2017

Data Analysis of Transmission Line Restoration Times

Sameera Kancherla
Iowa State University

Data analysis of transmission line restoration times

by

Sameera Kancherla

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Electrical Engineering

Program of Study Committee:
Ian Dobson, Major Professor
James D. McCalley
Venkataramana Ajjarapu

The student author and the program of study committee are solely responsible for the content of this thesis. The Graduate College will ensure this thesis is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

DEDICATION

To the memory of my Grandmother!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to use this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis.

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Ian Dobson for his guidance, patience and support throughout this research and the writing of this thesis. His insights and words of encouragement have often inspired me and helped me find a new direction.

I am grateful to my committee members Dr. Venkataramana Ajjarapu and Dr.James Mc-Calley for their efforts and contributions to this work.

I would like acknowledge and thank Arend and Verna Sandbulte for their financial support.

I thank Bonneville Power Administration for making publicly available the outage data that made this thesis possible.

I am absolutely grateful to my family for their love and support throughout. A special thank you to my classmates Rajaz Khan Amitava, Nichelle'Le Carrington and Vikram Kumar for being supportive and encouraging me.

Lastly, I would like to thank every friend and associate who has helped me in writing this thesis and achieving a major milestone in my academic career.

ABSTRACT

The world is highly dependent on electricity, and any service interruptions can be contagious sometimes leading to devastating blackouts. This in turn have severe impacts on customers. Large service interruption can impact an entire region. Therefore, reliability is an integral part of the system operations for utilities. On an interruption of power supply caused by a transmission or distribution failure, measures taken to restore the service highly depend on the interruption duration of normal supply paths. This thesis is a systematic study of transmission line restoration times with statistics obtained from a utility's data. The empirical probability distribution of transmission line restoration times is obtained from 14 years of field data. The distribution of restoration times has a heavy tail that indicates that long restoration times, although less frequent, routinely occur. The heavy tail differs from the convenient assumption of exponentially distributed restoration times, impacts power system resilience, and makes estimates of the mean time to repair highly variable. The mean restoration time of the heavy tailed distribution and its confidence interval is estimated using special bootstrap methods and its implications are outlined. The heavy tail in transmission line restoration times is one factor to be considered in assessing power system resilience.

CHAPTER 1. INTRODUCTION

A current global challenge in modern society is to meet the demand of electricity. Large interconnected systems are prone to stress with the increased demand. This stress might result in cascading outages and eventually blackouts. Blackouts in North America, Europe and other parts of the world in the past few years have demonstrated severe economic losses. Natural calamities such as windstorms, floods, lightning have caused major transmission and distribution line outages and have had major impact on the nations' public security, infrastructure and economic prosperity. Researchers predict that due to global warming and climate change these events may occur more frequently causing more damage. The temperature of water and air increase due to greenhouse gas emissions has resulted in an increase in storms, hurricanes and floods. There is a need for enhancement of the existing grid infrastructure to be more resilient, so as to withstand the shock and recover instantly from hazards. Researchers are motivated by these factors to develop techniques which help assess the consequences of natural disasters in a more comprehensive and systematic way to improve the grid resilience.

Assessing the impact of blackouts on our society requires adequate modeling of the restoration of transmission lines after they are outaged. The restoration depends on many factors such as weather, location, type of failure, and crew availability. Timely restoration of electric power after a blackout depends on the quick restoration of the outaged lines, and even if a line outage does not lead to load shed, the resilience of the power transmission system to other contingencies decreases during the outage. Several papers take into account multiple factors such as weather, voltage levels, load shed, and cascading of outages for post outage/blackouts analysis(2)(3)(4)(15).

1.0.1 Motivation

Transmission line outages whether single, multiple or cascaded can significantly alter the normal operation between supply and load points. Often, these service interruptions result in longer restoration times and cause severe damage to the society. The restoration process largely depends on the severity of the event, and also on the local weather conditions for utilities to take action. Moreover, the majority of the outages are dominated by weather-related events which can delay the process even more than expected. Studies also considered the influence of voltage levels on the transmission line outages, it is a reasonable speculation to check the influence of causes for outages on the restoration time.

The intricacies arise when processing data which exhibits a heavy tail phenomenon; that is the distribution of the outage duration is a heavy-tail. In such systems the convergence properties of sums of heavy tailed random variables is different. The convergence of sample mean to the true mean is slow or sometimes may not be achieved. Since there must be enough small samples (outages) to offset the large but rare events, the large ones can have a dominating effect on the convergence. In reality, large rare events do happen and cannot be excluded from the data. It ultimately leads to high variability in the reliability metrics which makes restoration models and analysis more difficult and needs to be addressed. This complex scenario motivates analyzing the heavy tailed statistical data that is studied in this thesis.

1.0.2 Objective

A typical power system outage database contains information such as the area/region, cause, time of occurrence, but has limitations on it. The database might not be comprehensive as the outages are reported to the government by major electricity providers and operators. Sometimes the cause for an event may not be obvious or there could be missing information which is difficult to interpret. Hence, there is equal weight for the data collection as well as the analysis. Mean restoration time is the average time taken by a system component to be placed back in service after an outage. The restoration includes re-energizing a tripped undamaged component as well as repairing a damaged component. When a system component has a high mean time to repair, there is a long restoration time leaving a higher vulnerability which is undesired. The objective of this thesis is to analyze the restoration times of the transmission line outages in transmission line data. In particular, this thesis addresses the following:

- The complications involved in obtaining statistics from heavy-tailed distributions
- Importance of having a larger data set to draw precise conclusions
- Statistical variations in the mean restoration time
- Annual availability of the transmission lines
- Influence of the presence of rare, more events in the mean restoration time calculations

1.0.3 Organization of Thesis

This chapter has presented the motivation, and the objective of the thesis.
Chapter 2 is a literature survey on traditional reliability methods used for modeling the failure and repair rates of a system.
Chapter 3 presents the processing and filtering of the raw data, the data was checked for effects which caused longer restoration times using time correlations, cascading effects, influence of voltage levels and weather causes.
Chapter 4 shows the implications of the heavy tail distribution observed in the data and the complexity involved in dealing with the presence of extremely large values in the dataset which makes results sensitive.
Chapter 5 describes statistical inferences obtained from the distribution of and finds that the annual estimates of mean restoration time are highly variable.
Chapter 6 concludes the research

CHAPTER 2. REVIEW OF LITERATURE

2.1 Power Outages Causes

Natural causes, equipments failure, maintenance issues, human errors and power surges can be reasons for power outages. Depending on the cause, power outages last from a second to a few months. Even short interruptions in power may sometimes lead to public disruption, loss of production and revenues. It is important to identify the causes for power outages in order to safeguard the grid from devastating effects.

The following two sections describe some causes for power outages.

2.1.1 Weather

About 70 percent of power outages in the United States are related to weather conditions, such as lightning, trees, ice, storms, wind and tornadoes. Historically, natural calamities have caused severe power outages which often led to large blackouts and large economic losses.

- Lightning: About 30 percent of power outages are related to lightning. Lightning can affect the line in two ways:

  (1) it hits the line, the surge causes flash over and damages the equipment/line

  (2) it strikes the trees which are around a line causing damage to the line.

- Water and Dust: Moisture and dust cause short circuits and power failures. Transmission lines in the areas of dust, sandstorms or more exposure to water need to be protected with sealed circuits.

- Ice: Areas with more snow or ice storms affect transmission lines by the accretion of ice on the line thereby damaging the equipment. Sometimes ice on a tree branch can break it so that it falls on the line causing a power failure.

- Rain and Floods: Both overhead and underground transmission system are prone to flooding. The system needs to be shutdown to prevent further damage causing service interruption.

- Winds, Tornadoes, and Hurricanes: Very high winds might uproot the trees, break the poles and sometimes tree limbs. When these come in contact with the transmission line, they damage it and also the protection equipment.

- Heat: High temperatures effect the transmission line in many ways. Heavy load along with high temperatures, heats up the transformers and other equipment, and causes voltage sag resulting in outage. High currents may stretch the cables to the point where the line can no longer withstand the power flow when needed.

Environmental conditions play a critical role in the analysis of outages. Weather remains one of the major causes for power outages and there has been a lot of emphasis on modeling weather-related outages for both transmission and distribution reliability(2)(12). Ref.(15) explains the impact of geographical location on transmission line outages and their frequency.

2.1.2 Maintenance

Occasionally, transmission lines are subject to regular maintenance and upgrades. With this type of outage, the utility notifies their customers well in advance with an estimate of interruption time. Some of the causes are mentioned here.

- Line material failure and Equipment Failure: Heavy stress on the lines may cause even well maintained equipment to fail.

