**Title**
Statistical Analysis of Wildfire Count and Size Distributions



**Author**
Huang, Yu

**Publication Date**
2019

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

# Statistical Analysis of Wildfire Count and Size Distributions

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Applied Statistics

by

Yu Huang

2019

## Abstract of the Thesis

Statistical Analysis of Wildfire Count and Size Distributions

by

Yu Huang

Master of Science in Applied Statistics

University of California, Los Angeles, 2019

Professor Frederic R Paik Schoenberg, Chair

Forest fires are one of the biggest ecological disasters in Canada. Counts and sizes of fires vary substantially from year to year. In this study, the data is collected from Northwest Territories in Canada. We assume that fire counts appear to follow the Gamma-Poisson distribution, and fire sizes approximately follow the Gamma-Exponential distribution. The Maximum Likelihood Estimation and Random Search are used to estimate the parameters of two models. Identifiability issues regarding parameters in the two models are explored. The Kolmogorov–Smirnov test is used to check for goodness of fit. For fire sizes data, although the Kolmogorov–Smirnov test shows a low p-value, by plotting theoretical and empirical distribution, we can see that the Gamma-Exponential distribution fits adequately.

The thesis of Yu Huang is approved.

Vivian Lew

Hongquan Xu

Frederic R Paik Schoenberg, Committee Chair

University of California, Los Angeles

2019

*To my mother . . .*

*who—among so many other things—*

*saw to it that I learned to touch-type*

*while I was still in elementary school*

# Table of Contents

# List of Figures

# Acknowledgments

I would to thank Prof. Schoenberg for his generous support and help with all of the advices; Steven G. Cumming for helping me to understand fire distributions and providing me with the data. I also thank Vivian Lew and Prof. Xu for all of the comments that greatly improved my thesis.

# CHAPTER 1

# Introduction

As population increases in the world, human activities are affected directly when natural hazards happened(Grid-Arendal UNEP, 2002). Natural hazards cause hundreds thousands deaths and threaten people's daily life(Guha-Sapir et al., 2013). Fire disasters have become studies for numerous scholars(Wang et al., 2005; Cheng and Wang, 2008). In order to manage public expectations and to assist forest managers with wildfire preparedness, it is important to be able to estimate the distributions of wildfire sizes and counts as accurately as possible

The dataset that is analyzed in this paper is from Northwest Territories in Canada. It contains records for the recorded lightning-caused fires.

The Poisson distribution is widely used to investigate count data. In some previously studies, the Poisson distribution was used to estimate the distributions of fires counts and fire sizes(Mandallaz and Ye, 1997). However, sometimes this may not be very accurate because most of the data are heavily skewed, such as the data we used in this study. If we assume the distribution is the Poisson distribution, then the variance of data should be the same as the mean. However, variance is often much larger than the mean for the real data, meaning that the data in real world are often overdispersed. In this study, we will use mixture distributions to solve the issue with overdispersed data.

Many different models have also been used to investigate the distribution of forest fires. Some fire sizes distributions have been investigated with Pareto distribution(Robertson, 1972). In previous research, models have been used to deal with over-dispersion data, such as Negative Binomial and other compound Poisson models(Bliss and Fisher, 1953; Hinde, 1982). The assumptions for this fire data are that fire counts follow a Gamma-Poisson distribution, and fire sizes follow a Gamma-Exponential distribution. Lawless (1987) showed that the negative binomial distribution can arise from a mixture of the Gamma and Poisson distributions, and a mixture of Gamma and exponential distirbutions can result in the Pareto distribution.

The structure of the rest of this thesis is as follows. The second chapter introduces the maximum likelihood method to estimate the unknown parameters in Gamma-Poisson and Gamma-Exponential distribution, when one has observed data. Analysis on fire counts and fire sizes distribution includes fitting the proposed distribution to data and check for goodness of fit by using Kolmogorov–Smirnov test and graph method. The third chapter focuses on the analysis on distribution of fire counts, with the assumption of Gamma-Poisson distribution. In the fourth chapter, we have the analysis on distribution of fire sizes, with the assumption of Gamma-Exponential distribution. Conclusions and areas for further improvements are discussed in chapter 5.

# CHAPTER 2

# Maximum Likelihood Estimates for Gamma-Poisson and Gamma-Exponential Distributions

## 2.1 Maximum Likelihood Estimation

Maximum Likelihood Estimation is a commonly used method that estimates the unknown parameters of a distribution by using observations, and is advocated by numerous scholars(Laird, 1978; Cohen, 1965). Maximum Likelihood Estimation aims to maximize the likelihood function in order to find the unknown parameters. Suppose $x_1, x_2, x_3, x_4, ..., x_n$ is a random sample of size n from a distribution that has parameter $\theta$. The joint probability density of n random variables is:

$$L = f(x_1, x_2, ..., x_n; \theta)$$

In this case, $L$ is the likelihood function. In principle, Maximum Likelihood Estimation selects the value of $\theta$ to maximize the likelihood function $L$. If $X_1, X_2, ..., X_n$ are independent, the likelihood function can be expressed as:

$$L = f(x_1, x_2, ..., x_n; \theta) = f(x_1; \theta) \times f(x_2; \theta) \times ... \times f(x_n; \theta)$$

Now, we can maximize $L$ with respect to $\theta$. Usually, we can make the derivation easier if we take log of $L$. Then, we take the derivative of the loglikelihood function with respect to $\theta$ and solve for $\theta$.

## 2.2 Gamma-Poisson Distribution

Because real fire counts are overdispersed, the distribution for such data have been modeled by such distributions as the negative binomial, or Gamma-Poisson distribution(Lindén and Mäntyniemi, 2011). In this study, let $n_i$ denote the fire numbers in year $i$ in a particular region. Thus, the number of fires is modeled as a Poisson random variable with a mean that is in turn variable from year to year, and is some multiple of a variable drawn from a Gamma distribution. A possible justification for such a model may be that the variable $\nu_i$ represents the effect of the mean temperature in the region during year $i$. This mean temperature may vary substantially from year to year, and conditionally, given this temperature, the number of wildfires may be modeled as Poisson. Our assumption is that:

$$n_i \sim poisson(\lambda_i)$$

$$\lambda_i = \lambda \nu_i$$

$$\nu_i = \Gamma(1, \alpha)$$

The probability density function of $n_i$ given $\lambda = \lambda\nu_i$ is:

$$P(n_i|\lambda_i = \lambda\nu_i) = \frac{(\lambda\nu_i)^{n_i}}{n_i!}e^{-\lambda\nu_i}$$

$$\because f(\nu_i) = \alpha e^{-\alpha\nu_i}$$

$$\therefore P(n_i) = \int_0^\infty \frac{(\lambda\nu_i)^{n_i}}{n_i!}e^{-\lambda\nu_i}\alpha e^{-\alpha\nu_i}d\nu_i$$

$$= \frac{\lambda_i^n}{n_i!}\alpha\int_0^\infty \nu_i^{n_i}e^{-(\alpha+\lambda)\nu_i}d\nu_i$$

$$\because \int_0^\infty x^{\alpha-1}e^{-\beta x}dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$$

$$\therefore = \frac{\lambda^{n_i}\alpha}{n_i!}\frac{\Gamma(n_i+1)}{(\alpha+\lambda)^{n_i+1}}$$

$$= \frac{\lambda^{n_i}\alpha}{n_i!}\frac{n_i!}{(\alpha+\lambda)^{n_i+1}}$$

$$= (\frac{\lambda}{\alpha+\lambda})^{n_i}(\frac{\alpha}{\alpha+\lambda})$$

$$\therefore Ł = \left(\frac{\alpha}{\alpha+\lambda}\right)^N\prod_{i=1}^N\left(\frac{\lambda}{\alpha+\lambda}\right)^{n_i}$$

where $N$ is the number of observations.

Thus, the loglikelihood function is obtained:

$$l = \sum n_i ln\lambda - \sum n_i ln(\alpha+\lambda) + Nln\alpha - Nln(\alpha+\lambda)$$

$$= \sum n_i ln\lambda - \left(\sum n_i + N\right) ln(\alpha+\lambda) + Nln(\alpha)$$

Then, we take the derivatives of loglikelihood function and get:

$$\frac{\partial l}{\partial \lambda} = \frac{\sum n_i}{\lambda} - \frac{(\sum n_i + N)}{\alpha + \lambda} = 0$$

$$\Rightarrow \alpha \sum n_i + \lambda \sum n_i = \lambda \sum n_i + \lambda N$$

$$\Rightarrow \lambda = \frac{\alpha \sum n_i}{N}$$

$$\frac{\partial l}{\partial \alpha} = \frac{N}{\alpha} - \frac{(\sum n_i + N)}{\alpha + \lambda} = 0$$

$$\Rightarrow \alpha N + \lambda N = \alpha \sum n_i + \alpha N$$

$$\Rightarrow \alpha = \frac{\lambda N}{\sum n_i}$$

The resulting analytical solutions for $\lambda$ and $\alpha$ are from the same identity. Therefore, there are more than one solution for $\lambda$ and $\alpha$. Note that Negative Binomial distribution is also a Gamma mixture of Poisson(Joe and Zhu, 2019). Therefore, one can easily see that $n_i$ also follows Negative Binomial distribution with $r = 1$ and $p = \frac{\lambda}{\alpha+\lambda}$. Hence, for Gamma-Poisson distribution in this paper, we have:

$$E[n_i] = \frac{\lambda}{\alpha}$$

$$var[n_i] = \frac{\lambda}{\alpha}\left(1 + \frac{\lambda}{\alpha}\right)$$

Although the exact values for $\lambda$ and $\alpha$ are unidentifiable, we can still use the relationship to fit the distribution to data and check for goodness of fit. The likelihood function is maximized when $\frac{\lambda}{\alpha} = E[n_i]$, which is the mean of $n_i$. This relationship is very handy for analyzing the data. After deriving the analytical solution for Gamma-Poisson distribution, we now proceed to find the maximum likelihood estimates for the parameters in Gamma-Exponential distribution. The procedures are pretty similar.

