A Cluster Distribution as a Model for Estimating High-Order Event Probabilities in Power Systems

Qiming Chen
chenq@pjm.com
PJM Interconnection

James D. McCalley
jdm@iastate.edu
Iowa State University

***Abstract*— We propose the use of the cluster distribution, derived from a negative binomial probability model, to estimate the probability of high order events in terms of number of lines outaged within a short time, useful in long term planning and also in short-term operational defense to such events. We use this model to fit statistical data gathered for a 30 year period for North America. The model is compared against the commonly used Poisson model and the Power Law model. Results indicate that the Poisson model underestimates the probability of higher order events while the Power Law model overestimates it. We use the strict Chi-square fitness test to compare the fitness of these three models and find that the cluster model is superior to the other two models for the data used in the study.**

***Index Terms*—negative binomial distribution, Power Law, high-order, contingency, cascading, blackouts, rare events.**

I. Introduction

THIS paper presents a discrete probability model for high-order events in electric power systems. By "high-order," we mean events where multiple elements are lost. Such events are relatively rare but often of extremely high consequence. There are at least two applications for this probability model. The first is to estimate rare event probabilities for the transmission and generation planning process, where capital investments in new facilities must be weighed against the extent to which those facilities reduce risk associated with contingencies. The second is to estimate rare event probabilities in operations, for control-room decision-making. Here, *preventive actions*, which cost money and are routinely taken in anticipation of N-1 events, are not reasonable for a rare event, since the certain cost of the preventive action cannot be justified for an event that is so unlikely. A more appropriate strategy for dealing with rare events in operations is to identify, in advance, *corrective actions* to be taken if the rare event occurs. That is, use computational power to build operational defense procedures to be used, following occurrence of a rare event, as a decision-aid to operators in arresting cascading sequences and mitigating severity. Given that the number of rare events is excessively large, one needs a way to decide, on-line, "what to compute next" in developing operational defense procedures, i.e., one needs to prioritize the rare events for which defense plans are to be developed. The best way to prioritize is by event probability. Some work has been done in [1].

Qiming Chen is a system planning engineer with PJM Interconnection and he is also a Ph.D. candidate of the Department of Electrical and Computer Engineering of Iowa State University, (e-mail: chenq@pjm.com).

James D. McCalley is professor of Electrical and Computer Engineering at Iowa State University, Ames, IA 50011, USA. (e-mail: jdm@iastate.edu).

There are three ways to estimate power system rare event probabilities. The first is to fit an existing probability model to historical data; the second is to use physical attributes of each individual event; the third is to use Monte Carlo simulation with variance reduction. Some work has been done in [2][3]. In this paper, we report on investigations via the first approach. There are also different metrics to use in characterizing power system rare events, including number of customers interrupted, power interrupted, energy not served, and number of elements lost (*N-1, N-2*,…). We use the latter characterization in our probability model because it better conforms to planning and operating reliability criteria used in industry. For example, reliability standards performance criteria are often categorized based on the number of elements lost.

In Section II, we describe three possible probability models: Poisson, negative binomial, and power law. Section III develops and describes a specific form of the negative binomial distribution that we call the cluster model. Section IV uses the maximum likelihood estimation to estimate parameters for the cluster, Poisson, and power law distributions in describing outage statistics from North American power grid histories over a 30 year period. Section V uses a Chi-square test to compare the fitness of the three models. Section VI concludes.

II. Three Probability Models for Rare Events

We introduce three probability models in this section, each of which is used Section IV to fit outage statistics data.

A. Poisson distribution

We develop the Poisson distribution in a traditional way here. Consider the event of an individual line tripping within a fixed time period by a binary random variable T, such that $T\in\{0.1\}$, with $T=1$ representing line tripping and $Pr(T=1)=p$. The probability of tripping of each line follows a Bernoulli distribution according to

$$\Pr(T=t|p)=p^t(1-p)^{1-t},\ \text{with}\ t=0,1;\ 0\le p\le 1 \tag{1}$$

Suppose the total number of lines in a power system is N. Each line has the same probability p to be tripped within a fixed time period, and each trip event is independent of any other one. Define M as the total number of lines removed from the power system during the time period. The probability distribution of M is binomial, according to

$$\Pr(M=k)=\binom{N}{k}p^k(1-p)^{N-k} \tag{2}$$

where $k=0,1,2,3,\ldots,N$ and $p=Pr(T=1)$. Usually, p is small and N is large, in which case $k\ge 1$ becomes a rare event and can be approximated by the Poisson distribution [4], that is,

$$\Pr\left(M=k\right)=\binom{N}{k}p^{k}\left(1-p\right)^{N-k}\approx\Pr\left(M=k\left|\lambda\right.\right)=e^{-\lambda}\lambda^{k}/k! \qquad (3)$$

where $\lambda=np$, $k=0,1,2,3,\ldots\infty$. The Poisson distribution is sometimes called the distribution of rare events [4]. Both binomial distribution and Poisson distribution assume that the element events (the failure of a individual) are independent, i.e., the failure of one part of the system does not affect the failure probability of another component. This is a significant weakness of the Poisson distribution when using it to characterize rare event probabilities for power systems.

