日本オペレーションズ・リサーチ学会
2019年 秋季研究発表会

経験コピュラによる乱数と準乱数の生成について

01604870　政策研究大学院大学　*諸星穂積　MOROHOSI Hozumi

1. はじめに

確率変数間に従属性があるような場合のシミュレーションでは，コピュラを利用することが一般的である．パラメトリックなモデルを利用することが多いが，ここではノンパラメトリックな方法によりサンプルを生成する方法について，数値的な考察を行った結果を報告する．

コピュラ C は $[0,1]^d$ の確率分布で，d 次元の確率分布 F をその周辺分布 F_i に $F(x_1,\dots,x_d) = C(F_1(x_1),\dots,F_d(x_d))$ のように分解するのに利用される．本発表では，周辺分布の方には立ち入らず，$[0,1]^d$ 上のサンプルの生成について論じる．

普通のシミュレーションでは（疑似）乱数を使うが，準乱数を利用すると，高次元で与えられた分布に対してより忠実なサンプルを発生させることができるので，期待値の計算が高速にできる可能性がある．準乱数をコピュラに用いることは，[1] で試みられている．そこでは，パラメトリックなモデルに対して，準乱数が高性能を示すことが報告されている．一方でノンパラメトリックなコピュラは，[2] などで研究が進められていて，いくつかモデルが提案されている．本発表では，これらのノンパラメトリックなモデルに対して準乱数を適用してみる．

2. 経験コピュラ関数

サンプル $\boldsymbol{X}_i = (X_{i1},\dots,X_{id})$, $i=1,\dots,n$, が与えられたとして，各座標成分内での X_{ij} の順位を R_{ij} とする，つまり X_{ij} は $\{X_{1j},\dots,X_{nj}\}$ の中で小さいほうから数えて R_{ij} 番目にある．この順位を利用して，経験コピュラは次式で定義される．

$$C_n(u_1,\dots,u_d) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\mathbf{1}\{R_{ij}/n \le u_j\}. \quad (1)$$

これは階段関数であるが，多項式を使ってなめらかな関数を構成するのが経験ベータコピュラである．$[0,1]$ 上の n 個の一様乱数 $U_1,\dots,U_n$ の r 番目の順序統計量 $U_{(r)}$ の分布は，

$$F_{n,r}(u) = \sum_{k=r}^{n}\binom{n}{k}u^k(1-u)^{n-k} \quad (2)$$

で与えられる．これを利用して経験ベータコピュラは以下の式で定義される．

$$\begin{aligned} C_n^{\beta}(u_1,\dots,u_d) &= \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}F_{n,R_{ij}}(u_j) \\ &= \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\sum_{k=1}^{n}\mathbf{1}\{R_{ij}\le k\}p_{nk}(u_j) \end{aligned} \quad (3)$$

ただし，$p_{nk}(u) = \binom{n}{k}u^k(1-u)^{n-k}$ とおいた．

経験ベータコピュラは高い次数（サンプルサイズと同じ大きさ）の多項式になってしまうが，これをもっと低い次数の多項式で置き換えて，近似をしたものが経験 Bernstein コピュラである．多項式の次数を決めるパラメータ $\boldsymbol{m} = (m_1,\dots,m_d)$ を，通常は $m_j \mid n$ となるように決めて，次のように定義する．

$$\begin{aligned} C_{\boldsymbol{m}}^{B}(u_1,\dots,u_d) &= \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}F_{m_j,\lceil R_{ij}m_j/n\rceil}(u_j) \\ &= \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\sum_{k_j=1}^{m_j}\mathbf{1}\{R_{ij}/n \le k_j/m_j\}p_{m_jk_j}(u_j). \end{aligned} \quad (4)$$

今回の実験では，m_j はすべて同じ値にした．

3. コピュラに従う乱数の発生

多次元の確率分布 $C(u_1,\dots,u_d)$ が与えられたとき，この分布に従う乱数を発生させるのには，条件付き分布の逆関数を使って 1 次元づつ逐次的に発生させるのが，一番一般的な方法だろう．まず，一様乱数 $U_1,\dots,U_d$ を用意しておいて，C の第 1 成分の周辺分布 C_1 に従う乱数 X_1 を $X_1 = C_1^{-1}(U_1)$ として発生させる（コピュラの場合は，C_1 は一様分布なので，$X_1 = U_1$ になる）．つぎに第 1 成分が

