York, Wiley, 1984.

Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lütkepohl, and T. Lee, *Introduction to the Theory and Practice of Econometrics*, 2nd ed., New York: Wiley, 1988.

Klugman, S. A., H. H. Panjer, and G. E. Wilmot, *Loss Models: From Data to Decisions*, New York: Wiley, 1998.

Leemis, L., "Log-Gamma Distribution," www.math.wm.edu/~leemis/chart/UDR/PDFs/Loggamma.pdf.

Meyers, G., "The Skew Normal Distribution and Beyond," *Actuarial Review*, May 2013, p. 15, www.casact.org/newsletter/pdfUpload/ar/AR_May2013_1.pdf.

Shapland, M. R., *Using the ODP Bootstrap Model: A Practitioner's Guide*, CAS Monograph Series Number 4, Arlington, VA: Casualty Actuarial Society, 2016, www.casact.org/pubs/monographs/papers/04-shapland.pdf.

Venter, G. G., "Transformed Beta and Gamma Distributions and Aggregate Losses," *Proceedings of the Casualty Actuarial Society*, 70 (1983), pp. 156–193, www.casact.org/pubs/proceed/proceed83/83156.pdf.

Wikipedia contributors, "Bohr–Mollerup Theorem," https://en.wikipedia.org/wiki/Bohr–Mollerup_theorem (accessed March 2017).

Wikipedia contributors, "Cauchy Distribution," https://en.wikipedia.org/wiki/Cauchy_distribution (accessed March 2017).

Wikipedia contributors, "Gamma Function," https://en.wikipedia.org/wiki/Gamma_function (accessed March 2017).

Wikipedia contributors, "Generalized Logistic Distribution," https://en.wikipedia.org/wiki/Generalized_logistic_distribution (accessed March 2017).

Wikipedia contributors, "Logistic Distribution," https://en.wikipedia.org/wiki/Logistic_distribution (accessed March 2017).

Wikipedia contributors, "Polygamma Function," https://en.wikipedia.org/wiki/Polygamma_function (accessed March 2017).

Wikipedia contributors, "Riemann Zeta Function," https://en.wikipedia.org/wiki/Riemann_zeta_function (accessed March 2017).

Appendix A
The gamma, log-gamma, and polygamma functions

A.1. The gamma function as an integral

The modern form of the gamma function is $\Gamma(\alpha) = \int_{x=0}^{\infty} e^{-x}x^{\alpha-1}dx$. The change of variable $t = e^{-x}$ (or $x = -\ln t = \ln 1/t$) transforms it into:

$$\Gamma(\alpha) = \int_{t=1}^{0} t\left(\ln\frac{1}{t}\right)^{\alpha-1}\left(-\frac{dt}{t}\right) = \int_{t=0}^{1}\left(\ln\frac{1}{t}\right)^{\alpha-1}dt$$

According to Havil [2003, p. 53], Leonhard Euler used the latter form in his pioneering work during 1729–1730. It was not until 1809 that Legendre named it the gamma function and denoted it with the letter 'Γ'. The function records the struggle between the exponential function e^{-x} and the power function $x^{\alpha-1}$, in which the former ultimately prevails and forces the convergence of the integral for $\alpha > 0$.

Of course, $\Gamma(1) = \int_{x=0}^{\infty} e^{-x}\,dx = 1$. The well-known recurrence formula $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ follows from integration by parts:

$$\alpha\Gamma(\alpha) = \alpha\int_{x=0}^{\infty} e^{-x}x^{\alpha-1}\,dx = \int_{x=0}^{\infty} e^{-x}d\left(x^{\alpha}\right) = e^{-x}x^{\alpha}\Big|_0^{\infty}$$

$$-\int_{x=0}^{\infty} x^{\alpha}d\left(e^{-x}\right) = 0 + \int_{x=0}^{\infty} e^{-x}x^{\alpha+1-1}\,dx = \Gamma(\alpha+1)$$

It is well to note that in order for $e^{-x}x^{\alpha}\big|_0^{\infty}$ to equal zero, α must be positive. For positive-integral values of α, $\Gamma(\alpha) = (\alpha - 1)!$; so the gamma function extends the factorial to the positive real numbers. In this domain the function is continuous, even differentiable, as well as positive. Although $\lim_{\alpha\to 0^+}\Gamma(\alpha) = \infty$,

$\lim_{\alpha\to 0^+}\alpha\Gamma(\alpha)=\lim_{\alpha\to 0}\Gamma(\alpha+1)=\Gamma(1)=1$. So $\Gamma(\alpha)\sim\frac{1}{\alpha}$ as $\alpha\to 0^+$. As α increases away from zero, the function decreases to a minimum of approximately $\Gamma(1.462)=0.886$, beyond which it increases. So, over the positive real numbers the gamma function is 'U' shaped, or concave upward.

