第28卷 第3期 北京工业大学学报 Vol.28 No.3
2002年 9月 JOURNAL OF BEIJING POLYTECHNIC UNIVERSITY Sept. 2002

# 利用样本分位数的极值分布的参数估计

程维虎
（北京工业大学 应用数理学院，北京 100022）

**摘 要**：基于极值分布的若干个样本分位数，建立了分布参数的线性回归模型，得到了分布参数的渐近正态无偏估计，对分布参数进行了渐近置信估计．通过变量代换，将具有形状参数和刻度参数的两参数 Weibull 分布变成极值分布，进一步得到这类两参数 Weibull 分布参数的渐近置信估计．

**关键词**：极值分布；Weibull 分布；参数估计；渐近正态性
**中图分类号**：O 213.2 **文献标识码**：A **文章编号**：0254-0037(2002)03-0326-03

## 1 问题的提出

极值分布是一种常见的寿命分布．在工程实践中，许多产品、部件或系统的应力强度或载荷强度等都服从极值分布．

极值总体的分布函数为

$$F(x;\mu,\sigma)=1-\exp\{-\exp[(x-\mu)/\sigma]\},\quad -\infty<x<\infty \tag{1}$$

其中 $-\infty<\mu<\infty,\sigma>0$ 为常数，分别称分布的位置参数和尺度参数．由式(1)，得总体的分布密度函数

$$f(x;\mu,\sigma)=(1/\sigma)\exp[-(x-\mu)/\sigma]\exp\{-\exp[(x-\mu)/\sigma]\},\quad -\infty<x<\infty \tag{2}$$

设 $X_1$、$X_2$、…、$X_n$ 是抽自极值总体(1)的一个简单样本，将 $X_1$、$X_2$、…、$X_n$ 按从小到大的次序排列成 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$，则称 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$ 为样本的次序统计量．

取定一组正数 $p_1$、$p_2$、…、$p_k$，使其满足 $0<p_1<p_2<\cdots<p_k<1$，对 $i=1,2,\cdots,k$，令 $n_i=[np_i]+1$，其中 $[np_i]$ 为不超过 $np_i$ 的最大整数．记 $Y_i$ 为样本的 $p_i$ 分位数，即 $Y_i=X_{(n_i)}$，$\xi_i$ 为分布的 $p_i$ 分位数，即 $F(\xi_i;\mu,\sigma)=p_i$．由 $F(\xi_i;\mu,\sigma)=p_i$ 及公式(1)和(2)，可导出

$$\xi_i=\mu+\sigma\ln[\ln(1-p_i)^{-1}],\ f(\xi_i;\mu,\sigma)=\sigma^{-1}(1-p_i)\ln(1-p_i)^{-1},\ i=1,2,\cdots,k \tag{3}$$

由 $n\to\infty$ 时，$\sqrt{n}\,(Y_1-\xi_1,Y_2-\xi_2,\cdots,Y_k-\xi_k)'\xrightarrow{L}N(\boldsymbol{O},\boldsymbol{\Lambda})$，其中 $\boldsymbol{O}'=(0,0,\cdots,0)_{1\times k}$，并且

$$\boldsymbol{\Lambda}=(\lambda_{ij})_{k\times k},\ \lambda_{ij}=\sigma^2p_i/[(1-p_i)\ln(1-p_i)\ln(1-p_j)],\ 1\leqslant i\leqslant j\leqslant k,\ \lambda_{ji}=\lambda_{ij} \tag{4}$$

建立线性回归模型

$$Y_i=\mu+\sigma\ln[\ln(1-p_i)^{-1}]+e_i,\ i=1,2,\cdots,k,\ \sqrt{n}(e_1,e_2,\cdots,e_k)'\xrightarrow{L}N(\boldsymbol{O},\boldsymbol{\Lambda}) \tag{5}$$

对于模型(5)，若 $\mu$ 与 $\sigma$ 的最小二乘解 $\hat\mu$ 与 $\hat\sigma$ 分别为 $\mu$ 和 $\sigma$ 的渐近正态无偏估计量，可利用 $\hat\mu$ 和 $\hat\sigma$ 构造出 $\mu$ 和 $\sigma$ 的渐近置信区间．

## 2 问题的解决

欲构造极值分布位置参数及尺度参数的渐近置信估计，需了解如下内容．

**引理 1**[1] 设 $X_1$、$X_2$、…、$X_n$ 是抽自总体分布函数为 $F(x)$ 的一个简单样本，$F(x)$ 绝对连续．取定一组正数 $p_1$、$p_2$、…、$p_k$，使其满足 $0<p_1<p_2<\cdots<p_k<1$．对 $i=1,2,\cdots,k$，记 $Y_i$ 为样本的 $p_i$ 分位数，$\xi_i$ 为分

