The Harris Extended Rayleigh Distribution: Applications to Lifetime Data

Keerthana Udupa, Himanshu Pokhriyal

Department of Data Science, Manipal Academy of Higher Education, Manipal

Abstract

We propose a new three-parameter lifetime model called the Harris Extended Rayleigh (HER) distribution obtained from a mixture of the Rayleigh and Harris distributions given by Aly and Benkherouf (2011). Different estimation procedures, such as Maximum Likelihood (ML) Estimation, Least Squares (LS) Estimation, Maximum Product of Spacings (MPS) Estimation etc. have been used to estimate the unknown parameters and their performances are compared using simulations. The use of the model in lifetime modelling is established by fitting it to two real data sets, remission times of bladder cancer patients and lifetime of squamous cell carcinoma patients.

The Harris Extended Rayleigh Distribution

Let $f_0(x)$ be the baseline probability density function of a lifetime random variable X. The probability density function (pdf) of HE family of distribution is defined by

$$f(x) = \frac{\alpha^{\frac{1}{k}} f_0(x)}{[1-\bar{\alpha}\bar{F}_0(x)^k]^{\left(\frac{k+1}{k}\right)}} \tag{1}$$

where $x > 0,\ \alpha > 0,\ k > 0, \bar{\alpha} = 1-\alpha$ and $\bar{F}_0(x)$ is the survival function of the baseline distribution. The statistical properties and some general results about the HE family of distributions can be seen in Aly and Benkherouf (2011) and Batsidis and Lemonte (2015).

We generalize the Rayleigh distribution given by Rayleigh (1880), by using it as a baseline distribution in the HE family pdf given in (1). A continuous random variable X is said to follow HER distribution if its pdf is given by,

$$f_{HER}(x;\Theta) = \frac{\alpha^{\frac{1}{k}} \frac{x}{\theta^2} \exp\left(\frac{-x^2}{2\theta^2}\right)}{\left(1-\bar{\alpha}\exp\left(\frac{-kx^2}{2\theta^2}\right)\right)^{\left(\frac{k+1}{k}\right)}} \tag{2}$$

where $x \geq 0,\ \theta > 0,\ \alpha > 0,\ k > 0$ and $\bar{\alpha} = 1-\alpha$ and $\Theta = \{\theta, \alpha, k\}$.

The new parameters $\alpha > 0$ and $k > 0$ are additional shape parameters that aims to introduce greater flexibility. They are sought as a manner to furnish a more flexible distribution.

