*Probability in the Engineering and Informational Sciences*, **19**, 2005, 489–505. Printed in the U.S.A.

# A CLUSTER DISTRIBUTION AS A MODEL FOR ESTIMATING HIGH-ORDER-EVENT PROBABILITIES IN POWER SYSTEMS

QIMING CHEN AND JAMES D. MCCALLEY
*Iowa State University*
*Ames, Iowa*
*E-mail: qmchen@ieee.org; jdm@iastate.edu*

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 to the commonly used Poisson model and the power-law model. Results indicate that the Poisson model underestimates the probability of higher-order events, whereas 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.

## 1. 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, *preven-*

© 2005 Cambridge University Press 0269-9648/05 $16.00 

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

*tive 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 the 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, online, "what to compute next" in developing operational defense procedures; that is, 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 [6].

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 [4,7]. 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 the number of customers interrupted, the power interrupted, the energy not served, and the number of elements lost ($N-1, N-2, \ldots$). We use the latter characterization in our probability model because it better conforms to the 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 2 we describe three possible probability models: Poisson, negative binomial, and power law. Section 3 develops and describes a specific form of the negative binomial distribution that we call the cluster model. Section 4 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 5 uses a chi-square test to compare the fitness of the three models. Section 6 concludes the paper.

## 2. THREE PROBABILITY MODELS FOR RARE EVENTS

We introduce three probability models in this section, each of which is used in Section 4 to fit outage statistics data.

### 2.1. 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

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

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

Suppose that 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, $k \geq 1$ becomes a rare event and can be approximated by the Poisson distribution [8]; that is,

$$\Pr(M = k) = \binom{N}{k} p^k (1-p)^{N-k} \approx \Pr(M = k|\lambda) = e^{-\lambda}\lambda^k/k!, \tag{3}$$

where $\lambda = np$, $k = 0,1,2,3,\ldots,\infty$. The Poisson distribution is sometimes called the distribution of rare events [8]. Both the binomial distribution and Poisson distribution assume that the element events (the failure of a individual) are independent; that is, 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.

### 2.2. 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

$$\Pr(T = t|p,M) = \binom{M+t-1}{t} p^M (1-p)^t,$$

$$t = 0,1,2,\ldots,0 \leq p \leq 1, M > 0. \tag{4}$$

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 3.

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

### 2.3. Power-Law Distribution [2,3,4,7]

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

$$\Pr(X = x|p) = \frac{x^{-p}}{\int x^{-p}\,dx}, \qquad p > 0, \tag{5}$$

where $p$ is a constant and $X$ can either be a continuous variable or a discrete variable. If $x$ is a discrete variable, the denominator is replaced by $\sum x^{-p}$. Expression (5) is a proper probability density function (p.d.f.) if the sample space is limited to a finite region, since, in this case, its integral over the entire probability space is 1.0. If the sample space is infinite, $0 < x < \infty$, $p$ must be greater than 1.0 to be a proper p.d.f. 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 $-p/\int x^{-p}\,dx$; that is,

$$\log P(X = x) = \frac{-p\log(x)}{\int x^{-p}\,dx}. \tag{6}$$

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

## 3. 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 [10].

The 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

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

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 [10].

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 $(r+1)$st 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 $n \to \infty$, $m \to \infty$, 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 (3); that is,

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

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 parameters $p$ and $n$. The limiting distribution of $K$ with $n \to \infty$, $m \to \infty$, and $n/m = \lambda$ is Poisson$(\lambda)$. If we let $\alpha \to 0$, the Poisson distribution is derived as follows:

$$\begin{aligned} \lim_{\alpha \to 0} P\left(k \,\middle|\, \alpha^{-1}, \frac{\lambda}{\lambda + \alpha^{-1}}\right) &= \lim_{\alpha \to 0} \binom{\alpha^{-1} + k - 1}{k} \left(\frac{\lambda}{\lambda + \alpha^{-1}}\right)^k \left(\frac{\alpha^{-1}}{\lambda + \alpha^{-1}}\right)^{\alpha^{-1}} \\ &= \frac{e^{-\lambda} \lambda^k}{k!}. \end{aligned} \tag{8}$$