- Animals: small animals like squirrels climb up the equipment and chew some of the cables leading to outage. In addition, raccoons, snakes, bird droppings are also some of the causes.

- Excavations and Renovations: Building new and tall structures may sometimes disturb the transmission lines. On the other hand, digging into the ground in the wrong place may damage the underground equipment contributing to outage.

- Public: Theft of electricity by tapping at different location of the line, theft of electrical equipment can cause outage. Accidents - such as vehicles hitting the utility poles or transformers are also reasons for power outages.

During peak hours of high demand, service interruptions occur due to limited power supply. At the time of high stress, utilities institute “brownouts”; i.e, lowering the voltage levels. Brownouts damage electrical motors as they overheat their insulation. Brownouts are also a type of power outages. Another most common type of outage is momentary interruption. No matter how long the interruption is, it is very inconvenient for any industry, residence or utility. The protective devices on the power system respond immediately when they sense a fault and prevent potential damages resulting in momentary service interruption.

2.1.3 Transmission Line Restoration Time Distributions

There has been heavy emphasis on modeling the failure rate probabilities while restoration models have less attention. In fact power system restoration is an important factor while assessing power system resiliency. When computing mean steady state reliability parameters of a power system, it is customary to assume exponential restoration times for the components. Non-exponential down times can give significantly different results compared to exponential downtimes (5). When models include rare events, it is important to identify a suitable distribution to characterize repair times accurately (9). Non exponential and time-varying failure rates are approximated using Stochastic Point Process, Monte Carlo Simulation, Markov Renewal Process etc as discussed in (6), but may not be accurate as the simplification involves many assumptions.

Other standard non-exponential distributions for modeling restoration times include Gamma, Weibull, Normal, and Log-normal distributions. They have exponentially decaying tails that make very long restoration times vanishingly unlikely, whereas the log-normal distribution can produce long restoration times. In distribution systems, log-normal distributions of line restoration times are considered in (7)(10). In transmission systems, there are not many published sources available for line restoration data. Ref. (13) models field data for restoration times of 345 kV lines using gamma distributions with shape parameter less than one. Ref. (9) fits field data in England and Wales for restoration times of 275 kV and 400 kV lines with a log-normal distribution. With the exception of the log-normal distribution in (9), the distributions assumed for line restoration at the transmission system level are not heavy-tailed.

2.1.4 Power Laws

The dataset used in thesis finds a tail in the distribution of restoration times that is somewhat heavier than log-normal in transmission system field data from a North American utility, and outlines some statistical consequences of the high variability of the restoration times caused by the heavy tail of the distribution. The repair state steady state probability, frequency and mean repair time are independent of the distribution of restoration times in many useful cases such as independent components. However, the distribution of restoration times affects the mean repair time for some common-mode failures (8) or if there are duration dependent effects (5), and significantly impacts the distribution of reliability indices about their mean values (7). Some heavy tail distributions have infinite variance. Ref. (18) discusses a step by step procedure for analyzing heavy tails in empirical data elaborately. Sample mean is Studentized over the actual mean and special methods needed to estimate confidence intervals are mentioned in (21). This variability in the estimates greatly affects the annual availability of the transmission lines.

CHAPTER 3. THE LINE OUTAGE DATA

A North American utility's historical data for 14 years from 1999 to 2012, is used for analysis (14). This data includes the duration of outages with dates, length and voltage level of transmission lines, type of outage; i.e, automatic or planned (maintenance) and causes of all outages. Total reported outages are about 42561 out of which automatic outages are about 10942 for various causes. Some of the causes for automatic outages include weather, wind, lightning etc. as shown in Table I.

About 80 percent of outages in the data are weather related and the remaining are due to technical and maintenance issues. In addition, bird droppings, malicious tripping, fire, wire arc switching, line material failure, malicious tripping, fire etc. are also some of the causes. High speed winds accompanied with rain/storms can be devastating and cause damage to electric systems and these have been one of the major causes for service interruptions. For such outages, the mean restoration time, number of consumers affected and economic losses can be high since most of them carry bulk power over long distances.

Momentary outages- those with 0 minute duration and sustained outages -with 1 minute or above also exist. By neglecting the momentary outages, the total non-momentary automatic outages in the data are 5594.

3.1 Processing the data

In general outages are categorized into few types; i.e, planned outages, forced outages, semi forced outages etc. In the current dataset, there are three types of outages namely momentary, planned and automatic outages. Momentary outages (ones with zero minute duration) are eliminated from the data. Automatic outages happen without one's control, implying that

Table 3.1 Causes for Non-Momentary Automatic Outages

Cause	Number of outages
Foreign Trouble	2010
Unknown	661
Lightning	519
Tree Blown	213
Wind	140
Weather	101
Less Frequent Causes	1950

they occur due to random interval or external causes. On the other hand, planned outages are the ones scheduled for maintenance activity. Most of the automatic outages are repairable but some can be consequential failure events. Major outages or blackouts which occurred previously have been associated with multiple dependent failure events (cascading outages).

Conventionally, the models for restoration use an exponential distribution; i.e, a constant failure rate. This is not the case in data under study. The distribution of duration of all automatic outages is a survivor function (complementary cumulative distribution function). The empirical probability distribution of transmission line restoration times from 14 years of data from a large utility is obtained. It does not follow exponential distribution but exhibits a heavy tail as shown in the figure 3.2. The distribution of restoration times has a heavy tail that indicates that long restoration times, although less frequent, routinely occur. To understand the heaviness of the tail, it is reasonable to look for evidence for the factors which could cause heavy tailed distribution. For instance, one factor is the nominal operating voltage level of a line and Ref.(15) claimed that the outage duration associated with weather varied significantly with operating voltage level.

The following questions motivated to further process the data.

- Any influence of operating voltage level of the transmission line on the restoration time ?
- Is there any effect of cascading of outages ?
- Any specific cause which impacted the heavy tail ?
- Did the outages with similar causes overlap with each other ?

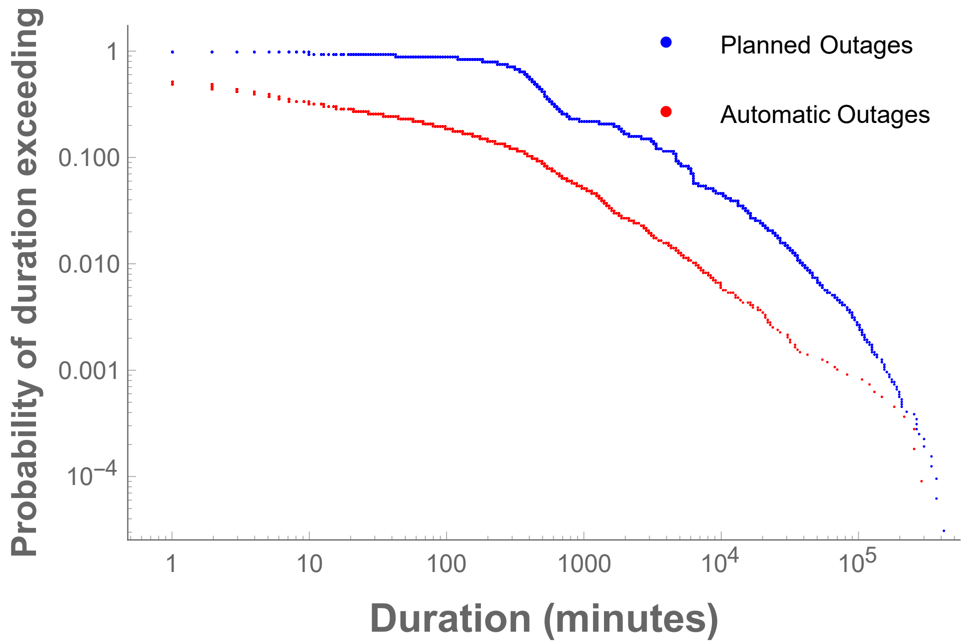

Figure 3.1 CCDF of Automatic and Planned Outages

The distribution of duration of outages is shown as a complementary cumulative distribution function (CCDF) or survivor function, which is the probability of outage exceeding against outage duration in minutes. The data contains both planned and the automatic outages, figure 3.1 shows the CCDF of both automatic and the planned outages which have heavy tails. Figure 3.2 shows the distribution of all non-momentary automatic outages. These non-momentary automatic outages are the focus of the rest of this thesis.

