正規二変量の比の信頼区間

城西大学 薬学部
長島 健悟

2012 年 10 月 30 日 改訂 (2)
2011 年 6 月 9 日 改訂 (1)
2009 年 4 月 30 日 作成

1　二つの平均の比の信頼区間

1.1　デルタ法による比の信頼区間

二つの確率変数に対するデルタ法を用いて, 二つの平均の比の期待値と分散の近似を求める.

X_1 および X_2 が確率変数のとき, 関数 $Y = h(X_1, X_2) = \frac{X_1}{X_2}$ も確率変数であり, Y に対する推測を考える. (20) 式を用いて, デルタ法による近似を行うと,

$$\mathrm{E}^A(Y) = \frac{\theta_1}{\theta_2} + \frac{\theta_1}{\theta_2^3}\mathrm{Var}(X_2) - \frac{1}{\theta^2}\mathrm{Cov}(X_1, X_2) \tag{1}$$

$$\mathrm{Var}^A(Y) = \left(\frac{1}{\theta_2}\right)^2 \mathrm{Var}(X_1) + \left(\frac{\theta_1}{\theta_2^2}\right)^2 \mathrm{Var}(X_2) - 2\left(\frac{\theta_1}{\theta_2^3}\right)\mathrm{Cov}(X_1, X_2), \tag{2}$$

となる. また, $\mathrm{Cov}(X_1, X_2) = \rho\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}$ として整理すると,

$$\mathrm{E}^A(Y) = \frac{\theta_1}{\theta_2} + \frac{\theta_1}{\theta_2^3}\mathrm{Var}(X_2) - \frac{1}{\theta^2}\rho\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)} \tag{3}$$

$$\mathrm{Var}^A(Y) = \frac{\theta_2^2\mathrm{Var}(X_1) + \theta_1^2\mathrm{Var}(X_2) - 2\theta_1\theta_2\rho\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}}{\theta_2^4} \tag{4}$$

となる.

Y は明らかに正規分布に従わないが, 漸近的に正規分布 $\mathrm{N}(\mathrm{E}^A(Y), \mathrm{Var}^A(Y))$ に従うと見なし, 信頼区間,

$$\mathrm{E}^A(Y) \pm z_{\alpha/2}\sqrt{\mathrm{Var}^A(Y)}, \tag{5}$$

を構成する. これがデルタ法による信頼区間である.

1.2　Fieller’s theorem による比の信頼区間 [1]

$Y = \frac{X_1}{X_2}$ を変形して整理すると $X_1 - YX_2 = 0$ である. ここで, Y を定数と見なし, $X_1 - YX_2$ の分布を考えてみよう. X_1, X_2 がどちらも正規分布に従う場合, $X_1 - YX_2$ は正規確率変数の線形結合だから, $\mathrm{N}(0, \mathrm{Var}(X_1) - 2Y\mathrm{Cov}(X_1, X_2) + Y^2\mathrm{Var}(X_2))$ である. したがって

$$\frac{X_1 - YX_2}{\sqrt{\mathrm{Var}(X_1) - 2Y\mathrm{Cov}(X_1, X_2) + Y^2\mathrm{Var}(X_2)}} \sim \mathrm{N}(0, 1) \tag{6}$$

である.

$X_1 - YX_2$ に対する $100(1-\alpha)\%$ 信頼限界は

$$\frac{X_1 - YX_2}{\sqrt{\mathrm{Var}(X_1) - 2Y\mathrm{Cov}(X_1, X_2) + Y^2\mathrm{Var}(X_2)}} = 0 \pm z_{\alpha/2} \tag{7}$$

であり, これについて 2 乗して Y について解くと,

$$\frac{2X_1X_2 + 2\mathrm{Cov}(X_1, X_2)z_{\alpha/2}^2 \pm K}{2\left(X_2^2 - \mathrm{Var}(X_2)z_{\alpha/2}^2\right)}, \tag{8}$$

$$K = \sqrt{\left(-2X_1X_2 - 2\mathrm{Cov}(X_1, X_2)z_{\alpha/2}^2\right)^2 - 4\left(X_1^2 - \mathrm{Var}(X_1)z_{\alpha/2}^2\right)\left(X_2^2 - \mathrm{Var}(X_2)z_{\alpha/2}^2\right)}, \tag{9}$$

である. これが Fieller's theorem による信頼区間である.