## 2.3 Gamma-Exponential Distribution

The most common distribution used for fire sizes seems to be the Pareto( Schoenberg, Peng, and Woods, 2003). Pareto distribution could also be written as a Gamma-Exponential mixture. Hence, Let $x_{ij}$ be the size of $j^{th}$ fire in year $i$, we have

$$x_{ij} \sim Exp(\mu_i)$$

$$\mu_i = \mu\xi_i$$

$$\xi_i \sim \Gamma(1, \alpha)$$

In a similar way, the probability density function of $X_{ij}$ can be obtained by:

$$\begin{aligned}
P(x_{i1}, x_{i2}, x_{i3}, ..., x_{im_i}|\mu, \alpha) &= \int P(x_{i1}, x_{i2}, x_{i3}, ..., x_{im_i}|\mu\xi_i)P(\xi_i) \\
&= \int \prod_{j=1}^{m_i} \mu\xi_i exp(-\mu\xi_i x_{ij}) \times \alpha exp(-\alpha\xi_i) d\xi_i \\
&= \int \mu^{m_i}\xi^{mi} exp(-\mu\xi_i \sum_{j=1}^{m_i} x_{ij})\alpha exp(-\alpha\xi_i) d\xi_i \\
&= \alpha\mu^{m_i} \int \xi_i^{m_i} exp[-(\mu \sum_{j=1}^{m_i} x_{ij} + \alpha)\xi_i] d\xi_i \\
&= \alpha\mu_{m_i} \frac{\Gamma(m_i + 1)}{\left(\mu \sum_{j=1}^{m_i} x_{ij} + \alpha\right)^{m_i+1}}
\end{aligned}$$

The likelihood function can be written as:

$$L = \prod_{i=1}^{N} P(x_{i1}, x_{i2}, x_{i3}, ..., x_{im_i}|\mu, \alpha)$$

$$= \prod_{i=1}^{N} \alpha\mu^{mi} \frac{\Gamma(m_i + 1)}{\left(\mu \sum_{j=1}^{m_i} x_{ij} + \alpha\right)^{m_i+1}}$$

After taking logs:

$$lnL = \sum_{i=1}^{N}[ln\alpha + m_i ln\mu + ln\Gamma(m_i + 1) - (m_i + 1)ln(\mu \sum_{j=1}^{m_i} x_{ij} + \alpha)]$$

$$= Nln\alpha + \sum_{i=1}^{N} m_i ln\mu + \sum_{i=1}^{N} ln\Gamma(m_i + 1) - \sum_{i=1}^{N}[(m_i + 1)ln(\mu \sum_{j=1}^{m_i} x_{ij} + \alpha)]$$

$$= Nln\alpha + tln\mu + c - \sum_{i=1}^{N}[(m_i + 1)ln(\mu X_i + \alpha)]$$

where $t = \sum_{i=1}^{N} m_i, c = \sum_{i=1}^{N} ln\Gamma(m_i + 1)$, and $X_i = \sum_{j=1}^{m_i} x_{ij}$. The meaning of each term here will be introduced in the next chapter.

In contrast with the fire counts distribution(Gamma-Poisson distribution), analytical solution is hard to find due to the complexity of loglikelihood function. In order to find the values of $\alpha$ and $\mu$, the optim function in R is very useful to find the parameters. The use of optim function will be explained in the next chapter as well.

# CHAPTER 3

# Fire Counts Distribution Analysis

The motivation for this study is both applied and theoretical. In this chapter, we will use the maximum likelihood estimates derived in section 2.1 to analyze the empirical fire counts data. The Gamma-Poisson distribution seems to be a good fit for our dataset. In future studies for fire counts dataset, we may apply similar statistical techniques used in this chapter to analyze other datasets and even simulate future fire data.

## 3.1 Likelihood Maximization

In this section, we will use the plot of negative loglikelihood function of Gamma-Poisson distribution to visualize where the minimum occurs. Recall from Chapter 2, where we have derived the following relationships:

$$E[n_i] = \frac{\lambda}{\alpha}$$

$$var[n_i] = \frac{\lambda}{\alpha}\left(1 + \frac{\lambda}{\alpha}\right)$$

A plot of $L$ function can visually check where the maximum value occurs. I also take negative value of loglikelihood function, so that we will need to find the minimum in the plot now. One can generate the negative loglikelihood function by:

```
l <- function(par, n) {
```

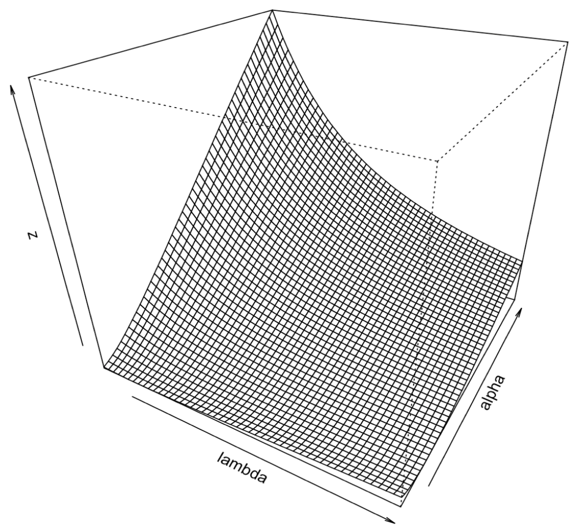


Figure 3.1: Plot of Negative Loglikelihood Function

```
  lambda <- par[1]
  alpha <- par[2]
  logl <- log(lambda)*sum(n) - (sum(n)+length(n))*log(alpha+lambda)
  +length(n)*log(alpha)
  return(-logl)
}
```

Figure 3.1 shows that the minimum value occurs at a straight line, in agreement with the analytical solution. Although $\alpha$ and $\lambda$ are not identifiable, the minimum of negative loglikelihood function can be achieved whenever $\frac{\lambda}{\alpha} = E[n_i]$. We may be able to identify the exact values by including a constraint on $\alpha$ or $\lambda$, but in this study we do not have such information. In following sections and codes, we denote $k = \frac{\lambda}{\alpha}$ for simplicity.

## 3.2 Goodness-of-fit Assessment

In many studies, Kolmogorov–Smirnov test has been introduced to check for goodness of fit for distributions(Massey Jr., 1951; Brown, Riccardo Barbieri, Ventura, Kass and Frank, 2002). K-S test is a non-parametric test to determine if two data samples come from the same distribution. In this section, Kolmogorov–Smirnov(K-S) Plot will be conducted to compare the theoretical cumulative probability function and empirical probability function. Figure 3.2 shows the K-S plot for fire counts of 1965 to 2008. The red 45-degree line means that the theoretical cumulative probability function is exactly the same as empirical probability function. The blue dashed line is 95 percent confidence interval, and green dashed line is 99 confidence interval. The 0.05 and 0.01 level significance for sample size greater than 35 are 1.36 and 1.63 (Massey Jr., 1951). If the black points fall into the confidence interval, then the sample and empirical distributions are not significantly different.

In Figure 3.2, overall the distribution fits fire counts data pretty well; however, there are some data points exceeding the confidence level, meaning that some of the data samples are not from the distribution we assumed. Figure 3.7 shows the histogram of fire counts for all years. The dashed blue line in Figure 3.7 is the theoretical distribution. We can see that the distribution fits data quite well, except for some departures from optimal fit. In order to investigate this issue further, I separate the fire counts into four time frames: fire counts from year 1965 to 1975, 1976 to 1985, 1986 to 1995 and 1996 to 2008. For each time frame, the K-S plot is used to test if the sample data follow the theoretical distribution. The value of $k$ should be re-calculate by taking the mean of fire counts for each time frame.

Figure 3.3 is the K-S plot for year from 1965 to 1975. All of points fall in 95 percent confidence interval, meaning that the Gamma-Poisson distribution we proposed fits the data very well. Same as year from 1996 to 2008(Figure 3.6), all of the points fall in 95 percent confidence interval. Therefore, our proposed distribution is well fitted to these two time frame. Figure 3.8 and Figure 3.11 show the histograms for these two time frames: although there are still some departures from optimal distribution, Gamma-Poisson fits empirical data quite well because variations in y-axis are smaller, showing the same results as K-S plots.

Figure 3.4 shows the K-S plot from 1976 to 1985. There are only few data fall out of 99 percent confidence interval, and those points are very close to confidence interval. Histogram in Figure 3.9 also shows that the theoretical distribution fits data quite adequately.

