II-448　限られた数のデータから得られた非超過確率水質値の推定精度に関する理論的検討

岡山大学工学部　正員　○河原長美
エイトコンサルタント　正員　細田信夫

1．はじめに

筆者らは、従来より、公共水域の水質環境基準の適合判定に用いられている非超過確率水質の推定精度に関して，旭川における観測値に基ずいて検討を加えてきているが，今回は，順序統計の理論を採用して，非超過確率水質の推定値が従う確率密度関数を誘導し，この結果に基ずいて検討を加える。

2．解析に用いたデータ

本研究で用いる水質データは、旭川兵団地点において，定時観測により得られたデータであり，SS（ガラスファイバー沪紙法），濁度（吸光光度法），COD（Mn）（酸性法），全窒素（還元法），全リン（混合試薬法）の5項目である。なお，データ数は，昭和54年～昭和56年の13～27カ月分であり，比較のために，採水地点より約2km上流における日流量についても，水質データと同様の検討を加えた。

3．非超過確率水質の推定値の確率分布関数

確率密度関数 $f(x)$ に従う母集団から得られた n 個の標本を考える。今，n 個の標本のうち，小さい側から k 番目の値を u，$k+1$ 番目の値を v とすると，トーマスプロットによれば，u，v，それぞれに対応する非超過確率は，$k/(n+1)$ および $(k+1)/(n+1)$ となる。

非超過確率 P% は，$k/(n+1) \leq P/100 \leq (k+1)/(n+1)$ を満足するものとし，P% に相当する水質値 w を，u，v の値とそれぞれに対応する非超過確率値から内挿して求めるものとし，$w = h(u, v, n, P)$ で表わす。なお，上述の不等式において，等号が成立する場合は，式の形が異なってくるので，分けて検討を加えることとし，ここでは，等号が成立しない場合のみを検討する。

非超過確率 P% の値が w になる確率 $g(w)\,dw$ は，内挿関数 h を定めると，小さい側から k 番目の値が u となり，$k+1$ 番目が v となる同時確率に等しい。ところで，互いに素な5領域 $I_1 \sim I_5$ に分割し，それぞれの領域に属する確率を $P_1 \sim P_5$ で表わすと次のようになる。

$I_1 = (-\infty, u)$	$P_1 = F(u)$	$k-1$ 個が属する
$I_2 = [u, u+du)$	$P_2 = F(u+du) - F(u) \cong f(u)\,du$	1 個が属する
$I_3 = [u+du, v)$	$P_3 = F(v) - F(u+du) \cong F(v) - F(u)$	0 個が属する
$I_4 = [v, v+dv)$	$P_4 = F(v+dv) - F(v) \cong f(v)\,dv$	1 個が属する
$I_5 = [v+dv, \infty)$	$P_5 = 1 - F(v+dv) \cong 1 - F(v)$	$n-k-1$ 個が属する

ここに，$F(x)$ は分布関数であり，$F(x) = \int_{-\infty}^{x} f(x)\,dx$ を満足する。

それ故，求める確率 $g(w)\,dw$ は次のように表わされる。

$$g(w)\,dw = \frac{n!}{(k-1)!\,(n-k-1)!}\,F(u)^{k-1}\,(1-F(v))^{n-k-1}\,f(u)\,f(v)\,du\,dv \tag{1}$$

ところで，n，P が与えられた場合に，ある一組の u，v の値，u_1，v_1 によって定まる w の値を w_1 とすると，$w_1 \leq w$ を満足する確率は，w の分布関数 $G(w)$ $\left(= \int_{-\infty}^{w} g(w)\,dw\right)$ に等しい。今，$-\infty < u_1 \leq w$ なる任意の u_1 が与えられたとすると，$w_1 \leq w$ となるためには，v_1 は $u_1 \leq v_1 \leq v_*(u_1, w, n, P)$ を満足する必要がある。ここに，$v_*(u_1, w, n, P)$ は，$w = h(u_1, v_1, n, P)$ を v_1 に関して解いた関数である。それ故，$G(w)$ は次のように表わされる。

$$G(w)=\frac{n!}{(k-1)!\,(n-k-1)!}\int_{-\infty}^{w}F(u_1)^{k-1}f(u_1)\left(\int_{u_1}^{v_*}(1-F(v_1))^{n-k-1}f(v_1)\,dv_1\right)du_1$$

$$=\frac{n!}{(k-1)!\,(n-k)!}\int_{-\infty}^{w}F(u_1)^{k-1}f(u_1)\left\{(1-F(u_1))^{n-k}-(1-F(v_*))^{n-k}\right\}du_1 \qquad (2)$$