B. Negative binomial distribution

Another family of discrete distributions is the negative binomial distribution models. We present this distribution because we are going to suggest a form of this distribution as our probability model for number of lost components. In contrast to the binomial distribution, where we count the number of successes after doing a predefined number of Bernoulli tests, the negative binomial distribution counts the total number of tests to get a predefined number of successes (M). Suppose T is the random variable that represents the number of failures before a predefined number of successes (M) are observed in a sequence of Bernoulli tests, then the distribution of T is:

$$\Pr(T=t\left|p,M\right.)=\binom{M+t-1}{t}p^{M}(1-p)^{t} \qquad (4)$$

$t=0,1,2,\cdots;\ 0\leq p\leq 1;\ M>0$

The distribution is called negative binomial *(M,p)*. The range of M here can be extended to include real numbers, a feature we use in the model introduced in Section III.

C. Power law distribution[2][3][5][6]

A random X variable that follows a power law has a normalized distribution as follows:

$$\Pr(X=x\left|p\right.)=x^{-p}\Big/\int x^{-p}dx;\ p>0 \qquad (5)$$

where p is a constant. X can either be a continuous variable or a discrete variable. In case x is a discrete variable, the denominator is replaced by $\sum x^{-p}$. The expression of (5) is a proper probability density function (pdf) if the sample space is limited to a finite region, since in this case, its integral over the entire probability space is 1.0. In case the sample space is infinite, $0<x<\infty$, p must be greater than 1.0 to be a proper PDF and p has to be greater than 2 so that the mean of X is bounded. If we draw the relationship of $P(X=x)$ and x on a log-log plot, we find a straight line with slope $-p\Big/\int x^{-p}dx$, that is,

$$\log P(X=x)=-p\times\log(x)\Big/\int x^{-p}dx \qquad (6)$$

This feature is unique to power law distribution among the many other pdfs that model rare event probability. For example, if we draw the relationship of $P(X=x)$ and x on a log-log plot, the Possion distribution is concave curve, which means the probability of large events decrease faster.

III. Cluster Model for High-order Transmission Outages

Students who do not know each other while in a library tend to avoid one another by choosing regularly spaced positions; but if some students are acquaintances, they tend to sit together. People in an elevator behave similarly. Molecules in a room repel each other, filling the room uniformly; however, bacteria on a plate reproduce themselves and tend to form colonies or ‘clusters’. Likewise, insects distribute eggs in a fashion that avoids placing too many eggs in one place [7].

A main theme in this paper is that loss of elements in power systems exhibit clustering phenomena. That is, the loss of one element immediately raises the likelihood of losing another element, which has a similar effect, and so on. A fault and the ensuing relay trip of one component causes transient oscillation throughout the power system and may cause other protection devices to operate. The forced outage of one generator or line changes the network flow pattern, and some circuits, being more loaded, may trip either by proper or unintended protection operation. The more severe the first event, the more likely an additional event will follow. This tendency is captured statistically using the “cluster” probability distribution, derived from the negative binomial distribution. We will develop the cluster distribution here, and in the process, show that the Poisson distribution may be derived from it as well. This development is quite different from what is typically found in most texts; it was first presented in Thompson’s monograph [7].

The negative binomial distribution can be derived for the case of n balls being placed into m cells consecutively so that the probability of transition from occupancy numbers $(r_1, r_2, \ldots, r_m)$ with $\sum r_i=r$ to $(r_1, \ldots, r_i+1, \ldots r_m)$, which is the same as the probability that the $(r+1)^{th}$ ball falls in the i^{th} cell, is $(\alpha\cdot r_i+1)/(\alpha\cdot r+m)$. When $\alpha>0$, the transition probability is a function of how the previous r balls that are already in the cells are located; the more balls in a cell, the more likely the next ball will fall into the cell. The element events (the action of placing a ball) becomes no longer independent, the succeeding event is dependent on what previously occurred. When $\alpha>0$ with $n\rightarrow\infty$, $m\rightarrow\infty$, and $n/m=\lambda$, the distribution of number of balls (K) in any cell follows a negative binomial distribution with parameters $M=\alpha^{-1}$ and $p=\lambda/(\lambda+\alpha^{-1})$, where M and p are defined in equation (3), that is

$$P(K=k\left|\alpha^{-1},\frac{\lambda}{\lambda+\alpha^{-1}}\right.)=\binom{\alpha^{-1}+k-1}{k}(\frac{\lambda}{\lambda+\alpha^{-1}})^{k}(\frac{\alpha^{-1}}{\lambda+\alpha^{-1}})^{\alpha^{-1}} \qquad (7)$$

where $k=0,1,2,\ldots$.

