周期のある母集団からのサンプリング（I）

久米　均*

1. まえがき

母集団に周期的な傾向，たとえば図1のような sine curve にしたがって動く傾向がある場合に，系統サンプリング（等間隔のサンプリング）を用いると，母集団の周期Tとサンプリング間隔のm間に

$$m=NT \qquad N：正整数$$

の関係がある場合には，サンプリング精度は非常に悪くなる。図1のAは$N=1$の場合を示したものであり，すべての観測値は同じ値になり，観測数をいくら増しても，その精度は1回の観測をランダムに行なった場合と同じである。逆に

$$m=(N-\tfrac{1}{2})T \qquad N：正整数$$

の関係がある場合にはサンプリング誤差は非常に小さく，観測値が偶数のときには，サンプリング誤差は0になる。図1のBは$N=1$の場合を示したもので，相隣る観測値の平均は，平均からのズレが相殺して母平均に一致する。サンプリング間隔を任意に定めた場合の誤差は上の二つの場合の間に存在する。以下において母集団周期T，サンプリング間隔m，観測数nとサンプリング誤差$V(\bar{x}_{sy})$との関係を吟味し，母集団に周期的な傾向があるときに系統サンプリングを実施することの妥当性について検討を行なう。

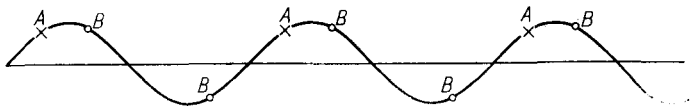

図　1

2. 母集団の周期と自己相関関数

母集団が次のような式にしたがう場合について考えてみよう。

$$x(t)=A_1\ sin(\lambda_1 t+\alpha_1)+A_2\ sin(\lambda_2 t+\alpha_2)+\cdots\cdots$$
$$+A_k sin(\lambda_k t+\alpha_k) \qquad \cdots\cdots(1)$$

$x(t)$は一般に概周期関数になり，各λ_iについて

$$T\ \lambda_i=2\pi\ N_i \qquad \cdots\cdots(2)$$

$$i=1,\ 2,\ \cdots\cdots k\cdot N_i：整数$$

を満足する実数Tが存在するとき$x(t)$は周期関数になる。周期は(2)式を満足するTの中の最小

*東大工学部　1963年11月14日受理「経営科学」第7巻3号

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

のものである。

$x(t)$の平均，分散および自己共分散を次のように定める。まず平均については

$$\mu=\lim_{h\to\infty}\frac{1}{2h}\int_{-h}^{h}x(t)\,dt$$

$$=\lim_{h\to\infty}\frac{1}{2h}\int_{-h}^{h}\sum A_i\sin(\lambda_i t+\alpha_i)\,dt=0 \qquad \cdots\cdots(3)$$

分散については

$$\sigma^2=\lim_{h\to\infty}\frac{1}{2h}\int_{-h}^{h}x^2(t)\,dt$$

$$=\lim_{h\to\infty}\frac{1}{2h}\int_{-h}^{h}[\sum A_i\sin(\lambda_i t+\alpha_i)]^2\,dt$$

$$=\frac{1}{2}\sum A_i^2 \qquad \cdots\cdots(4)$$

自己共分散については

$$Cov\,[x(t),\ x(t+\tau)]=\lim_{h\to\infty}\frac{1}{2h}\int_{-h}^{h}[x(t)\cdot x(t+\tau)]\,dt$$

$$=\lim_{h\to\infty}\frac{1}{2h}\int_{-h}^{h}[\sum A_i\sin(\lambda_i t+\alpha_i)]\,[\sum A_i\sin(\lambda_i t+\lambda_i\tau+\alpha_i)]\,dt$$

$$=\frac{1}{2}\sum A_i^2\cos\lambda_i\tau \qquad \cdots\cdots(5)$$

これより自己相関関数 $\rho(\tau)$は次式で与えられる。

$$\rho(\tau)=\frac{\sum\{A_i^2\cos\lambda_i\tau\}}{\sum A_i^2} \qquad \cdots\cdots(6)$$

