Generalized Erlang Derivation

Mike Wilensky, Jordan Brown

October 2018

1 Problem Statement

We concern ourselves with computing the probability density function for a sum of N exponential random variables with n distinct means β_j, where $n \leq N$. If $n < N$, then there are some β_j that have a multiplicity greater than 1. In other words, we allow for degeneracy. We denote the multiplicity of each β_j as m_j.

Let us denote the characteristic function of a random variable, X, as

$$g_X(t) = \int_{-\infty}^{+\infty} f_X(x) e^{itx} dx \tag{1}$$

where $f_X(x)$ is the probability density function (PDF) for X. In other words, $g_X(t)$ is the Fourier transform of $f_X(x)$. This means that equation 1 can be inverted with

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} g_X(t) e^{-itx} dt \tag{2}$$

2 A Single Exponential Random Variable

The PDF for an exponential random variable with mean β has the form

$$f_X(x) = \frac{1}{\beta} e^{-x/\beta} \tag{3}$$

The calculation will be slightly less messy if we consider the parameter

$$\lambda = \frac{1}{\beta}$$

in which case equation 3 looks like

$$f_X(x) = \lambda e^{-\lambda x} \tag{4}$$

We will refer to λ as the "rate constant." The characteristic function of an exponential random variable with rate constant λ has the form

$$g_X(t) = \frac{\lambda}{\lambda - it} = \frac{i\lambda}{i\lambda + t} \tag{5}$$

Note that if we consider t as a complex variable, then this function has a pole at

$$t = -i\lambda$$

3 Calculating the Sum PDF

To calculate the PDF of a sum of random variables, one must convolve their respective PDFs. By the convolution theorem, the resulting characteristic function is the product of the respective characteristic functions of the random variables. Our method is to calculate the characteristic function of the sum and then use an inverse Fourier transform (equation 2) to calculate the desired PDF.

Let us define the variable

$$Y = \sum_{k=1}^{N} X_k$$

where each X_k is exponential and there may be degeneracy in the means as discussed in the problem statement. Then the characteristic function of Y, using equation 5, is

$$g_Y(t) = \prod_{k=1}^{N} \frac{i\lambda_k}{i\lambda_k + t}$$

Using the multiplicities, we may write this as

$$g_Y(t) = \prod_{j=1}^{n} \left(\frac{i\lambda_j}{i\lambda_j + t} \right)^{m_j} = i^N \prod_{j=1}^{n} \left(\frac{\lambda_j}{i\lambda_j + 1} \right)^{m_j} \tag{6}$$

Applying equation 2, we see that the desired PDF is found by calculating

$$f_Y(y) = \frac{i^N}{2\pi} \int_{-\infty}^{+\infty} dt e^{-ity} \prod_{j=1}^{n} \left(\frac{\lambda_j}{i\lambda_j + t} \right)^{m_j} \tag{7}$$

This integral is evaluated simply using contour integration techniques. Consider t as a complex variable. Since for all j

$$\lambda_j > 0$$

then all of the poles of the integrand in equation 7 lie on the negative imaginary axis. Now consider when $y > 0$. Then consider a ***counter-clockwise*** infinite semi-circle contour enclosing the lower half of the complex plane. Temporarily dropping the integrand for ease of notation, we may write

$$\oint_C = -\int_{-\infty}^{+\infty} + \int_{\text{arc}} \tag{8}$$

where the minus sign comes in because of the ***counter-clockwise*** sense of the contour. The integrand satisfies Jordan's lemma[1] in the lower half-plane.

This mean's the integral over the arc vanishes. The integral over $\mathbb{R}$ is the one we desire, and the contour integral may be evaluated using the method of residues. This essentially states[1] for a complex function, $f(z)$,

$$\oint_C f(z)dz = \sum_{j=1}^{n} \oint_{C_j} f(z)dz \tag{9}$$

where the C_j enclose each of the poles of the integrand in a counter-clockwise sense. The integral over each pole may be evaluated with Cauchy's integral formula, whose generalization[1] states that for an analytic function $h(z)$ on a closed contour C, for *any* point z_0 interior to C,

$$h^{(m)}(z_0) = \frac{m!}{2\pi i} \oint_C \frac{h(z)}{(z-z_0)^{m+1}} dz \tag{10}$$

where $h^{(m)}(z_0)$ is the mth derivative of h evaluated at z_0.

