

〈総合報告〉

モンテカルロシミュレーションの  
統計的方法†

竹 内 啓\*

1. はじめに——モンテカルロ法の二つの目的

モンテカルロシミュレーションとは、何らかの形で確率的メカニズムをふくむシステムの性質を、コンピューターによって乱数を発生させることによって、シミュレートしてしらべることであると定義できよう。

このような実験を行なう目的は、一般的に2種類あるとよい。一つは教育的目的というべきもので、システムの変動するようすを目に見える形であらわすことによって、そのようなシステムに対する感覚を養うことである。この場合には、システムの数学的性質は、解析的にはあらかじめわかっていることも多く、ただそれがどんなことを意味するかについて、直感的に理解できるようになることが実験を行なう目的である。最も簡単な例をあげれば、ランダムな数の系列というものは、部分的にはかなり偏って見えることが多く、けっして直感的にそう思われるほど一様ではないということは、乱数を実際によく眺めてみて、はじめて理解できるようになるのである。このような教育的実験は、学生や企業の新人を対象とする、せまい意味の教育の場でたいせつであることはもちろん、実は研究の場においても、実際のシステムを眺めるときの感覚を養うために、しばしば重要になる。

モンテカルロシミュレーションの第二の目的は、複雑なシステムにおいて、解析的にシステムの特性値を求めることが不可能であるような場合に、実験によってその値を求めることである。以下ではもっぱら、この第二の場合を扱う。

特性値を求めるための実験においては、得られた値の精度と信頼性が最も重要である。この点では教育的実験の場合と非常に異なる。後者の場合には、特性値の“真の値”よりも、実験における変動性についての感覚を得ることのほうが重要であることが多い。したがってこの二つの種類の状況は、明確に区別して考えねばならない。

しかし、現実の場においては、この二つの問題は最初からはっきりと分離したものとして現われるとは限らない。非常に複雑なシステムにおいては、どのような特性値を指標としてえらんだ

† 1972年12月26日受理。1972年4月、月例講演会にて報告。

\* 東京大学経済学部。

らよいかも、事前には明らかでないこともある。そのときには、いわゆる教育的実験のなかからシステムのいわば“イメージ”をつかみ、それによって計算すべき特性値を定めるということが必要になる。しかしこのような場合でも、そのような特性値を信頼するに足る精度で求めるためには、とにかく第二の段階として、第二の種類の実験を行なわねばならない。教育的目的の実験だけで両者をかねるといことは、必要な計算量という観点からも一般に不可能である。

以下このような特性値を求めるための実験において、注意すべき点をいくつかとり上げたい。

## 2. 平均の存在

いま最も典型的な例として、ある確率変数  $Y$  の期待値を求める問題を考えよう。 $Y$  は、ある同時分布に従う確率変数  $X_1, \dots, X_n$  の関数として

$$Y=f(X_1, \dots, X_n)$$

と表わされるものとする。ここで  $Y$  として、 $X_1, \dots, X_n$  によって定められるある事象  $A$  が起これば  $Y=1$ 、起こらなければ  $=0$  という変数を考えれば、 $Y$  の期待値は、 $A$  が起こる確率に等しい、

$$E(Y)=P_r\{A\}$$

から、期待値を求める問題は、確率を求める問題もふくんでいる。

このような場合に、 $E(Y)$  を求める最も素朴な方法は、 $X_1, \dots, X_n$  の分布と同じ分布を持つ  $X'_1, \dots, X'_n$  を発生させて

$$Y'=f(X'_1, \dots, X'_n)$$

を計算し、その過程を何回もくり返して、計算された値の平均値

$$\bar{Y}'=\sum Y'/N \quad (N \text{ はくり返し回数})$$

を  $E(Y)$  の推定値とすることである。

このような計算を行なう場合に、まず第一に前提されねばならないことは、 $E(Y)$  の存在である。もちろん  $Y$  の値が有界であれば、それは自明であるし、またそうでなくても、解析的に平均値の存在が容易に示されることもあるが、そうでない場合には十分注意が必要である。実際世の中には、平均値が存在しないにもかかわらず、それを求めようとしている例もときどき見られる。さらになかには「 $E(Y)$  が理論的に存在しなくても、 $\bar{Y}'$  を標本から求めることはつねに可能なのだから、それはそれとして意味を持つ」というような主張をするものもあるが、これは無意味である。なぜならば、 $E(Y)$  が存在しないならば、 $\bar{Y}'$  の値は一般に  $N$  をいくら大きくしても一定の値に収束しない（大数法則の逆）から、 $\bar{Y}'$  の値を計算してみても、それが何も意味ある量を表わしていないことになる。

実際理論的に平均値が存在するかどうか明確でない場合には、 $N$  の増加に対する  $\bar{Y}'$  の値の変化を見ればよい。もし  $E(Y)$  が存在しないならば、その値は特定の値に収束しないで、どんどん絶対値が大きくなるか、あるいは振動するであろう。さらにとくにこの点を確かめたい場合には、 $|Y'|$  の平均値を計算して、それが有限の値に収束するか否かを確かめることができる。

