# The Continuous Bernoulli distribution software package ---transformation, simulator, testing and confidence interval.

**The probability density function of Continuous Bernoulli distribution,**
$X \sim CB(\lambda)$, this probability distribution for "machine learning".
(i)The probability density function,

$$f_X(x;\lambda)=C(\lambda)\lambda^x(1-\lambda)^{1-x},0<x<1,0<\lambda<1,$$

$$C(\lambda)=\begin{cases}\dfrac{2tanh^{-1}(1-2\lambda)}{1-2\lambda},\lambda\neq\dfrac{1}{2}\\ 2,\lambda=\dfrac{1}{2}\end{cases}$$

$$tanh^{-1}(x)=\frac{1}{2}log_e\left(\frac{1+x}{1-x}\right)=\frac{1}{2}ln\left(\frac{1+x}{1-x}\right),-1<x<1,$$

(ii)The distribution function,

$$F_X(x;\lambda)=\begin{cases}\dfrac{\lambda^x(1-\lambda)^{1-x}+\lambda-1}{2\lambda-1},\lambda\neq\dfrac{1}{2}\\ x,\lambda=\dfrac{1}{2}\end{cases},0<x<1$$

The following function is derived from the Continuous Bernoulli distribution which contains the probability distribution and statistical analysis.
Please see the free book--"Continuous_Bernoulli distribution-simulator and test statistic", (cloud drive )
if you want to understand the simulator transformation and mathematical formula and test statistic and confidence interval method.

The software program code is based on this book, the following is about the program designed explanation in brief.

1.The Continuous Bernoulli probability simulator,

$$x=\begin{cases}\dfrac{ln\left(F_X\left(x;\lambda\right)\times\left(2\lambda-1\right)-\left(\lambda-1\right)\right)-ln\left(1-\lambda\right)}{ln\left(\dfrac{\lambda}{1-\lambda}\right)},\lambda\neq\dfrac{1}{2}\\ F_X\left(x;\lambda\right),\lambda=\dfrac{1}{2}\end{cases}$$

The random number= $RND=F_X\left(x;\lambda\right)\sim Uniform(0,1)$,

$$x=\begin{cases}\dfrac{ln\left(RND\times\left(2\lambda-1\right)-\left(\lambda-1\right)\right)-ln\left(1-\lambda\right)}{ln\left(\dfrac{\lambda}{1-\lambda}\right)},\lambda\neq\dfrac{1}{2}\\ RND,\lambda=\dfrac{1}{2}\end{cases}$$

The program is
C:\C_Bernoulli\C_Bernoulli_01.cpp,
which is the generator of Continuous Bernoulli( $\lambda$ ),

The input data of C:\C_Bernoulli\C_Bernoulli_01.cpp,
lamda=?
the simulated data number=?

The output data,
The simulated data will output to console and filename
c:\C_Bernoulli\tep\simulated_data.txt.

2.Continuous Bernoulli moment and moment estimated valueand frequency distribution,

Computing the E(X),Var(X), skewed coefficient and kurtosis coefficient of Continuous Bernoulli( $\lambda$ ) using the simulated data from simulator and the data number=100,000,000. This number of sample data is closed to the population distribution.

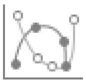
The program is
C:\C_Bernoulli\C_Bernoulli_02.cpp,

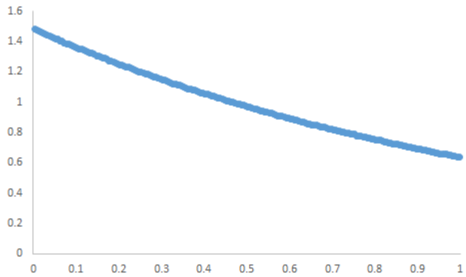
The input data of C:\C_Bernoulli\C_Bernoulli_02.cpp,
lamda=?

The output data,
computing the sample mean, sample variance, skewed coefficient and kurtosis coefficient and the simulated data size=100,000,000, the sample coefficient is approached to the population coefficient.
The frequency table of 100,000,000 data filename is c:\C_Bernoulli\tep\frequency_table.txt which can use the excel.exe to plot the probability distribution and distribution function.