1.3 正規確率変数の比の分布に基づく信頼区間 [2]

$X_i \sim \mathrm{N}(\theta_i, \sigma_i^2)$ であり互いに独立な場合, $Y = X_1/X_2$ の確率密度関数は

$$f_Y(y) = \frac{b(y)c(y)}{a^3(y)} \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2} \left\{2\Phi\left(\frac{b(y)}{a(y)}\right) - 1\right\} + \frac{1}{a^2(y)\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2}\right)\right\}, \tag{10}$$

$$a(y) = \sqrt{\frac{1}{\sigma_1^2}y^2 + \frac{1}{\sigma_2^2}}, \;\; b(y) = \frac{\mu_1}{\sigma_1^2}y + \frac{\mu_2}{\sigma_2^2}, \;\; c(y) = \exp\left\{\frac{1}{2}\frac{b^2(y)}{a^2(y)} - \frac{1}{2}\left(\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2}\right)\right\}, \tag{11}$$

である. ただし, $\Phi(y)$ は標準正規分布の累積分布関数である.

この分布に基づいて,

$$\Pr\left(f_Y^{-1}(y_L) < Y < f_Y^{-1}(y_U)\right) = 1 - \alpha, \tag{12}$$

を満たす y_L, y_U を求めれば, $100(1-\alpha)\%$ 信頼区間を得ることができる.

2 数値例

$X_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$ のときの Y に対する信頼区間を考える. X_{ij} は互いに独立で, 期待値 θ_i, 分散 $\sigma_i^2 < \infty$ の確率分布に従うとする. 中心極限定理から, n_i が十分に大きければ $X_i \sim \mathrm{N}(\theta_i, \sigma_i^2/n_i)$ である.

$\theta_i, \mathrm{Var}(X_i)$ の推定量,

$$\hat{\theta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \tag{13}$$

$$\hat{\mathrm{Var}}(X_i) = \frac{1}{n_i(n_i - 1)}\sum_{j=1}^{n_i}(X_{ij} - \hat{\theta}_i)^2 \tag{14}$$

を代入した場合の, 3 つの信頼区間を計算してみよう.

2.1 データ

$\hat{\theta}_1 = 9.08, \hat{\theta}_2 = 4.89, \hat{\mathrm{Var}}(X_1) = 0.062, \hat{\mathrm{Var}}(X_2) = 0.048, n_1 = 20, n_2 = 30$ とする. なお, 二群は独立である場合とする.

2.2 計算結果

(5) 式のデルタ法による信頼区間を求めると,

$$[1.668, 2.051] \tag{15}$$

であった.

(8) 式の Fieller's theorem による比の信頼区間を求めると,

$$[1.678, 2.063] \tag{16}$$

であった.

(12) 式の正規確率変数の比の分布に基づく信頼区間を求めると,

$$[1.678, 2.063] \tag{17}$$

であった.

2.3 考察

Fieller's theorem による信頼区間か, 正規確率変数の比の分布に基づく信頼区間を用いればよいだろう. デルタ法による近似信頼区間は, 他の二手法とは若干異なる値であった.

なお, 独立ではない場合は検討していないが, Fieller's theorem による信頼区間において共分散が 0 ではない場合か, 多変量正規分布を元に導いた比の分布から求めれば良いと考えられる. 文献値などの要約データからの計算を行う場合には, $\mathrm{Cov}(X_1, X_2) = \rho\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}$ の相関係数 ρ の値をいくつか与えて求めた $\hat{\mathrm{Var}}(Y)$ 値を見て, 参考情報として役立てることが出来ればよいだろう.