We may see that when $\alpha=0$, there is no dependence of ball placement probability on the way previous balls were placed, and the transition probability is just $p=1/m$. The distribution of the number of balls (K) in any of the cells is just like the case of the binomial distribution with parameter p and n. The limiting distribution of K with $n\rightarrow\infty$, $m\rightarrow\infty$, and $n/m=\lambda$, is Poisson(λ). If we let $\alpha\rightarrow 0$, the Poisson distribution is derived as follows:

$$\lim_{\alpha\to 0} \mathrm{P}(k|\alpha^{-1}, \frac{\lambda}{\lambda+\alpha^{-1}}) = \lim_{\alpha\to 0} \binom{\alpha^{-1}+k-1}{k} (\frac{\lambda}{\lambda+\alpha^{-1}})^k (\frac{\alpha^{-1}}{\lambda+\alpha^{-1}})^{\alpha^{-1}} = \frac{e^{-\lambda}\lambda^k}{k!} \quad (8)$$

Here, the number (N) of circuits tripped in each event must be greater than 1. We reparameterize (7) by $Y=X+1$ so that the sample space of random variables is {1,2,3,…}. This is necessary because the sample space of contingencies contained in our data set does not include the event $k=0$ (loss of no elements). We also reparameterize λ by $\mu=\lambda+1$ so that $E(Y)=\mu$ still holds as $E(X)=\lambda$ in (7). We use the notation $Cluster(Y=y|\mu,\alpha)$ to represent the new reparameterized distribution, defined as

$$Cluster(Y=y|\mu,\alpha)=$$
$$\begin{cases} \binom{\alpha^{-1}+y-2}{y-1}(\frac{\mu-1}{\mu-1+\alpha^{-1}})^{y-1}(\frac{\alpha^{-1}}{\mu-1+\alpha^{-1}})^{\alpha^{-1}} & \text{if } \alpha>0, \mu>1, \text{ and } y=1,2,3\cdots \\ \frac{e^{-(\mu-1)}(\mu-1)^y}{(y-1)!} & \text{if } \alpha=0, \mu>1, \text{ and } y=1,2,3\cdots \end{cases} \quad (9)$$

We call α the *affinity factor;* it will be shown to capture the tendency of the power system to have a cascading event. To compare the shapes of Cluster pdfs having different α with the Power law pdf, Table I summarizes convergence rates when the random variable (an index representing the size of contingencies) approaches infinity.

TABLE I: Comparing convergence rate with Cluster & power law pdf shapes

pdf Family	Poisson(λ) Or Cluster ($\alpha=0$)	Cluster ($1>\alpha>0$)	Geometric Or Cluster ($\alpha=1$)	Cluster ($\alpha>1$)	Power Law ($c>0$)
$\frac{\Pr(X=x+1)}{\Pr(X=x)}$	$\frac{\lambda}{x+1}$	$\frac{\alpha_1^{-1}+x-1}{x} \times \frac{\mu-1}{\mu-1+\alpha_1^{-1}}$	$\frac{\mu-1}{\mu}$	$\frac{\alpha_1^{-1}+x-1}{x} \times \frac{\mu-1}{\mu-1+\alpha_1^{-1}}$	$\left(\frac{1+x}{x}\right)^{-c}$
$\frac{\Pr(X=x+1)}{\Pr(X=x)}$ with $x\to\infty$	0	$\frac{\mu-1}{\mu-1+\alpha_1^{-1}}$	$\frac{\mu-1}{\mu}$	$\frac{\mu-1}{\mu-1+\alpha_2^{-1}}$	1
Shape of *x v.s. Pr(x)*	CC	CC	CC	CC	CC
Shape of *x v.s. logPr(x)*	CC	CC	SL	CV	CC
Shape of log(*x*) *v.s. logPr(x)*	CV	CV	CV	CV	SL

Legend: CV=Convex CC=Concave SL=StraightLine

From the table, we see that when $x\to\infty$, i.e. when events became large or 'rare', we have

$$0 < \frac{\mu-1}{\mu-1+\alpha_1^{-1}} < \frac{\mu-1}{\mu} < \frac{\mu-1}{\mu-1+\alpha_2^{-1}} < 1 \quad (10)$$

which means the relation between the different probability models, in terms of convergence rate for rare events, is:

$$\{Poisson\} > \{Cluster(1>\alpha>0)\} > \{Geometric\} > \{Cluster(\alpha>1)\} > \{PowerLaw\}$$

The convergence rate is actually an index representing the heaviness of the pdf's tail, or the likelihood of rare events. The higher the convergence rate, the less likely a rare event occurs. For Poisson distribution, since it assumes the independence of events, it converges faster than any of the five distributions described in Table I. The convergence rate of the Cluster distribution depends on α, which is just a parameter showing the interdependency of events, or the tendency of clustering. The larger the α, the slower the convergence and the heaver the tail of the cluster distribution.

IV. Cluster Model Applied To Outage statistics

The data we analyzed is the total number of elements lost in each contingency in North America from 1965 to 1985 [8], as indicated in Table II. The last two columns give a summary by voltage level. According to [8], the data reported in Table II adheres to the following:

- Each individual component tripping in a multi-component outage event must occur within a 1 minute interval; otherwise, it is considered a separate outage event.
- Whenever an event involves components of different voltage levels, it will be counted as one instance only with a specific voltage level.

