The International Conference on Trends and Perspectives in Linear Statistical Inference

A COMPARISON OF COMPOUND POISSON CLASS DISTRIBUTIONS

Dr. Deniz INAN
Oykum Esra ASKIN (PHD.CANDIDATE)
Dr. Ali Hakan BUYUKLU

«SURVIVAL ANALYSIS» »FAILURE TIME ANALYSIS» «EVENT TIME ANALYSIS»

- In many applications the primary endpoint of interest is survival time.

  - Medicine, Biology, Public health, Epidemiology, Engineering…

- We may be interested in characterizing the distribution of survival time (such as death, going out remission…etc) for a given population;

  - Comparing survival times among different groups
  - Modelling relationship between survival time and observable covariates

PARAMETRIC SURVIVAL

- In parametric survival model is one in which survival time (the outcome) follow a known distribution;
  - Weibull
  - Exponential
  - Log-logistic
  - Lognormal
  - Generalized gamma
  - …
- homogeneous population


If we study with heterogeneous population, can we rely on results when we assume the survival time follow pure classical distributions mentioned before?

In recent years, new classes of distributions have been proposed to deal with hardness of modelling heterogeneous data.

Some Examples

- Decreasing Failure Rate
  - exponential-geometric (Adamidis and Loukas, 1998)
  - exponential-Poisson (E-P)(Kus, 2007)
  - exponential-logarithmic (Tahmasbi and Rezaei, 2008)

- Failure Rate with decreasing, increasing and monotone decreasing
  - extended exponential-geometric (Adamidis et al. 2005)
  - weibull-geometric (Barreto-Souza et al., 2010)
  - weibull-logarithmic (Ciumara and Preda, 2009)
  - weibull-Poisson (W-P)(Lu and Shi, 2012)

SIMILAR MIXING PROCEDURE INTRODUCED BY ADAMIDIS AND LOUKAS

OUTLINE OF PRESENTATION

- Compound Poisson Class of Distributions
  - Exponential-zero truncated Poisson (E-P)
  - Weibull-zero truncated Poisson (W-P)
  - Rayleigh-zero truncated Poisson (RAY-P)
- Methodology
  - EM Algorithm
- Application
- Results
- Discussion

COMPOUND POISSON CLASS

- Think about a situation where failure (of a device for example) occurs due to the precence of an unkown number , Z, of same kind initial defects. Let us define Z as a zero truncated Poisson distributed.

- Then let W's represent the failure times of a unit caused by initial defects and each defect can be detected only after causing failure, in which case it is repaired perfectly (Adamidis and Loukas, 1998).

- According to W's distributional assumptions (W1, W2,…,Wz), we can model time to first failure X =Min(W1, W2,…,Wz).

- In this study, we will take W's as exponential, weibull and rayleigh distributed randoms

E-P (KUS,2007)

- Let $W_1, W_2, \ldots W_Z$ be iid random variables with the following pdf;

$$f(w,\beta) = \beta e^{-\beta w} \qquad (1)$$

- Also, Z is a zero truncated Poisson variable with following pdf;

$$f(z,\lambda) = \frac{e^{-\lambda}\lambda^{z}}{\Gamma(z+1)\left(1-e^{-\lambda}\right)} \qquad (2)$$

- Let us define X=min($W_1, W_2, \ldots W_Z$). Then, the marginal pdf of X; $\theta = (\lambda, \beta)$

$$f(x,\theta) = \frac{\lambda\beta}{\left(1-e^{-\lambda}\right)} \exp\left(-\lambda - \beta x + \lambda \exp(-\beta x)\right) \qquad (3)$$

W-P (Lu and Shi, 2012)

$X=\min(W_1, W_2, \ldots W_Z) \quad \theta=(\lambda,\beta,\alpha)$

$$f(x,\theta)=\frac{\lambda\beta\alpha x^{\alpha-1}}{\left(1-e^{-\lambda}\right)}\exp\left(-\lambda-\beta x^{\alpha}+\lambda\exp\left(-\beta x^{\alpha}\right)\right) \quad (4)$$

