Journal of Water Resources Research 水资源研究, 2013, 2, 389-394
http://dx.doi.org/10.12677/jwrr.2013.26055 Published Online December 2013 (http://www.hanspub.org/journal/jwrr.html)

An Application of Cross Entropy Method to the Parameter Estimation in Flood Frequency Analysis*

Linsen Niu, Songbai Song#

College of Water Resources and Architectural Engineering, Northwest A&F University, Xianyang
Email: mail:nls1990@163.com, #ssb6533@nwsuaf.edu.cn

Received: Sep. 30th, 2013; revised: Nov. 20th, 2013; accepted: Nov. 26th, 2013

Abstract: This paper studies on the application of fractile constrained cross-entropy to the estimation of distribution parameters in flood frequency analysis. Based on the principle of minimum cross-entropy, two annual maximum flood peak series respectively in Feather River in Canada and Zhangcunyi Station in northern Shaanxi province with Gumbel distribution and Gamma distribution were employed to the parameter estimation of the four distribution functions. Four frequency curves with the estimated parameters were also plotted. Then, comparing the calculated cross-entropy values with those that are derived by traditional methods-MOM and MLM, it turned out that: by using cross entropy method, we got the minimum cross entropy values. The plotted theoretical frequency curves fit well with the empirical frequency curves. So, we can conclude that the quantile constrained cross-entropy method has the considerable merit in the flood frequency parameter estimation and is superior to the traditional methods-MOM and MLM.

Keywords: Cross Entropy; Fractile Constraints; Parameter Estimation; Flood Frequency Analysis

交互熵法在洪水频率分布参数估计中的应用研究*

牛林森，宋松柏#

西北农林科技大学，水利与建筑工程学院，咸阳
Email: mail:nls1990@163.com, #ssb6533@nwsuaf.edu.cn

收稿日期：2013 年 9 月 30 日；修回日期：2013 年 11 月 20 日；录用日期：2013 年 11 月 26 日

摘　要：本文研究基于分位数对约束下的交互熵进行洪水频率分布参数估计方法。以加拿大 Feather 河和陕北地区张村驿站年最大洪峰流量序列为例，选取 Gumbel 和 Gamma 分布，基于最小交互熵原理，进行年最大洪峰流量序列分布参数估计，并根据估计参数推求洪峰流量频率曲线图。与矩法和极大似然法所求熵值比较，结果表明：交互熵法获得最小熵值，频率点距拟合亦取得满意效果。因此，在考虑分位数对约束的条件下，交互熵法能有效的估计分布参数，且较矩法和极大似然法优越。

关键词：交互熵；分位数对约束；参数估计；洪水频率分析

1. 引言

目前，在洪水频率分析中，虽已出现了许多非参数方法[1]，但是，在大多数研究和实际工作中，洪水频率分析仍然作为一个参数推断问题来处理[2]。传统的参数估计方法有矩法，极大似然法，概率权重矩法等，它们的优劣性已被众多研究者讨论过[3]。为了改进矩法估计，马秀峰于 1984 年提出了权函数法，在样本矩的计算中引入正态概率密度函数作为权函数，增

*基金项目：国家自然科学基金(51179160, 50879070, 50579065)；高等学校博士学科点专项科研基金(20110204110017)。
作者简介：牛林森(1990-)，女，河南驻马店人，在读硕士生，主要从事流域水文模拟及水文预报研究。宋松柏(1965-)，男，陕西咸阳人，教授，博士，主要从事水文水资源教学与研究工作。
#通讯作者。

