Chapter 4

Inliers estimation in Weibull models

4.1 Introduction

One of the most important and widely used distributions to study lifetime of any component is Weibull distribution. The Weibull distribution is appropriate to describe the variation in the lifetimes of many different types of components. It has been used as model of lifetimes with diverse types of items such as Vacuum tubes, ball bearings and electrical insulation. In survival analysis, the Weibull distribution is better suited than the Gaussian distribution, because, it is defined only for positive time (engines fail after assembly), the mathematical operations particular to reliability theory are simpler (e.g. the function is easy to integrate analytically) and the shape of the function is more flexible, it ranges from a close Gaussian resemblance, to a skewed Gaussian, to a pure exponential distribution. It is also widely used in biomedical applications for e.g. in studies on time to occurrence of tumors in human population or in laboratory animals etc. It includes exponential distribution as special case. Also exponential distribution has been widely used as model in areas ranging from studies on the lifetimes of manufactured items to research involving survival or remission times in chronic diseases. In all the above examples we can get inliers and target observations as discussed in chapter 1.

As discussed in the previous chapters, here also our objective is to study the inliers and their detection procedure in Weibull distributions. This chapter deals with competition between two Weibull mixture models representing inliers and the target distribution.

If we denote $\underline{X}=(X_1,X_2,...X_n)$ as realizations of a life test, then $\underline{X}=(X_I \cup X_T)$ where X_I is set of inlier observations (instantaneous and early failures) and X_T is set of observations coming from a target population. Since failure pattern of this situations usually discard the assumption of unimodal distribution, the usual method of modeling and inference procedures may not be accurate in practice. The prior objective in such situations is to decide how many inliers are present in the underlying model, and then study their inferences.

This article is organized as follows: In section (4.2) and (4.3), discussion of the UMVUE and identified inliers model, assuming both inlier distribution and target distribution as Weibull distribution is considered. The inference procedures when some of the parameters are known and unknown are considered. Section (4.3.3) deals with the inlier detection for labeled slippage model. The detection using information criterion, goodness of fit and data analysis are given in the subsequent sections.

4.2 Uniformly minimum variance unbiased estimator (UMVUE)

The UMVUE of mixture density of instantaneous and positive observation taken from Weibull distribution is obtained in this section. Based on above families a new family of df $\Im=\{F(x;\theta,p):x\geq 0,\theta\in\Omega,0<p<1\}$ is defined, such that

$$f(x;\theta,p)=\begin{cases}1-p+pf(x;\theta), & x=0\\ pf(x;\theta), & x>0\end{cases}$$

Hence the pdf of mixture family of instantaneous and Weibull distribution is obtained as

$$f(x;\theta,p)=(1-p)^{\delta}\,p^{(1-\delta)}\left[\beta\theta x^{\beta-1}\,e^{-\theta x^{\beta}}\right]^{(1-\delta)}$$

$$=\frac{\left(\beta x^{\beta-1}\right)^{(1-\delta)}\left[\exp(-\theta)\right]^{(1-\delta)x^{\beta}}\left[\frac{(1-p)}{p\theta}\right]^{\delta}}{\left[\frac{1}{p\theta}\right]} \qquad (4.2.1)$$

which is a member of exponential family with $a(x)=\beta x^{\beta-1}$, $h(\theta)=\exp(-\theta)$, $g(\theta)=\frac{1}{\theta}$ and $d(x)=x^{\beta}$. We have $z=\sum_{x>0}(1-\delta)x^{\beta}$ and $n-r=\sum_{x>0}\delta_j$ which are jointly complete sufficient statistics for (θ, p). Since x^{β} has exponential distribution with parameter θ. The UMVUE of mixture density given by Singh (2007) is defined as

$$\varphi_x(z,r,n)=\begin{cases}\frac{B(z,r,n-1)}{B(z,r,n)}=\frac{n-r}{n}, & x=0,\, r=0,1,2....n-1\\ a(x)\frac{B(z-d(x),r-1,n-1)}{B(z,r,n)}, & x>0,\, z>d(x),\, r=1,2....n\end{cases}$$

which simplifies in Weibull as

$$\varphi_x(z,r,n)=\begin{cases}\frac{n-r}{n}, & x=0\\ \frac{r(r-1)\beta x^{\beta-1}}{nz}\left[1-\frac{x^{\beta}}{z}\right]^{r-2}, & 0<x^{\beta}<z,\, n>1\end{cases} \qquad (4.2.2)$$

If $\beta=1$ in equation (4.2.2) one gets UMVUE of mixture density of instantaneous and positive observation from exponential distribution as

$$\varphi_x(z,r,n)=\begin{cases}\frac{n-r}{n}, & x=0\\ \frac{r(r-1)}{nz}\left[1-\frac{x}{z}\right]^{r-2}, & 0<x<z,\, n>1\end{cases} \qquad (4.2.3)$$

4.3 Weibull identified inliers model

Weibull distribution is used medical studies dealing with fatal diseases, where one is interested in the survival time of individual with the disease, measured either from the date of diagnosis or some other starting point. It is possible that patient dies without getting treatment or has smaller survival rate than target group which has on average longer survival rate. The inlier detection is done for the following two models:

Model-1: Shape parameter β is same for both inliers and target distribution.

Model-2: Scale parameter θ is same for both inliers and target distribution.

4.3.1 Inlier detection when the shape parameter is identical

If we take the distribution function of inliers as

$$G(x)=1-\exp\left(-\phi x^{\beta}\right), \qquad x>0,\ \phi>0,\ \beta>0. \tag{4.3.1}$$

and the distribution of target population is

$$F(x)=1-\exp\left(-\theta x^{\beta}\right), \qquad x>0, \theta>0, \beta>0. \tag{4.3.2}$$

Then the likelihood of model can be written as

$$L(x\,|\,\phi,\theta,\beta)=\prod_{i=1}^{r}\beta\phi x^{\beta-1}e^{-\phi x^{\beta}}\prod_{i=r+1}^{n}\beta\theta x^{\beta-1}e^{-\theta x^{\beta}}$$

The estimates of the parameters are found by solving the following likelihood equations:

$$\frac{\partial \ln L}{\partial \phi}=0 \Rightarrow \frac{r}{\phi}-\sum_{i=1}^{r}x_i^{\beta}=0$$

$$\frac{\partial \ln L}{\partial \theta}=0 \Rightarrow \frac{n-r}{\theta}-\sum_{i=r+1}^{n}x_i^{\beta}=0$$

and

$$\frac{\partial \ln L}{\partial \beta} = 0 \Rightarrow \phi \sum_{i=1}^{r} x_i^{\beta} \ln x_i + \theta \sum_{i=r+1}^{n} x_i^{\beta} \ln x_i - \frac{n}{\beta} - \sum_{i=1}^{n} \ln x_i = 0$$

As discussed earlier, the instantaneous failures are already identified and hence the proportion of such observations is not considered in the model. Using Newton Raphson method the estimates of ϕ, θ and β can be found.