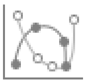
Continuous Bernoulli( $\lambda$ =0.3) and running this program to get frequency_table.txt.
Please use excel.exe plot the probability distribution and the distribution function,
Running excel.exe
Reading the c:\C_Bernoulli\tep\frequency_table.txt

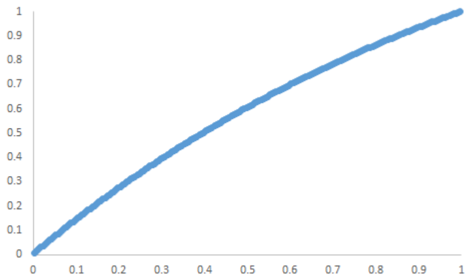
Marking A1 B1 and






Marking C1 D1 and






Note:
frequency_table.txt which class number=200,
$1^{st}$ line is class mid-point , $2^{nd}$ line is the relative frequency,
$3^{rd}$ line is class mid-point , $4^{th}$ line is the cumulative relative frequency,

3. Continuous Bernoulli sample mean sampling distribution,

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} CB(\lambda),$$

X bar=the sample mean= $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$,

Computing the E(X bar),Var(X var), skewed coefficient and kurtosis coefficient of using the simulated data from simulator and the data number=100,000,000. This number of sample data is closed to the population distribution.

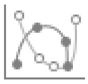
The program is
C:\C_Bernoulli\C_Bernoulli_03.cpp,

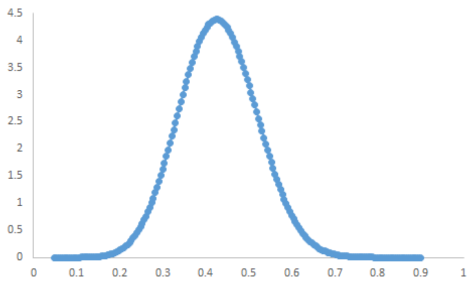
The input data of C:\C_Bernoulli\C_Bernoulli_03.cpp,
lamda=?
sample size=?

The output data,
computing the mean, variance, skewed coefficient and kurtosis coefficient of sample mean and the simulated data size=100,000,000, the sample coefficient is approached to the population coefficient.
The frequency table of 100,000,000 data filename is
c:\C_Bernoulli\tep\frequency_table.txt which can use the excel.exe to plot the probability distribution and distribution function.

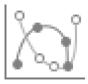
Continuous Bernoulli( $\lambda$ =0.3), sample size=10 and running this program to get frequency_table.txt.
Please use excel.exe plot the probability distribution and the distribution function,
Running excel.exe
Reading the c:\C_Bernoulli\tep\frequency_table.txt

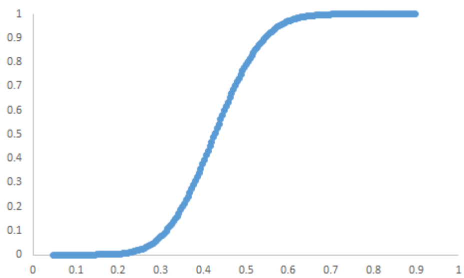
Marking A1 B1 and






Marking C1 D1 and






Note:
frequency_table.txt which class number=200,
1st line is class mid-point , 2nd line is the relative frequency,
3rd line is class mid-point , 4th line is the cumulative relative frequency,

4. Continuous Bernoulli lamda point estimator(MLE estimated equation) moment and sampling distribution frequency distribution,

The $\hat{\lambda}$ is from the following function,

$\lambda = \phi^*(\mu)$, $\phi^*(\mu)$ estimated equation,

(1)The estimated equation,

X= -0.596698+2.193196$\times\mu$,

$\lambda$=0.49997386580423608+1.36802409685464270*(X-0.5056)^1+
-0.00092474767006933689 0*(X-0.5056)^2+-2.73607823707760640*(X-0.5056)^3+
0.09510904364287853 2*(X-0.5056)^4+5.7483773675921839*(X-0.5056)^5+
-1.8419988453388214*(X-0.5056)^6+-12.357242575206328*(X-0.5056)^7+
16.361405849456787*(X-0.5056)^8+26.41792850010097*(X-0.5056)^9+
-80.02126121520996*(X-0.5056)^10+-48.621550429612398*(X-0.5056)^11+
228.76872253417969*(X-0.5056)^12+64.702439151704311*(X-0.5056)^13+
-380.75874328613281*(X-0.5056)^14+-51.895506033673882*(X-0.5056)^15+
341.66360473632812*(X-0.5056)^16+18.360968290828168*(X-0.5056)^17+
-127.70810317993164*(X-0.5056)^18,

