Resilience risk metrics calculated from utility data that address the huge variations in blackout cost

Ian Dobson

Arslan Ahmad

IOWA STATE UNIVERSITY

power group seminar November 2025

Funding from Iowa State EPRC, PSerc, NSF, EARNEST, and Argonne National Lab is gratefully acknowledged

Motivation and questions

- Extreme weather events are gradually increasing in frequency and size, and this continues to pose a risk of blackouts.
  This talk focuses on distribution; similar comments apply to transmission.
- To do resilience engineering and influence resilience investments and policy, we need to **quantify the risk to customers** of the larger blackouts.
- Risk combines probability and cost. As blackout cost increases, the probability goes down – **But how fast does probability go down**?
- Are large blackouts rare but expected (heavy tails), or are they outliers?

utility outage data → group outages into events → customer hours for each event → customer cost for each event → customer event cost probabilities

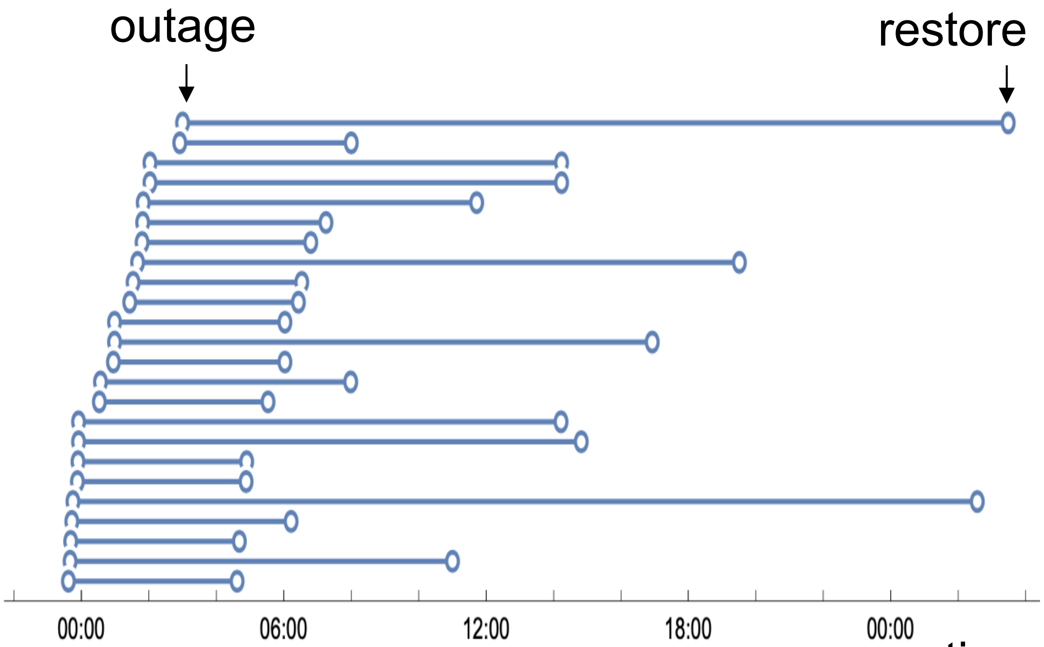
Detailed distribution system outage data from Outage Management Systems OMS

- OMS data include outage and restore times to the nearest minute, customers out, and other descriptions of each outage

- Utilities have this data and there is also some public data:
  ENW in UK, Massachusetts, Brazil, EAGLE-I; see last slide

Extracting Events from Utility Data

- Larger events are usually caused by extreme weather
- Events characterized by an overlapping accumulation of outages bunching up in time
- Grouping outages into events is the basis for resilience analysis driven by data

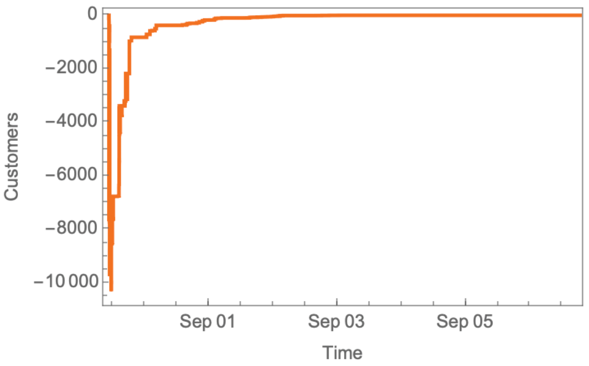


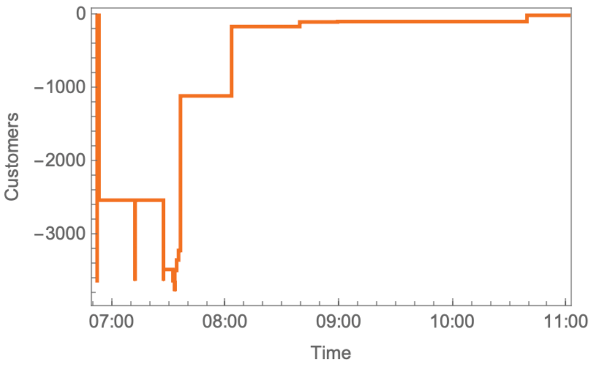
outage bunching and overlapping

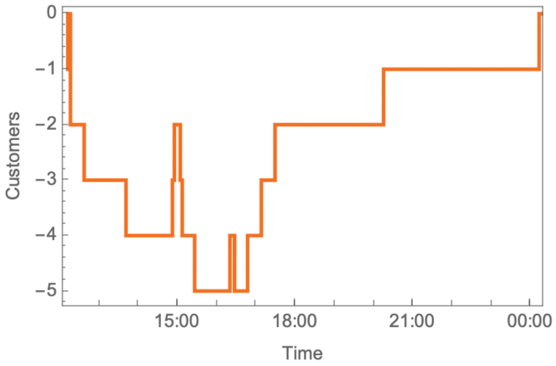
Key point is to have algorithms to automatically extract events from data based on outage timings
The larger events can be confirmed by finding the associated weather.
This gives events of all sizes and causes.

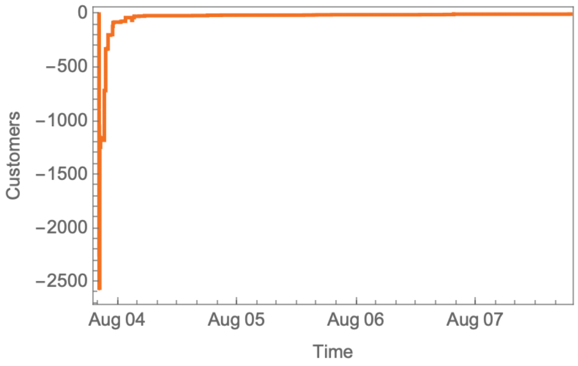
Distribution system event performance curves tracking customers out

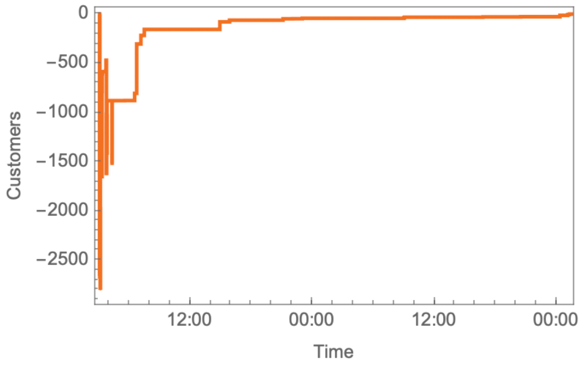
Performance curve = number of customers out in event at time t