しかし単に  $E(Y)$  が存在するだけでは、実は十分でない。というのは、もし  $E(Y^2)=\infty$ 、す

なわち分散が有限でなければ、 $N$ が増加するにつれて、 $\bar{Y}'$ の値が $Y$ の平均値に収束する速さが定められないからである。したがってそのときには、真の値に対する信頼区間を求めることができなくなる。これに対して $E(Y^2) = \sigma^2 < \infty$ ならば、 $N$ が十分大きいとき $\bar{Y}'$ は平均 $E(Y)$ 、分散 $\sigma^2$ の正規分布に従うと考えるとよいから、 $\sigma^2$ を

$$\sigma^2 = \sum (Y' - \bar{Y}')^2 / N$$

として、たとえば95%信頼区間を

$$\bar{Y}' - 1.96\hat{\sigma}/\sqrt{N} < E(\bar{Y}') < \bar{Y}' + 1.96\hat{\sigma}/\sqrt{N}$$

という形で与えることができる（一般に $N=1000$ というような大きさであれば、 $\hat{\sigma}^2$ を求めるときに $N-1$ でなく $N$ で割り、また $t$ 分布の代わりに正規分布を用いてさしつかえない）。

モンテカルロ実験の結果において、このような信頼区間が与えられていない場合が多いが、もしその結果によって特性値の大きさを論ずるとすれば、このように推定値の信頼度を与えておくことは絶対に必要である。

ところで、このような議論からも明らかなように、実験による推定値の精度は $\sqrt{N}$ に比例する。したがって、単にくり返し数を増すことのみによって精度を上げようとするのは、非常に効率が悪い。そこで効率を上げるための種々の方法が考えられている。

### 3. 乱数発生の問題

効率を上げる方法を論ずる前に、一つ注意すべき点として、 $X_1', \dots, X_n'$ および $Y'$ の分布が $X_1, \dots, X_n$ と $Y$ の分布に、正確にあるいは十分高い精度で一致しなければならないということがある。ふつうある同時分布に従う乱数 $X_1', \dots, X_n'$ は、一様分布に従う乱数 $U_1, \dots, U_m$ から( $m \geq n$ であるが、 $m=n$ とは限らない)

$$X_i' = g_i(U_1, \dots, U_m) \quad i=1, \dots, n$$

という形で計算される。そこで問題には一様乱数の発生、および $g_i$ と $f$ の計算の精度の二つの側面が生ずる。

一様乱数の発生については、ふつう剰余法やその変種が用いられるが、私の経験、あるいは見聞した限りでは、その点から問題が生ずることはあまりないようである。というのは、 $n$ があまり大きくない限り（すなわち長い時系列から計算される値の特性を求めたりするのでなければ）、乱数のわずかな周期性はそれほど重大な問題にならず、頻度が一樣でさえあれば偏りのない結果が得られるからである。ただし非常に大きい $N$ に対して $\bar{Y}'$ の分散が $N^{-1}$ に比例しなくなるかもしれないから、信頼区間については問題が生ずる危険性があるが、このような点については、まだあまりしらべられていないようである。

これに対して計算の精度は重要である。まず $U$ は本来 $[0, 1]$ 区間の連続な変量と考えられているが、現実には有限のケタ数しか求められていないから、実際には $0.0 \dots 0$ から $0.99 \dots 9$ までの値を等確率でとる離散分布に従っているにすぎない。この分布が連続分布を十分よく近似していると考えられる限りは、そのことから問題が生ずることはないが、もし $U$ が1に近いとこ

ろのわずかな変動が  $X$  の分布に大きい影響を与えるとすれば、ケタ数をかなり大きくとっても、なおかつ結果に偏りを生ずるかもしれない。この点を一つの例で示そう。

いま自由度 2 の  $t$  分布を考えよう。その密度は、

$$(3.1) \quad f(x) = (1/2)(1+x^2)^{-3/2}$$

で与えられる（実は簡単のために尺度を変えてあるが）。このとき累積分布関数は

$$U = F(x) = \int_{-\infty}^x f(x) dx = (1/2) \{1 + x/\sqrt{1+x^2}\}$$

となるから、 $U$  を一様分布に従う乱数として、

$$X = F^{-1}(U) = (U - 1/2) / \sqrt{U(1-U)}$$

とすれば、 $X$  が (3.1) を密度とする分布に従うことになる。

ところでこの分布において、分散は、

$$\frac{1}{2} \int_{-\infty}^{\infty} x^2 (1+x^2)^{-3/2} dx = \infty$$

であるが、 $U'$  を 0.0……0 から 0.99……9 までの値を一様にとる離散変量として

$$X' = F^{-1}(U')$$

とすると、 $X'$  の分散は有限になる。しかも  $U'$  を  $k$  ケタとったとすると、 $X'$  の分散は  $k \log_e 10$  に比例することがわかる。したがって  $k$  をかなり大きくとっても、 $X$  の真の分散が無限大になることを示唆するような大きい値は得られないであろう。