加了靠近均值部分的权重，减小了两端部分的权重，但是权函数属于单参数估计，不能全面地解决洪水频率参数估计问题[4]。为了避免计算对数 P-Ⅲ分布中的 C_s，Rao 提出了混合矩法，在所提出的四种混合矩法中，选择对数数据的均值、方差及原始数据均值作为混合矩时，统计性能最优[2]。Hosking 在 Greenwood (1979)定义概率权重矩之后，于 1990 年提出了线性矩，该法是概率权重矩的线性组合，其最大特点是对洪水系列中的极大值和极小值远没有常规矩那么敏感，所以估计结果受样本中个别点据误差的影响较小，但该法以概率作为权重来求矩，在各阶矩中，变数均为一次幂，其阶次主要由其相应的概率(作为权重)来反映，在推求参数时会引起误差和灵敏性问题[5]。另外，在实际拟合中，常常出现中间部分拟合较好，但在两端尤其是高尾部分拟合难以取得满意效果。针对这种情况，Wang (1997)建议使用高阶概率权重矩，为序列中的较大流量赋予更高的权重，该法求得频率曲线中上部拟合效果良好[6,7]。Cohn 于 1997 年提出了 EMA (Expected Moments Algorithm)算法，该法的主要思想是在具有历史洪水的情形下，以矩法估计结果为初值，充分考虑小于门限值 T (一般认为很大的洪峰流量)的洪水资料，通过高效迭代提高洪水分位点估计精度和收敛速度[8]。洪水本身就是一种非常复杂的水文现象，准确描述其规律需要大量且较长的观测数据，在目前的资料技术水平基础上，寻求更为有效的洪水频率参数估计方法，提高洪水频率分析成果的合理性和可靠性是一项非常重要的课题。

Kullback (1959)提出了交互熵的概念和最小交互熵原理。与贝叶斯方法类似，采用先验分布和后验分布，它们之间的概率距离(Probabilistic Distance)或直接偏差(Directed divergence)称为交互熵。最小交互熵原理即先验分布已知，且约束条件一定，候选分布中能使交互熵函数最小化值的分布即为所求分布[9-12]。Lind 和 Solana 于 1988 年提出用分位数对作为约束条件的交互熵法，该法将观测数据直接进行数据编码，从而产生分位数对约束，并在该约束条件下使交互熵最小化。这种方法计算简单，后验分布函数满足单调性和不变性的要求[13-15]。目前我国缺乏该法的应用研究。本文以加拿大 Feather 河和陕北地区张村驿站年洪峰流量序列为例，研究最小交互熵原理进行洪水分布参数估计的普适性。

2. 分位数对约束下的交互熵法

该法包括两步：第一步，选定参考分布，并从样本数据中推求出分位数对约束。第二步，选定一个后验分布，使分位数约束下的交互熵最小化。

2.1. 分位数对约束

将随机变量 X 的观测序列 x_i 按升序值进行排列，得到 $S=\{x_i\}$，$i=1,2,\cdots,r$，$x_i\in R$。假设 x 是一个可能未来发生值。那么该值落在 X 由 x_i 分成的 $r+1$ 个子区间 $[x_0,x_1),[x_1,x_2),\cdots,[x_r,x_{r+1})$ 内是等可能的，根据样本规则，对应的分位数概率为 $i/r+1$。假设 $Q(x|x_1,x_2,\cdots,x_r)$ 是由数据集 S 推断 X 的分布函数，$q(x|x_1,x_2,\cdots,x_r)$ 为相应密度函数。则分位数对约束为

$$\left(x;Q\left(x|x_1,x_2,\cdots,x_r\right)\right)_i=\left(x_i,\frac{i}{r+1}\right),\ i=1,2,\cdots,r \quad (1)$$

2.2. Kullback 最小交互熵原理

给定一个待选分布函数 $P(x)$ 和密度函数 $p(x)$，在分位数 $x=x_i$ 上，$P(x)$ 值记为 p_i。通过寻求后验分布函数 $Q(x|x_1,x_2,\cdots,x_r)$，使 Kullback-Leibler 交互熵函数

$$D(q,p)=\int q\left(x|x_1,x_2,\cdots,x_r\right)\log\frac{q\left(x|x_1,x_2,\cdots,x_r\right)}{p(x)}\mathrm{d}x \quad (2)$$

满足分位数对约束(1)的条件下最小化。式(1)可写为期望值的形式，即

$$g_i:\int_I f_i(x)q\left(x|x_1,x_2,\cdots,x_r\right)\mathrm{d}x-\frac{1}{r+1}=0,\quad i=1,2,\cdots,r \quad (3)$$

式中，$f_i(x)$ 为指示函数，$f_i(x)=\begin{cases}1; & x\in I_i\\ 0; & \text{其它}\end{cases}$。