The simple algebra $1=\frac{\Gamma(\alpha)}{\Gamma(\alpha)}=\frac{\int_{x=0}^{\infty}e^{-x}x^{\alpha-1}\,dx}{\Gamma(\alpha)}=\int_{x=0}^{\infty}\frac{1}{\Gamma(\alpha)}e^{-x}x^{\alpha-1}dx$ indicates that $f(x)=\frac{1}{\Gamma(\alpha)}e^{-x}x^{\alpha-1}$ is the probability density function of what we will call a *Gamma*(α, 1)-distributed random variable, or simply *Gamma*(α)-distributed with $\theta=1$ understood.[16] For $k>-\alpha$, the kth moment of $X\sim Gamma(\alpha)$ is:

$$E[X^k]=\int_{x=0}^{\infty}x^k\frac{1}{\Gamma(\alpha)}e^{-x}x^{\alpha-1}dx$$

$$=\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\int_{x=0}^{\infty}\frac{1}{\Gamma(\alpha+k)}e^{-x}x^{\alpha+k-1}dx$$

$$=\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}$$

Therefore, for $k=1$ and 2, $E[X]=\alpha$ and $E[X^2]=\alpha(\alpha+1)$. Hence, $Var[X]=\alpha$.

A.2. The gamma function as an infinite product

Euler found that the gamma function could be expressed as the infinite product:

$$\Gamma(x)=\lim_{n\to\infty}\frac{n!n^x}{x(1+x)\dots(n+x)}$$

$$=\lim_{n\to\infty}\frac{n^x}{x\left(1+\frac{x}{1}\right)\dots\left(1+\frac{x}{n}\right)}$$

As a check, $\Gamma(1)=\lim_{n\to\infty}\frac{n!n}{(n+1)!}=\lim_{n\to\infty}\frac{n}{n+1}=1$. Moreover, the recurrence formula is satisfied:

$$\Gamma(x+1)=\lim_{n\to\infty}\frac{n!n^{x+1}}{(1+x)\dots(n+x)(n+1+x)}$$

$$=x\lim_{n\to\infty}\frac{n!n^x}{x(1+x)\dots(n+x)}$$

$$\lim_{n\to\infty}\frac{n}{(n+1+x)}=x\Gamma(x)$$

Hence, by induction, the integral and infinite-product definitions are equivalent for positive-integral values of x. It is the purpose of this section to extend the equivalence to all positive real numbers.

The proof is easier in logarithms. The log-gamma function is $\ln\Gamma(\alpha)=\ln\int_{x=0}^{\infty}e^{-x}x^{\alpha-1}\,dx$. The form of its recursive formula is $\ln\Gamma(\alpha+1)=\ln(\alpha\Gamma(\alpha))=\ln\Gamma(\alpha)+\ln\alpha$. Its first derivative is:

$$(\ln\circ\Gamma)'(\alpha)=\frac{d}{d\alpha}\ln\Gamma(\alpha)\equiv\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$=\frac{\int_{x=0}^{\infty}e^{-x}x^{\alpha-1}\ln x\,dx}{\Gamma(\alpha)}$$

$$=\int_{x=0}^{\infty}\frac{1}{\Gamma(\alpha)}e^{-x}x^{\alpha-1}\ln x\,dx=E[Y]$$

So the first derivative of $\ln\Gamma(\alpha)$ is the expectation of $Y\sim Log\text{-}Gamma(\alpha)$, as defined in Section 2. The second derivative is:

$$(\ln\circ\Gamma)''(\alpha)=\frac{d}{d\alpha}\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$=\frac{\Gamma(\alpha)\Gamma''(\alpha)-\Gamma'(\alpha)\Gamma'(\alpha)}{\Gamma(\alpha)\Gamma(\alpha)}$$

