

RESEARCH

Open Access

Describing the Flexibility of the Generalized Gamma and Related Distributions



Matthew Matheson*, Alvaro Muñoz and Christopher Cox

* Correspondence:
mmathes4@jhu.edu
Johns Hopkins Bloomberg School
of Public Health, 615 N Wolfe St.,
Baltimore, MD 21205, USA

Abstract

The generalized gamma (GG) distribution is a widely used, flexible tool for parametric survival analysis. Many alternatives and extensions to this family have been proposed. This paper characterizes the flexibility of the GG by the quartile ratio relationship, $\log(Q2/Q1)/\log(Q3/Q2)$, and compares the GG on this basis with two other three-parameter distributions and four parent distributions of four or five parameters. For most parameter combinations of other distributions, a very similar GG, as assessed by the Kullback-Liebler distance, can be found by matching the three quartiles; extreme cases where this fails are examined. Limited additional flexibility is observed, supporting the basic GG family as an ideal platform for parametric survival analysis.

Keywords: Generalized gamma, Parametric survival analysis, Hazard functions, Kullback-Liebler distance

Introduction

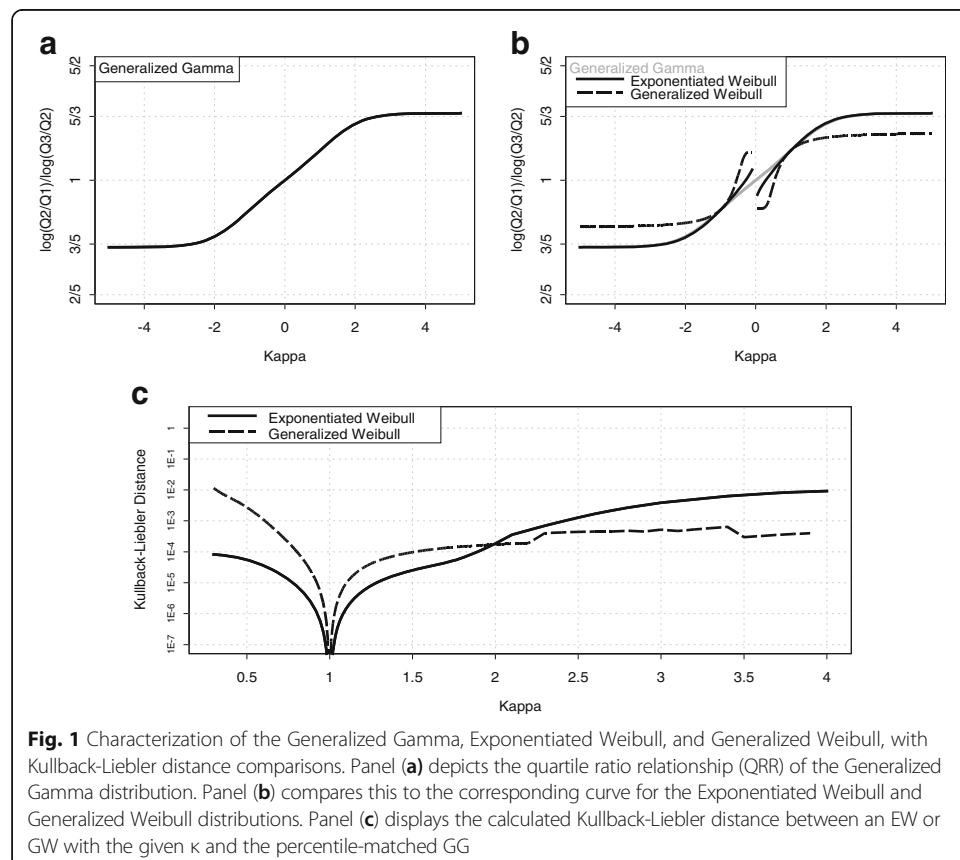
Parametric survival analysis has been the source for the development of distributions with richness and flexibility for modeling time-to-event data. Simple, familiar distributions such as the lognormal and Weibull have been extended, transformed, and combined into myriad new and complex distributions. Distributions with more than three parameters remain relatively rarely applied, with the Generalized Gamma (GG) being among the most popular choices at the three-parameter level. The major appeal of this distribution is its hazard behavior, which includes all four basic hazard shapes (increasing, decreasing, bathtub, and arc-shaped), as well as its ready implementation in standard statistical software packages. These features are described at length in a tutorial by Cox et al. (2007).