However, from 1986 to 1995, the histogram shows that there are large amount of departures from optimal distribution(Figure 3.10). The K-S plot also shows the same results because lots of points fall very far from confidence intervals. This may explain why the fit of the given distribution is quite poor during this time frame. Parametric model is very hard to capture local variations in data. There may be some missing information for year 1986 to 1995 that is not captured by our theoretical distribution. Overall, however, the Gamma-Poisson distribution fits well to our fire counts data.

A sample code to produce the K-S plot and histogram is following:

```
########### Construct Histogram versus Theoretical Distribution
    ##########

k_65_75 = mean(fc_65_75)
h <- hist(fc_65_75,breaks = 0:200, xlab ="Fire␣Number",ylab="Counts",
    main =
        "Histogram␣against␣Theoretical␣Distribution␣for␣1965␣to␣1975")
x = 0:200
y = (k_65_75/(1+k_65_75))^x * (1/(1+k_65_75))
N = length(fc_65_75)
lines(x,y*N,lty = 2, col=4,lwd=1.5)

########### Construct KS Plot and Confidence Interval #########

Fx_hat = Fx = cumsum(y)
ni = fc_65_75
N = length(fc_65_75)
for (i in 0:200){
Fx_hat[i+1] = sum(ni<=i)/N
}
upperconfidence_99 = Fx + 1.63/(N^(1/2))
lowerconfidence_99 = Fx - 1.63/(N^(1/2))
upperconfidence_95 = Fx + 1.36/(N^(1/2))
lowerconfidence_95 = Fx - 1.36/(N^(1/2))
```

```
plot(Fx , Fx_hat,pch=16, xlim=c(0,1), ylim=c(0,1),cex = .4, xlab = '
    Empirical␣CDF',
        ylab = 'Theoratical␣CDF', main = 'KS␣Plot␣with␣Parameter␣k␣=␣
            45.62␣for␣1965␣to␣1975')
lines(x=0:1,y=0:1, col='red',lty=5)
lines(x=Fx,y=upperconfidence_99, col ='green',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_99, col = 'green',cex = 1.5,lty=2)
lines(x=Fx,y=upperconfidence_95, col ='blue',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_95, col = 'blue',cex = 1.5,lty=2)
# Add a legend
legend("topleft", legend=c("95%␣Confidence␣Bounds", "99%␣Confidence␣
    Bounds","45␣Degree␣Line"), col=c("blue", "green", "red"), lty=c
    (2,2,5), cex=0.8)
```