Here, the number $(N)$ of circuits tripped in each event must be greater than one. We reparameterize (7) by $Y = X + 1$ so that the sample space of random variables is $\{1, 2, 3, \ldots\}$. This is necessary because the sample space of contingencies contained in our dataset does not include the event $k = 0$ (loss of no elements). We also reparameterize $\lambda$ 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

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

$$\text{Cluster}(Y=y|\mu,\alpha)=\begin{cases}\dbinom{\alpha^{-1}+y-2}{y-1}\left(\dfrac{\mu-1}{\mu-1+\alpha^{-1}}\right)^{y-1}\left(\dfrac{\alpha^{-1}}{\mu-1+\alpha^{-1}}\right)^{\alpha^{-1}} & \text{if } \alpha>0,\mu>1, \text{ and } y=1,2,3,\ldots \\ \dfrac{e^{-(\mu-1)}(\mu-1)^{y}}{(y-1)!} & \text{if } \alpha=0,\mu>1, \text{ and } y=1,2,3,\ldots.\end{cases} \tag{9}$$

We call $\alpha$ the *affinity factor*. This value captures the tendency of the power system to have a cascading event, as we will show later in the paper. To compare the shapes of Cluster p.d.f.s having different $\alpha$'s with the power-law p.d.f., Table 1 summarizes convergence rates when the random variable (an index representing the size of contingencies) approaches infinity.

From Table 1, we see that when $x \to \infty$ (i.e., when events became large or "rare"), we have

$$0<\frac{\mu-1}{\mu-1+\alpha_1^{-1}}<\frac{\mu-1}{\mu}<\frac{\mu-1}{\mu-1+\alpha_2^{-1}}<1, \tag{10}$$

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

$$\{\text{Poisson}\}>\{\text{Cluster}(1>\alpha>0)\}>\{\text{Geometric}\}>\{\text{Cluster}(\alpha>1)\}$$
$$>\{\text{Power law}\}.$$

The convergence rate is actually an index representing the heaviness of the p.d.f.'s tail, or the likelihood of rare events. The higher the convergence rate, the less likely a rare event occurs. For the Poisson distribution, since it assumes the independence of events, it converges faster than any of the five distributions described in Table 1. The convergence rate of the cluster distribution depends on $\alpha$, which is just a parameter showing the interdependency of events, or the tendency of clustering. The larger the $\alpha$, the slower the convergence and the heavier the tail of the cluster distribution.

## 4. CLUSTER MODEL APPLIED TO OUTAGE STATISTICS

The data we analyzed include the total number of elements lost in each contingency in North America from 1965 to 1985 [1], as indicated in Table 2. The last two columns give a summary by voltage level. According to [1], the data reported in Table 2 adhere to the following:

- Each individual component tripping in a multicomponent outage event must occur within a 1-min 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.

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

**Table 1.** Comparing Convergence Rate with Cluster and Power-Law p.d.f. Shapes

| p.d.f. Family | Poisson($\lambda$) 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 $x$ versus $\Pr(x)$ | CC | CC | CC | CC | CC |
| Shape of $x$ versus $\log \Pr(x)$ | CC | CC | SL | CV | CC |
| Shape of $\log(x)$ versus $\log \Pr(x)$ | CV | CV | CV | CV | SL |

*Note:* CV = convex; CC = concave; SL = straight line.

**Table 2.** High-Order Transmission Outages Statistics

| Contingency Type | No. of Contingences by Line Voltage Levels | | | | Total No. | Percentage |
|---|---|---|---|---|---|---|
| | 230 kV | 345 kV | 500 kV | 765 kV | | |
| N-1 | 3,320 | 5,807 | 721 | 295 | 10,143 | 89.85% |
| 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 | 0 | 1 | 0 | 1 | 2 | 0.02% |
| N-7 | 3 | 1 | 0 | 0 | 4 | 0.04% |
| N-8 | 2 | 0 | 0 | 0 | 2 | 0.02% |

*Source*: Reference [1].