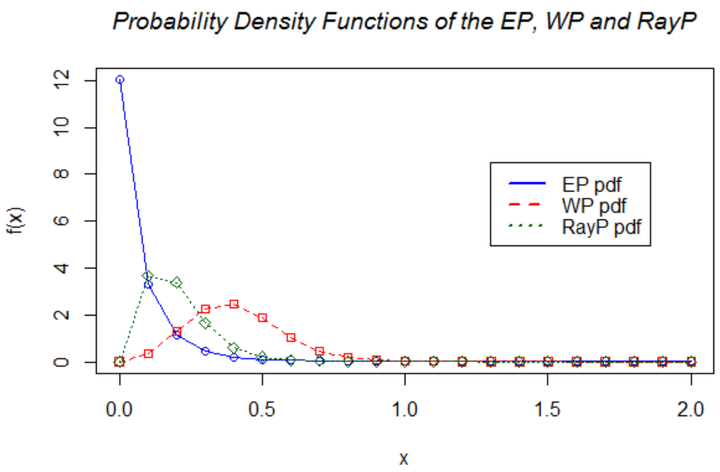
RAY-P (Hemmatı et al., 2011)

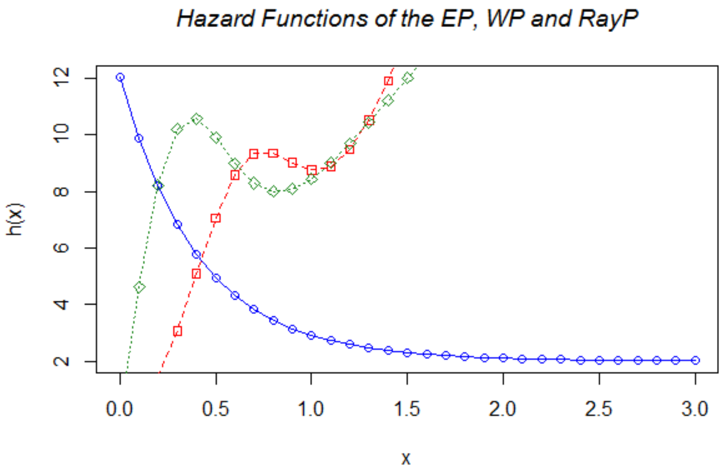
$X=\min(W_1, W_2, \ldots W_Z)$

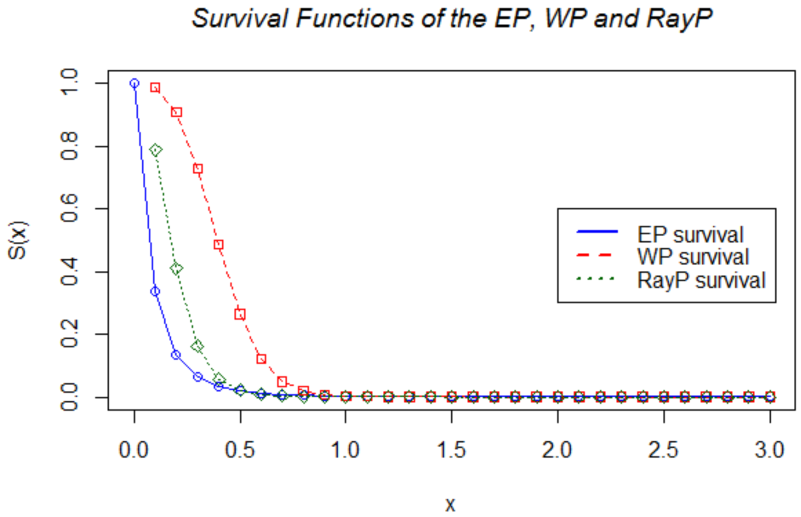
$$f(x,\theta)=\frac{2\lambda x\beta^{2}}{\left(1-e^{-\lambda}\right)}\exp\left(-\lambda-(\beta x)^{2}+\lambda\exp\left(-(\beta x)^{2}\right)\right) \quad (5)$$

for EP and RayP $\lambda = 6; \beta = 2$

for WP $\lambda = 6; \beta = 2; \alpha = 3$







To summarize….

