International Journal of Scientific & Engineering Research Volume 8, Issue 7, July-2017 105
ISSN 2229-5518

Modeling Extreme ASA & AHT values using Extreme Value theory

Ashish Pandey

Abstract---EVTdeals with the stochastic behavior of the extreme values in a process which is in contrast to the Central limit theory where we model central tendency. Paper applies the framework on 2 key metrics (ASA & AHT) of a call center. ASA is the average speed of answering an incoming call that measures the service level from the customer's perspective. AHT is the average handle time to service calls and directly relates to the cost of operations as an additional second of AHT can have significant impact on the costs. Statistical inference about extreme events deals with the estimation of the probability of occurrence of extreme events.Article gives the initial overview on how to estimate the probability of extreme stress on AHT & ASA that directly relates to the risk of increased costs and customer dissatisfaction. Objective of the paper is to describe how Extreme value theory can be applied in the Contact center space where Average Handling Time and Average speed to answer an incoming call are the key KPI's.

Data is hypothecated for illustration purpose & tool used is R software

Content

1) POT/GPD Theory: Model fit on AHT values
2) GEV Theory: Model fit on ASA values
3) Measurement of high Quantiles & Conditional Expectations

STATISTICAL THEORY (GPD DISTRIBUTION FUNCTION)

K: Stochastic variable> target AHT & ASA measure

Function of distribution of Variable K: $F(K) = P(K \leq k)$

'u' :Threshold level where the excess events are given by $Y = K - u$ and have the following distribution

(1) $Fu(y) = P(K - u \leq y | K > u) = (F(y+u) - F(u))/(1 - F(u))$ As 'u' increases the distribution of stochastic variable converges to GPD.

- *Ashish Pandey is an FRM (Passed Part 1 & Part 2) and an MBA from Symbiosis International University with extensive experience into Business Analytics & Research across Financial Services domain and Financial Operations, India, PH-917709045800. E-mail: ashish.pandeysims@gmail.com*

CDF (Cumulative distribution function) of GPD

(2) $G\varepsilon, \beta(y) = [1 - \left(1 + \frac{\varepsilon y}{\beta}\right)^{-\frac{1}{\varepsilon}}, \varepsilon \neq 0]$

(3) $G\varepsilon, \beta(y) = [1 - exp\left(-\frac{y}{\beta}\right), \varepsilon = 0]$

(4) $Fu(y) = G\varepsilon, \beta(y)$

$'\varepsilon'$= tail index parameter

$'\beta'$= scale parameter.

Modeling Extreme values of Average Handling time (AHT) using GPD distribution in R software

Dataset is aht having individual observations of average time spent in handling calls measured in seconds.

Target AHT: 180 sec. Data consists of values above 180 seconds as the focus is on the time spent beyond the target.

No. of observations: 3000

Dataset ="aht" Variable="aht"

R Packages: "evir" to model GPD distribution and "fBasics" for descriptive analysis.

Descriptive statistics

basicStats(aht$aht)

Min	Mean	Std.dev	Kurtosis	Max
401 sec	639.18 sec	309.33 sec	39.82	5054

Histogram of the "aht"

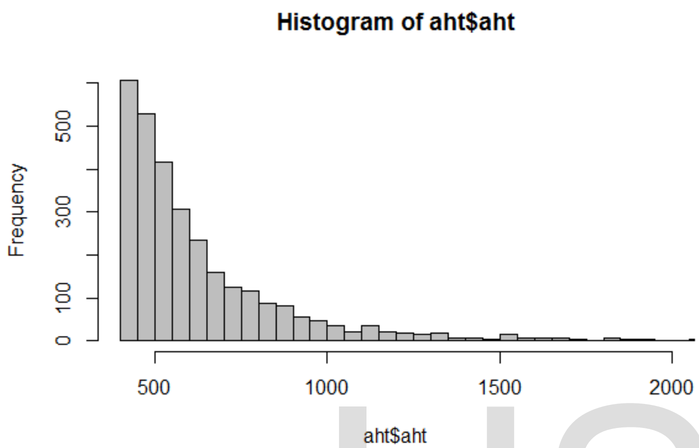

Fig. 1. Histogram of Average Handling time.

Inference: Values are skewed to the extreme right tail. Large deviations on the right tail are few in percentage terms but the magnitude impact can be significant on the Operational costs.

