



中华人民共和国气象行业标准

QX/T 529—2019

气候可行性论证规范 极值概率统计分析

Specifications for climatic feasibility demonstration—Probability and statistic
analysis of extremum

2019-12-26 发布

2020-04-01 实施

中 国 气 象 局 发 布

目 次

前言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 资料收集与处理	2
4.1 资料收集	2
4.2 极值序列的均一性处理	2
5 极值分布类型选取	2
6 极值分布参数估计	2
6.1 经验概率分布绘点	2
6.2 参数估计	2
6.2.1 耿贝尔(Gumbel)分布	2
6.2.2 皮尔逊Ⅲ型(Pearson Ⅲ)分布	2
6.2.3 威布尔(Weibull)分布	3
6.2.4 广义极值(GEV)分布	3
6.2.5 泊松-耿贝尔(Poisson-Gumbel)复合分布	3
6.2.6 其他分布	3
7 拟合优度综合分析	3
7.1 分布函数符合性检验	3
7.2 拟合优度分析	3
7.3 综合判定	3
7.4 其他方法	3
8 极值重现期	4
8.1 累积概率	4
8.2 重现期计算	4
附录 A(资料性附录) 极值序列及其经验概率分布	5
A.1 极值序列	5
A.2 极值序列的经验概率分布	5
附录 B(资料性附录) 极值分布及其参数估计方法	6
B.1 耿贝尔(Gumbel)分布及其参数估计方法	6
B.2 皮尔逊Ⅲ型(Pearson Ⅲ)分布及其参数估计方法	7
B.3 威布尔(Weibull)分布及其参数估计方法	12
B.4 广义极值(GEV)分布及其参数估计方法	13
B.5 泊松-耿贝尔(Poisson-Gumbel)复合分布及其参数估计方法	14
附录 C(资料性附录) 拟合优度检验方法	16
C.1 柯尔莫哥洛夫-斯米诺夫(Kolmogorov-Smirnov)检验法	16
C.2 剩余方差	16
C.3 拟合相对偏差	16

QX/T 529—2019

参考文献	17
------------	----

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准由全国气候与气候变化标准化技术委员会(SAC/TC 540)提出并归口。

本标准起草单位:中国气象局沈阳大气环境研究所、沈阳区域气候中心、广东省气象防灾技术服务中心。

本标准主要起草人:汪宏宇、龚强、黄浩辉。

气候可行性论证规范 极值概率统计分析

1 范围

本标准规定了气候可行性论证中涉及气象要素极值的资料收集与处理、分布类型选取、分布参数估计、拟合优度综合分析、重现期计算等概率统计分析方法及要求。

本标准适用于工程项目规划、设计、建设等的气候可行性论证。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

QX/T 457—2018 气候可行性论证规范 气象资料加工处理

QX/T 469—2018 气候可行性论证规范 总则

3 术语和定义

下列术语和定义适用于本文件。

3.1

极值 extremum

在一定时间段内某要素的极大值或极小值。

注:改写 GB/T 34293—2017,定义 2.4。

3.2

概率分布 probability distribution

用以表述随机变量取值的概率规律。

注:根据随机变量所属类型的不同,概率分布取不同的表现(函数)形式。

3.3

总体分布 population distribution

随机变量总体的概率分布。

注:研究对象(随机变量)的全体为总体。

3.4

参数估计 parameter estimation

利用实测有限样本及概率分布函数,估计概率分布函数中待定参数的值的过程。

3.5

重现期 recurrence interval

统计量的特定值重复出现的统计时间间隔。

注 1:也称平均再现间隔,单位一般为年(a)。即可能出现两次大于或等于某特定强度值的极端事件之间的平均间隔时间。该强度事件常被称为“××年一遇”事件。

注 2:改写 GB/T 34293—2017,定义 2.9。

QX/T 529—2019

4 资料收集与处理

4.1 资料收集

根据气候可行性论证项目的要求,宜按年极值法收集气象要素极值序列,如最大风速、极大风速、最高温度、最低温度、不同历时最大降水量、最大雪深、最大冻土深度、热带气旋中心最低气压等。极值序列应采用当地观测最长历史记录,时间序列应不少于 30 年。资料收集应满足 QX/T 469—2018 第 6 章的要求。

4.2 极值序列的均一性处理

气象站的观测场址、观测仪器型号、观测方式等在各历史阶段有变动时,在做极值统计分析之前应进行必要的一致性检验、订正和插补等处理。包括 10 分钟和 2 分钟最大风速换算、观测高度标准化修正、极大风速和最大风速换算以及观测场环境变化影响的修正等。资料应按照 QX/T 469—2018 第 9 章和 QX/T 457—2018 第 5 章的要求和方法处理。

5 极值分布类型选取

5.1 应首先按照论证项目相关的工程标准和规范,选取其中推荐的极值分布类型。

5.2 如无相关标准和规范,一般采用耿贝尔(Gumbel)分布作为初选;对随机性强的要素(如极端风速、极端降水等)宜采用皮尔逊Ⅲ型(Pearson Ⅲ)分布作为初选;对于极值序列有间断的(如区域台风极端风速等),宜采用泊松-耿贝尔(Poisson-Gumbel)复合分布作为初选。

5.3 标准和规范推荐的和初选的极值分布,都应按照第 7 章的要求进行拟合优度综合分析。

5.4 当 5.1 和 5.2 中极值分布不适合于所分析气象要素极值序列或有更适合的其他极值分布时,可选用威布尔(Weibull)分布、广义极值(GEV)分布等其他极值分布,但应在进行拟合优度综合分析基础上对比结果并说明。

