幾つかの指数型分布モデルにおける
統計量の漸近近似の良さのシミュレーション

2009SE297　脇田雅樹

指導教員：白石高章

1　はじめに

代表的な指数型分布として，二項分布，ポアソン分布，指数分布について考察する．1標本のデータを解析するために，漸近的に正規分布に従う統計量が使われる．その場合，中心極限定理をストレートに当てはめられる統計量と分散安定化変換による統計量の正規分布への近似の良さを，モンテカルロシミュレーションにより比較する．

2　擬似乱数生成法

擬似乱数生成法は様々な手法が知られている．
その中の一部を紹介する．
・線形合同法
・線形帰還シフトレジスタ
・メルセンヌ・ツイスタ
・セルオートマトン法
・xorshift

本研究では，メルセンヌ・ツイスタを用いて乱数を生成させる. メルセンヌ・ツイスタは既存の乱数生成アルゴリズムの欠点を改良して，高品質の乱数を高速に生成するように設計されている．

3　指数型分布に従う乱数の生成

[3] のメルセンヌ・ツイスタを用いて生成した一様乱数から分布に従う乱数を生成する．

3.1　ベルヌーイ乱数

二項分布 $B(1,p)$ に従う乱数の生成を行うための変換を考える．

U を $U(0,1)$ に従う確率変数とし,$0<p<1$ となる定数 p に対して，$U \le p$ ならば $X=1, U>p$ ならば $X>0$ で確率変数 X を定義すれば，すなわち，$X \equiv I_{(0,p]}(U)$ とおくならば，確率変数 X は二項分布 $B(1,p)$（ベルヌーイ試行）に従う．

3.2　指数乱数

平均 μ の密度関数 $f(x) = \frac{1}{\mu}e^{-\frac{x}{\mu}}$ をもつ指数分布 $EX(1/\mu)$ に従う乱数の生成を行うための変換を考える．
この指数分布の分布関数 $F(x)$ は，

$$F(x) = \int_0^x \frac{1}{\mu}e^{-\frac{x}{\mu}} = [-e^{-\frac{x}{\mu}}]_0^x = 1-e^{-\frac{x}{\mu}} \qquad (1)$$

である．

$F(x)$ は指数分布 $EX(1/\mu)$ の分布関数なので，$F(-\mu\log(1-u)) = u$ より，$F^{-1}(u) = -\mu\log(1-u)$ ゆえに U は $U(0,1)$ に従う確率変数とし，$Y \equiv -\mu\log(1-U)$ とおけば，$-\mu\log(1-U)$ は $EX(1/\mu)$ に従う. さらに，$1-U$ と U は同じ分布であるので，$Y \equiv -\mu\log(U)$ の U に一様乱数を入れることで平均 μ の指数乱数が得られる．
参考文献 [1] の命題 5.2(p.158 ∼ 160) を使うと

$$\begin{aligned} P(Y \le x) &= P(-\mu\log(1-U) \le x) \\ &= F(\frac{x}{\mu}) \\ &= 1-e^{-\frac{x}{\mu}} \end{aligned} \qquad (2)$$

となり，(1) より Y は $EX(1/\mu)$ に従う確率変数となる．

3.3　ポアソン乱数

平均 μ のポアソン分布の確率関数は，

$$f(x|\mu) = \frac{\mu^x}{x!}e^{-\mu},\ \ x=0,1,2,\cdots,\ \ \mu>0 \qquad (3)$$

である．(3) 式のポアソン分布 $P_0(\mu)$ において

$$f(0|\mu) = e^{-\mu}$$
$$F(x|\mu) = \sum_{i=0}^{x} f(i|\mu)$$

一様乱数 U を生成し

$$\begin{cases} (I) \quad U \le f(0|\mu) = e^{-\mu} \Rightarrow X = 0 \\ (II) \quad F(x-1|\mu) < U \le F(x|\mu),\ \ x \ge 1 \\ \qquad\quad \text{となる } x \text{ に対して、} X = x \end{cases}$$

とすると，$X \sim P_0(\mu)$ となる．

4　シミュレーションによる近似の比較

ベルヌーイ乱数, 指数乱数, ポアソン乱数から中心極限定理により, 標準正規分布に収束することを確かめる．ただし，$z(\alpha)$ は標準正規分布の上側 $100\alpha\%$ 点とする．
$\alpha = 0.05$，標本サイズを 100, 繰り返し回数を $10,000$ とした．

4.1　ベルヌーイ乱数

$\hat{p}_1 \equiv X/n$, $\hat{p}_2 \equiv (X+0.5)/(n+1)$ とする．
中心極限定理とスラツキーの定理より，$\hat{p} = \hat{p}_1$ or $\hat{p}_2$ に対して，n→∞ として，

$$\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \qquad (4)$$

である．参考文献 [1] の (7.25)，(7.27) 式より，分散安定化変換による統計量は，

$$2\sqrt{n}\{\arcsin(\sqrt{\hat{p}}) - \arcsin(\sqrt{p})\} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \qquad (5)$$

となる．ここで，

$$Z_{1n} = 2\sqrt{n}\{\arcsin(\sqrt{\hat{p}_1}) - \arcsin(\sqrt{p})\}$$
$$Z_{2n} = 2\sqrt{n}\{\arcsin(\sqrt{\hat{p}_2}) - \arcsin(\sqrt{p})\}$$
$$Z_{3n} = \frac{\sqrt{n}(\hat{p}_1-p)}{\sqrt{\hat{p}_1(1-\hat{p}_1)}}, \qquad Z_{4n} = \frac{\sqrt{n}(\hat{p}_2-p)}{\sqrt{\hat{p}_2(1-\hat{p}_2)}}$$
とする．