[16] Similarly, Venter [1983, p. 157]: "The percentage of this integral reached by integrating up to some point x defines a probability distribution, i.e., the probability of being less than or equal to x." The multiplication of this random variable by a scale factor $\theta>0$ is *Gamma*(α, θ)-distributed with density $f_{Gamma(\alpha,\theta)}(x)=\frac{1}{\Gamma(\alpha)}e^{-\frac{x}{\theta}}\left(\frac{x}{\theta}\right)^{\alpha-1}\frac{1}{\theta}$, as found under equivalent forms in Hogg [1984, p. 226] and Klugman [1998, Appendix A.3.2.1]. Because this paper deals with logarithms of gamma random variables, in which products turn into sums, we will often ignore scale factors.

$$= \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \left(\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right)^2$$

$$= \int_{x=0}^{\infty} \frac{1}{\Gamma(\alpha)} e^{-x} x^{\alpha-1} (\ln x)^2 \, dx - E[Y]^2$$

$$= E\left[Y^2\right] - E[Y]^2 = Var[Y] > 0$$

Because the second derivative is positive, the log-gamma function is concave upward over the positive real numbers.

Now let $n \in \{2, 3, \ldots\}$ and $x \in (0, 1]$. So $0 < n-1 < n < n+x \le n+1$. We consider the slope of the log-gamma secants over the intervals $[n-1, n]$, $[n, n+x]$, and $[n, n+1]$. Due to the upward concavity, the order of these slopes must be:

$$\frac{\ln\Gamma(n) - \ln\Gamma(n-1)}{n-(n-1)} < \frac{\ln\Gamma(n+x) - \ln\Gamma(n)}{(n+x)-n} \le \frac{\ln\Gamma(n+1) - \ln\Gamma(n)}{(n+1)-n}$$

Using the recurrence formula and multiplying by positive x, we continue:

$$x\ln(n-1) < \ln\Gamma(n+x) - \ln\Gamma(n) \le x\ln(n)$$

$$x\ln(n-1) + \ln\Gamma(n) < \ln\Gamma(n+x) \le x\ln(n) + \ln\Gamma(n)$$

But $\ln\Gamma(n) = \sum_{k=1}^{n-1} \ln k$ and $\ln\Gamma(n+x) = \ln\Gamma(x) + \ln x + \sum_{k=1}^{n-1} \ln(k+x)$. Hence:

$$x\ln(n-1) + \sum_{k=1}^{n-1} \ln k < \ln\Gamma(x) + \ln x + \sum_{k=1}^{n-1} \ln(k+x) \le x\ln(n) + \sum_{k=1}^{n-1} \ln k$$

$$x\ln(n-1) + \sum_{k=1}^{n-1} \ln k - \ln x - \sum_{k=1}^{n-1} \ln(k+x) < \ln\Gamma(x) \le x\ln(n) + \sum_{k=1}^{n-1} \ln k - \ln x - \sum_{k=1}^{n-1} \ln(k+x)$$

$$x\ln(n-1) - \ln x - \sum_{k=1}^{n-1} \ln\left(1+\frac{x}{k}\right) < \ln\Gamma(x) \le x\ln(n) - \ln x - \sum_{k=1}^{n-1} \ln\left(1+\frac{x}{k}\right)$$

This string of inequalities, in the middle of which is $\ln\Gamma(x)$, is true for $n \in \{2, 3, \ldots\}$. As n approaches infinity, $\lim_{n\to\infty}[x\ln(n) - x\ln(n-1)] = x\lim_{n\to\infty} \ln\left(\frac{n}{n-1}\right) = x\ln\left(\lim_{n\to\infty} \frac{n}{n-1}\right) = x\ln 1 = 0$. So $\ln\Gamma(x)$ is sandwiched between two series that have the same limit. Hence:

$$\ln\Gamma(x) = -\ln x + \lim_{n\to\infty}\left\{x\ln(n-1) - \sum_{k=1}^{n-1} \ln\left(1+\frac{x}{k}\right)\right\}$$

$$= -\ln x + \lim_{n\to\infty}\left\{x\ln(n) - \sum_{k=1}^{n} \ln\left(1+\frac{x}{k}\right)\right\}$$

If the limit did not converge, then $\ln\Gamma(x)$ would not converge. Nevertheless, Appendix C constructs a proof of the limit's convergence. But for now we can finish this section by exponentiation:

$$\Gamma(x) = e^{\ln\Gamma(x)}$$

$$= \frac{1}{x} e^{\lim_{n\to\infty}\left\{x\ln(n) - \sum_{k=1}^{n} \ln\left(1+\frac{x}{k}\right)\right\}}$$

$$= \frac{1}{x} \lim_{n\to\infty}\left\{e^{x\ln(n) - \sum_{k=1}^{n} \ln\left(1+\frac{x}{k}\right)}\right\}$$

$$= \frac{1}{x} \lim_{n\to\infty} \frac{n^x}{\prod_{k=1}^{n}\left(1+\frac{x}{k}\right)}$$

$$= \lim_{n\to\infty} \frac{n^x}{x\left(1+\frac{x}{1}\right)\ldots\left(1+\frac{x}{n}\right)}$$