同じことは、 $U$  から  $X'$  を、あるいは  $X'$  から  $Y'$  を計算するときの計算精度の問題と関連して生ずる。 $Y'$  の分布の、いわば中央の値を計算する場合には、計算上の誤差は、偶然変動（確率変動）と打ち消し合うことが多いから、あまり問題にならないが、 $X'$  の分布の両端ではそれは結果に偏りを生ぜしめる可能性がある。したがって分布の両端によって大きく影響されるような特性値を計算する場合には、十分な注意が必要である。

#### 4. モンテカルロ法の非効率性

ところで一般に単純なモンテカルロ実験は、精度の高い推定値を求める方法としては、あまり効率の高いものではない。

今度はその点を積分を求める問題として考えてみよう（もし  $Y$  の密度関数を求めることができれば、期待値を求めることは積分の計算に帰着する）。

いま  $g(x)$  を実変数  $x$  の関数とし、定積分

$$I = \int_a^b g(x) dx$$

を求めることを考える。これをモンテカルロ法によって計算する最も簡単な方法は、区間  $(a, b)$  のなかに、ランダムに  $N$  個の点  $x_1, \dots, x_N$  をえらんで

$$\hat{I}_1 = \{(b-a)/N\} \sum_{i=1}^N g(x_i)$$

とすることである。

このとき、いうまでもなく

$$E(\hat{I}_1) = I$$

すなわち  $\hat{I}_1$  は  $I$  の不偏な推定量となり、またその分散は

$$\sigma_{\hat{I}_1}^2 = 1/N \sigma_1^2 \quad \sigma_1^2 = \int_a^b (b-a)g^2(x)dx - I^2$$

となる。したがって誤差は  $\sqrt{N}$  に逆比例する。

ところが  $I$  を数値積分によって求めるとすると、最も簡単な台形公式

$$I_1 = \{(b-a)/N\} \{(1/2)g(a) + (1/2)g(b) + \sum_{i=1}^{n-1} g(h_i)\}$$

$$h_i = a + \{(b-a)/N\}i \quad i=1, \dots, N$$

を用いても、その誤差はよく知られているように一般に  $N^{-2}$  に比例する ( $g(x)$  が 2 回微分可能であればそうなる)。  $g(x)$  が十分なめらかであるとしてシンプソンの公式を適用すれば、誤差は  $N^{-4}$  に比例する。したがってモンテカルロ法による推定はきわめて能率が悪い。

効率を高める一つの方法は、区間をいくつかに分割して、それぞれの区間からランダムに点をえらぶことである、最も極端な場合には、  $N$  個の区間  $(h_{i-1}, h_i)$   $i=1, 2, \dots, N$  からそれぞれ 1 個ずつ点  $x_i'$  をえらんで

$$\hat{I}_2 = (b-a) \sum g(x_i')/N$$

とすることである。このとき  $\hat{I}_2$  の分散は

$$\sigma_{\hat{I}_2}^2 = (1/N^2) \sum \sigma_i^2$$

$$\sigma_i^2 = \{(b-a)/N\} \int_{h_{i-1}}^{h_i} g^2(x)dx - \left( \int_{h_{i-1}}^{h_i} g(x)dx \right)^2$$

となる。  $g(x)$  が微分可能とすれば、  $\sigma_i^2$  はほぼ  $N^{-2}$  に比例するから、  $\sigma_{\hat{I}_2}^2$  は  $N^{-3}$  に比例する大きさとなる。したがって  $\hat{I}_2$  の誤差はほぼ  $N^{-3/2}$  に比例する大きさとなり、なお台形公式よりも能率が悪い。

なお効率を高める二つの方法がある。一つは適当な変数変換  $x=h(y)$  を行なって、

$$I = \int_c^d g\{h(y)\} h'(y) dy$$

$$\text{ただし、 } h(c)=a \quad h(d)=b$$

とすることである。ここで  $g\{h(y)\} h'(y)$  がほぼ一定になるようにすることができれば、  $y$  の値をランダムにえらんで推定を行なうとき、推定値の分散が小さくなる。

もう一つの方法は、  $g(x)$  にほぼ等しく、かつ解析的に積分の値が求められるような関数  $\hat{g}(x)$  を用いて、

$$I = \int_a^b \{g(x) - \hat{g}(x_1)\} dx + \int_a^b \hat{g}(x) dx$$

とし、 $g(x) - \hat{g}(x)$  の積分をモンテカルロ法によって求めることである。もし  $g(x) - \hat{g}(x)$  の値がほぼ 0 に等しいならば、推定値の分散は小さくなるであろう。