4.3.2 Inlier detection when the scale parameters are identical

Here we consider the detection of inliers when shape parameters of both inlier and target distribution as same. The failure distribution for inliers is assumed to be

$$G(x) = 1 - \exp\left(-x^{\beta}\theta\right), \qquad x > 0,\ \theta > 0,\ \beta > 0. \tag{4.3.3}$$

and the distribution of target population is

$$F(x) = 1 - \exp(-x\theta), \qquad x > 0\ ,\theta > 0. \tag{4.3.4}$$

The likelihood estimates in this case are the solutions of

$$\frac{\partial \ln L}{\partial \theta} = 0 \Rightarrow \frac{n}{\theta} - \sum_{i=1}^{r} x_i^{\beta} + \sum_{i=r+1}^{n} x_i = 0$$

and

$$\frac{\partial \ln L}{\partial \beta} = 0 \Rightarrow \sum_{i=1}^{r} x_i^{\beta} \ln x_i - \frac{r}{\beta} - \sum_{i=1}^{r} \ln x_i = 0$$

Since all the likelihood equations are non linear, they may be solved using Newton Raphson method, to get estimates of θ and β.

4.3.3 Labeled slippage inliers model for Model-1

With $g(x)$ and $f(x)$ as described above, the likelihood under labeled slippage model referring to section (2.5) and substituting in equation (2.5.1), gives

$$\ln L=r_0\ln(1-p)+(n-r_0)\ln p-\ln\varphi_{r_1}(\phi,\theta)+r_1\ln\phi+(\beta-1)\sum\log x_{(i)}-\phi\sum_{i=1}^{r_1}x_{(i)}^{\beta}$$

$$+(n-r_0-r_1)\ln\theta-\theta\sum_{i=r_1+1}^{n}x_{(i)}^{\beta}$$

and the corresponding likelihood equations are

$$\frac{\partial\ln L}{\partial p}=\frac{-r_0}{(1-p)}+\frac{(n-r_0)}{p}=0 \tag{4.3.5}$$

$$\frac{\partial\ln L}{\partial\phi}=-\frac{\partial}{\partial\phi}\ln\varphi_{r_1}(\phi,\theta)+\frac{r_1}{\phi}-\sum_{i=1}^{r_1}x_{(i)}^{\beta} \tag{4.3.6}$$

$$\frac{\partial\ln L}{\partial\theta}=-\frac{\partial}{\partial\theta}\ln\varphi_{r_1}(\phi,\theta)+\frac{n-r_0-r_1}{\theta}-\sum_{i=r_1+1}^{n}x_{(i)}^{\beta} \tag{4.3.7}$$

and

$$\frac{\partial\ln L}{\partial\beta}=0\Rightarrow\phi\sum_{i=1}^{r}x_i^{\beta}\ln x_i+\theta\sum_{i=r+1}^{n}x_i^{\beta}\ln x_i-\frac{n}{\beta}-\sum_{i=1}^{n}\ln x_i=0 \tag{4.3.8}$$

Here (4.3.5) can be solved to get the estimate of p as $\hat{p}=(n-r_0)/n$. The equations (4.3.6) to (4.3.8) contains gamma and digamma functions. Solving (4.3.6) and (4.3.7) simultaneously we get the estimate of ϕ and θ. The parameter p is orthogonal to $(\phi,\theta)'$. Now

$$\varphi_{r_1}(\phi,\theta)=(n-r_0-r_1)\int_0^{\infty}\left\{1-e^{-x^{\beta}\phi}\right\}^{r_1}\left[e^{-x^{\beta}\theta}\right]^{n-r_0-r_1}\beta\theta x^{\beta-1}dx$$

$$=\frac{(n-r_0-r_1)\theta}{\phi}\beta\left(r_1+1,\frac{(n-r_0-r_1)\theta}{\phi}\right)$$

$$=\left[\frac{(n-r_0-r_1)\theta}{\phi}\right]\left[\frac{\Gamma(r_1+1)\Gamma\left(\frac{(n-r_0-r_1)\theta}{\phi}\right)}{\Gamma\left(\frac{(n-r_0-r_1)\theta}{\phi}+r_1+1\right)}\right]$$

and

$$\ln\varphi_{r_1}(\phi,\theta)=C+\ln\theta-\ln\phi+\ln\Gamma(z)-\ln\Gamma((z+r_1+1))$$

where

$$z=([n-r_0-r_1]\theta)/\phi.$$

hence

$$\frac{\partial}{\partial\phi}\ln\varphi_{r_1}(\phi,\theta)=-\frac{1}{\phi}+\frac{\partial}{\partial\phi}\ln\Gamma(z)\frac{\partial z}{\partial\phi}-\frac{\partial}{\partial\phi}\ln\Gamma(z+r_1+1)\frac{\partial z}{\partial\phi}$$

$$=-\frac{1}{\phi}+\left[\psi(z)-\psi(z+r_1+1)\right]\left(-\frac{(n-r_0-r_1)\theta}{\phi^2}\right)$$

and

$$\frac{\partial}{\partial\theta}\ln\varphi_{r_1}(\phi,\theta)=-\frac{1}{\theta}+\frac{\partial}{\partial\theta}\ln\Gamma(z)\frac{\partial z}{\partial\theta}-\frac{\partial}{\partial\theta}\ln\Gamma(z+r_1+1)\frac{\partial z}{\partial\theta}$$
$$=-\frac{1}{\theta}+\left[\psi(z)-\psi(z+r_1+1)\right]\left(\frac{(n-r_0-r_1)}{\phi}\right)$$

where

$$\psi(z)=\frac{\partial}{\partial\phi}\ln\Gamma(z) \quad \text{and} \quad \Gamma(z)=\int_0^\infty x^{z-1}e^{-x}dx.$$