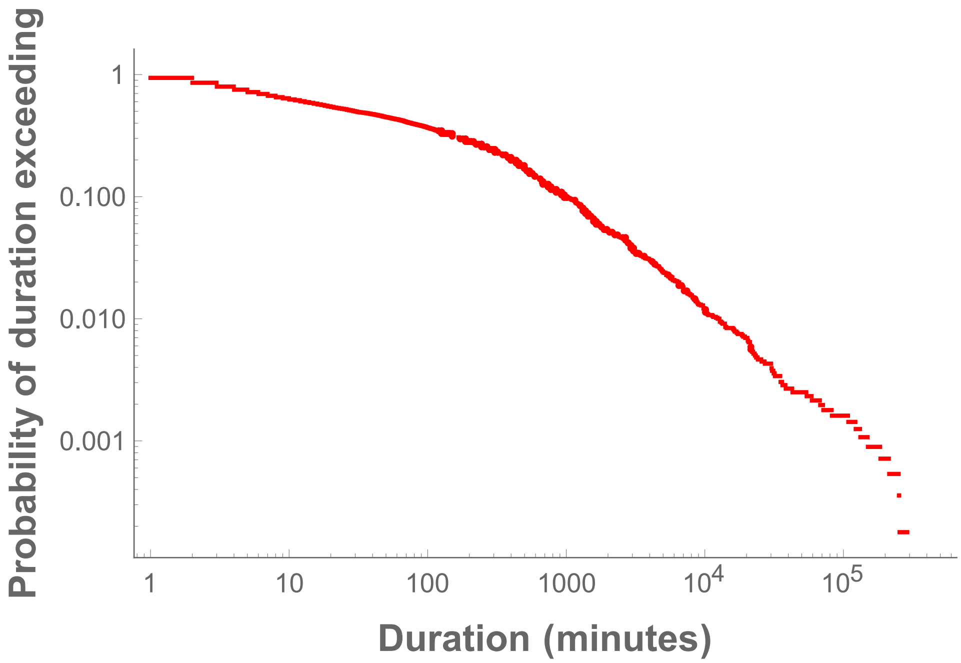

Figure 3.2 Survivor function of outage duration in utility data

3.1.1 Outages Classified by Causes

As discussed in the previous section, weather plays a dominant role in power outages. To reaffirm this fact, major causes (based on the number of outages listed in Table 3.1) in the data are considered and a CCDF plot is produced to spot the influence. Individual CCDF plots are shown in Figure 3.3 to understand the influence of each major cause.

Foreign trouble has substantial influence on the duration. It is noted that the "Foreign Trouble" follows the same pattern of the original CCDF of automatic outages. There are two interpretations for the definition of Foreign Trouble NERC's Transmission Availability Data System (TADS) defines Foreign Trouble as (16): "Automatic Outages caused by foreign interference from such objects such as an aircraft, machinery, a vehicle, a train, a boat, a balloon, a kite, a bird (including streamers), an animal, flying debris not caused by wind, and falling conductors from one line into another". NERC clarifies that these outages are not caused by a utility employee. Note that BPA includes "Outages caused by a non-BPA utility"(14) in the definition of Foreign Trouble.

From Figure 3.3 of individual Complementary Cumulative Distribution Function (CCDF)

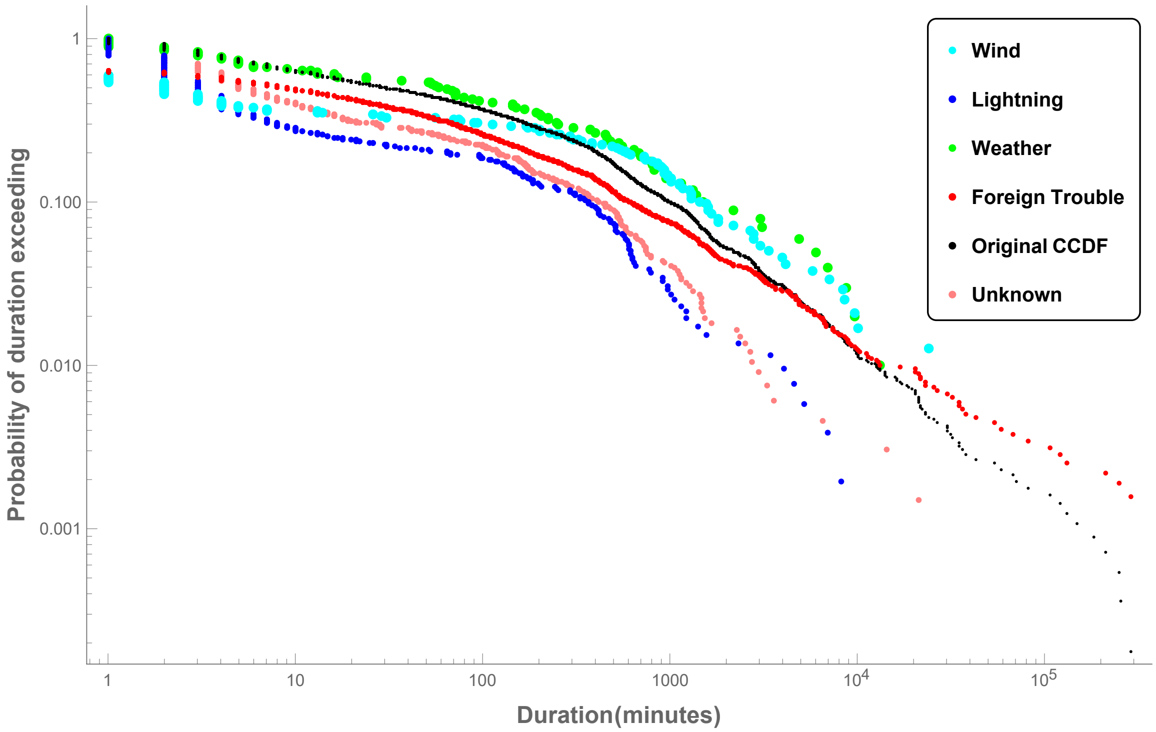

Figure 3.3 CCDF plot of outages by causes

for each cause it is observed that "Foreign Trouble" has higher influence on the overall CCDF of the outages.

3.1.2 Voltage Levels

There are 600 transmission lines in the data. The transmission line operating voltage levels are 69kV, 115kV, 230kV, 287kV, 345kV, 500kV. Table 3.2 shows the number of outages categorized by their voltage levels. The durations of outages are grouped according to their voltage levels separately. Ref.(15) emphasized the need to categorize transmission lines by voltage levels instead of grouping them to single category. However, the individual CCDF plots of voltage levels did not show any substantial difference in the pattern except that there may be more of the longer 500kV outages. These individual plots followed the original automatic outage CCDF as shown in the Figure 3.4.

To further define the variability in the numbers, basic statistics are calculated based on the cause and individual operating voltage levels. From Table 3.1 the causes Foreign Trouble,

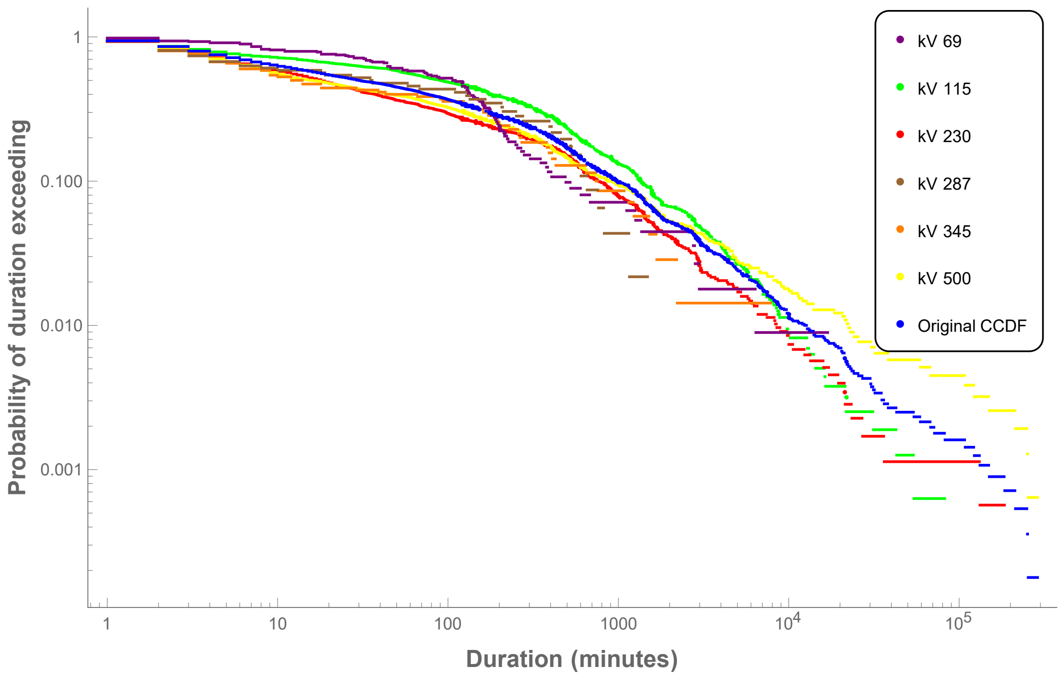

Figure 3.4 CCDF of the operating voltage levels of transmission lines

Table 3.2 Number of outages categorized by voltage levels

Voltage Level	No of outages
69 kV	112
115 kV	1586
230 kV	1761
287 kV	46
345 kV	70
500 kV	1559