Shape of the Density

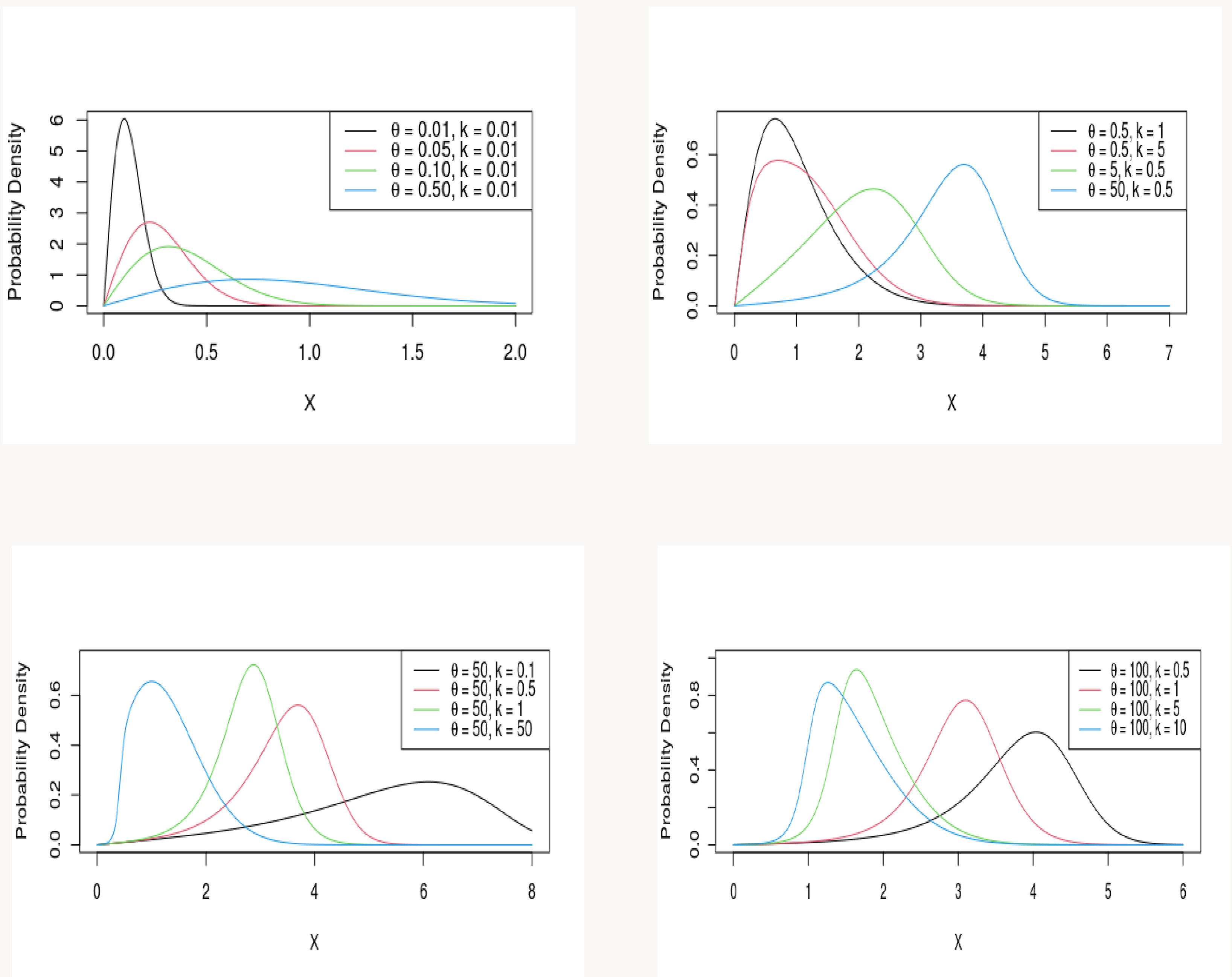


Figure 1: Shape of the HER Density for varying values of θ, k and $\alpha = 1$

Maximum Likelihood Estimation

The Log-Likelihood Function for a random sample of size n from HER distribution is given by,

$$L = \frac{n}{k} log\alpha - \frac{\sum_{i=1}^{n} x_i^2}{2\theta^2} + \frac{\sum_{i=1}^{n} logx_i}{\theta^2} - \frac{k+1}{k}\sum_{i=1}^{n} log\left(1-\bar{\alpha}\exp\left(\frac{-kx_i^2}{2\theta^2}\right)\right) \tag{3}$$

The value of Θ which maximizes the log likelihood function is the ML estimate, denoted by $\hat{\Theta}$. Using this log-likelihood function, we can also obtain the Asymptotic Confidence Intervals for Θ.

Other estimation techniques like LS Estimation, MPS Estimation etc. can also be used to estimate the Θ.

Real Data Analyses: The Bladder Cancer Data and The Lagakos Data

The Bladder Cancer data set given in Jose and Sivadas (2015), represents the uncensored remission times (in months) of the Bladder Cancer patients. The Lagakos data set given in Lagakos (1978), arose from a clinical trial conducted to study the survival time (in months) of squamous cell carcinoma patients.

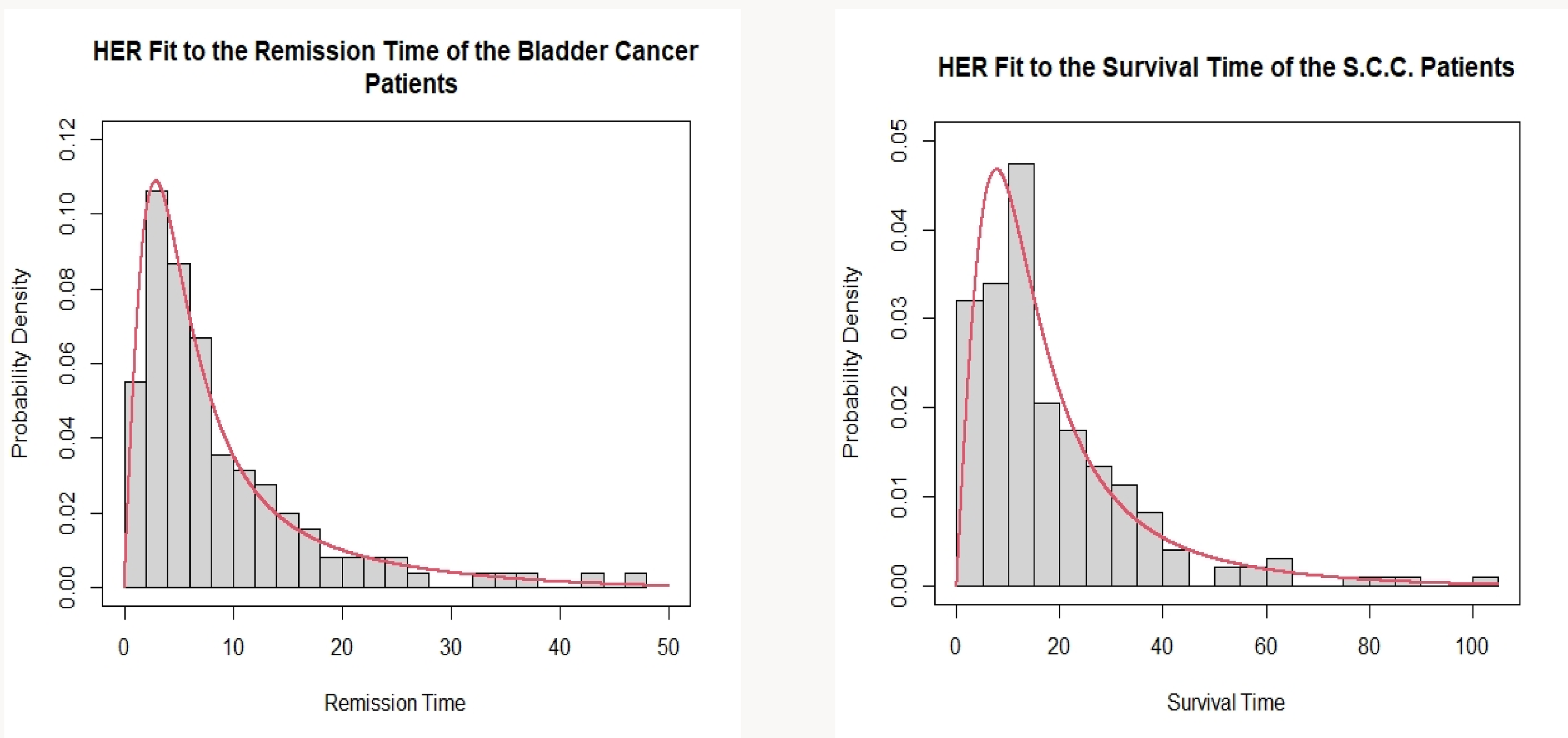
	N	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	S.D.
Bladder Cancer Data	127	0.08	3.34	6.25	8.82	11.72	46.12	8.51
Lagakos Data	194	1	8	14	18.81	24.75	101	16.54