收稿日期：2002-04-08．
基金项目：北京市优秀青年骨干基金资助项目．
作者简介：程维虎(1962-)，男，副教授，在职博士生．

布的 $p_i$ 分位数，设 $f(x)=F'(x)$ 在点 $\xi_1、\xi_2、\cdots、\xi_k$ 处连续且不为 0．则当 $n\to\infty$ 时，有

$$\sqrt{n}(Y_1-\xi_1,Y_2-\xi_2,\cdots,Y_k-\xi_k)'\xrightarrow{L}N(\boldsymbol{O},\boldsymbol{\Lambda})\tag{6}$$

此处 $\boldsymbol{O}'=(0,0,\cdots,0)_{1\times k}$，$\boldsymbol{\Lambda}=(\lambda_{ij})_{k\times k}$，其中

$$\lambda_{ij}=p_i(1-p_j)/[f(\xi_i)f(\xi_j)],\ 1\leqslant i\leqslant j\leqslant k,\ \lambda_{ji}=\lambda_{ij}\tag{7}$$

**引理 2**[12]　设 $\{Z_n\}$、$\{U_n\}$ 是两个随机变量序列，$Z$ 为随机变量，$c$ 为常数，$Z_n\xrightarrow{L}Z$，$U_n\xrightarrow{P}c$．则有：①$Z_n+U_n\xrightarrow{L}Z+c$；②$U_nZ_n\xrightarrow{L}cZ$；③$Z_n/U_n\xrightarrow{L}Z/c$，当 $c\neq0$ 时．

设 $X_1、X_2、\cdots、X_n$ 为抽自极值总体的一个简单样本，其总体的分布函数由式(1)确定，则 $F(x;\mu,\sigma)$ 关于 $x$ 绝对连续．取定一组正数 $0<p_1<p_2<\cdots<p_k$，对 $i=1,2,\cdots,k$，记 $Y_i$ 为样本的 $p_i$ 分位数，$\xi_i$ 为分布的 $p_i$ 分位数，则 $f(x;\mu,\sigma)=F'(x;\mu,\sigma)$ 在点 $\xi_1、\xi_2、\cdots、\xi_k$ 处连续且不为 0．于是，当 $n\to\infty$ 时，式(6)成立．

对 $i=1,2,\cdots,k$，由 $F(\xi_i;\mu,\sigma)=p_i$ 及公式(1)和(2)，易导出式(3)成立．将式(3)代入到式(7)，可得式(4)成立．再由式(6)成立，可知 $n\to\infty$ 时，$\sqrt{n}(Y_1-\xi_1,Y_2-\xi_2,\cdots,Y_k-\xi_k)'\xrightarrow{L}N(\boldsymbol{O},\boldsymbol{\Lambda})$．

由式(3)、(4)及上述结论，可建立线性模型式(5)．由文献[3]可知，$\mu$ 和 $\sigma$ 的最小二乘解 $\hat{\mu}$ 与 $\hat{\sigma}$ 可分别表示成

$$\begin{aligned}\hat{\mu}&=\Big(\sum_{i=1}^{k}Y_i\sum_{i=1}^{k}a_i^2-\sum_{i=1}^{k}a_i\sum_{i=1}^{k}a_iY_i\Big)\Big/\Big[k\sum_{i=1}^{k}a_i^2-\Big(\sum_{i=1}^{k}a_i\Big)^2\Big]\\ \hat{\sigma}&=\Big(k\sum_{i=1}^{k}a_iY_i-\sum_{i=1}^{k}a_i\sum_{i=1}^{k}Y_i\Big)\Big/\Big[k\sum_{i=1}^{k}a_i^2-\Big(\sum_{i=1}^{k}a_i\Big)^2\Big]\end{aligned}\tag{8}$$

其中 $a_i=\ln[\ln(1-p_i)^{-1}]$，$Y_i=X_{([np_i]+1)}$，$i=1,2,\cdots,k$．令

$$\begin{gathered}\boldsymbol{X}=\begin{pmatrix}1&\cdots&1\\a_1&\cdots&a_k\end{pmatrix}',\ \ \boldsymbol{V}=(b_{ij})_{k\times k},\\ b_{ij}=p_i/[(1-p_i)\ln(1-p_i)\ln(1-p_j)],\ 1\leqslant i\leqslant j\leqslant k,\ b_{ji}=b_{ij}\end{gathered}\tag{9}$$

当 $0<p_1<p_2<\cdots<p_k<1$ 时，$\boldsymbol{X}'\boldsymbol{X}$ 及 $\boldsymbol{V}$ 均可逆．如再令

$$\boldsymbol{C}=(c_{ij})_{2\times2}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1},\ \ \boldsymbol{O}=(0,0)'\tag{10}$$

