Analysis of the Oil Spills from Tanker Ships

Ringo Ching and T. L. Yip
The Data

Included accidents in which International Oil Pollution Compensation (IOPC) Funds were involved, up to October 2009.

In this study the spill amounts in tonnes recorded in 1992 Fund and 1971 fund are combined according to cases.
Background

According to annual report of the IOPC fund:

The 1992 Fund pays compensation when:

- the damage exceeds the limit of the shipowner’s liability under the 1992 Civil Liability Convention,
- the shipowner is exempt from liability under the 1992 Civil Liability Convention,
- the shipowner is financially incapable of meeting his obligations in full under the 1992 Civil Liability Convention, and the insurance is insufficient to pay valid compensation claims.
Background

An overall average does not fully describes the situation, especially for this compensation fund which is responsible for the major spills in excess of the liability limit of ship owners.
Background

Precise analysis should be done on the larger major spills

If premium is too large, the cost of the business will be increased unreasonably and lower the profit

If the premium was too low, the fund will go bankrupt, the risk sharing mechanism would not work

Accurate estimation would lead to more reasonable premium, making the fund more efficient
Summary Statistics

- Year 1979-2008
- Number of accidents=105, expected value = 4296.99
- Maximum = 84000
- Skewness= 4.43, Kurtosis= 19.46 (normal distribution has skewness = 0; kurtosis = 3)

Most spills are small in amount while some spills are in another extreme
Fitting with a single distribution

Weibull and lognormal distributions are fitted to the spilled amount

Log-Likelihood of fitted lognormal: -785.72
Log-Likelihood of fitted Weibull: -791.17
Observed spill amount: 4282.11 tonnes
Expected spill amount: **11731.70** tonnes
Expected vs observed: **173.9%** error

Single distribution is not working well
Possible solution: 2 distributions
Peaks-Over-Threshold Method

A method used widely in field of hydrology and insurance

Our random variable $X$ would be the spill amount in tonnes

The approximate distribution $F(x)$ of those $X$ larger than $u$, would be generalized Pareto distribution (GPD) [1]:

$$G_{\xi,\sigma,u}(x) = \begin{cases} 1 - \left(1 + \frac{\xi [x - u]}{\sigma}\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp\left(-\frac{[x - u]}{\sigma}\right) & \xi = 0 \end{cases}$$
Peaks-Over-Threshold Method

Oil spilled (tonnes)

Years

Threshold

Oil spilled (tonnes)
The plot was obtained by matching observed data to the exponential distribution. Since it is not linear the data cannot be modeled by exponential distribution (GPD with ξ=0).
Peaks-Over-Threshold Method

For GPD, if we keep on rising our threshold $R$ larger than the suitable threshold $u$, the average value of those spills minus $R$ (mean excess) would increase linearly with $R$ with slope

$$\frac{\xi}{1-\xi}$$

$$E(X - R | X > R) = \frac{\sigma + \xi(R - u)}{1 - \xi}, \quad R \geq u$$
Peaks-Over-Threshold Method

An example would be claim data from motor insurance portfolio consists of 172,161 policies, studied by P. Gigante, L. Picech and L. Sigalotti[2]
The linear pattern after reaching the threshold 6000 shows that the data can be modeled by GPD with threshold at around 6000.
Peaks-Over-Threshold Method

In other words, the major spill amount can be modeled by GPD by choosing a high enough threshold $u$

The overall spill amount is represented by 2 distributions, with the GPD responsible for the large spill
Results

Castillo and Hadi [3] compared the methods for estimating the generalized Pareto distribution. They suggested that for small sample, probability weighted moment method should be employed when there is reason to believe $0 \leq \xi \leq 0.5$.

From the linear part of empirical mean excess function, its slope $\frac{\xi}{1 - \xi}$ is positive, such that $\xi \geq 0$ and it is approximately 0.12.
Results

GPD has finite expectation and variance if and only if $\xi$ is smaller than 0.5. As the amount of spill is limited by the capacity, the expectation and variance of the spill amount should be finite.
## Results

<table>
<thead>
<tr>
<th>Thresholds $u$ (tonnes)</th>
<th>No. of exceedances</th>
<th>Average (tonnes) of those spills larger than:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>3900</td>
</tr>
<tr>
<td>3090</td>
<td>15</td>
<td>0.3489</td>
</tr>
<tr>
<td>3800</td>
<td>14</td>
<td>0.3197</td>
</tr>
<tr>
<td>5700</td>
<td>12</td>
<td>0.2650</td>
</tr>
<tr>
<td>5900</td>
<td>12</td>
<td>0.2769</td>
</tr>
<tr>
<td>6100</td>
<td>12</td>
<td>0.2888</td>
</tr>
<tr>
<td>6200</td>
<td>11</td>
<td>0.1780</td>
</tr>
<tr>
<td>6500</td>
<td>11</td>
<td>0.1954</td>
</tr>
<tr>
<td>6800</td>
<td>11</td>
<td>0.2128</td>
</tr>
<tr>
<td>7000</td>
<td>10</td>
<td>0.0751</td>
</tr>
<tr>
<td>7200</td>
<td>10</td>
<td>0.0864</td>
</tr>
<tr>
<td>7500</td>
<td>10</td>
<td>0.1034</td>
</tr>
</tbody>
</table>

**Observed Values**

|                    | 29556.79 | 36096.36 | 39006.00 | 42451.11 | 42451.11
Results

Averages of spills larger than $R$ (tonnes)

$R$ (tonnes)

- $u=2000$
- $u=2250$
- $u=2500$
- $u=3800$
- $u=3090$
- $u=6200$

Observed Values
Results

Averages of spills larger than \( R \) (tonnes)

\( R \) (tonnes)

Observed Values

\( u=6200 \)
\( u=6500 \)
\( u=7000 \)
\( u=7500 \)
Results

Thresholds 6200 and 7000 would be compared.