The result from Abramovitz and Stegun (1965) is

$$\left[\psi(z)-\psi(z+r_1+1)\right]=-\sum_{j=1}^{r_1}\frac{1}{z+j}. \tag{4.3.9}$$

Using the above results, one obtain the likelihood equations as

$$\frac{\partial\ln L}{\partial\phi}=\frac{r_1+1}{\phi}-\frac{\theta(n-r_0-r_1)}{\phi^2}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right]-\sum_{i=1}^{r_1}x_i^\beta=0 \tag{4.3.10}$$

and

$$\frac{\partial\ln L}{\partial\theta}=\frac{(n-r_0-r_1-1)}{\theta}+\frac{(n-r_0-r_1)}{\phi}\left[\sum_{j=1}^{r_1}\frac{1}{z+j}\right]-\sum_{i=r_1+1}^{n}x_i^\beta=0 \tag{4.3.11}$$

Now (4.3.8), (4.3.10) and (4.3.11) can be solved using Newton-Raphson method to get the estimates of ϕ, θ and β.

4.4 Inliers detection using information criterion

Here three information criteria are used to detect inliers, which are already discussed in chapters 2, section (2.5) such as Schawarz's Information criterion ($SIC = -2\ln L(\Theta) + p\ln n$), the Schawarz's Bayesian Information criterion ($BIC = -\ln L(\Theta) + 0.5\frac{p\ln n}{n}$) and the Hannan-Quinn criterion defined as ($HQ = -\ln L(\Theta) + p\ln\left[\ln(n)\right]$). Here $L(\Theta)$ the maximum likelihood function and p is the number of free parameters that need to be estimated under the model. Below we develop the procedure for SIC scheme. The following model of no inliers for Model-1 is given by

$$SIC(0) = -2n\ln\theta - 2n\ln\beta + 2\theta\sum_{i=1}^{n} x_{(i)}^{\beta} - 2(\beta-1)\sum_{1}^{n}\ln x_{(i)} + 2\ln n \tag{4.4.1}$$

and the corresponding model with r inliers is

$$SIC(r) = -2r\ln\phi - 2(n-r)\ln\theta - 2n\ln\beta + 2\phi\sum_{i=1}^{r} x_{(i)}^{\beta} + 2\theta\sum_{i=r+1}^{n} x_{(i)}^{\beta} - 2(\beta-1)\sum_{1}^{n}\ln x_{(i)} + 3\ln n \tag{4.4.2}$$

Similarly, for Model-2, the model with no inliers is

$$SIC(0) = -2n\ln\theta + 2\theta\sum_{1}^{n} x_{(i)} + \ln n \tag{4.4.3}$$

and corresponding model with r inliers is

$$SIC(r) = -2n\ln\theta - 2r\ln\beta - 2(\beta-1)\sum_{1}^{r}\ln x_{(i)} + 2\theta\sum_{1}^{r} x_{(i)}^{\beta} + 2\theta\sum_{r+1}^{n} x_{(i)} + 2\ln n \tag{4.4.4}$$

The estimate of inliers say r is such that $SIC(r) = \min_{1\le r\le n} SIC(r)$, where $r, 1 \le r \le n-1$, is the unknown index of the inliers. According to the procedure, the Model with no inlier is selected if $SIC(0) < \min_{1\le r\le n-1} SIC(r)$. And the Model with r inlier is selected if $SIC(0) > \min_{1\le r\le n-1} SIC(r)$.

4.4.1 Simulation study

To illustrate the method of identifying inliers model the random samples of size 15 have been generated from Weibull distribution. The data under two models are as follows:

Model-1: Five observations are generated from Weibull with parameter $\phi = 0.50$ and $\beta = 1.1$ and remaining ten observations from Weibull distribution with parameter $\theta = 0.25$ and $\beta = 1.1$. The ordered observations are 0.1475, 0.4076, 0.5435, 0.676, 1.0885, 2.662, 2.662, 2.7381, 2.9781, 3.1589, 4.1746, 4.3598, 4.8724, 9.5612 and 10.2065.

Model-2: Here five observations are generated from Weibull with parameter $\theta = 0.1$ and $\beta = 3$. The remaining ten observations from exponential distribution with parameter $\theta = 0.1$. The ordered observations are 0.7418, 1.3926, 1.4866, 1.5082, 1.5279, 2.1699, 3.0111, 3.1058, 3.4249, 5.6212, 6.5393, 9.1629, 10.2165, 22.0727 and 32.1888.

The identification is done as follows we evaluate for each fixed r the maximum likelihood equation $\hat{L}_r$, and then consider $\hat{r}$ being that value of r for which likelihood is maximum. The estimates are presented in table(4.4.1) and (4.4.2) for model-1 and model-2 respectively. The *SIC*(0) under Model-1 and Model-2 are 74.22128 and 93.55538 respectively. *BIC* and *HQ* are also found for both the models with the following values.

Table 4.4.1. The Likelihood, parameter estimates and information criterion for Model-1.

r	$\hat{\phi}$	$\hat{\theta}$	$\hat{\beta}$	L	SIC	BIC	HQ
1	11.9537	0.171279	1.29631	-30.204	68.53215	30.47481	33.19269
2	5.92071	0.123884	1.441	-28.4621	65.04835	28.73291	31.45079
3	4.63278	0.081892	1.62678	-26.5204	61.16495	26.79121	29.50909
4	3.87879	0.050043	1.84369	-24.4484	57.02095	24.71921	27.43709
5	**2.31486**	**0.039079**	**1.92499**	**-23.6473**	**55.41875**	**23.91811**	**26.63599**
6	0.864944	0.074318	1.55034	-26.8847	61.89355	27.15551	29.87339
7	0.615074	0.070581	1.53585	-27.5234	63.17095	27.79421	30.51209
8	0.48757	0.061317	1.56115	-27.6898	63.50375	27.96061	30.67849
9	0.400866	0.052183	1.58932	-27.7814	63.68695	28.05221	30.77009
10	0.336613	0.042551	1.62689	-27.7629	63.64995	28.03371	30.75159
11	0.293833	0.042395	1.56627	-28.4133	64.95075	28.68411	31.40199
12	0.260573	0.03771	1.53975	-28.8061	65.73635	29.07691	31.79479
13	0.235881	0.031794	1.50512	-29.215	66.55415	29.48581	32.20369