ANOVA

| Source | df | SS | MS |
|---|---|---|---|
| Regression | 18 | 83.0834922851 | 4.6157495714 |
| Error | 980 | 0.0000077149 | 0.0000000079 |
| Total | 998 | 83.0835000000 | |

H0:slope1=....=slope18=0, test statistic=586328245.808614, p value=0.000000, sample size=999, R2=1.000000, R2(adj)=1.000000,MSE=0.000000,

The R2$\rightarrow$1 and MSE=0, $\phi^*(\ )$ is not error when $\phi^*(\mu)$ converting $\lambda$.
$\phi(\ )$ estimated equation is $\phi^*(\ )$, the MLE of $\lambda$ which estimated equation is $\hat{\lambda} = \phi(\bar{x}) = \phi^*(\bar{x})$.

(2)The $\hat{\lambda} = \phi(\bar{x}) = \phi^*(\bar{x})$ error,

$\overline{X} = \mu + \varepsilon, E(\varepsilon) = 0, E(\varepsilon^2) = \frac{Var(X)}{n} \xrightarrow[n\to\infty]{}, \varepsilon \xrightarrow[n\to\infty]{} 0$.

The $\lambda = \phi^*(\mu), \phi^*(\overline{X} - \varepsilon) \xrightarrow[n\leftarrow\infty]{} \phi^*(\overline{X})$, $\lambda$ MLE=$\phi(\overline{X}) = \phi^*(\overline{X})$.

$\phi(\overline{X})$ hqw asymptotic unbiased, $E(\phi(\overline{X})) \neq \lambda$, but $E(\phi(\overline{X})) \xrightarrow[n\to\infty]{} \lambda$, the estimated error is very small can be seen as 0.

But $\lambda$=0.5, $E(\overline{X})$=$\lambda$=0.5, the $\lambda$ MLE=$\overline{X}$ is unbiased estimator if $\lambda$=0.5.

(3)The limitation of estimated equation, $\phi(\overline{X})$,

0.143853919$\leq \overline{X} \leq$0.856221427, the $\hat{\lambda} = \phi(\overline{X})$ could be reasonable number which is 0.001$\leq \hat{\lambda} \leq$0.999.

$X_1, X_2, ..., X_n \overset{iid}{\sim} CB(\lambda)$,

X bar=the sample mean= $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, lamda hart= $\hat{\lambda} = \phi(\overline{x}) = \phi^*(\overline{x})$

Computing the E(lamda hat),Var(lamda hat), skewed coefficient and kurtosis coefficient of using the simulated data from simulator and the data number=100,000,000. This number of sample data is closed to the population distribution.

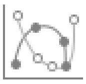
The program is
C:\C_Bernoulli\C_Bernoulli_04.cpp,

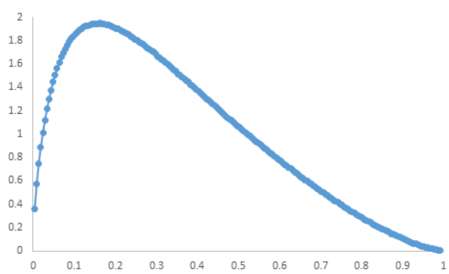
The input data of C:\C_Bernoulli\C_Bernoulli_04.cpp,
lamda=?
sample size=?

The output data,
computing the mean, variance, skewed coefficient and kurtosis coefficient of $\hat{\lambda} = \phi(\overline{x}) = \phi^*(\overline{x})$ and the simulated data size=100,000,000, the sample coefficient is approached to the population coefficient.
The frequency table of 100,000,000 data filename is
c:\C_Bernoulli\tep\frequency_table.txt which can use the excel.exe to plot the probability distribution and distribution function.

