Accurate likelihood inference for the generalized exponential distribution

Paulo V. C. Pereira and Isolde T. S. Previdelli[1]

[1]Universidade Estadual de Maringá

Abstract

Our aim is to help popularize accurate likelihood inference and show some unconsidered aspects of its use, specifically for the generalized exponential distribution. This two-parameter distribution is similar to the gamma and Weibull distributions, and presents itself as a good competitor for analyzing lifetime (or skewed) data which have monotonic hazard rate function. Since the median lifetime is usually of interest in survival analysis, we also consider higher order inference for the quantiles of the generalized exponential.
Key-words: modified likelihood root; skewed data; small sample; tangent exponential model.

1 Introduction

The two-parameter generalized exponential distribution, also known as the exponentiated exponential distribution, is obtained by simply exponentiating with a constant the exponential distribution function,

$$F(y) = (1 - e^{-\sigma y})^{\alpha}, \quad y, \alpha, \sigma > 0. \tag{1}$$

This gives the advantage of a more general risk function,

$$h(y) = \frac{f(y)}{1 - F(y)} = \frac{\alpha\sigma(1 - e^{-\sigma y})^{\alpha-1} e^{-\sigma y}}{1 - (1 - e^{-\sigma y})^{\alpha}}, \quad y, \alpha, \sigma > 0. \tag{2}$$

whose shape does not depend on the scale parameter σ, and $\lim_{y \to +\infty} h(y) = \sigma$. For some fixed value of σ, the risk function is increasing if $\alpha > 1$, decreasing if $0 < \alpha < 1$, and constant at σ when $\alpha = 1$. The quantile function is

$$q_p = F^{-1}(p) = -\frac{1}{\sigma} \log(1 - p^{1/\alpha}), \quad p \in [0, 1). \tag{3}$$

We investigate the performance of some inference procedures for the parameters of the generalized exponential when the sample is small. Large sample theory is often applied to small samples with no second thoughts. As mentioned by Lozada-Can & Davison (2012), "it is not widely appreciated that standard likelihood theory can be readily improved and that the corresponding computations are relatively straightforward." We use simulation to compare, in terms of bias and root mean square error, different estimators for the shape and scale parameters. We consider maximum likelihood estimation, the method of moments, the method of L-moments, least squares and percentile estimators. Even though maximum likelihood estimators have the smallest possible asymptotic variances, the variance of the other estimators could be smaller for small or moderate samples. So, the estimators different from maximum likelihood may have good small-sample properties, but they only yield point estimates, not interval estimates, and they are not easy to extend to more complex settings. We describe improved inference using maximum likelihood estimation in the next section, whereas description of the other estimators can be found in Gupta & Kundu (2007).

2 Higher order inference

The following discussion is based on the work of Brazzale et al. (2007), Brazzale & Davison (2008) and Lozada-Can & Davison (2012). Given a random sample $y = (y_1, \dots, y_n)$ from a model with density function $f(y;\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}^d$ is a d-dimensional parameter, our interest relies on improved inference for one of the components of θ based on the log likelihood $l(\theta) = \log f(y;\theta)$. Let $\theta = (\psi, \lambda)$, where ψ is the scalar interest parameter, and λ the possibly vector nuisance parameter. Under standard regularity conditions, the distribution of the maximum likelihood estimator, $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$, converges to a standard normal distribution with mean θ and variance $j(\hat{\theta})^{-1}$, where $j(\theta) = -\partial^2 l(\theta)/\partial\theta\partial\theta^\mathsf{T}$ is the observed information function. The Wald statistic, $t(\theta) = j(\theta)^{1/2}(\hat{\theta} - \theta)$, is commonly used to test hypothesis and build confidence intervals on θ. However, the distribution of the estimator may not be symmetric for small samples, and so resulting confidence intervals tend to be nonsensical. An alternative is based on the likelihood root statistic. Let $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$ denote the maximum likelihood estimator when ψ is held fixed, and let $l_p(\psi) = l(\hat{\theta}_\psi)$ be the profile log likelihood for ψ. The likelihood root,

$$r(\psi) = \mathrm{sign}(\hat{\psi} - \psi)\left[2\left\{l_p(\hat{\psi}) - l_p(\psi)\right\}\right]^{1/2},$$