このような方法は、確率変数の期待値を求める計算にもしばしば応用されるので、後にさらにのべる。

## 5. 高次元の場合

しかしながら、モンテカルロ法は高次元の積分を求めるには有効な方法である。

いま

$$I = \int_S \cdots \int_S g(x_1, \dots, x_m) dx_1 \cdots dx_m.$$

を求めることを考えよう。積分範囲  $S$  は、簡単のために長さ 1 の超立方体としよう。

このとき  $I$  を数値積分によって求める最も簡単な方法は、積分範囲を長さ  $h$  の小さい超立方体に分割し、その中心における  $g$  の値を加えて平均することであろう。 $g$  が連続 2 回微分可能であるとすると、このような方法による計算値の誤差は  $h^2$  に比例する。ところで計算する点の数を  $N$  とすると  $N = h^{-m}$  だから、結局誤差は  $N^{-2/m}$  に比例することになる。

これに対して  $S$  からランダムに  $N$  個の点をえらんで  $I$  の推定値を求めれば、その分散はやはり  $N^{-1}$  に比例するから、誤差はほぼ  $N^{-1/2}$  に比例することになる。したがって  $m$  が大きい ( $m > 4$ ) ならば、モンテカルロ法によるほうが、精度が高くなる。

もちろんこの場合にも、数値積分においていろいろな方法を工夫することは可能であるが、いずれにしても区間の幅  $h$  に対して、点の数  $N$  は  $h^{-m}$  に等しくなるから、 $m$  が大きくなるにつれて、モンテカルロ法が相対的にだんだん有力な方法となることには変わりはない。

ただしここにのべたことは、積分の次元をへらすことが不可能な場合についてのみにあてはまることであって、何らかの工夫によって次元をへらすことができれば、数値積分が有力になることはいうまでもない。たとえば、もし

$$g(x_1, \dots, x_m) = g_1(x_1) \cdots g_m(x_m)$$

と分解できるならば

$$I = \int_a^b g_1(x_1) dx_1 \times \cdots \times \int_a^b g_m(x_m) dx_m$$

というように表わされるから、それぞれの項を数値積分によって求めて、その積をつくれれば、容易に精度の高い計算を行なうことができる。

もっと複雑な場合でも、適当な変数変換等により次元をへらすことができることも少なくない。

## 6. 効率を高める方法

ここで再び2節で与えたような確率変数の期待値を計算する問題を考えよう。

まず、もし  $Y$  の密度関数が比較的容易に求められるならば、それを  $p(y)$  とするとき、 $Y$  の関数  $g(Y)$  の期待値を求めるには、モンテカルロ法によるよりも、直接数値積分によって

$$E\{g(Y)\} = \int g(y)p(y)dy$$

と計算するほうが効率がよい。

モンテカルロ法が有効に利用されるのは、 $X_1, \dots, X_m$  の分布から  $Y=f(X_1, \dots, X_m)$  の分布が容易に導けないような場合である。この場合にも、効率を高め、精度を上げるようないくつかの方法がある。それには次のようなものがある。

1. 層別抽出
2. 分布の変換
3. 補助変量の利用
4. 統計的性質の応用

層別抽出とは、 $X_1, \dots, X_m$  の変域をいくつかの部分に分け、それぞれの部分から別個にランダムに点をえらぶ方法であり、さらに各部分からえらぶ点の数は必ずしもそれぞれの確率に比例することなく、全体の推定量の分散が小さくなるように配分することができる。

分布の変換というのは、次のようなものである。いま  $X_1, \dots, X_m$  の同時密度関数を  $p(x_1, \dots, x_m)$  とするとき、 $Y=f(X_1, \dots, X_m)$  の期待値は次のように表わされる。

$$E(Y) = \int \dots \int f(x_1, \dots, x_m) p(x_1, \dots, x_m) dx_1 \dots dx_m$$

ところが  $p^*(x_1, \dots, x_m)$  を他の密度関数とすると

$$E(Y) = \int \dots \int f(x_1, \dots, x_m) \frac{p(x_1, \dots, x_m)}{p^*(x_1, \dots, x_m)} p^*(x_1, \dots, x_m) dx_1 \dots dx_m$$

と変形することができるから、

$$g(X'_1, \dots, X'_m) = f(X'_1, \dots, X'_m) \frac{p(X'_1, \dots, X'_m)}{p^*(X'_1, \dots, X'_m)}$$

とおき、 $X'_1, \dots, X'_m$  が  $p^*(x'_1, \dots, x'_m)$  を密度関数とする確定変数とすれば

$$E(Y) = E\{g(X'_1, \dots, X'_m)\}$$

となるから、 $X'_1, \dots, X'_m$  を乱数表から生成して実験をくり返すことにより、 $E(Y)$  を求めることができる。

ここでもし  $g(X'_1, \dots, X'_m)$  の値が、 $f(X_1, \dots, X_m)$  の値にくらべて、一定の値に近くなっていならば、このような変換を行なうことによって、推定量の分散を小さくすることができる。

実は変数変換の効果は、単に分散を小さくすることに限られない。 $X_1, \dots, X_m$  の分布に従う

乱数を生成することが困難である場合に、それに近い分布で乱数をつくりやすい  $X_1', \dots, X_m'$  をとって上記のような計算を行えば、計算時間が短縮される。