The infinite-product form converges for all complex numbers $z \notin \{0, -1, -2, \ldots\}$; it provides the

analytic continuation to the integral form. [17] With the infinite-product form, Havil [2003, pp. 58f] easily derives the complement formula $\Gamma(x)\Gamma(1-x) = \pi/\sin(\pi x)$. Consequently, $\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}) = \pi/\sin(\pi/2) = \pi$.

A.3. Derivatives of the log-gamma function

The derivative of the gamma function is $\Gamma'(\alpha) = \frac{d}{d\alpha}\int_{x=0}^{\infty} e^{-x}x^{\alpha-1}\,dx = \int_{x=0}^{\infty} e^{-x}\frac{d\left(x^{\alpha-1}\right)}{d\alpha}dx = \int_{x=0}^{\infty} e^{-x}x^{\alpha-1}\ln x\,dx.$

So, for $X \sim Gamma(\alpha)$:

$$E[\ln X] = \int_{x=0}^{\infty} \ln x \frac{1}{\Gamma(\alpha)}e^{-x}x^{\alpha-1}dx$$

$$= \frac{1}{\Gamma(\alpha)}\int_{x=0}^{\infty} e^{-x}x^{\alpha-1}\ln x dx$$

$$= \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \frac{d}{d\alpha}\ln\Gamma(\alpha)$$

Because the derivative of $\ln \Gamma(\alpha)$ has proven useful, mathematicians have named it the "digamma" function: $\psi(\alpha) = \frac{d}{d\alpha}\ln\Gamma(\alpha)$. Therefore, $E[\ln X] = \psi(\alpha)$. Its derivatives are the trigamma, tetragamma, and so on. Unfortunately, Excel does not provide these functions, although one can approximate them from GAMMALN by numerical differentiation. The R and SAS programming languages have the trigamma function, but stop at the second derivative of $\ln \Gamma(\alpha)$. The remainder of this appendix deals with the derivatives of the log-gamma function.

According to the previous section:

$$\ln\Gamma(x) = -\ln x + \lim_{n\to\infty}\left\{x\ln(n) - \sum_{k=1}^{n}\ln\left(1+\frac{x}{k}\right)\right\}$$

Hence the digamma function is:

$$\psi(x) = \frac{d}{dx}\ln\Gamma(x) = -\frac{1}{x} + \lim_{n\to\infty}\left\{\ln n - \sum_{k=1}^{n}\frac{1}{k+x}\right\}$$

$$= \lim_{n\to\infty}\left\{\ln n - \sum_{k=0}^{n}\frac{1}{k+x}\right\}$$

Its recurrence formula is:

$$\psi(x+1) = \lim_{n\to\infty}\left\{\ln n - \sum_{k=0}^{n}\frac{1}{k+1+x}\right\}$$

$$= \lim_{n\to\infty}\left\{\ln n - \sum_{k=0}^{n}\frac{1}{k+x}\right\} + \frac{1}{0+x} = \psi(x) + \frac{1}{x}$$

Also, $\psi(1) = \lim_{n\to\infty}\left\{\ln n - \sum_{k=0}^{n}\frac{1}{k+1}\right\} = \lim_{n\to\infty}\left\{\ln n - \sum_{k=1}^{n+1}\frac{1}{k}\right\} = -\lim_{n\to\infty}\left\{\sum_{k=1}^{n}\frac{1}{k} - \ln n\right\} = -\gamma$. Euler was the first to see that $\lim_{n\to\infty}\left\{\sum_{k=1}^{n}\frac{1}{k} - \ln n\right\}$ converged.[18] He named it 'C' and estimated its value as 0.577218. Eventually it was called the Euler-Mascheroni constant γ, whose value to six places is really 0.577216.