6 极值分布参数估计

6.1 经验概率分布绘点

经验概率分布的曲线绘点一般可采用威布尔(Weibull)概率均值法确定,参见附录 A 的 A.2。

6.2 参数估计

6.2.1 耿贝尔(Gumbel)分布

耿贝尔分布函数中的常用参数估计方法有矩法、耿贝尔法、极大似然法、最小二乘法等。宜采用耿贝尔法或最小二乘法进行参数估计,参见附录 B 的 B.1。

注:耿贝尔分布(又称极值 I 型概率分布)是一个较完全的极值理论分布,是在样本容量很大时的极限分布。

6.2.2 皮尔逊Ⅲ型(Pearson Ⅲ)分布

皮尔逊Ⅲ型分布函数中的常用参数估计方法有矩法、线性矩法、极大似然法、数值积分单权函数法、数值积分双权函数法、适线法、极大值调整适线法等。宜采用线性矩法、数值积分单(或双)权函数法或适线法进行参数估计,参见附录 B 的 B.2。

注:有相当多的自然现象符合皮尔逊Ⅲ型分布,在降水和水文领域中应用较多。

6.2.3 威布尔(Weibull)分布

威布尔分布函数中的常用参数估计方法有矩法、极大似然法、最小二乘法等。宜采用最小二乘法或矩法进行参数估计,参见附录 B 的 B.3。

注:威布尔分布(韦伯分布)是一个极值理论分布,常被应用于风速的概率分布和风能资源的研究中。

6.2.4 广义极值(GEV)分布

将耿贝尔分布、弗雷歇分布、威布尔分布整合到一个分布函数中,称之为广义极值分布。其分布函数中的常用参数估计方法有线性矩法、极大似然法等,参见附录 B 的 B.4。

6.2.5 泊松-耿贝尔(Poisson-Gumbel)复合分布

某随机事件属于离散型分布,在该类事件影响下的某要素极值可以构成连续型分布。若该随机事件符合泊松分布,且该类事件影响下的某要素极值符合耿贝尔分布,则可应用泊松-耿贝尔复合分布对其进行分析。其分布函数中的参数估计方法可采用耿贝尔法等,参见附录 B 的 B.5。

6.2.6 其他分布

当需采用 6.2.1—6.2.5 分布及其参数估计方法之外的方法时,应在进行拟合优度综合分析基础上与上述分布及其参数估计方法进行对比。

7 拟合优度综合分析

7.1 分布函数符合性检验

可采用柯尔莫哥洛夫-斯米诺夫(Kolmogorov-Smirnov)检验法确定极值序列是否服从某型极值分布,参见附录 C 的 C.1。

7.2 拟合优度分析

对于超过两种通过符合性检验的备选分布函数,可用剩余方差、拟合相对偏差等方法进行拟合优度分析,参见附录 C 的 C.2 和 C.3。

也可比较极值序列(按从小到大顺序排列)最后 15%样本(大值样本)的剩余方差,其值小者为优;若剩余方差一致,则可再比较 15%大值样本的拟合相对偏差,其值小者为优,参见附录 B 的 B.2.2.6 和 B.2.2.7。

7.3 综合判定

应绘制极值分布拟合曲线和实测数据经验概率分布对比图,若相关工程规范规定的极值分布函数拟合效果不佳,可采用多种极值分布函数进行拟合试验,选取最佳函数结果,并结合工程项目安全性和经济性需求,综合分析确认适合该气象要素极值序列的极值分布类型和相应参数。

7.4 其他方法

当需采用 7.1 和 7.2 之外的检验分析方法时,应对比分析检验结果并说明。

QX/T 529—2019

8 极值重现期

8.1 累积概率

把随机变量 X 不超过某个定值 x_p 发生的概率叫作累积概率(左侧概率),见式(1):

$$P(X \leq x_p) = F(x_p) \quad \dots\dots\dots(1)$$

式中:

$P(X \leq x_p)$ —— X 不超过某个定值 x_p 发生的概率。

$F(x_p)$ —— 累积概率(左侧概率)。

8.2 重现期计算

极值的重现期 $T(x_p)$ 的计算见式(2):

$$T(x_p) = \frac{1}{1 - P(X \leq x_p)} = \frac{1}{1 - F(x_p)} \quad \dots\dots\dots(2)$$

附录 A

(资料性附录)

极值序列及其经验概率分布

A.1 极值序列

A.1.1 序列

设序列 $\{x_i\}$, x_i 为序列中第 i 个样本的值。

$\bar{x} = E(x)$ 是序列 $\{x_i\}$ 的数学期望,即均值。

$\sigma = \sigma(x)$ 是序列 $\{x_i\}$ 的标准差。

在实际计算中可用有限样本容量的均值和标准差作为 $E(x)$ 和 $\sigma(x)$ 的估计值。对待分析极值序列进行排序后得: $x_1 \leq x_2 \leq \cdots \leq x_n$, n 为该序列样本总数,作为序列 $\{x_i\}$ 的一个有限容量且有序的样本序列。

A.1.2 样本序列均值的估计值

样本序列均值 \bar{x} 的估计值 $\hat{\bar{x}}$ 见式(A.1):

$$\hat{\bar{x}} = \frac{1}{n} \sum_{i=1}^n x_i \quad \cdots \cdots \cdots (A.1)$$