ベルヌーイ乱数によるシミュレーションにより，$P(Z_{in} > z(\alpha))$ $(1 \leqq i \leqq 4)$ の値の表を表 1 に与えた．

表 1 シミュレーションによる $P(Z_{in} > z(\alpha))$ $(1 \leqq i \leqq 4)$ の値の表

p	Z1n ratio1	Z2n ratio2	Z3n ratio3	Z4n ratio4
0.1	0.0370	0.0370	0.0689	0.0689
0.2	0.0576	0.0576	0.0576	0.0576
0.3	0.0548	0.0548	0.0548	0.0548
0.4	0.0444	0.0444	0.0444	0.0444
0.5	0.0496	0.0496	0.0496	0.0496
0.6	0.0620	0.0412	0.0412	0.0412
0.7	0.0499	0.0499	0.0499	0.0499
0.8	0.0473	0.0473	0.0473	0.0473
0.9	0.0602	0.0602	0.0602	0.0249

4.2 指数乱数

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ とする．

中心極限定理による統計量は，
$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\bar{X}_n} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \tag{6}$$
である．分散安定化変換による統計量は，
$$\sqrt{n}\{log(\bar{X}_n) - log(\mu)\} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \tag{7}$$
である．ここで，
$$Z_{5n} = \sqrt{n}\{log(\bar{X}_n) - log(\mu)\}$$
$$Z_{6n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\bar{X}_n}$$
とする．

指数乱数によるシミュレーションにより，$P(Z_{in} > z(\alpha))$ $(5 \leqq i \leqq 6)$ の値の表を表 2 に与えた．

表 2 シミュレーションによる $P(Z_{in} > z(\alpha))$ $(5 \leqq i \leqq 6)$ の値の表

μ	Z5n ratio1	Z6n ratio2
1.0	0.0408	0.0284
2.0	0.0407	0.0288
3.0	0.0442	0.0299
4.0	0.0401	0.0275
5.0	0.0442	0.0307
6.0	0.0431	0.0304
7.0	0.0428	0.0309
8.0	0.0408	0.0288
9.0	0.0415	0.0295
10.0	0.0437	0.0316

4.3 ポアソン乱数

μ の点推定量は $\hat{\mu} = \frac{W}{n}$ で与えられる．

また，σ の推定量は

$\hat{\sigma}_1 \equiv \sqrt{\hat{\mu}}$，$\hat{\sigma}_2 \equiv \frac{1}{2}\{\sqrt{\frac{W+1}{n} + \frac{W}{n}}\}$，$\hat{\sigma}_3 \equiv \sqrt{\frac{W}{n} + \frac{3}{8n}}$

である，$\hat{\sigma} = \hat{\sigma}_1$，$\hat{\sigma}_2$，$\hat{\sigma}_3$ とすると，参考文献 [1] の p.234 より，
$$\frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\hat{\mu}}} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \tag{8}$$
$$2\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{\mathcal{L}} Z \sim N(0,1) \tag{9}$$
ここで，
$$Z_{7n} = 2\sqrt{n}(\hat{\sigma}_1 - \sigma), \qquad Z_{8n} = 2\sqrt{n}(\hat{\sigma}_2 - \sigma),$$
$$Z_{9n} = 2\sqrt{n}(\hat{\sigma}_3 - \sigma), \qquad Z_{10n} = \frac{\sqrt{n}(\hat{\mu} - \mu)}{\sqrt{\hat{\mu}}}$$
とする．

Z_{7n}, Z_{8n}, Z_{9n} は分散安定化変換の統計量である．Z_{10n} は中心極限定理の統計量である．

ポアソン乱数によるシミュレーションにより，$P(Z_{in} > z(\alpha))$ $(7 \leqq i \leqq 10)$ の値の表を表 3 に与えた．

表 3 シミュレーションによる $P(Z_{in} > z(\alpha))$ $(7 \leqq i \leqq 10)$ の値の表

μ	Z7n ratio1	Z8n ratio2	Z9n ratio3	Z10n ratio4
1.0	0.0446	0.0556	0.0556	0.0446
2.0	0.0488	0.0488	0.0488	0.0430
3.0	0.0444	0.0498	0.0498	0.0444
4.0	0.0449	0.0449	0.0449	0.0405
5.0	0.0463	0.0514	0.0463	0.0413
6.0	0.0482	0.0482	0.0482	0.0441
7.0	0.0436	0.0463	0.0463	0.0436
8.0	0.0478	0.0514	0.0514	0.0478
9.0	0.0472	0.0506	0.0506	0.0472
10.0	0.0522	0.0522	0.0522	0.0485

5 おわりに

一様乱数からいくつかの指数分布モデルに従う乱数を生成した．(4),(6),(8) の漸近的正規性が教科書等の文献で通常記述されている．(5),(7),(9) は分散安定化変換に基づく統計量の漸近的正規性を示している．第 4 節のシミュレーション比較により，(4),(6),(8) の収束よりも，分散安定化変換による漸近収束 (5),(7),(9) の方が良いことが解った．参考文献 [2] の表 1 で $P(|Z_{in}| > z(\alpha/2))$ $(1 \leqq i \leqq 4)$ についてのシミュレーションによる値が載せられている．この場合も分散安定化変換による漸近収束 (4) が速いことが参考文献 [2] より分かっている．

参考文献

[1] 白石高章:『統計科学の基礎—データと確率の結びつきがよくわかる数理—』．日本評論社，東京，2012.

[2] 白石高章:『多郡の 2 項モデルとポアソンモデルにおけるすべてのパラメータの多重比較法』. 日本統計学会誌，第 42 巻，第 1 号, 55~90 項，2012.

[3] MersenneTwisterのWebPage http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/mt.html.