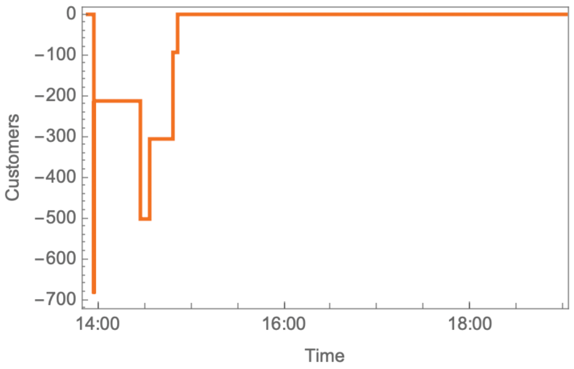






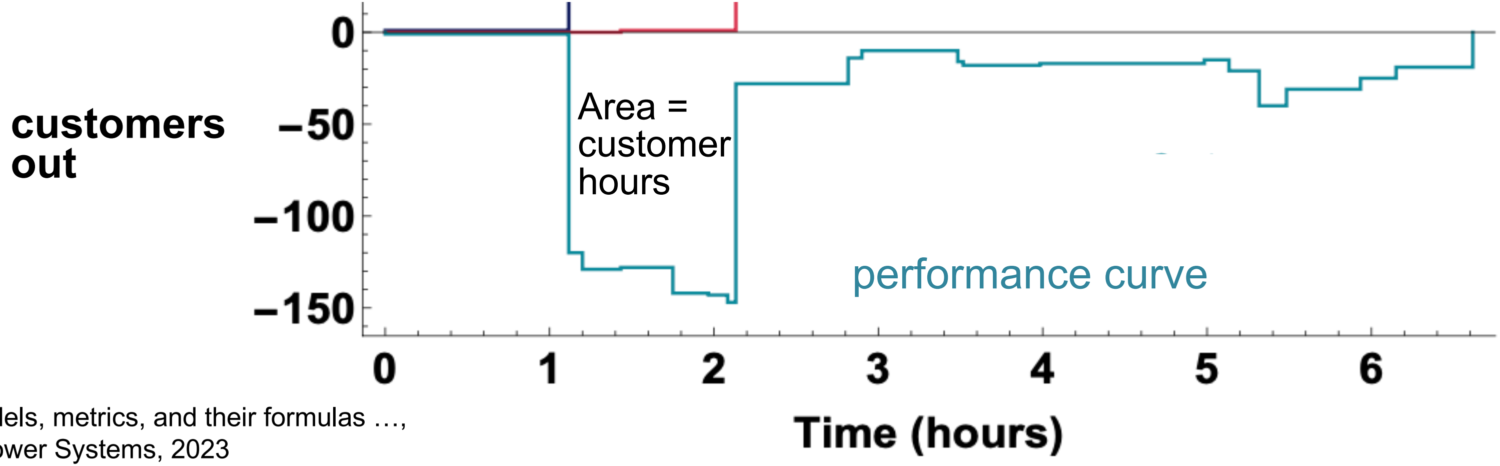






Customer cost of events

- Customer hours of an event = Area under the performance curve
  = Sum of [customers out x duration]* = what event would contribute to numerator of SAIDI
- **Customer cost of an event $\propto$ customer hours lost in event**
  - state of the art (but needs further advances)
  - proportionality constant combines the different effects of industrial, commercial, and residential loads
- Then get the empirical distribution of customer cost of all the events



*I Dobson, Models, metrics, and their formulas …,
IEEE Trans. Power Systems, 2023

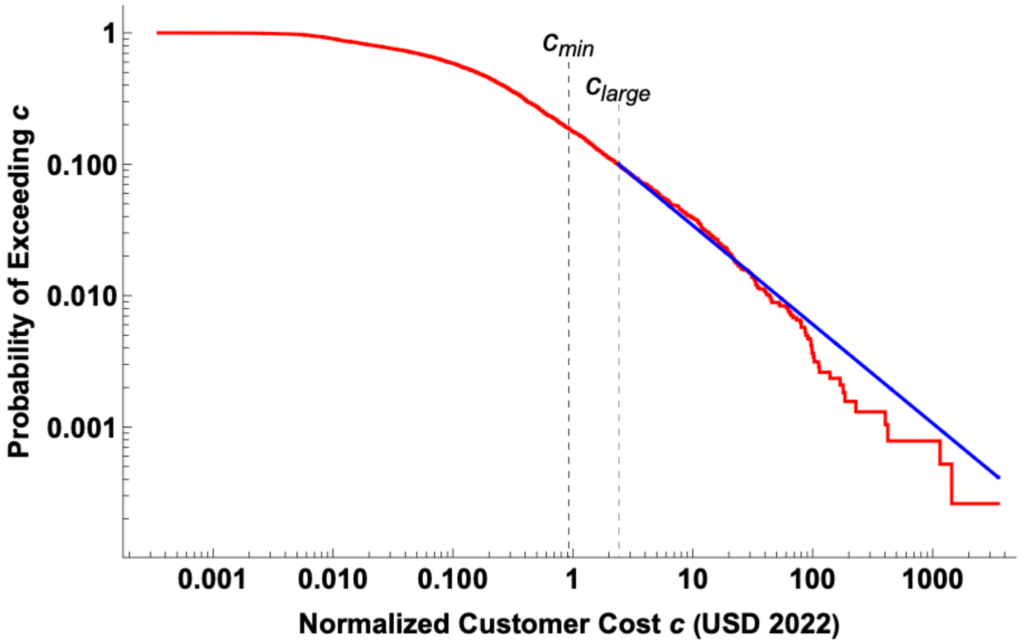
Customer risk described by distribution of event cost

Risk is given by Exceedance function P[event customer cost > c] (Kaplan 1981)

- Exceedance function has very heavy power law tail with
  $$P[\text{event cost} > c] \propto c^{-\alpha}$$
  α=0.75 is slope magnitude on log-log plot
  … probability goes down very slowly

- Very heavy tail implies
  - high risk of large blackout
  - no typical large blackout
  - mean of large blackouts is ill defined or impractical to estimate.

LOG-LOG PLOT OF EXCEEDANCE FUNCTION



(Exceedance function = CCDF = 1-CDF = Survival function)

estimate mean with
exponential decrease
in probability with cost

works with usual statistics

estimate mean with
power law decrease
in probability with cost

does not work for very heavy tail

solve it with logarithms …
logarithms convert power law to exponential

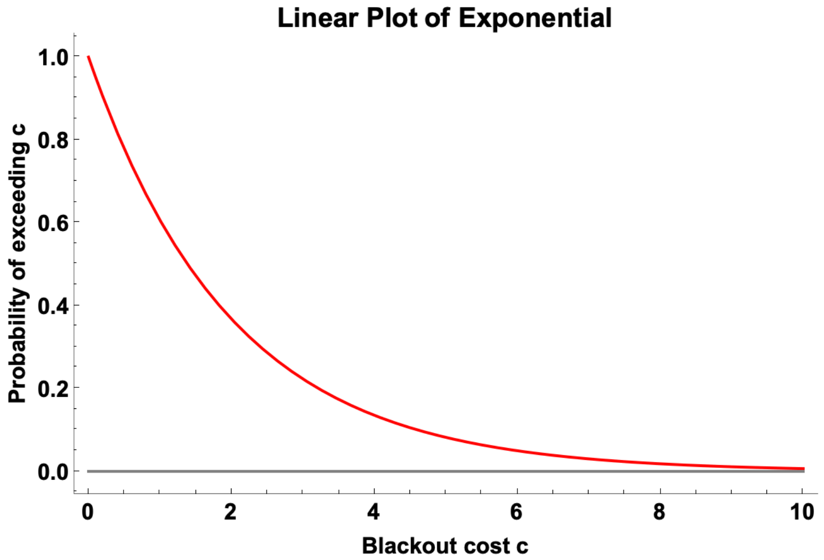
Blackout probability decreasing exponentially with cost

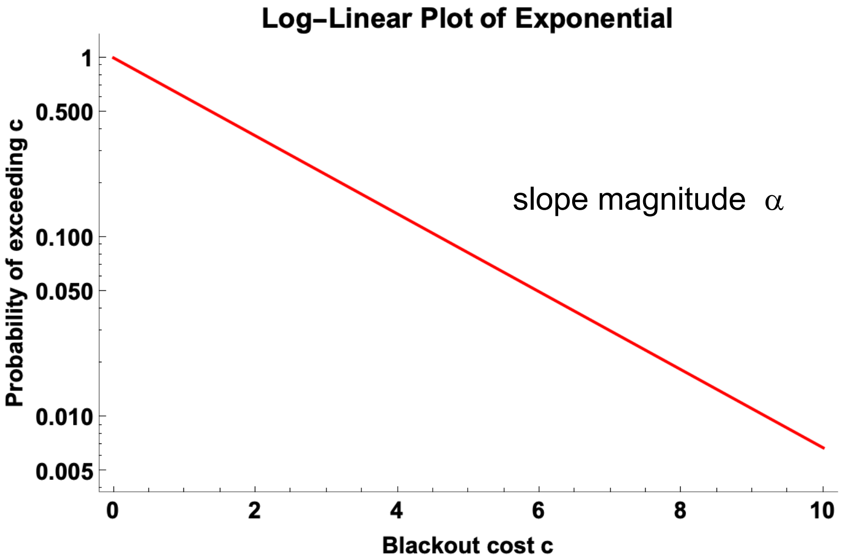
Probability of a blackout with cost X exceeding c = P[X>c] = $e^{-\alpha c}$

If you double the cost c, P[X>c] is squared: $e^{-\alpha 2c} = (e^{-\alpha c})^2$