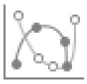
Continuous Bernoulli( $\lambda$ =0.3), sample size=10 and running this program to get frequency_table.txt.
Please use excel.exe plot the probability distribution and the distribution function,
Running excel.exe
Reading the c:\C_Bernoulli\tep\frequency_table.txt

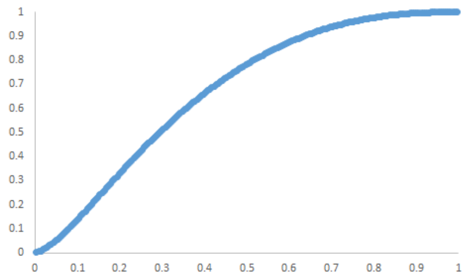
Marking A1 B1 and






Marking C1 D1 and






Note:
frequency_table.txt which class number=200,
1st line is class mid-point , 2nd line is the relative frequency,
3rd line is class mid-point , 4th line is the cumulative relative frequency,

5.Continous Bernoulli parameter lamda test statistic using input data,

$X_1, X_2, ..., X_n \overset{iid}{\sim} CB(\lambda)$,

The test statistic, $\frac{\sqrt{n}\left(\overline{X}-\mu(X)\right)}{\sigma(X)}$ is better than $\frac{\hat{\lambda}-E(\hat{\lambda})}{\sqrt{Var(\hat{\lambda})}}$ because

$E\left(\frac{\sqrt{n}\left(\overline{X}-\mu(X)\right)}{\sigma(X)}\right)=0, Var\left(\frac{\sqrt{n}\left(\overline{X}-\mu(X)\right)}{\sigma(X)}\right)=1$.

$\frac{\sqrt{n}\left(\overline{X}-\mu(X)\right)}{\sigma(X)} \xrightarrow[n\to\infty]{} Normal(0,1)$ which sample size is not too much.

(1)The Z test statistic for large sample,
$n \geq 6+250 \times |\lambda$ -0.5|, if $0.1 \leq \lambda \leq 0.9$,
$n \geq 100+2000 \times (\lambda$ -0.1), if $\lambda$ <0.1,
$n \geq 100+2000 \times (\lambda$ -0.9),if $\lambda$ >0.9,
$\frac{\sqrt{n}\left(\overline{X}-\mu(X)\right)}{\sigma(X)} \longrightarrow Normal(0,1)$,
$H_0 : \lambda = c$ $H_0 : \lambda = c$,
$Z^* = \frac{\overline{X}-G_1(c)}{\sqrt{\frac{G_2(c)}{n}}} \rightarrow Z \sim Normal(0,1), |Z^*| > Z_{\alpha/2}$ rejected $H_0$ and $P(Z > Z_{\alpha/2}) = \frac{\alpha}{2}$.
$G_1(\lambda)$ is $E(X)$ estimated equation and $G_2(\lambda)$ is $Var(X)$ estimated equation.

(2)The test statistic sampling distribution from simulator for small sample,
$n$<6+250 $\times |\lambda$ -0.5|, if $0.1 \leq \lambda \leq 0.9$,
$n$<100+2000 $\times (\lambda$ -0.1), if $\lambda$ <0.1,
$n$<100+2000 $\times (\lambda$ -0.9),if $\lambda$ >0.9,
The critical value of test statistic is computed by the simulated sampling distribution of $\frac{\sqrt{n}\left(\overline{X}-\mu(X)\right)}{\sigma(X)}$.

$H_0 : \lambda = c$ $H_0 : \lambda = c$, the test statistic value= $\frac{\overline{X}-G_1(c)}{\sqrt{\frac{G_2(c)}{n}}}$,

$G_1(\lambda)$ is $E(X)$ estimated equation and $G_2(\lambda)$ is $Var(X)$ estimated equation.

The program is
C:\C_Bernoulli\C_Bernoulli_05.cpp,

The input data of C:\C_Bernoulli\C_Bernoulli_05.cpp,
The data filename and $H_0 : \lambda = c$, $c$=?