A　R スクリプト

ダミーデータの作製と解析に用いた R スクリプトを示す.

```
set.seed(758923)
n1 <- 20
n2 <- 30
x1 <- rnorm(n1, 9, 1)
x2 <- rnorm(n2, 5, 1)
theta1 <- mean(x1)
theta2 <- mean(x2)
vhat1 <- var(x1)/n1
vhat2 <- var(x2)/n2
alpha <- 0.05
z <- qnorm(1 - alpha/2)

# Delta method
EA <- theta1/theta2 + theta1/theta2^3*vhat2
VA <- (1/theta2)^2 * vhat1 + (theta1/theta2^2)^2 * vhat2
Dl <- EA - z*sqrt(VA)
Du <- EA + z*sqrt(VA)

# Fieller's theorem
K <- sqrt((-2*theta1*theta2)^2 - 4*(theta1^2-vhat1*z^2)*(theta2^2-vhat2*z^2))
Fl <- (2*theta1*theta2 - K) / (2*(theta2^2-vhat2*z^2))
Fu <- (2*theta1*theta2 + K) / (2*(theta2^2-vhat2*z^2))

# Ratio distribution
fy <- function(y, mu1, mu2, sigma1, sigma2) {
  a <- function(...) sqrt(y^2/sigma1^2 + 1/sigma2^2)
  b <- function(...) y*mu1/sigma1^2 + mu2/sigma2^2
  c <- function(...) exp(b(y)^2/a(y)^2/2 - (mu1^2/sigma1^2 + mu2^2/sigma2^2) / 2)
  b(y)*c(y)/a(y)^3 * (2*pnorm(b(y)/a(y))-1) / (sqrt(2*pi)*sigma1*sigma2) +
  exp(-(mu1^2/sigma1^2 + mu2^2/sigma2^2)/2) / (a(y)^2*pi*sigma1*sigma2)
}
f <- function(y) fy(y, theta1, theta2, sqrt(vhat1), sqrt(vhat2))
objfl <- function(x) (integrate(f, 0, x)$value - 0.025)^2
objfu <- function(x) (integrate(f, x, +Inf)$value - 0.025)^2
Rl <- optimize(objfl, c(1, 3))$minimum
Ru <- optimize(objfu, c(1, 3))$minimum

Dl; Du
Fl; Fu
Rl; Ru
```

```
> Dl; Du
[1] 1.682642
[1] 2.035469
> Fl; Fu
[1] 1.690572
[1] 2.045355
> Rl; Ru
[1] 1.690579
[1] 2.045372
```