EP

exponential + zero truncated Poisson

WP

weibull +zero truncated Poisson distributions,

RayP

Rayleigh + zero truncated Poisson

with the same mixing procedure

Methodology

- To find MLE's of distribution parameters, Newton Raphson algorithm is one of the standard methods which is widely used. To employ the algorithm, second derivates of the log-likelihood are required.

- However EM algorithm is useful when maximizing observed log likelihood can be difficult then maximizing the complete data log likelihood.

- Recently, EM algorithm has been used by several authors such to find the ML estimations of compound distributions' parameters.

- We will show the steps of EM algorithm for only WP distribution because of the limited time…

- To find hypothetical complete data distribution, it is well known that the conditional density function can be defined as in equation (6). (Alkarni, and Oraby, 2012). Here, τ is the parameter vector of the weibull distribution.

$$f(x \backslash z;\tau) = z\, f(x;\tau)\left[1 - F(x;\tau)\right]^{z-1} \qquad (6)$$

$$= z\alpha\beta x^{\alpha-1} \exp\left(-\beta z x^{\alpha}\right)$$

- Using (6), the hypothetical complete data distribution is given by (7). Here, θ is the parameter vector of weibull and zero truncated Poisson distributions;

$$f(x,z;\theta) = f(x \backslash z;\tau)\, p(z;\lambda) = \frac{\alpha\beta z x^{\alpha-1} \exp\left(-\beta z x^{\alpha}\right)\lambda^{z}}{\Gamma(z+1)\left(\exp(\lambda)-1\right)} \qquad (7)$$

$$x > 0, \quad z = 1, 2, \ldots, \quad \lambda, \beta > 0$$

- E-step of EM cycle requires the computation of the conditional expectation of Z, which is given below;

$$E\left(Z \setminus X; \theta^{(k)}\right)$$

- Here, $\theta^{(k)} = \left(\lambda^{(k)}, \beta^{(k)}, \alpha^{(k)}\right)$ is the current estimate of θ. Conditional probabilty of Z can be given as in equation (8).

$$P\left(z \setminus x; \theta\right) = \frac{f\left(x, z; \theta\right)}{f\left(x; \theta\right)} = \frac{\lambda^{z-1} \exp\left(-\beta z x^{\alpha} + \beta x^{\alpha} - \lambda \exp\left(-\beta x^{\alpha}\right)\right)}{\Gamma\left(z\right)} \quad (8)$$

- Using equation (8), we can find the conditional expectation of Z for WP distribution as in equation (9).

$$E\left(z \setminus x; \theta\right) = \sum_{z=1}^{\infty} z\, P\left(z \setminus x; \theta\right) = 1 + \lambda e^{-\beta x^{\alpha}} \quad (9)$$

- The EM cycle is completed with M-step. In this step, missing Z's in complete data likelihood (given in equation (10)) are replaced by their conditional expectations .

$$\text{C. D. Likelihood}: \prod_{i=1}^{n} \frac{\lambda^{z-1} \exp\left(-\beta z x^{\alpha} + \beta x^{\alpha} - \lambda \exp\left(-\beta x^{\alpha}\right)\right)}{\Gamma(z)} \qquad (10)$$