$\frac{1}{2}$ $\frac{1}{4}$ $\frac{1}{16}$ $\frac{1}{256}$ $\frac{1}{65\,536}$ $\frac{1}{4\,294\,967\,296}$ $\frac{1}{18\,446\,744\,073\,709\,551\,616}$ $\frac{1}{340\,282\,366\,920\,938\,463\,463\,374\,607\,431\,768\,211\,456}$

Mean = $1/\alpha$



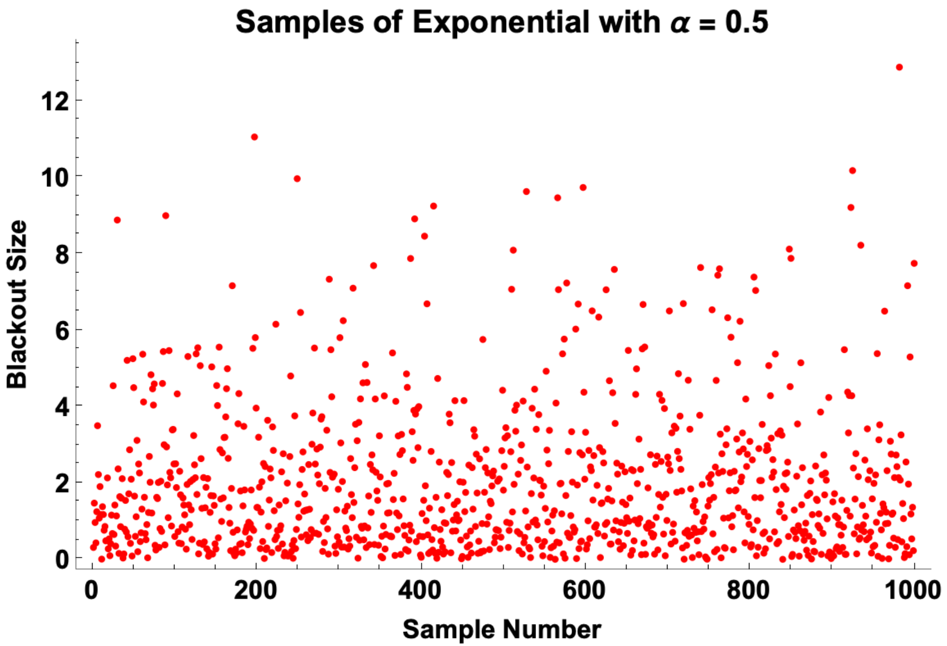


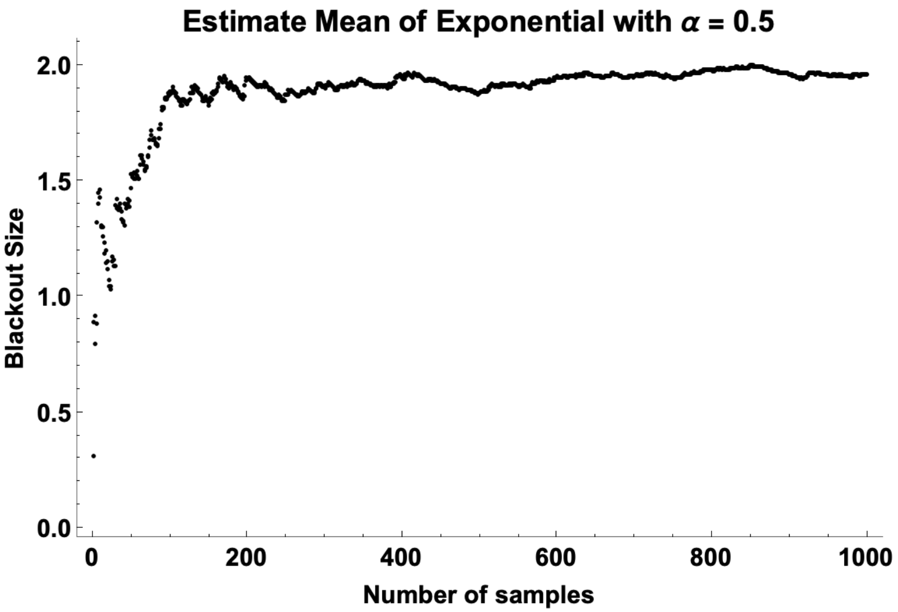
Samples from Exponential and estimating mean

Take 1000 samples from Exponential with $\alpha = 0.5$; mean = 2; mean is typical

Strong Law of Large Numbers: If and only if the mean is finite,

$\dfrac{\text{sum of samples}}{\text{number of samples}}$ tends to mean as number of samples increases

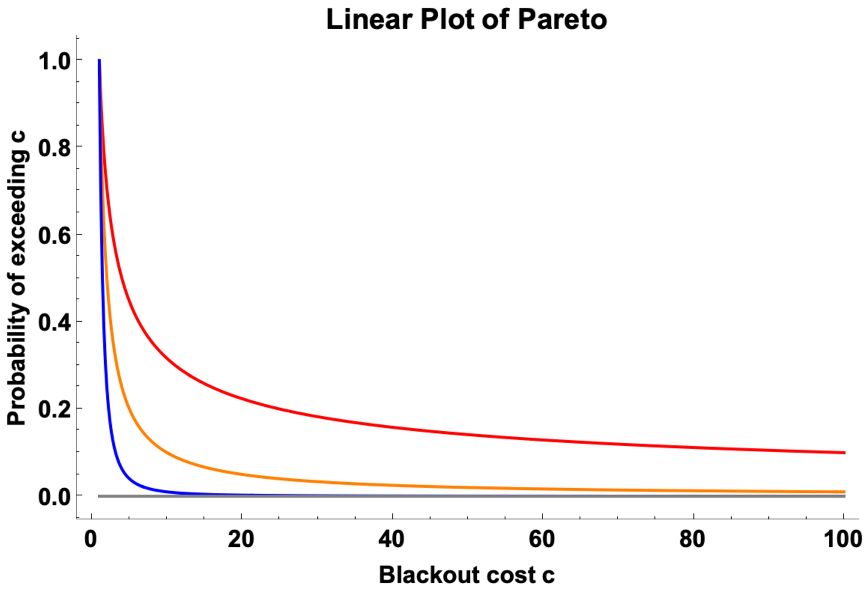


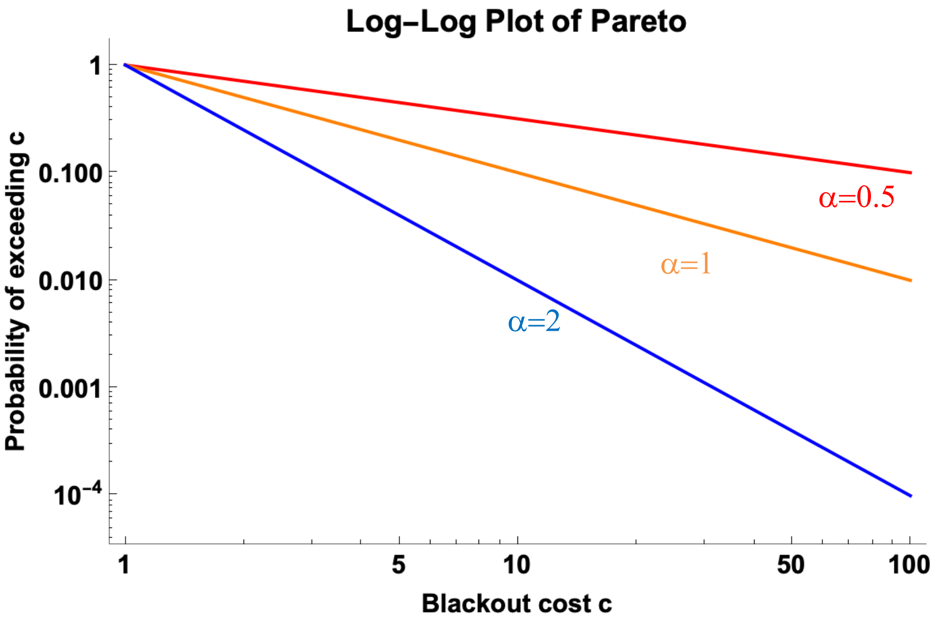


Blackout probability decreasing as a power law with cost

Probability of a blackout with cost X exceeding c = P[X>c] = $c^{-\alpha}$, c≥1

This is a Pareto distribution





α is slope magnitude on log-log plot

Blackout probability decreasing as a power law with cost c

Probability of a blackout with size exceeding c = $P[X>c] = c^{-\alpha}$, $c \geq 1$