二，三の簡単な例について自己相関関数を求めてみよう。

〔例1〕

$$x(t)=\begin{cases}-1 & -\pi\leqq t<0\\ 1 & 0\leqq t<\pi\end{cases}$$

$x(t)$を Fourier 展開すれば

$$x(t)=\frac{4}{\pi}\sum\frac{\sin(2n-1)t}{2n-1}$$

$$\rho(t)=\frac{\sum\frac{\cos(2n-1)t}{(2n-1)^2}}{\sum\frac{1}{(2n-1)^2}}=\frac{\frac{\pi}{4}\left(\frac{\pi}{2}-|t|\right)}{\frac{\pi^2}{8}}$$

$$=\left(1-\frac{2}{\pi}|t|\right)$$

〔例2〕

$$x(t)=\begin{cases}-t-\frac{\pi}{2} & -\pi\leqq t\leqq 0\\ t-\frac{\pi}{2} & 0\leqq t\leqq\pi\end{cases}$$

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

$x(t)$を Fourier 展開すれば

$$x(t)=-\frac{4}{\pi}\sum\frac{\cos(2n-1)t}{(2n-1)^2}$$

$$\rho(t)=\frac{\sum\frac{\cos(2n-1)t}{(2n-1)^4}}{\sum\frac{1}{(2n-1)^4}}=\frac{\frac{\pi}{96}(\pi^3-6\pi t^2+4t^3)}{\frac{\pi^4}{96}}$$

$$=1-6\left(\frac{t}{\pi}\right)^2+4\left(\frac{t}{\pi}\right)^3$$

〔例3〕

$$x(t)=\pi-x \qquad 0<x<2\pi$$

$x(t)$を Fourier 展開すれば

$$x(t)=2\sum\frac{\sin nt}{n}$$

$$\rho(t)=\frac{\sum\frac{4}{n^2}\cos nt}{\sum\frac{4}{n^2}}=\frac{\frac{1}{4}(t-\pi)^2-\frac{\pi^2}{12}}{\frac{\pi^2}{6}}=\frac{3}{2}\left(\frac{t}{\pi}-1\right)^2-\frac{1}{2}$$

以上を図に示すと，図2，3，4が得られる。

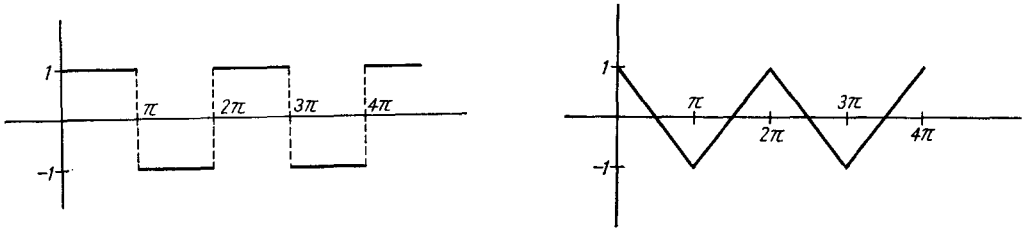

図 2

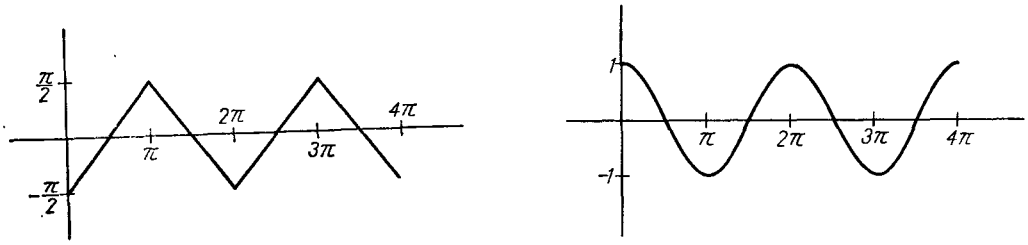

図 3

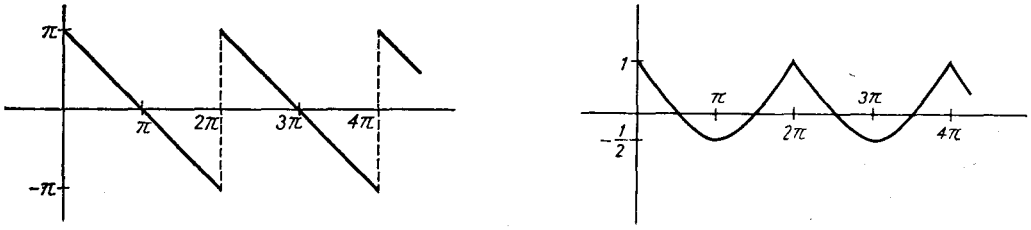

図 4

© 日本オペレーションズ・リサーチ学会. 無断複写・複製・転載を禁ず.