The output data,
The test result.

example, the sample data is simulated from $CB(\lambda = 0.8)$ and data size=200,
$H_0 : \lambda = c$, $c$=0.8,
the output is

```
sample size=200, the requirement size=81 for Z distribution,
the big sample, the Z distribution,
please wait a moment to compute the test statistic database.
H0:lamda=0.800000,
sample size=200, sample mean=        0.6080133352
sample variance=        0.0758751167,
lamda estimated value=        0.7914521738
H0:lamda=0.800000,
The population mean=        0.6122350304,
population variance=        0.0758751167 under H0
test statistic=       -0.2167465456, p value=0.828610,
```

$H_0 : \lambda = c$, $c$=0.5,
the output is

```
the big sample, the Z distribution,
please wait a moment to compute the test statistic database.
H0:lamda=0.500000,
sample size=200, sample mean=        0.6080133352
sample variance=        0.0833475335,
lamda estimated value=        0.7914521738
H0:lamda=0.500000,
The population mean=        0.5000586945,
population variance=        0.0833475335 under H0
test statistic=        5.2882251615, p value=0.000000,
```

6.Continous Bernoulli parameter lamda confidence interval,

$X_1, X_2, ..., X_n \overset{iid}{\sim} CB(\lambda)$,

The statistic = $\dfrac{\sqrt{n}\left(\overline{X} - \mu(X)\right)}{S(X)}, \overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}, S(X) = \sqrt{\dfrac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}}$ ,

$\dfrac{\sqrt{n}\left(\overline{X} - \mu(X)\right)}{S(X)} \xrightarrow[n\to\infty]{} Normal(0,1)$.

The computing the sample size by $\hat{\lambda} = \phi\left(\overline{X}\right)$,

The confidence interval is from Z distribution when the sample size is large sample and the confidence interval is from sampling distribution of $\overline{X}$ when sample size is small sample.

(1) The confidence interval of $\lambda$ for large sample,

$\hat{\lambda} = \phi\left(\overline{X}\right)$,$0.143853919 \le \overline{X} \le 0.856221427$ and $0.001 \le \hat{\lambda} \le 0.999$.

The large sample is $n \ge 33+350\times|\hat{\lambda} = \phi\left(\overline{X}\right)-0.5|$, if $0.1 \le \hat{\lambda} \le 0.9$,

$n \ge 500+15000\times(0.1-\hat{\lambda} = \phi\left(\overline{X}\right))$, if $\hat{\lambda} = \phi\left(\overline{X}\right) < 0.1$,

$n \ge 500+15000\times(\hat{\lambda} = \phi\left(\overline{X}\right)-0.9)$,if $\hat{\lambda} = \phi\left(\overline{X}\right) > 0.9$,

$(1-\alpha)\times 100\%$ C.I. for $E\left(\overline{X}\right) = \mu$

$\overline{X} - Z_{\alpha/2} \times S\left(\overline{X}\right) \le \mu \le \overline{X} + Z_{\alpha/2}\sqrt{S^2\left(\overline{X}\right)}$,

$P(Z > Z_\alpha) = \alpha$, $Z$ is the standard normal distribution,

$(1-\alpha)\times 100\%$ C.I. for $\lambda$

$\phi\left(\overline{X} - Z_{\alpha/2} \times S\left(\overline{X}\right)\right) \le \lambda \le \phi\left(\overline{X} + Z_{\alpha/2} \times S\left(\overline{X}\right)\right)$

(2) The confidence interval of $\lambda$ for small sample,

$n < 33+350\times|\hat{\lambda} = \phi\left(\overline{X}\right)-0.5|$, if $0.1 \le \hat{\lambda} \le 0.9$,

$n < 500+15000\times(0.1-\hat{\lambda} = \phi\left(\overline{X}\right))$, if $\hat{\lambda} = \phi\left(\overline{X}\right) < 0.1$,

$n < 500+15000\times(\hat{\lambda} = \phi\left(\overline{X}\right)-0.9)$,if $\hat{\lambda} = \phi\left(\overline{X}\right) > 0.9$,