This is a **Pareto distribution** with constant α

If you double the cost c, $P[X>c]$ is multiplied by $2^{-\alpha}$: $(2c)^{-\alpha} = 2^{-\alpha} c^{-\alpha}$

$$\alpha = 0.5 \qquad \frac{1}{2} \quad \frac{1}{2\sqrt{2}} \quad \frac{1}{4} \quad \frac{1}{4\sqrt{2}} \quad \frac{1}{8} \quad \frac{1}{8\sqrt{2}} \quad \frac{1}{16} \quad \frac{1}{16\sqrt{2}}$$

$$\alpha = 1 \qquad \frac{1}{2} \quad \frac{1}{4} \quad \frac{1}{8} \quad \frac{1}{16} \quad \frac{1}{32} \quad \frac{1}{64} \quad \frac{1}{128} \quad \frac{1}{256}$$

$$\alpha = 2 \qquad \frac{1}{2} \quad \frac{1}{8} \quad \frac{1}{32} \quad \frac{1}{128} \quad \frac{1}{512} \quad \frac{1}{2048} \quad \frac{1}{8192} \quad \frac{1}{32768}$$

WARNING FROM MATH about Pareto with slope magnitude α

$\alpha \leq 1$: Mean is infinite;
estimates of mean do not converge

$\alpha > 1$: Mean is $\frac{\alpha}{\alpha-1}$

estimates of mean do converge, but how fast?

```
{3.4, 1.4, 2., 2.9, 1., 1.3, 8.3, 1.4, 220., 18., 1.8, 1.1, 3.4, 1.3, 29., 4.2, 2.6, 36.,
 680., 19., 7.4, 1.1, 2.4, 160., 2.9, 1.2, 3.5, 12., 2.4, 57., 61., 4.2, 1.1, 5.3, 7.5,
 2.6, 2.6, 13., 1.1, 3.7, 2.4, 3., 3., 1.1, 19., 1.7, 19., 4., 1.1, 1.7, 2.5, 2.2, 1.6,
 3.2, 12., 3.6, 1.8, 3.3, 3.9, 10., 350., 14., 2.4, 27000., 2.8, 8.9, 1.1, 15., 1.,
 1.5, 36., 1.3, 1.1, 1.2, 22., 1.1, 2.6, 120., 3.5, 1.6, 1.3, 32., 3.6, 4.6, 23., 27.,
 120., 35., 3.6, 2.4, 1., 4.1, 1.4, 3.3, 6.7, 1.8, 40., 12., 5900., 1.2, 11., 750.,
 8.1, 2.2, 1.6, 20., 21., 560., 3.2, 36., 2.1, 2.3, 7.5, 3.4, 1., 4.7, 1., 220., 1.2,
 13., 3.8, 160., 38., 5., 1.4, 1.7, 1.4, 72., 1.8, 10., 1.2, 360., 11., 4.2, 19., 1.6,
 17., 4.8, 110., 10., 310., 5.3, 1.6, 1.9, 1.4, 25000., 4.4, 1.3, 4.5, 5.6, 12000.,
 2.8, 43., 1.8, 1.4, 3.7, 11., 12., 11., 3500., 140., 2.9, 2.1, 1., 460000., 6.5, 53.,
 8.2, 46., 1.1, 19., 94., 6.8, 4.1, 13., 2.6, 9.8, 130., 140., 1.9, 1.6, 1.9, 6900.,
 1.8, 2.2, 4.2, 43., 2.5, 1.4, 110., 1.5, 1.5, 1.2, 3.4, 12., 2., 4.8, 5.3, 3.4, 170.,
 1.5, 10., 220., 100., 2., 5.2, 4.1, 12., 1.3, 19., 1.2, 24., 5.8, 1.5, 5.7, 10., 6.,
 1500., 2., 170., 3.8, 23., 120., 1.6, 2.2, 5.6, 3.5, 12000., 1.9, 1.6, 1.2, 2., 17.,
 1.4, 3.3, 1.4, 4900., 11., 4.6, 120., 1.2, 5.3, 350., 120., 2.9, 5.1, 1200., 11.,
 190., 1.3, 1.9, 1.5, 14., 6.2, 91., 1., 3.7, 25., 2.4, 31., 1., 6.5, 1.1, 42., 1.9,
 11., 1.6, 4.4, 6.7, 1100., 1.4, 5.2, 1.6, 12., 3.9, 9.9, 3.1, 56., 1.2, 91., 8.7, 1.5,
 1.4, 6.7, 6.7, 1.9, 6.2, 15., 1.8, 1.3, 70., 1.2, 2.6, 2.6, 120., 2.6, 1.9, 1.3, 6.1,
 23., 1., 1., 6800., 1.9, 1., 2.5, 35., 26., 1.6, 1.6, 1.7, 40., 23., 9.9, 2.2, 1.1,
 6.2, 4.3, 1.8, 1.6, 1.8, 1.6, 2.8, 23., 2.7, 1.5, 6.9, 1.3, 1., 1.7, 2., 7.1, 1600.,
 1.3, 5.8, 1.4, 120., 11., 13., 1., 5.6, 8., 1.2, 28., 1.1, 3., 2.1, 5.8, 360., 1., 1.5,
 1.1, 2.3, 1.1, 1.9, 22., 2.5, 2.2, 83., 3.6, 1.8, 39., 1.7, 2.6, 3.1, 6.3, 1.5, 1.4,
 50., 7.6, 1.2, 9.5, 230., 5300., 1.9, 260., 4.9, 3.4, 72., 5.1, 4.3, 1.1, 23., 29.,
 6.5, 10., 830., 570., 12., 6., 1.8, 1.2, 18., 26000., 21., 2.2, 110., 1.9, 170., 8.4}
```

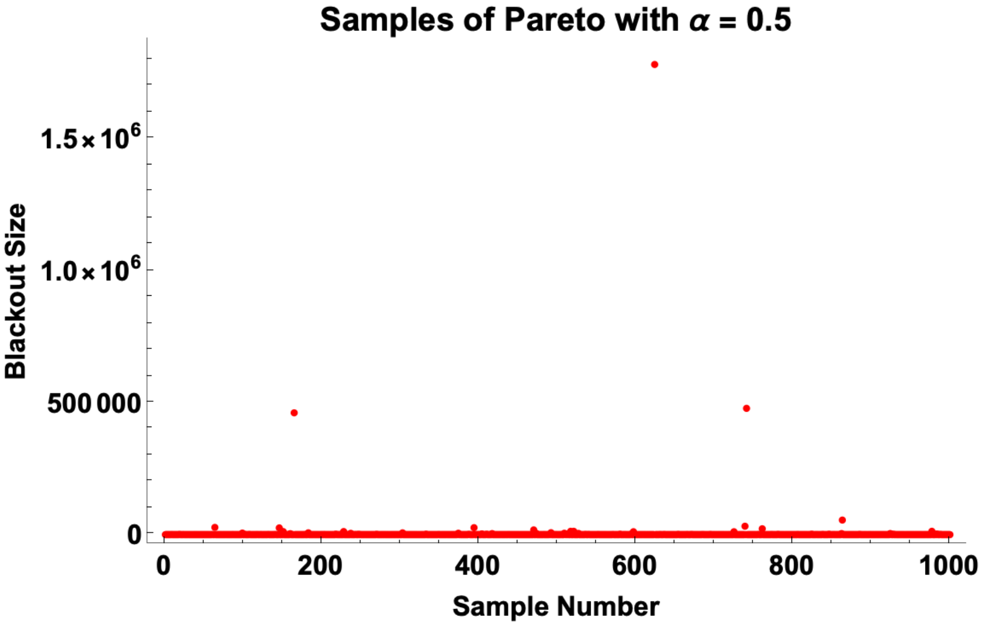
400 samples of Pareto with $\alpha = 0.5$
(2 significant figures given)

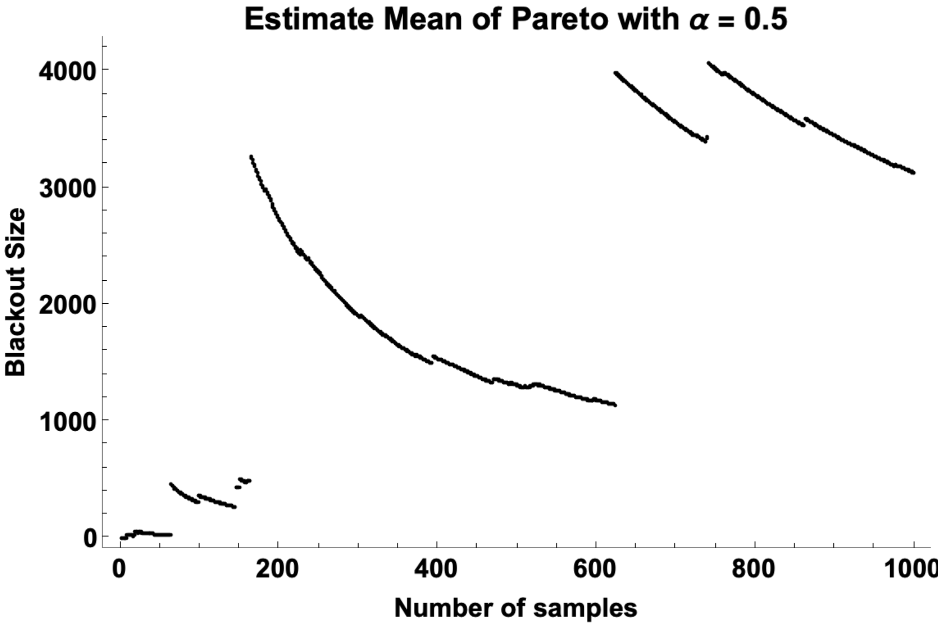
mean is infinite
median = 4

Samples from Pareto and trying to estimate mean

Take 1000 samples from Pareto with $\alpha = 0.5$; Cannot estimate the infinite mean

