



# 基于双截尾-GPD分布的 船舶溢油损失分布拟合研究

田荣洁<sup>1</sup>, 田晓洁<sup>2</sup>, 李金林<sup>1</sup>, 杜晓川<sup>3</sup>

(1. 北京理工大学管理与经济学院, 北京 100081; 2. 交通运输部水运科学研究院, 北京 100088;  
3. 长沙理工大学城南学院, 湖南 长沙 410076)

**摘要:** 以全国船舶溢油事故损失数据为样本, 针对船舶溢油损失数据的主体部分和厚尾部分分段建模, 并将双截尾-GPD分布引入船舶溢油损失序列的分布拟合中, 对船舶溢油损失序列的分布特征进行研究。实证研究发现, 船舶溢油损失数据不服从正态分布, 具有“尖峰、厚尾”的特性; 分段拟合则进一步提高了损失分布的拟合准确性。

**关键词:** 船舶溢油损失分布; 双截尾-GPD分布; 拟合研究

中图分类号: U 698.7

文献标志码: A

文章编号: 1002-4972(2014)03-0041-04

## Fitting research on ship's oil spill loss based on DTD-GPD distribution

TIAN Rong-jie<sup>1</sup>, TIAN Xiao-jie<sup>2</sup>, LI Jin-lin<sup>1</sup>, DU Xiao-chuan<sup>3</sup>

(1. Beijing Institute of Technology of Management and Economics, Beijing 100081, China;

2. China Waterborne Transport Research Institute, Beijing 100088, China;

3. College of Chengnan, Changsha University of Science & Technology, Changsha 410076, China)

**Abstract:** According to the samples of oil spills loss data from ships in China, this paper considers to separate the data into the main part and the tail part, and constructs a doubly-truncated distribution - GPD (DTD-GPD) model to model the two parts of the loss data. The results show that the loss data of oil spills from ship are not normally distributed but has a “sharp peak, heavy tail” feature, and the DTD-GPD distribution fitting effect is better than the traditional distribution.

**Key words:** ship's oil spill distribution; DTD-GPD distribution; fitting research

随着我国港航大国的崛起和海洋经济的迅猛发展, 各类溢油事故甚至重大溢油事故时有发生。因此, 深入研究船舶溢油损失分布规律对船舶溢油风险评估和各类溢油事故环境影响评价具有现实意义。

我国船舶溢油损失数据存在着数据收集阈值具有左截断的特性。目前研究中多采用的对数正态、Gamma分布等传统损失分布的随机变量取值范围都是 $[0, +\infty]$ 或 $[-\infty, +\infty]$ , 并未考虑到溢油损失数据的截断特性。另外, 船舶溢油损失数据多呈现出“尖峰厚尾”特点<sup>[1-2]</sup>, 传统损失分布虽然可以对船舶溢油损失的主体部分的“尖峰”特

性进行较好的描述, 却无法有效地对“厚尾”特性进行描述, 因此常常会低估溢油风险。部分学者运用极值理论对损失分布的尾部进行研究来捕捉溢油损失数据的厚尾特征, 却因为只关注尾部而舍弃了大部分珍贵的损失数据, 存在一定的片面性。

为了对船舶溢油损失分布进行完整准确的拟合研究, 本文提出分段建模的思路: 即针对溢油损失的主体部分, 通过传统分布来进行度量; 针对厚尾部分, 引入极值理论进行分析。这种分段建模的思路可以针对损失主体部分和厚尾部分的各自特点选用合适的分布, 从而实现对整个分布

收稿日期: 2013-06-03

作者简介: 田荣洁(1984—), 女, 博士研究生, 主要从事交通环保、溢油应急等方向的研究。

的准确度量。

### 1 双截尾分布理论

溢油损失主体部分是在船舶溢油事故门槛值 $d$ 和极端损失阈值 $u$ 之间的一个两端截断的分布，即船舶溢油损失整体分布的 $[d,u]$ 部分的条件分布。假设 $F(x, \theta)$ 是溢油损失的分布函数， $f(x, \theta)$ 是相应的概率密度函数，其中 $\theta$ 为需要估计的参数。那么双截尾分布函数和密度函数如式(1)和式(2)所示。

$$F^*(X \leq x; \theta | d \leq x \leq u) = \frac{F(x, \theta) - F(d, \theta)}{F(u, \theta) - F(d, \theta)} \quad (1)$$

$$f^*(X \leq x; \theta | d \leq x \leq u) = \frac{f(x, \theta)}{F(u, \theta) - F(d, \theta)} \quad (2)$$