たとえば、 $X_1, \dots, X_m$  が互いに独立に平均 0、分散 1 の正規分布に従うとされる場合、もし正規乱数の生成が困難と思われる（もちろんこのことはそれほど困難ではないが）ならば、次のようにすることも可能である。 $X_1', \dots, X_m'$  を次のような密度関数を持つ、両側指数分布に従う確率変数とする：

$$p^*(x) = e^{-|x|} / 2$$

そうすると、 $U_1, \dots, U_m$  を一様分布に従う乱数とするとき、

$$\begin{aligned} X_i' &= \log 2U_i & U_i < 1/2 \text{ のとき} \\ &= -\log 2(1-U_i) & U_i > 1/2 \text{ のとき} \end{aligned}$$

とすれば、仮定された分布に従う乱数が得られる。したがって  $X_1, \dots, X_m$  のある関数  $Y=f(X_1, \dots, X_m)$  の期待値を求めるには、

$$g(X_1', \dots, X_m') = f(X_1', \dots, X_m') \times (\sqrt{2/\pi})^m \exp\{\sum (X_i') - (1/2) \sum X_i'^2\}$$

の平均を計算すればよい。

## 7. 補助変量の利用

補助変量の利用の方法には、いろいろな形がある。最も簡単な方法は、 $Y$  と非常に相関が高く、かつ  $Y$  の期待値が知られているような  $Y^* = f^*(X_1, \dots, X_m)$  を用いて、

$$E(Y - Y^*)$$

をモンテカルロ法によって求め、それに  $E(Y^*)$  を加えたものを  $E(Y)$  の推定値とするのである。これは差推定ともいふべき方法である。

これに対して“比推定”その他も考えられる。もし比  $Y/Y^*$  がほぼ一定と考えられ、かつ  $Y^*$  がつねに正ならば、 $E(Y)$  と  $E(Y^*)$  の比を

$$\sum Y / \sum Y^*$$

によって推定し、 $E(Y)$  の推定値を

$$E(Y^*) \times (\sum Y / \sum Y^*)$$

という形で与えることも考えられる。

しかしながら、このような推定量は不偏にならない。不偏な推定量を求めるには“不偏比推定量”を計算すればよい。それは次の式で与えられる。

$$(7.1) \quad \{N/(N-1)\} (\bar{Y} - \bar{R}\bar{Y}^*) + \bar{R}E(Y^*)$$

ただしここに、 $\bar{Y} = \sum Y/N$ 、 $\bar{Y}^* = \sum Y^*/N$  であり、また  $\bar{R}$  は比  $Y/Y^*$  の平均、すなわち

$$\bar{R} = (1/N) \sum Y/Y^*$$

である。 $N$  が大きければ、(7.1) はほぼ

$$\bar{Y} - \bar{R} \{\bar{Y}^* - E(Y^*)\}$$

に等しいものとしてよい。

またいくつかの補助変数  $Y_1^*, \dots, Y_k^*$  が与えられている場合に、

$$Y = a_1 Y_1^* + \dots + a_k Y_k^*$$

として、 $E(Y)$  を

$$E(Y) = a_1 E(Y_1^*) + \dots + a_k E(Y_k^*)$$

から計算することが考えられる。ここに  $E(Y_1^*), \dots, E(Y_k^*)$  はすべて既知とし、また係数  $a_1, \dots, a_k$  は最小 2 乗法を適用して

$$Q = \sum (Y - a_1 Y_1^* - \dots - a_k Y_k^*)^2 = \text{最少}$$

となるように計算すればよい。このような推定方式は“回帰推定”と呼ばれている。このような推定量も不偏ではないが、 $Q$  が小さければ、偏りはほとんど問題にならない。

補助変数による方法の一種として、先にのべた変数変換を利用することもできる。すなわち  $Y = f(X_1, \dots, X_m)$  の期待値を求めるとき、もし  $X_1, \dots, X_m$  に近い分布を持つ  $X_1', \dots, X_m'$  に対して  $Y' = f(X_1', \dots, X_m')$  の期待値は知られているならば、

$$\begin{aligned} Y &= f(X_1', \dots, X_m') \frac{p(X_1', \dots, X_m')}{p^*(X_1', \dots, X_m')} \\ &= Y' \frac{p(X_1, \dots, X_m)}{p^*(X_1', \dots, X_m')} \end{aligned}$$

として、 $Y$  と  $Y'$  の比を計算し、上にのべた比推定の方法を適用して  $E(Y')$  から  $E(Y)$  を計算することができる。このような方法は、とくに同じ統計量のいろいろな分布の下での性質を比較するのに有力である。