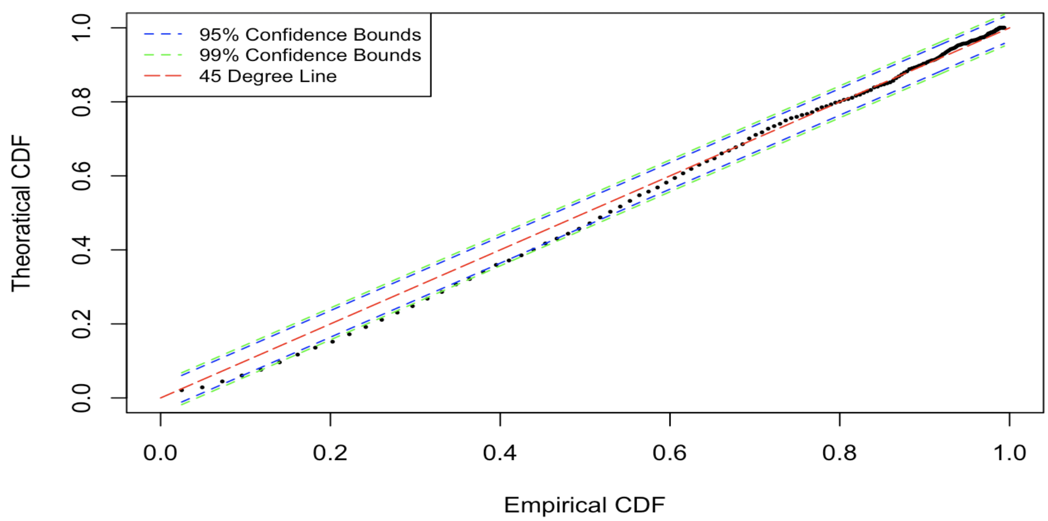


Figure 3.2: K-S Plot for all Years

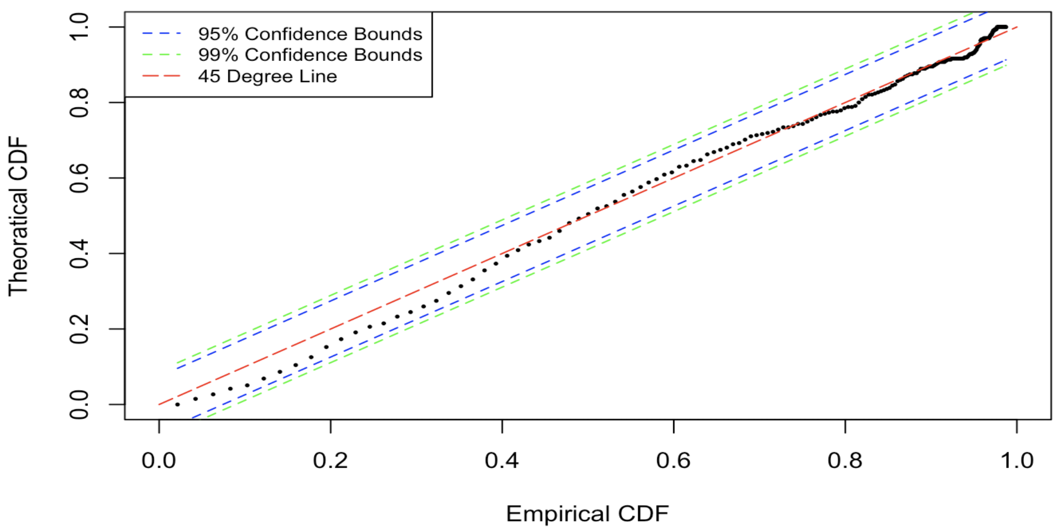


Figure 3.3: K-S Plot for 1965 to 1975

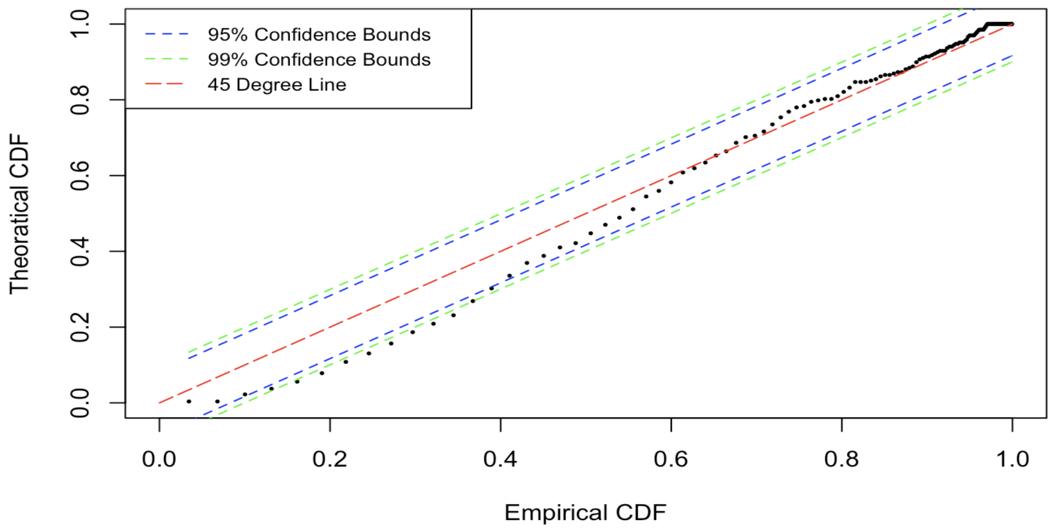


Figure 3.4: K-S Plot for 1976 to 1985

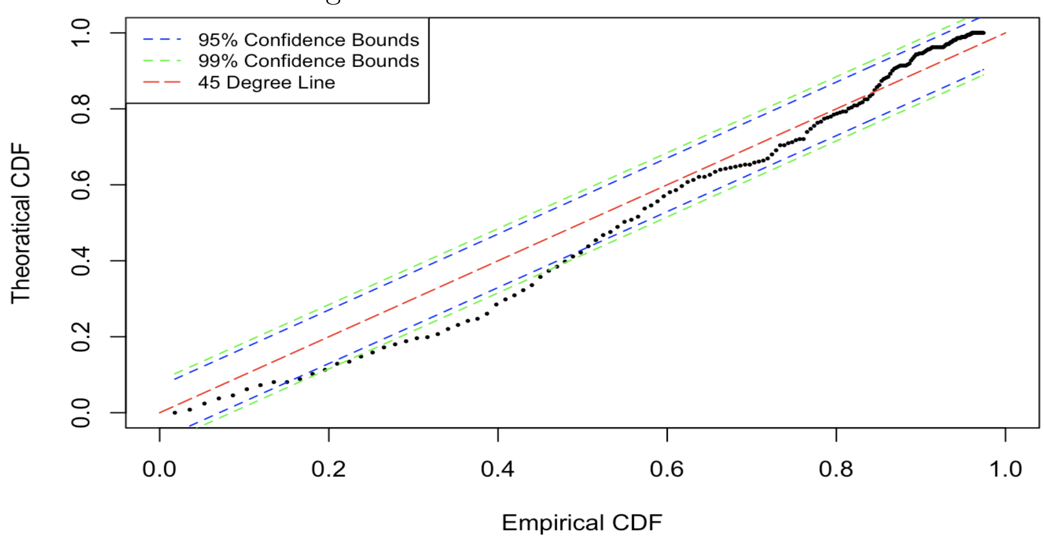


Figure 3.5: K-S Plot for 1986 to 1995

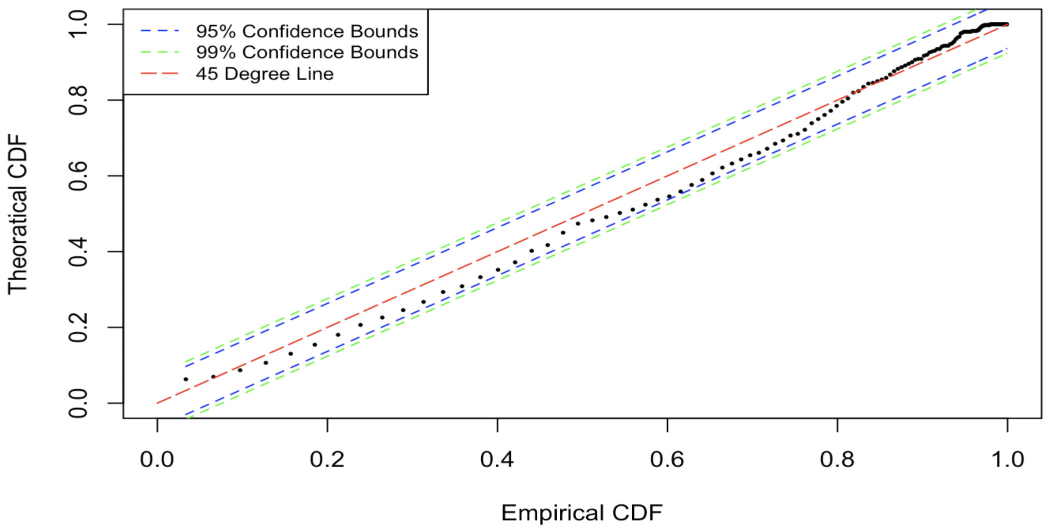


Figure 3.6: K-S Plot for 1996 to 2008

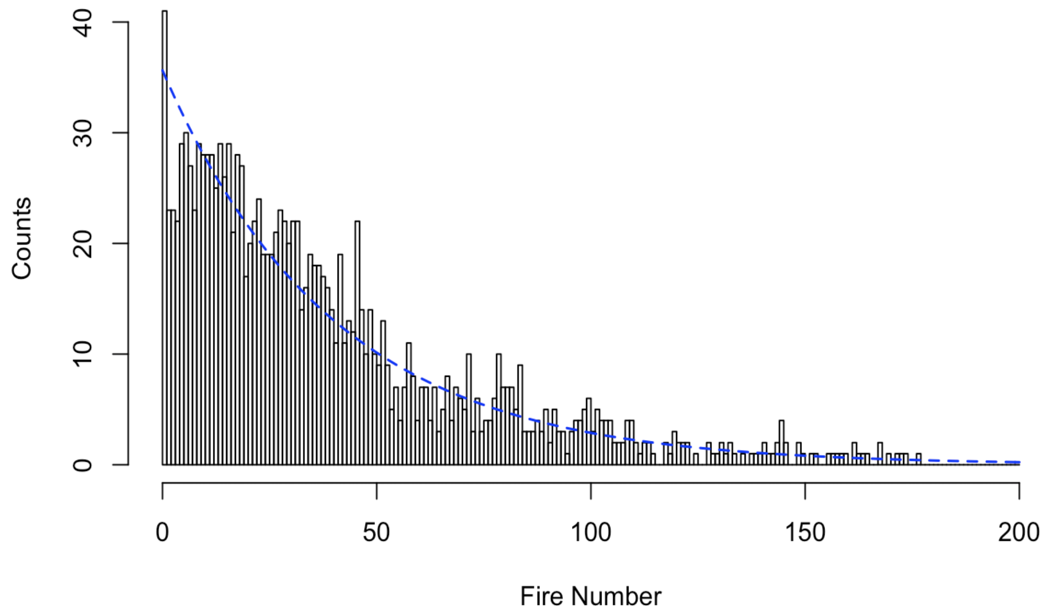


Figure 3.7: Fire Counts Histogram for all Years

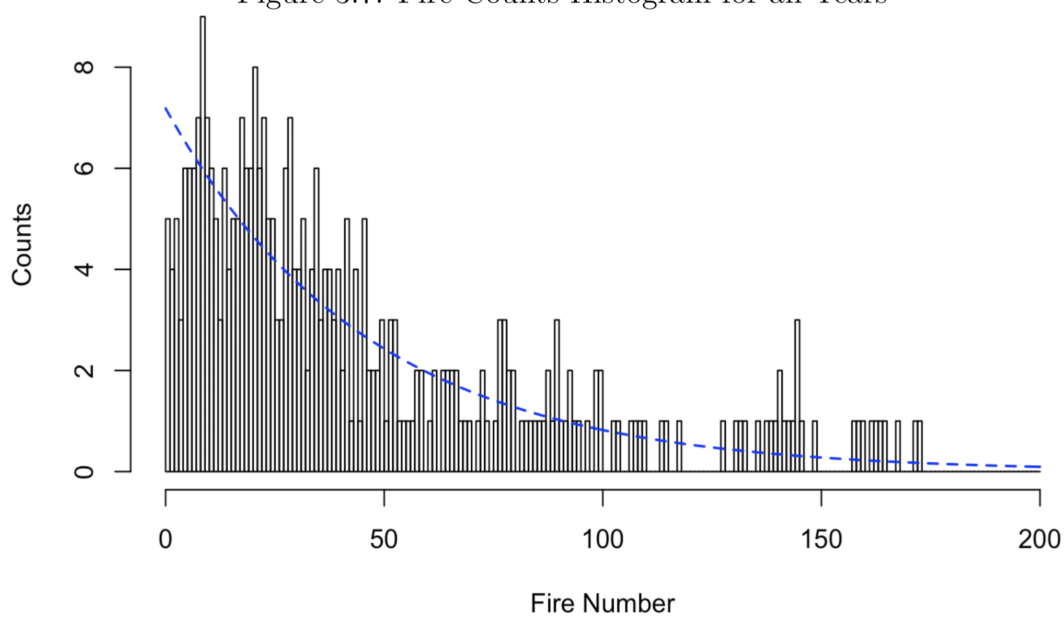


Figure 3.8: Fire Counts Histogram for 1965 to 1975

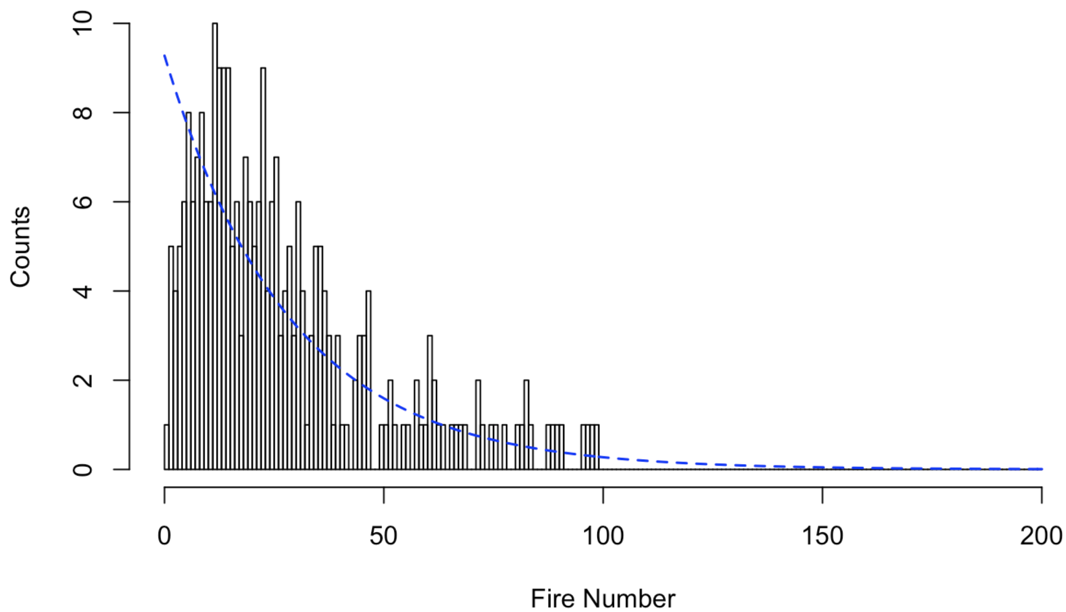


Figure 3.9: Fire Counts Histogram for 1976 to 1985

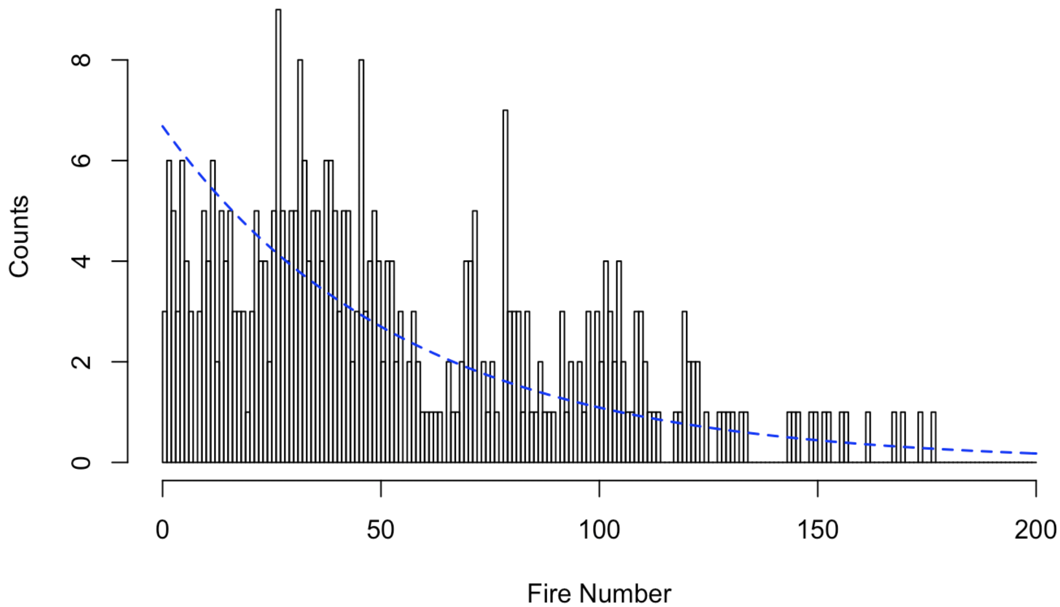


Figure 3.10: Fire Counts Histogram for 1986 to 1995

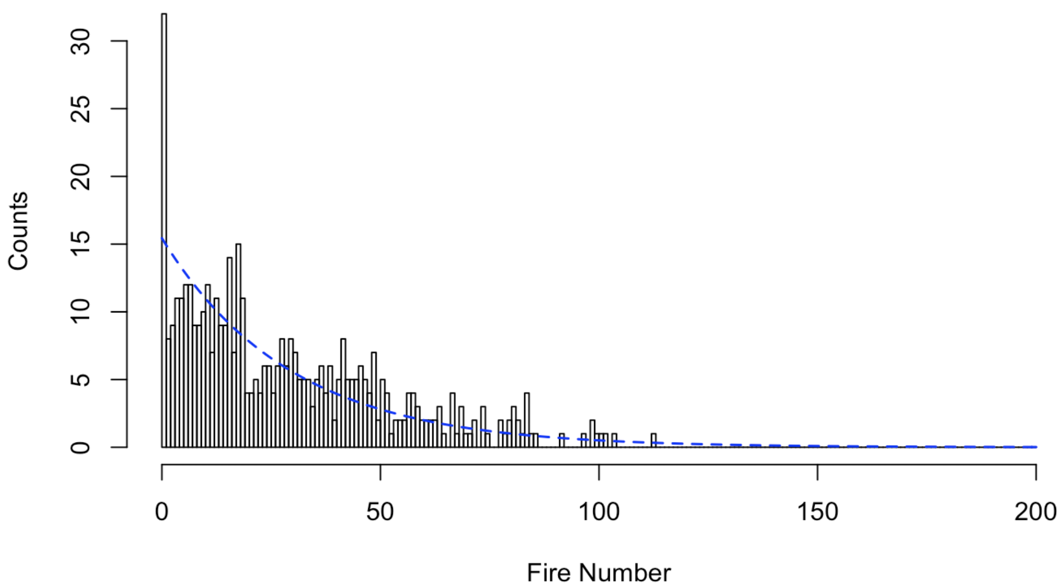


Figure 3.11: Fire Counts Histogram for 1996 to 2008

# CHAPTER 4

# Fire Sizes Distribution Analysis

In addition to the distribution of fire counts, the distribution of wildfire sizes is also of great interest to practitioners. We want to see if the Gamma-Exponential distribution fits the empirical data. In this chapter, we will derive a method combining Random Search and the optim function in R to estimate the unknown parameters. Using this method, the parameters are found and goodness-of-fit is assessed. The motivation of this analysis is to provide a statistical procedure that we can also apply to other fire sizes dataset in the future.

## 4.1 Negative Loglikelihood Function

In order to investigate the unknown parameters in fire size distribution, we derive the negative loglikelihood function because the optim in R minimize the function by default. Recall the distribution assumption for fire sizes:

$$x_{ij} \sim Exp(\mu_i)$$

$$\mu_i = \mu \xi_i$$

$$\xi_i \sim \Gamma(1, \alpha)$$

The loglikehood function is:

$$lnL = l = Nln\alpha + tln\mu + c - \sum_{i=1}^{N}[(m_i + 1)ln(\mu X_i + \alpha)]$$

where $t = \sum_{i=1}^{N} m_i, c = \sum_{i=1}^{N} ln\Gamma(m_i + 1)$, and $X_i = \sum_{j=1}^{m_i} x_{ij}$.

- $x_{ij}$ : fire size of $j^{th}$ fire in year $i$
- $N$: length of years
- $m_i$: number of records in year $i$
- $X_i$: total fire size in year $i$, length of $X_i$ is $N$
- $t$: total number of records
- $c$: constant

Since $c$ is a constant, $l$ is proportional to: $Nln\alpha + tln\mu - \sum_{i=1}^{N}[(m_i+1)ln(\mu X_i + \alpha)]$, and the negative loglikehood function is simply:

$$-lnL = -l = -Nln\alpha - tln\mu + \sum_{i=1}^{N}[(m_i + 1)ln(\mu X_i + \alpha)]$$

## 4.2 OPTIM Function

Negative loglikelihood function is useful since the built in function optim in R minimize the function by varying parameters. Optim is a general-purpose optimization takes initial values of parameters and function, and returns the minimum of the function with value of the parameters. The basic syntax is:

**optim(init, f)**

The next step is to simulate some data using random $\alpha$ and $\mu$, and use the optim function to see if it can recover the same value of $\alpha$ and $\mu$. The reason to perform this procedure is to make sure the likelihood function is correct.