Occasional samples are gigantic; there is no typical large sample

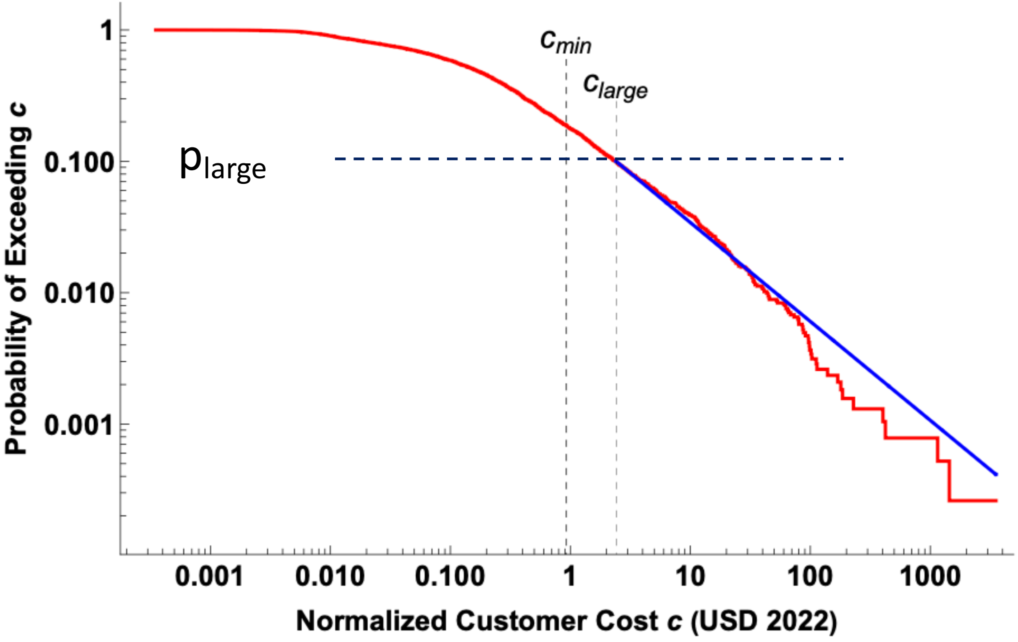




Large event customer risk

- c_{large} is large event threshold = 2.4 \$ (costs are normalized by number of customers served)
- p_{large} = P[event cost > c_{large}] = 0.1
  350 events per year so
  annual large event frequency
  f_{large} = 35 per year

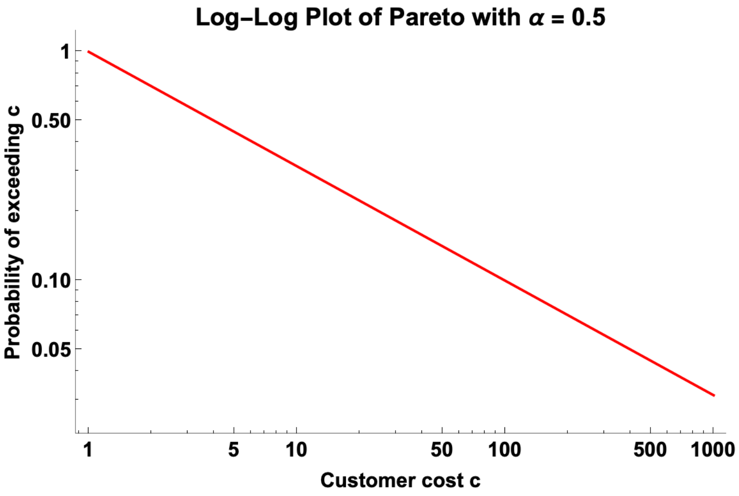
Risk is given by P[event customer cost > c]



- Exceedance curve has very heavy tail with P[event cost > c] $\propto c^{-\alpha}$
  α=0.75 is slope magnitude on log-log plot
  … probability goes down very slowly

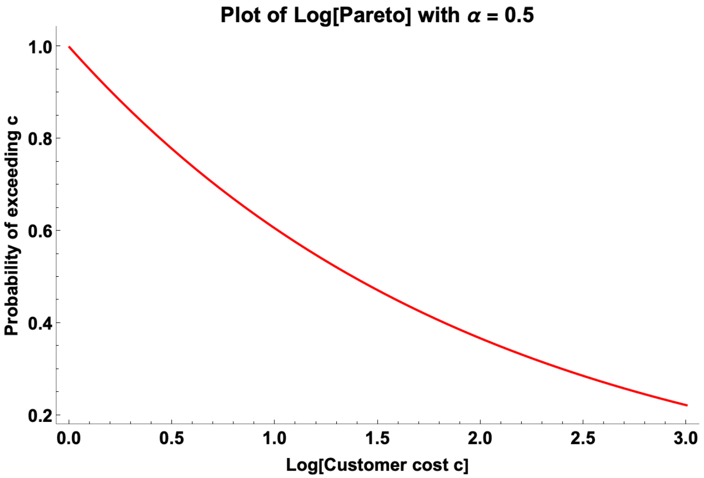
- Heavy tail implies no typical large blackout and mean of large blackouts and CVAR are ill defined and/or impractical to estimate.

- p_{large} and α metrics describe heavy tail and large event risk.



Log[Pareto] = Exponential

Therefore
Mean of Log cost data = $1/\alpha$
and estimates of mean converge!



ALEC = Average Log Event Cost

IDEA: Take log of the large event data before taking the mean
Taking log converts heavy tail into light tail and enables mean to be estimated.

large event normalized costs = $c_1, c_2, c_3, \ldots, c_{nlarge}$

$$\mathrm{ALEC} = \frac{1}{n_{\mathrm{large}}} \sum_{i=1}^{n_{\mathrm{large}}} \log_{10} c_i$$ = mean of log of large event costs

ALEC determines estimates of slope magnitude: $$\alpha = \frac{(\ln 10)^{-1}}{\mathrm{ALEC} - \log_{10} c_{\mathrm{large}}}$$

Annual frequency of large blackouts f_{large} and ALEC (implies slope magnitude α) determine large blackout risk even with the heavy tails

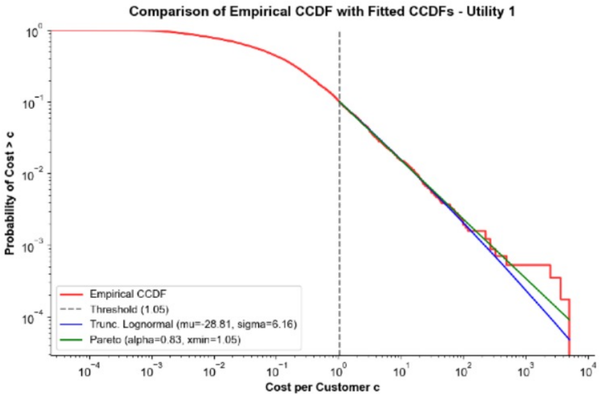
ALEC is average log cost of large events per event

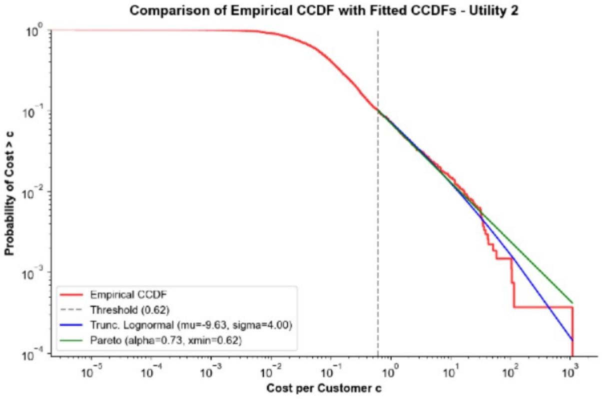
ALCRI = Annual Log Cost Resilience Index is per year