- Thus, an EM iteration, taking $\theta^{(k)}$ into $\theta^{(k+1)}$ is given by;

$$\alpha^{(k+1)} = n \Big/ \left(\sum_{i=1}^{n} \beta^{k+1} \log(x_i) x_i^{\alpha^{(k+1)}} w_i^k - \log(x_i) \right)$$

$$\beta^{(k+1)} = n \Big/ \left(\sum_{i=1}^{n} w_i^{(k)} x_i^{\alpha^{(k+1)}} \right)$$

$$\lambda^{(k+1)} = n \Big/ \left[\left(1 - e^{-\lambda^{(k+1)}}\right) \sum_{i=1}^{n} w_i^{(k)} \right]$$

$$w_i^k = 1 + \lambda^{(k)} e^{-\beta^{(k)} x_i^{(k)}}$$

THE FIRST DATA SET

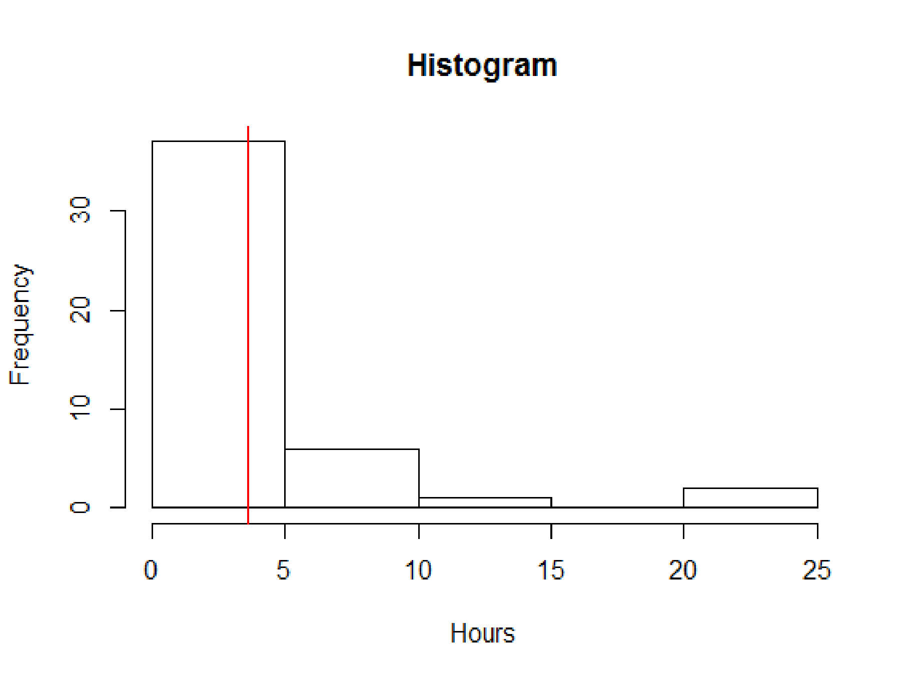