## 4.3 Random Search

Random Search is very competitive when we want to find the global minimum of a function having many local minimums and when we don't have much information of unknown parameters(Solis and Wets, 1981). In our case, the optim function needs to take some initial values of parameters; however, we have no assumptions about the initial values for parameters $\alpha$ and $\mu$. In addition, our likelihood function may have many local minimums. Hence, a useful way to find the global minimum is to randomly generate a large number of $\alpha$ and $\mu$, and set them as initial values for optim function. If the optim function returns a smaller value, then we keep the resulting parameters from the optim function.

For example, if we assume that $\alpha = 10.6$ and $\mu = 15.7$. I simulate 100 years and 10,000 fire records for each year, meaning that $i$ is from 1 to 100 and $j$ is from 1 to 10,000. In R:

```
a = 10.6
u = 15.7
vi = rgamma(100,1,1/a)
ui = (1/u)*vi
x = matrix(nrow= 100, ncol = 10000)
for (i in 1:100) {
  x[i,]=rexp(10000,1/ui)
}
Xi = rowSums(x)
N = 100
t = sum(100*10000)
mi = rep(10000,100)
```

Negative likelihood function is generated by:

```
f <- function(par) {
  s = numeric(N) #this is the last term in the summation of
                       #negative loglikelihood function
  u <- par[1]
  a <- par[2]
  for (i in 1:N) {
    s[i] = (mi[i]+1) * log(u*Xi[i]+a) #this loop generates the last term
```

```
                                              #of negative loglikelihood function
  }
  logL = (N*log(a) + t*log(u) - sum(s))
  return(-logL)
}
```

If we use optim function directly to find the parameters, we need to set initial values for optim function. Suppose we do not know the initial values, as in our fire sizes case. Set initial $\alpha = 8$ and $\lambda = 3$(this is a guess because we do not know where to start). The optim function returns $\alpha = 4.0$ and $\lambda = 6.4$, which are not even close to the true value($\alpha = 10.6$ and $\lambda = 15.7$). Therefore, optim function cannot optimize correctly if there are local minimums. Therefore, Random search is very useful in this case.

Next step is to create a random amount of sets of $\alpha$ and $\lambda$. Here I create 500 sets of parameters:

```
n = 500
a = runif(n, 5, 20)
u = runif(n, 13, 19)

a0 = runif(1, 5, 20) #random intial number
u0 = runif(1, 13, 19) #random initial number
```

Then, for each set of $\alpha$ and $\lambda$, we use optim to find the local minimum. If the new local minimum is smaller than current local minimum, we replace current local minimum with the new one. After 500 times, we can approximately find the global minimum.

```
temp = optim(par=c(u0,a0), fn=f)
logL_value = temp$value
estimates = temp$par
logL = matrix(0, n, n)
k = 0 #check the number of replacements made
for (i in 1:n){
  for (j in 1:n){
    temp = optim(par=c(u[i], a[j]),fn=f)
    if (temp$value < logL_value) {
```

```
      logL_value = temp$value
      estimates = temp$par
      k = k+1
      print(k)
      print(logL_value)
      print(estimates)
    }
  }
}
logL_value
estimates
```

Random search returns $\alpha = 9.72$ and $\lambda = 15.75$, which almost recover the true values. Therefore, a combination of random search and optim function is very pragmatic to find the parameters for our distribution.