Table 3.3 Outage Statistics for cause "Foreign Trouble"

Statistic (duration in minutes)	69 kV	115 kV	230 kV	287 kV	345 kV	500 kV
Mean	680.02	885.95	749.2	140.14	88.54	2602.12
Standard Deviation	2752.15	4866.87	5734.33	133.019	118.06	19077.1
Median	50	72.5	44	110	16	81

Table 3.4 Outage Statistics for cause "Unknown"

Statistic (duration in minutes)	69 kV	115 kV	230 kV	287 kV	345 kV	500 kV
Mean	255.3	491.4	137.92	71.5	193	47.39
Standard Deviation	626.49	2012.6	381.83	124.05	534.98	176.4
Median	121	16	6	4.5	4	3

Unknown, Lightning, Wind, Trees and Weather are noted as causes with more number of outages. Hence, the statistics are computed for these causes based on their voltage levels. Similar calculations can be done for other less frequent causes.

From tables 3.3, 3.4, 3.5, 3.6, 3.7, 3.8 it is evident that for different weather causes and operating voltage levels the mean, median and standard deviation of transmission line outages varied significantly. The distribution of duration of outages is skewed and this influences the variability of mean restoration times.

Table 3.5 Outage Statistics for cause "Lightning"

Statistic (duration in minutes)	69 kV	115 kV	230 kV	287 kV	345 kV	500 kV
Mean	143.64	259.976	125.615	123.25	169.22	66.29
Standard Deviation	149.59	916.065	629.215	333.147	402.11	197.13
Median	131	4	3	4	2	2.5

Table 3.6 Outage Statistics for cause "Wind"

Statistic (duration in minutes)	69 kV	115 kV	230 kV	345 kV	500 kV
Mean	162	1057.18	902.38	111.28	1450.22
Standard Deviation	123.03	1913.91	2079.9	276.53	5647.89
Median	162	317.5	7	3	3

Table 3.7 Outage Statistics for cause "Tree blown"

Statistic (duration in minutes)	115 kV	230 kV	287 kV	345 kV	500 kV
Mean	924.2	907.7	752.6	701.33	854.33
Standard Deviation	1349.39	992.03	423.93	413.12	1476.1
Median	450.5	558.5	532	755	279

Table 3.8 Outage Statistics for cause "Weather"

Statistic (duration in minutes)	115 kV	230 kV	345 kV	500 kV
Mean	988.43	729.7	219	148.3
Standard Deviation	2184.76	2493.07	369.79	303.127
Median	146	3	6	4.5

3.1.3 Density of Outages

If more outages occur together, that could result in longer restoration times. This could be due to cascading of outages themselves or cascading of restoration as there could be limited crew to work; i.e, a jammed queue of outages. Figure 3.5 shows the plot for density (number of outages per day) of non-momentary automatic outages against the duration (restoration time). Correlation between the density and duration of outages is inconclusive.

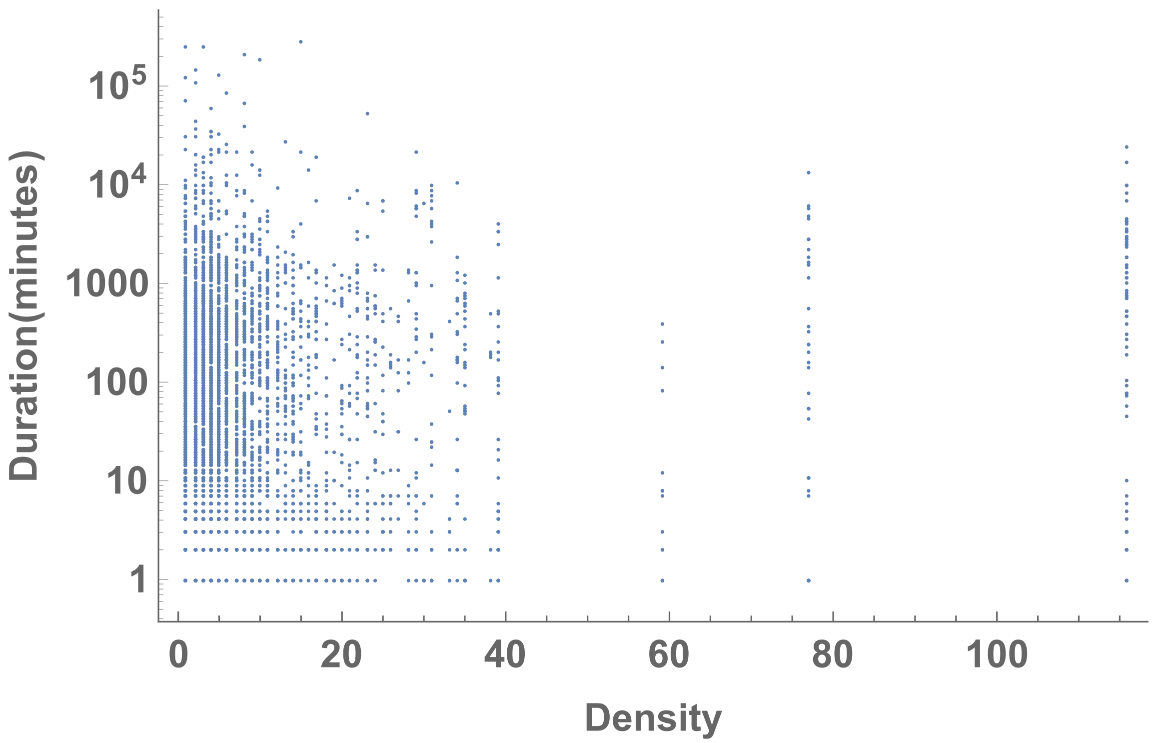

Figure 3.5 Density of non-momentary Automatic outages

3.1.4 Overlap of Outages

Definition for overlap in the analysis is as follows:

Let the start time for an outage be t_1 and end time is t_2; i.e, the interval $[t_1, t_2]$. Another outage $[t_3, t_4]$ is said to overlap with the first if:

1. $[t_3, t_4]$ occurs between the interval $[t_1, t_2]$

2. t_3 starts before t_1 and t_4 ends before t_2

3. t_3 starts after t_1 and t_4 ends after t_2

4. t_3 starts before t_1 and t_4 ends after t_2

Data was checked for all possible cases. Figure 3.6 shows the number of overlaps of non-momentary automatic outages against the duration (restoration time). Each outage has a time interval $[t_1, t_2]$, the number of intervals overlapping a given interval gives the number of overlaps.

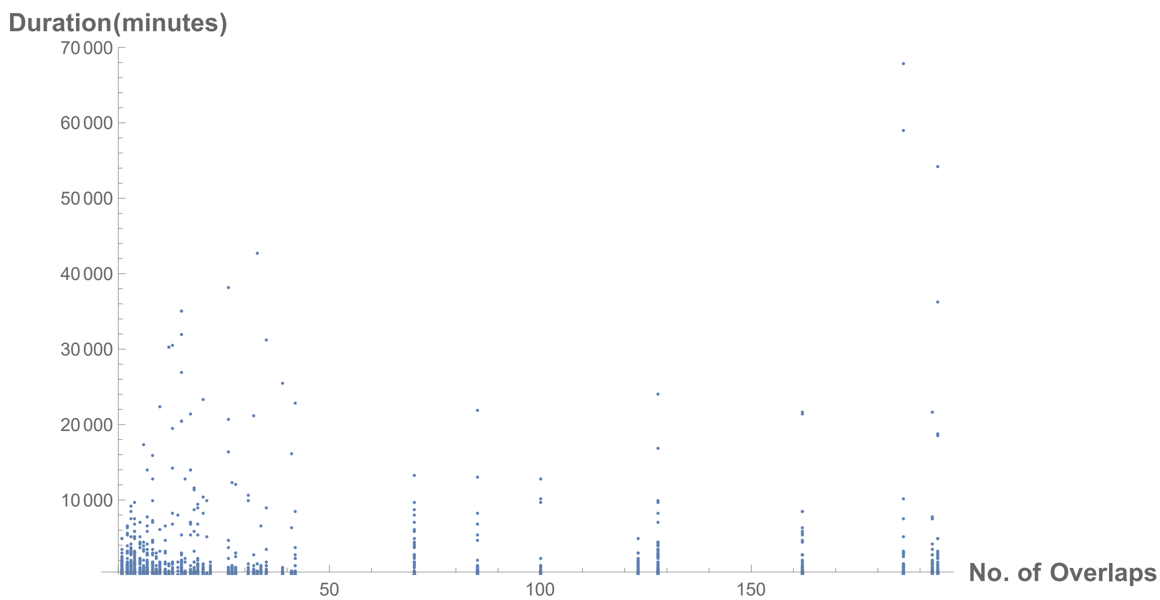

Figure 3.6 No. of Overlaps of non-momentary Automatic outages

To search for a firm evidence from the overlaps of non-momentary automatic outages plot (Figure: 3.5) the overlaps are separated by their causes. Figure 3.6 shows the plot for overlap of outages by five major causes. The interval [1, 100] minutes has more number of outages with causes "Lightning", "Tree blown" and "Wind". It is unclear if the outages with causes