被覆確率のシミュレーションに用いた R スクリプトおよび結果を示す.

```
deltaci <- function(theta1, theta2, vhat1, vhat2, alpha) {
  z <- qnorm(1 - alpha/2)
  EA <- theta1/theta2 + theta1/theta2^3*vhat2
  VA <- (1/theta2)^2 * vhat1 + (theta1/theta2^2)^2 * vhat2
  Dl <- EA - z*sqrt(VA)
  Du <- EA + z*sqrt(VA)
  return(list(Dl=Dl,Du=Du))
}

fiellerci <- function(theta1, theta2, vhat1, vhat2, alpha) {
  z <- qnorm(1 - alpha/2)
  K <- sqrt((-2*theta1*theta2)^2 - 4*(theta1^2-vhat1*z^2)*(theta2^2-vhat2*z^2))
  Fl <- (2*theta1*theta2 - K) / (2*(theta2^2-vhat2*z^2))
  Fu <- (2*theta1*theta2 + K) / (2*(theta2^2-vhat2*z^2))
  return(list(Fl=Fl,Fu=Fu))
}
```

```r
ratioci <- function(theta1, theta2, vhat1, vhat2, alpha) {
  fy <- function(y, mu1, mu2, sigma1, sigma2) {
    a <- function(...) sqrt(y^2/sigma1^2 + 1/sigma2^2)
    b <- function(...) y*mu1/sigma1^2 + mu2/sigma2^2
    c <- function(...) exp(b(y)^2/a(y)^2/2 - (mu1^2/sigma1^2 + mu2^2/sigma2^2) / 2)
    b(y)*c(y)/a(y)^3 * (2*pnorm(b(y)/a(y))-1) / (sqrt(2*pi)*sigma1*sigma2) +
    exp(-(mu1^2/sigma1^2 + mu2^2/sigma2^2)/2) / (a(y)^2*pi*sigma1*sigma2)
  }
  f <- function(y) fy(y, theta1, theta2, sqrt(vhat1), sqrt(vhat2))
  objfl <- function(x) (integrate(f, 0, x)$value - alpha/2)^2
  objfu <- function(x) (integrate(f, x, +Inf)$value - alpha/2)^2
  Rl <- optimize(objfl, c(1, 3))$minimum
  Ru <- optimize(objfu, c(1, 3))$minimum
  return(list(Rl=Rl,Ru=Ru))
}

set.seed(758923)

rep <- 10000

# Set simulation parameters
n1 <- 20; n2 <- 30
mu1 <- 9; mu2 <- 5
sigma1 <- 1; sigma2 <- 1
ytrue <- 9/5

# Start a simulation
dcr <- 0; fcr <- 0; rcr <- 0
theta1 <- rnorm(rep, mu1, sigma1/sqrt(n1))
theta2 <- rnorm(rep, mu2, sigma2/sqrt(n2))
vhat1 <- rchisq(rep, n1-1) * sigma1^2 / (n1-1) / n1
vhat2 <- rchisq(rep, n2-1) * sigma2^2 / (n2-1) / n2

d <- deltaci(theta1, theta2, vhat1, vhat2, 0.05)
f <- fiellerci(theta1, theta2, vhat1, vhat2, 0.05)

# Create the progress bar.
pb <- txtProgressBar(min = 1, max = rep, style=3)

for(i in 1:rep) {
  r <- ratioci(theta1[i], theta2[i], vhat1[i], vhat2[i], 0.05)
  if (ytrue > d$Dl[i] && ytrue < d$Du[i]) dcr <- dcr + 1
  if (ytrue > f$Fl[i] && ytrue < f$Fu[i]) fcr <- fcr + 1
  if (ytrue > r$Rl && ytrue < r$Ru) rcr <- rcr + 1
  setTxtProgressBar(pb, i)
}

# Close the progress bar.
close(pb)

# coverage
dcr / rep * 100
fcr / rep * 100
rcr / rep * 100
```

```
> # coverage
> dcr / rep * 100
[1] 95.13
> fcr / rep * 100
[1] 95.03
> rcr / rep * 100
[1] 95.03
```

デルタ法の被覆確率は若干大きい. やはり, Fieller's theorem による信頼区間か, 正規確率変数の比の分布に基づく信頼区間を用いればよいだろう.

B　二つの確率変数のデルタ法

二つの確率変数 $\mathbf{X} = (X_1, X_2)^T$ に対する関数 $h(\mathbf{X})$ のデルタ法を考える.

まず $h(\mathbf{X})$ を, $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ の周りで 2 次までテーラー展開すると

$$\begin{aligned} h(\mathbf{X}) =& h(\boldsymbol{\theta}) + h'_1(\boldsymbol{\theta})(X_1 - \theta_1) + h'_2(\boldsymbol{\theta})(X_2 - \theta_2) \\ &+ \frac{1}{2}\left\{h''_{11}(\boldsymbol{\theta})(X_1 - \theta_1)^2 + 2h''_{12}(\boldsymbol{\theta})(X_1 - \theta_1)(X_2 - \theta_2) + h''_{22}(\boldsymbol{\theta})(X_2 - \theta_2)^2\right\} \\ &+ o\left((X_1 - \theta_1)^2 + (X_2 - \theta_2)^2\right) \end{aligned} \tag{18}$$

が得られる[†]. ただし,

$$h'_i(\boldsymbol{\theta}) = \left.\frac{\partial h(\mathbf{X})}{\partial X_i}\right|_{\mathbf{X}=\boldsymbol{\theta}},\ h''_{ij}(\boldsymbol{\theta}) = \left.\frac{\partial^2 h(\mathbf{X})}{\partial X_i \partial X_j}\right|_{\mathbf{X}=\boldsymbol{\theta}},$$

とした.