While the GG is broadly applicable and flexible, there are still many kinds of hazards, even among the four basic shapes, which it cannot accommodate. We have investigated competing distributions including the three-parameter Exponentiated Weibull (EW; Cox and Matheson 2014) and a family that includes the GG as a special case, the five-parameter Beta-Generalized Gamma (Matheson and Cox 2017). Both of these families include all four of the four basic hazard shapes. Our approach was to attempt to find a closely-matching GG (including both the survival and hazard functions) for any given member of these two families, using the Kullback-Liebler distance to assess the closeness of the match. Each of these comparisons has led to the somewhat surprising conclusion that the GG itself continues to be a good choice for modeling data,

in the sense that given any member of the either of these two families, a GG can be found whose survival and hazard functions are very similar and in many cases, including the EW, indistinguishable. In this paper, we characterize an aspect of the GG's flexibility by interpreting the shape parameter, κ , in terms of the relationship between the three quartiles of the distribution. This gives us a valuable tool for comparing the GG to other distributions using our matching approach, which we pursue with six competing distributions of three to five parameters. Special cases of competing distributions which may not be well-approximated by a GG are evaluated.

A More Complete Characterization of the Shape of the GG

The standard accelerated failure time model involves location (β) and scale (σ) parameters, and in the case of the GG, an additional 'shape' parameter κ . Here we propose a more concrete interpretation of κ as governing the **quartile ratio relationship (QRR)**, defined as $\log(Q2/Q1)/\log(Q3/Q2)$. For any GG, the QRR curve depends solely on κ , such that the curve defined by $QRR = \log(t_{\kappa}(0.5)/t_{\kappa}(0.25))/\log(t_{\kappa}(0.75)/t_{\kappa}(0.5))$, where $t_{\kappa}(p)$ represents the p^{th} percentile of a $GG(0, 1, \kappa)$, is identical for any given β and σ . Indeed, there is a one-to-one correspondence between κ (independent of β and σ) and the QRR; this function is shown in Fig. 1, panel a, with the y-axis logarithmic to highlight the function's symmetry (namely, that the QRR for any negative κ is the reciprocal of the QRR for the corresponding positive κ value).



The implications of this graph are substantial, but several in particular are relevant to our current investigation. First, if three quartiles of data are known, one can determine the κ of the GG that most appropriately fits the data, if one exists, using this function (Cox and Matheson, 2014). As long as the QRR lies between roughly $3/5$ and $5/3$, the appropriate κ for the given QRR can be found; then, given this value of κ , σ is simply $\log(Q3/Q1)/\log(t_{\kappa}(0.75)/t_{\kappa}(0.25))$ and β is $\log(Q2) - \sigma \log t_{\kappa}(0.5)$. If the QRR falls outside this range, κ should be restricted to 4 for positive $\log(QRR)$ values and -4 for negative ones, as more extreme values of κ will not bring the GG any closer to the desired QRR; σ and β can still be solved for as shown above. This leads to the second point, which is that the flexibility of other distributions can be compared to the GG by plotting their QRR. If the QRR of another distribution falls entirely between the limits of the QRR of the GG, that distribution is not likely to have any additional flexibility beyond what the GG already provides. However, if the competing distribution's QRR extends outside this vertical range, the GG may be limited in its capacity to approximate it.

Matching a GG to a Competitor Distribution

Given any parametric family, one can choose parameter values, calculate the three quartiles of the resulting distribution, evaluate the QRR, and determine whether there is a GG with the same QRR as described above. This process is entirely independent of data, simulation, or considerations of censoring; it is a purely theoretical exercise for matching two distributions. Cox and Matheson (2014) have shown that percentile matching in this way compares favorably to generating simulated data for determining matching GGs. The resulting matched distributions will not necessarily be mathematically identical. The Kullback-Liebler distance (KLD) is used to measure the closeness between a competing distribution and its matched GG (Cox and Matheson 2014).