Understanding the nature of distribution of tail values of the Average Handling Time (AHT)

Empirical complementary cumulative distribution function (ccdf, that is, the empirical probability of the AHT exceeding any given threshold) CCDF measurement estimates the CCDF of the random variable $'x'$as defined by the following equation:

(5) $x = p/E(p)$

The following equation defines the CCDF of$'x'$

(6) $CCDFx(x) = \Pr\{x > X\} = \Pr\{p > X.E(p)\}$

Pr {e}: Probability of an event e (probability of the event that instantaneous value of AHT exceeds the mean AHT by at least X dB)

The below function plots the empirical ccdf of AHT values with both the axis on logarithmic scales. Vertical axis models probability of exceeding AHT levels (on Log scale), Horizontal axis models log of AHT levels

emplot(aht$aht,alog = "xy")

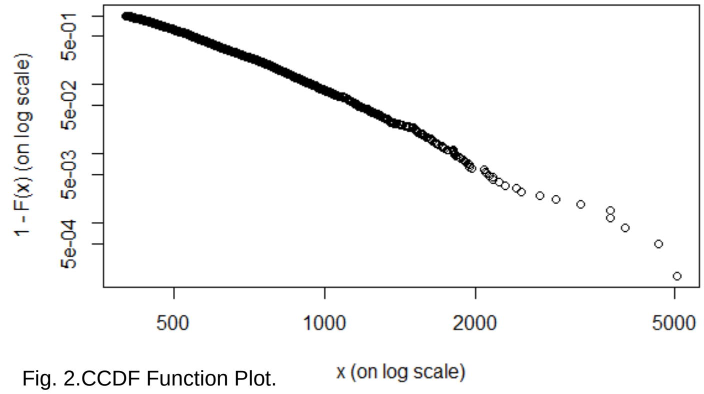

Fig. 2.CCDF Function Plot.

The linear plot of the ccdf function describes the fat-tailed nature of the data and Pareto-type distribution of AHT's.

The graph of QQ plot below describes the Fat-tailed nature with hypothesized exponential distribution.

qplot(aht$aht)

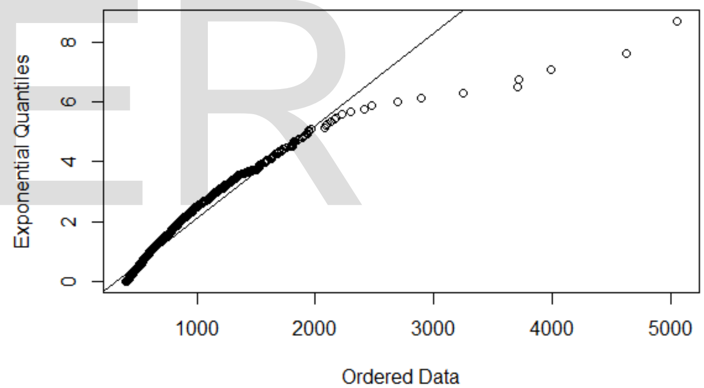

Fig. 3. QQ Plot reflects fat tails.

Fitting the GPD to threshold exceedances

Fitting the GPD distribution requires the Threshold level.

Suppose $'K'$is the random loss and$u > 0$the mean excess loss variable is the conditional variable $K - u|K > u$and the mean excess loss function $eK(d)$is defined by:

(7) $eK(d) = E(K - u|K > u)$

meplotfunction plots the Mean excess on the threshold levels. A positive gradient & linear nature of the mean excess plot indicates fat tails.

meplot(aht$aht,omit=50)

ISSN 2229-5518

Reasonable threshold =1000 seconds with 247 observations beyond the threshold.

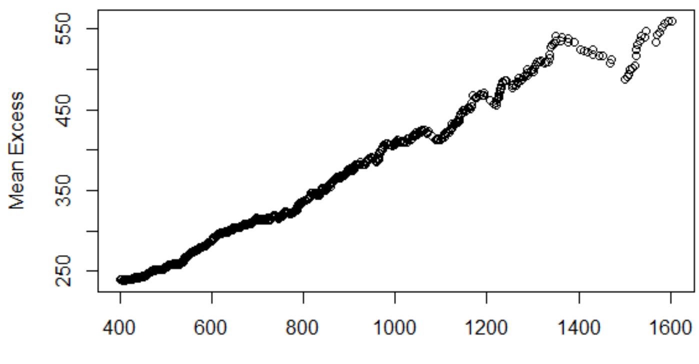

Fig. 4.Mean Excess Vs threshold.