また補助量による方法の変種として、 $Y$  と同じ期待値を持つことが知られている  $Y_2 = f_2(X_1, \dots, X_m), \dots, Y_h = f_h(X_1, \dots, X_m)$  が存在するとき、 $Y$  の期待値を

$$(1/h)(\bar{Y} + Y_2 + \dots + \bar{Y}_h)$$

で推定することが考えられる。この方法はとくに  $Y, Y_2, \dots, Y_h$  の間に負の相関が存在すれば、精度が高くなる。

とくに分布の対称性を利用して、適当な  $Y_2$  以下を見いだすことができる。すなわち  $X_1, \dots, X_m$  が互いに独立に (0,1) 区間の一様分布に従うとすれば、 $1 - X_1, \dots, 1 - X_m$  も同じ分布になる。したがって

$$Y_2 = f(1 - X_1, \dots, 1 - X_m)$$

とおけば、 $E(Y_2) = E(Y)$  となり、また  $f$  が各変数に関して単調ならば、 $Y$  と  $Y_2$  は負の相関を持つであろう。したがって  $(\bar{Y} + \bar{Y}_2)/2$  によって  $E(Y)$  を推定すれば、分散が小さい推定量が得られるであろう。このような方法はしばしば“antithetic variable”による方法と呼ばれている。

## 8. 数理統計理論の利用

最後に、統計理論の応用として、最も量要なものは、いろいろな統計量の間での独立性を利用することである。たとえば次のことが成り立つ。

$$1. \circ X_1, \dots, X_n \text{ が互いに独立に平均 } 0, \text{ 分散 } \sigma^2 \text{ の正規分布に従うとき、 } \bar{X} = \sum X_i/n \quad S =$$

$\sqrt{\sum (X_i - \bar{X})^2 / (n-1)}$  とおく.  $T = t(X_1, \dots, X_n)$  を任意の統計量とすると、 $T$  が位置変換に関しても不変、すなわち

$$t(X_1 + a_1, \dots, X_n + a_n) \equiv t(X_1, \dots, X_n)$$

ならば、 $T$  は  $\bar{X}$  と独立であり、さらに尺度変換についても不変、すなわち

$$t(cX_1 + a, \dots, cX_n + a) \equiv t(X_1, \dots, X_n)$$

ならば、 $T$  は  $X$  および  $S$  と独立である。

このことを利用して、たとえば

$$R = \max X_i - \min X_i$$

の分布の特性値を求めてみよう。

まず  $R$  の期待値を求めるのに、上記の命題から  $R/S$  と  $S$  とが独立であることを利用すれば

$$(8.1) \quad E(R/S)E(S) = E(R)$$

となる。  $E(S)$  の値は知られているから、  $R/S$  の期待値を求めればよい。  $R/S$  の変動係数（分散と平均の二乗との比）は  $R$  自体の変動係数の比より小さいから、  $R/S$  の平均値を正規乱数から求めて計算すれば、比較的精度の高い推定値が得られるであろう。

(8.1) の  $S$  は、  $S' = \sqrt{\sum X_i^2 / n}$  でおきかえてもよい。 さらにこの場合、次のようにしてより精度の高い方法が得られる。いま  $X$  の絶対値の集合  $|X_1|, \dots, |X_n|$  が与えられたものとして、条件付分布を考える。  $\varepsilon_i = 1, X_i > 0$  のとき、  $\varepsilon_i = -1, X_i < 0$  のとき、とおくと、  $\varepsilon_1, \dots, \varepsilon_n$  は互いに独立、かつ  $|X_i|$  とも無関係に、  $P_r\{\varepsilon_i = 1\} = 1/2$  となるから、  $|X_i|$  が与えられたときの任意の総計量の条件付分布は、これから計算することができる。いま  $|X_{(1)}| < \dots < |X_{(n)}|$  とすると、

$$P_r\{X_{\max} = |X_{(i)}|\} = 2^{-n+i-1} \quad i=1, \dots, n$$

$$P_r\{X_{\max} = -|X_{(1)}|\} = 2^{-n}$$

となるから、  $X_{\max}$  の条件付期待値は

$$\begin{aligned} E\{X_{\max} \mid |X_{(1)}|, \dots, |X_{(n)}|\} \\ = \sum_{i=2}^n 2^{-n+i-1} |X_{(i)}| \end{aligned}$$

となる。  $X_{\min}$  の条件付分布も同様に求められるから、結局

$$(8.2) \quad E\left\{\frac{R}{S'} \mid |X_{(1)}|, \dots, |X_{(n)}|\right\} = \frac{1}{S'} \sum_{i=2}^n 2^{-n+i} |X_{(i)}|$$

を得る。そこで正規乱数から (8.2) の値を計算し、その平均を求めれば、  $R/S'$  の期待値の推定値が得られる。

$R$  の確率分布は次のように求められる。  $a$  を定数とすると、

$$P_r\{R < a\} = P_r\left\{\frac{R}{S'} < \frac{a}{S'}\right\} = P_r\left\{S' < \frac{aS'}{R}\right\}$$

ところで  $S'/R$  と  $S'$  とは独立だから、  $S'$  の分布関数を  $\phi(s) = P_r\{S' < s\}$  とすると、

$$(8.3) \quad P_r\left\{S' < \frac{aS'}{R}\right\} = E\left[P_r\left\{S' < \frac{aS'}{R} \mid \frac{S'}{R}\right\}\right] = E\left\{\phi\left(\frac{aS'}{R}\right)\right\}$$