$(1-\alpha)\times 100\%$ C.I. for $E\left(\overline{X}\right) = \mu$

$\overline{X} - W_{\alpha/2} \times S\left(\overline{X}\right) \le \mu \le \overline{X} + W_{\alpha/2}\sqrt{S^2\left(\overline{X}\right)}$,

$P(W > W_\alpha) = \alpha$, $W$ is the sampling distribution of $\dfrac{\left(\overline{X} - \mu(X)\right)}{S\left(\overline{X}\right)}$ which can be simulated using the Continuous Bernoulli distribution simulator. The $\lambda$ and sample size will be a specific sampling distribution, the software computing critical value is a essentially way.

The program is
C:\C_Bernoulli\C_Bernoulli_06.cpp,

The input data of C:\C_Bernoulli\C_Bernoulli_06.cpp,
The data filename,

The output data,
The confidence interval of $\lambda$.

example, the sample data is simulated from $CB(\lambda = 0.8)$ and data size=200,

```
sample data size=200
sample size=200, sample mean=        0.6080133352
sample variance=        0.0752478124,
lamda estimated value=        0.7914521738
sample size=200, the requirement size=135 for Z distribution,
the big sample, the Z distribution,
90% C.I. for lamda         0.7161964287<= lamda <=         0.8537022639
95% C.I. for lamda         0.7004447207<= lamda <=         0.8640825009
99% C.I. for lamda         0.6685540313<= lamda <=         0.8829691142
```

example, the sample data is simulated from $CB(\lambda = 0.1)$ and data size=80,

```
sample data size=80
sample size=80, sample mean=        0.3292495861
sample variance=        0.0679395409,
lamda estimated value=        0.0987511464
sample size=80, the requirement size=518 for Z distribution,
the small sample, the testing sampling distribution,
please wait a moment to collect the sampling distribution
90% C.I. for lamda         0.0451617516<= lamda <=         0.1748318583
95% C.I. for lamda         0.0373468527<= lamda <=         0.1920987471
99% C.I. for lamda         0.0241198888<= lamda <=         0.2287191354
```

7.The goodness of fit when H0: Continuous Bernoulli distribution,
lamda known from input data,

$X_1, X_2, ..., X_n \overset{iid}{\sim} CB(\lambda)$, ,

$H_0$:X~Continuous Bernoulli( $\lambda$ ) and $\lambda$ is known,

$H_1$:against $H_0$,

The test process,
The frequency distribution setting,
(i)The class number and the probability of each class,

The class number= $k = log_2(n) + 1$, each class probability is setting to $\frac{1}{k}$.

(ii)The class limit,
The first class lower limit=0 and the last class upper limit=1.

$$c_j = \begin{cases} \dfrac{log_e\left(\dfrac{j}{k} \times (2\lambda - 1) - (\lambda - 1)\right) - log_e(1-\lambda)}{log_e\left(\dfrac{\lambda}{1-\lambda}\right)}, \hat{\lambda} \neq \dfrac{1}{2} \\ \dfrac{j}{k}, \hat{\lambda} = \dfrac{1}{2} \end{cases}, j = 1,2,...,k-1, ,$$

The first class upper limit= $c_1$ =the second class lower limit,…..,
The $j-th$ class upper limit= $c_j$ =the $(j+1)-th$ class lower limit, $j = 1,2,...,k-1$.

(iii)The frequency table for testing and computing the observed number and expected number,

| class | class limit | frequency= $O$ | $E = n \times \frac{1}{k}$ |
|---|---|---|---|
| 1 | 0~ $c_1$ | $O_1$ | $E_1$ |
| 2 | $c_1$ ~ $c_2$ | $O_2$ | $E_2$ |
| … | | | |
| k | $c_{k-1}$ ~1 | $O_k$ | $E_k$ |

The chi square test statistic,

$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} > \chi^2_{\alpha,k-1}$, rejected $H_0$.

The program is
C:\C_Bernoulli\C_Bernoulli_07.cpp,

The input data of C:\C_Bernoulli\C_Bernoulli_07.cpp,
The data filename and Continuous Bernoulli( $\lambda$ ) $\lambda$ =?

