II-289　ごみ分析におけるサンプリング法に関する一考察

京都大学　工学部　学生員　○藤田勝良
正員　平岡正勝
正員　武田信生

1　はじめに　現在、都市ごみのサンプリング法は、四分法により行なわれている。これは、図1に示すような操作を繰り返して、その量が、適当なもの（現在では、約20kg）となると、これを試料として採取する段階的な縮分・サンプリング法である。ところで、近年のように都市ごみを袋に入れて収集する方式で、このサンプリング法を用いるには、このような操作の前に袋を破らなければならず、現在では、本サンプリング法は、非常に時間や労力を要するサンプリング法となっている。そこで、袋をいくつか試料として採取する方法（袋サンプリング）が、考えられる。ここでは、種々の仮定を設けることにより、四分法と袋サンプリングについて、袋破り、混合，切断，段階的といった操作をモデル化し、近似的なものであるが、推定値の精度の定式化（ここでは、推定値の分散）を行なった。つぎに実験より得られたデータをこの式に適用することにより、両者を比較し、四分法の妥当性や、新しいサンプリング法について、検討を行なった。

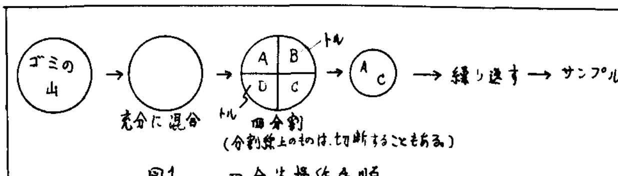

図1　四分法操作手順

2　各種サンプリング法　四分法に関連すると思われる各種サンプリング法について、簡単に述べる。

a　(1段)単純ランダムサンプリング　N個の単位からなる母集団から、n個の試料を、乱数表やカードを用いてランダムに選ぶサンプリング法である。推定値$\bar{y}$とその分散$V_R(\bar{y})$は、それぞれ次のようになる。

$\bar{y}=\frac{1}{n}\sum_{i}^{n}y_i$，$V_R(\bar{y})=\frac{1-f}{n}S^2$　（y_i：i番目の単位の特性値，$S^2=\frac{1}{N-1}\sum_{i}^{N}(y_i-\bar{Y})^2$：母分散，　$f=\frac{n}{N}$：抽出率）

b　2段サンプリング　母集団がN個の1次抽出単位からなり、各1次抽出単位には、それぞれM個の2次抽出単位が含まれているとする。このときN個の1次抽出単位からn個を選び、さらにその各々からm個ずつの2次抽出単位をサンプルとして抽出するようなサンプリング法である。推定値とその分散は、次のようになる。

$\bar{y}=\frac{1}{nm}\sum_{i}^{n}\sum_{j}^{m}y_{ij}$

$V(\bar{y})=\frac{1-f_1}{n}S_1^2+\frac{1-f_2}{nm}S_2^2$

（y_{ij}：i番目の1次抽出単位のj番目の2次抽出単位の特性値
$S_1^2=\frac{1}{N-1}\sum_{i}(\bar{Y}_i-\bar{\bar{Y}})^2$，$S_2^2=\frac{1}{N(M-1)}\sum_{i}\sum_{j}(y_{ij}-\bar{Y}_i)^2$，$\bar{Y}_i=\frac{1}{M}\sum_{j}y_{ij}$，$f_1=\frac{n}{N}$，$f_2=\frac{m}{M}$）

c　(1段)クラスターサンプリング　bでM＝mすなわち、1次サンプルに含まれる2次抽出単位をすべてサンプルとするようなサンプリング法である。推定値およびその分散は、bで$m=M$とすればよい。

$\bar{y}=\frac{1}{nM}\sum_{i}^{n}\sum_{j}^{M}y_{ij}$，$V(\bar{y})=\frac{1-f_1}{n}S_1^2$