Next step is to apply the GPD distribution on the tail values of the AHT dataset

gpd:fits the GPD distribution Parameter estimates **Tail index** $'\varepsilon'$and **Scale parameter** $'\beta'$.

GPDdistribution=gpd(aht$aht,threshold =1000)
GPDdistribution$par.estsgives the estimates of $'\varepsilon'$ and $'\beta'$

ε	β
0.2161968	319.1114115

GPDdistribution$par.sesgives the standard error for estimates of$'\varepsilon'$ and $'\beta'$

ε	β
0.07349516	30.70797805

To test how well the GPD distribution fits the empirical excess distribution we will use theplot(GPDdistribution)functionon the tail of the original distribution and fitted GPD distribution.

Below plot shows that GPD distribution fits nearly well to the excess empirical Values

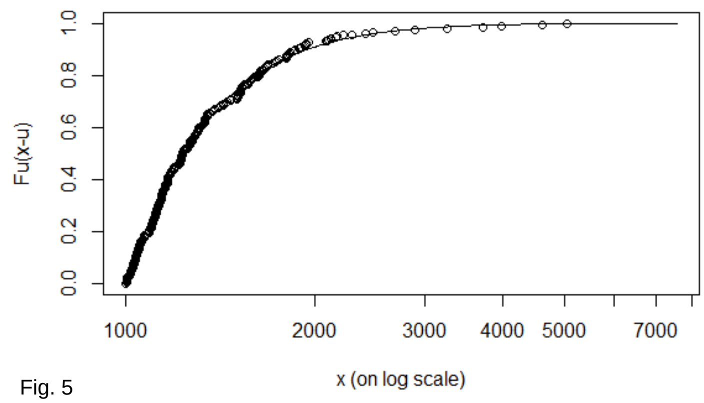

Fig. 5

Basis the Generalized Pareto distribution function that gives the cumulative probability distribution we can find the cumulative probability value for the extreme events.

$$1 - G\varepsilon,\beta(y) = 1 - [1 - \left(1 + \frac{\varepsilon y}{\beta}\right)^{-\frac{1}{\varepsilon}}, \varepsilon \neq 0] \quad (8)$$

1-K1200=1-[1-(1+0.2161968*1200/319.1114115)^(-1/0.2161968)]=6.37%
1-K1500=1-[1-(1+0.2161968*1500/319.1114115)^(-1/0.2161968)]=3.9%%

Probability of 1200<ASA<1500 = K1500-K1200=2.47%

Estimating the large quantiles using the fitted GPD model

Function:gpd.qon the tail plot of the fitted GPD distribution.Thetailplotfunction plots the tail of AHT distribution while gpd.qalong with pp=required quantile&p=Confidence level

TailPlot=tailplot(gpdfit)

gpd.q(TailPlot,pp=0.99,ci.p=0.95)
Lower CI Estimate Upper CI
1733.302 1850.596 2003.859

There's a 1% probability that the AHT will cross 1850.596. Conditional quantile (sfall)estimate gives the conditional expected value of AHT beyond the 99% quantile level.

gpd.sfall(TailPlot,0.99)
Lower CI Estimate Upper CI
2220.630 2492.348 3016.419

If the 99% quantile limit of AHT 1850.596 seconds is breached then the Expected value of AHT is 2492.348 seconds