A.1.3 样本序列标准差的估计值

样本序列标准差 σ 的估计值 $\hat{\sigma}$ 见式(A.2):

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\bar{x}})^2} \quad \cdots \cdots \cdots (A.2)$$

A.1.4 样本序列偏度系数的估计值

样本序列偏度系数 r_1 的估计值 \hat{r}_1 见式(A.3):

$$\hat{r}_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\bar{x}})^3}{\hat{\sigma}^3} \quad \cdots \cdots \cdots (A.3)$$

A.2 极值序列的经验概率分布

$F_w(x_i)$ 是极值有序序列在 x_i 处的经验概率值,决定了序列在概率图上的绘点位置,有多种公式可选。选择何种公式对采用适线法进行参数估计时有较大影响。一般情况下,经验概率曲线绘点采用威布尔(Weibull)概率均值法(即常见的经验分布算法),见式(A.4):

$$F_w(x_i) = \frac{i}{n+1} \quad (i = 1, 2, \cdots, n) \quad \cdots \cdots \cdots (A.4)$$

QX/T 529—2019

附录 B

(资料性附录)

极值分布及其参数估计方法

B.1 耿贝尔(Gumbel)分布及其参数估计方法

B.1.1 耿贝尔分布

耿贝尔分布的概率密度函数 $f(x)$ 和累积分布函数 $F(x)$ 分别见式(B.1)和式(B.2):

$$f(x) = ae^{-a(x-b)-e^{-a(x-b)}} \quad (-\infty < x < +\infty, a > 0, -\infty < b < +\infty) \quad \text{.....(B.1)}$$

$$F(x) = e^{-e^{-a(x-b)}} \quad \text{.....(B.2)}$$

式中:

a ——尺度参数;

b ——位置参数。

可设转换变量 $y = a(x - b)$ 以方便计算。

已知累积概率 $P(x_P)$, 按式(B.3)求极端事件的极值:

$$x_P = -\ln(-\ln(P(x_P)))/a + b \quad \text{.....(B.3)}$$

B.1.2 耿贝尔分布常用参数估计方法

B.1.2.1 矩法

通过积分可得转换变量 y 的数学期望和方差, 见式(B.4):

$$\begin{cases} E(y) = 0.5772 \\ \sigma(y)^2 = \frac{\pi^2}{6} \end{cases} \quad \text{.....(B.4)}$$

由 y 和 x 的关系可得:

$$\begin{cases} a = \frac{\sigma(y)}{\sigma(x)} = \frac{1.28255}{\sigma(x)} \\ b = E(x) - \frac{E(y)\sigma(x)}{\sigma(y)} = E(x) - 0.45004\sigma(x) \end{cases} \quad \text{.....(B.5)}$$

将附录 A 中 A.1 的序列均值和标准差的估计值(参见式(A.1)和式(A.2))代入, 得到参数 a 和 b 的估计值。

B.1.2.2 耿贝尔法

耿贝尔法也属于矩估计法。

按式(B.6)构造一个新序列 $\{y_{G,i}\}$, 其中:

$$y_{G,i} = -\ln(-\ln(F_W(x_i))) \quad (i = 1, 2, \dots, n) \quad \text{.....(B.6)}$$

式中:

$F_W(x_i)$ ——经验概率分布函数, 见式(A.4)。

可得:

$$\begin{cases} a = \frac{\sigma(y_G)}{\sigma(x)} \\ b = E(x) - \frac{E(y_G)\sigma(x)}{\sigma(y_G)} \end{cases} \dots\dots\dots (B.7)$$

式中:

$\sigma(x)$ ——序列 $\{x_i\}$ 的标准差;

$\sigma(y_G)$ ——序列 $\{y_{G,i}\}$ 的标准差;

$E(x)$ ——序列 $\{x_i\}$ 的数学期望;

$E(y_G)$ ——序列 $\{y_{G,i}\}$ 的数学期望。

利用附录 A 中 A.1 的方法可分别得到它们的估计值,参见式(A.1)和式(A.2),代入式(B.7)中,得到参数 a 和 b 的估计值。

B.1.2.3 极大似然法

在统计学理论上,在知道总体分布类型的情况下,极大似然估计是一种较优的参数估计方法。首先根据概率密度函数得似然函数,见式(B.8):

$$L = \sum_{i=1}^n \ln f(x_i) = na - \sum_{i=1}^n a(x_i - b) - \sum_{i=1}^n e^{-a(x_i - b)} \dots\dots\dots (B.8)$$

将 a 和 b 看作变量,将式(B.8)分别对 a 和 b 求导并令其为 0,得极大似然方程组,见式(B.9):

$$\begin{cases} \frac{\partial L}{\partial a} = n - \sum_{i=1}^n (x_i - b) + \sum_{i=1}^n (x_i - b)e^{-a(x_i - b)} = 0 \\ \frac{\partial L}{\partial b} = a(n - \sum_{i=1}^n e^{-a(x_i - b)}) = 0 \end{cases} \dots\dots\dots (B.9)$$

式(B.9)是一非线性方程组,可在适当的初值下用迭代法求解,即可得到参数 a 和 b 的估计值。

B.1.2.4 最小二乘法

对耿贝尔分布函数两边取两次对数,得式(B.10):

$$-\ln(-\ln(F(x))) = a(x - b) \dots\dots\dots (B.10)$$

令 $y_G = -\ln(-\ln(F(x)))$,则得式(B.11):

$$y_G = ax - ab \dots\dots\dots (B.11)$$

式(B.11)是一直线方程,对 B.1.2.2 中的序列 $\{x_i\}$ 和 $\{y_{G,i}\}$ 用最小二乘法进行拟合,即可求出参数 a 和 b 的估计值。