4．流量および各水質指標の確率密度関数

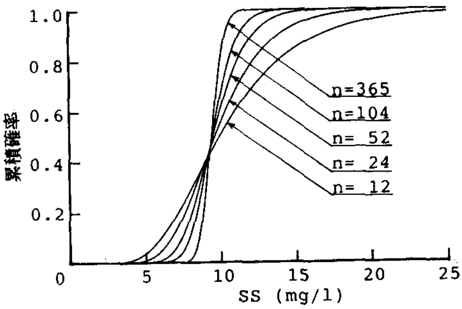
各項目について，対数正規分布，ピアソンⅢ型分布（1母数，2母数，3母数），ならびに対数ピアソンⅢ型分布に対する適合性を検討した結果，最も適合性が良好な場合を表－1に示す。なお，K－S検定の結果，5%水準で帰無仮説が否定されたのは，濁度，TN，CODであり，帰無仮説が否定されなかった場合についても，最も適合性が良好な場合を示している。なお，TNの場合は，対数正規分布もK－S検定を満足するが，表－1に示すピアソンⅢ型2母数分布のほうが適合性が良好であった。

表－1　確率密度関数

Q	$f(x)=\frac{1}{0.8126x(2\pi)^{1/2}}\exp\{-\frac{(\ln x-3.711)^2}{2\times0.6603}\}$
SS	$f(x)=\frac{1}{1.039x(2\pi)^{1/2}}\exp\{-\frac{(\ln x-1.570)^2}{2\times1.078}\}$
TUR	$f(x)=\frac{1}{0.9807x(2\pi)^{1/2}}\exp\{-\frac{(\ln x-2.307)^2}{2\times0.9617}\}$
COD	$f(x)=\frac{1}{0.6261x(2\pi)^{1/2}}\exp\{-\frac{(\ln x-0.5749)^2}{2\times0.3920}\}$
TP	$f(x)=\frac{1}{0.7460x(2\pi)^{1/2}}\exp\{-\frac{(\ln x+3.586)^2}{2\times0.5565}\}$
TN	$f(x)=\frac{1}{0.2270x\Gamma(2.936)}\left(\frac{x}{0.2272}\right)^{2.936}\exp-\left(\frac{x}{0.2272}\right)$

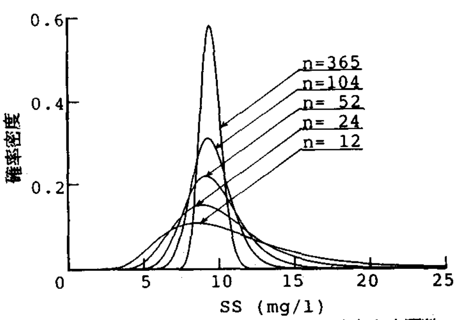
5，非超過確率水質の推定値の分布

図－1，2には，SSを例にとって，非超過確率75%に関する分布関数および確率密度関数を示す。なお，この場合，線形な内挿関数 $w=au+bv$（$a=1-P(n+1)+[P(n+1)]$，$b=P(n+1)-[P(n+1)]$ ここに，[ ]はガウスの記号）を用いている。



図－1　非超過確率75%のSS濃度の分布関数

図－1，2より，データ個数が多くなると推定値の分布が集中すること，ならびに，データ個数が多くなると正規分布に似た形になることがうかがえよう。



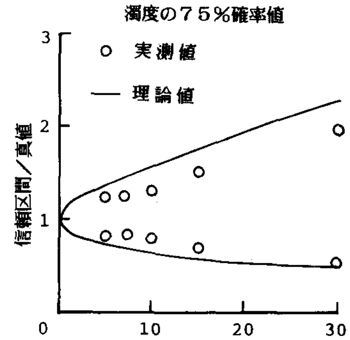
図－2　非超過確率75%のSS濃度の確率密度関数

次に，濁度を例として，推定値の95%信頼区間に関して，(2)式に基ずいて算出した値と実測値から直接算出した値の間で比較すると，図－3に示す通りである。なお，実測値から算出する場合には，一年間のデータから与えられたサンプリング間隔ごとにデータを抽出し，抽出された各データ群を用いて非超過確率水質を推定し，この値に基ずいて分布を推定した。両者の傾向はほぼ一致しているが，信頼区間の上限値については幾分差が大きい。内挿関数の形も影響していると考えられ，更に検討を加える予定である。

ところで，対数正規分布に従う場合，信頼区間／真値の比の値に影響をおよぼすのは，対数値の分散である。水質も流量も対数正規分布に従い，$C=aQ^b$ の関係が成立すると仮定すると，$\ln C=\ln a+b\ln Q$ であるから，$\ln Q$ が $N(\alpha,\beta^2)$ に従うなら，$\ln C$ は $N(\ln a+b\alpha,b^2\beta^2)$ に従うことになり，bの値の大小によって，各河川における水質指標間の推定巾の大小関係を判断できることになる。

紙面の関係で説明不足の点が多々認められるが，詳細に関しては講演時に発表する予定である。

濁度の75%確率値



図－3　推定精度の理論値と実測値との比較