[17] Here is an example of the increasing standard of mathematical rigor. Euler assumed that the agreement of the integral and the infinite product over the natural numbers implied their agreement over the positive reals. After all, the gamma function does not "serpentine" between integers, for it is concave. Only later was proof felt necessary, especially after Cantor demonstrated that $\aleph_0 + \aleph_0 = \aleph_0$. This implies that the agreement of two analytic functions at $\aleph_0$ points does not require the $\aleph_0$ derivatives in their Taylor series to be identical. (It would if $\aleph_0$ of those points were contained within a finite interval. But here the points of agreement are at least one unit distant from one another.) According to Wikipedia [Bohr–Mollerup theorem], when Harald Bohr and Johannes Mollerup published a proof of the integral/infinite-product equivalence in their 1922 textbook on complex analysis, they did not realize that their proof was original. Harald Bohr was the younger brother of Niels Bohr. A decade later, in the preface to his classic monograph on the gamma function, Emil Artin [2015 (1931)] acknowledged the fundamental nature of this theorem.

[18] Reformulate it as: $\lim_{n\to\infty}\left\{\sum_{k=1}^{n}\frac{1}{k} - \ln n\right\} = \lim_{n\to\infty}\left\{\sum_{k=1}^{n}\frac{1}{k} - \ln(n+1)\right\} = \lim_{n\to\infty}\left\{\sum_{k=1}^{n}\frac{1}{k} - \int_{x=k}^{k+1}\frac{dx}{x}\right\} = \sum_{k=1}^{\infty}\int_{x=k}^{k+1}\left(\frac{1}{k} - \frac{1}{x}\right)dx$. The value of each unit-integral is positive, but less than $1/k - 1/(k+1) = 1/k(k+1) < 1/k^2$. So the series increases, but is bounded by $\sum_{k=1}^{\infty} 1/k^2 = \zeta(2) = \pi^2/6$, where ζ is the zeta function [Wikipedia, Riemann zeta function]. See also Appendix C.

The successive derivatives of the digamma function are:

$$\psi'(x) = \frac{d}{dx}\lim_{n\to\infty}\left\{\ln n - \sum_{k=0}^{n}\frac{1}{k+x}\right\} = \sum_{k=0}^{\infty}\frac{1}{(k+x)^2}$$

$$\psi''(x) = -2\sum_{k=0}^{\infty}\frac{1}{(k+x)^3}$$

$$\psi'''(x) = 6\sum_{k=0}^{\infty}\frac{1}{(k+x)^4}$$

The general formula for the nth derivative is $\psi^{[n]}(x) = (-1)^{n-1}n!\sum_{k=0}^{\infty}\frac{1}{(k+x)^{n+1}}$ with recurrence formula:

$$\psi^{[n]}(x+1) = (-1)^{n-1}\,n!\sum_{k=0}^{\infty}\frac{1}{(k+1+x)^{n+1}}$$

$$= (-1)^{n-1}\,n!\sum_{k=0}^{\infty}\frac{1}{(k+x)^{n+1}} - \frac{(-1)^{n-1}\,n!}{x^{n+1}}$$

$$= \psi^{[n]}(x) + \frac{(-1)^n\,n!}{x^{n+1}}$$

Finally, $\psi^{[n]}(1) = (-1)^{n-1}n!\sum_{k=0}^{\infty}\frac{1}{(k+1)^{n+1}} = (-1)^{n-1}\,n!\sum_{k=1}^{\infty}\frac{1}{k^{n+1}} = (-1)^{n-1}\,n!\,\zeta(n+1)$. A graph of the digamma through pentagamma functions appears in Wikipedia [Polygamma function].

Appendix B
Moments versus cumulants[19]

Actuaries and statisticians are well aware of the moment generating function (mgf) of random variable X:

$$M_X(t) = E\left[e^{tX}\right] = \int_X e^{tx} f(x)\,dx$$

Obviously, $M_X(0) = E[1] = 1$. It is called the moment generating function because its derivatives at zero, if convergent, equal the (non-central) moments of X, i.e.:

$$M_X^{[n]}(0) = \frac{d^n}{dt^n}E\left[e^{tX}\right]\Big|_{t=0} = E\left[\frac{d^n e^{tX}}{dt^n}\Big|_{t=0}\right]$$

$$= E\left[X^n e^{0\cdot X}\right] = E\left[X^n\right].$$

The moment generating function for the normal random variable is $M_{N(\mu,\sigma^2)}(t) = e^{\mu t+\sigma^2 t^2/2}$. The mgf of a sum of independent random variables equals the product of their mgfs:

$$M_{X+Y}(t) = E\left[e^{t(X+Y)}\right] = E\left[e^{tX}e^{tY}\right]$$

$$= E\left[e^{tX}\right]E\left[e^{tY}\right] = M_X(t)\,M_Y(t)$$

The cumulant generating function (cgf) of X, or $\psi_X(t)$ is the natural logarithm of the moment generating function. So $\psi_X(0) = 0$; and for independent summands:

$$\psi_{X+Y}(t) = \ln M_{X+Y}(t) = \ln M_X(t)\,M_Y(t)$$

$$= \ln M_X(t) + \ln M_Y(t) = \psi_X(t) + \psi_Y(t)$$

The nth cumulant of X, or $\kappa_n(X)$, is the nth derivative of the cgf evaluated at zero. The first two cumulants are identical to the mean and the variance:

$$\kappa_1(X) = \frac{d\ln M_X(t)}{dt}\Big|_{t=0} = \frac{M_X'(t)}{M_X(t)}\Big|_{t=0} = E[X]$$

$$\kappa_2(X) = \frac{d}{dt}\frac{M_X'(t)}{M_X(t)}\Big|_{t=0} = \frac{M_X(t)M_X''(t) - M_X'(t)M_X'(t)}{M_X^2(t)}\Big|_{t=0}$$

$$= E\left[X^2\right] - E[X]E[X] = Var[X]$$

One who performs the somewhat tedious third derivative will find that the third cumulant is identical to the skewness. So the first three cumulants are equal to the first three central moments. This amounts to a proof that the first three central moments are additive.

[19]For moment generating functions see Hogg [1984, pp. 39f] and Klugman [1998, §3.9.1]. For cumulants and their generation see Daykin [1994, p. 23] and Halliwell [2011, §4].

However, the fourth and higher cumulants do not equal their corresponding central moments. In fact, defining the central moments as $\mu_n = E[(X-\mu)^n]$, the next three cumulant relations are:

$$\kappa_4 = \mu_4 - 3\mu_2^2 = \mu_4 - 3\sigma^4$$

$$\kappa_5 = \mu_5 - 10\mu_2\mu_3$$

$$\kappa_6 = \mu_6 - 15\mu_2\mu_4 - 10\mu_3^2 + 30\mu_2^3$$

The cgf of a normal random variable is $\psi_{N(\mu,\sigma^2)}(t) = \ln e^{\mu t+\sigma^2 t^2/2} = \mu t + \sigma^2 t^2/2$. It first two cumulants equal μ and σ^2, and its higher cumulants are zero. Therefore, the third and higher cumulants are relative to the zero values of the corresponding normal cumulants. So third and higher cumulants could be called "excess of the normal," although this in practice is done only for the fourth cumulant. Because $\mu_4 = E[(X-\mu)^4]$ is often called the kurtosis, ambiguity is resolved by calling κ_4 the "excess kurtosis," or the kurtosis in excess of $\mu_4 = E[N(0,\sigma^2)^4] = E[N(0,1)^4]\cdot\sigma^4 = 3\sigma^4$.

Appendix C
Log-gamma convergence

In Appendix A.2 we proved that for real $\alpha > 0$:

$$\ln\left(\int_{x=0}^{\infty} e^{-x}x^{\alpha-1}\,dx\right) = \ln\Gamma(\alpha) = -\ln\alpha + \lim_{n\to\infty}\left\{\alpha\ln n - \sum_{k=1}^{n}\ln\left(1+\frac{\alpha}{k}\right)\right\}$$

There we argued that the "real-ness" of the left side for all $\alpha > 0$ implies the convergence of the limit on the right side; otherwise, the equality would be limited or qualified. Nevertheless, it is valuable to examine the convergence of the log-gamma limit, as we will now do.

Define $a_n(x) = \frac{x}{n} - \ln\left(1+\frac{x}{n}\right)$ for natural number n and real number x. To avoid logarithms of negative numbers, $x > -1$, or $x \in (-1, \infty)$. The root of this group of functions is $f(x) = x - \ln(1+x)$. Since the two functions $y = x$ and $y = \ln(1+x)$ are tangent to each other at $x = 0$, $f(0) = f'(0) = 0$. Otherwise, the line is above the logarithm and $f(x \neq 0) > 0$. Since $f''(x) = 1/(1+x)^2$ must be positive, $f(x)$ is concave upward with a global minimum over $(-1, \infty)$ at $x = 0$.