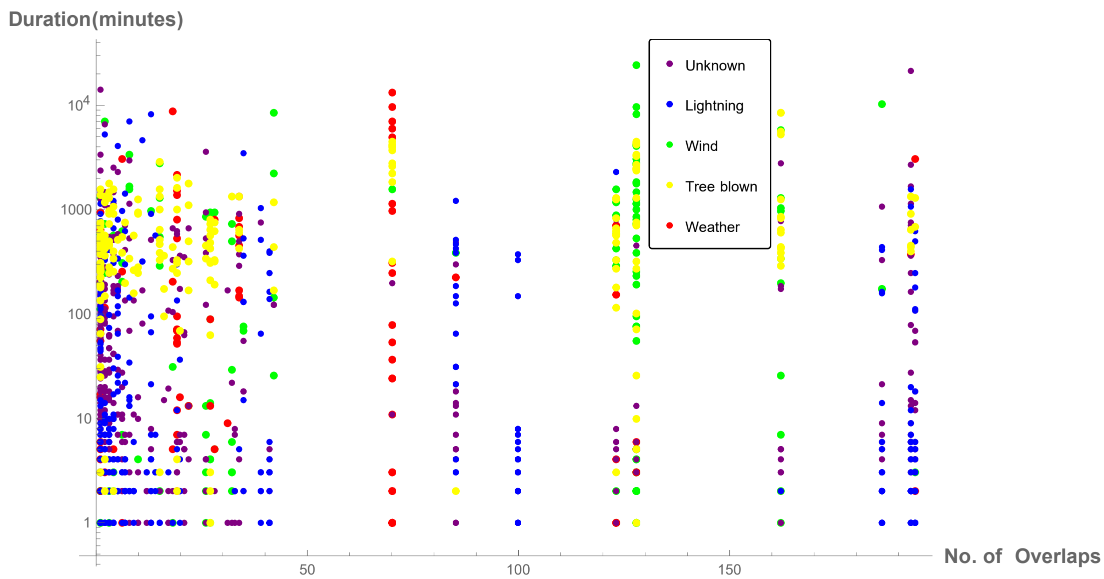

Figure 3.7 No. of Overlaps of non-momentary automatic outages by causes

in particular had more impact on the mean restoration time of the transmission lines. There is no clear evidence for overlaps. Correlation between the overlap of outages and duration of outages is inconclusive. The analysis in this chapter searched for, but provides no evidence of any influence on the duration of transmission line restoration in terms of operating voltages, cascading affects and the density of outages. The only firm conclusion is that the foreign interference significantly affected the restoration times.

CHAPTER 4. HEAVY-TAILED DISTRIBUTIONS

4.1 Definitions and Properties

Heavy tailed distributions, such as Power law probability distributions are ubiquitous. They commonly appear in diverse areas such as finance, biology, computer science and social sciences etc. Distributions of earthquakes, people's income, hurricanes, moon craters etc are examples of heavy tailed distributions. Unlike Normal or Exponential distributions (which have tails decaying faster), heavy tailed distributions converge relatively slower. When the probability of measuring a particular value of some quantity varies inversely as a power of that value, the quantity is said to follow a power law, also known variously as Zipf's law or the Pareto distribution (17).

Mathematically, power law is probability distribution of the form,

$$P(x) \propto x^{-\alpha} \tag{4.1}$$

More precisely, let X be a non-negative random variable, $F(x)$ is the cumulative density function $F(x) = P[X \leq x]$ and its complement $\bar{F}(x) = 1 - F(x) = P[X > x]$. Then,

$$\bar{F}(x) \sim Ax^{-\alpha} \qquad 0 < \alpha \leq 2 \tag{4.2}$$

where A is a positive constant,and α is the power law exponent. The complementary cumulative distribution function of the restoration times of utility data is shown in Figure 3.2. It is the probability that duration of an outage exceeds the given size.

Power law distributions for non negative random variables can be categorized into two types based on the data: Discrete, with strictly discrete values which are normally non-negative integers, and continuous, with real numbers greater than or equal to zero. The probability density function [PDF] $p(x)$ of a continuous power law distribution is expressed as $p(x)Ax^{-\alpha}$ where X is the observed value and A is a constant. On the other hand for discrete values the PDF is expressed as $p(x) = Pr(x = X) = Ax^{-\alpha}$ and there must be a minimum value of x (x_{min}) on the power law behavior. The following Figure 4.2 shows typically used statistical distributions for continuous and discrete cases (18).

	name	distribution $p(x) = Cf(x)$	
		$f(x)$	C
continuous	power law	$x^{-\alpha}$	$(\alpha - 1)x_{\min}^{\alpha-1}$
	power law with cutoff	$x^{-\alpha}e^{-\lambda x}$	$\frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha,\lambda x_{\min})}$
	exponential	$e^{-\lambda x}$	$\lambda e^{\lambda x_{\min}}$
	stretched exponential	$x^{\beta-1}e^{-\lambda x^{\beta}}$	$\beta\lambda e^{\lambda x_{\min}^{\beta}}$
	log-normal	$\frac{1}{x}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right]$	$\sqrt{\frac{2}{\pi\sigma^2}}\left[\operatorname{erfc}\left(\frac{\ln x_{\min}-\mu}{\sqrt{2}\sigma}\right)\right]^{-1}$
discrete	power law	$x^{-\alpha}$	$1/\zeta(\alpha, x_{\min})$
	Yule distribution	$\frac{\Gamma(x)}{\Gamma(x+\alpha)}$	$(\alpha - 1)\frac{\Gamma(x_{\min}+\alpha-1)}{\Gamma(x_{\min})}$
	exponential	$e^{-\lambda x}$	$(1 - e^{-\lambda})\,e^{\lambda x_{\min}}$
	Poisson	$\mu^x/x!$	$\left[e^{\mu} - \sum_{k=0}^{x_{\min}-1}\frac{\mu^k}{k!}\right]^{-1}$

Figure 4.1 Power Law Distribution and other Statistical Distributions (18)

Several other distributions are sometimes referred to as power law distributions and are used in the reliability models. Pareto distribution, Cauchy distribution, Log-normal distribution and Weibull distribution are some examples of heavy tailed distributions.

4.1.1 Analyzing Power laws

A basic tool to study the empirical distributions which exhibit power law behaviors is an empirical complementary cumulative distribution function. It gives a rough estimate of the scaling parameter α and the lower bound; i.e, x_{min}. A naive way to determine whether a dataset follows heavy tailed distribution is: if X is a random variable whose CCDF is a function of form $f(x) = Ax^{-\alpha}$, take logarithm on both sides of this equation, log$f(x)$ = log A - α log$x$. If log$f(x)$ is plotted as a function of log$x$, then a straight line should appear with $-\alpha$ as slope and log A as y- intercept. That implies, the log-log plot of the complementary cumulative distribution function, $Pr[X \geq x]$ has asymptotic behavior and will be a straight line. In addition, this method gives an estimate of the scaling parameter α. Accurate ways of determining tail exponent are described in (19). The following steps give a broad outline for the analysis of power law data:

1. Find the rough estimates of the parameters α and x_{min} for a given sample size. Verify the values using methods mentioned in (19), maximum likelihood estimator or the Hills estimator.
2. Calculate Goodness-of-Fit statistical test between the data and power law. The power law is plausible hypothesis for the data if the probability p-value which is returned by the statistical test is greater than 0.1.
3. Validate the power law using likelihood ratio test by comparing it with alternate hypotheses. The higher the likelihood, the better the fit.