Below, we apply this process to six competitor distributions. For each distribution, we examine its QRR and compare the QRR to that of the GG; select several parameter combinations to capture different hazard shapes and QRR values; find matching GGs for each competitor; and assess the closeness of the match numerically (using the KLD) as well as visually.

3. Competitors to the Generalized Gamma

Alternate Three-Parameter Distributions

Cox and Matheson (2014) previously investigated the exponentiated Weibull, another family having all four of the basic hazard shapes, as a competitor to the GG. They found that given any member of the EW family, a matching GG can be found whose survival and hazard functions are indistinguishable. Another three-parameter family also having the four basic hazard shapes is the Generalized Weibull (GW), which is most easily defined by its CDF:

$$F_{GW}(t) = 1 - \exp\left(1 - \left(1 + (e^{-\beta}t)^{1/\sigma\kappa}\right)^{\kappa^2}\right)$$

This expression gives the CDF for positive values of κ . Both the GW and the previously discussed EW can be extended to $\kappa < 0$ by $F(t; \kappa < 0) = 1 - F(t; -\kappa)$, that is, the

complement of the CDF with the corresponding positive value of κ . We can then easily plot the QRR of the EW and GW distributions for a range of values of κ ; this is shown in Fig. 1, panel b. The QRR of the EW very closely parallels that of the GG as κ gets further away from 0, appearing to have the same limit, thus supporting the results of our previous investigation. The GW proves to be even more limited, with a smaller QRR range that lies completely within the range of the GG. Consistent with this limited range, the GW distribution also fails to offer any features beyond the GG since a matching GG can again be found whose survival and hazard functions are indistinguishable. Figure 1, panel c displays the KLD between the EW and GW, for a range of absolute values of their κ parameter, and their percentile-matched GGs. While the largest discrepancies are observed for large values of κ for the EW and small values of κ for the GW, the majority of KLDs are on the order of 10^{-3} or lower, which is very good agreement; for such a KLD, the survival and hazard functions of the competing distribution and its matched GG are visually indistinguishable. For comparison, the KLD between the standard normal and standard logistic distributions, commonly regarded as “close” matches, is 1.436×10^{-2} . A slightly but noticeably raised section in the KLD curve for the GW, from roughly $\kappa = 2.3$ to $\kappa = 3.4$, shows the difficulty in calculating the numerical integral for the KLD in some extreme cases.

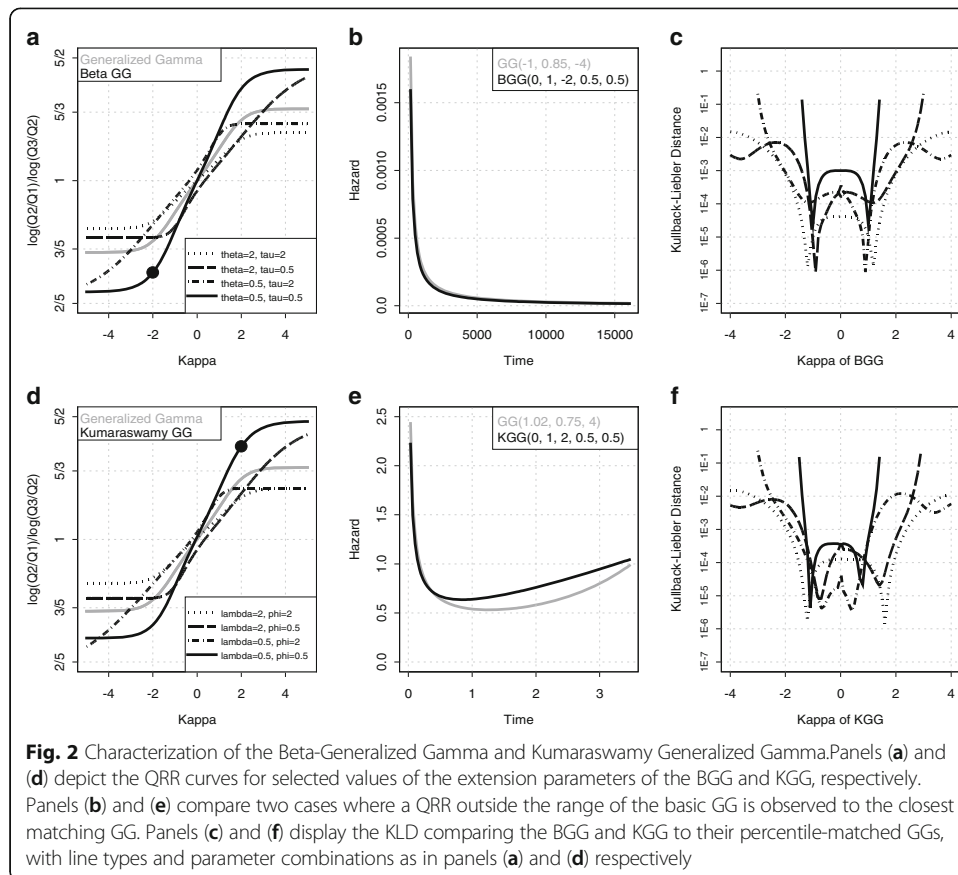
Four- and Five-Parameter Extensions of the GG