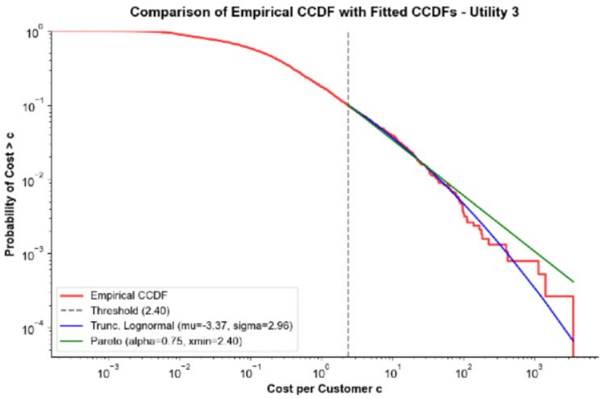
$$\mathrm{ALCRI} = f_{\mathrm{large}}\,\mathrm{ALEC} = \frac{f_{\mathrm{large}}}{n_{\mathrm{large}}} \sum_{i=1}^{n_{\mathrm{large}}} \log_{10} c_i = \frac{1}{n_{\mathrm{year}}} \sum_{i=1}^{n_{\mathrm{large}}} \log_{10} c_i$$

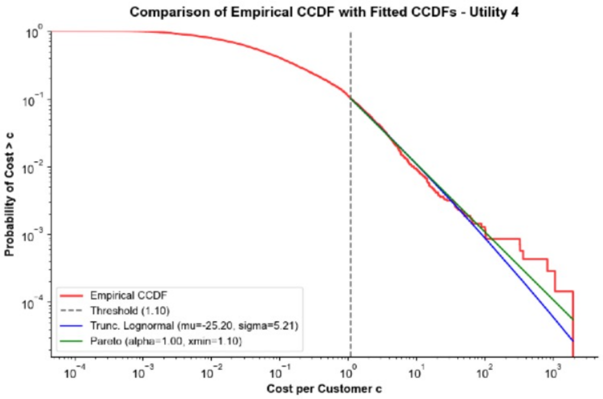
ALCRI combines frequency f_{large} and log cost into a single resilience risk index
This is especially useful when optimizing resilience

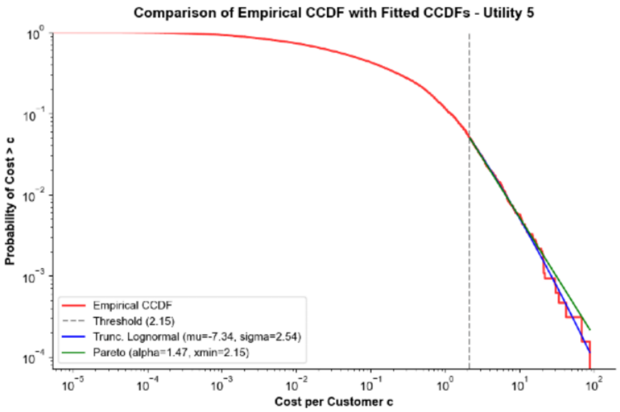
Exceedance curves for the 5 utilities







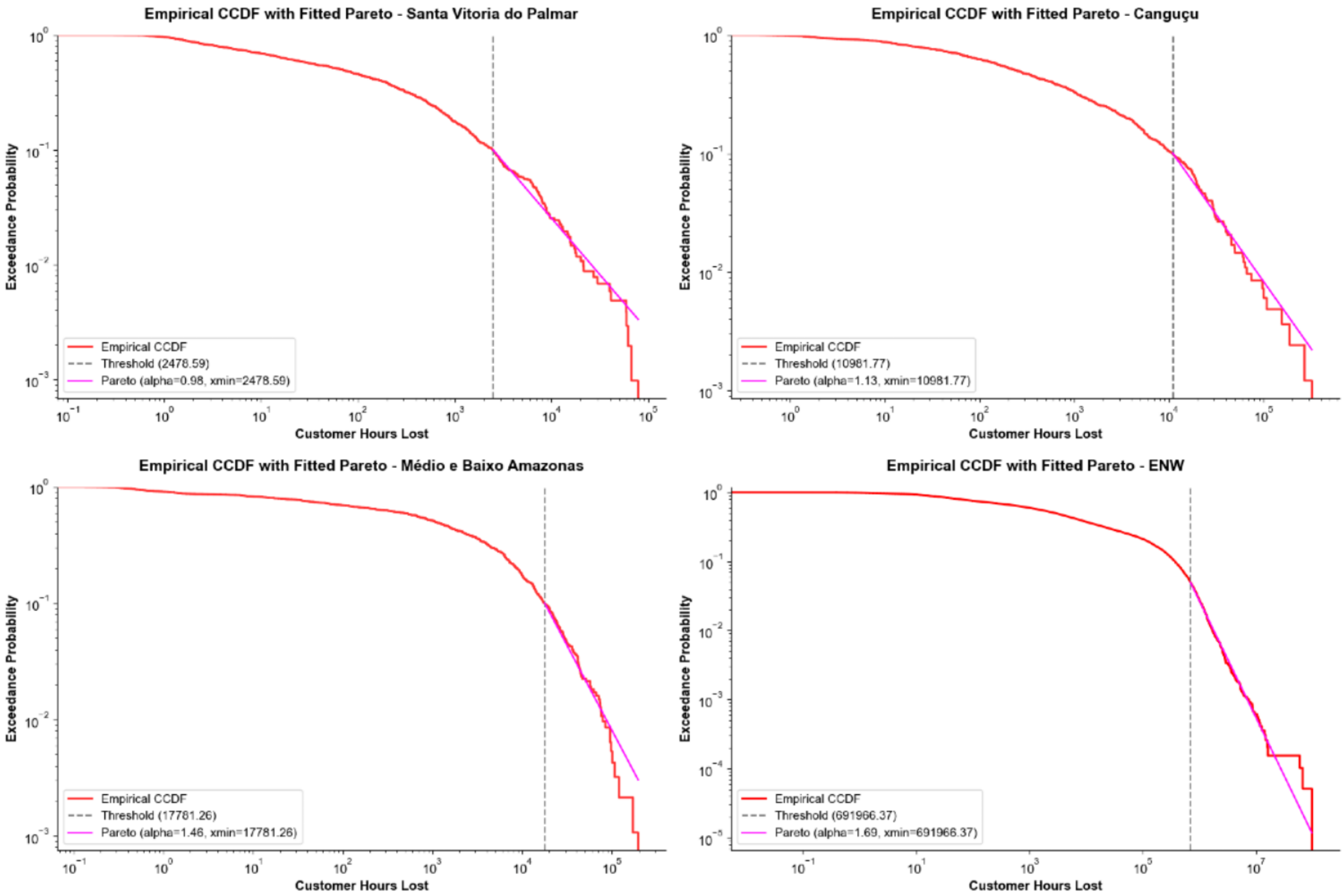




Data and Metrics for 5 USA Distribution Utilities

	Utility-1	Utility-2	Utility-3	Utility-4	Utility-5
n	5716	2706	3830	7000	6485
n_{year}	6	17.4	11	10	11
k	370.2	228.2	323.0	339.9	339.9
c_{maxobs}	\$5063	\$1095	\$3523	\$1954	\$88
c_{large}	\$1.05	\$0.62	\$2.40	\$1.10	\$2.15
n_{large}	572	271	384	701	325
p_{large}	0.1	0.1	0.1	0.1	0.05
f_{large}	96	16	35	70	30
α	0.83	0.73	0.75	1.00	1.47
CI_{α}	(.76,.89)	(.65,.82)	(.68,.83)	(.93,1.1)	(1.3,1.6)
ALEC	0.55	0.38	0.96	0.48	0.63
$\mathrm{CI}_{\mathrm{ALEC}}$	(.50,.59)	(.31,.46)	(.90,1.0)	(.44,.51)	(.60,.66)
$\mathrm{RSE}_{\mathrm{ALEC}}$	0.040	0.094	0.031	0.034	0.026
$\mathrm{RSE}_{\mathrm{Pb}}$	1.44	1.70	1.05	1.46	0.66
$\mathrm{RSE}_{\mathrm{LNb}}$	1.38	1.53	0.65	0.89	0.13
ALCRI	52.8	6.08	33.6	33.6	18.9