B.2 皮尔逊 III 型(Pearson III)分布及其参数估计方法

B.2.1 皮尔逊 III 型分布

皮尔逊 III 型分布的概率密度函数 $f(x)$ 和累积分布函数 $F(x)$ 分别见式(B.12)和式(B.13):

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}(x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} \quad (\alpha, \beta > 0; x \geq a_0) \dots\dots\dots (B.12)$$

$$F(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{a_0}^x (t - a_0)^{\alpha-1} e^{-\beta(t-a_0)} dt \dots\dots\dots (B.13)$$

式中:

α ——形状参数;

β ——比例参数(尺度参数);

QX/T 529—2019

a_0 ——位置参数。

皮尔逊Ⅲ型分布其他常用参数有：

- a) \bar{x} : 均值；
- b) C_V : 离(变)差系数；
- c) C_S : 偏态系数；
- d) σ : 标准差。

参数间换算关系见式(B.14)和式(B.15)：

$$\begin{cases} \alpha = \frac{4}{C_S^2} \\ \beta = \frac{2}{\bar{x} \cdot C_V \cdot C_S} \\ a_0 = \bar{x} \cdot (1 - \frac{2C_V}{C_S}) \end{cases} \dots\dots\dots (B.14)$$

$$\begin{cases} \bar{x} = \frac{\alpha}{\beta} + a_0 \\ C_S = \frac{2}{\sqrt{\alpha}} \\ C_V = \frac{\sqrt{\alpha}}{\alpha + a_0 \cdot \beta} \\ \sigma = \bar{x} \cdot C_V \end{cases} \dots\dots\dots (B.15)$$

已知累积概率 $P(x_P)$, 求极值 x_P 时, 可据分布函数进行变量转换并按伽玛函数积分求得, 也可查皮尔逊Ⅲ型分布的离均系数表求得。

B.2.2 皮尔逊Ⅲ型分布常用参数估计方法

B.2.2.1 矩法

把极值序列的最小值作为 a_0 的估计值, 则得式(B.16)：

$$\begin{cases} a_0 = \min(x_i) \\ C_V = \frac{\sigma}{\bar{x}} \\ C_S = \frac{2C_V}{1 - a_0/\bar{x}} \end{cases} \dots\dots\dots (B.16)$$

把附录 A 中 A.1 的 \bar{x} 和 σ 的估计值代入, 再据参数间关系式(B.14), 可得参数 α , β , a_0 的估计值。

B.2.2.2 线性矩法

与线性矩对应的前三阶样本矩为 l_1 、 l_2 、 l_3 , 见式(B.17)：

$$\begin{cases} l_1 = b_0 \\ l_2 = 2b_1 - b_0 \\ l_3 = 6b_2 - 6b_1 + b_0 \\ \tau_3 = l_3/l_2 \end{cases} \dots\dots\dots (B.17)$$

其中：

$$\begin{cases} b_0 = \frac{1}{n} \sum_{i=1}^n x_i \\ b_1 = \frac{1}{n} \sum_{i=2}^n \frac{i-1}{n-1} x_i \\ b_2 = \frac{1}{n} \sum_{i=3}^n \frac{(i-1)(i-2)}{(n-1)(n-2)} x_i \end{cases} \dots\dots\dots (\text{B. 18})$$

根据 τ_3 的不同取值,可计算得到中间变量 d 的数值。

如 $0 < |\tau_3| < 1/3$, 令 $Z = 3\pi\tau_3^2$, 则有

$$d = \frac{1 + 0.2906Z}{Z + 0.1882Z^2 + 0.0442Z^3} \dots\dots\dots (\text{B. 19})$$

如 $1/3 \leq |\tau_3| < 1$, 令 $Z = 1 - |\tau_3|$, 则有

$$d = \frac{0.36067Z - 0.59567Z^2 + 0.25361Z^3}{1 - 2.78861Z + 2.56096Z^2 - 0.77045Z^3} \dots\dots\dots (\text{B. 20})$$

进而据式(B. 21)及参数间关系式(B. 14),可得参数 α, β, a_0 的估计值。

$$\begin{cases} \hat{x} = l_1 \\ \hat{C}_V = \frac{l_2}{l_1} \pi^{\frac{1}{2}} d^{\frac{1}{2}} \frac{\Gamma(d)}{\Gamma(d + \frac{1}{2})} \\ \hat{C}_S = 2d^{\frac{1}{2}} \text{sign}(\tau_3) \end{cases} \dots\dots\dots (\text{B. 21})$$

B. 2. 2. 3 极大似然法

据皮尔逊Ⅲ型分布的概率密度函数,可得其极大似然方程组,见式(B. 22):

$$\begin{cases} \frac{\partial L}{\partial a_0} = -(\alpha - 1) \sum_{i=1}^n \frac{1}{x_i - a_0} + \beta = 0 \\ \frac{\partial L}{\partial \alpha} = -n \ln \beta - n \Psi(\alpha) + \sum_{i=1}^n \ln(x_i - a_0) = 0 \\ \frac{\partial L}{\partial \beta} = -n \ln(\alpha \beta) + \beta^2 \sum_{i=1}^n (x_i - a_0) = 0 \end{cases} \dots\dots\dots (\text{B. 22})$$