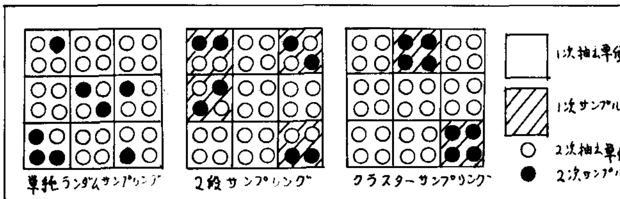

図2　各種サンプリング法について

3　モデル化および定式化　四分法と袋サンプリングとの違いは、①袋破り、②混合，③切断，④多段の4点である。ここでは、これらの行為をモデル化し、推定値の精度の定式化を行なう。

a　袋破り，混合　四分法では、サンプルをつねにクラスターとして採取する。したがって、袋中に、いくつかの独立挙動単位（要素）が含まれているものと考えれば、袋サンプリングおよび袋を破っても、全く混合しない場合は、それぞれ一種のクラスターサンプリング、混合を完全に行なった場合は、要素を抽出単位とする単純ランダムサンプリングであると考えられる。

b．切断　切断とは、隣接要素間で、特性値が平均化される行為とみればよい。特性値が平均化されるので

母分散S^2は、減少する。この減少率は、平均化される範囲や要素の隣接のしかたによって変化する。この両者について平均化した減少率を、$(1-\beta)$とすれば、母分散は、平均的にβS^2となったと考えればよい。

c 多段　aで述べたことより、混合を全く行なわない四分法は、多段クラスターサンプリング、完全に混合を行なっていれば、ランダム性は、この時点で確立されるので、以後何段で行なっても、1段の単純ランダムサンプリングであると考えればよい。切断については、各段で前述のβなるものを考えればよい。

以上のようなモデルを用い、また、つぎに示す2つの仮定を設けることにより、四分法および袋サンプリングについて、サンプルを母集団の1/16（四分法の操作で4回）採取する場合について、推定値の精度（分散）の定式化を行なう。

仮定 ① 各クラスターは、同数の要素より成る。

② 各要素の総重量は、一定である。

この2つの仮定により、特性値y_iとしては、各要素の任意の組成の重量を用いても、重量百分率を用いても、推定値は不偏となる。よって推定値の精度、すなわち、サンプリング法の良否は、推定値の分散の大小の比較のみで行ない得る。表1に、四分法および袋サンプリングについて、そのモデルおよび推定値の分散の式を示す。ただし、記号は、四分法については表2に、袋サンプリングについては表3に示す通りであり、要素の大きさは、スコップ1杯の量とする。ここで、各βの関係を用いれば、V_1〜V_5の大小は、比較できる。しかし、V_4のサンプリング法は、最も時間、労力を要するサンプリング法であり、V_5は、最も容易なものである。したがって、各方法の比較には、このようなことも考慮しなければならないと思われる。

表1　各種サンプリング法のモデルとその精度の式

サンプリング法	モデル	推定値の分散の式
四分法 4段，無混合 無切断	4段クラスターサンプリング 特性値は変化しない。	$V_1 = \frac{1-f_1}{n} S_1^2 + \frac{1-f_2}{nm} S_2^2 + \frac{1-f_3}{nml} S_3^2 + \frac{1-f_4}{nmlp} S_4^2$
四分法 4段，無混合 切断	4段クラスターサンプリング 特性値は平均化される。	$V_2 = \frac{1-f_1}{n} \beta_1 S_1^2 + \frac{1-f_2}{nm} \beta_2 S_2^2 + \frac{1-f_3}{nml} \beta_3 S_3^2 + \frac{1-f_4}{nmlp} \beta_4 S_4^2$
四分法 4段，完全混合 無切断	1段単純ランダムサンプリング 特性値は変化しない。	$V_3 = \frac{1-f_1 f_2 f_3 f_4}{nmlpQ} S^2$
四分法 4段，完全混合 切断	1段単純ランダムサンプリング 特性値は平均化される。	$V_4 = \frac{1-f_1 f_2 f_3 f_4}{nmlpQ} \beta_5 S^2$
袋サンプリング	1段クラスターサンプリング 特性値は変化しない。	$V_5 = \frac{1-f_1'}{n'} S_1'^2$