airborne communication transceiver

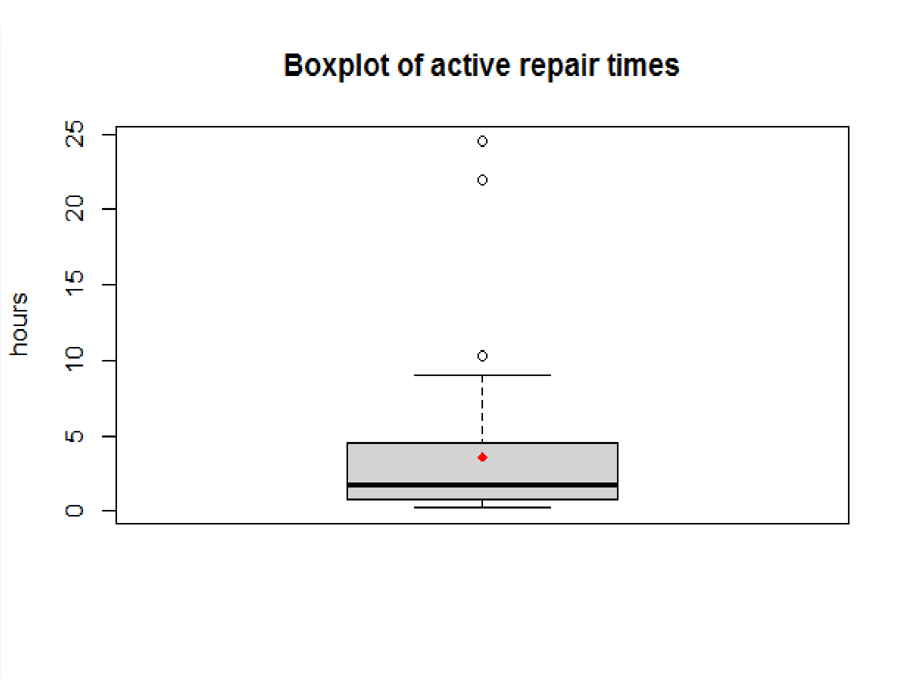
BWP (Burbank Water and Power) model

- The data concerns 46 observations reported on active repair times (hours) for an airborne communication transceiver.
- Data set is used as a lifetime distribution by many authors.

Data characteristics

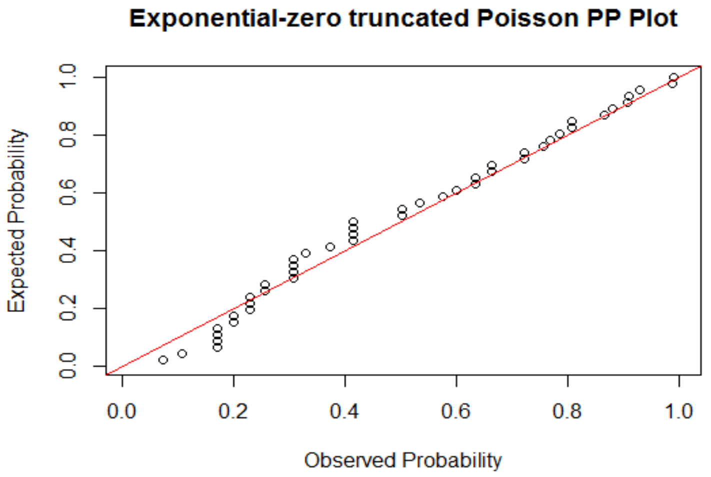
Minimum	Maximum	Mean	Median	1st quartile	3rd quartile	Skewness	Kurtosis
0.2	24.5	3.607	1.75	0.800	4.375	2.794666	8.294985