3. 系統サンプリングの誤差

母平均 μ, 母分散 σ^2, 自己相関関数 $\rho(t)$ なる母集団から, ランダムスタートにより間隔 m で n 回の観測を行なった場合のサンプリング誤差 $V(\bar{x}_{sy})$ は, 次のようにして求めることができる。

観測値を $x_1, x_2, \cdots\cdots, x_n$ とすれば

$$\bar{x}_{sy}=\frac{1}{n}\sum x_i$$

$$\sum(x_i-\mu)^2=\sum(x_i-\bar{x}_{sy})^2+n(\bar{x}_{sy}-\mu)^2$$

$$\sum_{i=1}^{n}(x_i-\bar{x}_{sy})^2=\frac{1}{n}\sum_{i=1}^{n}\sum_{j>i}(x_j-x_i)^2=\frac{1}{n}\sum_{i=1}^{n}\sum_{j>i}\{(x_j-\mu)-(x_i-\mu)\}^2$$

$$=\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}(x_j-\mu)^2-\frac{2}{n}\sum_{i=1}^{n}\sum_{j>i}(x_j-\mu)(x_i-\mu)$$

$$E\sum(x_i-\mu)^2=E\sum(x_i-\bar{x}_{sy})^2+nE(\bar{x}_{sy}-\mu)^2$$

$$E\sum(x_i-\mu)^2=\sum E(x_i-\mu)^2=n\sigma^2$$

$$E(\bar{x}_{sy}-\mu)^2=V(\bar{x}_{sy})$$

$$E\sum(x_i-\bar{x}_{sy})^2=\frac{1}{n}\sum\sum E(x_i-\mu)^2-\frac{2}{n}\sum\sum E(x_j-\mu)(x_i-\mu)$$

$$=\frac{1}{n}n(n-1)\sigma^2-\frac{2}{n}\sum_{u=1}^{n-1}(n-u)\rho(mu)\sigma^2$$

$$\therefore\ V(\bar{x}_{sy})=\frac{1}{n}E\sum(x_i-\mu)^2-\frac{1}{n}E\sum(x_i-\bar{x}_{sy})^2$$

$$=\sigma^2-\frac{n-1}{n}\sigma^2+\frac{2}{n^2}\sum_{u=1}^{n-1}(n-u)\rho(mu)\sigma^2$$

$$=\frac{\sigma^2}{n}\left\{1+\frac{2}{n}\sum_{u=1}^{n-1}(n-u)\ \rho(mu)\right\}\qquad\cdots\cdots(7)$$

したがって母集団が式(1)で表わされる場合に, 間隔 m で n 回観測を行ない, 母平均を推定したとすれば

$$E(\bar{x}_{sy})=0$$

$$V(\bar{x}_{sy})=\frac{\sigma^2}{n}\left\{1+\frac{2}{n}\sum_{u}(n-u)\ \rho(mu)\right\}$$

$$=\frac{1}{2n}\left\{\sum_i A_i^2\right\}\left[1+\frac{2}{n}\sum_u(n-u)\frac{\sum\{A_i^2\cos\lambda_i mu\}}{\sum A_i^2}\right]$$

$$=\frac{1}{2n}\sum_i A_i^2+\frac{1}{n^2}\sum_i A_i^2\left[\sum_u(n-u)\cos\lambda_i mu\right]\qquad\cdots\cdots(8)$$

ここで

© 日本オペレーションズ・リサーチ学会. 無断複写・複製・転載を禁ず.

$$S_i=\sum_u (n-u)\cos\ \lambda_i\ mu \qquad \cdots\cdots(9)$$

とおくと，三角関数を含む有限級数の公式から，

$$S_i=\frac{1}{2}\ \frac{\sin^2(\lambda_i\ nm/2)}{\sin^2\ (\lambda_i m/2)}-\frac{n}{2} \qquad \cdots\cdots(10)$$