Heavy tailed distributions behave differently than other distributions; i.e, the probability of occurrence of large observations when sampling a random variable is not negligible. One way to deal with this is to eliminate the large observations, however then the statistical estimates may ignore some of the most impactful data. When the large observations are included, this naive method may have some errors associated with it, though it has been heavily used in the literature (18). The new fits obtained by using regression methods usually do not follow basic requirements of probability distribution, such as normalization(18).

4.1.2 Characteristics of Heavy tail Distributions

Large data sets can be analyzed either by descriptive statistics such as graphs and scattered plots or by fitting metrics to power law. The two main attributes for analyzing any data distribution are mean and standard deviation estimates. The question is how the mean and standard deviation can be estimated.

Estimation of mean can be tricky for heavy tailed distributions. The mean value x in a power law distribution is given by (17);

$$\langle x \rangle = \int_{xmin}^{\infty} xp(x)dx = C\int_{xmmin}^{\infty} X^{-\alpha+1}dx = \frac{C}{2-\alpha}[x^{-\alpha+2}]_{xmin}^{\infty} \quad (4.3)$$

The distribution will have not have finite mean if the value of $\alpha \leq 2$ and will have finite mean otherwise. This means that, as the data size increases the estimates of mean will also increase. Further, if the mean square is calculated using(17): $\langle x^2 \rangle = \frac{C}{3-\alpha}[x^{-\alpha+3}]_{xmin}^{\infty}$ and this diverges if $\alpha \leq 3$ and the standard deviation of heavy tailed distribution is infinite. The mean and standard deviation calculated directly from the dataset may not be well behaved. Estimation of confidence interval of the mean of heavy tailed distribution is discussed in the later sections.

4.2 Processing Heavy tailed Restoration Times

Earlier sections of this chapter outline some properties and characteristics of heavy tails. The goal of this section is to analyze the non-exponential transmission line restoration times observed in the data. As mentioned in the literature, common distributions which are employed to model down times are Gamma, Normal, Lognormal, and Weibull (20). Five typically used distributions are discussed below(12).

- Lognormal Distribution: Lognormal distribution assumes that the natural logarithm of random variable is normally distributed with a mean μ and standard deviation σ. The density function is given as:

  $$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} exp[-\frac{ln(t-\mu)^2}{2\sigma^2}] \quad (4.4)$$

  where $t > 0$ and μ and σ^2 are not the mean and variance of the random variable t but of its natural logarithm. The mean, E and variance, V of lognormal distribution are given,

  $$ET = exp(\mu + \frac{\sigma^2}{2}) \quad (4.5)$$

  $$V = exp(2\mu + \sigma^2)[exp(\sigma^2) - 1] \quad (4.6)$$

- Weibull Distribution: Weibull distribution is a two parameter distribution with α being the scale parameter and β being the shape parameter. the density function is given by:

$$f(t) = \frac{\beta t^{\beta} - 1}{\alpha^{\beta}} exp[-(\frac{t}{\alpha})^{\beta}] \tag{4.7}$$

The cumulative distribution function is

$$F(t) = 1 - exp[-(\frac{t}{\alpha})^{\beta}] \tag{4.8}$$

where $t > 0$, $\alpha > 0, \beta > 0$

- Gamma Distribution This is also a two parameter distribution like the Weibull distribution and has similar properties. Scaling parameter α and shape parameter β can be adjusted to fit the data. The density function is given as:

$$f(t) = \frac{t^{\beta} - 1}{\alpha^{\beta}\Gamma(\beta)} exp[-\frac{t}{\alpha}] \tag{4.9}$$

- Exponential Distribution The density function :

$$f(t) = \lambda exp(-\lambda t) \tag{4.10}$$

and the cumulative distribution function is:

$$F(t) = 1 - exp(-\lambda t) \tag{4.11}$$

The mean of exponential distribution is $1/\lambda$ and the variance is $1/\lambda^2$.

- Pareto Distribution This is the simplest heavy tailed distribution, which is power law over its entire range. Its probability density function is:

$$f(t) = \alpha k^{\alpha} t^{-\alpha-1} \tag{4.12}$$

and the cumulative density function is:

$$F(t) = P[X \leq t] = 1 - (\frac{k}{t})^{\alpha} \tag{4.13}$$

where k is the smallest positive value of the random variable and α is the shape parameter. If X has power law distribution it has infinite variance for $0 < \alpha \leq 2$, and infinite mean for $\alpha \leq 1$.

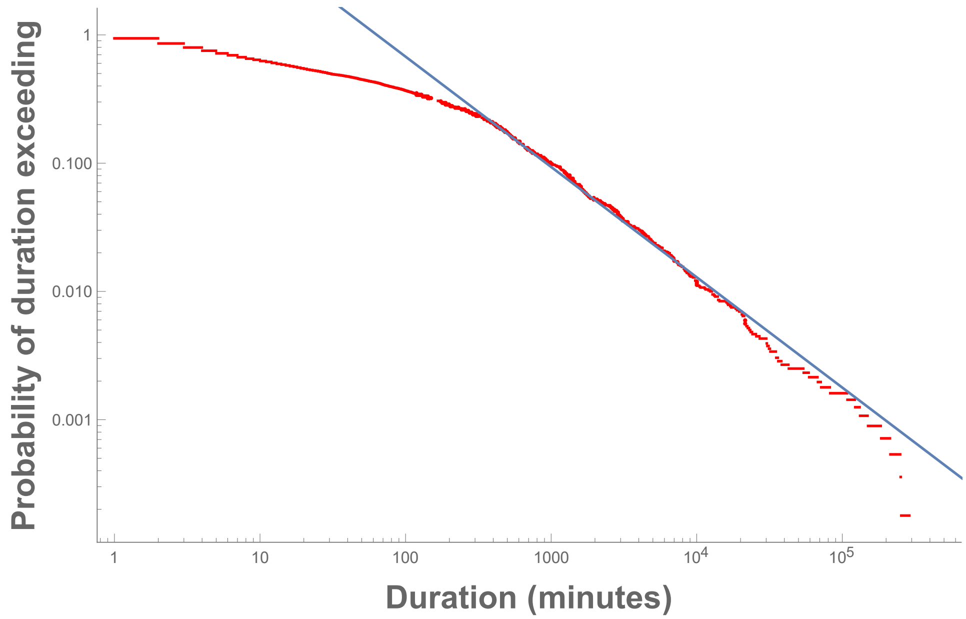

Figure 4.2 Survivor function of outage duration with slope of tail indicated in utility data

4.2.1 Heavy tail in the Line Outage Data

The distribution of transmission line restoration times shows a heavy tail due to the approximately linear behavior of the longer outages on the log-log plot as shown in figure 4.2. The slope of the heavy tailed linear region is -0.84, which implies that probability of the outage duration exceeding time t varies as $t^{-0.84}$. (The corresponding tail of the probability distribution function of outage duration varies as $t^{-1.84}$). The heavy tail implies that long restoration times are rare and highly variable, and, in contrast to distributions with exponentially decaying tails, occur routinely. Another way to see the effect of the power law decay of the tail of the probability distribution is that the repair intensity is proportional to $1/t$. This deterioration in the repair intensity over time shows the impact of some very long restoration times.

Except for a certain range, the curve is approximately linear with a slope of 0.84 and this is regarded as a heavy tail; i.e, tail is exponentially unbounded. It means that long restoration times although rare, but occur routinely. For heavy tails the tail exponent is $\gamma > 1$, and it indicates that if extrapolated indefinitely there would be a definite mean but variance is undefined. In practice, with this finite distribution the mean estimates behave erratically.