To proceed, we evaluate each residue for the pole of order m_j one at a time using this formula, setting

$$h_j(t) = \frac{1}{2\pi}(i\lambda_j + t)^{m_j} e^{-ity} g_Y(t) \tag{11}$$

so that

$$\oint_C = 2\pi i \sum_{j=1}^{n} \frac{1}{(m_j-1)!} h_j^{(m_j-1)}(-i\lambda_j) \tag{12}$$

which may be written

$$\oint_C = i^{N+1} \sum_{j=1}^{n} \frac{\lambda_j^{m_j}}{(m_j-1)!} \left(\frac{d}{dt}\right)^{m_j-1} \left(e^{-ity} \prod_{l\neq j} \left(\frac{\lambda_l}{i\lambda_l + t}\right)^{m_l} \right)\Bigg|_{t=-i\lambda_j} \tag{13}$$

We can expand the derivatives using the generalization of the product rule. We split the term into two factors: the exponential factor, and the product of rational functions, obtaining

$$\oint_C = i^{N+1} \sum_{j=1}^{n} \frac{\lambda_j^{m_j}}{(m_j-1)!} e^{-ity} \sum_{m=0}^{m_j-1} \left[\binom{m_j-1}{m} (-iy)^{m_j-m-1} \frac{d^m}{dt^m} \left(\prod_{l\neq j} \left(\frac{\lambda_l}{i\lambda_l + t}\right)^{m_l} \right) \right]\Bigg|_{t=-i\lambda_j} \tag{14}$$

Now we need to collect our factors of i. When the evaluation is made at $t = -i\lambda_j$, a factor of i can be taken out for each factor of $i\lambda_l + t$. There are $N - m_j$ such factors before differentiation, and differentiation adds m such factors. Pulling all these out and remembering the factors attached to the powers of y we have a total factor of

$$(-i)^{N-m_j+m+m_j-m-1} = (-i)^{N-1}$$

Combining these with the factors of i out front gives

$$\oint_C = -\sum_{j=1}^{n}\sum_{m=0}^{m_j-1}\frac{\lambda_j^{m_j}}{(m_j-m-1)!m!}y^{m_j-m-1}e^{-\lambda_j y}\Psi_{mj}(-\lambda_j) \tag{15}$$

where

$$\Psi_{mj}(t) = \frac{d^m}{dt^m}\left[\prod_{l\neq j}\left(\frac{\lambda_l}{\lambda_l+t}\right)^{m_l}\right]$$

Recalling equation 8, we have, for $y > 0$,

$$f_Y(y) = \sum_{j=1}^{n}\sum_{m=0}^{m_j-1}\frac{\lambda_j^{m_j}}{(m_j-m-1)!m!}y^{m_j-m-1}e^{-\lambda_j y}\Psi_{mj}(-\lambda_j) \tag{16}$$

For $y < 0$, we instead choose a counter-clockwise infinite semicircle contour enclosing the upper half-plane. The integral over the arc vanishes the same as before, but the contour encloses no poles. This means that The contour integral evaluates to zero, and so also the integral over the real line evaluates to zero. Thus, we summarize this general result using a Heavisde step function, $\theta(y)$

$$f_Y(y) = \theta(y)\sum_{j=1}^{n}\sum_{m=0}^{m_j-1}\frac{\lambda_j^{m_j}}{(m_j-m-1)!m!}y^{m_j-m-1}e^{-\lambda_j y}\Psi_{mj}(-\lambda_j) \tag{17}$$

This makes sense, since we are evaluating the density for a sum of non-negative random variables.