ISSN 2229-5518

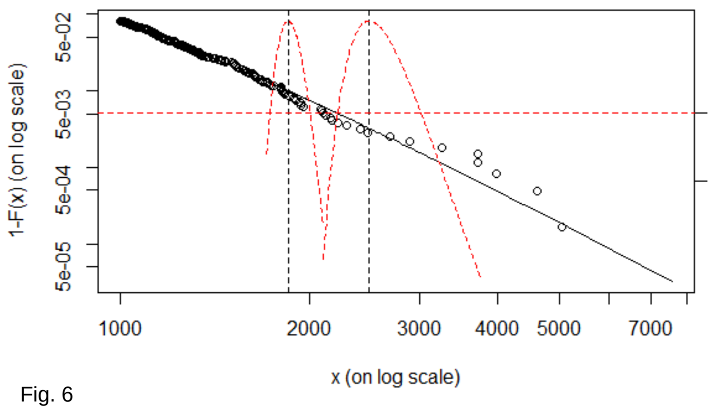

Fig. 6

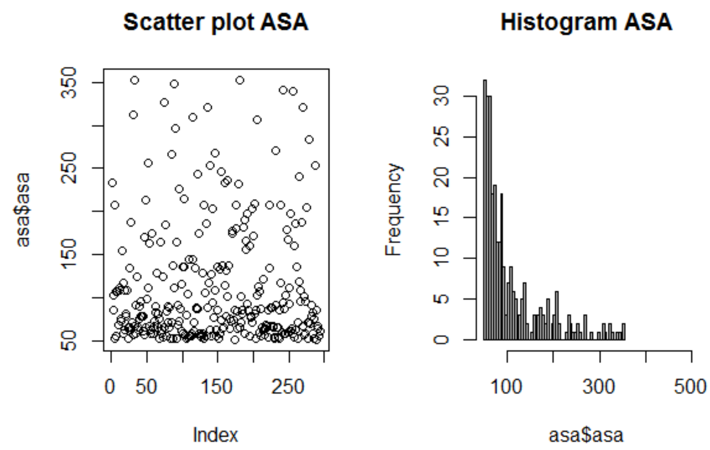

Fig. 7

STATISTICAL THEORY (GEV DISTRIBUTION FUNCTION)

Continuous probability distribution function that uses Block maxima approach of a sequence of iid random variables

The GEV distribution

(9) $G_{\varepsilon,\sigma,\mu}(y) = \exp[-\left(1+\frac{\varepsilon(y-\mu)}{\sigma}\right)^{-\frac{1}{\varepsilon}}, (1+\frac{\varepsilon(y-\mu)}{\sigma}) > 0, if\, \varepsilon \neq 0]$

(10) $G_{\varepsilon,\sigma,\mu}(y) = \exp[-\exp(-\frac{y}{\sigma}), y \in R \text{ if } \varepsilon = 0]$

$G_{\varepsilon,\sigma,\mu}(y)$ Incorporates the three (Fisher-Tippett) type's families

Frechet:$\varepsilon > 0$ Fat tails
Gumbel:$\varepsilon = 0$light tails
Weibul:$\varepsilon = 0$ lighter tails

Descriptive analysis: ASA Data (Average Speed to answer the calls).

R "evid" package to model the ASA values.

basicStats(asa$asa)

Min	Mean	Std.dev	Kurtosis	Max
51 sec	179.64sec	12.29 sec	7.92	1286 sec

Kurtosis>3 indicates fat tails

Modeling ASA extreme values using GEV distribution

fgev(asa$asa)function gives the estimates of GEV distribution as Location (μ), Scale (σ),tail index (ε).

Estimates
locscale shape
72.514 30.377 1.196

Standard Errors
locscaleshape
1.97085 2.88006 0.08836

confint(fgev(asa$asa))gives 95% confidence level of parameter estimates. Confidence interval of the tail index does not includes 0 so we can reject the null hypothesis of shape (ε) =0

2.5 % 97.5 %
loc68.651631 76.377230
scale24.732403 36.022047
shape1.023238 1.369593

Since (ε) >0Frechet distribution is the possible candidate to model maximum ASA values. Greater accuracy for the confidence intervals is usually attained by the profile likelihood. plot(profile(fgev(asa$asa)),ci = c(0.95, 0.99)

ISSN 2229-5518

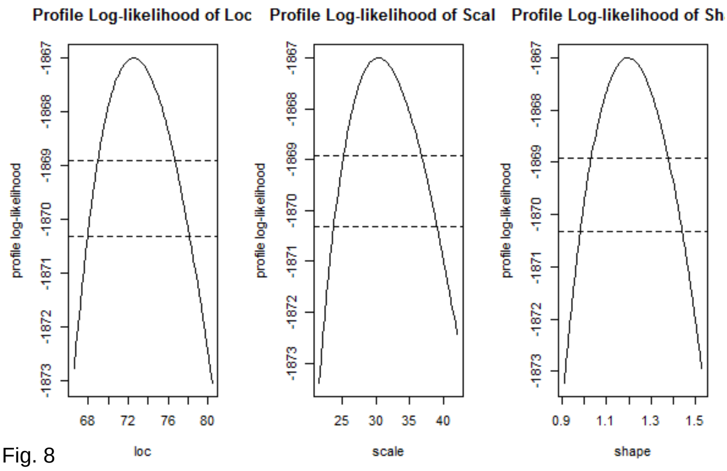

Fig. 8

Diagnostic plots of the fitted GEV distribution

```
gev.diag(gev.fit(asa$asa))
```