There are many extensions of the GG that transform the cumulative distribution function (CDF) using one or two additional parameters. Matheson and Cox (2017) previously investigated the Beta-Generalized Gamma distribution (BGG), failing to find a completely matching GG only for small values of the beta parameters θ and τ , when the behavior of the BGG becomes somewhat unusual. Other candidates include the five-parameter Kumaraswamy GG (KGG; de Pascoa et al. 2011), the four-parameter transmuted GG (TGG; Lucena et al. 2015), and the four-parameter Marshall-Olkin GG (MOGG; Tahir and Nadarajah 2015). Simplified versions of the CDFs of these distributions, illustrating how they transform the basic GG, are shown in Table 1. In order to evaluate the QRR of these extended distributions, we must fix values of the additional parameters.

As shown by Matheson and Cox (2017) the Beta-GG, for $\theta > 1$ and $\tau > 1$, is well-approximated by the GG; consistent with this, we note here that the QRR range is compressed for these parameter values. However, allowing one or both parameters to fall below 1 expands the range; lower QRRs are possible if $\theta < 1$, while higher QRRs are possible if $\tau < 1$. The QRR curves for several combinations of (θ, τ) are shown in Fig. 2, panel a to illustrate this; panel b compares the hazard functions of the BGG(0, 1, -2, 0.5, 0.5) to the closest approximating GG, GG(-1, 0.85, -4). Despite the necessity of

Table 1 Selected Extensions of the Generalized Gamma Distribution

Distribution	CDF	Additional parameters
Generalized Gamma	$F_{GG}(t)$	n/a
Beta GG	$F_{BGG}(t) = F_B(F_{GG}(t)) = \frac{B(F_{GG}(t); \theta, \tau)}{B(\theta, \tau)}$	$\theta > 0, \tau > 0$
Kumaraswamy GG	$F_{KGG}(t) = 1 - (1 - F_{GG}(t))^\varphi$	$\lambda > 0, \varphi > 0$
Transmuted GG	$F_{TGG}(t) = (1 + \lambda)F_{GG}(t) - \lambda F_{GG}(t)^2$	$-1 \leq \lambda \leq 1$
Marshall-Olkin GG	$F_{MOGG}(t) = 1 - \frac{\alpha(1-F_{GG}(t))}{1-(1-\alpha)(1-F_{GG}(t))}$	$\alpha > 0$