## 4.4 Finding $\alpha$ and $\mu$ for Fire Sizes Distribution

In this section, we use the combination of random search and optim function to find estimates of parametes $\alpha$ and $\mu$. Using the codes in section 4.3, we get $\alpha = 0.067$ and $\mu = 0.038$. One can simulate data by:

```
set.seed(1024)
u = .038
a = .067
vi = rgamma(N,1,1/a)
ui = u*vi
xi = numeric(0)
for (i in 1:N) {
  c = rexp(mi[i],ui[i])
  xi = c(xi,c)
}
```

A way to check goodness of fit is to compare the simulated data with observed data. To perform a direct K-S test, one can use the built in function ks.test in R:

```
ks.test(xi,tdata$Hectares[tdata$Hectares]) #xi is simulated data,
                                           #hectares is the observed data
```

The p-value is 3.441e-6, indicating a statistically significant departure from the modelled distribution, although visually the fit does not appear to be poor. Let fire sizes be x, a better way to visualize the goodness of fit is to take log of fire sizes, then plot it against $logS(x)$, where $S(x) = 1 - F(x)$ and $F(x) = P(y \leq x)$.

```
y1 = sort(xi)
n = length(y1)
plot(y1,(n:1)/n,log = "xy", type='l',col ='red',ylab = "logS(x)",
     xlab = "log(x)" )

y2 = sort(tdata$Hectares)
lines(y2,(n:1)/n,log = "xy",col ='black', lty =1)
```

Figure 4.1 shows the plot of comparison between simulated and observed fire sizes. Simulated and observed data seems to follow the same trend and they are pretty close. According to the plot, our proposed distribution fits the data relatively well. From Figure 4.1, two distributions start to deviate at $log(x) = 1e3$

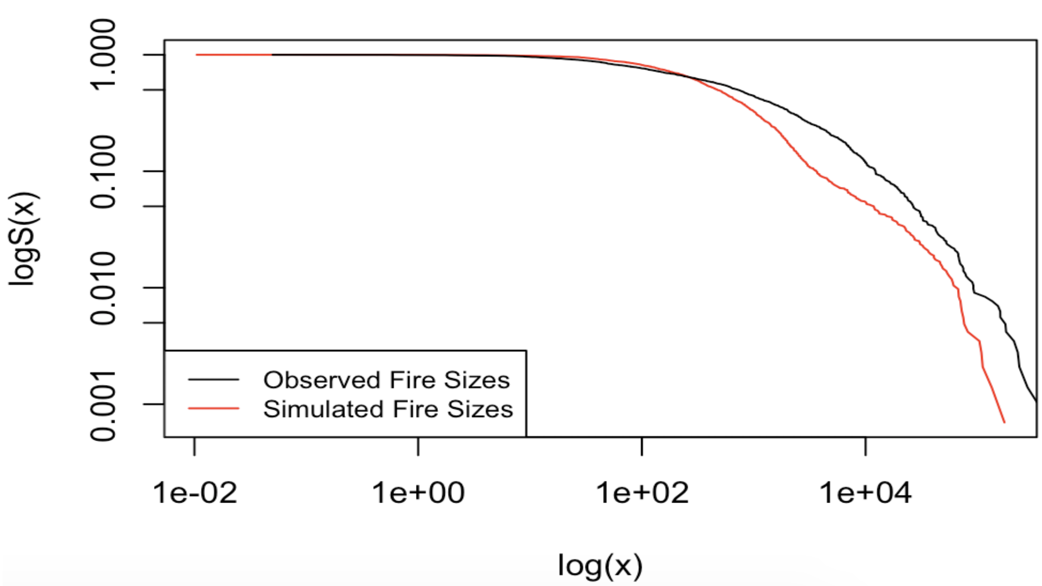


Figure 4.1: Comparsion between Simulated and Observed data

approximately. If we set $x = 10^3$, we are observing 635 points in real dataset, and

469 points in simulated dataset, meaning that we observe an excess of approximately 166 wildfires above this size threshold in our dataset relative to what we would expect according to the fitted distribution.

# CHAPTER 5

# Future Discussion and Conclusion

## 5.1 Future Improvement and Discussion

### 5.1.1 Outliers

Although the Gamma-Exponential distribution fits our fire sizes pretty well, there are still some deviations between empirical and theoretical distributions. One of the potential problems may be that there are some outliers of records in certain years. For example, the fire sizes data in our study contains 44 years in total. Each year there are many different records: some fire sizes are very big and some are very small. Outliers may cause a poor fit between distribution and empirical data. Therefore, in order to eliminate the effect of outliers, one can use Q-Q plot and box plot to detect the outliers, or built in function outlier.test in R. However, one should be always be cautious with removing outliers in dataset. Before removing outliers, we should investigate the reason that cause the outliers, rather than deleting it directly. Moreover, if there are too many outliers, there may be something interesting going on with data that we should look further into.

### 5.1.2 Hidden Features

In this paper, we investigate if Gamma-Poisson and Gamma-Exponential distributions fit to empeirical fire data. In real life, there may be some hidden features in the model, such as human accident, time, locations and etc. One may develop a better model and understand the dataset better by fitting a regression model to data, including other hidden factors. Other models may be helpful such as random effect model.

### 5.1.3 Transformation

In some studies, transformations of the count data are conducted for a better fit(Schoenberg, Peng and Woods, 2003). Various models and transforms could be used to examine data, such as lognormal, half-normal, exponential and etc. In addition to Kolmogorov–Smirnoff test that we mainly used in this study, other goodness of fit criteria can also be used and compared, such as Cramer–von Mises statistics.

### 5.1.4 Gradient Descent

In addition to random search, gradient descent is another good method to find the unknown parameters. Gradient descent is used to minimize likelihood function by iteratively moving in the direction of steepest descent. We can continuously update parameters and eventually find the parameters in our model. A sample code that I used in R is provided here, but it needs some further developments:

```
l <- function(u, a) {
  s = numeric(N)
  for (i in 1:N) {
    s[i] = (mi[i]+1) * log(u*Xi[i]+a)
  }
  logL = (N*log(a) + t*log(u) - sum(s))
  return(-logL)
}

grad_a = function(u,a){
  N/a - sum((mi+1)/(u*Xi+a))
}

grad_u = function(u,a){
  t/u - sum((mi+1)*Xi/(u*Xi+a))
}
```

```
gradient = function(u,a){
  c(grad_u(u,a),grad_a(u,a))
}
h = .0000001 #learning rate
a0 = u0 = 5
k=1
while (abs(grad_a(u0,a0)) >1e-3 & k < 50000) {
  l_old = l_new
  a_new = abs(a0 - h*grad_a(u=u0,a=a0))+1e-4
  u_new = abs(u0 - h*grad_u(u=u0,a=a0))+1e-4
  l_new = l(u_new,a_new)
  a0 = a_new
  u0 = u_new
  k = k+1
}
print(c(a0,u0))
l_new
```

### 5.1.5 Evaluating the Goodness of Maximum Likelihood Estimators

Besides the maximum likelihood estimates derived in chapter 2, we also want to know how good are our parameters, which is usually measured by the MSE. We know that $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = bias^2 + var(\hat{\theta})$. By central limit theorem for maximum likelihood estimators, we have: $\hat{\theta} \sim N(\theta, I(\theta)^{-1})$, where $I(\theta)$ is the fisher information and could be estimated by Hessian matrix (Lehmann, 2004). Hessian matrix is the second derivative of the log-likelihood function(Gourieroux, Monfort and Trognon, 1984).

However, sometimes the Hessian matrix may be too complicated to derive and compute, like in our case. We could also appeal to Bootstrap method to get the variance of our estimators. To simply implement it, we sample with replacement from our observations, and find the corresponding MLE to the sample data (e.g repeat 1000 times), and the variance of the computed MLE will be unbiased estimate of the variance of MLE.

## 5.2 Conclusion

In this study, we propose two distributions: Gamma-Poisson and Gamma-Exponential distribution. Two distributions are used to fit fire counts and fire sizes data from Northwest Territories in Canada. Northwest Territories is the second-largest and the most populous in Northern Canada. Understanding the fire distributions can help prevent fire hazard in a certain way.

For fire counts distribution, we propose Gamma-Poisson distribution and derive the negative loglikelihood function in order to estimate unknown parameters. K-S plots are used to check for goodness of fit. In four time frames, year from 1965 to 1975 and 1996 to 2008 are well fitted by proposed distribution. There are some deviations between theoretical distribution and fire counts from 1976 to 1985 and 1986 to 1995. Overall, Gamma-Poisson distribution is a good fit for this empirical data.

