Computing Exact Confidence Intervals for the Difference of Two Proportions Using R

Guogen Shan (guogen.shan@unlv.edu) and Weizhen Wang (weizhen.wang@wright.edu)

We here show how to compute exact confidence intervals (lower one-sided, upper one-sided and two-sided intervals) for the difference of two proportions, $p_1 - p_2$, in two cases.

CASE i). Observe (n_{12}, t, n_{21}), where $t = n_{11} + n_{22}$, from a matched pair experiment. The function PairedCI() in the **ExactCI** package is for calculating the one-sided confidence intervals for the difference of two paired proportions p_1 and p_2. The basic usage of this function is provided with complete arguments as the following,

```
PairedCI(n12,t,n21,confidence.level=0.95,alternative='Lower',
         precision=0.00001,grid.one=30,grid.two=20)
```

The arguments $n12, t$, and $n21$ in the function take the measurements from the experiment. The value of confidence.level is the level of confidence, $1 - \alpha$. Precision of the confidence interval, default is 0.00001 rounded to 5 decimals. The values of grid.one and grid.two are the number grid points in the two-step grid search algorithm for the global minimum. The higher values of grid.one and grid.two, the more accurate of the solution. Based on our extensive numerical study, we found that grid.one=30 and grid.two=20 are sufficient enough for the problem.

Example 1. (Karacan et al., 1976) We illustrate to show the usage of the PairdCI() function to calculate the exact smallest lower one-sided confidence interval. In this study, 32 marijuana users are compared with 32 matched controls with respect to their sleeping difficulties, with $n_{11} = 16, n_{12} = 9, n_{21} = 3$, and $n_{22} = 4$. The second argument in the function is $t = n_{11} + n_{22} = 20$. The researchers wish to see how much more help the marijuana use provides for sleeping by using a lower one-sided 95% confidence interval $[L_P(n_{12}, t), 1]$ for $\theta_P = p_1 - p_2$ at $(n_{12}, t) = (9, 20)$, where p_1 is the proportion of marijuana users who have sleeping improved, and p_2 is the proportion in the controls. First download ExactCI_1.1.zip at http://www.wright.edu/~weizhen.wang/software/ExactTwoProp/ExactCI_1.1.zip and install it. Then try the following.

```
> library(ExactCI)
> lciall=PairedCI(9,20,3,confidence.level=0.95)  # store relevant quantities
> lciall                                         # print lciall
$confidence.level
[1] 0.95                                         # confidence level
$alternative
[1] "Lower"                                      # the lower one-sided interval
$estimate
[1] 0.1875                                       # the mle of p1-p2
$ExactOneCI
[1] 0.00613 1.00000      # the computed lower one-sided interval in characters
> lci=lciall$ExactOneCI  # the computed 95% lower one-sided interval in numbers
> lci
[1] 0.00613 1.00000
```

The use of marijuana helps sleeping because interval $[0.00613, 1.00000]$ is positive. The upper one-sided 95% interval and the two-sided 95% interval are given below.

```
> library(ExactCI)
> uci=PairedCI(9,20,3,confidence.level=0.95,alternative='Upper')$ExactOneCI
> uci                    # the computed 95% upper one-sided interval in numbers
[1] -1.00000  0.36234
> u95=PairedCI(9,20,3,confidence.level=0.975, alternative='Upper')$ExactOneCI
> u95                    # the 97.5% upper one-sided interval in numbers
[1] -1.00000  0.39521
```

```
> l95=PairedCI(9,20,3,confidence.level=0.975)$ExactOneCI
> l95                    # the 97.5% lower one-sided interval in numbers
[1] -0.03564  1.00000
> ci95=c(l95[1],u95[2])
> ci95                   # the 95% two-sided interval in numbers
[1] -0.03564  0.39521
```

The 95% upper one-sided interval and the 95% two-sided interval for $p_1 - p_2$ are $[-1.00000 0.36234]$ and $[-0.03564 0.39521]$, respectively.