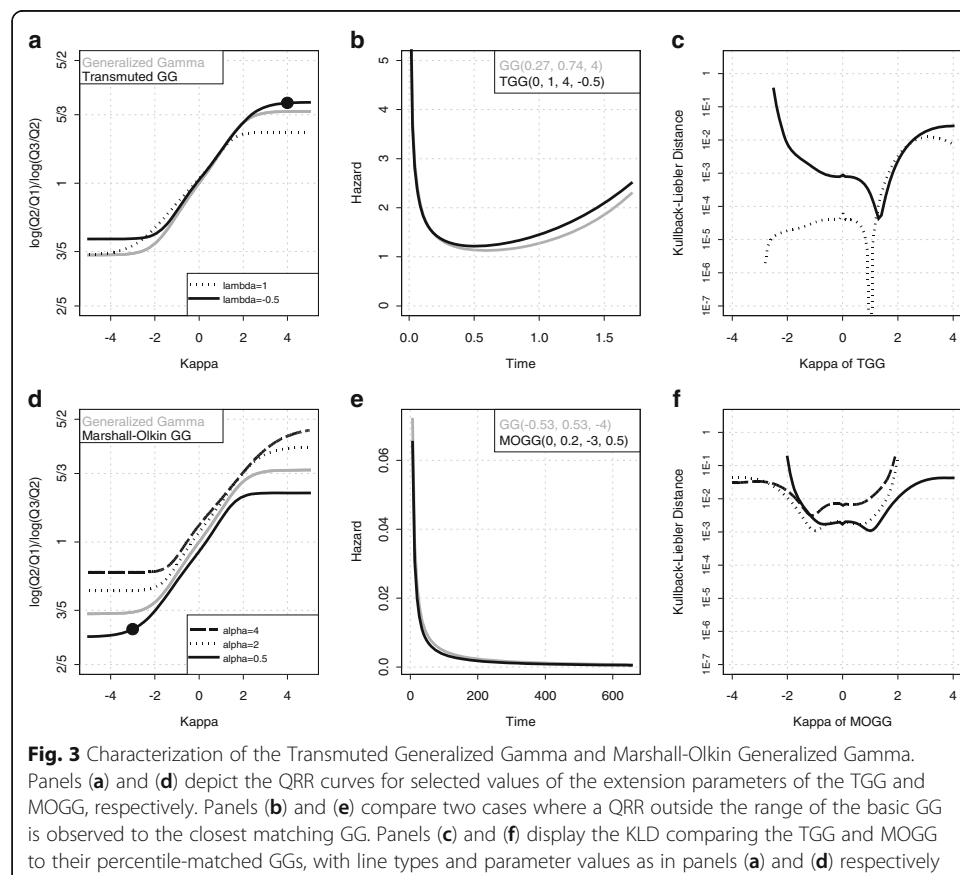
settling for an imperfect percentile match, the hazard functions are remarkably similar. The KLD between these two distributions is calculated as 1.702, a relatively large value for this measure. Figure 2, panel c displays curves of the KLD for the various combinations of (θ, τ) utilized in panel a. For $(\theta = 2, \tau = 2)$, the case for which the QRR lies fully within the range of the GG, the matched GG maintains a good agreement across values of the BGG's κ , with values less than 10^{-4} for small absolute values of κ , rising to around 10^{-2} for large absolute values of κ . For $(\theta = 2, \tau = 0.5)$, positive values of κ enable close GG matches, with the largest KLDs around 10^{-2} ; however, negative values of κ begin to become unmatchable, leading to relatively large or incomputable KLDs. The reverse is true for $(\theta = 0.5, \tau = 2)$, as the KLD curve is a mirror image of the previous case. However, for $(\theta = 0.5, \tau = 0.5)$, only very small absolute values of κ can be matched well, with KLDs around 10^{-3} ; once $|\kappa|$ becomes greater than 1, matching cannot be achieved and the KLD of the “closest” fits are relatively poor. For clarity, we display the KLD lines only where good matches could be achieved.

The Kumaraswamy GG behaves very similarly. Values of $\lambda > 1$ or $\phi > 1$ compress the lower and upper end of the QRR, respectively, while values < 1 expand the corresponding tail. Figure 2, panel d illustrates this, and panel e compares the hazard of the KGG(0, 1, 2, 0.5, 0.5) to the closest approximating GG, GG(1.02, 0.75, 4); again we note the hazard functions are quite similar. The KLD between these two distributions is 2.060. Figure 2, panel f gives a fuller examination of the KLDs between many KGGs

and their matched GGs; the patterns are strikingly similar to those of the BGG matches in panel c, such that a detailed description would be effectively redundant.