となる。ところで  $S'$  の分布はカイ二乗分布から直接導かれるから、 $\phi$  の値は直接計算することができる。そこで正規乱数にもとづく標本から  $I$  回ごとに  $\phi(aS'/R)$  を求め、その平均を計算すれば (8.3) の推定値が得られる。このような方法によれば  $R$  の分布を直接標本から求める、すなわち  $R < a$  となる回数を数える場合にくらべると、非常に精度が高くなる。

同じような命題として、次のことが成り立つ。

2.°  $X_1, \dots, X_n$  が互いに独立に指数分布に従うとき、 $T$  が尺度変換に関して不変ならば、 $T$  は  $\bar{X}$  と独立であり、また  $T$  が位置変換に関して不変ならば、 $T$  は  $\min X_i$  と独立である。

あるいは、

3.°  $X_1, \dots, X_n$  がガンマ分布 (密度関数  $f(x) = cx^{a-1}e^{-x}$ ) に従うとき、 $T$  が尺度不変な統計量ならば、 $T$  は  $\bar{X}$  と独立であり、またもし

$$t(X_1^c, \dots, X_n^c) \equiv t(X_1, \dots, X_n)$$

となるならば、 $T$  は  $\pi X_i$  と独立になる。

このような性質をうまく利用すると、次にのべるように、きわめて高い効率を上げることができる。

## 9. 一つの例

巧妙な方法によって効率を著しく高めることに成功した例として、Hodges による Hodges-Lehmann 推定量の分散、および分布の推定 (5th Berkeley Symposium) の例をあげよう。

$X_1, \dots, X_n$  が互いに独立に未知母数  $\theta$  を中心とする対称な分布に従うとき、 $\theta$  の推定量として Hodges および Lehmann によって提案されたのが

$$\hat{\theta}_H = (X_i + X_j) / 2 \text{ という形の } n(n+1)/2 \text{ 個の値の中央値}$$

である。この推定量は  $\theta$  に関する符号つき順位と検定から導かれるものであるが、 $n$  が大きいとき漸近的には、 $X$  の分布がロジスティック分布の場合には有効推定量になり、また正規分布の下ではほぼ 95% の効率を持つ。すなわち最もよい推定量  $\bar{X}$  とくらべて

$$\frac{V(\bar{X})}{V(\hat{\theta}_H)} \doteq 0.95$$

となることが知られている。この推定量は  $X$  の分布が正規分布よりややスソの長い分布の下でよい推定量である。

そこで有限の  $h$  に対して  $\hat{\theta}_H$  の分散を求めることが望まれる。Hodges は、正規分布の仮定の下で  $n=18$  の場合について計算した。

$\theta=0$  としてよいから、問題は  $E(\hat{\theta}_H^2)$  を求めることである。ところが先にのべた 1° によって  $\hat{\theta}_H - \bar{X}$  と  $\bar{X}$  とは独立になるから次の式を得る。

$$E(\hat{\theta}_H^2) = E(\hat{\theta}_H - \bar{X})^2 + E(\bar{X}^2) = E(\hat{\theta}_H - \bar{X})^2 + \frac{\sigma^2}{n}$$

そこで  $\Delta = \hat{\theta}_H - \bar{X}$  とおいて、 $\Delta$  の値を正規乱数を用いた標本からくり返し計算して、 $E(\hat{\theta}_H^2)$  を  $\sigma^2/n + (1/N)\sum \Delta^2$

という形で求めることができる。そうするとこのような推定量の分散は、

$$(1/N)V(D^2) = 1/N \{E(D^4) - (E(D^2))^2\}$$

となる。いま  $\sigma^2 = 1$  とする。  $\hat{\theta}_H$  の分布はかなり早く漸近分布に近づくことが知られているから

$$E(\hat{\theta}_H^2) \doteq \frac{1}{0.95} \frac{1}{n} \doteq 1.05/n$$

となるであろう。したがって

$$E(D^2) \doteq 1/20n$$

また  $\hat{\theta}_H - \bar{X}$  の分布は、ほぼ正規分布になると見なすことができるから

$$V(D^2) \doteq 2(E(D^2))^2$$

したがって  $E(\hat{\theta}_H^2)$  の推定量の標準偏差は、ほぼ

$$\sqrt{2/NE(D^2)} \doteq \frac{1}{20n} \sqrt{2/N}$$

となるであろう。したがって変動係数はだいたい

$$\sqrt{2}/21\sqrt{N}$$

に等しくなる。そこでたとえば  $N=100$  とすると、変動係数はだいたい 0.7% となり、95% 信頼区間を求めても、相対誤差が  $\pm 1.4\%$  にとどまる。したがって 200~300 回のくり返し程度で十分信頼性のある結果が得られることになる。Hodges はさらに層別抽出を適用して、 $N=100$  で相対誤差が 1% より小さい信頼区間を求めた。

ここでもし  $\hat{\theta}_H^2$  の平均から  $E(\hat{\theta}_H^2)$  を求めたら、そのときの推定量の標準偏差は

$$\sqrt{1/N} \times \sqrt{V(\hat{\theta}_H^2)} \doteq \sqrt{2/NE(\hat{\theta}_H^2)}$$