双截尾分布函数可利用传统的极大似然估计法来对参数 $\theta$ 进行估计，其极大似然函数如式(3)所示，其中 $x_i \in [d,u]$ 。

$$l(x, \theta) = \sum_{i=1}^n \ln \left[ \frac{f(x_i, \theta)}{F(u, \theta) - F(d, \theta)} \right] \quad (3)$$

根据式(2)可以计算出传统损失分布的双截尾分布。下面给出正态分布和指数分布在双截尾条件下的密度函数。

双截尾正态分布：

$$f^*(X \leq x | d \leq x \leq u) = \frac{1}{\left[ \phi\left(\frac{u-\mu}{\sigma}\right) - \phi\left(\frac{d-\mu}{\sigma}\right) \right] \sqrt{2\pi} \sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (4)$$

双截尾指数分布：

$$f^*(X \leq x | d \leq x \leq u) = \frac{\lambda e^{-\lambda x}}{e^{-\lambda d} - e^{-\lambda u}} \quad (5)$$

### 2 GPD分布介绍

极值理论是次序统计理论的一个分支，它主要用来研究极端值(极大值或极小值)的分布情况。对于极端风险的度量，极值理论的优势在于它具有超越样本数据的估计能力，不必事先对整个分布做出假设，而是利用数据本身说话。

#### 2.1 GPD分布函数

极值理论的GPD分布最大的特点就是仅描述损失数据中超过阈值 $u$ 的部分，而非描述整个损失

数据。假设船舶溢油损失数据序列 $X_1, \dots, X_n$ 的分布函数为 $F(x)$ ，则GPD的分布函数如式6所示<sup>[3]</sup>：

$$G_{\xi, \sigma}(x) = \begin{cases} 1 - (1 + \frac{\xi x}{\sigma})^{-\frac{1}{\xi}} & \xi > 0 \\ 1 - e^{-x/\sigma} & \xi = 0 \end{cases} \quad (6)$$

其中 $\sigma > 0$ ，并且当 $\xi \geq 0$ 时 $x > 0$ ，当 $\xi < 0$ 时 $0 \leq x \leq -\frac{\sigma}{\xi}$ 。称 $\sigma$ 为分布的尺度参数， $\xi$ 为形状参数，用以描述分布尾部的消失速度。当 $\xi > 0$ 时， $G_{\xi, \sigma}(x)$ 为厚尾分布，如Pareto分布、Loggamma分布和Cauchy分布等。

#### 2.2 阈值的确定

1) 平均超出量函数。

平均超出量函数<sup>[4]</sup>是所有大于阈值 $u$ 的溢油损失值与阈值 $u$ 的差的条件期望。据此定义，可得到GPD分布的平均超出量函数为

$$e(u) = \frac{\sigma + \xi u}{1 - \xi} \quad (7)$$

其中 $0 < \xi < 1$ 时， $\sigma + \xi u > 0$ 。由式(7)可知，GPD分布的平均超出量函数是 $u$ 的线性函数，且函数的斜率可以用来描述分布尾部的厚度。当样本平均函数尾部在超过某一点后呈现为一个正斜率的直线时，这个分界点即为阈值。

2) Hill图。

Hill图就是不同的阈值与相对应的Hill估计统计量 $H_k$ 绘制的图形。Hill指数图中阈值从大到小变化时 $H_k$ 出现的第一个平稳区域的分界点即为阈值。Hill估计是Bruce.M.Hill<sup>[5]</sup>在 $\gamma > 0$ 的条件下，构造的形状参数的非参数化估计方法。令 $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$ 表示独立同分布的次序统计量，固定一个阈值 $u$ ，且假设此时共有 $k$ 个损失大于阈值， $H_k$ 如式(8)所示：

$$\frac{1}{\xi_k} = H_k = \frac{1}{k} \sum_{i=1}^k [\ln X_{(i)} - \ln X_{(k)}] = \frac{1}{k} \sum_{i=1}^k \ln \frac{X_{(i)}}{X_{(k)}} \quad k = 1, 2, \dots, n-1 \quad (8)$$

#### 2.3 参数估计

极大似然估计是近20多年来最重要的点估计方法。对于GPD分布函数，在给定样本 $(X_i, i=1, 2, \dots, n)$ 时，容易得到极大似然函数的表达式<sup>[6]</sup>：

$$\lg L = n \lg \xi - n \lg \sigma - \left( \frac{1}{\xi} + 1 \right) \cdot \sum_{i=1}^n \lg \left( 1 + \frac{\xi}{\sigma} x_i \right) \quad (\xi \neq 0) \quad (9)$$