TABLE II High order transmission outages statistics
An IEEE survey of US and Canadian overhead transmission outages at 230kv and above, 1965-1985, [8]

Cont. Type	Number of Contingences By Line Voltage Levels				Total No & Perc	
	230kv	345kv	500kv	765kv		
N-1	3320	5807	721	295	10143	89.84 %
N-2	303	577	35	36	951	8.42%
N-3	39	99	3	2	143	1.27%
N-4	18	16	0	2	36	0.32%
N-5	7	1	0	0	8	0.07%
N-6	3	1	0	1	5	0.04%
N-7	0	1	0	0	1	0.01%
N-8	3	0	0	0	3	0.03%

We draw the relationship of $Pr(X=x)$ vs x as Fig. 1-3 below. The shapes (concave or convex) of Fig. 1-3 match the cluster model in the third column of Table I, which suggests cluster model with $1>\alpha>0$ could be better than others.

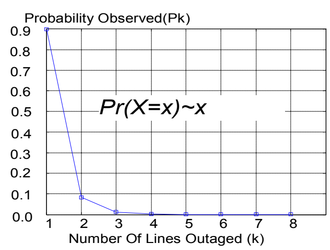

Fig. 1. Pr(k) v.s. k plot

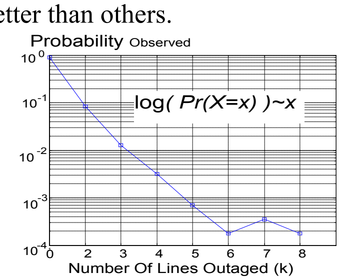

Fig. 2. log{Pr(k)} v.s. k plot

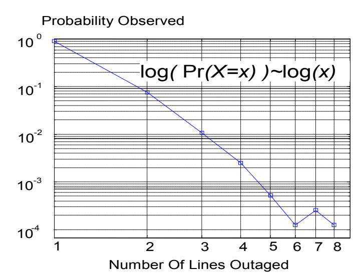

Fig. 3. log{$Pr(k)$} v.s. log(k) plot

A. Maximum Likelihood estimation (MLE)

We fit the three probability models, Poisson, cluster, and Power Law to the data in Table II. We use maximum likelihood estimation (MLE) to estimate the parameter(s) for each model. In MLE [9], if $(X_1, X_2, X_3, \ldots, X_n)$ are one independent identical sample from the space X with probability distribution function $f(x|\theta_1, \theta_2, \ldots, \theta_m)$, where the θ_i's are the model parameters to be estimated, the joint probability distribution function (PDF) is

$$\Pr\left(X_1 = x_1, \cdots, X_n = x_n \middle| \theta_1, \cdots, \theta_m\right) = \prod_{i\in\{1,2,\cdots,n\}} f(x_i|\theta_1, \cdots, \theta_m) \quad (11)$$

The ML approach defines a likelihood function

$$L(\theta_1,\cdots, \theta_m|x_1,\cdots,x_n) = \prod_{i\in\{1,2,\cdots,n\}} f(x_i|\theta_1,\cdots, \theta_m) \quad (12)$$

which is equal to the joint PDF but switches the parameters and variables, i.e., it takes the sample value $(x_1,x_2,...,x_n)$ as parameters and $(\theta_1,\theta_2,...,\theta_m)$ as variables. If we transform (12) by taking the logarithm of both sides, then we get the following log likelihood equation

$$\log L(\theta_1,\cdots, \theta_m|x_1,\cdots,x_n) = \sum_{i\in\{1,2,\cdots,n\}} \log f(x_i|\theta_1, \theta_2, \cdots, \theta_m) \quad (13)$$

Define $\underline{\theta}=\theta_1,\theta_2,...,\theta_m$ and $\underline{x}=(x_1,x_2,...,x_n)$. The $\underline{\theta}'$ that maximizes $L(\underline{\theta}|\underline{x})$ is called a MLE of the parameter $\underline{\theta}$. It should be noted that $\underline{\theta}'$ must be a global maxima. Because it is easier to find the maxima of (13) than that of (12) by differentiation, we use the log likelihood function.

B. Estimating the parameters of Poisson distribution

If we assume our data follows a Poisson distribution, then only parameter λ need be estimated. The Poisson distribution is a special case of the negative binomial in the form of (7) where $\alpha=0$, indicates Poisson has no clustering. Suppose we observe N_k samples of $K=k$ from the $Poisson(\lambda)$ such as in (3) and the sample space of K is $\{0,1,2,3,...\}$. The likelihood function is

$$\log L\left(\lambda|N_0,\cdots,N_K,\cdots\right) = \sum_{k=0,1,\cdots} \log\left(\frac{e^{-\lambda}\lambda^k}{k!}\right)^{N_k} = \sum_{k=0,1,\cdots} N_k \log\left(\frac{e^{-\lambda}\lambda^k}{k!}\right) \quad (14)$$