For fire sizes distribution, we propose Gamma-Exponential distribution and derive the negative loglikelihood function as well. However, fire sizes distribution is more complex than fire counts distribution. Built in function optim in R and method of random search were used to find the parameters. We check for goodness of fit by simulating fire sizes data and using K-S test. Although p-value is somehow small, it still shows that the Gamma-Exponential distribution fits our data pretty well. In addition, Figure 4.1 shows that empirical and theoretical function are pretty similar, indicating proposed distribution is a good fit to our data.

# Appendix

```
#read data to R
tdata <- read.csv("~/desktop/thesis/tdata.csv")

########### Fire Count Analysis ###########

#neg_l function returns negative loglikelihood
neg_l <- function(par, n) {
  lambda <- par[1]
  alpha <- par[2]
  logl <- log(lambda)*sum(n) - (sum(n)+length(n))
        *log(alpha+lambda)+length(n)*log(alpha)
  return(-logl)
}

#likehood estimation by plot
lambda <- seq(30,100, len=50)
alpha <- seq(1,10, len=50)
z <- outer(lambda, alpha, neg_l)
persp(lambda, alpha, z,theta=30,phi=30)

#find estimation of fire number
sum(tdata$FIRE_NUM)/length(tdata$FIRE_NUM)
k = mean(tdata$FIRE_NUM)

#Construct KS plot: compare theoretical cdf and empirical cdf, overall
x = 0:200
y = (k/(1+k))^x * (1/(1+k))
Fx_hat = Fx = cumsum(y)
ni = tdata$FIRE_NUM
N = length(tdata$FIRE_NUM)
for (i in 0:200){
Fx_hat[i+1] = sum(ni<=i)/N
}
upperconfidence_99 = Fx + 1.63/(N^(1/2))
lowerconfidence_99 = Fx - 1.63/(N^(1/2))
upperconfidence_95 = Fx + 1.36/(N^(1/2))
lowerconfidence_95 = Fx - 1.36/(N^(1/2))
plot(Fx , Fx_hat,pch=16, xlim=c(0,1), ylim=c(0,1),cex = .4,
        xlab = 'Empirical CDF', ylab = 'Theoratical CDF',
         main = 'KS Plot with Parameter k = 39.266')
lines(x=0:1,y=0:1, col='red',lty=5)
lines(x=Fx,y=upperconfidence_99, col ='green',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_99, col = 'green',cex = 1.5,lty=2)
```

```
lines(x=Fx,y=upperconfidence_95, col ='blue',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_95, col = 'blue',cex = 1.5,lty=2)
# Add a legend
legend("topleft", legend=c("95%␣Confidence␣Bounds", "99%␣Confidence␣
    Bounds",
        "45␣Degree␣Line"), col=c("blue", "green", "red"), lty=c(2,2,5),
            cex=0.8)

#Histogram against theoretical
h <- hist(tdata$FIRE_NUM,breaks = 0:200, xlab ="Fire␣Number",ylab="
    Counts", main = "Fire␣Counts␣Histogram␣against␣Theoretical␣
    Distribution␣for␣All␣Years",plot = TRUE)
x = 0:200
y = (k/(1+k))^x * (1/(1+k))
N = 1435 #length of data
lines(x,y*N,lty = 2, col=4,lwd=1.5)

#seperate data in different time frames
summary(tdata$FIRE_YEAR)
fc_65_75 = tdata$FIRE_NUM[tdata$FIRE_YEAR <= 1975]
fc_76_85 = tdata$FIRE_NUM[tdata$FIRE_YEAR >1975 & tdata$FIRE_YEAR <=
    1985]
fc_86_95 = tdata$FIRE_NUM[tdata$FIRE_YEAR >1985 & tdata$FIRE_YEAR
    <=1995]
fc_96_08 = tdata$FIRE_NUM[tdata$FIRE_YEAR >=1996]

######Real data vs. distribution between 1965-1975#####
## Construct Histogram versus Theoretical Distribution ##

k_65_75 = mean(fc_65_75)
h <- hist(fc_65_75,breaks = 0:200, xlab ="Fire␣Number",ylab="Counts",
    main = "Histogram␣against␣Theoretical␣Distribution␣for␣1965␣to␣1975"
    )
x = 0:200
y = (k_65_75/(1+k_65_75))^x * (1/(1+k_65_75))
N = length(fc_65_75)
lines(x,y*N,lty = 2, col=4,lwd=1.5)

## Construct KS Plot and Confidence Interval ##

Fx_hat = Fx = cumsum(y)
ni = fc_65_75
N = length(fc_65_75)
for (i in 0:200){
Fx_hat[i+1] = sum(ni<=i)/N
}
```

```
upperconfidence_99 = Fx + 1.63/(N^(1/2))
lowerconfidence_99 = Fx - 1.63/(N^(1/2))
upperconfidence_95 = Fx + 1.36/(N^(1/2))
lowerconfidence_95 = Fx - 1.36/(N^(1/2))
plot(Fx , Fx_hat,pch=16, xlim=c(0,1), ylim=c(0,1),cex = .4, xlab = '
    Empirical␣CDF', ylab = 'Theoratical␣CDF', main = 'KS␣Plot␣with␣
    Parameter␣k␣=␣45.62␣for␣1965␣to␣1975')
lines(x=0:1,y=0:1, col='red',lty=5)
lines(x=Fx,y=upperconfidence_99, col ='green',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_99, col = 'green',cex = 1.5,lty=2)
lines(x=Fx,y=upperconfidence_95, col ='blue',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_95, col = 'blue',cex = 1.5,lty=2)
# Add a legend
legend("topleft", legend=c("95%␣Confidence␣Bounds", "99%␣Confidence␣
    Bounds","45␣Degree␣Line"), col=c("blue", "green", "red"), lty=c
    (2,2,5), cex=0.8)

######Real data vs. distribution between 1976-1985#####
## Construct Histogram versus Theoretical Distribution ##

k_76_85 = mean(fc_76_85)
h <- hist(fc_76_85,breaks = 0:200, xlab ="Fire␣Number",ylab="Counts",
    main = "Histogram␣against␣Theoretical␣Distribution␣for␣1976␣to␣1985"
    )
x = 0:200
y = (k_76_85/(1+k_76_85))^x * (1/(1+k_76_85))
N = length(fc_76_85)
lines(x,y*N,lty = 2, col=4,lwd=1.5)

## Construct KS Plot and Confidence Interval##

Fx_hat = Fx = cumsum(y)
ni = fc_76_85
N = length(fc_76_85)
for (i in 0:200){
Fx_hat[i+1] = sum(ni<=i)/N
}
upperconfidence_99 = Fx + 1.63/(N^(1/2))
lowerconfidence_99 = Fx - 1.63/(N^(1/2))
upperconfidence_95 = Fx + 1.36/(N^(1/2))
lowerconfidence_95 = Fx - 1.36/(N^(1/2))
plot(Fx , Fx_hat,pch=16, xlim=c(0,1), ylim=c(0,1),cex = .4, xlab = '
    Empirical␣CDF', ylab = 'Theoratical␣CDF', main = 'KS␣Plot␣with␣
    Parameter␣k␣=␣27.89␣from␣1976␣to␣1985')
lines(x=0:1,y=0:1, col='red',lty=5)
lines(x=Fx,y=upperconfidence_99, col ='green',cex = 1.5,lty=2)
```

```
lines(x=Fx,y=lowerconfidence_99, col = 'green',cex = 1.5,lty=2)
lines(x=Fx,y=upperconfidence_95, col ='blue',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_95, col = 'blue',cex = 1.5,lty=2)
# Add a legend
legend("topleft", legend=c("95%␣Confidence␣Bounds", "99%␣Confidence␣
    Bounds","45␣Degree␣Line"), col=c("blue", "green", "red"), lty=c
    (2,2,5), cex=0.8)

#####Real data vs. distribution between 1986-1995#####
## Construct Histogram versus Theoretical Distribution ##

k_86_95 = mean(fc_86_95)
h <- hist(fc_86_95,breaks = 0:200, xlab ="Fire␣Number",ylab="Counts",
    main = "Histogram␣against␣Theoretical␣Distribution␣for␣1986␣to␣1995"
    )
x = 0:200
y = (k_86_95/(1+k_86_95))^x * (1/(1+k_86_95))
N = length(fc_86_95)
lines(x,y*N,lty = 2, col=4,lwd=1.5)

## Construct KS Plot and Confidence Interval##
Fx_hat = Fx = cumsum(y)
ni = fc_86_95
N = length(fc_86_95)
for (i in 0:200){
Fx_hat[i+1] = sum(ni<=i)/N
}
upperconfidence_99 = Fx + 1.63/(N^(1/2))
lowerconfidence_99 = Fx - 1.63/(N^(1/2))
upperconfidence_95 = Fx + 1.36/(N^(1/2))
lowerconfidence_95 = Fx - 1.36/(N^(1/2))
plot(Fx , Fx_hat,pch=16, xlim=c(0,1), ylim=c(0,1),cex = .4, xlab = '
    Empirical␣CDF', ylab = 'Theoratical␣CDF', main = 'KS␣Plot␣with␣
    Parameter␣k␣=␣54.68␣from␣1986␣to␣1995')
lines(x=0:1,y=0:1, col='red',lty=5)
lines(x=Fx,y=upperconfidence_99, col ='green',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_99, col = 'green',cex = 1.5,lty=2)
lines(x=Fx,y=upperconfidence_95, col ='blue',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_95, col = 'blue',cex = 1.5,lty=2)
# Add a legend
legend("topleft", legend=c("95%␣Confidence␣Bounds", "99%␣Confidence␣
    Bounds","45␣Degree␣Line"), col=c("blue", "green", "red"), lty=c
    (2,2,5), cex=0.8)

#####Real data vs. distribution between 1996-2008#####
## Construct Histogram versus Theoretical Distribution ##
```

```
k_96_08 = mean(fc_96_08)
h <- hist(fc_96_08,breaks = 0:200, xlab ="Fire␣Number",ylab="Counts",
    main = "Histogram␣against␣Theoretical␣Distribution␣for␣1996␣to␣2008"
    )
x = 0:200
y = (k_96_08/(1+k_96_08))^x * (1/(1+k_96_08))
N = length(fc_96_08)
lines(x,y*N,lty = 2, col=4,lwd=1.5)

## Construct KS Plot and Confidence Interval##

Fx_hat = Fx = cumsum(y)
ni = fc_96_08
N = length(fc_96_08)
for (i in 0:200){
Fx_hat[i+1] = sum(ni<=i)/N
}
upperconfidence_99 = Fx + 1.63/(N^(1/2))
lowerconfidence_99 = Fx - 1.63/(N^(1/2))
upperconfidence_95 = Fx + 1.36/(N^(1/2))
lowerconfidence_95 = Fx - 1.36/(N^(1/2))
plot(Fx , Fx_hat,pch=16, xlim=c(0,1), ylim=c(0,1),cex = .4, xlab = '
    Empirical␣CDF', ylab = 'Theoretical␣CDF', main = 'KS␣Plot␣with␣
    Parameter␣k␣=␣28.8␣from␣1996␣to␣2008')
lines(x=0:1,y=0:1, col='red',lty=5)
lines(x=Fx,y=upperconfidence_99, col ='green',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_99, col = 'green',cex = 1.5,lty=2)
lines(x=Fx,y=upperconfidence_95, col ='blue',cex = 1.5,lty=2)
lines(x=Fx,y=lowerconfidence_95, col = 'blue',cex = 1.5,lty=2)
# Add a legend
legend("topleft", legend=c("95%␣Confidence␣Bounds", "99%␣Confidence␣
    Bounds","45␣Degree␣Line"), col=c("blue", "green", "red"), lty=c
    (2,2,5), cex=0.8)

############ Fire Size Analysis ############

N = 2008-1965+1 # Total Years: length of i
mi = NA # Number of records in year i
Xi = NA # Total Fire Size in year i

for (i in 1965:2008) {
  mi[i] = length(tdata$Hectares[tdata$FIRE_YEAR == i])
}
mi = as.vector(na.omit(mi))
length(mi) == N
```

```
t = sum(mi)
c = sum(gamma(mi+1)) # c is 3rd term in loglikelihood function

for (i in 1965:2008) {
  Xi[i] = sum(tdata$Hectares[tdata$FIRE_YEAR == i])
}
Xi = as.vector(na.omit(Xi))
length(Xi) == N

f <- function(par) {
  s = numeric(N)
  u <- par[1]
  a <- par[2]
  for (i in 1:N) {
    s[i] = (mi[i]+1) * log(u*Xi[i]+a)
  }
  logL = (N*log(a) + t*log(u) - sum(s))
  return(-logL)
}

# random search
n = 100
a = runif(n, .003, .1)
u = runif(n, .005, .09)

a0 = runif(1, 0.03, .1)
u0 = runif(1, 0.05, .09)
#temp = optim(a[i], mu[j], logL, ...)

temp = optim(par=c(u0,a0), fn=f)
logL_value = temp$value
estimates = temp$par
logL = matrix(0, n, n)
dim(logL)

k = 0
options(error = expression(NULL))

for (i in 1:n){
  for (j in 1:n){
    temp = optim(par=c(u[i], a[j]),fn=f)
    if (temp$value < logL_value) {
      logL_value = temp$value
      estimates = temp$par
      k = k+1
      print(k)
```

```
      print(logL_value)
      print(estimates)
    }
    #logL_value = logL_value + (temp$function_value < logL_value)*(temp$
        function_value - logL_value)
  }
}
logL_value
estimates

####Simulate Fire size data and check goodness of fit####

set.seed(1024)
u = .038
a = .067
vi = rgamma(N,1,1/a)
ui = u*vi
xi = numeric(0)
for (i in 1:N) {
  c = rexp(mi[i],ui[i])
  xi = c(xi,c)
}
hist(xi, nclass =500)
hist(tdata$Hectares,nclass =500)
ks.test(xi,tdata$Hectares[tdata$Hectares])

y1 = sort(xi)
n = length(y1)
plot(y1,(n:1)/n,log = "xy", type='l',col ='red',ylab = "logS(x)", xlab =
     "log(x)" )

y2 = sort(tdata$Hectares)
lines(y2,(n:1)/n,log = "xy",col ='black', lty =1)
legend("bottomleft", legend=c("Observed␣Fire␣Sizes","Simulated␣Fire␣
    Sizes"),
       col=c("black", "red"), lty=c(1,1), cex=0.8)

data_point_for_sim_greater_1e3 = y1[y1>1e3]
data_point_for_observ_greater_1e3 = y2[y2>1e3]

length(data_point_for_sim_greater_1e3) #=469
length(data_point_for_observ_greater_1e3) #635
```