其中 $\Psi(\alpha) = \frac{\partial}{\partial \alpha} \ln \Gamma(\alpha)$ 为双 Γ 函数,经整理得式(B. 23):

$$\begin{cases} \frac{\alpha}{\beta} = \bar{x} - a_0 \\ \alpha = \frac{\frac{\bar{x} - a_0}{n} \sum_{i=1}^n \frac{1}{x_i - a_0}}{\frac{\bar{x} - a_0}{n} \sum_{i=1}^n \frac{1}{x_i - a_0} - 1} \\ \Psi(\alpha) = \ln \alpha - \ln(\bar{x} - a_0) + \frac{1}{n} \sum_{i=1}^n \ln(x_i - a_0) \end{cases} \dots\dots\dots (\text{B. 23})$$

据式(B. 23)中后两式通过迭代可解出 a_0 和 α ,再据第一式可得 β ,三个参数的估计值均得到。

B. 2. 2. 4 数值积分单权函数法

为了提高参数计算精度,提出了数值积分单权函数法,求解 C_S 时进行加权积分。

令权函数 $\varphi(x)$ 为标准正态密度函数,见式(B. 24):

QX/T 529—2019

$$\varphi(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \quad \dots\dots\dots (\text{B. 24})$$

经推导可得式(B. 25):

$$C_S = -4\sigma \frac{U(x)}{H(x)} \quad \dots\dots\dots (\text{B. 25})$$

其中:

$$\begin{cases} U(x) = \int_{a_0}^{\infty} f(x)(x-\bar{x}) \varphi(x) dx \\ H(x) = \int_{a_0}^{\infty} f(x)(x-\bar{x})^2 \varphi(x) dx \end{cases} \quad \dots\dots\dots (\text{B. 26})$$

利用数值积分加权求得:

$$\begin{cases} \bar{x} = \int_{-\infty}^{\infty} xf(x) dx = \int_0^1 x dp \approx \frac{1}{W} \sum_{i=1}^n w(i) x_i \\ \sigma = \sqrt{\int_{-\infty}^{\infty} (x-\bar{x})^2 f(x) dx} \approx \sqrt{\frac{1}{W} \sum_{i=1}^n w(i) (x_i - \bar{x})^2} \end{cases} \quad \dots\dots\dots (\text{B. 27})$$

$$\begin{cases} U \approx \frac{1}{W} \sum_{i=1}^n w(i) (x_i - \bar{x}) \varphi(x_i) \\ H \approx \frac{1}{W} \sum_{i=1}^n w(i) (x_i - \bar{x})^2 \varphi(x_i) \end{cases} \quad \dots\dots\dots (\text{B. 28})$$

式中:

n ——极值序列样本容量;

$w(i)$ ——序号为 i 处的数值积分权重;

W —— n 个权重的总和;

U ——函数 $U(x)$ 的近似计算值;

H ——函数 $H(x)$ 的近似计算值。

选取合适的 $w(i)$ 数列确定权重,即可据式(B. 27)和式(B. 28)求得 \bar{x} , σ , U , H ,由 U 和 H 据式(B. 25)可求得 C_S ,由 \bar{x} , σ , C_S 及参数间关系式(B. 14)和式(B. 15),可得参数 α , β , a_0 的估计值。

B. 2. 2. 5 数值积分双权函数法

数值积分单权函数法提高了 C_S 的估计精度,但在 C_V 的估计上仍可继续优化,在其基础上又设计出了数值积分双权函数法。

令第一权函数 $\varphi(x)$ 和第二权函数 $\psi(x)$ 为式(B. 29):

$$\begin{cases} \varphi(x) = \frac{k}{x \sqrt{2\pi}} e^{-\frac{k^2(x-\bar{x})^2}{2x^2}} \\ \psi(x) = e^{-\frac{h(x-\bar{x})}{x}} \end{cases} \quad \dots\dots\dots (\text{B. 29})$$

式中:

h , k ——不依赖于分布参数的待定正常数。

则得式(B. 30):

$$\begin{cases} U(x) = \int_{a_0}^{\infty} f(x)(x - \bar{x}) \varphi(x) dx \\ H(x) = \int_{a_0}^{\infty} f(x)(x - \bar{x})^2 \varphi(x) dx \\ A(x) = \int_{a_0}^{\infty} f(x) \psi(x) dx \\ D(x) = \int_{a_0}^{\infty} f(x)(x - \bar{x}) \psi(x) dx \end{cases} \dots\dots\dots (B. 30)$$

经推导可得式(B. 31):

$$\begin{cases} C_V^2 = \frac{\frac{1}{h\bar{x}} - \frac{1}{k^2} \cdot \frac{U(x)}{H(x)}}{-\frac{A(x)}{D(x)} + \frac{U(x)}{H(x)}} \\ C_S = -\frac{2\bar{x}(k^2 C_V^2 + 1)}{k^2 C_V} \cdot \frac{U(x)}{H(x)} \end{cases} \dots\dots\dots (B. 31)$$