$\sin(\lambda_i m/2)=0$，すなわち $\lambda_i m=2N_i\pi$（N_iは正整数）のときは

$$S_i=\frac{1}{2}(n^2-n)$$

となるが，

$$\lim_{\lambda_i m\to 2N_i\pi}\left[\frac{1}{2}\ \frac{\sin^2\ (\lambda_i nm/2)}{\sin^2(\lambda_i m/2)}-\frac{n}{2}\right]=\frac{1}{2}(n^2-n)$$

となり，式(10)の極限の値と一致するので，分母が0の場合は極限の値をとることに定めておけば S_i の値は一般に式(10)で表わすことができる。

$$\therefore\ V(\bar{x}_{sy})=\frac{1}{2n}\sum_i A_i^2+\frac{1}{n^2}\sum_i A_i^2 S_i$$

$$=\frac{1}{2}\sum_i\left\{\frac{A_i\ \sin(\lambda_i nm/2)}{n\ \sin\ (\lambda_i m/2)}\right\}^2 \qquad \cdots\cdots(11)$$

と書くことができる。

4. $\rho(t)=\cos\ \lambda t$ の場合

式(1)において $k=1$ の場合，すなわち

$$x(t)=A\ \sin(\lambda t+\alpha) \qquad \cdots\cdots(12)$$

のときは

$$\mu=0,\ \sigma^2=A^2/2,\quad \rho(t)=\cos\ \lambda t$$

となる。サンプリング間隔m，サンプル数nでランダムスタートで系統サンプリングを行なうものとすれば，サンプリング誤差は式(11)から

$$V(\bar{x}_{sy})=\frac{A^2}{2}\left\{\frac{\sin(\lambda nm/2)}{n\ \sin(\lambda m/2)}\right\}^2 \qquad \cdots\cdots(13)$$

となる。

$$\lambda=\frac{2\pi}{T}\qquad m=(N+r)T\qquad (N:\ 0\text{または正整数}\ 0\leqq r<1)$$

とおくと

$$V(\bar{x}_{sy})=\frac{A^2}{2}\left\{\frac{\sin(nr\pi)}{n\sin(r\pi)}\right\}^2 \qquad \cdots\cdots(14)$$

となる。以上より$V(\bar{x}_{sy})$ の大きさは A^2，n，r により定まる。A^2 は母数で一定と考えられるから，n，r によって $V(\bar{x}_{sy})$ がどのように変わるかということに興味が残される。

$$f\ (n,\ r)=\left\{\frac{\sin(nr\pi)}{n\sin(r\pi)}\right\}^2 \qquad \cdots\cdots(15)$$

とおいて，n，rの値をいろいろ変えて$f(n,\ r)$の値を計算する表1が得られる。

$f(n,\ r)=f(n\,1-r)$であるから計算は $0\leqq r\leqq 1/2$について行なった。一番右の列にある数字は$1/n$でランダムサンプリングの誤差に対応するものである。