In order to find the value of $\lambda(N_1,N_2,...)$ that maximizes $\log\{L(\lambda|N_1,N_2,...,N_k,...)\}$, we need to solve

$$\partial \log L(\lambda|N_0,N_1,\cdots,N_K,\cdots)/\partial\lambda = 0 \quad (15)$$

Substitution of (14) in (15) results in (16), the MLE of λ.

$$\hat{\lambda} = \sum_{k=0}^{\infty} kN_k \Big/ \sum_{k=0}^{\infty} N_k \quad (16)$$

which is the sample average. Recalling that the sample space of the standard Poisson distribution is $\{0,1,2,3,...\}$ while the range of the data we have is {1, 2, 3, 4, 5, 6, 7, 8}, which does not contain zero. The MLE $\hat{\lambda}$ is:

$$\hat{\lambda} = \sum_{k=1}^{8}(k-1)N_k \Big/ \sum_{k=1}^{8} N_k \approx 0.12657 \quad (17)$$

The estimated distribution is as (18) where $k=1,2,3\ldots$

$$\Pr(Y=k) = e^{-\hat{\lambda}}\hat{\lambda}^{k-1}/(k-1)! = e^{-0.12657}0.12657^{k-1}/(k-1)! \quad (18)$$

C. Estimating the parameters of Cluster Model

For the cluster model, we will deduce the MLE of λ and α.. We consider only $\alpha>0$, since if $\alpha=0$, the cluster distribution is Poisson. The cluster distribution is given as:

$$Cluster(Y=y|\alpha,\mu) = \binom{\alpha^{-1}+y-2}{y-1} \times \left(\frac{\mu-1}{\mu-1+\alpha^{-1}}\right)^{y-1} \times \left(\frac{\alpha^{-1}}{\mu-1+\alpha^{-1}}\right)^{\alpha^{-1}} \quad (19)$$

for $y=1,2,3,...$.Given N_k samples for each K, the likelihood function is

$$\log L(\alpha,\mu|N_1,N_2,\cdots N_K,\cdots) = \sum_{k\in\{1,2,\cdots\}} \log\left\{\binom{\alpha^{-1}+k-2}{k-1} \times \left(\frac{\mu-1}{\mu-1+\alpha^{-1}}\right)^{k-1} \times \left(\frac{\alpha^{-1}}{\mu-1+\alpha^{-1}}\right)^{\alpha^{-1}}\right\}^{N_k}$$
$$= \sum_{k\in\{1,2,\cdots\}}\left\{N_k \times \log\binom{\alpha^{-1}+k-2}{k-1}\right\} + \log\left(\frac{\mu-1}{\mu-1+\alpha^{-1}}\right) \times \left(\sum_{k\in\{1,2,\cdots\}}\left[(k-1)\cdot N_k\right]\right)$$
$$+\log\left(\frac{\alpha^{-1}}{\mu-1+\alpha^{-1}}\right)^{\alpha^{-1}} \times \left(\sum_{k\in\{1,2,\cdots\}} N_k\right) \quad (20)$$

To find the candidate pair $(\hat{\alpha},\hat{\mu})$ that maximize the function $\log L(\alpha,\mu|N_1,N_2,...,N_K,...)$, we need to solve the following:

$$\begin{cases} \partial \log L(\alpha,\mu|N_1,N_2,\cdots,N_K,\cdots)/\partial\alpha\big|_{(\alpha,\mu)=(\hat{\alpha},\hat{\mu})} = 0 \\ \partial \log L(\alpha,\mu|N_1,N_2,\cdots,N_K,\cdots)/\partial\mu\big|_{(\alpha,\mu)=(\hat{\alpha},\hat{\mu})} = 0 \end{cases} \quad (21)$$

These two equations are far more complex than (15). However we can still solve for $\hat{\lambda}$. Substituing (20) in (21) , we get a closed form solution for $\hat{\mu}$, the same formula as that of Possion.

$$\hat{\mu} = \sum_{k\in\{1,2,\cdots\}} k\cdot N_k \Big/ \sum_{k\in\{1,2,\cdots\}} N_k \quad (22)$$

However we have difficulty in finding a closed form solution for $\hat{\alpha}$. We will search for the maxima pair $(\hat{\alpha},\hat{\mu})$ for $\log(L(\alpha,\mu|N_1,N_2,...,N_k...)$ directly using the contour graph function in Matlab. It is much easier to understand in this case with the aid of graph. Substituting the k's and the N_k's into the likelihood function of (20), we get the graphs of the likelihood function $\log(L(\alpha,\mu|N_1,N_2,...,N_k...)$ as Fig. 4 and Fig. 5 . The MLE of α and μ are approximately

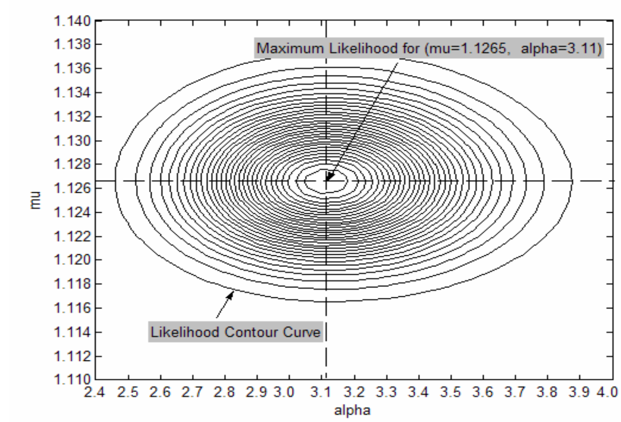

Fig. 4. Contour plot of maximum likelihood function (20)