利用数值积分加权求得式(B. 32):

$$\begin{cases} \bar{x} \approx \frac{1}{W} \sum_{i=1}^n w(i) x_i \\ U \approx \frac{1}{W} \sum_{i=1}^n w(i) (x_i - \bar{x}) \varphi(x_i) \\ H \approx \frac{1}{W} \sum_{i=1}^n w(i) (x_i - \bar{x})^2 \varphi(x_i) \\ A \approx \frac{1}{W} \sum_{i=1}^n w(i) \psi(x_i) \\ D \approx \frac{1}{W} \sum_{i=1}^n w(i) (x_i - \bar{x}) \psi(x_i) \end{cases} \dots\dots\dots (B. 32)$$

式中:

- n ——极值序列样本容量;
- $w(i)$ ——序号为 i 处的数值积分权重;
- W —— n 个权重的总和;
- U ——函数 $U(x)$ 的近似计算值;
- H ——函数 $H(x)$ 的近似计算值;
- A ——函数 $A(x)$ 的近似计算值;
- D ——函数 $D(x)$ 的近似计算值。

选取合适的 $w(i)$ 数列确定权重,即可据式(B. 32)求得 \bar{x}, U, H, A, D ;选用合适的系数 h 和 k ,即可求得 C_V 和 C_S ,由 \bar{x}, C_V, C_S 及参数间关系式(B. 14),可得参数 α, β, a_0 的估计值。

B. 2. 2. 6 适线法

适线法是指在一定的寻优准则下,求解与经验概率分布绘图点拟合最优的概率分布曲线对应的分布参数。常用的适线准则有离差平方和最小准则、离差绝对值和最小准则及相对离差平方和最小准则。

应用适线法估计参数时,参数初值及其范围的划分很重要,一般采用其他参数估计方法获得初值;也应注意其对经验概率分布绘点位置很敏感,即要求适当选取经验概率曲线绘点方法。

QX/T 529—2019

B.2.2.7 极大值调整适线法

在采用适线法进行皮尔逊Ⅲ型分布参数估计时,从安全保守性考虑,有时需要向特别突出的极大值倾斜,这就要求对适线法进行调整,在适线准则判断时,增加极大的几个值的权重,以达到拟合线向极大值点靠近的目的,称之为极大值调整适线法。

B.3 威布尔(Weibull)分布及其参数估计方法

B.3.1 威布尔分布

三参数威布尔分布的概率密度函数 $f(x)$ 和累积分布函数 $F(x)$ 分别见式(B.33)和式(B.34):

$$f(x) = \frac{c}{b} \left(\frac{x-a}{b} \right)^{c-1} e^{-\left(\frac{x-a}{b} \right)^c} \quad (x \geq a) \quad \dots\dots\dots (B.33)$$

$$F(x) = 1 - e^{-\left(\frac{x-a}{b} \right)^c} \quad \dots\dots\dots (B.34)$$

式中:

a ——位置参数;

b ——比例参数(尺度参数);

c ——形状参数。

当 $a = 0$ 时,即为两参数威布尔分布。

已知累积概率 $P(x_p)$,按式(B.35)求极端事件的极值:

$$x_p = a + b(\ln(P(x_p)))^{\frac{1}{c}} \quad \dots\dots\dots (B.35)$$

B.3.2 威布尔分布常用参数估计方法

B.3.2.1 矩法

用矩法估计威布尔分布的三个参数要用到前三阶矩,推导可得式(B.36):

$$r_1 = \frac{\Gamma(1 + \frac{3}{c}) - 3\Gamma(1 + \frac{2}{c})\Gamma(1 + \frac{1}{c}) + 2\Gamma^3(1 + \frac{1}{c})}{\left[\Gamma(1 + \frac{2}{c}) - \Gamma^2(1 + \frac{1}{c}) \right]^{\frac{3}{2}}} \quad \dots\dots\dots (B.36)$$

利用样本资料计算附录 A 中 A.1 的偏度系数 r_1 的估计值,参见式(A.3),按式(B.36)可解出 c 。

令

$$\begin{cases} d = \Gamma(1 + \frac{1}{c}) - \Gamma^2(1 + \frac{1}{c}) \\ t = \Gamma(1 + \frac{1}{c}) \end{cases} \quad \dots\dots\dots (B.37)$$

由式(B.38):

$$\begin{cases} b = \frac{\sigma}{\sqrt{d}} \\ a = \bar{x} - bt \end{cases} \quad \dots\dots\dots (B.38)$$

可得出 a 和 b 。这样 a, b, c 三个参数的估计值均已得到。

B.3.2.2 极大似然法

威布尔分布的极大似然方程组见式(B.39):

$$\begin{cases} \frac{\partial L}{\partial a} = -(c-1) \sum_{i=1}^n (x_i - a)^{-1} + \frac{c}{b} \sum_{i=1}^n \left(\frac{x_i - a}{b}\right)^{c-1} = 0 \\ \frac{\partial L}{\partial b} = \frac{c}{b} \left[\sum_{i=1}^n \left(\frac{x_i - a}{b}\right)^c - n \right] = 0 \\ \frac{\partial L}{\partial c} = \frac{n}{c} + \sum_{i=1}^n \ln\left(\frac{x_i - a}{b}\right) - \sum_{i=1}^n \left(\frac{x_i - a}{b}\right)^c \ln \frac{x_i - a}{b} = 0 \end{cases} \quad \dots\dots\dots (B. 39)$$

