2 標本指数分布型モデルにおける統計量の漸近近似の良さ

2010SE070 伊藤圭亮

指導教員：白石高章

1 はじめに

本稿では指数型分布として, 二項分布, 指数分布, ポアソン分布を用いて 2 標本モデルの解析を目的とする. そのため漸近的な正規分布に従う統計量として, 中心極限定理に当てた統計量と分散安定化変換による統計量を正規分布への近似の良さを考察する. 1 標本の場合は, 脇田 [3] の卒業論文で示されている. 本稿はその続きの研究内容となる.

2 乱数発生法

乱数発生法には様々な種類がある. 本稿では Website[5] のメルセンヌ・ツイスタを用いて, 一様乱数を生成する.

3 指数型分布に従う乱数生成

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変換の考え, U を U(0,1) に従う確率変数として, $0 < p < 1$ となる定数 p に対して, U≤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≡ $I_{(0,p]}(U)$ とおくと, 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 $f(x) = (1/\mu)e^{-\frac{x}{\mu}}$ の指数分布 $EX(1/\mu)$ に従う変換を考える. この分布関数 F(x) は,

$$F(x) = \int_0^x \frac{1}{\mu}e^{-\frac{x}{\mu}}dx = 1 - e^{-\frac{x}{\mu}}$$

となる. F(x) は $EX(1/\mu)$ の分布関数より, $F(-\mu\log(1-\mu)) = \mu$ となり, 逆関数は $F^{-1}(\mu) = -\mu\log(1-\mu)$ となる. ここで U を一様分布に従う確率変数とし, $X = -\mu\log(1-U)$ とすれば、X は EX($1/\mu$) に従う. さらに, $1-U$ と U は同分布なので $X = -\mu\log(U)$ の U に一様乱数を入れることで平均μ の指数乱数を得る.

3.3 ポアソン乱数

平均 μ のポアソン分布の確率関数は,

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

である. ここで, 一様乱数を生成し, 系列 $u_0, u_1, \cdots$ からポアソン乱数系列 $y_0, y_1, \cdots$ を

$$y_i \equiv e^{\mu}\prod_{m=0}^{i} u_m \qquad (i = 0,1,2,\cdots)$$

とする. $u_i < 1$ より $y_0 > y_1 > y_2 > \cdots$ となる. ここで, 初めて $y_i \le 1$ となる i に対して, $X = i$ 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, ポアソン乱数を生成し, 正規分布へと収束するか確かめる. ただし, $z(\alpha)$ は正規分布の上側 $100\alpha\%$ 点とし, $\alpha = 0.05$, それぞれの乱数, 標本サイズ n_1, n_2 を 100, 繰り返し回数は 100 万回とした. ポアソン分布は他の分布に比べ収束速度が速いため, 標本サイズ n_1, n_2 を 30 とした.

4.1 ベルヌーイ乱数

標本サイズ n_1, n_2 の確率変数をそれぞれ $X \equiv X_1 + \cdots + X_{n_1}$, Y≡ $Y_1 + \cdots + Y_{n_2}$ と定義し, p_1, p_2 点推定量を

$$\hat{p}_1(1) \equiv X/n_1, \ \ \hat{p}_1(2) \equiv (X+0.5)/(n_1+1)$$
$$\hat{p}_2(1) \equiv Y/n_2, \ \ \hat{p}_2(2) \equiv (Y+0.5)/(n_2+1)$$

とする. ここで $n = n_1 + n_2$ とおき, n_1, $n_2 \to \infty$ として,

$$\frac{n_1}{n_1+n_2} \to \lambda_1, \quad \frac{n_2}{n_1+n_2} \to \lambda_2$$

となり, $0 < \lim_{x\to\infty} n_1/n = \lambda < 1$ と仮定する.
中心極限定理とスラツキーの定理より, $\hat{p}_1 = \hat{p}_1(1)$ or $\hat{p}_1(2)$, $\hat{p}_2 = \hat{p}_2(1)$ or $\hat{p}_2(2)$ に対して, n_1,$n_2 \to \infty$ として,

$$\frac{\hat{p}_1 - \hat{p}_2}{\sigma_n(i)} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \qquad (i = 1,2) \tag{1}$$

となる. ただし,

$$\sigma_n(1) \equiv \sqrt{\frac{1}{n_1}\hat{p}_1(1)\{1-\hat{p}_1(1)\} + \frac{1}{n_2}\hat{p}_2(1)\{1-\hat{p}_2(1)\}}$$
$$\sigma_n(2) \equiv \sqrt{\frac{1}{n_1}\hat{p}_1(2)\{1-\hat{p}_1(2)\} + \frac{1}{n_2}\hat{p}_2(2)\{1-\hat{p}_2(2)\}}$$

とする.
白石 [1] の (7.31), (7.32) 式, スラツキーの定理とデルタ法より, 分散安定化変換による統計量は,
$\hat{p}_1 = \hat{p}_1(1)$ or $\hat{p}_1(2)$, $\hat{p}_2 = \hat{p}_2(1)$ or $\hat{p}_2(2)$ に対して,

$$\frac{2\{\arcsin(\sqrt{\hat{p}_1}) - \arcsin(\sqrt{\hat{p}_2})\}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \tag{2}$$

となる. よって, これらより用いる統計量は,

$$Z_{1n} = \frac{\hat{p}_1(1) - \hat{p}_2(1)}{\sigma_n(1)}, \quad Z_{2n} = \frac{\hat{p}_1(2) - \hat{p}_2(2)}{\sigma_n(2)}$$
$$Z_{3n} = \frac{2\{\arcsin(\sqrt{\hat{p}_1(1)}) - \arcsin(\sqrt{\hat{p}_2(1)})\}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$
$$Z_{4n} = \frac{2\{\arcsin(\sqrt{\hat{p}_1(2)}) - \arcsin(\sqrt{\hat{p}_2(2)})\}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

とする.
ベルヌーイ乱数の解析結果を表 1 に示す.
ただし, $p_1 = p_2 = P$ とする.