The output data,
The chi square test result.
example, the sample data is simulated from $CB(\lambda = 0.8)$ and data size=200,

input $\lambda = 0.8$,

```
sample data size=200
H0:Continuous Bernoulli(lamda),lamda=0.8
Collecting chi square(df=7) database,...
the sample mean=        0.6080133352
the lamda estimated=0.791452.

 pearson goodness of fit
H0:X~Continuous Bernoulli(lamda),
H1:against H0,
lamda under H0=        0.8000000000
 pearson goodness of fit

degree of freedom=7
chi square test=5.760000
p value=0.568421
```

example, the sample data is simulated from $CB(\lambda = 0.8)$ and data size=200,
input $\lambda = 0.58$,

```
sample data size=200
H0:Continuous Bernoulli(lamda),lamda=0.58
Collecting chi square(df=7) database,...
the sample mean=        0.6080133352
the lamda estimated=0.791452.

 pearson goodness of fit
H0:X~Continuous Bernoulli(lamda),
H1:against H0,
lamda under H0=        0.5800000000
 pearson goodness of fit

degree of freedom=7
chi square test=17.600000
p value=0.014054
```

8.The goodness of fit when H0: Continuous Bernoulli distribution,
lamda unknown and MLE estimated equation using input data,

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} CB(\lambda), \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n},$$

$H_0$:Continuous Bernoulli( $\hat{\lambda}$ ), $H_1$:against $H_0$,

$\hat{\lambda} = \phi(\overline{X})$ is the estimated equation of the $\lambda$.

The test process,

(i)The class number and the probability of each class,

The class number= $k = log_2(n) + 1$, each class probability is setting to $\frac{1}{k}$.

(ii)The class limit,

The first class lower limit=0 and the last class upper limit=1.

$$c_j = \begin{cases} \dfrac{log_e\left(\dfrac{j}{k} \times (2\hat{\lambda} - 1) - (\hat{\lambda} - 1)\right) - log_e(1 - \hat{\lambda})}{log_e\left(\dfrac{\hat{\lambda}}{1 - \hat{\lambda}}\right)}, \hat{\lambda} \neq \dfrac{1}{2} \\ \dfrac{j}{k}, \hat{\lambda} = \dfrac{1}{2} \end{cases}, j = 1,2,\ldots,k-1,,$$

The first class upper limit= $c_1$ =the second class lower limit,…..,

The $j - th$ class upper limit= $c_j$ =the $(j+1) - th$ class lower limit, $j = 1,2,\ldots,k-1$.

(iii)The frequency table for testing and computing the observed number and expected number,

| class | class limit | frequency= $O$ | $E = n \times \frac{1}{k}$ |
|---|---|---|---|
| 1 | 0~ $c_1$ | $O_1$ | $E_1$ |
| 2 | $c_1$ ~ $c_2$ | $O_2$ | $E_2$ |
| … | | | |
| k | $c_{k-1}$ ~1 | $O_k$ | $E_k$ |

The chi square test statistic,

$\chi^2_{k-2} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} > \chi^2_{\alpha,k-2}$ , rejected $H_0$.

The program is
C:\C_Bernoulli\C_Bernoulli_08.cpp,

The input data of C:\C_Bernoulli\C_Bernoulli_08.cpp,
The data filename,

The output data,
The chi square test result.

example, the sample data is simulated from $CB(\lambda = 0.8)$ and data size=200,

```
sample data size=200
Collecting chi square(df=6) database,...
the sample mean=        0.6080133352
the lamda estimated=0.791452.

 pearson goodness of fit
H0:X~Continuous Bernoulli(lamda),lamda unknown
H1:against H0,
lamda estimated=0.791452
 pearson goodness of fit

degree of freedom=6
chi square test=4.880000
p value=0.559800
```

example, the sample data is simulated from $CB(\lambda = 0.1)$ and data size=80,

```
sample data size=80
Collecting chi square(df=5) database,...
the sample mean=        0.3292495861
the lamda estimated=0.098751.

 pearson goodness of fit
H0:X~Continuous Bernoulli(lamda),lamda unknown
H1:against H0,
lamda estimated=0.098751
 pearson goodness of fit

degree of freedom=5
chi square test=3.300000
p value=0.653653
```