则有 $\sqrt{n}(\hat{\mu}-\mu,\hat{\sigma}-\sigma)'\xrightarrow{L}N(\boldsymbol{O},\sigma^2\boldsymbol{C})$．于是

$$\sqrt{n}(\hat{\mu}-\mu)/(\sigma\sqrt{c_{11}})\xrightarrow{L}N(0,1),\ \ \sqrt{n}(\hat{\sigma}-\sigma)/(\sigma\sqrt{c_{22}})\xrightarrow{L}N(0,1)\tag{11}$$

故 $\hat{\mu}$ 和 $\hat{\sigma}$ 分别为 $\mu$ 和 $\sigma$ 的渐近正态且渐近无偏估计量．

由式(11)及引理可得如下结果．

**定理 1**　设 $X_1、X_2、\cdots、X_n$ 为抽自极值总体(1)的一个简单样本，记样本的次序统计量为 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$，任给一组正数 $0<p_1<p_2<\cdots<p_k<1$，对 $i=1,2,\cdots,k$，记 $Y_i=X_{([np_i]+1)}$，$a_i=\ln[\ln(1-p_i)^{-1}]$，对给定的置信水平 $0<\alpha<1$，$\sigma$ 的置信度为 $1-\alpha$ 的渐近置信区间为 $[\sqrt{n}\hat{\sigma}/(\sqrt{n}+\sqrt{c_{22}}Z_{\alpha/2}),\sqrt{n}\hat{\sigma}/(\sqrt{n}-\sqrt{c_{22}}Z_{\alpha/2})]$．其中 $\hat{\sigma}$ 由式(8)给出，$c_{22}$ 由式(9)、(10)给出，$Z_{\alpha/2}$ 为标准正态分布的右 $\alpha/2$ 分位点．

**证**　由式(11)的后半部，得 $\sigma$ 的置信度为 $1-\alpha$ 的渐近置信区间为满足不等式 $-Z_{\alpha/2}\leqslant[\sqrt{n}(\hat{\sigma}-\sigma)/(\sigma\sqrt{c_{22}})]\leqslant Z_{\alpha/2}$ 的 $\sigma$ 所构成的区间．作等价变换，得 $\sqrt{n}\hat{\sigma}/(\sqrt{n}+\sqrt{c_{22}}Z_{\alpha/2})\leqslant\sigma\leqslant\sqrt{n}\hat{\sigma}/(\sqrt{n}-\sqrt{c_{22}}Z_{\alpha/2})$．定理 1 得证．

**定理 2**　设 $X_1、X_2、\cdots、X_n$ 为抽自极值总体(1)的一个简单样本，记样本的次序统计量为 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$，任给一组正数 $0<p_1<p_2<\cdots<p_k<1$，对 $i=1,2,\cdots,k$，记 $Y_i=X_{([np_i]+1)}$，$a_i=\ln[\ln(1-p_i)^{-1}]$，对给定的置信水平 $0<\alpha<1$，$\mu$ 的置信度为 $1-\alpha$ 的渐近置信区间为 $[\hat{\mu}-\sqrt{c_{11}/n}\,\hat{\sigma}Z_{\alpha/2},\hat{\mu}+\sqrt{c_{11}/n}\,\hat{\sigma}Z_{\alpha/2}]$．其中 $\hat{\mu}$ 和 $\hat{\sigma}$ 由式(8)给出，$c_{11}$ 由式(9)、(10)给出，$Z_{\alpha/2}$ 为标准正态分布的右 $\alpha/2$ 分位点．

**证**　由式(11)的后半部，得 $\hat{\sigma}\xrightarrow{P}\sigma$．结合式(11)的前半部，利用引理 2 中③的结论，得 $\sqrt{n}(\hat{\mu}-\mu)/(\hat{\sigma}\sqrt{c_{11}})\xrightarrow{L}N(0,1)$．从而，$\mu$ 的置信度为 $1-\alpha$ 的渐近置信区间为满足不等式 $-Z_{\alpha/2}\leqslant\sqrt{n}(\hat{\mu}-\mu)/(\hat{\sigma}\sqrt{c_{11}})\leqslant Z_{\alpha/2}$ 的 $\mu$ 所构成的区间．对上式作等价变换，得 $\hat{\mu}-\sqrt{c_{11}/n}\,\hat{\sigma}Z_{\alpha/2}\leqslant\mu\leqslant\hat{\mu}+\sqrt{c_{11}/n}\,\hat{\sigma}Z_{\alpha/2}$．定理 2 得证．

## 3 Weibull 分布参数的渐近置信估计

Weibull 分布是重要的寿命分布，许多电子产品或元件的寿命都服从这一分布．所以，它在可靠性理论中占有非常重要的地位．

Weibull 总体分布函数为

$$F(x;\alpha,\beta)=1-\exp(-x^{\alpha}/\beta),\quad 0<x<\infty \tag{12}$$

其中$\alpha>0,\beta>0$为常数，分别称分布的形状参数和尺度参数．由式(12)，得 Weibull 分布的分布密度函数为

$$f(x;\alpha,\beta)=(\alpha/\beta)x^{\alpha-1}\exp(-x^{\alpha}/\beta),\quad 0<x<\infty \tag{13}$$