表2　四分法についての記号

	総単位数	サンプル数	抽出率	平均値	母分散，クラスター間分散
1次抽出単位	N	n	$f_1 = \frac{n}{N}$	$\bar{\bar{\bar{Y}}} = \frac{1}{N}\sum_k^N \bar{\bar{Y}}_k$	$S_1^2 = \frac{1}{N-1}\sum_k^N (\bar{\bar{Y}}_k - \bar{\bar{\bar{Y}}})^2$
2次抽出単位	M	m	$f_2 = \frac{m}{M}$	$\bar{\bar{Y}}_k = \frac{1}{M}\sum_i^M \bar{\bar{Y}}_{ki}$	$S_2^2 = \frac{1}{N(M-1)}\sum_k^N\sum_i^M (\bar{\bar{Y}}_{ki} - \bar{\bar{Y}}_k)^2$
3次抽出単位	L	l	$f_3 = \frac{l}{L}$	$\bar{\bar{Y}}_{ki} = \frac{1}{L}\sum_j^L \bar{\bar{Y}}_{kij}$	$S_3^2 = \frac{1}{NM(L-1)}\sum_k^N\sum_i^M\sum_j^L (\bar{\bar{Y}}_{kij} - \bar{\bar{Y}}_{ki})^2$
4次抽出単位	P	p	$f_4 = \frac{p}{P}$	$\bar{\bar{Y}}_{kij} = \frac{1}{P}\sum_h^P \bar{Y}_{kijh}$	$S_4^2 = \frac{1}{NML(P-1)}\sum_k^N\sum_i^M\sum_j^L\sum_h^P (\bar{Y}_{kijh} - \bar{\bar{Y}}_{kij})^2$
5次抽出単位（要素）	Q	Q	$f_5 = 1$	$\bar{Y}_{kijh} = \frac{1}{Q}\sum_t^Q y_{kijht}$	$S^2 = \frac{1}{NMLPQ-1}\sum_k^N\sum_i^M\sum_j^L\sum_h^P\sum_t^Q (y_{kijht} - \bar{\bar{\bar{Y}}})^2$

表3　袋サンプリングについての記号

	総単位数	サンプル数	抽出率	平均値	母分散，クラスター間分散
1次抽出単位	N'	n'	$f_1' = \frac{n'}{N'}$	$\bar{\bar{Y}} = \frac{1}{N'}\sum_u^{N'} \bar{Y}_u$	$S_1'^2 = \frac{1}{N'-1}\sum_u^{N'} (\bar{Y}_u - \bar{\bar{Y}})^2$
2次抽出単位（要素）	Q'	Q'	$f_2' = 1$	$\bar{Y}_u = \frac{1}{Q'}\sum_u^{N'}\sum_v^{Q'} y_{uv}$	$S'^2 = \frac{1}{N'Q'-1}\sum_u^{N'}\sum_v^{Q'} (y_{uv} - \bar{\bar{Y}})^2$

4 実験　前述のモデルを用いて、約50kgの都市ごみから、約3kgをサンプルとして採取する場合について、ごみ中の種々の組成重量を要素ごとに求める実験を行ない、推定値の分散を求めた。表4に、都市ごみのいくつかの代表組成について、V_1を1とした場合の値を示した。同表より、m, n等を一定とした場合、分散が最小となるのは、V_4の方法であることがわかる。また、V_2とV_3を比較すると切断よりも混合の効果の方が大きいことがわかる。ただ、今回の実験では、ごみ袋1袋がスコップ1杯の量となっていたので、袋1つ1つが要素となっており、袋破りの効果は、現われていない。

表4　推定値の分散の比較

組成	V_1/V_1	V_2/V_1	V_3/V_1	V_4/V_1	V_5/V_1
軟質プラスチック	1	0.46	0.80	0.32	0.80
硬質プラスチック	1	1	0.79	0.79	0.79
紙（軟）	1	1	0.66	0.66	0.66
ボール紙	1	1	0.54	0.54	0.54
ちゅう芥	1	1	0.69	0.69	0.69
せんいくず	1	0.80	0.69	0.38	0.69
平均	1	0.88	0.70	0.56	0.70

5 今後の方向　以上、四分法と袋サンプリングについて、それぞれの行為が、完全に行なえたとして、各方式の精度の比較を行なった。しかし、実際の四分法は、これらの中間的なものであるから、その精度の式は、

$$V = \psi\varphi_1 V_1 + \psi(1-\varphi_1)V_2 + (1-\psi)\varphi_2 V_3 + (1-\psi)(1-\varphi_2)V_4$$

（φ, ψは、各行為の達成率を示す0〜1の係数）

となる。今後は、これらの係数を求め、より現実的な比較を行ない、また、仮定の妥当性や計算法の検討もあわせて行なって、モデルをもっと精密にし、時間の労力＝一定とした場合の、より厳密な解析が必要となろう。

本研究は、坪田正人との共同研究である。