For now, let us investigate the convergence of $\varphi(x) = \sum_{n=1}^{\infty} a_n(x) = \sum_{n=1}^{\infty}\left\{\frac{x}{n} - \ln\left(1+\frac{x}{n}\right)\right\}$. Obviously, $\varphi(x=0) = 0$. But if $x \neq 0$, for some $\xi \in (\min(n, n+x), \max(n, n+x))$:

$$a_n(x) = \frac{x}{n} - \ln\left(1+\frac{x}{n}\right)$$

$$= \frac{x}{n} - \ln\left(\frac{n+x}{n}\right)$$

$$= \frac{x}{n} - \int_{u=n}^{n+x}\frac{du}{u}$$

$$= \frac{x}{n} - ((n+x)-n)\frac{1}{\xi}$$

$$= \frac{x}{n} - \frac{x}{\xi}$$

$$= x\left(\frac{1}{n} - \frac{1}{\xi}\right)$$

Regardless of the sign of x:

$$|x|\left|\left(\frac{1}{n}-\frac{1}{n}\right)\right| < |a_n(x)| < |x|\left|\left(\frac{1}{n}-\frac{1}{n+x}\right)\right|$$

The lower bound reduces to zero; the upper to $|x|\left|\frac{x}{n(n+x)}\right| = \frac{|x|^2}{|n||n+x|} = \frac{x^2}{n|n+x|}$. But $n+x$ must be positive, since $x > -1$. Therefore, for $x \neq 0$:

$$0 < |a_n(x)| < \frac{x^2}{n(n+x)}$$

For $x = 0$, equality prevails. And so for all $x > -1$:

$$0 \le |a_n(x)| \le \frac{x^2}{n(n+x)}$$

Now *absolute* convergence, or the convergence of $\sum_{n=1}^{\infty}|a_n(x)|$, is stricter than simple convergence. If $\sum_{n=1}^{\infty}|a_n(x)|$ converges, then so too does $\varphi(x) = \sum_{n=1}^{\infty}a_n(x)$. And $\sum_{n=1}^{\infty}|a_n(x)|$ converges, if it is bounded from above. The following string of inequalities establishes the upper bound:

$$\begin{aligned}\sum_{n=1}^{\infty}|a_n(x)| &\le \sum_{n=1}^{\infty}\frac{x^2}{n(n+x)}\\ &\le \frac{x^2}{1(1+x)}+\sum_{n=2}^{\infty}\frac{x^2}{n(n+x)}\\ &\le \frac{x^2}{1+x}+x^2\sum_{n=2}^{\infty}\frac{x^2}{n(n+x)}\\ &\le \frac{x^2}{1+x}+x^2\sum_{n=2}^{\infty}\frac{1}{(n-1)^2}\\ &\le \frac{x^2}{1+x}+x^2\left(\frac{\pi^2}{6}=\varsigma(2)=\sum_{n=1}^{\infty}\frac{1}{n^2}\right)\end{aligned}$$

Thus have we demonstrated the convergence of $\varphi(x)=\sum_{n=1}^{\infty}a_n(x)=\sum_{n=1}^{\infty}\left\{\frac{x}{n}-\ln\left(1+\frac{x}{n}\right)\right\}$.

As a special case:[20]

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=1}^{n}a_k(1) &= \lim_{n\to\infty}\sum_{k=1}^{n}\left\{\frac{1}{k}-\ln\left(1+\frac{1}{k}\right)\right\}\\ &= \lim_{n\to\infty}\sum_{k=1}^{n}\left\{\frac{1}{k}-\ln\left(\frac{k+1}{k}\right)\right\}\\ &= \lim_{n\to\infty}\left\{\sum_{k=1}^{n}\frac{1}{k}-\ln(n+1)\right\}\\ &= \lim_{n\to\infty}\left\{\sum_{k=1}^{n}\frac{1}{k}-\ln(n)\right\}=\gamma\approx 0.577\end{aligned}$$

Returning to the limit in the log-gamma formula and replacing 'α' with 'x', we have:

$$\begin{aligned}&\lim_{n\to\infty}\left\{x\ln n-\sum_{k=1}^{n}\ln\left(1+\frac{x}{k}\right)\right\}\\ &= \lim_{n\to\infty}\left\{x\ln n+\sum_{k=1}^{n}\left[-\frac{x}{k}+\frac{x}{k}-\ln\left(1+\frac{x}{k}\right)\right]\right\}\\ &= \lim_{n\to\infty}\left\{x\ln n-\sum_{k=1}^{n}\frac{x}{k}\right\}+\sum_{k=1}^{n}\left\{\frac{x}{k}-\ln\left(1+\frac{x}{k}\right)\right\}\\ &= -x\lim_{n\to\infty}\left\{\sum_{k=1}^{n}\frac{1}{k}-\ln(n)\right\}+\sum_{k=1}^{n}\left\{\frac{x}{k}-\ln\left(1+\frac{x}{k}\right)\right\}\\ &= -\gamma x+\sum_{k=1}^{n}a_k(x)\end{aligned}$$

