

アプリケーション搭載用 Dirichlet 乱数発生プログラムモジュールの試作とその性能評価について

井上弘樹

新潟医療福祉大学 医療情報管理学科

【背景・目的】Dirichlet 分布は多項モデルのベイズ推定で3つ以上の状態・段階の多項比率の解析を行う場合¹⁾の事前分布として適しており、事後確率の計算に当たって多数のDirichlet 乱数を発生し、シミュレーション・可視化を行う方法が取られている²⁾。多項モデルを自己開発アプリケーション上で実装・再現する場合、Dirichlet 乱数を発生する関数もしくはライブラリ、パッケージなどが必要になってくるが、当該プログラミング言語でそれらがメーカーにより公式にサポートされなかった場合、何らかの同乱数を発生させる仕組みが必要となる。そこでメーカーによるガンマ乱数発生関数がサポートされていないプログラミング言語について、同乱数を一樣乱数から発生させるコードを新たに記述し、その性能評価を試みた。

【方法】マイクロソフト社製の統合開発環境である Visual Studio 内の、Visual Basic.NET を使用した。要素数 k 個の Dirichlet 分布に従う乱数のベクトル列 $X_i (i=1, \dots, k)$ は、 k 個の標準ガンマ乱数 Y_i (母数 $\gamma_i (i=1, \dots, k)$) とすると、 $X_i = Y_i / (Y_1 + \dots + Y_k)$ で算出される³⁾。 Y_i の発生は Best、Cheng の方法を用いた⁴⁾⁶⁾。今回作成されたプログラムは要素数3のDirichlet 分布に従う乱数のベクトル列を発生するものとし、ウィンドウズ PC のデスクトップ上で表示され、 γ_i の値、求める乱数列の項数を入力により設定できる様にした(図1)。プログラム全体のボリュームはおよそ20キロバイトであった。プログラムの評価指標は乱数列の作成に所要した時間を用いることとし、対象には R のパッケージ gtools⁷⁾ 上の Dirichlet 乱数発生関数 rdirichlet0 (以下 R の関数) を用いた。同一条件を期するため、同一の PC 端末上で本プログラムと R を動作させ、所要時間を比較することとした。

【結果】計算結果の例を示す。 $\gamma_1=0.2$ 、 $\gamma_2=0.3$ 、 $\gamma_3=0.5$ 、乱数 50,000 個で乱数を発生させた場合、本ソフトの発生時間は、3.355 秒、R の関数の発生時間は 3.332 秒であった。 $\gamma_1=2$ 、 $\gamma_2=3$ 、 $\gamma_3=4$ 、乱数 100,000 個の設定で乱数を発生させた場合、本ソフトの発生時間は、6.826 秒、R の関数の発生時間は 6.2614 秒であった。

【考察】当初 R の関数の方が高速に乱数ベクトル列を発生できることが予想された。しかし、実際の乱数発生時間を計測したところ、本プログラムモジュールはやや劣るものの R の関数とそれほど変わらない発生時間となった。



図 1

ちなみに乱数列の項数を 50 万、100 万と増加させると、本プログラムモジュールと R の関数はそれぞれ 30 秒と 30 秒、69 秒と 60 秒となり、著しい差は見られなかった。今回使用したアルゴリズムではガンマ乱数を 3 種類発生させることから発生に時間がかかるものと考えられるが、R の関数も本プログラムモジュールと同様のガンマ乱数を発生させるアルゴリズムなのかもしれない。また、あまり項数の多い乱数列(ベクトル数が数 100 万個レベル)を発生させようとするとハングアップを生じる。バグフィックスやコードの見直しの他に、今回使用した言語がどの程度数値計算に適しているのかを検討し、その動作限界の範囲内で使用することを考慮する必要があるのかもしれない。

【結論】事後確率の計算に当たって多数の Dirichlet 乱数を発生する機能を実装したアプリケーションを開発するに当たり、必要となる Dirichlet 乱数発生のためのプログラムモジュールを試作した。

【文献】

- 1) McLean, N: Ecology and management of overabundant koala (*Phascolarctos cinereus*) populations, PhD thesis, Department of Zoology, The University of Melbourne, 229-238, 2003.
- 2) McCarthy, 野間口真太郎: 生態学のためのベイズ法, 共立出版, 初版, 93-97, 東京, 2009.
- 3) 蓑谷千風彦: 統計分布ハンドブック, 朝倉書店, 初版, 469-472, 東京, 2003.
- 4) Best, DJ: A note on gamma variate generators with shape parameter less than unity, *Computing*, 30: 185-188, 1983.
- 5) Cheng, RCH: The generation of gamma variables with non-integral shape parameter, *Applied Statistics*, 26: 71-75, 1977.
- 6) 井上弘樹: アプリケーション搭載用ガンマ乱数発生プログラムモジュールの試作とその性能評価について, *新潟医療福祉学会誌*, 17: 73, 2017.
- 7) gtools v2.2.2, <https://www.rdocumentation.org/packages/gtools/versions/2.2.2>, 2018 年 8 月 26 日.