表 1　ベルヌーイ乱数 $n_1 = 100, n_2 = 100$

P	Z_{1n}	Z_{2n}	Z_{3n}	Z_{4n}
0.1	0.0537	0.0464	0.0541	0.0472
0.3	0.0527	0.0501	0.0524	0.0485
0.5	0.0458	0.0444	0.0448	0.0443
0.7	0.0527	0.0501	0.0523	0.0485
0.9	0.0531	0.0459	0.0535	0.0467

4.2 指数乱数

標本サイズ n_1, n_2 標本平均をそれぞれ $\hat{\mu}_1 \equiv \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$, $\hat{\mu}_2 \equiv \frac{1}{n_2}\sum_{j=1}^{n_2} Y_i$ とする.

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\mu}_1^2}{n_1} + \frac{\hat{\mu}_2^2}{n_2}}} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \tag{3}$$

$$\frac{\log(\hat{\mu}_1) - \log(\hat{\mu}_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \tag{4}$$

よって, これらより用いる統計量は,
白石 [2](p.8, p.18) より,

$$Z_{5n} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\mu}_1^2}{n_1} + \frac{\hat{\mu}_2^2}{n_2}}}$$

$$Z_{6n} = \frac{\log(\hat{\mu}_1) - \log(\hat{\mu}_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

指数乱数の解析結果を表 2 に示す.
ただし, $\mu_1 = \mu_2 = \mu$ とする.

表 2　指数乱数 $n_1 = 100, n_2 = 100$

μ	Z_{5n}	Z_{6n}
1.0	0.0479	0.0497
3.0	0.0482	0.0498
5.0	0.0489	0.0500
7.0	0.0488	0.0501
9.0	0.0469	0.0492

4.3 ポアソン乱数

標本サイズ n_1, n_2 の確率変数をそれぞれ $W_1 \equiv X_1 + \cdots + X_{n_1}$, $W_2 \equiv Y_1 + \cdots + Y_{n_2}$ と定義し, μ_i の点推定量は $\hat{\mu}_i = W_i/n_i$(i=1,2) で与えられる．また σ の推定量は i=1, 2 に対して,

$$\hat{\sigma}_i(1) \equiv \sqrt{\hat{\mu}_i},\ \hat{\sigma}_i(2) \equiv \frac{1}{2}\left\{\sqrt{\frac{W_i+1}{n_i}} + \sqrt{\frac{W_i}{n_i}}\right\},$$

$$\hat{\sigma}_i(3) \equiv \sqrt{\frac{W_i}{n_i} + \frac{3}{8n_i}}$$

である.
$\hat{\sigma}_1 = \hat{\sigma}_1(1),\ \hat{\sigma}_1(2),\ \hat{\sigma}_1(3),\quad \hat{\sigma}_2 = \hat{\sigma}_2(1),\ \hat{\sigma}_2(2),\ \hat{\sigma}_2(3)$ に対して, $\mu_1 = \mu_2$ の下で

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\mu}_1}{n_1} + \frac{\hat{\mu}_2}{n_2}}} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \tag{5}$$

$$\frac{2(\hat{\sigma}_1 - \hat{\sigma}_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \xrightarrow{\mathcal{L}} Z \sim N(0,1) \tag{6}$$

となる．ここで, これらより用いる統計量は,

$$Z_{7n} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\mu}_1}{n_1} + \frac{\hat{\mu}_2}{n_2}}},\qquad Z_{8n} = \frac{2(\hat{\sigma}_1(1) - \hat{\sigma}_2(1))}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$Z_{9n} = \frac{2(\hat{\sigma}_1(2) - \hat{\sigma}_2(2))}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},\quad Z_{10n} = \frac{2(\hat{\sigma}_1(3) - \hat{\sigma}_2(3))}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

とする.
ポアソン乱数の解析結果を表 3 に示す.
ただし, $\mu_1 = \mu_2 = \mu$ とする.

表 3　ポアソン乱数 $n_1 = 30, n_2 = 30$

μ	Z_{7n}	Z_{8n}	Z_{9n}	Z_{10n}
1.0	0.0496	0.0511	0.0497	0.0498
3.0	0.0497	0.0505	0.0501	0.0501
5.0	0.0496	0.0505	0.0500	0.0501
7.0	0.0494	0.0503	0.0501	0.0501
9.0	0.0499	0.0500	0.0502	0.0503

5 おわりに

一様乱数から指数型分布に従う乱数を生成することが出来た. 求めた統計量を導くシミュレーションを作成し, それぞれの比較より, 分散安定化変換による正規分布への近似の方が良いとわかる. さらに, 3 種類の分布の中でポアソン分布の収束は, 他の分布より速度が速いことがわかった.

参考文献

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

[2] 白石高章（2013）. 多群指数モデルにおける平均パラメータの多重比較法. 計量生物学. 34 1-20.

[3] 脇田雅樹. 『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』. 2012 年度南山大学情報理工学部情報システム数理卒業論文.

[4] Shiraishi, T. (2012). Multiple Compaprison Procedures for Poisson Parameters in Multi-Sample Models. Behaviormetrika. 39 167-182.

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