Thus, independently of the integral definition, we have proven for $\alpha > 0$ the convergence of $\ln\Gamma(\alpha)=-\ln\alpha+\lim_{n\to\infty}\left\{\alpha\ln n-\sum_{k=1}^{n}\ln\left(1+\frac{\alpha}{k}\right)\right\}$.[21]

This is but one example of a second-order convergence. To generalize, consider the convergence of $\varphi(x)=\lim_{n\to\infty}\sum_{k=1}^{n}f\left(\frac{x}{k}\right)$. As with the specific $f(x) = x - \ln(1 + x)$ above, the general function f must be analytic over an open interval about zero, $a < 0 < b$. Likewise, its value and first derivative at zero must be zero, i.e., $f(0) = f'(0) = 0$. According to the second-order mean value theorem, for $x \in (a, b)$, there exists a real ξ between zero and x such that:

$$f(x)=f(0)+f'(0)x+\frac{f''(\xi)}{2}x^2=\frac{f''(\xi)}{2}x^2$$

[20]See also Appendix A.3 on the Euler-Mascheroni constant γ.

[21]Actually. we proved the convergence of the second term, the one with the limit, for $\alpha > -1$. It is the first term, the logarithm, that contracts the domain of $\ln\Gamma(\alpha)$ to $\alpha > 0$. As a complex function $\ln\Gamma(z)$ converges for all complex $z \notin \{0, -1, -2, \ldots\}$. However, $\ln z$ is multivalent, or unique only to an integral multiple of $2\pi i$. That exponentiation removes this multivalence is the reason why $e^{\ln\Gamma(z)} = \Gamma(z)$ is analytic, except at $z \notin \{0, -1, -2, \ldots\}$. The infinite product $\Gamma(z)=\lim_{n\to\infty}\frac{e^{z\ln n}}{z\left(1+\frac{z}{1}\right)\cdots\left(1+\frac{z}{n}\right)}$ converges, unless some factor in the denominator equals zero, i.e., unless $z = \{0, -1, -2, \ldots\}$.

Therefore, there exist ξ_k between zero and $\frac{x}{k}$ such that $\varphi(x) = \lim_{n\to\infty}\sum_{k=1}^{n} f\left(\frac{x}{k}\right) = \lim_{n\to\infty}\sum_{k=1}^{n}\frac{f''(\xi_k)}{2}\left(\frac{x}{k}\right)^2$.

Again, we appeal to absolute convergence. If $\lim_{n\to\infty}\sum_{k=1}^{n}\left|\frac{f''(\xi_k)}{2}\left(\frac{x}{k}\right)^2\right| = \frac{x^2}{2}\lim_{n\to\infty}\sum_{k=1}^{n}\frac{|f''(\xi_k)|}{k^2}$ converges, then $\varphi(x)$ converges. But every ξ_k belongs to the closed interval Ξ between 0 and x. And since f is analytic over $(a, b) \supset \Xi$, f'' is continuous over Ξ. Since a continuous function over a closed interval is bounded, $0 \le |f''(\xi_k)| \le M$ for some real M. So, at length:

$$0 \le \lim_{n\to\infty}\sum_{k=1}^{n}\frac{|f''(\xi_k)|}{k^2} \le \lim_{n\to\infty}\sum_{k=1}^{n}\frac{M}{k^2} = M\lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{k^2}$$

$$= M\zeta(2) = M\frac{\pi^2}{6}$$

Consequently, $\lim_{n\to\infty}\sum_{k=1}^{n} f\left(\frac{x}{k}\right)$ converges to a real number for every function f that is analytic over some domain about zero and for which $f(0) = f'(0) = 0$. The convergence must be at least of the second order; a first-order convergence would involve $\lim_{n\to\infty}\sum_{k=1}^{n}\frac{M}{k}$ and the divergence of the harmonic series.

The same holds true for analytic functions over complex domains. Another potentially important second-order convergence is $\varphi(z) = \lim_{n\to\infty}\sum_{k=1}^{n}\left\{e^{\frac{z}{k}} - \left(1 + \frac{z}{k}\right)\right\}$.