其中,  $X_i \in D(\sigma, \xi)$ , 找出令  $L(x)$  最大的  $\hat{\xi}$  和  $\hat{\sigma}$ 。对上述公式分别关于  $\hat{\xi}$  和  $\hat{\sigma}$  求偏导, 并令偏导数为零, 化简后可得它的数值解。

### 3 实证分析

收集 1973—2007 年间共 88 个重大船舶溢油事故案例 (不含环境损失), 数据阈值  $d$  设为 1 000 万元。为了能对所搜集到的船舶溢油损失案例有个整体上的认识, 用一般描述对船舶溢油损失数据的统计特征进行了分析, 结果见表 1。

表 1 溢油损失描述性统计结果 万元

均值	标准差	最大值	最小值	偏度	峰度	25%分位数	75%分位数
6 984.62	11 846	77 599	1 000	4.57	26.05	1 755.51	7 421.34

从表 1 可知, 这 84 个溢油损失案例造成的平均损失为 6 984.62 万元, 且可以看出实际溢油损失数据的分布比正态分布有更尖的峰 (峰度系数  $26.05 > 3$ )。实际分布不服从对称分布, 而是明显右偏态分布 (偏度  $4.57 > 0$ ), 均值 6 984 万元和 75% 分位数 7 421 万元仅相差 437 万元, 呈现明显的厚尾性。溢油损失分布“尖峰、厚尾”的特点在溢油事故损失金额的直方图 (图 1) 中也得以验证。

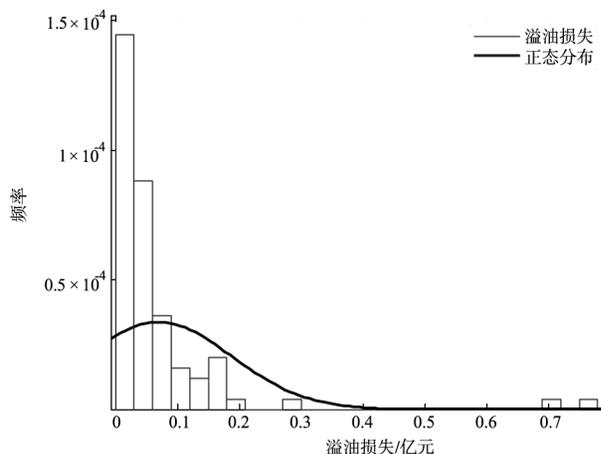


图 1 溢油损失数据直方图

#### 3.1 阈值的确定

依照上文介绍的 GPD 模型阈值选取的原理, 利用式 (7) 和式 (8), 分别绘制了溢油损失平均超出量函数图 (图 2) 和 Hill 图 (图 3)。

从图 2 中可以看出, 平均超出量函数图中图形的斜率发生明显的变化且大致呈线性的地方大约在 1.2 亿左右。从图 3 中看出, 自竖线标记处起, 数据出现第一个平稳区域, 此时分界点大约