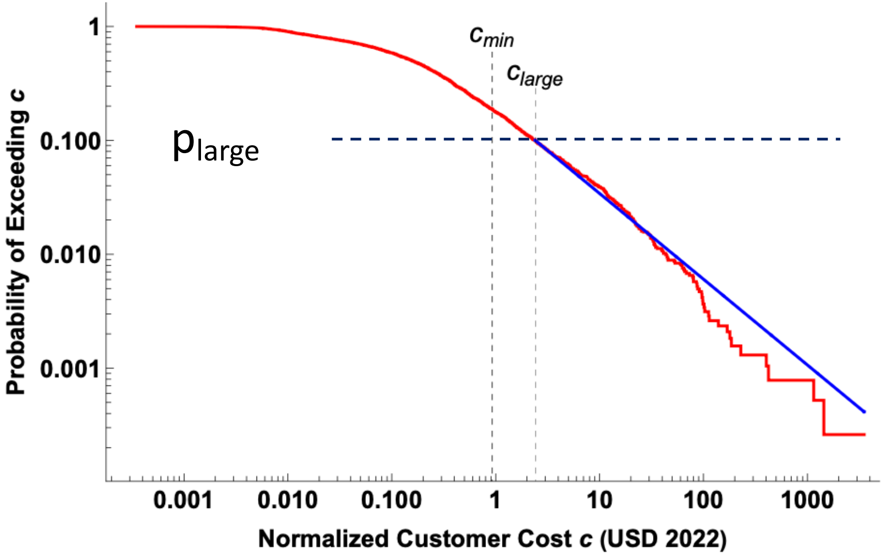
all costs in 2022 USD; f_{large} in per year.



Exceedance curves for Brazilian and English utilities

Relation to VaR and CVaR

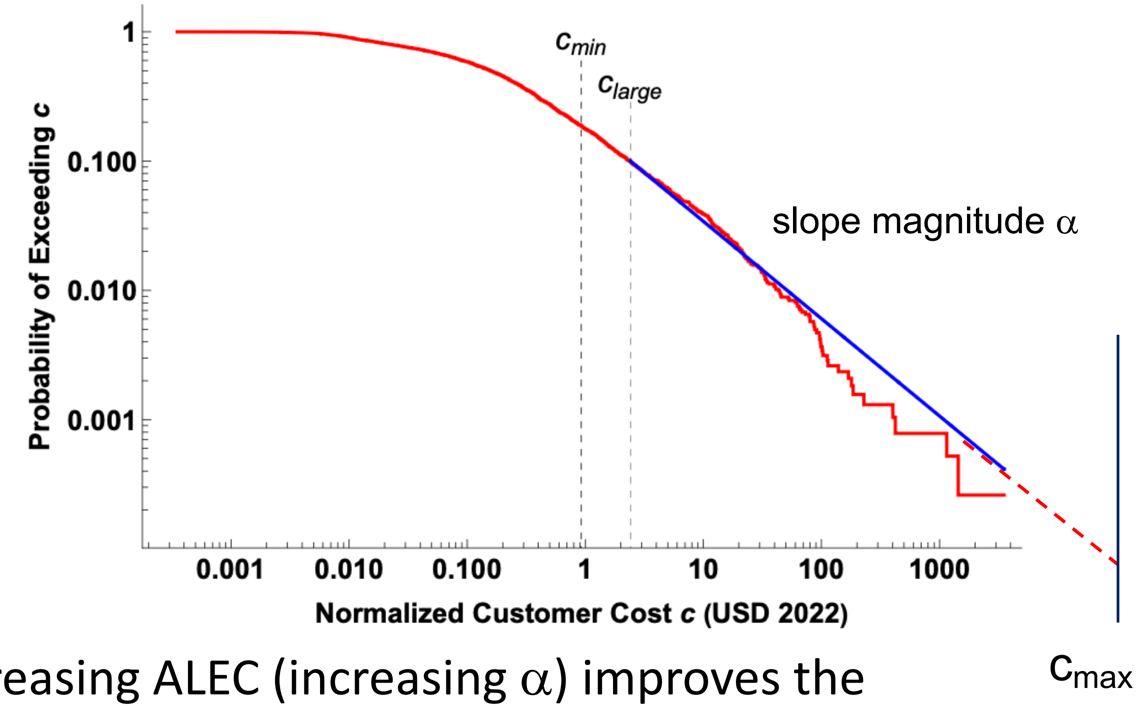
- We fix large cost threshold c_{large} and evaluate p_{large}=P[cost>c_{large}]
- VaR fixes p_{large} and evaluates corresponding c_{large}
- CVaR is mean of costs exceeding c_{large} and for our data CVaR is impractical to estimate with the limited number of large cost events



S Poudel, A Dubey, A Bose, Risk-based probabilistic quantification of power distribution system operational resilience, IEEE Systems Journal, 2020

Extrapolating beyond the largest observed blackout and the question of the largest possible blackout

- Assuming a maximum blackout cost c_{max} gives a finite mean, but requires extrapolation beyond the largest observed blackout cost.
- One choice of c_{max} is the cost equivalent to a one month blackout of the entire system
- We can extrapolate the largest blackout cost trend to c_{max} under a linear or other assumption, and our results show that the mean is still impractical to estimate.



Decreasing p_{large} and decreasing ALEC (increasing α) improves the large cost blackout trend and its linear extrapolation up to c_{max}

Major event days (MED) … and heavy tails

It is well known that customer hours are heavy tailed and yearly SAIDI is erratic unless major event days are excluded.

SAIDI excluding major event days measures reliability (that is, SAIDI usefully measures normal conditions over the year)

ALEC and ALCRI focus on and quantify the heavy tails and the large events to measure resilience risk

Comparing ALCRI and SAIDI

$$\text{ALCRI} = \frac{1}{n_{\text{year}}} \sum_{i=1}^{n_{\text{large}}} \log_{10} c_i$$

If we set cost per Customer Minute Interrupted = 1 and take n_{year} = 1, normalized event costs c_i become normalized CMI and ALCRI becomes

$$\text{SALIDI} = \sum_{\text{large events}} \log_{10} \frac{\text{event CMI}}{\#\ \text{customers served}}$$

a resilience metric

whereas

$$\text{SAIDI} = \sum_{\substack{\text{events} \\ \text{except MED}}} \frac{\text{event CMI}}{\#\ \text{customers served}}$$

a reliability metric

MED = Major Event Day

Conclusions

- Huge variability in customer event costs - well known from major event days
  - very heavy tails in cost exceedance function; slope magnitude $\alpha < 2$
  - no typical or representative large cost event
  - mean values of event costs do not converge and impractical to estimate
  - high risk of large cost events
- Logarithmic metrics can describe the large event customer risk
  - group outages into events and calculate customer hours and customer cost for each event
  - set a threshold so that large cost events have cost > c_{large}
  - take log of large event costs
  - ALEC metric is the Average Log large Event Cost (per event)
  - ALCRI metric is the Annual Log Cost Resilience Metric (per year)
- Can plausibly quantify, monitor, optimize resilience risk with these metrics

References

Distribution system blackout risk

A Ahmad, I Dobson, Logarithmic resilience risk metrics that address the huge variations in blackout cost, IEEE Trans. Power Systems, vol. 40, no. 6, November 2025, pp. 5507-5510.

A Ahmad, I Dobson, Quantifying distribution system resilience from utility data: large event risk and benefits of investments, to appear in IET Generation, Transmission & Distribution.

Transmission system blackout risk

**BA Carreras, DE Newman, I Dobson,
North American blackout time series statistics and implications for blackout risk,
IEEE Trans. Power Systems, vol. 31, no. 6, November 2016, pp. 4406--4414**

Heavy tails (math and intuition)

J Nair, A Wierman, B Zwart, The fundamentals of heavy tails: properties, emergence, and estimation, Cambridge University Press, UK, 2022

PUBLIC OUTAGE DATA: A Gold Mine for Resilience!

