日本オペレーションズ・リサーチ学会
2021年 春季研究発表会

経験ベータコピュラによる乱数の生成と応用

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

1. はじめに

コピュラは，確率変数間の従属性を記述する一般的な方法である．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乱数を生成する方法を提案し，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 $F(x_1,\ldots,x_d) = C(F_1(x_1),\ldots,F_d(x_d))$ 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．

2. 経験コピュラ関数

サンプル $\boldsymbol{X}_i = (X_{i1},\ldots,X_{id})$, $i=1,\ldots,n$, 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 $\{X_{1j},\ldots,X_{nj}\}$ 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．

$$C_n(x_1,\ldots,x_d) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\mathbf{1}\{R_{ij}/n \le x_j\}. \quad (1)$$

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 [3]．$[0,1]$ 上の n 個の一様乱数に対する，r 番目の順序統計量の分布は，

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

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．

$$C_n^{\beta}(x_1,\ldots,x_d) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}F_{n,R_{ij}}(x_j). \quad (3)$$

式 (3) を見ると，経験ベータコピュラ分布関数 C_n^{β} は，分布関数 $\prod_{j=1}^{d}F_{n,R_{ij}}(x_j)$, $i=1,\ldots,n$, を等重み $1/n$ で混合した分布と見なせる．この観察に基づいて，C_n^{β} に従う乱数を発生する算法を以下に提案する．

3. 混合分布としての乱数の発生

混合分布 $f(x) = \sum_{i=1}^{n}w_i f_i(x)$ に従う乱数の発生は簡単で，確率 w_i で分布 f_i を選び，選ばれた f_i に従う乱数を発生させればよい．経験ベータコピュラの場合は $w_i = 1/n$ なので，等確率で $\prod_{j=1}^{d}F_{nR_{ij}}(x_j)$, $i=1,\ldots,n$, から 1 つ分布を選び，これに従う乱数を発生させる．この関数は具合の良いことに積の形をしているので，次元ごとに逆関数法を使うことができる．すなわち，まず $[0,1]^d$ 上の一様乱数 $(U_1,\ldots,U_d)$ を発生させ，$X_j = F_{n,R_{ij}}^{-1}(U_j)$, $j=1,\ldots,d$, とする．算法をまとめると以下のようになる．

経験ベータコピュラに従う乱数の発生

入力をサンプル $(X_{i1},\ldots,X_{id})$ とし，その各成分内での順位 $(R_{i1},\ldots,R_{id})$, $i=1,\ldots,n$ とする．サンプルに基づいて (3) の経験ベータコピュラ C_n^{β} を求めておく．

1. $d+1$ 組の一様乱数 $(U_1,\ldots,U_{d+1})$ を発生させる．
2. $I = \lceil nU_{d+1}\rceil$ として，I 番目の分布 $\prod_{j=1}^{d}F_{n,R_{Ij}}(x_j)$ を選ぶ．
3. $X_j = F_{n,R_{Ij}}^{-1}(U_j)$, $j=1,\ldots,d$, とする．

上記の算法で，一様乱数の代わりに準乱数を用いることも考えられる．[1] では，パラメトリックなコピュラの場合に，準乱数の使用による収束の向上が報告されている．

4. 数値実験

経験コピュラの応用として，t-コピュラを利用したパラメータのブートストラップ推定の実験を以下のような手順で行った．

1. パラメータ ν と ρ を決め，d 次元の t-コピュラに従う乱数 $x_1,\ldots,x_n$ を発生させる．
2. 発生したサンプルから経験ベータコピュラ C_n^{β} を求める．
3. C_n^{β} を使ってサンプル $y_1,\ldots,y_n$ を発生させ，これを用いて t-コピュラのパラメータ ν の推定値 $\hat{\nu}$ を求める．これを繰り返して，パラメータの推定値 $\hat{\nu}$ を複数求める．

4. 求められた複数の推定値 $\hat{\nu}$ から，ν の区間推定を行う．

比較検証のために，パラメトリックブートストラップ法（cf. [2]）による推定を行った．これは，上記の手順のうち 2, 3 で，C_n^β の代わりに，サンプルから推定した $\hat{\nu}$ と $\hat{\rho}$ を使った t-コピュラからサンプル $y_1, \ldots, y_n$ を発生させて，ν の区間推定を行うものである．実験は $\nu = 4$, $\rho = 0.7$, $d = 5$ として，n の値を変えて行った．結果を表 1 にまとめる．

表 1: パラメトリックブートストラップ (p.) と経験ベータコピュラ (β) によるパラメータ推定実験．95%信頼区間の下限 (low) と上限 (up)．

n	p. low	p. up	β low	β up
120	2.71	9.86	3.49	32.53
240	2.81	6.84	3.66	11.90
360	2.91	5.87	3.39	8.51
480	3.05	5.35	3.41	6.78
600	3.48	5.27	3.44	6.56

実験結果をみると，どちらの方法でも推定された区間に真の値 $\nu = 4$ が含まれるので，区間推定には成功しているようである．パラメトリックブートストラップ法のほうが信頼区間の幅が小さいが，これは一般的な現象であろう．サンプルサイズが大きくなると，両者の差は縮まってくる．

表 2: 準乱数によるパラメトリックブートストラップ (p.) と経験ベータコピュラ (β) によるパラメータ推定実験．95%信頼区間．

n	p. low	p. up	β low	β up
120	2.76	7.61	3.37	22.78
240	3.20	7.02	3.71	10.03
360	3.28	5.21	3.61	7.64
480	3.25	5.72	3.81	6.87
600	3.48	4.72	3.69	6.18

同じ実験を準乱数を使って行った結果が表 2 である．こちらの場合も区間推定には成功しているようである．また，区間幅は乱数を使った場合より若干小さくなっている．これを確認するためグラフ化したものが図 1 である．パラメトリックのほうが区間幅が小さく，乱数を使った場合 (MC) より準乱数を使った場合 (QMC) のほうが区間幅が小さいことが見て取れる．

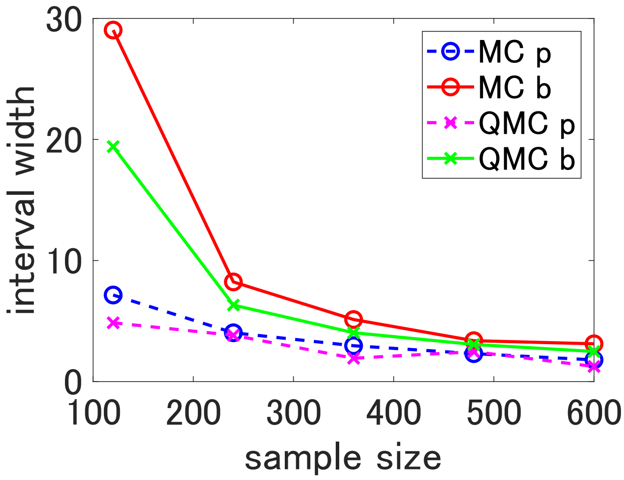

図 1: 信頼区間幅の比較．

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，もう少しほかのコピュラで試してみること，より高い次元で実験してみること，などが課題である．．また，ベータコピュラ生成には逆関数法を使っているので，かなり計算時間がかかる．計算の効率化も課題である．

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] M. Hofert, I. Kojadinovic, M. Mächler, and J. Yan. *Elements of Copula Modeling with R*, Springer, 2018.

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