若总体 $X$ 的分布函数由式(12)给出，令 $Y=\ln X,\mu=\alpha^{-1}\ln\beta,\sigma=\alpha^{-1}$，则 $Y$ 的分布函数可由式(1)给出．于是，由本文的定理 1、定理 2 及引理 2，不难得到如下结果．

**定理 3**　设$X_1$、$X_2$、…、$X_n$为抽自 Weibull 总体式(12)的一个简单样本，记样本的次序统计量为$(X_{(1)},X_{(2)},\cdots,X_{(n)})$，任给一组正数$0<p_1<p_2<\cdots<p_k<1$，对$i=1,2,\cdots,k$，记 $Y_i=\ln X_{([np_i]+1)}$，$a_i=\ln[\ln(1-p_i)^{-1}]$．对给定的置信水平$0<\gamma<1$，$\alpha$ 的置信度为$1-\gamma$ 的渐近置信区间为$[(1-\sqrt{c_{22}/n}\ Z_{\gamma/2})\hat{\alpha},(1+\sqrt{c_{22}/n}\ Z_{\gamma/2})\hat{\alpha}]$．其中：$\hat{\alpha}=[k\sum_{i=1}^{k}a_i^2-(\sum_{i=1}^{k}a_i)^2]/[k\sum_{i=1}^{k}a_iY_i-\sum_{i=1}^{k}a_i\sum_{i=1}^{k}Y_i]$；$c_{22}$由式(9)、(10)给出；$Z_{\gamma/2}$为标准正态分布的右$\gamma/2$分位点．

**定理 4**　设 $X_1$、$X_2$、…、$X_n$为抽自 Weibull 总体式(12)的一个简单样本，记样本的次序统计量为$(X_{(1)},X_{(2)},\cdots,X_{(n)})$，任给一组正数$0<p_1<p_2<\cdots<p_k<1$，对$i=1,2,\cdots,k$，记 $Y_i=\ln X_{([np_i]+1)}$，$a_i=\ln[\ln(1-p_i)^{-1}]$；对给定的置信水平$0<\gamma<1$，$\beta$的置信度为$1-\gamma$的渐近置信区间为$[\hat{\beta}\exp(-\sqrt{c_{11}/n}\ Z_{\gamma/2}),\hat{\beta}\exp(\sqrt{c_{11}/n}\ Z_{\gamma/2})]$．其中：$\hat{\beta}=\exp[\sum_{i=1}^{k}Y_i\sum_{i=1}^{k}a_i^2-\sum_{i=1}^{k}a_i\sum_{i=1}^{k}a_iY_i]/[k\sum_{i=1}^{k}a_iY_i-\sum_{i=1}^{k}a_i\sum_{i=1}^{k}Y_i]$；$c_{11}$由式(9)、(10)给出；$Z_{\gamma/2}$为标准正态分布的右$\gamma/2$分位点．

利用 $n\to\infty$时，$\exp(-\sqrt{c_{11}/n}\ Z_{\gamma/2})\approx 1-\sqrt{c_{11}/n}\ Z_{\gamma/2}$，还可得$\beta$的置信度为$1-\gamma$的渐近置信区间为$[(1-\sqrt{c_{11}/n}\ Z_{\gamma/2})\hat{\beta},(1+\sqrt{c_{11}/n}\ Z_{\gamma/2})\hat{\beta}]$．其中$\hat{\beta}$、$c_{11}$、$Z_{\gamma/2}$同定理 4．

**参考文献：**

[1] 陈希孺．数理统计引论[M]．北京：科学出版社，1997．

[2] 茆诗松，濮晓龙，王静龙．高等数理统计[M]．北京：高等教育出版社，1998．

[3] 王松桂，陈立萍，陈敏．线性统计模型[M]．北京：高等教育出版社，1999．

# Parameter Estimation for Extreme-value Distribution Based on the $p_i$-th Quantiles of Samples

CHENG Wei-hu

( College of Applied Sciences, Beijing Polytechnic University, Beijing 100022, China )

**Abstract:** The asymptotic normal and asymptotic unbiased estimation of distribution parameter and scale parameter is proposed by linear regression model, based on $k$ $p_i$-th quantiles of extreme-value distribution simple sample. The asymptotic confidence estimation of the parameter is given for extreme-value distribution. The Weibull distribution of shape parameter and scale parameter is turned into an extreme-value distribution. And thus the asymptotic confidence estimation for the Weibull distribution of this kind of parameters is also obtaind.

**Key words:** extreme-value distribution; Weibull distribution; parameter estimation; asymptotic normality