DISTRIBUTION SYSTEMS

- Electricity Northwest UK: register to get access
  https://electricitynorthwest.opendatasoft.com/pages/homepage/
- Brazilian Electric Regulatory Agency ANEEL
  https://dadosabertos.aneel.gov.br/dataset/interrupcoes-de-energia-eletrica-nas-redes-de-distribuicao
- Massachusetts data from Unitil, Eversource Energy, National Grid:
  https://www.mass.gov/info-details/power-outages#historic-power-outages
- EAGLE-I USA Data at DOE: https://eagle-i.doe.gov/login

TRANSMISSION SYSTEMS

- Bonneville Power Admin. BPA www.bpa.gov/energy-and-services/transmission/operations-information
- NYISO http://mis.nyiso.com/public/P-54Blist.htm
  see NK Carrington, I Dobson, Z Wang, Transmission grid outage statistics extracted from a web page logging outages in Northeast America, NAPS, College Station, TX USA, Nov 2021
  https://iandobson.ece.iastate.edu/PAPERS/carringtonNAPS21preprint.pdf
- SPP https://transoutage.spp.org

EXTRA SLIDES FOLLOW

Pareto = $e^{\text{Exponential}}$; Exponential = ln Pareto

Let Y be Exponential so that $P[Y>y] = e^{-\alpha y}$, $y \geq 0$

Let $X = e^Y$ and $x = e^y$

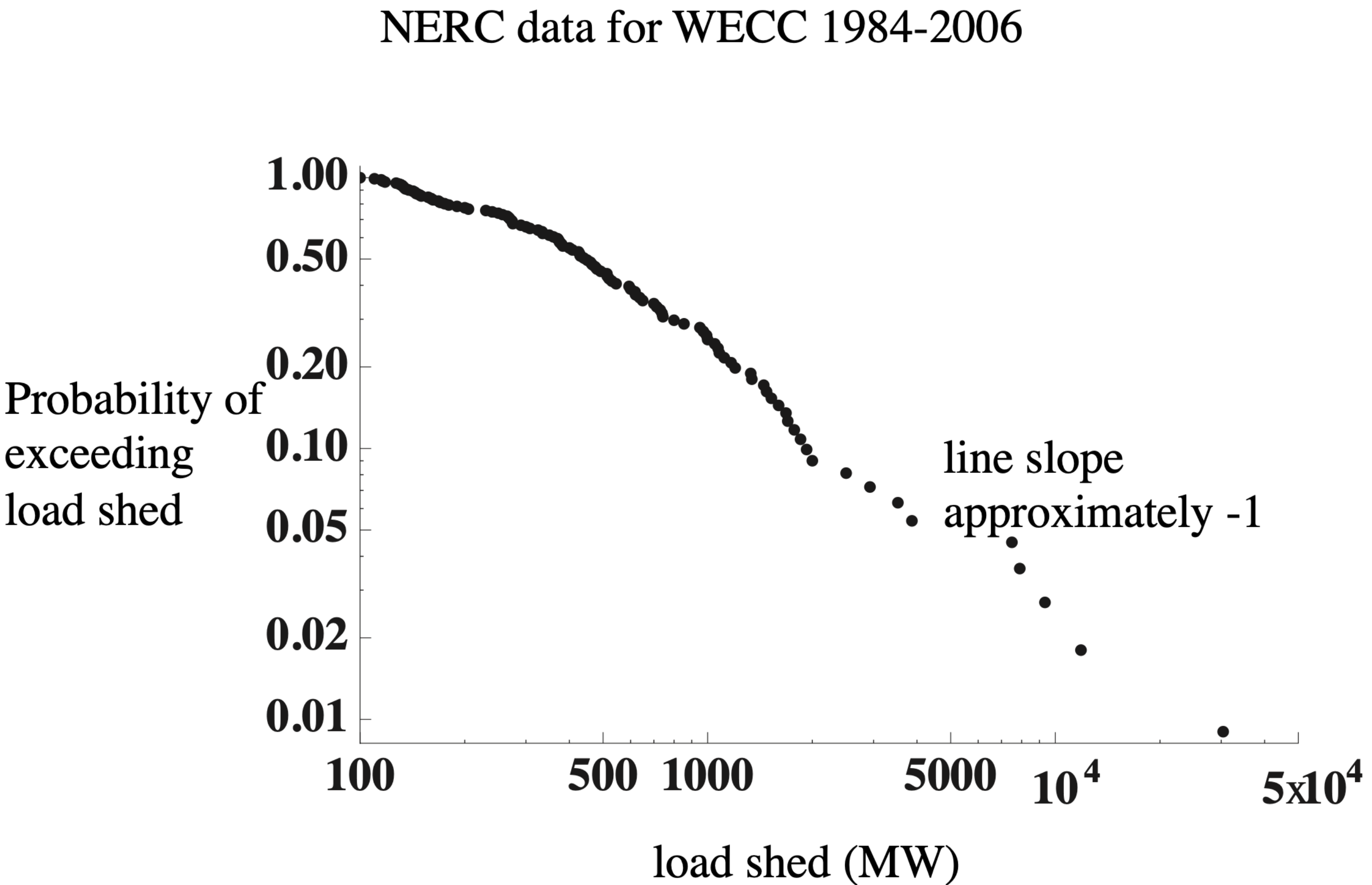
Then $P[X>x] = P[e^Y > e^y] = P[Y>y] = e^{-\alpha y} = x^{-\alpha}$, so X is Pareto

There is a similar relation between normal and lognormal distributions:

If N is normal, then e^N is lognormal

Exceedance function and PDF slopes on log-log plot differ by one:
If $P[X>x] \sim x^{-\alpha}$, then $PDF[X] \sim x^{-(\alpha+1)}$

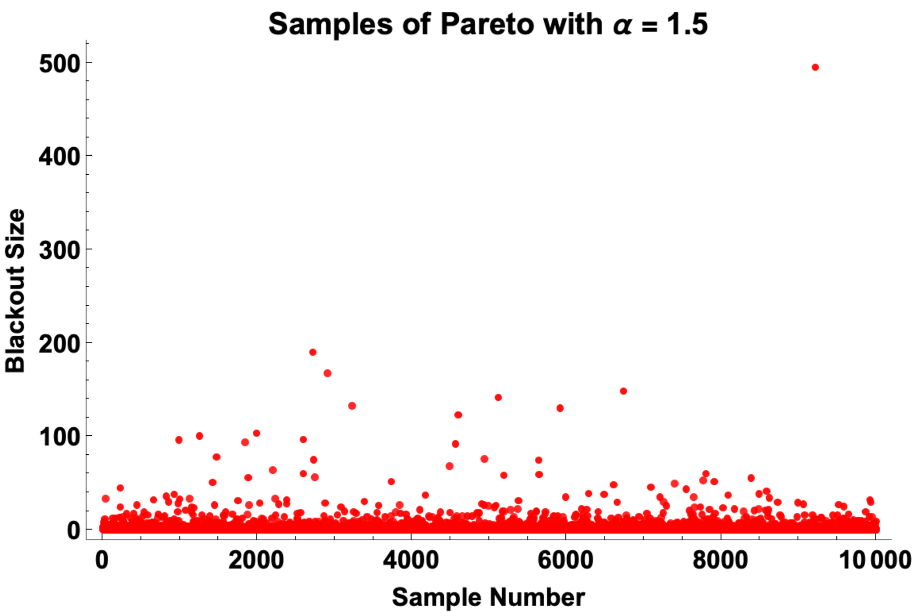
Heavy tails in probability distribution of blackout size

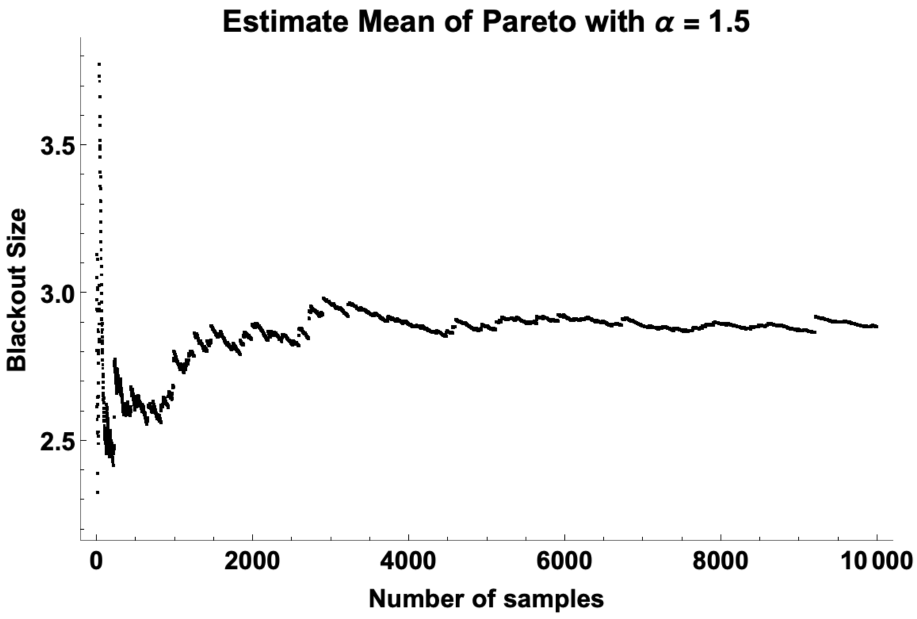


Samples from Pareto and trying to estimate mean

Take 10000 samples from Pareto with α = 1.5; Occasional samples are large;

Can estimate the mean but with large number of samples





Samples from Pareto and trying to estimate mean

Take 1000 samples from Pareto with α = 3; Samples are bounded;

Can estimate the mean with hundreds of samples