式(B. 39)是非线性方程组,可采用牛顿迭代法求解,得到 a , b , c 三个参数的估计值。但该方程组对初值的要求很高,实际计算有时难以做到。

B. 3. 2. 3 最小二乘法

由威布尔分布的分布函数得式(B. 38):

$$1 - F(x) = e^{-(\frac{x-a}{b})^c} \quad \dots\dots\dots (B. 40)$$

对两边取两次对数,则得式(B. 39):

$$\ln(-\ln(1 - F(x))) = c \ln(x - a) - c \ln b \quad \dots\dots\dots (B. 41)$$

令

$$\begin{cases} d = -c \ln b \\ t = \ln(x - a) \\ y_w = \ln(-\ln(1 - F(x))) \end{cases} \quad \dots\dots\dots (B. 42)$$

则得式(B. 43):

$$y_w = ct + d \quad \dots\dots\dots (B. 43)$$

式(B. 43)是一直线方程。用 $\{x_i\}$ 序列构建 $\{t_i\}$ 序列,用式(A. 4)中经验概率分布 $\{F_w(x_i)\}$ 序列构建 $\{y_{w,i}\}$ 序列,再对 $\{t_i\}$ 和 $\{y_{w,i}\}$ 序列用最小二乘法估计其中的 c 和 d ,进而用式(B. 42)求出 b 。

实际计算时,先选定 a 的初值,由上法可得一组 (a, b, c) 值,再调整 a 值,得到多组 (a, b, c) 值,将拟合误差最小的一组值作为威布尔分布 a, b, c 三个参数的估计值。

B. 4 广义极值(GEV)分布及其参数估计方法

B. 4. 1 广义极值分布

广义极值分布的概率密度函数 $f(x)$ 和累积分布函数 $F(x)$ 分别见式(B. 44)和式(B. 45):

$$\begin{cases} f(x) = \frac{1}{b} \left(1 + c \frac{x-a}{b}\right)^{-1-\frac{1}{c}} e^{-(1+c\frac{x-a}{b})^{-\frac{1}{c}}} & (c \neq 0, 1 + c \frac{x-a}{b} > 0) \\ f(x) = \frac{1}{b} e^{-e^{-\frac{x-a}{b}} - \frac{x-a}{b}} & (c = 0) \end{cases} \quad \dots\dots\dots (B. 44)$$

$$F(x) = e^{-(1+c\frac{x-a}{b})^{-\frac{1}{c}}} \quad (c \neq 0, 1 + c \frac{x-a}{b} > 0) \quad \dots\dots\dots (B. 45)$$

式中:

a ——位置参数;

b ——比例参数(尺度参数);

c ——形状参数。

当 $c > 0$ 时,对应Ⅱ型极值分布;当 $c < 0$ 时,对应Ⅲ型极值分布即威布尔分布;当 $c = 0$ 时,对应Ⅰ型极值分布即耿贝尔分布。

已知累积概率 $P(x_p)$,按式(B. 46)求极端事件的极值:

QX/T 529—2019

$$x_P = a + \frac{b}{c}((-\ln P(x_P))^{-c} - 1) \quad \dots\dots\dots (B.46)$$

B.4.2 广义极值分布常用参数估计方法

B.4.2.1 线性矩法

与线性矩对应的前三阶样本矩为 l_1, l_2, l_3 , 见式(B.47):

$$\begin{cases} l_1 = b_0 \\ l_2 = 2b_1 - b_0 \\ l_3 = 6b_2 - 6b_1 + b_0 \\ \tau_3 = l_3/l_2 \end{cases} \quad \dots\dots\dots (B.47)$$

其中:

$$\begin{cases} b_0 = \frac{1}{n} \sum_{i=1}^n x_i \\ b_1 = \frac{1}{n} \sum_{i=2}^n \frac{i-1}{n-1} x_i \\ b_2 = \frac{1}{n} \sum_{i=3}^n \frac{(i-1)(i-2)}{(n-1)(n-2)} x_i \end{cases} \quad \dots\dots\dots (B.48)$$

由于无显式表达式, 只能用近似方法求解 c 。当 $-0.5 \leq \tau_3 \leq 0.5$ 时, 式(B.48)可将计算误差控制在 9×10^{-4} 以下:

$$c = 7.8590d + 2.9554d^2, \quad \text{其中 } d = \frac{2}{3 + \tau_3} - \frac{\ln 2}{\ln 3} \quad \dots\dots\dots (B.49)$$

进而可据式(B.50):

$$\begin{cases} b = \frac{l_2 c}{(1 - 2^{-c})\Gamma(1 + c)} \\ a = l_1 - \frac{b}{c}(1 - \Gamma(1 + c)) \end{cases} \quad \dots\dots\dots (B.50)$$

得到广义极值分布 a, b, c 三个参数的估计值。

B.4.2.2 极大似然法

据广义极值分布的概率密度函数, 可得其极大似然函数, 见式(B.51):

$$L = -\ln(b) - (1 + \frac{1}{c}) \sum_{i=1}^n \ln(1 + c \frac{x_i - a}{b}) - \sum_{i=1}^n (1 + c \frac{x_i - a}{b})^{-\frac{1}{c}} \quad \dots\dots\dots (B.51)$$

令 $\frac{\partial L}{\partial a} = 0, \frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial c} = 0$ 可得极大似然方程组, 其为非线性方程组, 用牛顿迭代法可得 a, b, c 三个参数的估计值。

B.5 泊松-耿贝尔(Poisson-Gumbel)复合分布及其参数估计方法

B.5.1 泊松-耿贝尔复合分布

在某随机事件出现频次 m 的概率分布为 P_j 的情况下, 其影响下某要素 x_m 的分布函数为 $G(x)$ 。 P_j 为离散型, 符合泊松分布, 见式(B.52):

$$P_j = \frac{\lambda^j}{j!} e^{-\lambda} \quad (j = 0, 1, 2, \dots) \quad \dots\dots\dots (B.52)$$