根据极值原理，引入拉格朗日乘子 λ_i，有

$$\begin{aligned}L&=D(q,p)+\sum_{i=1}^{r}\lambda_i g_i\\&=\int q\left(x|x_1,x_2,\cdots,x_r\right)\left[\log q\left(x|x_1,x_2,\cdots,x_r\right)-\log p(x)\right]\\&\quad+\sum_{i=1}^{r}\lambda_i\left[\int_I f_i(x)q\left(x|x_1,x_2,\cdots,x_r\right)\mathrm{d}x-\frac{1}{r+1}\right]\end{aligned} \quad (4)$$

为了获得 $p(x)$，使交互熵最小，根据变分法中的 Euler-Lagrange 方程，若 $I=\int_a^b G\left(x,q(x),q'(x)\right)\mathrm{d}x$，其中 G 已知，则 $p(x)$ 满足

$$\frac{\partial G}{\partial q(x)}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial G}{\partial q'(x)}\right)=0 \tag{5}$$

在式(4)中没有 $q'(x)$ 项，仅为 $q(x)$ 的函数。式(4)仅对 $q(x)$ 求偏导，并令其为 0，有

$$\frac{\partial L}{\partial q(x)}=\log\frac{q\left(x|x_1,x_2,\cdots,x_r\right)}{p(x)}+1+\sum_{i=1}^{r}\lambda_i f_i(x)=0 \tag{6}$$

进一步整理，有

$$\begin{aligned}q\left(x|x_1,x_2,\cdots,x_r\right)&=p(x)\exp\left(-1-\sum_{i=1}^{r}\lambda_i f_i(x)\right)\\&=\mu(x)p(x)\end{aligned} \tag{7}$$

式中，$\mu(x)=\exp\left(-1-\sum_{i=1}^{r}\lambda_i f_i(x)\right)$。当 $x\in I_i$，$\mu(x)=\mu(x_i)=\mu_i=\exp(-1-\lambda_i)$。

把式(7)代入式(3)的左边，考虑 $x\in I_i$，$f_i(x)=1$ 有

$$\begin{aligned}&\int_I f_i(x)q\left(x|x_1,x_2,\cdots,x_r\right)\mathrm{d}x-\frac{1}{r+1}\\&=\int_I f_i(x)\mu(x)p(x)\mathrm{d}x-\frac{1}{r+1}=\int_I \mu_i p(x)\mathrm{d}x-\frac{1}{r+1}\end{aligned}$$

有

$$\mu_i=\frac{1}{(r+1)\left(P_{i+1}-P_i\right)} \tag{8}$$

因此，当 $x\in I_i$，有

$$q\left(x|x_1,x_2,\cdots,x_r\right)=\mu_i p(x) \tag{9}$$

并可通过代换，有

$$Q\left(x|x_1,x_2,\cdots,x_r\right)=\frac{i}{r+1}+\mu_i\left[P(x)-P_i\right] \tag{10}$$

把式(8)和式(9)代入式(2)中，有

$$D(q,p)=\sum_{i=0}^{r}\mu_i p(x)\log\frac{\mu_i p(x)}{p(x)},$$

化简整理，有

$$D(q,p)=-\log(r+1)-\frac{1}{r+1}\sum_{i=0}^{r}\log\left(P_{i+1}-P_i\right)。$$

令 $c=\log(r+1)$，c 为非负数。令 $S(P)=\sum_{i=0}^{r}\left(P_{i+1}-P_i\right)$。则

$$D_{\min}(q,p)=-c+S(P)/(r+1) \tag{11}$$

由式(11)可知，最小交互熵可以写成参考分布的函数，并且 $S(P)$ 值越小，式(11)的值越小，如果参考分布已经选定，那么根据最小交互熵原理，说明该分布的参数值最优。

3. 参考分布的选择

对于一个给定的随机样本来说，分位数对约束可以通过数据编码求得，利用最小交互熵法将后验分布 $Q\left(x|x_1,x_2,\cdots,x_r\right)$ 看作是先验分布 $P(x)$ 的一个代数转化。因此，参考分布必须尽可能的靠近后验分布和样本点。本文选择 Gumbel 分布和 Gamma 分布为参考分布。

4. 应用实例

以加拿大的 Feather River 57 年洪峰流量数据和中国陕北的张村驿站 37 年洪峰流量数据为例，选择 Gumbel 和 Gamma 分布为参考分布，进行参考分布参数的估计，并与矩法和极大似然法相对比。