与えられたときの第 2 成分の条件付き周辺分布を $C_{2|1}$ として，$C_{2|1}$ に従う乱数 X_2 を $X_2 = C_{2|1}^{-1}(U_2)$ と発生させる．以下順次 $C_{j|1\ldots j-1}$ を $1, \ldots, j-1$ 成分が与えられたときの j 成分の条件付き周辺分布として，これから $X_j = C_{j|1\ldots j-1}^{-1}(U_j)$ を発生させる．このようにして生成した $(X_1, \ldots, X_d)$ は C に従うことが示されている．分布 C が微分可能であれば，条件付き周辺分布は以下の式により求めることができる．

$$C_{j|1\ldots j-1} = \frac{\partial_{j-1\ldots 1} C_{1\ldots j}}{\partial_{j-1\ldots 1} C_{1\ldots j-1}},$$

ただし，$\partial_{j-1\ldots 1} C = \partial^{j-1} C / \partial u_1 \ldots \partial u_{j-1}$ であり，$C_{1\ldots j}$ は C の第 1 から j 成分への周辺分布である．

4.　数値実験

最初に [2] にならって，経験コピュラのバイアスを計算した．バイアスを測る尺度として平均 2 乗誤差を利用する．もとになるコピュラ C を決めて，乱数と準乱数で n 点からなるサンプルを発生させ，それぞれのサンプルから経験ベータまたは Bernstein コピュラ C_n^* をつくり，乱数と準乱数の違いによるコピュラのバイアスを求めるため以下の値を計算した．

$$\int_{[0,1]^d} \left[\mathrm{E} \left\{ C_n^*(\boldsymbol{u}) - C(\boldsymbol{u}) \right\}^2 \right] d\boldsymbol{u} \tag{5}$$

準乱数には Sobol' 列を使い，サンプルサイズを変えてベータと Bernstein で，乱数と比較計算した結果が図 1 である．もとのコピュラは 5 次元の t-コピュラである．

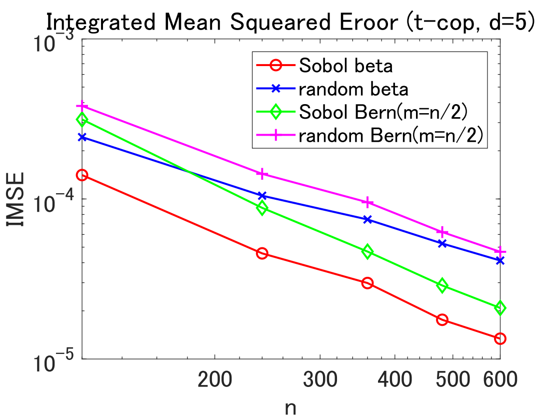

図 1: 経験コピュラのバイアス．

ベータと Bernstein の両方で，準乱数のほうが乱数より小さいバイアスをもつことが示されている．

次に，一般的な数値計算としてバスケットオプションの価格計算を行った．こちらの実験では，最初に n 点のサンプルを t-コピュラに従って発生させ，それから作った経験ベータコピュラに従う乱数と準乱数を N 点発生させる．それぞれを対数正規分布に変換して，成分の平均をとり，バスケットコールオプションの利得関数を計算する．興味があるのは，計算値の標準偏差がどのくらいの速度で収束するかである．準乱数には Sobol' 列と Halton 列を使った．図 2 が結果である．

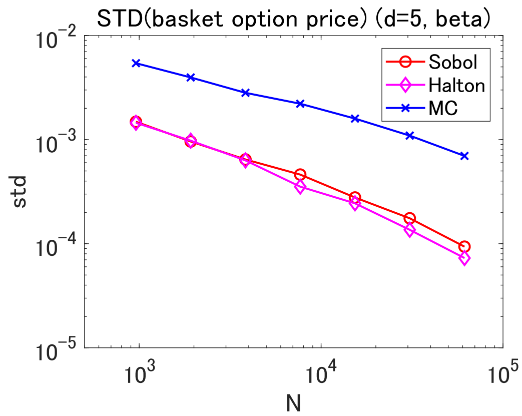

図 2: バスケットオプション価格の標準偏差の変化．

一般的な準乱数のシミュレーションと同様に，高速な収束が確認された．

5.　まとめ

ノンパラメトリックなコピュラに準乱数のシミュレーションを利用した結果を紹介した．今後はもう少しサンプルサイズを大きくすること，次元を上げることなどが課題である．

参考文献

[1] M. Cambou, M. Hofert, and C. Lemieux. Quasi-random numbers for copula models, *Statistics and Computing*, 27(2017), 1307–1329.

[2] J. Segers, M. Sibuya, and H. Tsukahara. The empirical beta copula. *Journal of Multivariate Analysis*, 155(2017), 35–51.