Table 4.4.2. The Likelihood, parameter estimates and information criterion for Model-2.

r	$\hat{\beta}$	$\hat{\theta}$	L	SIC	BIC	HQ
2	6.25818	0.136196	-41.0662	87.5485	41.15647	42.06243
3	6.25055	0.12443	-38.5097	82.4355	38.59997	39.50593
4	6.32895	0.112701	-35.8889	77.1939	35.97917	36.88513
5	**6.44645**	**0.101003**	**-33.1883**	**71.7927**	**33.27857**	**34.18453**
6	4.47787	0.099347	-33.5527	72.5215	33.64297	34.54893
7	3.25159	0.096469	-34.7522	74.9205	34.84247	35.74843
8	3.03171	0.086918	-34.0851	73.5863	34.17537	35.08133
9	2.8828	0.07774	-33.4219	72.2599	33.51217	34.41813
10	2.28964	0.078485	-35.5625	76.5411	35.65277	36.55873
11	2.06388	0.073388	-36.518	78.4521	36.60827	37.51423
12	1.80399	0.072127	-38.2551	81.9263	38.34537	39.25133
13	1.68415	0.067832	-39.2468	83.9097	39.33707	40.24303

One can observe that the likelihood is maximum and $\min_{1\le r\le n-1} BIC(r)$ $\min_{1\le r\le n-1} SIC(r)=SIC(5)<SIC(0)$, and $\min_{1\le r\le n-1} HQ(r)$ corresponds to $r=5$, which was expected. The corresponding estimates of the parameters are shown in the tables (4.4.1) and (4.4.2). The graphical representations of the likelihood plot are given in figure (4.4.1) and (4.4.2).

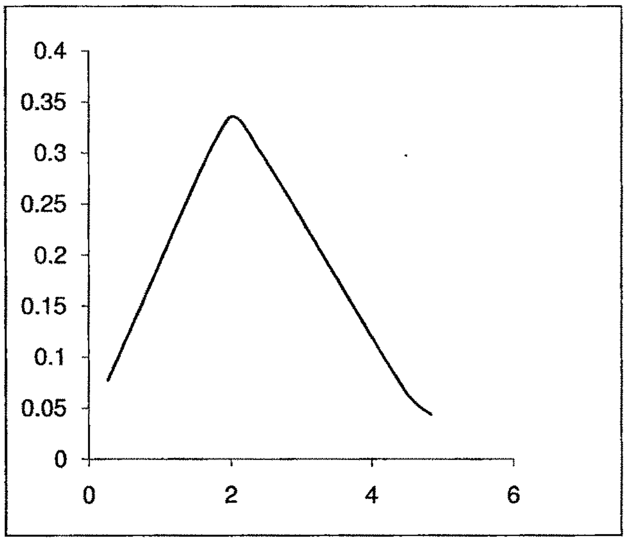

Fig. 4.4.1. The likelihood plot for Model-1

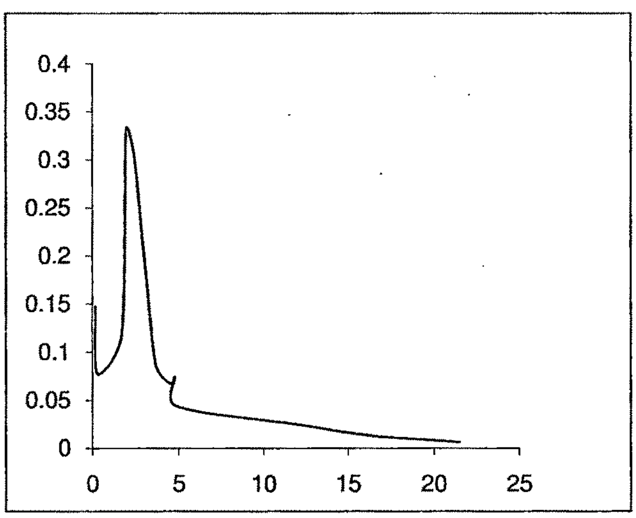

Fig. 4.4.2. The likelihood plot for Model-2

4.5 Data Example:

The example is based on Vanmann's (1991) data on drying of woods under different experiments and schedules. It is the example given in appendix, numbered E-3 S-1.

Under model-1 : The computed value $SIC(0)$ = 133.2468 > $SIC(9)$ = min SIC (r) = 98.46836. Also the likelihood is maximum for $\hat{r} = 9$. The corresponding estimates of the parameters are $\hat{\phi}$ =1.40087, $\hat{\beta}$ = 1.96982 and $\hat{\theta}$= 0.015968 as given in the table (4.5.1) below.

Under Model-2: The computed value $SIC(0)$ = 130.3241 and > SIC (13) = min SIC (r) = 125.7627. Hence value of $\hat{r}$ = 13. Similarly other information criteria and likelihood function gives us the same result. The estimates of the parameters are given in table (4.5.2).

Table 4.5.1. Estimates of parameters, likelihood, information criterion under M_1.

r	$\hat{\phi}$	$\hat{\theta}$	$\hat{\beta}$	L	SIC	BIC	HQ
1	17.0411	0.15391	1.1227	-58.6091	126.7524	58.67531	59.76537
2	6.62578	0.125022	1.20295	-56.8636	123.2614	56.92981	58.01987
3	5.61105	0.092835	1.32332	-54.588	118.7102	54.65421	55.74427
4	5.13406	0.06591	1.46092	-52.1563	113.8468	52.22251	53.31257
5	3.87184	0.050274	1.5623	-50.4173	110.3688	50.48351	51.57357
6	3.14867	0.036469	1.68353	-48.5128	106.5598	48.57901	49.66907
7	2.24189	0.029789	1.74933	-47.4397	104.4136	47.50591	48.59597
8	1.77073	0.021963	1.85802	-45.9382	101.4106	46.00441	47.09447
9	**1.40087**	**0.015968**	**1.96982**	**-44.4671**	**98.46836**	**44.53331**	**45.62337**
10	0.789435	0.029368	1.67797	-47.7531	105.0404	47.81931	48.90937
11	0.583143	0.032049	1.61366	-49.0733	107.6808	49.13951	50.22957
12	0.458334	0.036603	1.53082	-50.6362	110.8066	50.70241	51.79247
13	0.383572	0.037393	1.49579	-51.5492	112.6326	51.61541	52.70547
14	0.332632	0.040501	1.43709	-52.7273	114.9888	52.79351	53.88357
15	0.296617	0.041975	1.39664	-53.6335	116.8012	53.69971	54.78977
16	0.271311	0.04458	1.34706	-54.6304	118.795	54.69661	55.78667
17	0.252993	0.047369	1.29805	-55.6037	120.7416	55.66991	56.75997
18	0.240966	0.051993	1.23932	-56.7067	122.9476	56.77291	57.86297
19	0.230694	0.054786	1.19563	-57.6047	124.7436	57.67091	58.76097
20	0.221885	0.055836	1.16131	-58.3575	126.2492	58.42371	59.51377
21	0.215411	0.05678	1.12541	-59.1278	127.7898	59.19401	60.28407
22	0.210003	0.055551	1.09305	-59.8413	129.2168	59.90751	60.99757