4.1. 参考分布

Gumbel 密度与分布函数分别为

$$p(x)=\alpha\exp\left(-\alpha(x-u)-\mathrm{e}^{-\alpha(x-u)}\right);\quad -\infty<x<\infty \tag{12}$$

$$P(x)=\exp\left(-\mathrm{e}^{-\alpha(x-u)}\right) \tag{13}$$

Gamma 密度与分布函数分别为

$$p(x)=\frac{\lambda(\lambda x)^{\kappa-1}\mathrm{e}^{-\lambda x}}{\Gamma(\kappa)} \tag{14}$$

$$P(x)=1-\frac{\Gamma(\kappa,\lambda x)}{\Gamma(\kappa)} \tag{15}$$

4.2. 参数估计

首先，分别根据矩法和极大似然法求出估计分布的参数，将矩法所求的参数作为初始值，然后利用优化函数确定交互熵函数最小值及参考分布的分布参数。

经计算，参数计算结果见表 1~2。

4.3. 结果分析

表 1 显示，Feather 河选用 Gumbel 分布作为参考

分布进行参数估计时，矩法、极大似然法、交互熵法分别求得的 $S(P)$ 值和 $D_{\min}(q,p)$ 值呈减小的趋势，而 Gamma 分布所求 $S(P)$ 值和 $D_{\min}(q,p)$ 值同 Gumbel 分布规律一致，且所求值比 Gumbel 分布小；而参数估计值的计算结果分别为：Gumbel 分布中，极大似然法计算所得 u 值最大，交互熵法其次，矩法最小，α 值的计算结果为交互熵法所得值最大，矩法其次，极大似然法最小；Gamma 分布中，三个方法计算所得 u 值相差不大，α 值的计算结果为矩法最大，极大似然法其次，交互熵法最小。表 2 显示，张村驿站所求 $S(P)$ 值和 $D_{\min}(q,p)$ 值结果与 Feather 河存在相同的规律，且 Gumbel 分布中，极大似然法计算所得 λ 值最大，交互熵法其次，矩法最小，κ 值的计算结果为极大似然法所得值最大，交互熵法其次，矩法最小。

由以上分析结果可知，选用上述两种参考分布时，交互熵法所求得的 $S(P)$ 值和 $D_{\min}(q,p)$ 值均是最小，所以运用该法所求参数绘制频率曲线图，结果见图 1 和图 2。

图 1 显示，对于 Feather 河，整体来说，两种分布下，经验点据与分位数估计值拟合结果均比较理想，但在高尾部拟合中，Gamma 分布的拟合效果优于 Gumbel 分布，即分位数估计值更接近实测值。图 2 显示，张村驿站在两种分布下拟合结果亦较好，但 Gamma 分布的拟合效果整体优于 Gumbel 分布。

5. 结论与展望

本文以加拿大 Feather 河和陕北张村驿站为例，探讨了分位数对约束条件下的交互熵法在洪水频率分析参数估计中的应用，并与传统的矩法和极大似然法相对比。结论与展望如下：

1) 在选择 Gumbel 和 Gamma 两种参考分布时，交互熵法求得的 $S(P)$ 值和 $D_{\min}(q,p)$ 值均小于由其它两种方法所求值，且采用 Gamma 分布求得的 $S(P)$ 值和 $D_{\min}(q,p)$ 值较 Gumbel 分布所求值小，说明在洪水频率参数估计时，交互熵法优于传统的矩法和极大似然法。

2) Feather 河和张村驿站的理论频率曲线与经验点据的拟合效果均良好，Gamma 分布的两条拟合曲线均较相应站点 Gumbel 分布下的拟合效果好，没有负值段，说明本文的两个实例选用 Gamma 分布线型更合理。

