二項分布，ポアソン分布，指数分布における統計量の正規近似の良さについて

2020SS064 下瀬悠輔

指導教員：白石高章

1　はじめに

本研究の目的は，二項分布，ポアソン分布，指数分布における正規近似の精度を調査することである．一言で近似といってもどういった条件下で，どれほどの精度の近似値を得られるのかを調査する．本研究では Excel を用い計算結果を得る．得られたデータから条件と近似の良さの関係を考察する．

2　式の導出

2.1　二項分布の正規近似

独立な n 回のベルヌーイ試行を $X_1, \cdots, X_n$ とする．確率変数 X は $X = X_1 + \cdots + X_n = n\bar{X}_n$ である．このとき，X は二項分布 $B(n,p)$ に従う．
X_i は二項分布 $B(1,p)$ に従い平均, 分散は次のようになる．

$$E(X_i) = p, \quad V(X_i) = p(1-p)$$

中心極限定理より，

$$\lim_{n\to\infty} P\left(\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \leqq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

が成り立つ．

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \leqq x \Longleftrightarrow X \leqq np + x\sqrt{np(1-p)}$$

の関係がある．

$s \equiv np + x\sqrt{np(1-p)}$ とおくと，参考文献 [1] の補題 7.5 より，

$$\begin{aligned} &P\big(X \leqq np + x\sqrt{np(1-p)}\big) \\ &= P(X \leqq [s]) \\ &= P\left(F_{2(n-[s])}^{2([s]+1)} \geqq \frac{(n-[s])p}{([s]+1)(1-p)}\right) \end{aligned}$$

と変換できるから，二項分布の正規近似による誤差は，

$$\left|P\left(F_{2(n-[s])}^{2([s]+1)} \geqq \frac{(n-[s])p}{([s]+1)(1-p)}\right) - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy\right| \tag{1}$$

となる．

ただし，自然数 m_1, m_2 に対して，$F_{m_2}^{m_1}$ を自由度 (m_1, m_2) の F 分布に従う確率変数とし，自然数 m に対して，$F_0^m = 1$ とする．
$[x]$ は x を超えない最大の整数である．

2.2　ポアソン分布の正規近似

$X_1, \cdots, X_n$ を平均 μ のポアソン分布に従う確率変数とする．また $X = X_1 + \cdots + X_n = n\bar{X}_n$ とすると，X の平均と分散は,

$$E(X) = n\mu, \quad V(X) = n\mu$$

である．
中心極限定理より，

$$\lim_{n\to\infty} P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}} \leqq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

が成り立つ．

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\mu}} \leqq x \Longleftrightarrow X \leqq n\mu + x\sqrt{n\mu}$$

の関係がある．
$t \equiv n\mu + x\sqrt{n\mu}$ とおくと，参考文献 [1] の補題 8.3 より，

$$\begin{aligned} &P(X \leqq n\mu + x\sqrt{n\mu}) \\ &\qquad = P(X \leqq [t]) \\ &\qquad = P\big(\chi^2_{2([t]+1)} \geqq 2n\mu\big) \end{aligned}$$

と変換できるから，正規近似による誤差は,

$$\left|P\big(\chi^2_{2([t]+1)} \geqq 2n\mu\big) - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy\right|$$

となる．

ただし，自然数 m に対して，χ^2_m を自由度 m の χ^2 分布に従う確率変数とし，$\chi^2_0 = 0$ とする．

2.3　指数分布の正規近似

$X_1, X_2, \cdots, X_n$ は互いに独立で指数分布 $EX\left(\frac{1}{\mu}\right)$ に従うとする. このとき,

$$T \equiv \left(\frac{2n}{\mu}\right)\bar{X}_n \sim \chi^2_{2n}$$

が成り立つ．ここで，χ^2_{2n} は自由度 $2n$ の χ^2 分布である．ただし，$\bar{X}_n \equiv \left(\frac{1}{n}\right)\sum_{i=1}^{n} X_i$ とする．

証明は参考文献 [2] の 5.1 で述べられている．

$X_1, \cdots, X_n$ が指数分布 $EX\left(\frac{1}{\mu}\right)$ に従うとき，X_i の平均と分散は次のようになる．

$$E(X_i) = \mu, \quad V(X_i) = \mu^2$$

中心極限定理より，

$$P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\mu} \leqq x\right) \approx \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

が成り立つ． $T \equiv \left(\frac{2n}{\mu}\right)\bar{X}_n$ より，

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\mu} \leqq x \Longleftrightarrow T \leqq 2n + 2x\sqrt{n}$$