Let us now consider two simpler cases. First, let us consider the case where there is only one distinct rate constant. Then the products and sums over j are only over one term. In equation 13, the product over $l \neq j$ is just set equal to 1, and the derivatives only hit the exponential term, so the generalized product rule does not follow through. We are left with

$$f_Y(y) = \frac{\lambda^N y^{N-1}}{(N-1)!}e^{-\lambda y}\theta(y) \tag{18}$$

which is the Erlang distribution that we were already familiar with.

Second, let us consider when all the rate constants are distinct. Then $m_j = 1$ for all j. All the poles are simple poles, so all the derivatives are of 0th order. The result is then

$$f_Y(y) = \theta(y)\sum_{j=1}^{N}\lambda_j e^{-\lambda_j y}\prod_{l\neq j}\frac{\lambda_l}{\lambda_l-\lambda_j} \tag{19}$$

4 Calculation of the Means and Variances

We will calculate the means and variances separately for each case.

4.1 The Fully Degenerate Case

We begin from equation 18. The mean is given by

$$E[Y] = \int_0^\infty \frac{(\lambda y)^N}{(N-1)!} e^{-\lambda y} dy = \frac{N}{\lambda} \tag{20}$$

The second moment is given by

$$E[Y^2] = \int_0^\infty \frac{\lambda^N y^{N+1}}{(N-1)!} e^{-\lambda y} dy = \frac{N(N+1)}{\lambda^2} \tag{21}$$

Therefore the variance is given by

$$E[Y^2] - E[Y]^2 = \frac{N}{\lambda^2} \tag{22}$$

4.2 The Nondegenerate Case

Now we begin with equation 19. The integrations are simple. The results are as follows.

$$E[Y] = \sum_{j=1}^{N} \frac{1}{\lambda_j} \prod_{l \neq j} \frac{\lambda_l}{\lambda_l - \lambda_j} \tag{23}$$

$$E[Y^2] = \sum_{j=1}^{N} \frac{2}{\lambda_j^2} \prod_{l \neq j} \frac{\lambda_l}{\lambda_l - \lambda_j} \tag{24}$$

So the variance is given by

$$E[Y^2] - E[Y]^2 = \sum_{j=1}^{N} \frac{2}{\lambda_j^2} \prod_{l \neq j} \frac{\lambda_l}{\lambda_l - \lambda_j} - \sum_{p=1,q=1}^{N} \frac{1}{\lambda_p \lambda_q} \prod_{m \neq p, n \neq q} \frac{\lambda_m \lambda_n}{(\lambda_m - \lambda_p)(\lambda_n - \lambda_q)} \tag{25}$$

where all possible pairs of p and q must be evaluated, and all possible pairs of m and n must likewise be evaluated for each pair of p and q.

4.3 The General Case

We must now start with equation 17. The results from integration are

$$E[Y] = \sum_{j=1}^{n} \sum_{m=0}^{m_j - 1} \frac{(m_j - m)\lambda_j^{m-1}}{m!} \Psi_{mj}(-\lambda_j) \tag{26}$$

$$E[Y^2] = \sum_{j=1}^{n} \sum_{m=0}^{m_j - 1} \lambda_j^{m-2} \frac{(m_j - m + 1)(m_j - m)}{m!} \Psi_{mj}(-\lambda_j) \tag{27}$$

The full expression for the variance is cumbersome and does not simplify further. Simply calculate $E[Y^2]$ and $E[Y]$ as above and use the expression $E[Y^2] - E[Y]^2$. A quick inspection will show that this case easily simplifies down to the nondegenerate case if $m_j = 1$ for all j.

5 Convergence to a Gaussian Distribution

To prove convergence to a Gaussian distribution in the large N limit, we will make use of the Lévy continuity theorem, which states that if a sequence of characteristic functions of a random sequence converges pointwise to the characteristic function of a given random variable, then the random sequence converges in distribution to that given random variable.

Let us consider our N exponential random variables, X_i, with rate constants, $\{\lambda_j\}$. Let us consider the random variables defined by

$$Z_j = \frac{(X_j - \mu_j)}{\sqrt{N}\sigma_j} = \frac{\lambda_j X_j - 1}{\sqrt{N}} \tag{28}$$

We will show that the variable defined by

$$Z = \sum_{j=1}^{N} Z_j \tag{29}$$

converges to a standard normal random variable (zero-mean, unit width, Gaussian) in the large N limit.