式中:

λ ——事件发生总次数与总年数的比值。

该事件影响下的某要素符合耿贝尔分布,见式(B.53):

$$G(x) = e^{-e^{-a(x-b)}} \quad \dots\dots\dots (B.53)$$

可得泊松-耿贝尔复合分布的累积分布函数 $F(x)$, 见式(B.54):

$$F(x) = \sum_0^j P_j(G(x))^j = e^{-\lambda(1-G(x))} = e^{-\lambda(1-e^{-e^{-a(x-b)}})} \quad \dots\dots\dots (B.54)$$

已知累积概率 $P(x_p)$, 则得式(B.55):

$$G(x_p) = 1 + \frac{1}{\lambda} \ln P(x_p) \quad \dots\dots\dots (B.55)$$

再取两次对数可得求要素极值的公式(B.56):

$$x_p = b + \frac{-\ln(-\ln(1 + \frac{1}{\lambda} \ln P(x_p)))}{a} \quad \dots\dots\dots (B.56)$$

泊松-耿贝尔复合分布参数估计可参考耿贝尔分布的参数估计方法。

B.5.2 耿贝尔法参数估计

令

$$y_{B,i} = -\ln(-\ln(1 + \frac{1}{\lambda} \ln(F_w(x_i)))) \quad (i = 1, 2, \dots, n) \quad \dots\dots\dots (B.57)$$

式中:

$F_w(x_i)$ ——经验概率分布函数,见式(A.4)。

可得式(B.58):

$$\begin{cases} a = \frac{\sigma(y_B)}{\sigma(x)} \\ b = E(x) - \frac{E(y_B)\sigma(x)}{\sigma(y_B)} \end{cases} \quad \dots\dots\dots (B.58)$$

式中:

$\sigma(x)$ ——序列 $\{x_i\}$ 的标准差;

$\sigma(y_B)$ ——序列 $\{y_{B,i}\}$ 的标准差;

$E(x)$ ——序列 $\{x_i\}$ 的数学期望;

$E(y_B)$ ——序列 $\{y_{B,i}\}$ 的数学期望。

利用附录 A 中 A.1 的方法可分别得到它们的估计值,参见式(A.1)和式(A.2),代入式(B.58)中,得到参数 a 和 b 的估计值。

QX/T 529—2019

附 录 C

(资料性附录)

拟合优度检验方法

C.1 柯尔莫哥洛夫-斯米诺夫 (Kolmogorov-Smirnov) 检验法

柯尔莫哥洛夫统计量见式(C.1):

$$D_n = \max(|F(x_i) - F_W(x_i)|) \quad \dots\dots\dots (C.1)$$

式中:

D_n ——柯尔莫哥洛夫统计量,表示在所有各点上,假设理论分布与经验概率分布之差的最大值;

$F(x_i)$ ——概率分布函数在 x_i 处的概率值;

$F_W(x_i)$ ——有序序列在 x_i 处的经验概率值,参见附录 A 中 A.2。

对于不同的显著水平 α 和 n ,可查表获得柯尔莫哥洛夫检验中的临界值 $\lambda_{\alpha,n}$ 。若 $D_n \sqrt{n} < \lambda_{\alpha,n}$,则接受原假设,认为样本符合假设理论分布;否则,拒绝原假设。

C.2 剩余方差

剩余方差见式(C.2):

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad \dots\dots\dots (C.2)$$

式中:

S^2 ——剩余方差;

n ——序列样本总数,见附录 A 中 A.1;

x_i ——序列中第 i 个样本的值,见附录 A 中 A.1;

\hat{x}_i ——拟合估计值。

剩余方差越小拟合越优。

C.3 拟合相对偏差

拟合相对偏差见式(C.3):

$$R = 100 \times \frac{1}{n} \sum_{i=1}^n \frac{|x_i - \hat{x}_i|}{x_i} \quad \dots\dots\dots (C.3)$$

式中:

R ——拟合相对偏差。

拟合相对偏差越小拟合越优。

参 考 文 献

- [1] GB/T 34293—2017 极端低温和降温监测指标
 - [2] GB/T 35227—2017 地面气象观测规范 风向和风速
 - [3] HAD101/10—1991 核电厂厂址选择的极端气象事件
 - [4] 陈家鼎,郑忠国. 概率与统计[M]. 北京:北京大学出版社,2007
 - [5] 马开玉,张耀存,陈星. 现代应用统计学[M]. 北京:气象出版社,2004
 - [6] 黄嘉佑,李庆祥. 气象数据统计分析方法[M]. 北京:气象出版社,2015
-

中 华 人 民 共 和 国
气 象 行 业 标 准
气候可行性论证规范 极值概率统计分析
QX/T 529—2019

*

气象出版社出版发行
北京市海淀区中关村南大街 46 号
邮政编码:100081
网址: <http://www.qxcbs.com>
发行部:010-68408042
北京中科印刷有限公司印刷

*

开本:880 mm×1230 mm 1/16 印张:1.5 字数:45 千字
2020 年 1 月第 1 版 2020 年 1 月第 1 次印刷

*

书号:135029-6109 定价:22.00 元

如有印装差错 由本社发行部调换
版权专有 侵权必究
举报电话:(010)68406301