CASE ii). Observe two independent binomial variables: $X \sim Bin(n_1, p_1)$ and $Y \sim Bin(n_2, p_2)$. The function BinomCI() in the **ExactCI** package is for calculating the one-sided confidence intervals for the difference of two independent proportions p_1 and p_2. First download and install ExactCI_1.1.zip if not done yet. The basic usage of this function is provided with complete arguments as the following,

```
BinomCI(n1,n2,x,y,confidence.level=0.05,alternative='Lower',
         precision=0.00001,grid.one=30,grid.two=20)
```

The arguments n_1, n_2, x, and y are the observations from the experiment.

Example 2. There is a two-arm randomized clinical trial for the effect of tobacco smoking on mice by Essenberg (1952). In the treatment (smoking) group, the number of mice is $n_1 = 23$, and the number of mice developed tumor is $x = 21$; in the control group, $n_2 = 32$ and $y = 19$. The mle for the difference between two proportions is

$$\theta_I = \frac{x}{n_1} - \frac{y}{n_2} = 0.319293.$$

The lower confidence interval for θ_I is given in the following.

```
> library(ExactCI)
```

```
> lciall=BinomCI(23,32,21,19)      # store relevant quantities
> lciall                           # print lciall
$confidence.level
[1] 0.95                           # confidence level
$alternative
[1] "Lower"                        # the lower one-sided interval
$estimate
[1] 0.319293                       # the mle of p1-p2
$ExactOneCI
[1] 0.133 1.00000         # the computed lower one-sided interval in characters
> lci=lciall$ExactOneCI  # the computed 95% lower one-sided interval in numbers
> lci
[1] 0.133 1.00000
```

The lower one-sided 95% confidence interval is [0.133,1]. Therefore, the tumor rate in the smoking group is higher than that of the control group. The following code is for the upper one-sided and two-sided 95% confidence intervals.

```
> library(ExactCI)
> uci=BinomCI(23,32,21,19,confidence.level=0.95,alternative='Upper')$ExactOneCI
> uci                     # the computed 95% upper one-sided interval in numbers
[1] -1.00000  0.48595
> u95=BinomCI(23,32,21,19,confidence.level=0.975,alternative='Upper')$ExactOneCI
> u95                     # the 97.5% upper one-sided interval in numbers
[1] -1.00000  0.51259
> l95=Binom(23,32,21,19,confidence.level=0.975)$ExactOneCI
> l95                     # the 97.5% lower one-sided interval in numbers
[1] 0.09468 1.00000
```

```
> ci95=c(l95[1],u95[2])
> ci95
[1] 0.09468  0.51259     # the 95% two-sided interval in numbers
```

The 95% upper one-sided interval and the 95% two-sided interval for $p_1 - p_2$ are $[-1.00000, 0.48595]$ and $[0.09468, 0.51259]$, respectively.

REFERENCES:

Essenberg, J. M. (1952). Cigarette smoke and the incidence of primary neoplasm of the lung in albino mice. *Science* 116, 561-562.

Karacan, I., Fernandez-Salas, A., Coggins, W. J., Carter, W. E., Williams, R. L., Thornby, J. I., Salis, P. J., Okawa, M., and Villaume, J. P. (1976). Sleep Electrocephalographic-Celectrooculographic Characteristics of Chronic Marijuana Users: Part 1. *Annals of New York Academy of Sciences*, 282, 348-374.

Shan, G. and Wang, W. (2012). ExactCI: An R Package for Computing Exact Confidence Intervals for the Difference of Two Proportions. Manuscript.

Wang, W. (2010). On construction of the smallest one-sided confidence interval for the difference of two proportions. *The Annals of Statistics* 38, 1227-1243.

Wang, W. (2012). An inductive order construction for the difference of two dependent proportions. *Statistics and Probability Letters* **82**, 1623-1628.