# References


[1] A. Clifford Cohen. 1965. *Maximum Likelihood Estimation in the Weibull Distribution Based On Complete and On Censored Samples*. Technometrics, 7:4, 579-588, DOI: 10.1080/00401706.1965.10490300.

[2] A. Lindén, and Samu Mäntyniemi. 2011. *Using the Negative Binomial Distribution to Model Overdispersion in Ecological Count Data*. Ecology, vol. 92, no. 7, pp. 1414–1421, JSTOR.

[3] Bliss, C. I., and R. A. Fisher. 1953. *Fitting the Negative Binomial Distribution to Biological Data.*. Biometrics, vol. 9, no. 2, pp. 176–200. JSTOR, www.jstor.org/stable/3001850.

[4] Cheng T, Wang J. 2008. *Integrated spatio-temporal data mining for forest fire prediction*. Trans GIS 12(5):591–611.

[5] C. Gourieroux, A. Monfort and A. Trognon. 1984. *Pseudo Maximum Likelihood Methods: Applications to Poisson Models*. Econometrica, Vol. 52, No. 3, pp. 701-720

[6] D Mandallaz and R Ye. 1997. *Prediction of forest fires with Poisson models*. Canadian Journal of Forest Research, 27(10): 1685-1694, https://doi.org/10.1139/x97-103

[7] Emery N. Brown , Riccardo Barbieri , Valerie Ventura , Robert E. Kass , and Loren M. Frank. 2002. *The Time-Rescaling Theorem and Its Application to Neural Spike Train Data Analysis*. Neural Computation, 14:2, 325-346

[8] E.L. Lehmann. 2004. *Elements of Large-Sample Theory*. Springer Science & Business Media.

[9] Frank J. Massey Jr. 1951. *The Kolmogorov-Smirnov Test for Goodness of Fit*. Journal of the American Statistical Association, 46:253, 68-78.

[10] Frederic Paik Schoenberg, Roger Peng, James Woods. 2003. *On the Distribution of Wildfire Sizes*. Environmetrics 14.6 (2003): 583-92.

[11] Francisco J. Solis and Roger J-B. Wets. 1981. *Minimization by Random Search Techniques*. Mathematics of Operations Research, vol. 6, no. 1, pp. 19–30. JSTOR.

[12] Guha-Sapir D, Hoyois P, Below R. 2013. *Annual disaster statistical review 2012: the numbers and trends*. Centre for Research on the Epidemiology of Disasters (CRED).

[13] Grid-Arendal UNEP. 2002 *State of the environment and policy retrospective.* (Norway) United Nations Environment Programme (UNEP).

[14] Hinde J. 1982. *Compound Poisson Regression Models.* In: Gilchrist R. (eds) GLIM 82: Proceedings of the International Conference on Generalised Linear Models. Lecture Notes in Statistics, vol 14. Springer, New York, NY.

[15] Harry Joe and Rong Zhu. 2019 *Generalized Poisson Distribution: The Property of Mixture of Poisson and Comparison with Negative Binomial Distribution.* Biometrical Journal 47.2 (2005): 219-29.

[16] Lawless JF. 1987. *Negative binomial and mixed Poisson regression.* The Canadian Journal of Statistics, Vol. 15, No. 3, 1987, Pages 209-225.

[17] Nan Laird, 1978 *Nonparametric Maximum Likelihood Estimation of a Mixing Distribution.* Journal of the American Statistical Association, 73:364, 805-811, DOI: 10.1080/01621459.1978.10480103.

[18] Robertson C. 1972. *Analysis of forest fire data in California.* Technical Report No. 11, Department of Statistics, University of California, Riverside.

[19] Wang F, Lu S, Li C. 2005. *Analysis of fire statistics of China: fire frequency and fatalities in fires.* International Association for Fire Safety Science: Fire Safety Science.