The transformation induced by the additional parameter of the Transmuted GG is interesting. At the null value of $\lambda = 0$, the original GG – and therefore QRR – are maintained. At either extreme value, $\lambda = \pm 1$, the QRR is compressed in one tail and equal at the limit of the other. But other positive or negative values of lambda shift the QRR down or up, respectively, enabling some QRRs to be outside the range of the GG. Notably, though, these novel QRRs are only achieved for fairly extreme values of κ . The largest divergence comes around values of $\lambda = \pm 0.5$. This is shown in Fig. 3, panel a. Because the curves shift very little, we show only those for $\lambda = 1$ and $\lambda = -0.5$; the curves for $\lambda = -1$ and $\lambda = 0.5$ are symmetric, as the QRR for any given (κ, λ) is the reciprocal of the QRR for $(-\kappa, -\lambda)$. Figure 3, panel b compares the hazard functions of the TGG(0, 1, 4, -0.5) to the closest approximating GG, GG(0.27, 0.74, 4). Again, the hazards are very similar; the KLD between these distributions is 0.363, much closer than for the previous examples, which is consistent with the closeness of the QRR. Figure 3, panel c displays the KLDs for $\lambda = 0.5$ and $\lambda = 1$; because of the symmetry of the transformation, the KLD for any (κ, λ) is equal to the KLD for $(-\kappa, -\lambda)$. We can see that the KLD is quite small for positive values and small negative values of κ ; only when κ dips below -2 do the matches and distances start to lose quality.

The “tilt” parameter of the Marshall-Olkin GG is much more powerful for expanding the QRR range. Any value of $\alpha > 1$ produces an upward shift in the whole curve



(although $\alpha \gg 1$ will bend the curve in such a way that part of it will lie below the GG's curve), while $\alpha < 1$ produces a corresponding downward shift (with the corresponding caveat for $\alpha \ll 1$). Figure 3, panel d illustrates these shifts, while panel e compares the hazard of the MOGG(0, 1, -3, 0.5) to the closest approximating GG, GG(-0.53, 0.53, -4). The agreement is considerable, and the KLD between these distributions is 1.867. Figure 3, panel f shows the KLD for a range of κ ; because of the symmetry of this transformation, the KLD for any (κ, α) is equal to the KLD for $(-\kappa, 1/\alpha)$. For $\alpha = 2$, even the close matches are not great, with KLDs ranging from 10^{-3} up to 10^{-1} for negative values of κ , while $\kappa > 2$ leads to poor matches and high KLDs. As the tilt parameter increases, the matching becomes more difficult and the KLDs accordingly higher.

4. Discussion

In this manuscript, we have described the flexibility of the generalized gamma distribution in terms of the relationship among its three quartiles, and demonstrated how this quantity can aid in simple GG matching and quickly allow comparisons of the similarity between the GG and other distributions. We compared a matching GG to members of six competing distributions via the QRR and KLD, and in particular highlighted several specific cases where the match was relatively poor but the hazard behavior of the extended distribution was only slightly different from that of the matched GG. Overall, the graphs of the KLD that we have provided show that the quality of the match can vary considerably, as the values of the KLD vary over several orders of magnitude. However, in the vast majority of cases the KLD was near or below 10^{-2} , the magnitude for the normal and logistic distributions, which are considered to be relatively close.

An investigation of this scope is necessarily limited, as there are literally infinitely many possible parameter values to evaluate. However, the simplicity of the QRR as a summary measure of the flexibility of a distribution combined with the plainly observable effect on the QRR of shifting extension parameters makes even this limited exploration of these families robust and illuminating. There are also certainly many other possible extensions of the GG or alternatives to it. About these we can only broadly say that, based on the results seen here, it will generally require fairly extreme parameter values to extend the QRR beyond that of the basic GG, and we welcome further research to explore this comparison in other distributions.

In previous work, we noted that extreme parameter values of the Beta-GG in the range that induce a QRR outside the range of the GG produce distributions which may be numerically unstable or difficult to estimate. The same is true for extreme cases of the Kumaraswamy, Transmuted, and Marshall-Olkin GGs; while each one has some flexibility beyond the GG, this often involves extreme values of κ as well as the additional parameters. It seems relatively rare to have a scenario that features both 1) data which is well-fit by, say, a Marshall-Olkin GG but not a traditional GG, and 2) sufficient observations to facilitate accurate estimation of the extreme parameter values. We have also seen that even cases which fall beyond the actual QRR range of the GG produce data which are still reasonably well estimated by a GG.