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

表1　$f(n, r)$ の値

	0	0.01	0.05	0.1	0.2	0.3	0.4	0.5	1/n
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	1.000	0.999	0.976	0.905	0.655	0.346	0.096	0.000	0.500
3	1.000	0.997	0.936	0.762	0.291	0.016	0.043	0.111	0.333
4	1.000	0.995	0.882	0.592	0.062	0.033	0.062	0.000	0.250
5	1.000	0.992	0.817	0.419	0.000	0.061	0.000	0.040	0.200
6	1.000	0.989	0.744	0.263	0.028	0.015	0.028	0.000	0.167
7	1.000	0.984	0.662	0.140	0.053	0.003	0.008	0.020	0.143
8	1.000	0.979	0.577	0.057	0.041	0.022	0.006	0.000	0.125
9	1.000	0.974	0.492	0.012	0.012	0.012	0.012	0.012	0.111
10	1.000	0.968	0.409	0.000	0.000	0.000	0.000	0.000	0.100
11	1.000	0.962	0.329	0.008	0.008	0.008	0.008	0.008	0.091
12	1.000	0.954	0.256	0.025	0.018	0.010	0.003	0.000	0.083
13	1.000	0.946	0.192	0.041	0.016	0.001	0.002	0.006	0.077
14	1.000	0.938	0.136	0.048	0.005	0.002	0.005	0.000	0.071
15	1.000	0.929	0.091	0.047	0.000	0.007	0.000	0.004	0.067
16	1.000	0.919	0.055	0.037	0.004	0.002	0.004	0.000	0.062
17	1.000	0.909	0.029	0.024	0.009	0.001	0.001	0.003	0.059
18	1.000	0.898	0.012	0.011	0.008	0.004	0.001	0.000	0.056
19	1.000	0.887	0.003	0.003	0.003	0.003	0.003	0.003	0.053
20	1.000	0.876	0.000	0.000	0.000	0.000	0.000	0.000	0.050
22	1.000	0.851	0.008	0.007	0.005	0.003	0.001	0.000	0.045
24	1.000	0.825	0.025	0.016	0.002	0.001	0.002	0.000	0.042
26	1.000	0.809	0.040	0.014	0.001	0.001	0.001	0.000	0.038
28	1.000	0.768	0.047	0.004	0.003	0.002	0.000	0.000	0.036
30	1.000	0.737	0.045	0.000	0.000	0.000	0.000	0.000	0.033
32	1.000	0.706	0.036	0.003	0.003	0.001	0.000	0.000	0.031
34	1.000	0.674	0.023	0.008	0.001	0.000	0.001	0.000	0.029
36	1.000	0.640	0.011	0.007	0.001	0.000	0.001	0.000	0.028
38	1.000	0.607	0.003	0.003	0.002	0.001	0.000	0.000	0.026
40	1.000	0.573	0.000	0.000	0.000	0.000	0.000	0.000	0.025
42	1.000	0.539	0.002	0.002	0.001	0.001	0.000	0.000	0.024
44	1.000	0.505	0.007	0.005	0.001	0.000	0.001	0.000	0.023
46	1.000	0.472	0.013	0.005	0.000	0.000	0.000	0.000	0.022
48	1.000	0.438	0.016	0.002	0.001	0.001	0.000	0.000	0.021
50	1.000	0.406	0.016	0.002	0.000	0.000	0.000	0.000	0.020

これを図に示すと，図5になる。

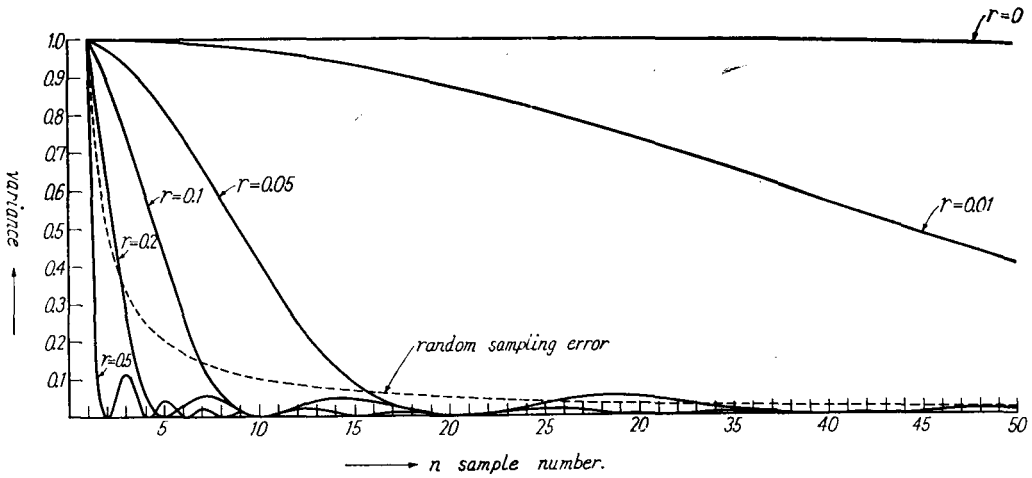

図　5

© 日本オペレーションズ・リサーチ学会. 無断複写・複製・転載を禁ず.