Table 4.5.2. Estimates of parameters, likelihood, information criterion under M_2.

r	$\hat{\beta}$	$\hat{\theta}$	L	SIC	BIC	HQ
2	2.45514	4.85387	-65.4516	137.2593	65.51781	67.76414
3	2.14020	4.84449	-64.8678	136.0917	64.93401	67.18034
4	1.94013	4.83674	-64.2642	134.8845	64.33041	66.57674
5	1.84068	4.83136	-63.5878	133.5317	63.65401	65.90034
6	1.76655	4.82876	-62.9371	132.2303	63.00331	65.24964
7	1.73299	4.83517	-62.2321	130.8203	62.29831	64.54464
8	1.70401	4.84410	-61.5473	129.4507	61.61351	63.85984
9	1.68347	4.85958	-60.8549	128.0659	60.92111	63.16744
10	1.60145	4.97633	-60.3106	126.9773	60.37681	62.62314
11	1.53127	5.07469	-59.8869	126.1299	59.95311	62.19944
12	1.43583	5.16651	-59.7701	125.8963	59.83631	62.08264
13	**1.36456**	**5.22507**	**-59.7033**	**125.7627**	**59.76951**	**62.01584**
14	1.28113	5.24744	-59.9096	126.1753	59.97581	62.22214
15	1.21049	5.23142	-60.1883	126.7327	60.25451	62.50084
16	1.13620	5.16086	-60.6704	127.6969	60.73661	62.98294
17	1.06222	5.02958	-61.3154	128.9869	61.38161	63.62794
19	0.90209	4.52081	-63.2469	132.8499	63.31311	65.55944
20	0.82158	4.16407	-64.5197	135.3955	64.58591	66.83224
21	0.72684	3.69465	-66.283	138.9221	66.34921	68.59554
22	0.61228	3.11046	-68.9122	144.1805	68.97841	71.22474

We can observe that, for mixture of two different distributions, we do not get same number of inliers. Now the next problem is to decide which of the model discussed above is better ?

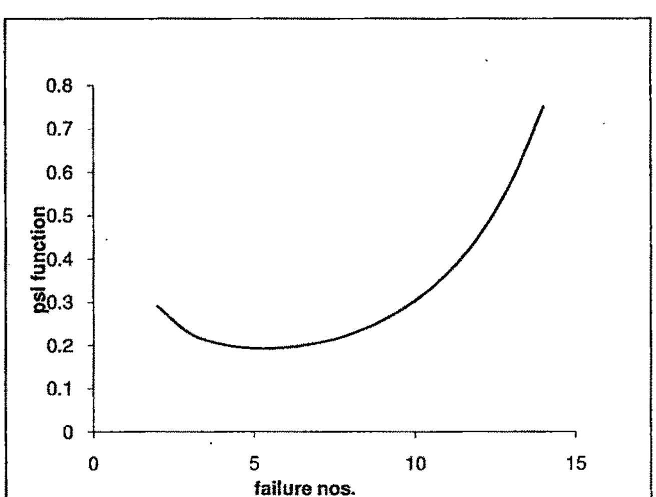

Fig. 4.4.3. The graph of $P[X_{(i)} < X_{(i+1)}]$ under model-1

4.6 Inlier detection using conditional distribution of total lives

This test makes use of basic properties of Poisson process. If one observes Poisson process for a fixed time T and if say n events occur in $[0,T]$ at times $0 \le x_{(1)} \le x_{(2)} \ldots\ldots \le x_{(n)} \le T$ then these times can be considered as ordered observation on a random variable uniformly distributed over $[0,T]$. Let $x_{(i)}$ = life time of i^{th} ordered unit. Then

$$P\Big[x_{(1)} \le X_{(1)} \le x_{(1)} + \Delta x_{(1)}, x_{(2)} \le X_{(2)} \le x_{(2)} + \Delta x_{(2)} \ldots\ldots\ldots x_{(n)} \le X_{(n)} \le x_{(n)} + \Delta x_{(n)}$$

$$| n \ events occur(0,T) \ \Big] =$$

$$= \lambda^n e^{-\lambda T} \prod_{i=1}^{n} \frac{\Delta x_{(i)}}{\left([\lambda T]^n e^{-\lambda T} / n!\right)}$$

$$= \frac{n!}{T^n} \prod_{i=1}^{n} \Delta x_{(i)}, \qquad x_{(1)} \le x_{(2)} \ldots\ldots \le x_{(n)} \tag{4.6.1}$$

For large value of n, $\bar{X}$ is approximately normal with mean $\frac{T}{2}$ and variance $\frac{T^2}{12n}$. It can be used to test, for large sample, whether or not the data is drawn from Poisson process. One can also show that if one observes a Poisson Process until exactly n events occur, then (n-1) r.v. can be considered as uniformly distributed over $\left(0, x_{(n)}\right)$.

In context of life testing if the failed items are not placed then all we need to do is to use total lives S_i where $S_i = \sum_{j=1}^{i} D_j$ and $D_i = (n-i+1)\left[x_{(i)} - x_{(i-1)}\right]$. Here D_i's are known as normalized spacing. If S_i is the total life observed in getting the i[th] failure then $S_1 \le S_2 \ldots\ldots \le S_n$.

$$S_r = X_{(1)} + X_{(2)} + \ldots\ldots + (n-r+1) X_{(r)} = \sum_{i=1}^{r} D_i, \quad r = 1,2 \ldots\ldots n. \tag{4.6.3}$$

Here also one can show that the total lives $S_1, S_2......, S_n$ can be considered as being drawn from a density function which is uniform over $(0,T)$. If the life test ends as soon as the first n failures occur, then the (n-1) r.v. $S_1, S_2......, S_{n-1}$ can be considered as being drawn from a density function which is uniform over $(0, S_n)$.