これを用いて, $h(\mathbf{X})$ の期待値は

$$\begin{aligned} \mathrm{E}\{h(\mathbf{X})\} =& h(\boldsymbol{\theta}) + \frac{1}{2}h''_{11}(\boldsymbol{\theta})\mathrm{Var}(X_1) + h''_{12}(\boldsymbol{\theta})\mathrm{Cov}(X_1, X_2) + \frac{1}{2}h''_{22}(\boldsymbol{\theta})\mathrm{Var}(X_2) \\ &+ o\left((X_1 - \theta_1)^2 + (X_2 - \theta_2)^2\right) \end{aligned} \tag{19}$$

と書くことができる. また, $h(\mathbf{X})$ の分散は

$$\begin{aligned} \mathrm{Var}\{h(\mathbf{X})\} =& \{h'_1(\boldsymbol{\theta})\}^2\,\mathrm{Var}(X_1) + \{h'_2(\boldsymbol{\theta})\}^2\,\mathrm{Var}(X_2) + 2h'_1(\boldsymbol{\theta})h'_2(\boldsymbol{\theta})\mathrm{Cov}(X_1, X_2) \\ &+ o\left((X_1 - \theta_1)^2 + (X_2 - \theta_2)^2\right) \end{aligned} \tag{20}$$

と書くことができる.

また, X_1 が確率変数列 X_{n_1}, X_2 が確率変数列 X_{n_2} のとき, $n = \min_i\{n_i\}$ とすると,

$$o\left((X_{n_1} - \theta_1)^2 + (X_{n_2} - \theta_2)^2\right) = o_p(n^{-1}) \tag{21}$$

である.

C　一般の場合のデルタ法

p 個の確率変数の関数 $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ を, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ の周りで 1 次のテーラー展開をすると,

$$\begin{aligned} h(\mathbf{X}) =& h(\boldsymbol{\theta}) + \sum_{i=1}^{p} h'_i(\boldsymbol{\theta})(X_i - \theta_i) + \frac{1}{2}\sum_{i=1}^{p} h''_{ii}(\boldsymbol{\theta})(X_i - \theta_i)^2 \\ &+ \sum_{i=1}^{p-1}\sum_{j=i+1}^{p} h''_{ij}(\boldsymbol{\theta})(X_i - \theta_i)(X_j - \theta_j) + o\left(\sum_{i=1}^{p}(X_i - \theta_i)^2\right) \end{aligned} \tag{22}$$

$h(\mathbf{X})$ の期待値は

$$\mathrm{E}\{h(\mathbf{X})\} = h(\boldsymbol{\theta}) + \frac{1}{2}\sum_{i=1}^{p} h''_{ii}(\boldsymbol{\theta})\mathrm{Var}(X_i) + \sum_{i=1}^{p-1}\sum_{j=i+1}^{p} h''_{ij}(\boldsymbol{\theta})\mathrm{Cov}(X_i, X_j) + o\left(\sum_{i=1}^{p}(X_i - \theta_i)^2\right) \tag{23}$$

と書くことができる. また, $h(\mathbf{X})$ の分散は

$$\mathrm{Var}\{h(\mathbf{X})\} = \sum_{i=1}^{p} h'_i(\boldsymbol{\theta})\mathrm{Var}(X_i) + 2\sum_{i=1}^{p-1}\sum_{j=i+1}^{p} h''_{ij}(\boldsymbol{\theta})\mathrm{Cov}(X_i, X_j) + o\left(\sum_{i=1}^{p}(X_i - \theta_i)^2\right) \tag{24}$$

と書くことができる.

[†]一変数ずつ順番にテーラー展開すればよい

参考文献

[1] Fieller EC. The distribution of the index in a normal bivariate population. *Biometrika* 1932;**3**(4):428–440.

[2] Hinkley DV. On the ratio of two correlated normal random variables. *Biometrika* 1969;**56**(3):635–639.