Table 1: Summary Statistics of the Two Real Datasets

The estimates of the parameters of the HER distribution were obtained using ML Estimation procedure. Akaike Information Criterion, AIC (see Akaike (1998)) was calculated to compare the HER model with some other lifetime models. Goodness of fit was assessed statistically by KS test and graphically visualized by constructing the histogram.

Model	Bladder Cancer Data			Lagakos Data		
	Estimate(s)	L	AIC	Estimate(s)	L	AIC
Harris Extended Rayleigh	(17.53, 0.05, 1.61)	-402.06	810.12	(37.04, 0.07, 1.21)	-753.61	1513.23
Log Normal	(1.73, 1.05)	-406.69	817.38	(2.57, 0.92)	-758.04	1520.08
Generalised Rayleigh	(17.13, 0.41)	-410.13	824.26	(32.91, 0.48)	-767.45	1538.89
Gumbel	(5.47, 5.02)	-416.50	836.99	(12.22, 10.16)	-769.80	1543.60
Inverse Gaussian	(8.82, 3.42)	-432.17	868.34	(18.82, 13.62)	-770.42	1544.83
Logistic	(7.39, 4.08)	-438.49	880.97	(16.29, 8.06)	-799.84	1603.69
Transmuted Rayleigh	(9.91, 0.71)	-439.32	882.64	(20.65, 0.72)	-790.04	1584.09
Rayleigh	(8.65)	-454.93	911.87	(17.69)	-809.79	1621.58

Table 2: Parameter Estimation and Goodness of Fit Comparison



(a) Blood Cancer Data **(b)** Lagakos Data

Figure 2: HER Fit to the two Real Datasets

Concluding Remarks

In the lifetime data analysis it is important to have more flexible models as the real data sets may have high degree of skewness and kurtosis.

The HER distribution is a three-parameter lifetime model which is very flexible with respect to its shape of the pdf.

The HER model had minimum value of AIC in both the real data analyses and hence can be considered as the best and more flexible model compared with the remaining models.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer.

Aly, E.-E. A. and Benkherouf, L. (2011). A new family of distributions based on probability generating functions. *Sankhya B*, 73(1):70–80.

Batsidis, A. and Lemonte, A. J. (2015). On the harris extended family of distributions. *Statistics*, 49(6):1400–1421.

Jose, K. and Sivadas, R. (2015). Negative binomial marshall–olkin rayleigh distribution and its applications. *Stochastics and Quality Control*, 30(2):89–98.

Lagakos, S. W. (1978). A covariate model for partially censored data subject to competing causes of failure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 27(3):235–241.

Rayleigh, F. L. (1880). Xii. on the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(60):73–78.

Presenting Author: Ms Keerthana Udupa

keerthana.udupa@learner.manipal.edu

M.Sc. Biostatistics, Department of Data Science, Manipal Academy of Higher Education, Manipal