II—56

母集団推定の限界性把握

信州大学工学部　正会員　荒木正夫
信州大学工学部　正会員　寒川典昭
信州大学工学部　　　　○仙石雄一

1．はじめに

確率水文量あるいはリターン・ピリオドの精度は，確率分布，パラメタ推定法，用いたデータ，の３者によって決まると著者達は考えている。[1)]　シミュレーションデータで検討するとき，仮定した母集団と同じタイプの分布形を用いれば，第１の問題は取り除かれる。第２の問題については，種々の推定法を用いて検討すべきであるが，ここでは，とりあえず第１段階の研究であるので最尤法に限る。以上より本稿では，データ数による母集団推定の限界を把握する。

一方，推定精度は，Kullback-Leibler情報量と指定した値に対するリターン・ピリオドで測る。前者は全体的な適合度を，後者は右裾の適合度と変動をみている。

2．検討方法

母集団分布が既知のとき，推定分布の適合度をみるには，次の Kullback-Leibler(K-L)情報量を用いるのがよい。

$$I(g;f)=\int \ln\{\frac{g(x)}{f(x)}\}g(x)dx \tag{1}$$

ここに，g(x)は真の密度関数，f(x)はモデルの密度関数である。一般に，$I(g;f)\geqq 0$ であり，等号は $g(x)=f(x)$ のとき成立する。

いま，$g(x):N(0,1)$，$f(x):N(\xi,\tau^2)$とすると，式(1) は

$$I(g;f)=-\frac{1}{2}\ln(2\pi\sigma^2)-\frac{1}{2} \tag{2}$$

となり，$g(x):E(1)$，$f(x):E(\beta)$ とすると

$$I(g;f)=\beta-\ln\beta-1 \tag{3}$$

となる。ただし，E(k)は1/k を期待値にもつ指数分布である。

右裾の適合度と変動は，水工計画に利用される範囲のリターン・ピリオド(T) でみることにする。

具体的には，次の手順で計算する。

1)　N(0,1),E(1) を母集団として，シミュレーションによりデータ数n(n=5,10,20,･･･,100,150,200) の5000組のデータセットを作る。

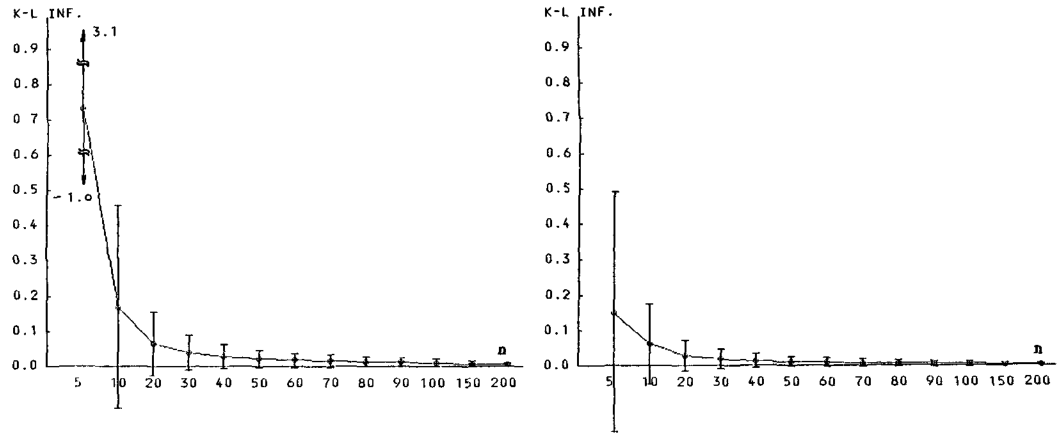
2)　各組ごとに，式(2) あるいは式(3) で K-L情報量を，さらに指定した複数個の値に対するリターン・ピリオドを計算する。

3)　2)で求めた K-L情報量とリターン・ピリオドのnごとの平均と分散を計算し，nの増加とともにこれらの値がどのように挙動するかを確かめる。

3．結果と考察

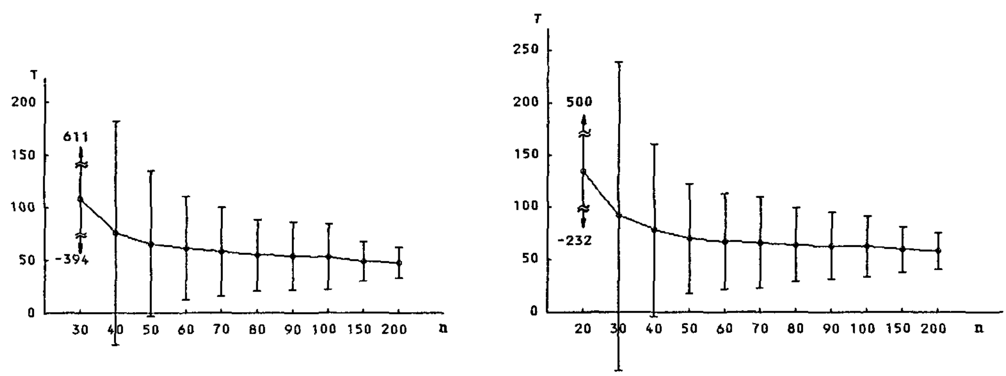
Fig.1 は3.の方法で計算した K-L情報量である。どちらの母集団の場合も，データ数が20～30個程度からK-L 情報量の平均値が零に近づき，標準偏差も小さくなってきている。全体的な適合度を議論するとき，少なくともこの程度のデータ数が必要であろう。Fig.2(a)は２を超過する正規分布のリターン・ピリオドを(b) は４を超過する指数分布のリターン・ピリオドをみたものである。平均値は30～40個程度から落ち着きみせているが，標準偏差はかなり大きい。データ数が増えると，標準偏差も徐々に減少するが 100個以上になってもまだこの程度の変動を残していることは予想以上である。また，掲載は省略しているが，対象とす

るリターン・ピリオドが大きい程，平均値の落ち着くデータ数は大きくなり，標準偏差の値もさらに大きくなる。



(a) 正規分布 (母集団:N(0,1))　　(b) 指数分布 (母集団:E(1))

Fig.1 K-L 情報量



(a) 正規分布 (母集団:N(0,1))　　(b) 指数分布 (母集団:E(1))

Fig.2 リターン・ピリオド

4．あとがき

ここで仮定した母集団では，水文統計に用いられる範囲のリターン・ピリオドの変動は，実在するデータ数程度の場合予想以上に大きかった。当然，分布形を変えれば，これらの挙動も変化するであろう。今後，(分布形，パラメタ推定法，データ数) をセットとして，種々の組み合わせごとに母集団推定の限界を把握していきたい。

1)寒川・荒木・渡辺：確率分布の推定母数の不確定性評価法，土木学会論文集，第 375号/ II-6,pp.133～141,1986年11月.