We draw the relationship of $\Pr(X = x)$ versus $x$ in Figures 1–3. The shapes (concave or convex) of Figures 1–3 match the cluster model in the third column of Table 1, which suggests that a cluster model with $1 > \alpha > 0$ could be better than others.

### 4.1. Maximum Likelihood Estimation

We fit the three probability models Poisson, cluster, and power law to the data in Table 2. We use the maximum likelihood estimation (MLE) to estimate the parameter(s) for each model. In MLE [5], 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 $\theta_i$'s are the model parameters to be estimated, the joint p.d.f. is

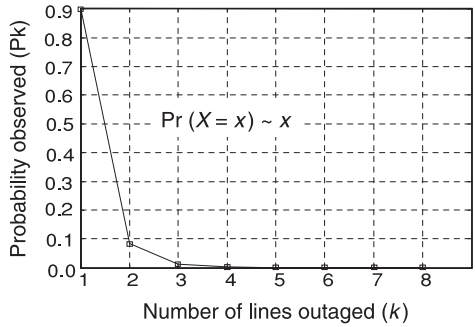



**Figure 1.** $\Pr(k)$ versus $k$.

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

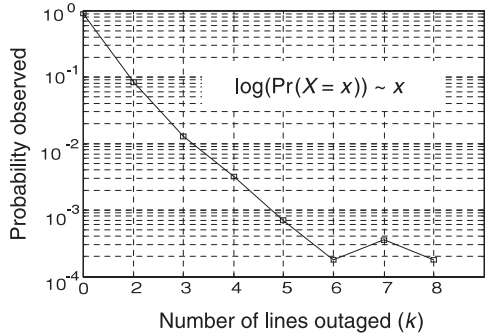



**FIGURE 2.** $\log\{\Pr(k)\}$ versus $k$.

$$\Pr(X_1 = x_1, \ldots, X_n = x_n | \theta_1, \ldots, \theta_m) = \prod_{i \in \{1,2,\ldots,n\}} f(x_i | \theta_1, \ldots, \theta_m). \tag{11}$$

The ML approach defines a likelihood function

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

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

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

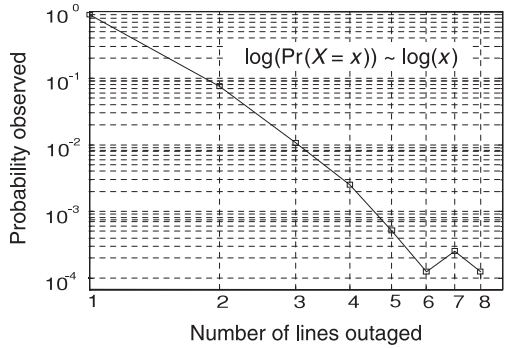



**FIGURE 3.** $\log\{\Pr(k)\}$ versus $\log(k)$.

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

Define $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ and $\underline{x} = (x_1, x_2, \ldots, x_n)$. The $\hat{\underline{\theta}}$ that maximizes $L(\underline{\theta}|\underline{x})$ is called an MLE of the parameter $\underline{\theta}$. It should be noted that $\hat{\underline{\theta}}$ 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.