5. ランダムサンプリグとの比較結論

図5からわかるようにサンプリング間隔が母集団周期の整数倍になっている場合はサンプル数を増加しても精度の向上はないが，サンプリング間隔が母集団周期の整数倍と異なる場合は，サンプリング精度は$(1/n^2)$の項によりnの増加とともに急速によくなっていく。rを縦軸に，nを横軸にとって系統サンプリングの方が精度がよくなる範囲を求めると，だいたい図6の斜線で示した区域になる。この図からわかるように大部分の場合に系統サンプリングの方がランダムサンプリングよりも精度がよい。

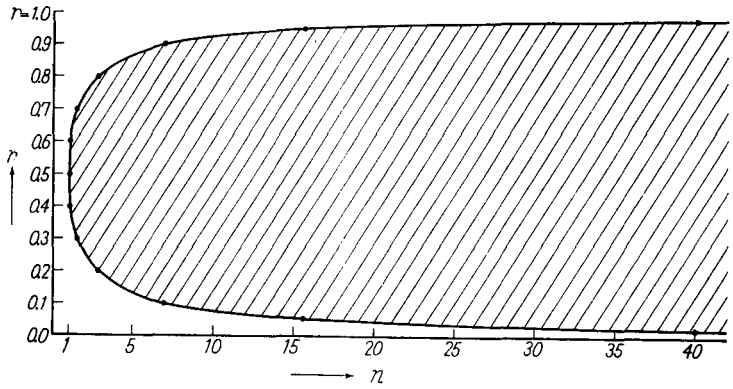

図 6

nがあまり大きくない場合には系統サンプリングの精度は $\{sin\ nr\pi/sin\ r\pi\}^2$ の大きさにより支配される。r が0，あるいは1に近いほど精度が悪くなる。

ある値のrで系統サンプリングを行なったとき，同じサンプル数でランダムサンプリングよりも精度がよくなるnの値を求めると近似的につぎのようになる。すなわち与えられたr($0<r<\frac{1}{2}$)に対して

$$\left\{\frac{sin\ nr\pi}{sin\ r\pi}\right\}^2 \frac{\sigma^2}{n^2} < \frac{\sigma^2}{n} \qquad \cdots\cdots(16)$$

が成り立つnの値を求めればよい。($\frac{1}{2}<r<1$のときは $r'=1-r$ としてr'について以下と同じ計算ができる)。

$$\left\{\frac{sin\ nr\pi}{sin\ r\pi}\right\}^2 = \frac{1-cos\,2nr\pi}{1-cos\,2r\pi}$$

$$= \frac{\frac{1}{2!}(2nr\pi)^2 - \frac{1}{4!}(2nr\pi)^4 + \cdots\cdots}{\frac{1}{2!}(2r\pi)^2 - \frac{1}{4!}(2r\pi)^4 + \cdots\cdots}$$

$$\fallingdotseq \frac{\frac{1}{2!}(2nr\pi)^2 - \frac{1}{4!}(2nr\pi)^4}{\frac{1}{2!}(2r\pi)^2} = n^2 - \frac{2!}{4!}(2r\pi)^2 n^4$$

$$= n^2\left(1 - \frac{1}{3}r^2\pi^2 n^2\right) \qquad \cdots\cdots(17)$$

© 日本オペレーションズ・リサーチ学会. 無断複写・複製・転載を禁ず.

これを(16)に代入して

$$n-\frac{1}{3}(r\pi)^2 n^3<1$$

したがって

$$n-\frac{1}{3}(r\pi)^2 n^3<0$$

であればよい。これより

$$n>\frac{\sqrt{3}}{\pi r}\fallingdotseq\frac{1.732}{3.14r} \qquad \cdots\cdots(18)$$

近似の程度を考慮して

$$n>\frac{1}{r} \qquad \cdots\cdots(19)$$

なら系統サンプリングの方がランダムサンプリングよりも精度がよいといえる。これは図5，図6からも大体正しいことがみられる。r が0に極めて近い場合には，たとえば $r=0.05$の場合について考えると $n=1/0.05=20$となるが，図5の $r=0.05$のグラフから$26\leqq n\leqq 33$の範囲で系統サンプリングよりランダムサンプリングの方がわずかに精度がよく，式(19)の近似は成りたたないが，実用上の目安を与える式として式(19)を利用しても問題はないと思われる。