It is important to recognize that there are indeed nonstandard hazard shapes achievable through specific (though often extreme) parameter combinations for some of these distributions. An extreme tilt parameter in the Marshall-Olkin GG can produce an arc-

bathtub or bathtub-arc shape (e.g., the MOGG(0, 1.2, 1, 3)); a Beta-GG with very small θ , τ , and σ can actually produce a double arc (or “m-shaped”) hazard (e.g., the BGG(0, 0.3, 0, 0.2, 0.2)). In each of these examples, the QRR of the MOGG or BGG with its unique hazard is well within the range of the GG, so a “good match” can easily be found, although it will have a bathtub or single-arc hazard shape, respectively. This highlights one limitation of the QRR as a comparison tool. We cannot guarantee that percentile-matched GGs will have the same hazard behavior as the distribution (or data) they are matched to, and we cannot discern the hazard taxonomy of a distribution based on its QRR curve. However interesting these cases are, again, this requires parameter values which might be difficult to estimate from actual data, as the estimation procedure would encounter numerical instability. Our focus is not on breaking out of the four standard hazard shapes, but rather finding potential additional richness within those hazard shapes, and our investigations, as presented here, have yielded very little in this respect.

The generalized gamma is a rich, robust, and flexible parametric distribution for modeling many types of data, and while many extensions and competitors have been studied, those we have examined add very little to its capabilities. Because of the complexity of estimating these extended distributions and the minimal potential benefit of doing so, we continue to recommend the three-parameter generalized gamma as the standard for parametric analysis of positive data.

Abbreviations

BGG: Beta-generalized gamma; CDF: Cumulative distribution function; EW: Exponentiated weibull; GG: Generalized gamma; GW: Generalized weibull; KGG: Kumaraswamy generalized gamma; KLD: Kullback-Liebler distance; MOGG: Marshall-Olkin generalized gamma; Q1: First quartile; Q2: Second quartile; Q3: Third quartile; QRR: Quartile ratio relationship; TGG: Transmuted generalized gamma

Acknowledgements

We thank Dr. Carl Lee and Dr. Felix Famoye for inviting the presentation of a part of this paper at the 2016 International Conference on Statistical Distributions and Applications.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Authors' contributions

Matthew B. Matheson researched the distributions and their relationships to the generalized gamma, wrote the manuscript, and created figures. Dr. Alvaro Muñoz is credited with the development of the quartile ratio relationship graph and reviewed the manuscript and figures. Dr. Christopher Cox initiated the investigation of the relationships between the generalized gamma and other distributions, provided suggestions for competing distributions, and reviewed the manuscript and figures. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

None to declare.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 31 March 2017 Accepted: 27 July 2017

Published online: 01 November 2017

References

- Cox, C., Matheson, M.B.: Comparison of the generalized gamma and exponentiated Weibull distributions. *Stat. Med.* **33**, 3772–3780 (2014)
- Cox, C., Chu, H., Schneider, M.F., Muñoz, A.: Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat. Med.* **26**, 4352–4374 (2007)
- de Pascoa, M.A.R., Ortega, E.M.M., Cordeiro, G.M.: The Kumaraswamy generalized gamma distribution with application in survival analysis. *Stat. Meth.* **8**, 411–433 (2011)
- Lucena, S.E.F., Silva, A.H.A., Cordeiro, G.M.: The transmuted generalized gamma distribution: properties and application. *J. Data Sci.* **13**, 409–420 (2015)
- Matheson, M.B., Cox, C.: The shape of the hazard function: does the generalized gamma have the last word? *Comm. Stat. Th. Meth.* (2017). doi:10.1080/03610926.2016.1277757
- Tahir, M.H., Nadarajah, S.: Parameter induction in continuous univariate distributions: well-established G families. *Ann. Bra. Ac. Sci.* **87**, 539–568 (2015)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