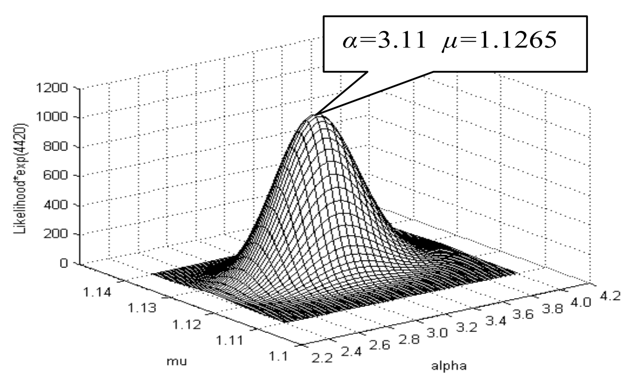

Fig. 5. Mesh plot of maximum likelihood function (20)

$$\hat{\mu} \approx 1.1265,\ \ \hat{\alpha} \approx 3.115 \quad (23)$$

We can also get the estimate of μ directly, using (22)

$$\hat{\mu} = \sum_{k=1}^{8} k \times N_k \Big/ \sum_{k=1}^{8} N_k = 1.12657\cdots \quad (24)$$

This is very close to our estimate of μ from contour plot in Fig. 4, which is evidence of the correctness of our method.

D. Estimating the parameters for Power Law distribution

The likelihood function for the power law distribution is

$$\begin{aligned} \log L\left(p \middle| N_1, N_2, \cdots\right) &= \sum_{k=1,\cdots,\infty} N_k \times \log\left(k^{-p} \Big/ \sum_{k=1,\cdots,\infty} k^{-p}\right) \\ &= (-p) \times \sum_{k=1}^{\infty} \left(N_k \times \log k\right) - \left(\log \sum_{k=1}^{\infty} k^{-p}\right) \times \left(\sum_{k=1}^{\infty} N_k\right) \\ &= (-p) \times \sum_{k=1}^{\infty} \left(N_k \times \log k\right) - \left(\log \sum_{k=1}^{\infty} k^{-p}\right) \times \left(\sum_{k=1}^{\infty} N_k\right) \end{aligned} \quad (25)$$

Like $\hat{\alpha}$ in the cluster model, we cannot get closed form solution for $\hat{p}$ since the following cannot be solved analytically:

$$0 = (-p) \times \left(\sum_{k=1}^{\infty} \left(N_k \times \log k\right)\right) - \left(\log \sum_{k=1}^{\infty} k^{-p}\right) \times \left(\sum_{k=1}^{\infty} N_k\right) \quad (26)$$

Since we have only one parameter to estimate, it is easier to find $\hat{p}$ than the $(\hat{\alpha}, \hat{\mu})$ pair for the cluster model. In order to estimate the Power Law parameter (as described in section I), we use the least square curve fitting to estimate the approximate range of the slope $-p \Big/ \int x^{-p} dx$. It is around 4.0. We draw the likelihood function (26) with p ranging from 3 to 6, and then we obtain the plot in Fig. 6. Using the bisection method to search the maximum, we find $p \approx 3.78$. Then the estimated power law PDF for the data in the last two columns of TABLE II is

$$\mathrm{P}(\mathrm{X} = x | p = 3.78) = x^{-3.78} \Big/ \sum_{k=1}^{\infty} k^{-3.78} = 0.9098 x^{-3.78} \quad (27)$$

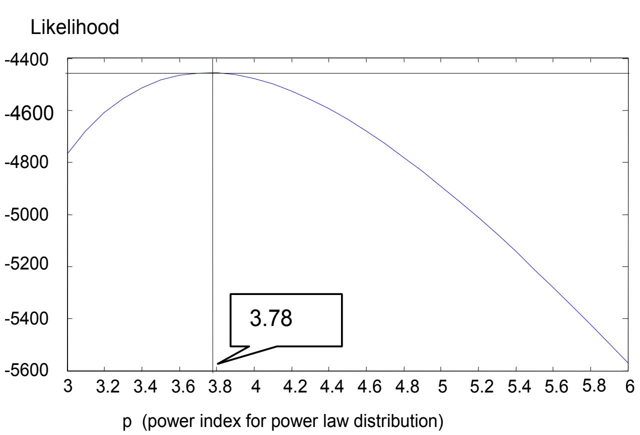

Fig. 6. Plot of maximum likelihood function (26)

The three distributions are summarized below, where $x \in \{1,2,3,\ldots\}$ for all the three models.