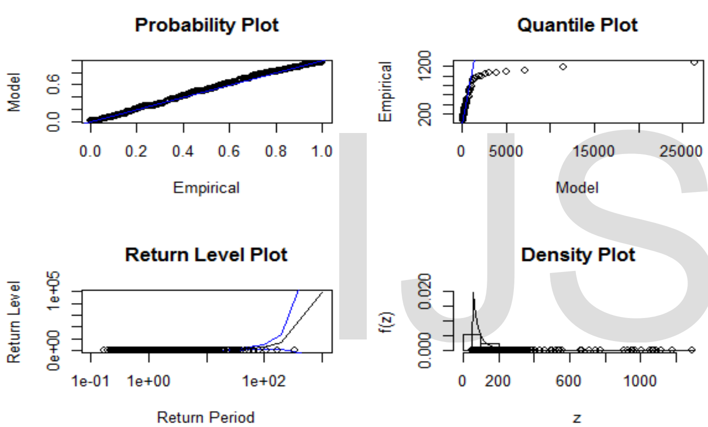

Fig. 9. Diagnostic Plots

Probability plot shows a reasonable GEV fit withFrechet distribution being nearly adequate.

Cumulative probability value for the extreme events can be calculated as.

$$(11) \quad 1 - G\varepsilon,\sigma,\mu(y) = 1 - exp[-\left(1 + \frac{\varepsilon(y-\mu)}{\sigma}\right)^{-\frac{1}{\varepsilon}}, (1 + \frac{\varepsilon(y-\mu)}{\sigma}) > 0, if\,\varepsilon \neq 0]$$

1-k200=1-exp(-(1+1.196*(200-72.514)/30.377)^(-1/1.196))
=9.01%
1-K250=1-exp(-(1+1.196*(250-72.514)/30.377)^(-1/1.196))
=6.35%

Probability of 200<ASA<250 = k250-k200=3.84%

CONCLUSION

The Primary objective of the paper was to describe how the Extreme Value Theory and the aligned distributions like Generalized extreme value distributions & Peak over threshold approach can be used to derive statistical inferences around extreme values of AHT & ASA KPI's. Central limit theorem would not be able to provide inferences around extreme events that may have a significant impact in terms of magnitude but can be frequency could be really low.

Extreme Value theory has its applications across various domains and with this paper the idea was to give an example of how it can be used in the contact center domain.

REFERENCES

[1] McNeil, Frey, and Embrechts (2005) "Quantitative Risk Management: Concepts, Techniques and Tools"

[2] Introduction to R for Quantitative Finance: By: Agnes Vidovics-Dancs; Daniel Havran; Dr. Edina Berlinger; GergelyDaróczi; MártonMichaletzky; Michael Puhle; Péter Csóka; Dr. Kata Váradi; ZsoltTulassay

[3] Topics in Data Analysis Using R in Extreme Value Theory : By Helena Penalva1,Manuela Neves2 and Sandra Nunes3

[4] Stephenson, A. G. (2002): evd: Extreme Value Distributions.

[5] Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values.Springer-Verlag, London

[6] Beirlant, J, Goegebeur, Y., Teugels, J. and Segers, J. (2004). Statistics of Extremes:Theory and Applications. Wiley, England.

[7] Stephenson, A. G. (2012): ismev: An Introduction to Statistical Modeling of Extreme Values. Original S functions written by Janet E. Heffernan with R port and R documentation. R package version 1.38, URL http://CRAN.Rproject.org/package=ismev.

[8] Reiss, R. D. and Thomas, M. (2007). Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields. Third Edition, Springer Verlag

[9] Embrechts, P., Kluppelberg, C. and Mikosch, T. (2003). ¨ Modelling Extremal Events for Insurance and Finance. Springer-Verlag, London.