### 4.2. Estimating the Parameters of the Poisson Distribution

If we assume that our data follow a Poisson distribution, then only parameter $\lambda$ needs be estimated. The Poisson distribution is a special case of the negative binomial in the form of (7), where $\alpha = 0$ indicates that 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 $\{0,1,2,3,\ldots\}$. The likelihood function is

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

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

$$\partial \log L(\lambda|N_0, N_1, \ldots, N_K, \ldots)/\partial\lambda = 0. \tag{15}$$

Substitution of (14) in (15) results in (16), the MLE of $\lambda$:

$$\hat{\lambda} = \sum_{k=0}^{\infty} kN_k \Big/ \sum_{k=0}^{\infty} N_k, \tag{16}$$

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

$$\hat{\lambda} = \sum_{k=1}^{8} (k-1)N_k \Big/ \sum_{k=1}^{8} N_k \approx 0.1262. \tag{17}$$

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

$$\Pr(Y = k) = e^{-\hat{\lambda}}\hat{\lambda}^{k-1}/(k-1)! = e^{-0.1262}0.1262^{k-1}/(k-1)! \tag{18}$$

### 4.3. Estimating the Parameters of the Cluster Model

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

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

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

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

$$\begin{aligned}
&\log L(\alpha,\mu|N_1,N_2,\ldots,N_K,\ldots)\\
&= \sum_{k\in(1,2,\ldots,)} \log\left\{\binom{\alpha^{-1}+k-2}{k-1}\left(\frac{\mu-1}{\mu-1+\alpha^{-1}}\right)^{k-1}\left(\frac{\alpha^{-1}}{\mu-1+\alpha^{-1}}\right)^{\alpha^{-1}}\right\}^{N_k}\\
&= \sum_{k\in(1,2,\ldots,)}\left\{N_k\log\binom{\alpha^{-1}+k-2}{k-1}\right\}\\
&\quad+\log\left(\frac{\mu-1}{\mu-1+\alpha^{-1}}\right)\left(\sum_{k\in(1,2,\ldots,)}[(k-1)N_k]\right)\\
&\quad+\log\left(\frac{\alpha^{-1}}{\mu-1+\alpha^{-1}}\right)^{\alpha^{-1}}\left(\sum_{k\in(1,2,\ldots,)}N_k\right).
\end{aligned} \tag{20}$$

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

$$\begin{aligned}
\partial\log L(\alpha,\mu|N_1,N_2,\ldots,N_K,\ldots)/\partial\alpha|_{(\alpha,\mu)=(\hat{\alpha},\hat{\mu})} &= 0,\\
\partial\log L(\alpha,\mu|N_1,N_2,\ldots,N_K,\ldots)/\partial\mu|_{(\alpha,\mu)=(\hat{\alpha},\hat{\mu})} &= 0.
\end{aligned} \tag{21}$$

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

$$\hat{\mu} = \sum_{k\in\{1,2,\ldots\}} kN_k \Big/ \sum_{k\in\{1,2,\ldots\}} N_k. \tag{22}$$

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 $\log(L(\alpha,\mu|N_1,N_2,\ldots,N_k,\ldots))$ directly using the contour graph function in Matlab. It is much easier to understand this case with the aid of graphs. Substituting the $k$'s and the $N_k$'s into the likelihood function of (20), we get the graphs of the likelihood function $\log(L(\alpha,\mu|N_1,N_2,\ldots,N_k,\ldots))$, as Figures 4 and 5. The MLE of $\alpha$ and $\mu$ are approximately

$$\hat{\mu}\approx 1.12623, \qquad \hat{\alpha}\approx 3.08. \tag{23}$$

We can also get the estimate of $\mu$ directly, using (22):

$$\hat{\mu} = \sum_{k=1}^{8} kN_k \Big/ \sum_{k=1}^{8} N_k = 1.1262907\ldots. \tag{24}$$

This is very close to our estimate of $\mu$ from the contour plot in Figure 4, which is evidence of the correctness of our method.