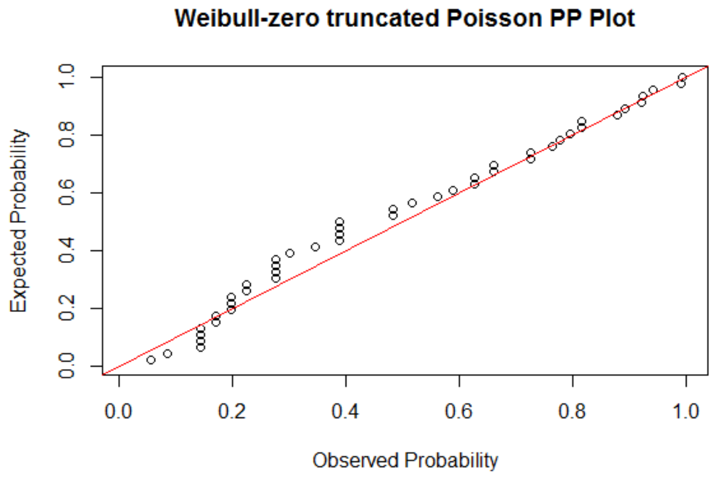


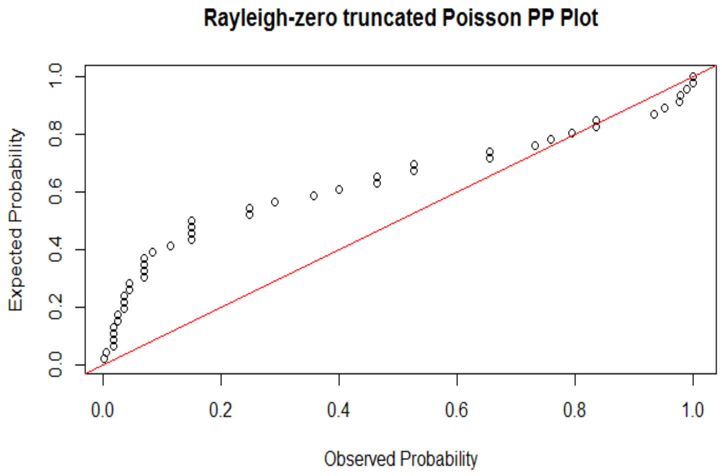


The first data set

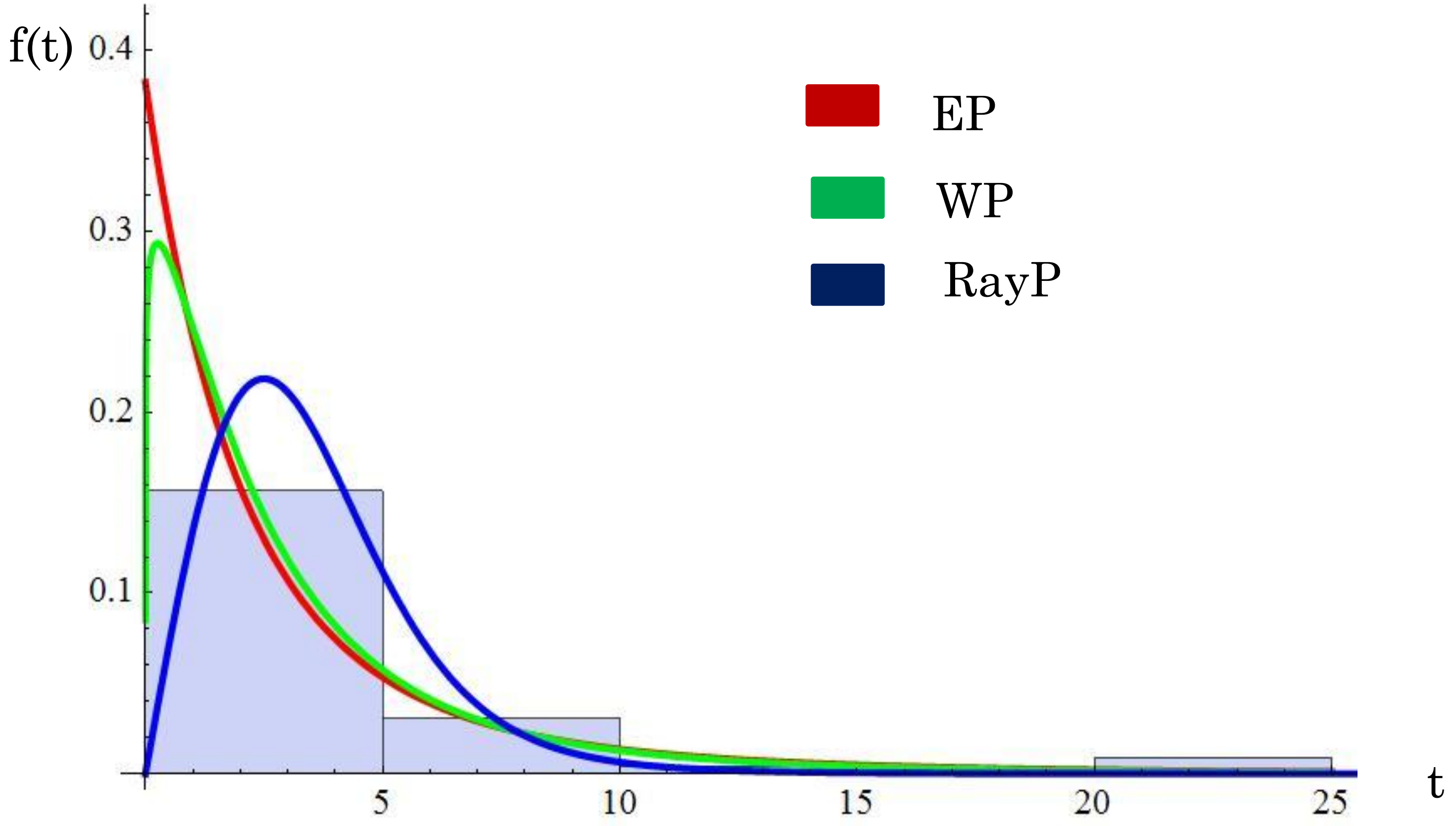
Distribution	Parameters	KS Test	p-value
EP	$\theta:(\lambda=3.41;\beta=0.108)$	0.1051	0.6891
WP	$\theta:(\lambda=3.52;\beta=0.09;\alpha=1.10)$	0.1111	0.6210
RP	$\theta:(\lambda=5.92;\beta=0.11)$	0.3498	2.5×10^{-5}