为 1.2 亿。因此, 本文选取 1.2 亿作为主体部分和厚尾部分的分界点, 即阈值。

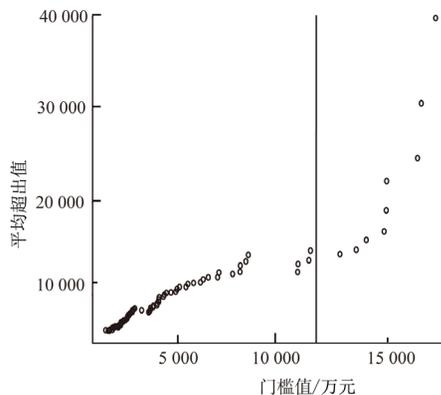


图 2 船舶溢油损失平均超出量函数图

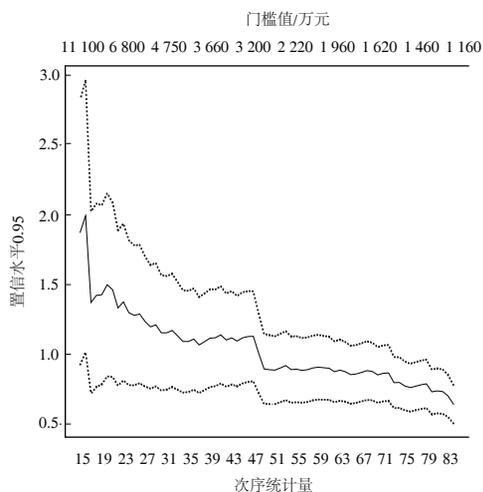


图 3 Hill 图

#### 3.2 基于常数双截尾分布的溢油损失主体部分的拟合

本文将对未截尾条件下 (表 2) 和常数截尾条件下 (表 3) 的两种情形分别建模, 以便能检验双截尾分布对船舶溢油损失分布的拟合效果。

表2 未截尾条件下传统损失分布对溢油损失整体的拟合效果

分布	参数名称	参数估计值	标准差	K-S检验值	P值
正态分布	$\mu$	1 933.6	134.65	0.356 7	0
	$\sigma$	3 498.6	345.66		
指数分布	$\lambda$	1 933.6	493.65	0.176 5	0.034 5
Gamma分布	$\alpha$	0.526 5	0.166 5	0.120 1	0.065 4
	$\beta$	8 654.77	1 226.77		
对数正态分布	$\mu$	7.56	0.89	0.086 0	0.316 5
	$\sigma$	1.39	0.58		
Weibull分布	$\alpha$	5 432.15	493.44	0.346 6	0.008 7
	$\beta$	0.947 6	0.043 2		
Pareto分布	$K$	0.44	0.032 2	0.134 5	0.079 7
	$\sigma$	284 9	145.21		

表2显示未截尾条件下正态分布、指数分布和Weibull分布均没有通过K-S检验。Gamma分布、Pareto分布和对数正态分布均通过了检验，其中对数正态分布的拟合效果较好。进一步对表2中拟合

效果较好的Gamma分布、对数正态分布和Pareto分布在常数截尾条件下进行分析，其参数估计结果见表3。

表3 常数截尾条件下Gamma分布、对数正态分布和Pareto分布对主体部分的拟合结果

分布	参数名称	参数估计值	标准差	K-S检验值	P值
Gamma分布	$\alpha$	0.254 7	0.034 3	0.049 3	0.645 6
	$\beta$	10 384.292 5	1 384.343 2		
对数正态分布	$\mu$	9.541 5	0.563 6	0.032 1	0.849 2
	$\sigma$	1.144 6	0.435 1		
Pareto分布	$K$	0.343 7	0.045 6	0.043 2	0.729 2
	$\sigma$	6 894.237 3	101.334 2		

从表3可以看出，在常数双截尾条件下，Gamma分布、对数正态分布和Pareto分布3个双截尾分布都对船舶溢油主体部分的损失强度有较佳的拟合，三者的拟合均通过了显著性检验。与未考虑截尾条件下的拟合结果对比（表2），可以发现双截尾分布的K-S检验值远小于未截尾的完整分布，其拟合的显著性水平也有明显提升。同时，在未截尾条件下，3个分布的拟合效果之间存在显著的差距；而在双截尾条件下，三者的拟合效果比较接近。

### 3.3 基于GPD模型的溢油损失厚尾部分的拟合

对大于阈值 $u$ 的损失数据用GPD分布来进行拟合，参数的估计值见表4。

表4 溢油损失厚尾部分的GPD分布的拟合结果

参数名称	参数估计值	标准差	K-S检验值	P值
形状参数	0.323 1	0.283 6	0.237 0	0.282 7
尺度参数	9 863.46	3 643.11		

从表4中可知，形状参数 $\xi > 0$ ，说明超过阈值1.2亿元的损失数据的尾部消失速度很慢，即可知溢油损失数据的分布具有“厚尾”的特性。同时，根据K-S检验值和P值可知，GPD分布对重大溢油损失分布拟合得较好，重大溢油损失数据和GPD的累积概率图也证明了这一点(图4)。

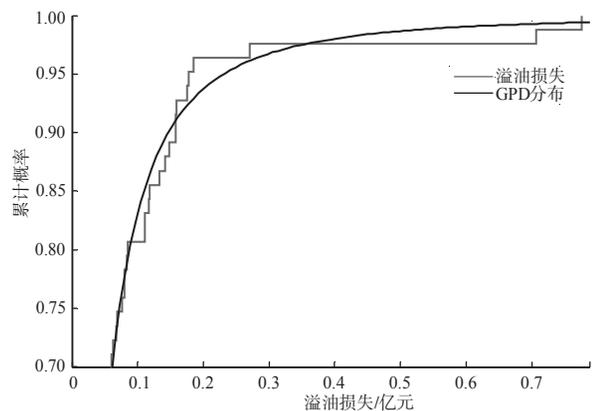


图4 GPD模型厚尾部分拟合效果 (累积概率)