Selecting between the generalized inverted Rayleigh and the generalized inverted half logistic distributions.

Bishal Diyali[1]

[1] Central University of Haryana, India – 123029.

Abbreviated abstract: *The generalized inverted Rayleigh distribution (GIRD) and the generalized inverted half logistic distribution (GIHLD) are two of the members from the generalized inverted scale family (GISF) of distributions which was recently introduced in the literature. These distributions can be used quite flexibly for the analysis of positively skewed lifetime datasets. If there is some prerequisite knowledge about the data regarding its positive skewness and the characteristic features indicating the possibility of using this family of distributions, then any of this two can be used for this purpose. Our priority will be to use the one that is most suitable for describing the data and for that reason we have performed this study. The method of maximized log-likelihood functions have been used in this study for the discriminating purpose and their asymptotic properties has been explored as well. Rigorous simulations are conducted for addressing the performance of the discrimination procedure with the help of probability of correct selection (PCS) and the asymptotic PCS are also computed.*

Related publications:

– Dumonceaux et al, Technometrics, 15(1),19– 27 (1973)

– Raqab et al, Communications in Statistics-Simulation and Computation, 47(5):1397–1419 ,(2018)

bishaldiyali09@gmail.com

4th Conference on
Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

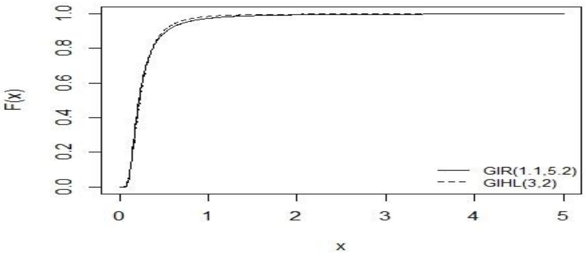
Problem Definition

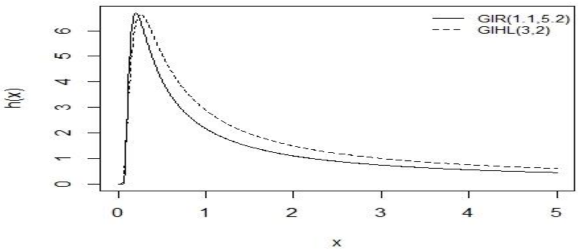
- Both of these distribution have unimodal hazard rates and they shows very similar fit for some values (range) of the parameters. The figures below shows the plots for cumulative distribution functions (CDFs) and the hazard functions (HFs) for GIRD(1.1,5.2) and GIHLD(3,2) which are showing very similar fits, so, any one can be used in place of the other one.

- The 90th, 95th and 99th percentile points of GIRD are 0.5302, 0.738 and 1.558 and that of GIHLD are 0.4974, 0.6467 and 1.1422. Clearly these values are very different. Many times, the researchers are interested in the extreme values rather than the central values for making inferential decisions . To make the correct decisions about the population characteristics, it is very important to make the right choice of the distribution as the wrong model selection can be misleading and can have severe effect on the decision making.

- So, if we know through preliminary analysis that the data is positively skewed with unimodal hazard rate, then any of these two distributions can be used for modeling the data but we would like to choose the one that best describes the data. And we address this problem in this work.

- One can refer to Kundu and Manglick (2004), Quesenberry and Kent (1982),Kundu et al. (2005),Dey and Kundu (2009), Abid and Kokonendji (2021), for details on such works on discrimination and model selection.





4th Conference on
Statistics and
Data Science
Salvador, Brazil (online)
December 1-3, 2022

Methods

- The PDF and CDF for X following GIRD with shape parameter α and scale parameter λ is given by

$$f_{GIRD}(x;\alpha,\lambda)=\frac{2\alpha}{\lambda^2x^3}e^{-\frac{1}{(\lambda x)^2}}\left(1-e^{-\frac{1}{(\lambda x)^2}}\right)^{\alpha-1}\quad and\quad F_{GIRD}(x;\alpha,\lambda)=1-\left(1-e^{-\frac{1}{(\lambda x)^2}}\right)^{\alpha};x>0,\alpha,\lambda>0$$

- The PDF and CDF for X following a GIHLD with shape parameter β and scale parameter ξ is given by

$$f_{GIHLD}(x;\beta,\xi)=\frac{2\beta}{\xi x^2}e^{-\frac{1}{\xi x}}\frac{\left(1-e^{-\frac{1}{\xi x}}\right)^{\beta-1}}{\left(1+e^{-\frac{1}{\xi x}}\right)^{\beta+1}}\quad and\quad F_{GIHLD}(x;\beta,\xi)=1-\left[\frac{1-e^{-\frac{1}{\xi x}}}{1+e^{-\frac{1}{\xi x}}}\right]^{\beta};x>0,\beta,\xi>0$$

- To address this problem, we used the method of the difference of the maximized log-likelihoods (DMLL) which states that the chosen distribution will be the one which will have the higher value of the log-likelihood function calculated at the maximum likelihood estimates (MLEs) of the parameters.

- We also explore the asymptotic distribution of the test statistics to explore the theoretical results.

- The extensive simulation studies (10000 times) have been performed by generating random samples of different sizes and for different sets of parameter values from both the distributions to see how many times the procedure correctly selects the parent distribution i.e. probability of correct selection (PCS).

Note:

- The log-likelihood functions are calculated using the expressions for the PDF.
- The MLEs are calculated by solving the likelihood equations formed from the likelihood function by differentiating with respect to the parameters and equating to zero.

4th Conference on
Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Results and Conclusions

PCS when parent distribution is GIRD (1,1) using Monte Carlo (MC) simulations and the asymptotic results (AR)							
	20	40	60	80	100	200	500
MC	0.7242	0.7585	0.7935	0.8209	0.8446	0.9151	0.9816
AR	0.6719	0.7355	0.7800	0.8133	0.8403	0.9204	0.9870

PCS when parent distribution is GIHLD(1,1) using Monte Carlo (MC) simulations and the asymptotic results (AR)							
	20	40	60	80	100	200	500
MC	0.4442	0.5446	0.6133	0.6594	0.694	0.8185	0.9389
AR	0.6129	0.6576	0.6904	0.717	0.7395	0.8179	0.9244

- The above table is the part of the simulation study with the scale and shape parameter taken to be 1 in both the cases. The first row shows the Monte Carlo simulation results and the second row show the asymptotic results.
- The DMLL method seems to work very well for the GIRD case, the PCS are good even for small sample sizes and are close to 1 for sample sizes of 500. For the GIHLD case the PCS are low for the small sample sizes but improves as the sample size increases.
- The asymptotic results seems to work well even for the small sample sizes and matches very well for the large sample sizes.
- Based on our observed results, we conclude that the DMLL and the asymptotic results both works well can be applied for the model selection problems.

FUTURE WORKS

- This work can be done for different distributions and can be extended for the cases of censoring.
- The minimum sample sizes required for user defined PCS can be explored as well.

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022