GRAPHS OF PROBABILITY DENSITY FUNCTIONS



Characteristics of EP distribution

E(t)	1st quartile	3rd quartile	Skewness	Kurtosis
3.558	0.780	4.388	2.893	9.297

Characteristics of WP distribution

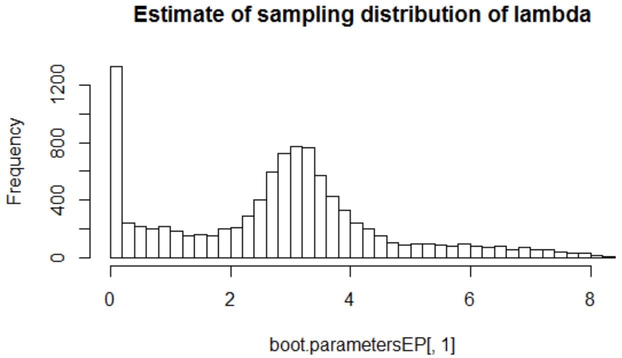
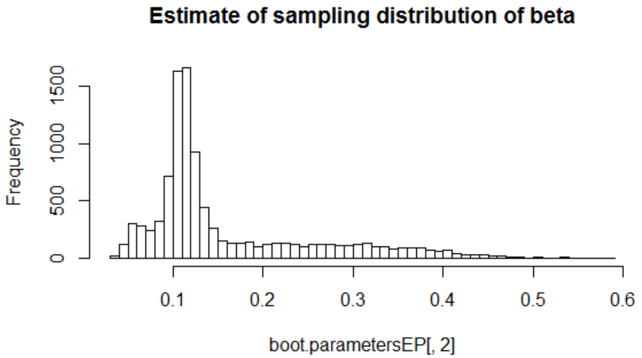
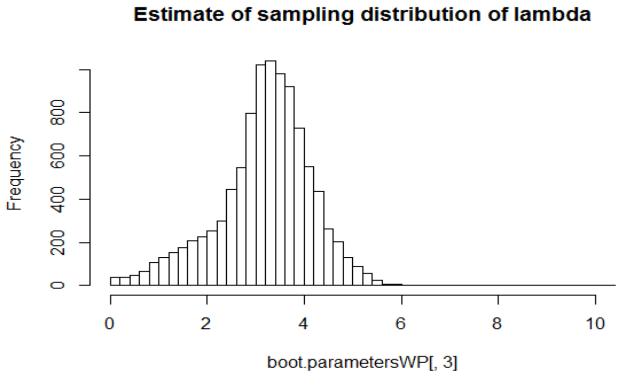
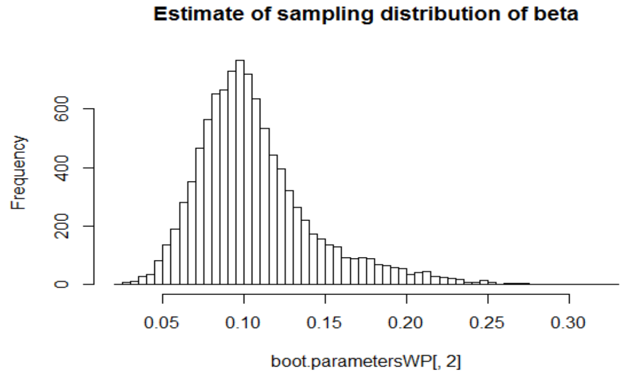
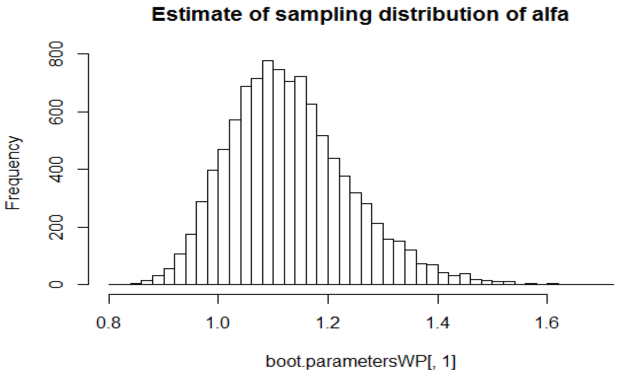
E(t)	1st quartile	3rd quartile	Skewness	Kurtosis
3.384	0.898	4.304	3.320	17.342