also has an asymptotic standard normal distribution, but it accounts for potential asymmetry of the log likelihood. (The χ_1^2 approximation to the likelihood ratio statistic, $w(\psi) = r^2(\psi)$, is more familiar.) Furthermore, apart from a possible sign change, $r(\psi)$ is invariant to interest-respecting reparameterizations. Nonetheless, the approximation to the distribution of both statistics, $t(\psi)$ and $r(\psi)$, is based on the central limit theorem and typically have an error of order $n^{-1/2}$. In practice, for small samples, the normal approximation to $r(\psi)$ might be biased, and the approximation to $t(\psi)$ completely off. Improved likelihood inferences may be based on the modified likelihood root

$$r^*(\psi) = r(\psi) + \frac{1}{q(\psi)}\log\left\{\frac{q(\psi)}{r(\psi)}\right\}, \tag{4}$$

where $q(\psi)$ is defined to be an approximate pivot, such that the modified likelihood root incorporates not only an improved approximation, but also an adjustment for the elimination of nuisance parameters. When the log likelihood is not of the exponential family form, one can use the tangent exponential model (Fraser & Reid, 2001) to obtain

$$q(\psi) = \frac{\left|\phi(\hat{\theta}) - \phi(\hat{\theta}_\psi) \quad \phi_\lambda(\hat{\theta}_\psi)\right|}{\left|\phi_\theta(\hat{\theta})\right|} \frac{\left|j(\hat{\theta})\right|^{1/2}}{\left|j_{\lambda\lambda}(\hat{\theta}_\psi)\right|^{1/2}}, \tag{5}$$

where $\phi(\theta)$ is the canonical parameter, ϕ_θ and ϕ_λ denote the $d \times d$ and $d \times (d-1)$ matrix of partial derivatives, $\partial\phi/\partial\theta^\intercal$ and $\partial\phi/\partial\lambda^\intercal$. For a sample of independent observations, ϕ is a sample space derivative of the log likelihood function, and it is defined as

$$\phi(\theta)^\intercal = \sum_{i=1}^{n} \frac{\partial l(\theta; y)}{\partial y_i}\Big|_{y=y^0} V_i, \tag{6}$$

where y^0 denotes the observed data, and V_i is computed by differentiating y_i with respect to θ, for a fixed pivotal z_i, which is always available through the probability integral transformation $F(y_i;\theta)$, although simpler alternatives may be available. So, we take derivatives of the pivotal z_i on the parameter space and on the sample space,

$$V_i = \frac{\mathrm{d}y_i}{\mathrm{d}\theta^\intercal}\Big|_{\theta=\hat{\theta}} = -\left(\frac{\partial z_i}{\partial y_i}\right)^{-1}\left(\frac{\partial z_i}{\partial \theta^\intercal}\right)\Big|_{\theta=\hat{\theta}}. \tag{7}$$

Confidence intervals using the modified likelihood root, $r^*(\psi)$, are obtained in a similar fashion as using the likelihood root, $r(\psi)$. The modified maximum likelihood estimate, $\hat{\psi}^*$, can be found by solving the equation $r^*(\hat{\psi}^*) = 0$.

3 Partial results and discussion

We simulated 1,000 random samples from the generalized exponential distribution for $\alpha = 0.5, 1, 1.5, 2, 2.5$ and $\sigma = 1$. We reparametrized the model in terms of $\log\alpha$ and $\log\sigma$, hoping to avoid numerical problems with the likelihood maximizations. The simulation and computations were performed using the statistical software R (R Core Team, 2016). For the tangent exponential model, we adapted function `tem` in package **hoa** (Brazzale, 2005). In order to compare the different estimators for α, we plot their bias and root mean square error according to sample size; see Figure 1. Although the least squares estimators have biases close to zero, only the modified maximum likelihood estimator was able of effectively reducing the root mean square error. In this study, we also compare the use of the Wald, lihelihood root, and modified likelihood root statistics to compute confidence intervals for α, σ, and $q_{0.5}$ (the median). For now, we have only simulated 1,000 samples, but our goal is to simulate at least 10,000 samples for all five scenarios, so the results are smoother.