From the density graph of the spills less than 6300, weibull, gamma and lognormal distributions were fitted to these smaller spill amounts.

The Log-Likelihood of the fitted distributions:

<table>
<thead>
<tr>
<th>Thresholds (tonnes)</th>
<th>Weibull</th>
<th>Gamma</th>
<th>Log-normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>6200</td>
<td>-628.01</td>
<td>-970.4774</td>
<td>-631.2586</td>
</tr>
<tr>
<td>7000</td>
<td>-640.7031</td>
<td>-1043.872</td>
<td>-643.5014</td>
</tr>
</tbody>
</table>
Results

Hypothesis tests were conducted on the overall fitness of the mixture distributions.

The Kolmogorov-Smirnov (KS) Test is based on the maximum difference between the observed distribution $F_n(x)$ and the estimated distribution $F(x)$ [4]:

$$\sup_x |F(x) - F_n(x)|$$

The Anderson-Darling (AD) Test is a modification which puts more weight on the large data:

$$n \int_{-\infty}^{\infty} \frac{(F(x) - F_n(x))^2}{F(x)[1 - F(x)]} dF(x)$$
## Results

<table>
<thead>
<tr>
<th>Threshold (tonnes)</th>
<th>Kolmogorov-Smirnov (KS) Test</th>
<th>Anderson-Darling (AD) Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>6200</td>
<td>0.0530</td>
<td>0.2257</td>
</tr>
<tr>
<td>7000</td>
<td>0.0565</td>
<td>0.2317</td>
</tr>
</tbody>
</table>

Critical values

0.1327

2.492

(5% level of significance)

The distribution with threshold 6200 have a slightly better fit to the observed spill amounts

Average spill amount given by this proposed distribution is 4307.08 tonnes with 0.58% percentage error

A log-normal distribution gives an estimate with 173.9% percentage error
Implication

Lognormal

Proposed

Spill Amount (tonnes)

$F(x)$

Observed

lognormal

Observed

proposed
Implications

Lognormal

Proposed

Where fits well
Implications

We further compare the performance of a single lognormal and the proposed distribution through hypothesis tests. The test statistics are given below

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov (KS) Test</th>
<th>Anderson-Darling (AD) Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log normal</td>
<td>0.0433</td>
<td>0.2386</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.0530</td>
<td>0.2257</td>
</tr>
<tr>
<td>Critical values (5% level of significance)</td>
<td>0.1327</td>
<td>2.492</td>
</tr>
</tbody>
</table>

The proposed distribution performs better when placing more emphasis on the large data.
Implications

From the prospective of funds, the estimated average spill amount larger than a level would be put to test

<table>
<thead>
<tr>
<th></th>
<th>3000</th>
<th>3900</th>
<th>6300</th>
<th>8000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>26524.68</td>
<td>28053.00</td>
<td>33671.67</td>
<td>42451.11</td>
<td>42451.11</td>
</tr>
<tr>
<td>Log normal</td>
<td>78776.71</td>
<td>89943.78</td>
<td>116155.39</td>
<td>132772.86</td>
<td>151009.2</td>
</tr>
<tr>
<td></td>
<td>(197%)</td>
<td>(216%)</td>
<td>(245%)</td>
<td>(213%)</td>
<td>(256%)</td>
</tr>
<tr>
<td>GPD (u=6200)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36218.54</td>
<td>38286.67</td>
<td>40719.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.56%)</td>
<td>(-9.81%)</td>
<td>(-4.08%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where the percentage errors compared with the observed values are in blankets
Implications

• Through separate treatment of the larger spill amounts with Peak-Over-Threshold method, a more accurate distribution for extreme oil spill data is obtained.

• This distribution can be used by funds which are responsible for accidents exceeded the liability limit of ship owners to determine more reasonable premium, making the whole business more efficient.
References

The Overall Distribution

A mixture distribution can be used for the overall spill amount, with the GPD responsible for the larger spill amounts \((X>R)\), the expectation would thus be given as

\[
E(X) = P(X \leq R)E(X|X \leq R) + P(X > R)E(X|X > R)
\]

\(E(X|X>R)\) given by GPD would be

\[
E(X|X > R) = \frac{\sigma + \xi(R - u)}{1 - \xi} + R, \quad R \geq u
\]
Appendix

Suggested by Castillo and Hadi [2]:

1. If the sample size is large (>500) and it is believed that \(0.5 > \xi > -0.5\), maximum likelihood estimation (MLE) method would be preferred.

2. If sample size is not large and it is believed that \(0.5 > \xi > 0\), probability weighted moment method (PWM) should be used.

3. In all other cases, used elemental percentile method (EPM).

4. In all cases, if MLE has convergence problems or if PWM gives nonsensical estimates, then use EPM.
Appendix

Probability weighted moment method:

$$\alpha_s = E[ X(1 - F(X))^s ] = \frac{\sigma}{(s + 1)(s + 1 + \xi)}$$

$$\hat{\alpha}_s = n^{-1} \sum_{j=1}^{n} (1 - p_{j:n})^s x_{j:n}$$

where $x_{j:n}$ is the jth largest sample and $p_{j:n} = \frac{(j - 0.35)}{n}$