The fact that the conditional distribution of total lives is uniform over suitable interval makes it quiet evident that one has a good tool for detecting whether the failure rate is indeed constant. Thus the contamination of a purely exponential distribution by early failure would manifest itself in the pronounced tendency to get too many clustering together in the early part of total life thus violating uniformity. If the failure rate changes, for example, it increases with time then this should result in a tendency for failures to cluster together as time goes on, again violating uniformity. If the amount of failure data observed is quiet small, then we can expect large changes from exponentiality. Otherwise one can use a chi-square to detect whether the conditional distribution of times to failure or total lives deviate excessively from being normal.

4.6.1 A test for abnormally early failures (inliers)

Suppose that $x_{(1)} \le x_{(2)} ...... \le x_{(n)}$ are the n ordered failures. If all the $x_{(i)}$ are drawn from a common exponential then S_1 the total life in $\left[0, x_{(1)}\right]$ and $S_n - S_1$, the total life in $\left[x_{(1)}, x_{(n)}\right]$ are distributed independently of each other, where $\frac{2S_1}{\theta} \sim \chi^2_{(2)}$ and $\frac{2(S_n - S_1)}{\theta} \sim \chi^2_{(2[n-1])}$ degrees of freedom each. Hence the ratio

$$R = \frac{(n-1)S_1}{S_n - S_1} \sim F_{(2,2n-2)}. \tag{4.6.4}$$

If the ratio is too small then we assert that $x_{(1)}$ is abnormally small. More precisely if α is the level of significance, we will say $x_{(1)}$ is an inlier if

$$R=\frac{(n-1)S_1}{S_n-S_1}<F_{(2,2n-2),\alpha}. \tag{4.6.5}$$

Suppose one wants to detect $x_{(1)}$ and $x_{(2)}$ are inliers and if all the $x_{(i)}$ are drawn from a common exponential then S_2, the total life in $\left[0, x_{(2)}\right]$ and $S_n - S_2$, the total life in $\left[x_{(1)}, x_{(n)}\right]$ are distributed independently of each other.

$$R=\frac{(2n-2)S_2}{(S_n-S_2)/2} \sim F\,(4,4n-4) \tag{4.6.6}$$

If this ratio R is too small then we can conclude $x_{(1)}$ and $x_{(2)}$ are inliers. One can continue in similar manner, to detect whether $x_{(1)}, x_{(2)}.....x_{(r)}$ are inliers, where $r=3,4,.....n$ till we get first ratio which is greater than tabulated value. Hence at this point one can conclude r observations, till which the hypothesis is accepted, are inliers and rest of the observations are from target population.

4.7 Predictive approach to inlier model detection

The use of predictive distributions has been recognized as the correct Bayesian approach to model determination. In particular, Box(1980) notes the complementary roles of posterior and predictive distributions stating that posterior is used for the "estimation of parameters conditional on the adequacy of the model" whereas the predictive distribution is used for "criticism of the entertained model in the light of the current data". In examining two models, it is clear that the predictive distributions will be comparable whereas the posterior will not.

In this case there are n models, such as model with number of inliers r = 0, 1, 2....n-1. M_1 is considered model with 0 inliers. M_2 can be considered model with r inliers and (n-r) target observations. The procedure is as follows:

Let model M_1 assume that the data X are samples from independent random variables having a target exponential distribution with density

$$f(x|\theta)=\theta e^{-\theta x}, \qquad \theta>0,\ x>0. \tag{4.7.1}$$

The model M_2 assume that there are two distinct labels so that data $X=(X_1,X_2)$ where X_1 and X_2 are sampled from independent random variables having inliers and target exponential distribution, having $n_1 = r$ and $n_2 = (n-r)$ observations respectively, with density function as

$$f(x|\theta_i)=\theta_i e^{-\theta_i x}, \qquad \theta_i>0,\ x>0,\ i=1,2. \tag{4.7.2}$$

where $\theta_1=\phi$ the parameter of inliers distribution and $\theta_2=\theta$ the parameter of target distribution. If assumption regarding the vague prior density of the form under M_1 is $g(\theta)\propto\frac{1}{\theta}$. The likelihood under model M_1 is as follows

$$L(\underline{X},M_1)=\prod_{i=1}^{n} f(x_i|\theta,M_1)$$
$$=\theta^n e^{-\theta\sum_{i=1}^{n} x_i}$$

Then predictive density of observation x under M_1 is given by

$$f(x|X,M_1)=\frac{\int L(x,\theta,M_1)L(\underline{X},\theta,M_1)g(\theta)d\theta}{\int L(\underline{X},\theta,M_1)g(\theta)d\theta} \tag{4.7.3}$$

The model M_2 assume that there are two distinct labels so that data $X=(X_1, X_2)$ where X_i are sampled from independent random variables having a distinct exponential distribution with density

$$f(x|\theta_i)=\theta_i e^{-\theta_i x}, \qquad i=1,2,\ \theta_i>0,\ x>0. \tag{4.7.4}$$

The vague prior densities of the form $g(\theta_i) \propto \theta_i^{-1}$ for both the parameters are assumed, then the respective predictive densities under M_1 and M_2 are

$$f(x \mid X, M_1) = n(n\bar{x})^n / (n\bar{x} + x)^{n+1} \tag{4.7.5}$$

$$f(x \mid X, M_2) = n_i(n_i\bar{x}_i)^{n_i} / (n_i\bar{x}_i + x)^{n_i+1}, \quad i = 1,2 \tag{4.7.6}$$

where

$$\bar{x} = n^{-1}(n_1\bar{x}_1 + n_2\bar{x}_2) \quad \text{and} \quad \bar{x}_i = \frac{\sum_{i=1}^{n_i} x_i}{n_i}, \quad i = 1,2. \tag{4.7.7}$$

The prior density yields the optimal estimate of the density, in the frequency sense, among all estimates that are invariant with regards to transformation of scale using Kullback- Leibler measure of divergence.