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

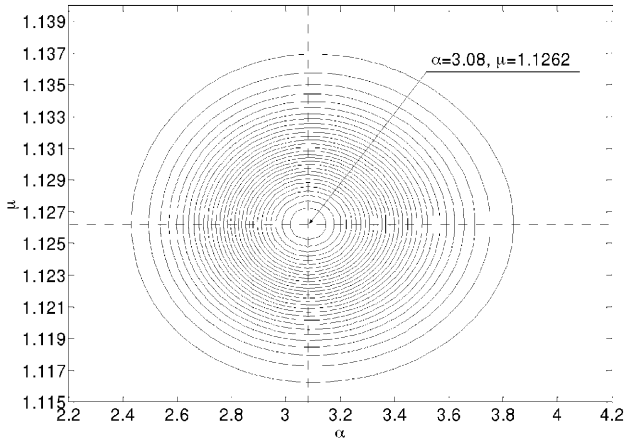



**FIGURE 4.** Contour plot of maximum likelihood function (20).

### 4.4. Estimating the Parameters for the Power-Law Distribution

The likelihood function for the power-law distribution is

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

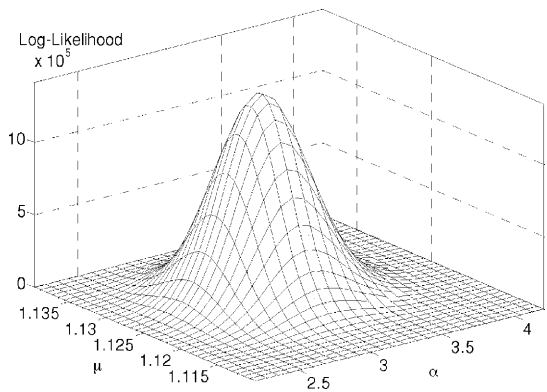



**FIGURE 5.** Mesh plot of maximum likelihood function (20).

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

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

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

Because 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 1), we use the least square curve fitting to estimate the approximate range of the slope $-p/\int x^{-p}\,dx$. 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 Figure 6. Using the bisection method to search the maximum, we find $\hat{p} = 3.75$. Then the estimated power-law p.d.f. for the data in the last two columns of Table 2 is

$$P(X = x | p = 3.75) = x^{-3.75} \Big/ \sum_{k=1}^{\infty} k^{-3.75} = 0.9080 x^{-3.75}. \tag{27}$$

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

$$\Pr(X = x | \alpha = 3.08, \mu = 1.1263) = \binom{3.08^{-1} + x - 2}{x - 1}\left(\frac{1.1263 - 1}{1.1263 - 1 + 3.08^{-1}}\right)^{x-1} \times \left(\frac{3.08^{-1}}{1.1263 - 1 + 3.08^{-1}}\right)^{3.08^{-1}} \tag{28}$$

for the cluster model;

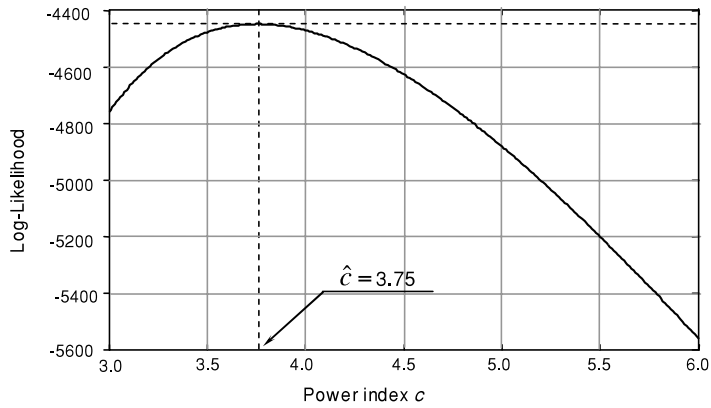



**FIGURE 6.** Plot of maximum likelihood function (26).

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

**TABLE 3.** Comparing the Fitness of Three Different Probability Models for the Distribution of Observed Multiple Line Outages