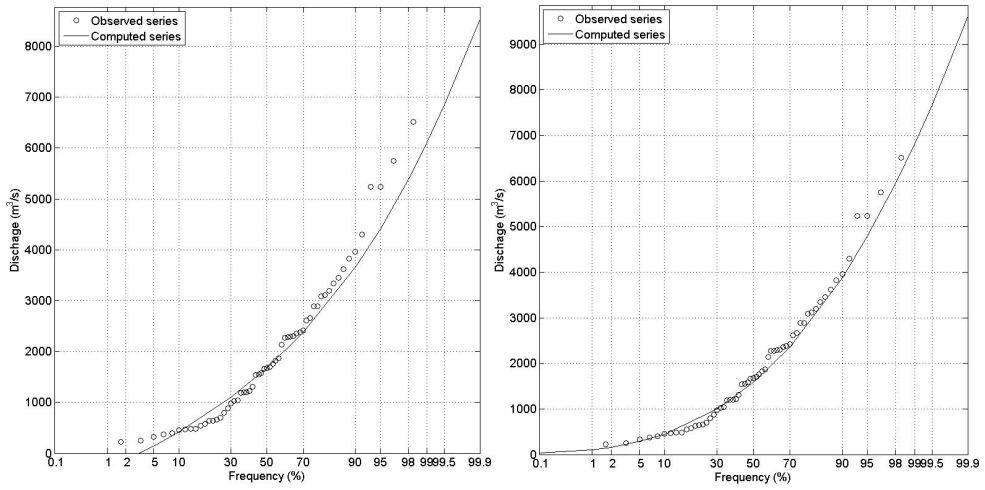
3) 本文主要研究了分位数对为约束条件下，利用

Table 1. The parameters of Feather River for the annual maximum flood peak discharge
表 1. Feather 河洪峰流量分布参数计算结果

选用分布	估算方法	$S(P)$	$H(D_{\min}(q,p))$	u 或 λ	α 或 κ
Gumbel	矩法	352.0545	1.7732	2652.3285	0.0009
	极大似然法	323.4119	1.2959	2785.0856	0.0006
	交互熵法	292.9187	0.7876	1293.0598	0.0010
Gamma	矩法	289.9769	0.7386	0.0009	1.8204
	极大似然法	289.9578	0.7383	0.0009	1.7933
	交互熵法	289.9000	0.7373	0.0009	1.7434

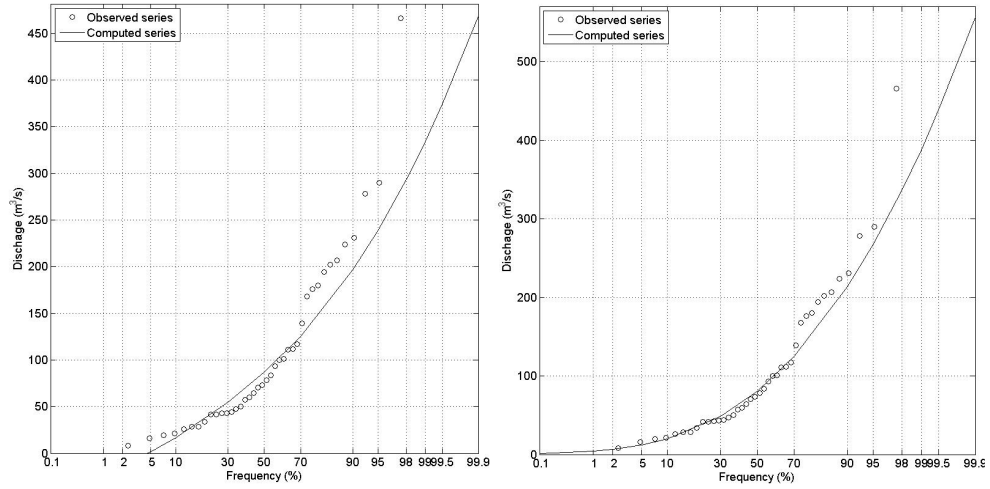
Table 2. The parameters of Zhangcunyi Station for the annual maximum flood peak discharge
表 2. 张村驿站年洪峰流量分布参数计算结果

选用分布	估算方法	$S(P)$	$H(D_{\min}(q,p))$	u 或 λ	α 或 κ
Gumbel	矩法	202.8352	1.2336	152.3158	0.0134
	极大似然法	188.8355	0.8922	162.6435	0.0078
	交互熵法	165.9407	0.3338	65.4722	0.0171
Gamma	矩法	163.6595	0.2781	0.0119	1.3020
	极大似然法	163.4675	0.2734	0.0139	1.5143
	交互熵法	163.3476	0.2705	0.0146	1.4960



(a) Gumbel 分布　　(b) Gamma 分布

Figure 1. The fitting results of Feather River for the annual maximum flood peak discharge
图 1. Feather 河年最大洪峰流量频率曲线拟合结果