The Predictive sample reuse (PSR) quasi–bayes criterion chooses the Larger of

$$L_1 = \frac{\prod_{i=1}^{2}\prod_{j=1}^{n_i}(n-1)(n\bar{x} - x_{ij})^{n-1}}{n\bar{x}} \qquad \text{for } x_{ij} > 0 \tag{4.7.8}$$

and

$$L_2 = \prod_{i=1}^{2}\prod_{j=1}^{n_i}\frac{(n_i-1)(n_i\bar{x}_i - x_{ij})^{n_i-1}}{(n_i\bar{x}_i)^{n_i}} \qquad \text{for } x_{ij} > 0 \tag{4.7.9}$$

The Predictive sample reuse (PSR) quasi–bayes criterion used by Geisser and Eddy (1979) chooses the model with Larger of

$$\hat{L}_1 = \prod_{i=1}^{2}\prod_{j=1}^{n_i}\frac{(n-1)}{(n\bar{x} - x_{ij})}\exp\left(-\frac{(n-1)x_{ij}}{(n\bar{x} - x_{ij})}\right) \qquad \text{for } x_{ij} > 0 \tag{4.7.10}$$

and

$$\hat{L}_2 = \prod_{i=1}^{2}\prod_{j=1}^{n_i}\frac{(n_i-1)}{(n_i\bar{x}_i - x_{ij})^{n_i-1}}\exp\left(-\frac{(n_i-1)x_{ij}}{(n_i\bar{x}_i - x_{ij})^{n_i-1}}\right) \qquad \text{for } x_{ij} > 0 \tag{4.7.11}$$

Above mentioned both the criterions are asymptotically equivalent to Akaike's criterion. One can use any of the above given criteria to obtain number of inliers in a given set of data.

4.8 Numerical illustration

The data represents ozone concentration in ppb monitored from morning 8 a.m. to evening 8 p.m. at express highway of Anand in the month of July on hourly basis. The data is collected by Dr. Sukalyan Chakraborty as a part of air pollution status monitoring of Anand district for his research. The observations arranged in increasing order of their magnitude are 14.00, 14.50, 15.00, 15.00, 17.00, 17.00, 19.00, 21.00, 21.80, 22.30, 23.00, 23.20 and 24.00. In table (4.8.1), *r* represents number of inliers observations to be considered. Level of significance is taken as 2.5 %.

Table 4.8.1. Inlier detection using Likelihood and Conditional method

r	Likelihood	$\hat{L}_2$	For Conditional Method				Conclusion
			D_i	S_i	Ratio	F-tab	
1	-51.2223	---------	182	182.0	1.5888	4.318725	Accept
2	-51.1812	5.7383E-23	174	356.0	1.6308	3.066233	Accept
3	-51.1434	5.9886E-23	165	521.0	1.6769	2.589498	Accept
4	-51.0982	6.3160E-23	150	671.0	1.7047	2.327027	Accept
5	-51.0853	6.3908E-23	153	824.0	1.7996	2.157011	Accept
6	**-51.0636**	**6.5672E-23**	**136**	**960.0**	**1.8773**	**2.036182**	**Accept**
7	-51.0712	6.5008E-23	133	1093.0	2.0208	1.944986	Reject
8	-51.1041	6.2351E-23	126	1219.0	2.2567	1.873191	
9	-51.1387	5.9862E-23	109	1328.0	2.5818	1.814874	
10	-51.1712	5.7747E-23	89.2	1417.2	3.0499	1.766351	
11	-51.2051	5.5642E-23	69.0	1486.2	3.8383	1.725199	
12	-51.2345	3.5233E-21	46.4	1532.6	5.3215	1.689750	
13	-52.2672	---------	24.0	1556.6			

For conditional method null hypothesis is rejected when r = 7 implies that number of inlier in the data set is 6 as shown in table (4.8.1). The likelihood is also maximum at $r = 6$. Using Predictive Method we have obtained $\hat{L}_1 = 5.21825\text{E}-23$ and maximum $\hat{L}_2 = 6.5672\text{E}-23$ corresponds to $r = 6$.

4.9 Goodness of fit

The problem of testing of goodness of fit to test whether the sample data is taken from modified mixture Weibull distribution against they are taken from single exponential or Weibull distribution is discussed in this section.

4.9.1 To test whether target observations are from Exponential

Our first test is

H_0 : the sample is from single population with exponential distribution i.e. $f(x,\theta)$

H_1 : the sample is from population with Modified Weibull distribution,

In terms of the MLE, the likelihood ratio test statistics for testing H_0 against H_1 is

$$\Lambda = \frac{L(\theta \mid H_0)}{L(\phi,\theta,\beta \mid H_1)} \tag{4.9.1}$$

$$\ln\Lambda = n\ln\theta - \theta\sum_{i=1}^{n} x_{(i)} - r[\ln\phi + \ln\beta_0] - (n-r)[\ln\theta + \ln\beta_1] - (\beta_0 - 1)\sum_{i=1}^{r}\ln x_{(i)} - (\beta_1 - 1)\sum_{i=r}^{n}\ln x_{(i)} + \phi\sum_{i=1}^{r} x_{(i)}^{\beta_0} + \theta\sum_{i=r+1}^{n} x_{(i)}^{\beta_1} \tag{4.9.2}$$

Under null hypothesis $Y_L = -2\ln(\Lambda) \sim \chi_4^2$. Reject H_0 for appropriate value of level of significance when $Y_L > \chi_{4,\alpha}^2$.

4.9.2 To test whether all observations are from single Weibull against they are from mixture of two (inliers and target) Weibull distributions.

Our second test is

$H_0: p=1$ the sample is from single (target) population from Weibull distribution with parameters $\beta \neq 1$ and $\theta > 0$.

$H_1: p<1$ the population distribution is Modified Weibull with parameters $\beta \neq 1$, $\phi > 0$ and $\theta > 0$.

In terms of the MLE, the likelihood ratio test statistics for testing H_0 against H_1, as used in test 1, is

$$\ln\Lambda = n[\ln\theta+\ln\beta]+(\beta-1)\sum_{i=1}^{n}\ln x_{(i)}-\theta\sum_{i=1}^{n}x_{(i)}^{\beta}-r[\ln\phi+\ln\beta_0]-(n-r)[\ln\theta+\ln\beta_1]$$
$$-(\beta_0-1)\sum_{i=1}^{r}\ln x_{(i)}-(\beta_1-1)\sum_{i=r}^{n}\ln x_{(i)}+\phi\sum_{i=1}^{r}x_{(i)}^{\beta_0}+\theta\sum_{i=r+1}^{n}x_{(i)}^{\beta_1}$$

(4.9.3)

Under null hypothesis $Y_L = -2\ln(\Lambda) \sim \chi_2^2$, then reject H_0 for appropriate value of level of significance when $Y_L > \chi_{2,\alpha}^2$.