| Contingency Type | $K$ | Observed Number | Observed Probability | Cluster | Power Law | Poisson |
|---|---|---|---|---|---|---|
| N-1 | 1 | 10,143 | 0.8985 | 0.8989 | 0.9080 | 0.8814 |
| N-2 | 2 | 951 | 0.0843 | 0.0817 | 0.0672 | 0.1113 |
| N-3 | 3 | 143 | 0.01267 | 0.01515 | 0.01467 | 0.007022 |
| N-4 | 4 | 36 | 0.003189 | 0.00329 | 0.004979 | 0.000296 |
| N-5 | 5 | 8 | 0.0007087 | 0.000766 | 0.002154 | 9.324E−06 |
| N-6 | 6 | 2 | 0.000177 | 0.000185 | 0.001086 | 2.354E−07 |
| N-7 | 7 | 4 | 3.543E−04 | 0.000046 | 0.000609 | 4.95E−09 |
| N-8 | 8 | 2 | 0.000177 | 1.166E−05 | 0.000369 | 8.93E−11 |

$$\Pr(X = x|\lambda = 0.1263) = e^{-0.1263}0.1263^{x-1}/(x-1)! \tag{29}$$

for the Poisson model;

$$\Pr(X = x|p = 3.75) = x^{-3.75}\Big/\sum_{k=1}^{\infty} k^{-3.75} = 0.9080x^{-3.75} \tag{30}$$

for the power-law model.

Evaluating the above three expressions for $k = \{1,2,3,4,5,6,7,8\}$, we obtain the results shown in Table 3. These results are plotted in Figure 7.

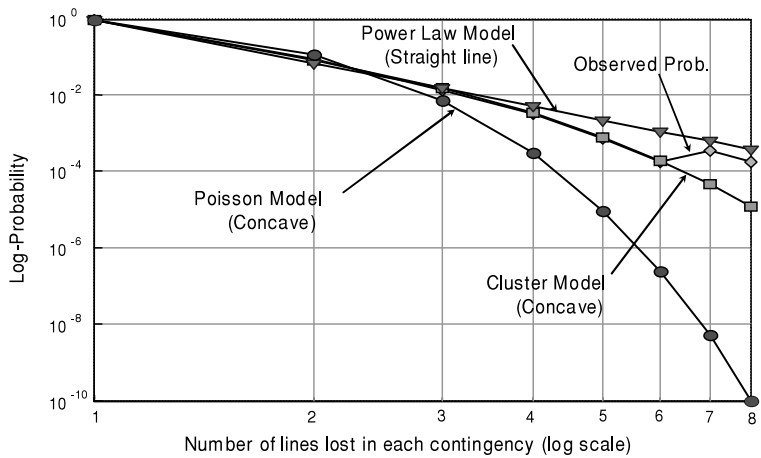



**FIGURE 7.** The log-log plot of p.d.f.s in (28)–(30).

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

By inspecting Figure 7, one concludes that the cluster model (the curve with squares) is superior to the other two models, as it fits almost perfectly for $k = 1,\ldots,7$. The cluster model and the power-law model agree for $k = 8$, for which there are 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 $10^{-5}$.

## 5. FITNESS TEST OF THREE DIFFERENT PROBABILITY MODELS

Figure 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 [9], 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 $p_i$, $i = 1,2,3,\ldots,k$, 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

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

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 p.d.f. as in (31) and define*

$$\chi^2 = \sum_{i=1}^{k} (N_i - np_i)^2/np_i, \tag{32}$$

*then when* $n \to \infty$, $\chi^2$ *follows the chi-square distribution* $\chi^2(k-1)$.

We can see from (32) that the statistic $\chi^2$ is an index showing how much the samples deviate from the polynomial distribution to be tested. The larger the statistics $\chi^2$, 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 $(1,2,3,\ldots)$ and it is an infinite set, we separate the sample space into $k$ exclusive sets denoted as $S_i = 1,2,3,\ldots,k$. Suppose $X$ is a random variable and its p.d.f. is $f(x) = \Pr(X = x), x \in \{1,2,3,\ldots\}$. We draw a total of $n$ samples of $X$ from p.d.f. $f(x)$ 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 $N_i$, $i = 1,2,\ldots,k$, follow the polynomial distribution of (31). The statistics $\chi^2$ defined in (32) follow the $\chi^2(k-1)$ distribution. If $\chi^2$ 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 five exclusive sets: $S_1 = \{1\}$, $S_2 = \{2\}$, $S_3 = \{3\}$, $S_4 = \{4\}$, and $S_5 = \{5\}$. The reason we