である．

$T \sim \chi^2_{2n}$ より，指数分布の正規近似による誤差は，

$$\left| P\left(\chi^2_{2n} \leqq 2n + 2x\sqrt{n}\right) - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \right|$$

となる．

3 結果と考察

3.1 二項分布

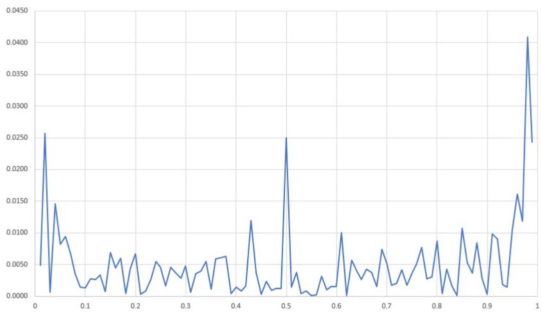

図 1 二項分布の正規近似における誤差

図 1 は試行回数 n を 500 回としたときの二項分布を正規近似したときの，F 分布と標準正規分布の最小の誤差の結果をグラフにしたものである．成功確率 p を $0 < p < 1.0$, $x = 1.0$ として計算結果を得た．グラフからは p が 0 と 1.0 の周辺と $p = 0.5$ のときに誤差がとても大きくなっていることが分かる．その他の成功確率の時はほとんどが誤差 0.005% 以内でありとても近似が良いといえる．結果から成功確率が極端に低い，あるいは極端に高い,$p = 0.5$ の場合は近似が悪くなると考えられる．

3.2 ポアソン分布

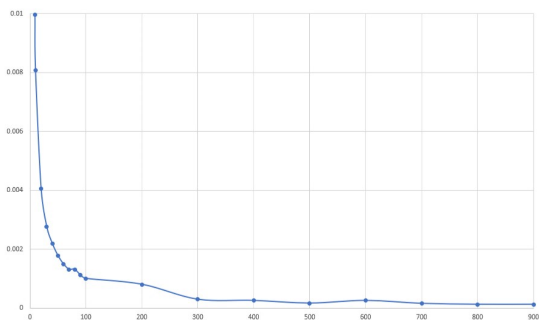

図 2 ポアソン分布の正規近似における誤差

図 2 は試行回数 n を 500 回としたときのポアソン分布を正規近似したときの，カイ二乗分布と標準正規分布の最小の誤差の結果をグラフにしたものである．$x = 1.0$ とし，μ を段階的に増加させ計算結果を得た．μ が 100 までの間に誤差が急激に小さくなり，それ以降も緩やかに小さくなっている．この結果からある程度 μ が大きければ近似が良いといえる．

3.3 指数分布

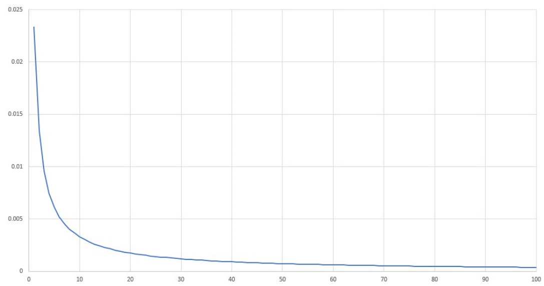

図 3 指数分布の正規近似における誤差

図 3 は試行回数 n を 100 回としたときの指数分布を正規近似したときの，自由度 $2n$ のカイ二乗分布と標準正規分布との誤差をグラフにしたものである．$x = 1.0$ とし n を 1 から 100 まで 1 ずつ増加させた．グラフはきれいな減衰曲線となり，カイ二乗分布の下側確率の方が $x = 1.0$ の標準正規分布の下側確率より常に大きい値をとっていた．指数分布の正規近似は n が増加するほど誤差が少なくなっていくことが分かる．

4 おわりに

本研究では二項分布，ポアソン分布，指数分布の正規近似の良さを調査した．もっとも良い結果が得られたのは指数分布で n の大きさ次第ではとても良い近似が得られる．二項分布，ポアソン分布はそれぞれのパラメータである p や μ の値によって結果が大きく変化することがわかった．二項分布では p が 0.5 のとき例外的に近似が良くない原因は未解決だ．今回の研究で用いた近似式においては良い結果が得られた．

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] 加藤駿介・丹羽雄士：『指数分布モデルにおける非劣等生の統計解析法』．南山大学 2013 年度卒業論文