$$\Pr(X = x | \alpha = 3.115, \mu = 1.12657) = \binom{3.115^{-1} + x - 2}{x - 1} \times \left(\frac{1.1266 - 1}{1.1266 - 1 + 3.115^{-1}}\right)^{x-1} \times \left(\frac{3.115^{-1}}{1.1266 - 1 + 3.115^{-1}}\right)^{3.115^{-1}} \quad (28)$$

for cluster model

$$\Pr\left(X = x | \lambda = 0.12657\right) = e^{-0.12657} 0.12657^{x-1} \big/ (x-1)! \quad (29)$$

for poisson model

$$\mathrm{P}(\mathrm{X} = x | p = 3.78) = x^{-3.78} \Big/ \sum_{k=1}^{\infty} k^{-3.78} = 0.9098 x^{-3.78} \quad (30)$$

for Power Law model

Evaluating the three expressions above for *k={1,2,3,4,5,6,7,8}*, we obtain the results shown in TABLE III. These results are plotted in Fig. 7.

By inspecting Fig. 7, one concludes that the Cluster model (the curve with squares) is superior to the other two models, as it fits almost perfectly for *k=1,…,7*. The Cluster model and the power law model agree for *k=8*, for which there is almost no data. The power law model overestimates the probability of large contingencies. The concave curve generated by the Poisson model deviates heavily from the observed data, underestimating the probability of large events (*k>3*) by a factor of 10^{-5}.

TABLE III Comparing the fitness of three different probability models for the distribution of observed multiple line outages

Cont. Type	K	Observed Number	Observed Prob.	Cluser	PowerLaw	Poisson
N-1	1	10143	0.898	0.899	0.91	0.881
N-2	2	951	0.084	0.082	0.066	0.112
N-3	3	143	0.0127	0.0152	0.0143	0.00706
N-4	4	36	0.00319	0.00333	0.00482	0.000298
N-5	5	8	0.000709	0.000783	0.00207	9.42E-06
N-6	6	5	0.000443	0.000191	0.00104	2.39E-07
N-7	7	1	8.86E-05	0.000048	0.000581	5.03E-09
N-8	8	3	0.000266	1.22E-05	0.000351	9.1E-11

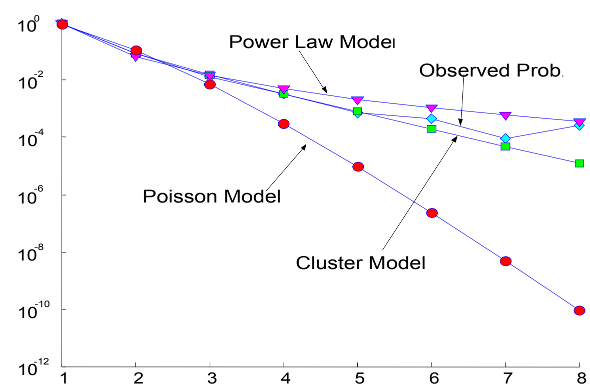

Fig. 7. The semi-log plot of PDF's in (28), (29),and (30)

V. Fitness Test of Three Different Probability Models

Fig. 7 provides qualitative evidence that the cluster model fits the data better than either the Poisson model or the power law model. In this section, we use the Chi-square test to provide quantitative evidence. The Chi-square test, based on the Pearson theorem [10], is widely used in statistics to test the fitness of a probability model to sample data. Suppose a certain random trial has k possible outcomes, the probability that each trial results in the i^{th} outcome is p_i, $i=1,2,3,\dots,k$, where $\sum p_i=1$. If we perform n trials, and the i^{th} outcome results N_i times, then the multivariate distribution of N_i is

$$\Pr\left(N_1=n_1,\cdots,N_k=n_k \left| p_1,\cdots p_k\right.\right)=\frac{n!}{n_1!n_2!\cdots n_k!}p_1^{n_1}p_2^{n_2}\cdots p_k^{n_k} \quad (31)$$

with $\sum_{i=1}^{k} n_i = n$ and $\sum_{i=1}^{k} p_i = 1$.

Pearson theorem: Suppose the parameters of a polynomial distribution has the pdf as in (31), and define

$$\chi^2=\sum_{i=1}^{k}\left(N_i-np_i\right)^2 \Big/ np_i \quad (32)$$

then when $n\to\infty$, χ^2 follows the chi-square distribution $\chi^2\ (k-1)$. We can see from (32) that the statistic χ^2 is an index showing how much the samples deviate from the polynomial distribution to be tested. The larger the statistics χ^2, the larger the deviation. In order to apply the Pearson theorem, we need to convert the distribution we are going to test into a polynomial distribution. Since the sample space of the three distributions we are going to test is $(1,2,3,\dots)$, and it is an infinite set, we separate the sample space into k exclusive sets denoted as $S_i=1,2,3,\dots,k$. Suppose X is a random variable and its pdf is $f(x)=\Pr(X=x), x\in\{1,2,3,\dots\}$. We draw a total of n samples of X from pdf $f(x)$ and count the number (denoted again as N_i) of samples that are members of the set S_i. Denote $p_i=Pr(X\in S_i)$. Then the random variables N_i, $i=1,2,\dots,k$ follow the polynomial distribution of (31). The statistics χ^2 defined in (32) follow $\chi^2(k-1)$ distribution. If χ^2 is too large, we have reason to doubt the fitness of our model with respect to the data. For this test we decompose the sample space into 5 exclusive sets. They are $S_1=\{1\}$, $S_2=\{2\}$, $S_3=\{3\}$, $S_4=\{4\}$, $S_5=\{5\}$. The reason we group this way is that for all S_i, and all test models (i.e. Poisson, cluster and power law) $P(X\in S_i)\times N_i$ (where N_i is the number of samples that fall in set S_i) is greater than 5, which is suggested for the credibility of the fitness test.