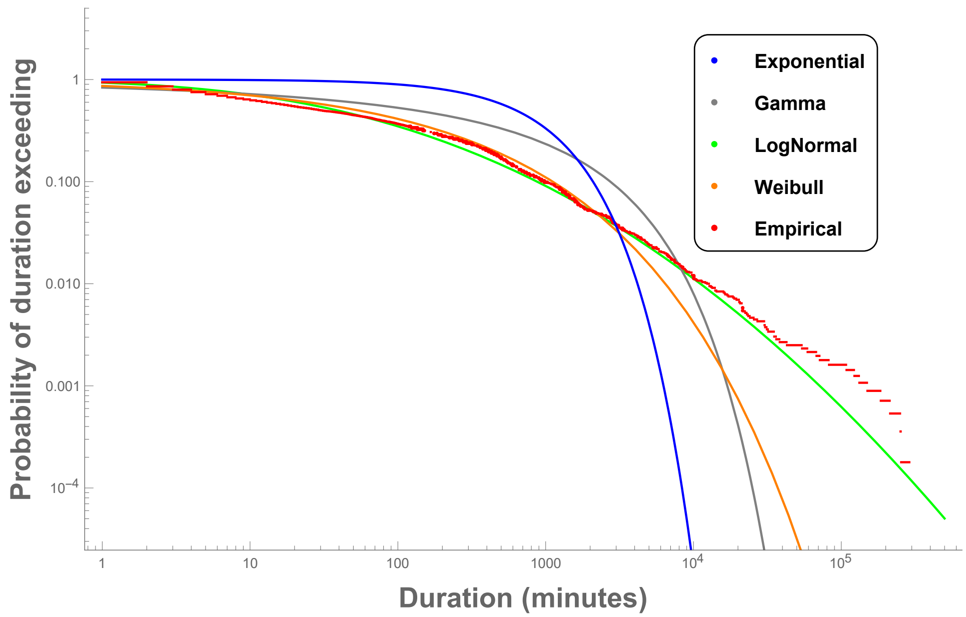

Figure 4.3 Survivor function of outage duration in utility data and some fitted ideal distributions

4.2.2 Comparison with Alternative Distributions

Section 4.1.1 explained a fundamental method to check for heavy-tailedness in the data, which might not be true in all cases. It is plausible that the data can be a good fit for other distributions as stated in the literature. Figure 4.3 shows several other fits for the data. Figure 4.3 shows that the tail (Empirical Distribution) is a bit heavier than log-normal with tail index $\gamma = 1.84$; i.e, power law is quite significant. The next step is to determine on what significance level does lognormal distribution assumption fit the data set. A goodness-of-fit test provides a solution and is discussed in the next section.

4.2.3 Goodness-of-fit Tests

A goodness-of-fit test is a standard approach which quantifies whether the dataset belongs to a certain distribution. It generates a probability value, p-value which expresses the plausibility of a hypothesis. A large p-value, closer to 1 indicates that the empirical data and the model is subject to statistical variations, and small p-value indicates that the distribution under question is not a good fit. It is intuitively obvious from figure 4.4 that the data could possibly be a lognormal distribution. Statistical test showed that the data cannot be confirmed to be lognormal. The p-value ($p < 0.1$) obtained from Kolmogorov Smirnov and Pearson ChiSquare tests is below the threshold value.

In addition, there may be infinite number of distributions, and it is not ideal to compare the power law fit with every distribution available. In data processing, finding a class of distributions to fit the data is initial step to narrow down the process. It is important to note that the data type whether continuous and discrete has to be treated separately to compare the p-values with competing distributions. In situation where two or more distributions pass the goodness-of fit test then one has to find a better fit comparing one against other. The Likelihood ratio tests determine the higher likelihood of the two competing distributions. The current data is only tested for Lognormal distribution.

CHAPTER 5. STATISTICS OF HEAVY-TAILED DISTRIBUTIONS

Most of the historical power outage data sets comprise the outage information related to causes, type of outages, duration of outages etc. What is not mentioned usually is the repair rate, failure rate, mean time to failure and mean time to repair. Each outage model is characterized by these parameters and needs to be calculated from the data. These core parameters play a vital role in performing the reliability and vulnerability assessment studies. In this chapter estimation of the basic reliability parameters such as the mean and confidence intervals for the heavy tailed distribution of transmission line restoration times is presented.

5.1 Estimation of Mean and Confidence interval of the Mean

The estimation of mean duration from the data is described in this section. Let $X_1, X_2....X_n$ be n independent and identically distributed random variables of a distribution. Estimation of mean is given by the sum of samples divided by the number of samples:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5.1}$$

Due to the presence of extremely large samples in the data, the mean and variance values fluctuate excessively. One way to deal with large samples is to exclude them and process the data. In fact, extreme events although rare, occur persistently, hence it is not ideal to exclude the extreme events from the data. Figure 5.1 shows the non-uniform variation in the mean in the raw data, for each year individually. To get a better estimate of the mean, it is necessary to have sufficiently large number of samples. Figure 5.2 shows how the mean varies as the number of samples taken from the data increases. Note the erratic form of convergence that is due to frequent small values of restoration time and occasional large restoration times that is inherent in the data. For comparison, Figure 5.2 also shows the convergence of an exponential

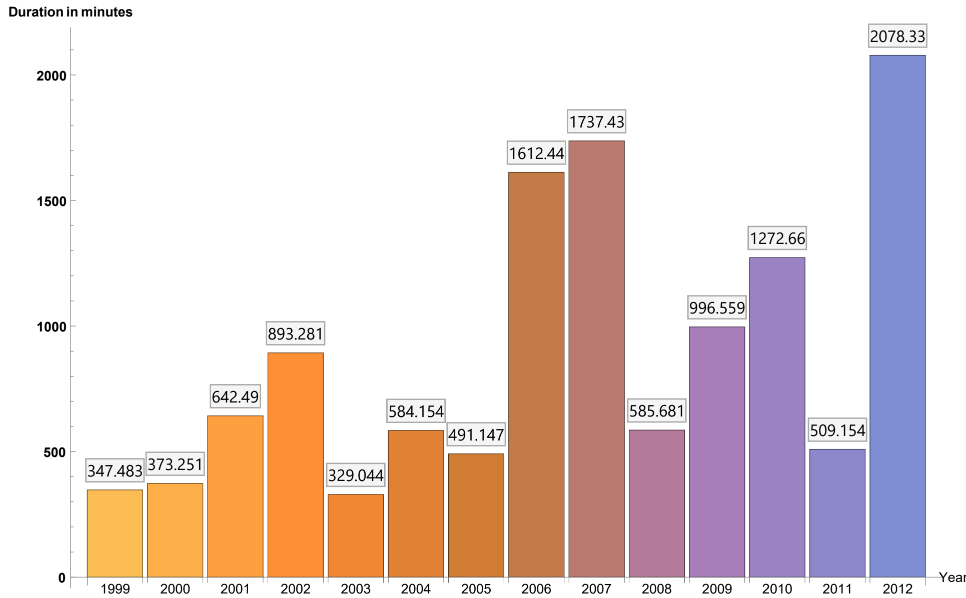

Figure 5.1 Annual mean duration of outages in utility data

distribution of restoration times with the same sample mean. It can be seen that it reaches steady state in a short time.

Over 14 years, 5594 non-momentary automatic outages are observed and with the average rate of 430 outages/year. The sample mean is $\bar{X}$= 907 minutes and the sample standard deviation is S= 8514 minutes for 14 years. The bootstrap approximation of the distribution of studentized sample mean provides further inference about the sample mean. In order to determine the confidence interval for the sample mean some studies recommend certain bootstrap sampling techniques such as the m out n bootstrap method, sub-sampling bootstrap method and empirical likelihood based confidence intervals which provide a good estimate (19). For heavy-tailed distributions, resampling with the full sample size may not yield better estimates (21). These techniques are used to approximate the Studentized mean constructed for original sample size. Among the established non parametric bootstrap methods which estimate the confidence interval, (19) suggests that the confidence intervals generated by m out of n bootstrap methods produces better confidence interval for $\bar{X}$ than other methods.

The procedure for m out of n bootstrap confidence interval is as follows (21): Let $X = X_1, X_2....X_n$ be n independent identically distributed random variables sampled from a dis-

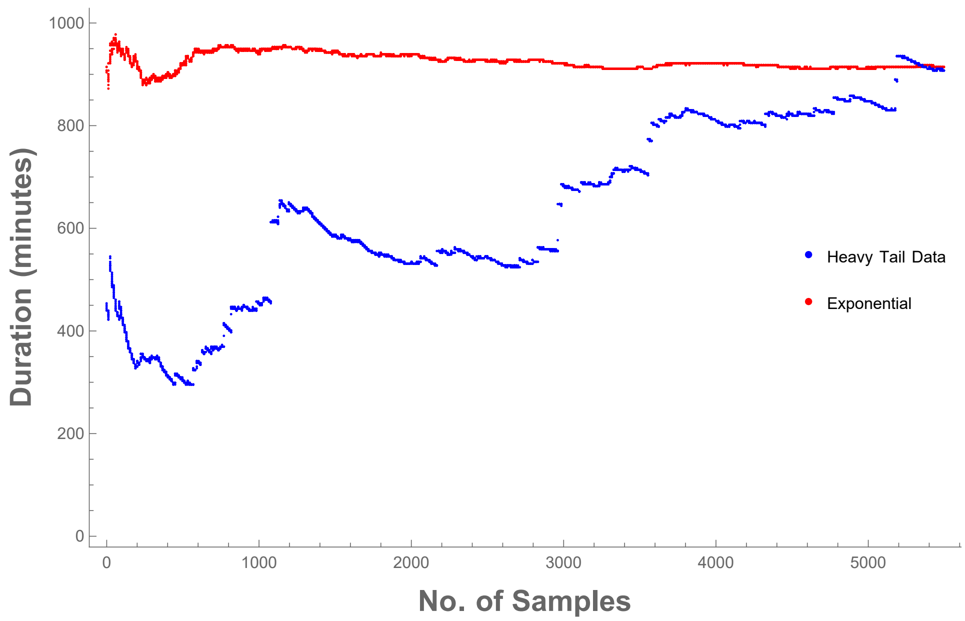

Figure 5.2 Mean Convergence for Exponential and Heavy tail Forms

tribution with a finite mean μ.