The characteristic function of Z_j, $\phi_{Z_j}(t)$, is given by

$$\phi_{Z_j}(t) = E[e^{itZ_j}]_{Z_j} = \int_{-\infty}^{\infty} dz_j e^{itz_j} f_{Z_j}(z_j). \tag{30}$$

We can express this as an expectation over X_j like so:

$$\phi_{Z_j}(t) = \int_{-\infty}^{\infty} dx_j e^{it(\lambda_j x_j - 1)/\sqrt{N}} f_{X_j}(x_j) = e^{-it/\sqrt{N}} \phi_{X_j}\left(\frac{\lambda_j t}{\sqrt{N}}\right) \tag{31}$$

Inspecting equation 5, we see that this is

$$\phi_{Z_j}(t) = e^{-it/\sqrt{N}} \frac{1}{1 - \frac{it}{\sqrt{N}}} \tag{32}$$

The characteristic function of Z is the product of these characteristic functions.

$$\phi_Z(t) = e^{-it\sqrt{N}} \left(\frac{1}{1 - \frac{it}{\sqrt{N}}}\right)^N \tag{33}$$

The logarithm of this function will be easier to work with.

$$\log(\phi_Z(t)) = -i\sqrt{N}t - N\log\left(1 - \frac{it}{\sqrt{N}}\right) \tag{34}$$

The second term can be expanded in a power series with a radius of convergence of $\sqrt{N}$. Therefore, in the large N limit, this series converges everywhere. This gives

$$\log(\phi_Z(t)) = -i\sqrt{N}t + N\sum_{k=0}^{\infty} \frac{(it/\sqrt{N})^{k+1}}{k+1} \tag{35}$$

The linear term in the power series exactly cancels the first term in the expression that came from the exponential, giving

$$\log(\phi_Z(t)) = -\frac{1}{2}t^2 + O(|t^3|/\sqrt{N}) \tag{36}$$

Then, in the limit as N tends to infinity, the higher order terms disappear, giving (undoing the log),

$$\lim_{N\to\infty} \phi_Z(t) = e^{-t^2/2} \tag{37}$$

which is the characteristic function of a standard normal random variable, as desired.

5.1 An Alternative Proof not Involving the Complex Logarithm

In order to avoid confusions about branch cuts, below is a development that keeps things in the exponent. Let us start with equation 33. Consider the complex number defined by

$$w = 1 - \frac{it}{\sqrt{N}}. \tag{38}$$

Since t is real-valued, w is always in the right-half of the complex plane. We may rewrite w in the following way using the principal branch of the arctangent:

$$w = \sqrt{1 + \frac{t^2}{N}} e^{-i\arctan(t/\sqrt{N})} \tag{39}$$

Notice that equation 33 may now be written

$$\phi_Z(t) = e^{-it\sqrt{N}} w^{-N}, \tag{40}$$

which may be further expanded to

$$\phi_Z(t) = e^{-it\sqrt{N}} \frac{1}{\sqrt{(1 + \frac{t^2}{N})^N}} e^{iN\arctan(t/\sqrt{N})} \tag{41}$$

Given a fixed t, for sufficiently large N, the following approximation holds well:

$$\arctan(t/\sqrt{N}) \approx \frac{t}{\sqrt{N}} \tag{42}$$

Ultimately we will take the limit as N tends to infinity to show convergence. If we use the above approximation, then the two exponentials in equation 41 cancel exactly, leaving

$$\lim_{N\to\infty} \phi_Z(t) = \lim_{N\to\infty} \frac{1}{\sqrt{(1 + \frac{t^2}{N})^N}} \tag{43}$$

This is none other than

$$\lim_{N\to\infty} \phi_Z(t) = \sqrt{e^{-t^2}} = e^{-t^2/2} \tag{44}$$

exactly as desired.

5.2 Rate of Convergence

6 References

[1] *Mathematics of Classical and Quantum Physics*, Frederick W. Byron, Jr. and Robert W. Fuller, Dover edition.