となるから、変動係数は  $\sqrt{2/N}$ 、すなわち上にのべた方法による場合の 21 倍にもなる。したがって同じ精度を得るために、くり返し数  $N$  は  $21^2 \doteq 441$  倍にしなければならないであろう。

次に  $\hat{\theta}_H$  の確率分布を求める。  $\hat{\theta}_H$  の分布が原点に関して対称になることは明らかだから、定数  $C > 0$  に対して  $P_r\{|\hat{\theta}_H| < C\}$  を求める。

$$\begin{aligned} P_r\{|\hat{\theta}_H| < C\} &= P_r\{-C < \bar{X} + D < C\} = P_r\{-C - D < \bar{X} < C - D\} \\ &= E[P_r\{-C - D < \bar{X} < C - D \mid D\}] \end{aligned}$$

ここで  $\bar{X}$  が  $D$  とは独立に平均 0、分散  $1/N$  の正規分布に従うことを用いれば

$$(9.1) \quad P_r\{-C - D < \bar{X} < C - D \mid D\} = \Phi\{\sqrt{n}(c - D)\} - \Phi\{-\sqrt{n}(c + D)\}$$

となる。ただし

$$\Phi(t) = \int_{-\infty}^{\infty} (1/\sqrt{2\pi}) \exp(-U^2/2) dU$$

である。そこで  $D$  に対応する (9.1) の値をくり返し計算して、その平均をとれば、求める確率を推定することができる。

ここで  $\sqrt{n}\Delta$  はあまり大きくないと考えられるから、(9.1) はほぼ次のように表わされる。

$$\Phi\{\sqrt{n}(c - \Delta)\} - \Phi\{-\sqrt{n}(c + \Delta)\} \doteq \Phi\{\sqrt{nc}\} - \Phi\{-\sqrt{nc}\} - \Phi'(\sqrt{nc})\sqrt{n}\Delta + \Phi'(-\sqrt{nc})\sqrt{n}\Delta$$

$$\begin{aligned}
 & + \frac{1}{2} \phi''(\sqrt{nc})n\Delta^2 - \frac{1}{2} \phi''(-\sqrt{nc})n\Delta^2 \\
 & = 2\phi(\sqrt{nc}) - 1 - \sqrt{nc}\phi'(\sqrt{nc})n\Delta^2
 \end{aligned}$$

ただし、 $\phi(U) = \phi'(U) = (2\pi)^{-1/2} e^{-U^2/2}$  である。

$\hat{\theta}_H$  がほぼ分散  $1/n$  の正規分布に従うことを考えて、たとえば  $\sqrt{nc} = 2$  とすれば、 $\sqrt{nc}\phi(\sqrt{nc}) = 0.108$  となるから、ほぼ

$$\begin{aligned}
 P_r\{|\hat{\theta}_H| < 2/\sqrt{n}\} & \doteq 0.955 - 0.108E\{n\Delta^2\} \\
 & \doteq 0.950
 \end{aligned}$$

となり、また推定量の分散はだいたい

$$(1/N)(0.108)^2 \times V(n\Delta^2) \doteq 1.16/20000N$$

となるであろう。

ここでもし  $|\hat{\theta}_H|$  が  $2/\sqrt{n}$  を越える回数を数えて、 $P_r\{|\hat{\theta}_H| < 2/\sqrt{n}\}$  を推定するとすれば、その推定量の分散はほぼ

$$(0.95 \times 0.05)/N \doteq 1/20N$$

となるから、上記のような方法によって、同じ精度を達成するのに必要な  $N$  の比率で計算した効率率は 1000 倍近くにもなることがわかる。Hodges は層別抽出との組み合わせによってもっと高い効率を得ている。

## 10. む す び

確率分布の特性値を求めるためのモンテカルロシミュレーションにおいては、以上にのべたような方法をうまく組み合わせることにより、著しく効率を高めることができる場合が少なくない。

もちろん効率というものは、単に一定の精度を持つ推定値を得るために必要なくり返し数  $N$  の大きさによってのみはかられるわけではない。計算時間は 1 回のくり返しにおける計算時間の長さ、プログラムの複雑さ等にも依存するし、また人間の時間と機械の時間との割り振りも考えねばならない。計算機の使用時間が十分あれば、効率のあまりよくない方法を用いても、人間が考える時間を節約したほうが有効かもしれない。

しかしながら、計算にあたって精度と効率を最初からはっきり考慮に入れておくこと、そうしてもし簡単な方法では十分な精度が得られないことがわかったならば、効率を高める工夫をすること、それだけは必要であり、そのためにはここにのべたようなことが参考になるであろう。

## 参 照 文 献

- Hammersley, J. M. & D. C. Handscomb, *Monte Carlo Methods*, Methuen, 1964.  
 Hodges, J. L., "Efficiencies in normal samples and tolerances of extreme values for some estimates of location," *Proceedings of 5th Berkeley Symposium on Probability and Statistics*, vol. 1, pp. 163–183, 1967.