Characteristics of data1

Mean	1st quartile	3rd quartile	Skewness	Kurtosis
3.607	0.800	4.375	2.794666	8.294985

Bootstrap Confidence Intervals

Parameters	Mean	Std.Err.	Bootstrap CI (95%)
EP Distribution			
$\theta = \lambda : \beta$	2.7569 0.1589	1.8192 0.09487	(0.0024, 6.949) (0.054, 0.405)
WP Distribution			
$\theta = \lambda : \beta : \alpha$	3.1947 0.1069 1.1303	0.9761 0.0367 0.1124	(0.9114, 4.9331) (0.0532, 0.2033) (0.9444, 1.3835)

Distribution	boot. lambda	boot.beta	boot.alpha
EP			
WP			

DiSCUSSiON

EP or WP ?

Ref.

- [1] Adamidis, K. Loukas, S., (1998). A life time distribution with decreasing failure rate. Statist. Probab. Lett. 39, 35-42.
- [2] Dempster, A.P., Laird, N.M. Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. Roy. Statist.Soc.Ser. B 39, pp. 1-38.
- [3] Kus C.(2007). A new lifetime distribution, Comput. Statist. Data Anal. 51,pp. 4497-4509.
- [4] Erisoglu U., Erisoglu M. Calis N. (2013). Heterogeneous data modeling with two-component Weibull-Poisson distribution, Journal of Applied Statistics, 40:11, pp. 2451□ 2461.
- [5] Lu W., Shi D., (2012). A new compounding life distribution: The Weibull-Poisson distribution, J. Appl. Stat. 39(1) , pp. 21-38.

- [6] Barreto-Souza, W. Cribari-Neto, F., (2009). A generalization of the exponential-Poisson distribution. Statistics and Probability Letters 79, pp.2493-2500.
- [7] Alkarni, S., Oraby, A., (2012). A compound class of Poisson and lifetime distribution. J. Stat.Appl.Pro 1, pp.45-51.
- [8] Hemmati, F., Khorram, E., Rezakhah, S., (2011). A new three-parameter ageing distribution 141, pp. 2266-2275.
- [9] Duchateau, L., Janssen, P., (2008). The frailty model. (New York : Springer).
- [10] Hanagal, D.D., (2011). Modeling survival data using frailty models.(Chapman and Hall/CRC)
- [11] Wienke, A., (2011). Frailty models in survival analysis. (Boca Raton: Chapman & Hall/CRC)
- [12] Kleinbaum, D.G., Klein, M., (2005). Survival Analysis: A self Learning Text, second edition. (New York : Springer).
- [13] R DEVELOPMENT CORE TEAM (2011).R: A Language and Environment for Statistical Computing. Vienna, Austria:R Foundation for Statistical Computing

THANK YOU