1. Calculate the sample mean $\bar{X} = n^{-1}\sum_{j=1}^{n} X_j$ and the sample variance $S^2 = n^{-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$.

2. Draw m samples with replacement from X, where $m < n$ and let it be $X_1^*, X_2^*, ...X_n^*$.

3. Calculate bootstrap mean $\bar{X}_m^* = m^{-1}\sum_{j=1}^{m} X_j^*$, variance $S_m^{*2} = m^{-1}\sum_{j=1}^{m}(X_j^* - \bar{X}_m^*)^2$, and Studentized mean $T_m^* = m^{\frac{1}{2}}(\bar{X}_m^* - \bar{X})/S_m^*$.

4. Repeat this for 100 000 times to obtain an empirical distribution for T_m^*.

5. Then $I_{95\%}$ is a nominal 95% - level confidence interval for μ.

$\hat{x}_{95\%} = sup\{x : P[|T_m^*| \leq x] \leq 0.95\}$ and $I_{95\%} = (\bar{X} - \hat{x}_{95\%}S/\sqrt{n}, \bar{X} + \hat{x}_{95\%}S/\sqrt{n})$.

5.1.1 Variability of means

From the approach mentioned in section 5.1, the 95% confidence interval for mean over 14 years is [191,1625]. This shows a substantial variation in the mean for annual estimates. Since the unavailability is small, the unavailability is very nearly proportional to the mean outage duration. The fraction of time (expressed in %)a typical transmission line is available on average an year is 99.87%. The average annual rate of a line going down is 0.7 times.

Since the availability of line and the mean restoration time are related, variability in the mean restoration time leads to the variability of the availability of a line. The confidence interval of annual availability of the system is [0.999, 0.997]. In terms of unavailability the confidence interval is [0.00026, 0.0022].

CHAPTER 6. CONCLUSION

Despite the large amount of ongoing research in strengthening the power grid resiliency, there has not been much published about the observed statistics of transmission line outages. This thesis analyzes the line outage data to determine the form and implications of the distribution of transmission line restoration times. The restoration time data showed undesirable variabilities including rare but persistent long restoration times. Excluding the extreme events from the study will not yield reliable statistics for analysis. The analysis showed that the longer restoration times were not influenced by cascading affect and the transmission line operating voltage levels. The only cause which had impact on the mean restoration time is the 'Foreign Trouble'.

This thesis presented the statistical variations of longer restoration times for some transmission line outage data and showed that the distribution of restoration times had a heavy tail. By including the outages with longer duration in computing the metrics, it is evident that the mean values are highly variable. Also, even if the mean can be accurately estimated and longer observation times, the mean and the unavailability are no longer representative values of the distribution that they summarize.

As more automatically processed and detailed data sets are becoming available to utilities, it seems appropriate to re-evaluate observed repair statistics. While heavy tails of log-normal form have appeared before, this data shows longer repair times with a power law region that is slightly heavier than log-normal. It is important to realize the complications involved in obtaining basic statistics from when heavy-tailed distributions are involved.

In particular, it reaffirms that processing the data plays a vital role to strengthen the underpinnings of reliability assessment studies. It is hoped that by providing insights into the work on mean time to repair process of outages, these conclusions would help researchers to develop robust models to understand post outage/blackout behaviors in the power system. Although, the results pertain to only one utility, all utilities in NERC collect TADS data that contains the needed outage and restoration times alongside many other utilities worldwide collect similar data. Therefore, similar calculations can be done by many utilities.

These conclusions are unique to transmission systems, for future study it will be interesting to study the restoration times of the distribution systems. The distribution systems have lower voltage levels and different geographic layout, so they have to be treated separately.

I gratefully thank Bonneville Power Administration for making publicly available the outage data that made this thesis possible. The analysis and any conclusions are strictly those of the author and not of Bonneville Power Administration.

BIBLIOGRAPHY

[1] D. Zhu. Electric distribution reliability analysis considering time-varying load, weather conditions and reconfiguration with distributed generation(2007), Ph.D thesis, *Department of Electrical Engineering, Virginia Tech, VA*

[2] Chow, Mo-yuen, Leroy S. Taylor, and Mo-Suk Chow(1996). Time of outage restoration analysis in distribution systems. *IEEE Transactions on Power Delivery 11*(3),1652-1658.

[3] Dobson, Ian, Carreras, B. A., Newman, D. E., & Reynolds-Barredo, J. M. (2016). Obtaining statistics of cascading line outages spreading in an electric transmission network from standard utility data. *IEEE Transactions on Power Systems 31*(6), 4831-4841.

[4] Carreras, B. A., Newman, D. E., & Dobson, I. (2016). North American Blackout Time Series Statistics and Implications for Blackout Risk. *IEEE Transactions on Power Systems 31*(6), 4406-4414.

[5] C.C. Fong (1989). Reliability evaluation of transmission and distribution configurations with duration-dependent effects, *Proceedings Part C Generation, Transmission and Distribution, 136*(2), 64-67.

[6] C. Zapata, S. Silva, H. Gonzalez, O. Burbano, and J. Hernandez (2008). Modeling the repair process of a power distribution system, *Transmission and Distribution Conference and Exposition: Latin America*,1-7.

[7] R. Billinton, E. Wojczynski (1985). Effect of distribution system reliability index distribution upon interruption cost/reliability worth estimates *IEEE Transactions on Power Apparatus and Systems 104*(11),3229-3235.

[8] C. Dichirico, C. Singh(1988). Reliability analysis of transmission lines with common mode failures when repair times are arbitrarily distributed *IEEE Transactions on Power Apparatus and Systems 3*(3),1012-1019.

[9] S.J. Argent, S.G. Ryan (1985). An analysis of supergrid transmission fault data in a specific area of England and Wales, *Reliability Engineering*, *12*(4), 205-216.

[10] S. Asgarpoor, M.J. Mathine (1997). Reliability evaluation of distribution systems with non-exponential down times, *IEEE Transactions on Power Systems*, *12*(2), 579-584.

[11] R. Billinton, P. Wang (1999). Teaching distribution system reliability evaluation using Monte Carlo simulation, *IEEE Transactions on Power Systems*, *14*(2), 97-403.

[12] R. Billinton, W. Li (1994). Reliability Assessment of Electric Power Systems Using Monte Carlo Methods, New York : Plenum Press,151-160.

[13] A.W. Schneider, G.L. Landgren (1982). Discussion of Non-Markovian models for common mode failures in transmission systems, *IEEE Transactions on Power Apparatus and Systems*, *101*(6),1545-1550.

[14] Bonneville Power Administration Transmission Services Operations & Reliability website http://transmission.bpa.gov/Business/Operations/Outages

[15] D. O. Koval, A. A. Chowdhury(2005). An investigation into extreme weather caused transmission line unavailability, *Proceedings IEEE Power Engineering Society General Meeting*, 2425-2428.

[16] NERC:Transmission Availability Data System (TADS) Definitions (2010). www.nerc.com

[17] Newman, Mark EJ (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323-351.

[18] Aaron Clauset, Cosma Rohilla Shalizi and M.E. J. Newman (2009). Power-Law Distributions in Empirical Data. *Society for Industrial and Applied Mathematics Review*, *51*(4), 661-703.

[19] Ainura Tursunalieva and Param Silvapulle (2009). Estimation of confidence intervals for the mean of heavy tailed loss distributions: a comparative study using a simulation method *Citeseer*,

[20] B. Retterath, S.S. Venkata, A.A. Chowdhury(2005). Impact of time-varying failure rates on distribution reliability, *International Conference on Probabilistic Methods Applied to Power Systems*, *27*(9), 953-958.

[21] P. Hall, R. LePage(1996), On bootstrap estimation of the distribution of the studentized mean, *Annals of the Institute of Statistical Mathematics*, *48*(3), 403-421.