The Pearson theorem assumes all parameters for the distribution to be tested are known. If there is any parameter that is unknown so that p_i's are estimates, we need to reduce the freedom number of the χ^2 distribution by one for each estimated parameter. The rule is: if there are a total of r estimated parameters, the freedom number of the χ^2 distribution is k-r-1. For the Poisson model, parameter λ is an estimate, so the Chi-square distribution has freedom number 5-1-1=3. For the cluster model, α and μ are estimates, so the freedom number is 2. For the power law model, the power index p is estimated, so the freedom number is 3. The test result is summarized in Table IV. The last row of the table lists the probability of getting a sample deviation larger than observed, assuming the sample comes from a certain probability model. The cluster model is far more fit than the other two.

VI. Conclusion

This paper proposes the cluster model for computing probabilities associated with high-order events. This model is very appealing because it provides the ability, through the affinity factor α, to capture the tendency of component outages

TABLE IV χ^2-Test results summary

i	n_i	Cluster		Poisson		Power Law	
		p_i	$p_i n_i$	p_i	$p_i n_i$	p_i	$p_i n_i$
1	10143	0.899	10147	0.881	9948	0.91	10271
2	951	0.082	921	0.112	1259	0.066	748
3	143	0.015	172	0.0071	80	0.014	161
4	36	0.0033	38	0.0003	3.36	0.0048	54
>=5	17	0.00103	11.68	9.67E-06	0.11	0.0049	55.3
$\chi^2=\sum_{i=1}^{k}(N_i-np_i)^2/np_i$		8.37		3060		91.76	
Prob.$\{\chi^2>\chi^2(k-r-1)\}$		0.0152		$\ll 10^{-15}$		$<10^{-15}$	

in power systems to increase the likelihood of successive component outages, through, for example, cascading phenomenon, so that they cluster. We have shown that the cluster model is actually a quite general model, with the familiar Poisson (complete independence between events) being a specific instance of it, where α=0. When α is very large, the power law and the cluster model both exhibit similar behavior in convergence speed as the event becomes very large. In our application to real data, we observed that Poisson underestimates rare event probabilities, power law overestimates them, and the cluster model captures them very well, and this observation was confirmed using a statistical test of model fitness. The results of this work will enhance decision making at both the planning and operational level. In particular, operational procedures for defending against large outages are of great interest to us, and the cluster model is a promising aid in directing computational resources as they are used on-line to develop defense strategies as real-time conditions change. Work to this effect has been submitted for publication.

References

[1] Q. Chen, J. McCalley, "Identification of High Order Initiating Events for Power System Security Assessment Considering Protection System Failure", *Proceedings of North American Power Symposium (NAPS) 2001*, pp 303-308, October 15-16, 2001.

[2] B. Carreras, D. Newman, I. Dobson, A. Poole, "Initial evidence for self-organized criticality in electric power system blackouts", *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, pp 1411–1416, 2000.

[3] I. Dobson, J. Chen, J. Thorp, B. Carreras, D. Newman, "Examining criticality of blackouts in power system models with cascading events", *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, 2002. HICSS, pp 803-812, January 7-10, 2002.

[4] M. Falk, J. Husler, R. Reiss, *Laws of small numbers: Extremes and rare events*, Birkhauser Verlag, Basel, Switzerland, 1994, p. 4.

[5] P. Bak, *How Nature Works*, Springer-Verlag New York, Inc., 1996.

[6] P. Bak, C. Tang, K. Wiesenfeld, "Self-Organized Criticality: An Explanation of 1/f Noise", *Phys. Rev. Lett.*, Vol. 59, pp 381-384, 1987.

[7] W.A. Thompson, Jr., *Point Process Models with Applications to Safety and Reliability*, Chapman and Hall, 1988.

[8] R. Adler, S. Daniel, C. Heising, M. Lauby, R. Ludorf, T. White, "An IEEE survey of US and Canadian overhead transmission outages at 230 kV and above", *IEEE Trans. on Pwr Del.*, Vol. 9, Issue 1, pp 21 -39, Jan. 1994.

[9] G. Casella, R. Berger, *Statistical Inference*, the Wadsworth Group, 2002.

[10] J. Neter, M. H. Kutner, C. J. Nachtsheim, W. Wasserman, *Applied Linear Statistical Models (4th edition),* R. D. Irwin, 1996