4.9.3 Sequential Probability ratio test (SPRT)

SPRT is used to find number of inliers in given data set for both the models as shown in the following sub sections.

Case 1 : SPRT for model-1

To test whether inliers and target population is from single Weibull distribtion against they are from two different Weibull population, i.e with reference to section (1.5). The SPRT test is given as follows

H_0: Sample observations are taken from inlier population with interest parameter $\xi = \phi$.

H_1: Sample observations are taken from target population with interest parameter

$\xi = \theta$..

and likelihood ratio λ_m is given by $\lambda_m = \frac{L_{1m}}{L_{0m}}$ or equivalently

$$\ln \lambda_m = \sum_{i=1}^{m} \ln \frac{f\left(x_{(i)},\theta\right)}{g\left(x_{(i)},\phi\right)} =$$

$$= m\left(\ln\theta + \ln\beta_1 - \ln\phi - \ln\beta_0\right) + \left[\beta_1 - \beta_0\right]\sum_{i=1}^{m}\ln x_{(i)} - \theta\sum_{i=1}^{m} x_{(i)}^{\beta_1} + \phi\sum_{i=1}^{m} x_{(i)}^{\beta_0} \qquad m = 1,2,\ldots.n$$

(4.9.5)

For deciding number of inliers *r*, first arrange the observations in ascending order and then continue to take likelihood ratio for *m* = 1, 2.... by including observations one by one till we reject H_0. That is

if $\sum_{i=1}^{m} z_{(i)} \leq \ln B$ accept H_0 and take the next observation.

and

if $\sum_{i=1}^{m} z_{(i)} \geq \ln A$ reject H_0 and stop.

The corresponding *m* represents the first observation from $f\left(x_{(i)},\theta\right)$ and number of inliers $\hat{r} = m - 1$. Also

$$B = \frac{\gamma}{1-\alpha} \qquad A = \frac{1-\gamma}{\alpha} \tag{4.9.6}$$

where α represents probability of type I error and γ represents probability of type II error. Arrange $X_{(1)} \leq X_{(2)} \leq \ldots\ldots X_{(n)}$ and apply SPRT process till the hypothesis H_0 is rejected.

Case 2: SPRT for model-2

To test whether observations follow Weibull distribution against they follow exponential distribution. The SPRT test is given as follows:

H_0: Inlier observations are taken from Weibull population

H_1: Inlier observations from Weibull and target from exponential population

and likelihood ratio λ_m is given by $\lambda_m = \frac{L_{1m}}{L_{0m}}$ or equivalently

$$\ln \lambda_m = \sum_{i=1}^{m} \ln \frac{f\left(x_{(i)},\theta\right)}{g\left(x_{(i)},\phi\right)} = -m \ln \beta_0 - (\beta_0 - 1)\sum_{i=1}^{m} \ln x_{(i)} + \theta\left(\sum_{i=1}^{m} x_{(i)}^{\beta} - \sum_{i=1}^{m} x_{(i)}\right)$$

where $m = 1, 2, \ldots n$ (4.9.7)

For deciding number of inliers *r*, first arrange the observations in ascending order and then we continue to take likelihood ratio for *m*= 1, 2....*n* by including observations one by one till we reject H_0. Arrange $X_{(1)} \leq X_{(2)} \leq \ldots\ldots X_{(n)}$ and apply SPRT process till the hypothesis *H_0* is rejected.

Test criteria for rejection of *H_0*, using $\ln \lambda_m$ as defined for case 1 and case 2 in equations (4.9.6) and (4.9.7) is to reject *H_0*, if

$$\ln \lambda_m > \ln A \quad (4.9.8)$$

Corresponding value of *m* for which *H_0* was accepted last becomes number of inliers *r*.

4.10 Conclusion

The Akaike information criterion is a measure of the relative goodness of fit of a statistical model.. It can be said to describe the tradeoff between bias and variance in model construction, or loosely speaking between accuracy and complexity of the model.

Given a data set, several candidate models may be ranked according to their *AIC* values. From the *AIC* values one may also infer that e.g. the top two models are roughly in a tie and the rest are far worse. Thus, *AIC* provides a means for comparison among models—a tool for model selection. In general *AIC* = $2k - 2 \ln L$, where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Hence AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters.

To compare above two models, defined in section (4.3.1) and (4.3.2), obtained value of AIC for Model-1 is 94.9342 and for Model-2 is 123.4066. Clearly we can observe Model-1 is better than Model-2. i.e. Model representing inliers and target observations as Weibull distribution with different scale parameters is better. For same example discussed in section (4.5), the Pareto distribution had also been applied in chapter 2. Hence comparing Weibull against Pareto model, it was noted that *AIC* for Weibull distribution is 127.7126 > AIC for Pareto distribution is 59.17455. Hence one can conclude for that example Pareto model is better than Weibull model. The Pareto distribution is a power-tailed distribution which is a special case of a heavy-tailed distribution whose tails go to zero more slowly than exponential. In particular, in the cases where initial defects are present causing early failures, the Pareto distribution is found adequate to model such phenomenon.

Above result is supported by Jian-ming Mo and Zong-Fang (2008) who compared the sensitivity of aggregate operational value-at-risk in the Pareto distribution with that in the Weibull distribution to select an optimal model from the loss severity distributions of approximate goodness-of-fit. After the aggregate operational value-at-risk is obtained, the sensitivities of aggregate operational value-at-risk are compared when the loss severity distribution are respectively the Pareto and Weibull. The authors have shown that the sensitivity of aggregate operational value-at-risk with the Pareto distribution is far better than that with the Weibull distribution.

Another paper that discussed the comparision of Pareto and Weibull model was by Li-Hua Lai, Khoo, Murlidharan and Xie (2007) and Pei-Hsuan Wu (2008) and Wo-Chiang Lee (2009) have shown that using extreme value theory, generalized

Pareto distribution (GPD) fits the heavy-tailed distribution better than the lognormal, gamma, Weibull and normal distributions. In an empirical study, they determine the thresholds of GPD through mean excess plot and Hill plot.