References

Brazzale, A. R. (2005). hoa: An R package bundle for higher order likelihood inference. *Rnews*, 5/1 May 2005, 20–27. ISSN 609-3631.

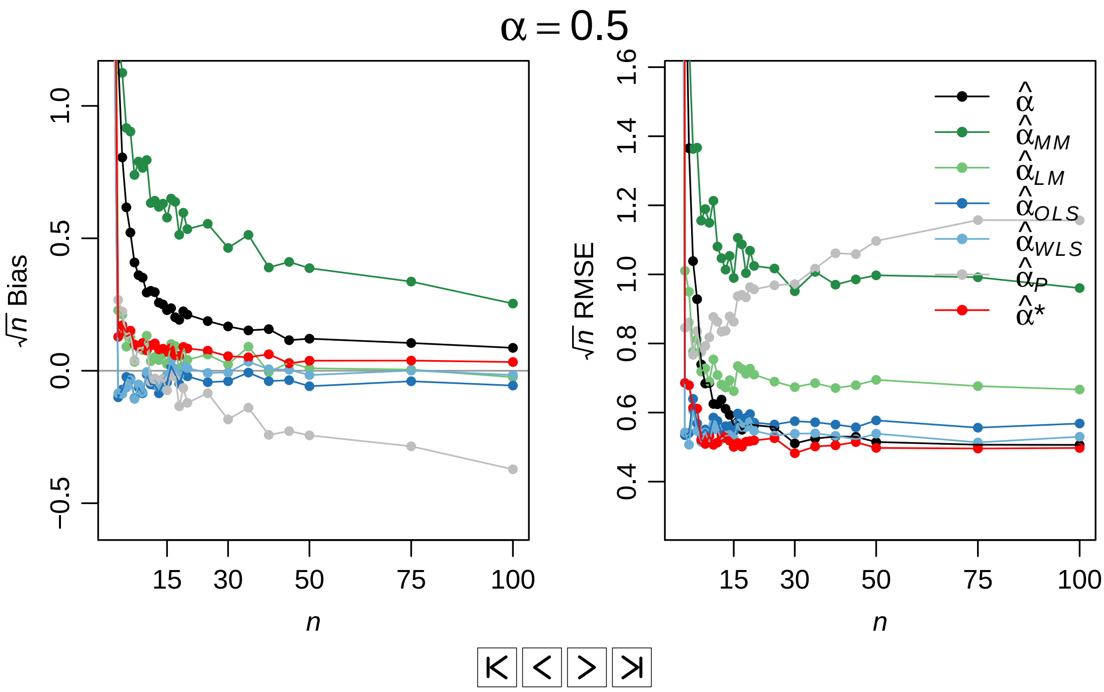

Figura 1: Animated figure (use the buttons or click on the buttom)! Bias (left) and root mean square error (RMSE – right) of the different estimators for α: the usual maximum likelihood estimator, $\hat{\alpha}$; the moments estimator, $\hat{\alpha}_{MM}$; the L-moment estimator, $\hat{\alpha}_{LM}$; the ordinary least squares estimator, $\hat{\alpha}_{OLS}$; the weighted least squares estimator, $\hat{\alpha}_{WLS}$; and percentiles estimator, $\hat{\alpha}_{P}$.

Brazzale, A. R. & Davison, A. C. (2008). Accurate parametric inference for small samples. *Statistical Science*, 23(4), 465–484.

Brazzale, A. R., Davison, A. C., & Reid, N. (2007). *Applied Asymptotics: Case Studies in Small-Sample Statistics*. Cambridge University Press.

Fraser, D. A. S. & Reid, N. (2001). Ancillary information for statistical inference. In S. E. Ahmed & N. Reid (Eds.), *Empirical Bayes and Likelihood Inference* (pp. 185–207). New York: Springer.

Gupta, R. D. & Kundu, D. (2007). Generalized exponential distribution: Existing results and some recent developments. *Journal of Statistical Planning and Inference*, 137(11), 3537–3547.

Lozada-Can, C. & Davison, A. C. (2012). Three examples of accurate likelihood inference. *The American Statistician*, 64, 131–139.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.