母集団の周期についての情報を持たない場合，任意にサンプリング間隔を定めて観測を行なうときの誤差の期待値について考えてみよう。母集団周期Tに対してサンプリング間隔mが全く無関係に，偶然的に定められるとすれば，r の値は $0\leqq r<1$の範囲で，偶然的に決まることになる。すなわちrは確率変数で，その分布は $0\leqq r<1$で一様分布になるとすれば $V(\bar{x}_{sy})$ の期待値は

$$E\left[V(\bar{x}_{ry})\right]=\int_0^1\left\{\frac{\sin nr\pi}{\sin r\pi}\right\}^2\frac{\sigma^2}{n^2}\,dr \qquad \cdots\cdots(20)$$

定積分の公式により

$$\int_0^\pi\left\{\frac{\sin nx}{\sin x}\right\}^2 dx=n\pi$$

であるから，$x=r\pi$ とおけば

$$E\left[V(\bar{x}_{sy})\right]=\frac{\sigma^2}{\pi n^2}\int_0^\pi\left\{\frac{\sin nx}{\sin x}\right\}^2 dx=\frac{\sigma^2}{n} \qquad \cdots\cdots(21)$$

となり，単純ランダムサンプリングの場合の誤差　注〕と一致する。母集団に周期のある場合，系統サンプリングを用いるとランダムサンプリングを用いた場合よりも精度の良くなる場合もあり悪くなる場合もあるが，平均的にはその精度は全く同じであるといえるのである。

注〕 観測を行なう期間を $0\leqq t\leqq L$と区間$(0,L)$で観測をn回ランダムな時刻に行なうとすれば，サンプリ

© 日本オペレーションズ・リサーチ学会. 無断複写・複製・転載を禁ず.

ング誤差は

$$\sigma_r^2=\frac{\sigma^2}{n}\left[1+(n-1)\left\{\frac{\sin \frac{1}{2}\lambda L}{\frac{1}{2}\lambda L}\right\}^2\right]$$

となり一般に σ_r^2はσ^2/nよりも大きくなるが，ここでは $\sin\frac{1}{2}\lambda L=0$，すなわち観測を行なう期間が母集団周期の整数倍にとられると仮定している。

6. 結　　論

従来母集団に周期がある場合には，系統サンプリングを用いることは余り好ましいことではないといわれているが，サンプリング間隔と母集団の周期（あるいはその整数倍）とが一致している場合や，両者の値が非常に近い場合にはサンプリングの誤差は大きくなるが，サンプリング間隔と母集団の周期（あるいはその整数倍）の差 r が母集団周期に対して10%以上異なっておれば，サンプルを10個もとれば系統サンプリングの方が精度がよくなる。

母集団の周期性についての情報を全然持たないときのサンプリング精度は，minimax 的な意味ではランダムサンプリングの方がよいが，平均的には全く同じであり，周期性について何らかの知識を持っておれば，その知識を用いてサンプリング間隔を適当に定めることにより非常に精度のよい結果を得ることが可能であり，このような場合には積極的に系統サンプリング法を用いる方がよい。

式(1)で表わされる一般の周期関数あるいは概周期関数について系統サンプリングを行なった場合の誤差は $A_i\ \sin(\lambda_i t+\alpha_i)\ i=1, 2\cdots\cdots$の各についての誤差を第4節の方法で求め，$(A_i^2)$の重みで平均して計算することができる。

他のサンプリング法，たとえば層別ランダムサンプリング，ジグザグ・サンプリングを用いた場合の測定誤差については次の機会に述べることにする。

参　考　文　献

〔1〕Madow, W.G. & L.H (1944). "On the theory of systematic sampling", Annals of Math. Stat., Vol. 15, pp. 1—24

〔2〕Cochran, W.G. (1946). "Relative accuracy of systematic and stratified random samples for a certain class of populations". Ann. Math. Stat., Vol. 17, pp. 164—177

〔3〕Cochran, W.G. (1953). ***Sampling Techniques***. John Wiley & Sons, Inc. pp. 160—188.

〔4〕Kendall, M.G. (1948). ***The Advanced Theory of Statistics*** II, Charles Grifin & Co., Ltd. London pp. 363—437

〔5〕日本応用力学会編(1949)，応用統計学，金原出版　7.01—7.44

〔6〕森口繁一他，(1957)．数学公式 I，II　岩波全書

〔7〕Tables of Natural Value of Trigonometric Function (1) (1961). Corona Publishing Co., Ltd.

〔8〕Barlow's Table of Squares Cubes etc.，森北出版

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．