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

**TABLE 4.** $\chi^2$ Test Results

| | | Cluster | | Poisson | | Power Law | |
|---|---|---|---|---|---|---|---|
| $i$ | $n_i$ | $p_i$ | $p_i n_i$ | $p_i$ | $p_i n_i$ | $p_i$ | $p_i n_i$ |
| 1 | 10,143 | 0.899 | 10,147 | 0.881 | 9,950 | 0.908 | 10,250 |
| 2 | 951 | 0.082 | 922.1 | 0.113 | 1,256 | 0.067 | 759 |
| 3 | 143 | 0.015 | 172.0 | 0.007 | 79 | 0.0147 | 165.6 |
| 4 | 36 | 0.0033 | 37.1 | 0.0003 | 3.33 | 0.00498 | 56.2 |
| ≧5 | 17 | 0.00101 | 11.38 | 9.56E−06 | 0.11 | 0.00422 | 47.6 |
| $\chi^2 = \sum_{i=1}^{k} (N_i - np_i)^2/np_i$ | | 7.4 | | 2,788 | | 81 | |
| Prob.$\{\chi^2 > \chi^2(k-r-1)\}$ | | 0.0298 | | $\ll 10^{-321}$ | | $< 10^{-42}$ | |

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 that 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 $\chi^2$ distribution by one for each estimated parameter. The rule is: If there are a total of $r$ estimated parameters, the freedom number of the $\chi^2$ distribution is $k - r - 1$. For the Poisson model, the parameter $\lambda$ is an estimate, so the chi-square distribution has freedom number $5 - 1 - 1 = 3$. For the cluster model, $\alpha$ and $\mu$ 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 4. The last row of Table 4 lists the probability of getting a sample deviation larger than observed, assuming that the sample comes from a certain probability model. The cluster model is far more fit than the other two.

## 6. 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 $\alpha$, to capture the tendency of component outages 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 $\alpha = 0$. When $\alpha$ is very large, the power-law and the cluster models both exhibit similar behavior in convergence speed as the event becomes very large. In our application to real data,

https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press

we observed that the Poisson model underestimates rare-event probabilities, the power-law model overestimates them, and the cluster model captures them very well. 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 levels. 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 online to develop defense strategies as real-time conditions change. Work to this effect has been submitted for publication.

### *References*


1. Adler, R., Daniel, S., Heising, C., Lauby, M., Ludorf, R., & White, T. (1994). An IEEE survey of US and Canadian overhead transmission outages at 230 kV and above. *IEEE Transactions on Power Delivery* 9(1): 21–39.
2. Bak, P. (1996). *How nature works*. New York: Springer-Verlag.
3. Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of 1/f noise. *Physical Review Letters* 59: 381–384.
4. Carreras, B., Newman, D., Dobson, I., & Poole, A. (2000). 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.
5. Casella, G. & Berger, R. (2002). *Statistical inference*. New York: Wadsworth.
6. Chen, Q. & McCalley, J. (2005). Identifying high-risk $N - k$ contingencies for on-line security assessment. *IEEE Transactions on Power Systems* 20(2): 823–834.
7. Dobson, I., Chen, J., Thorp, J., Carreras, B., & Newman, D. (2002). Examining criticality of blackouts in power system models with cascading events. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, pp. 803–812.
8. Falk, M., Husler, J., & Reiss, R. (1994). *Laws of small numbers: Extremes and rare events*. Basel: Birkhauser-Verlag, p. 4.
9. Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophy Magazine* 50: 157–172.
10. Thompson, W.A., Jr. (1988). *Point process models with applications to safety and reliability*. London: Chapman & Hall.


https://doi.org/10.1017/S0269964805050321 Published online by Cambridge University Press