(a) Gumbel 分布　　(b) Gamma 分布

Figure 2. The fitting results of Zhangcunyi Station for the annual maximum flood peak discharge
图 2. 张村驿站年最大洪峰流量频率曲线拟合结果

最小交互熵原理进行洪水频率分析参数估计的方法。对于我国常用的 P-III 型分布，目前缺少该法推求分布参数的研究，把概率权重矩作为约束时，采用交互熵法推求分布参数的方法有待进一步研究。

6. 致谢

本文系国家自然科学基金(51179160, 50879070, 50579065)和高等学校博士学科点专项科研基金(20110204110017)资助。在研究过程中，作者十分感谢 Niels Lind 教授的指导及提供文献资料，同时感谢第十一届中国水论坛推荐。

参考文献 (References)

[1] 董洁, 谢悦波, 翟金波. 非参数统计在洪水频率分析中的应用与展望[J]. 河海大学学报(自然科学版), 2004, 32(1): 23-26.
DONG Jie, XIE Yuebo and ZHAI Jinbo. Application of non-parametric statistic approach to flood frequency analysis and prospect of its development trend. Journal of Hohai University (Natural Sciences), 2004, 32(1): 23-26. (in Chinese)

[2] 丛树铮, 胡四一. 洪水频率分析的现状与展望[J]. 水文, 1987, 6: 52-58.
CONG Shuzheng, HU Siyi. Present situation and prospect of flood frequency analysis. Hydrology, 1987, 6: 52-58. (in Chinese)

[3] RAO, A.R., HAMED, K.H. Flood frequency analysis. New York: CRC Press LCC, 2000: 127-186.

[4] 马秀峰. 计算水文频率参数的权函数法[J]. 水文, 1984, 3: 1-11.
MA Xiufeng. The weighted function method applied in hydro-

logic frequency parameters calculation. Hydrology, 1984, 3: 1-11. (in Chinese)

[5] 陈元芳, 沙志贵, 陈剑池等. 具有历史洪水时 P-III 分布线性矩法的研究[J]. 河海大学学报, 2001, 29(4): 76-80.
CHEN Yuanfang, SHA Zhigui, CHEN Jianchi, et al. Study on L-moments estimation method for P-III distribution with historical flood. Journal of Hohai University, 2001, 29(4): 76-80. (in Chinese)

[6] 李扬, 宋松柏. 高阶概率权重矩在洪水频率分析中的应用[J]. 水力发电学报, 2013, 32(2): 14-21.
LI Yang, SONG Songbai. Application of higher-order probability-weighted moments to flood frequency analysis. Journal of Hydroelectric Engineering, 2013, 32(2): 14-21. (in Chinese)

[7] WANG, Q.J. Using higher probability weighted moments for flood frequency analysis. Journal of Hydrology, 1997, 194(1): 95-106.

[8] COHN, T.A., LANE, W.L. and BAIER, W.G. An algorithm for computing moments-based flood quantile estimates when historical flood information is available. Water Resources Research, 1997, 33(9): 2089-2096.

[9] DENG, J., PANDEY, M.D. and GU, D. Extreme quantile estimation from censored sample using partial cross-entropy and fractional partial probability weighted moments. Structural Safety, 2009, 31(1): 43-54.

[10] PANDY, M.D. Extreme quantile estimation using order statistics with minimum cross-entropy principle. Probabilistic Engineering Mechanics 2001, 16(1): 31-42.

[11] 茆诗松. 贝叶斯统计[M]. 北京: 中国统计出版社, 1999: 1-6.
MAO Shisong. Bayesian statistics. Beijing: China Statistics Press, 1999: 1-6. (in Chinese)

[12] SHORE, J.E., JOHNSON, R.W. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE Transformation on Information Theory, 1980, 26(1): 26-37.

[13] 黄振平. 水文统计学[M]. 南京: 河海大学出版社, 2008: 169-171.
HUANG Zhenping. Hydrological statistics. Nanjing: Hohai University Press, 2008: 169-171. (in Chinese)

[14] LIND, N.C. and HONG, H.P. A cross entropy method in flood frequency analysis. Stochastic Hydrology and Hydraulics, 1989, 3(3): 191-192.

[15] LIND, N.C. and SOLANA, V. Fractile constrained entropy estimation of distributions based on scarce data. Civil Engineering Systems, 1990, 7(2): 87-93.