9. Two independent Continuous Bernoulli distribution testing,
$H_0: \mu_1 = \mu_2$+c, c$\neq 0$,

$\mu_1$ is the 1st population mean, $\mu_2$ is the 2nd population mean.
There are two independent Continuous Bernoulli populations,

$$X_{1,1}, X_{\backslash 1,2}, ..., X_{1,n_1} \overset{iid}{\sim} CB(\lambda_1), \overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_{1,i}}{n_1}, S_1 = \sqrt{\frac{\sum_{i=1}^{n_1}\left(X_{1,i} - \overline{X}_1\right)^2}{n_1 - 1}},$$

$\mu_1 = G_1(\lambda_1)$,

$$X_{2,1}, X_{\backslash 2,2}, ..., X_{2,n_2} \overset{iid}{\sim} CB(\lambda_2), \overline{X}_2 = \frac{\sum_{j=1}^{n_2} X_{2,j}}{n_2}, S_2 = \sqrt{\frac{\sum_{j=1}^{n_2}\left(X_{2,j} - \overline{X}_2\right)^2}{n_2 - 1}},$$

$\mu_2 = G_1(\lambda_2)$,
but $\lambda_1$ and $\lambda_2$ are unknown, $\hat{\lambda}_1 = \phi(\overline{X}_1), \hat{\lambda}_2 = \phi(\overline{X}_2)$, and $\lambda_1 = \phi(\mu_1), \lambda_2 = \phi(\mu_2)$.

If $\mu_1 \neq \mu_2$, $\lambda_1 \neq \lambda_2$,
the test statistic= $\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

(1) The large sample,
$\hat{\lambda}_1 = \phi(\overline{X}_1)$,$0.143853919 \leq \overline{X}_1 \leq 0.856221427$ and $0.001 \leq \hat{\lambda}_1 \leq 0.999$.
The large sample is $n_1 \geq 33+350\times|\hat{\lambda}_1$-0.5|, if $0.1 \leq \hat{\lambda}_1 \leq 0.9$,
$n_1 \geq 500+15000\times(0.1-\hat{\lambda}_1)$, if $\hat{\lambda}_1 < 0.1$,
$n_1 \geq 500+15000\times(\hat{\lambda}_1-0.9)$,if $\hat{\lambda}_1 > 0.9$,
and
$\hat{\lambda}_2 = \phi(\overline{X}_2)$,$0.143853919 \leq \overline{X}_2 \leq 0.856221427$ and $0.001 \leq \hat{\lambda}_2 \leq 0.999$.
The large sample is $n_2 \geq 33+350\times|\hat{\lambda}_2$-0.5|, if $0.1 \leq \hat{\lambda}_2 \leq 0.9$,
$n_2 \geq 500+15000\times(0.1-\hat{\lambda}_2)$, if $\hat{\lambda}_2 < 0.1$,
$n_2 \geq 500+15000\times(\hat{\lambda}_2-0.9)$,if $\hat{\lambda}_2 > 0.9$,

$H_0: \mu_1 = \mu_2$+c, c$\neq 0$,

$$Z^* = \frac{\overline{X}_1 - \overline{X}_2 - c}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \longrightarrow Z \text{ (standard normal distribution)},$$

$Z^* > Z_{\alpha/2}$, $H_0$ is rejected.

p value= $2\times P(Z \leq Z^*)$, if $P(Z \leq Z^*) < 0.5$
p value= $2\times(1 - P(Z \leq Z^*))$, if $P(Z \leq Z^*) \geq 0.5$

(2) The small sample,

$\hat{\lambda}_1 = \phi\left(\overline{X}_1\right)$,0.143853919$\le \overline{X}_1 \le$0.856221427 and 0.001$\le \hat{\lambda}_1 \le$0.999.

The large sample is $n_1 <$33+350$\times|\hat{\lambda}_1$-0.5|, if 0.1$\le \hat{\lambda}_1 \le$0.9,

$n_1 <$500+15000$\times$(0.1-$\hat{\lambda}_1$), if $\hat{\lambda}_1 <$0.1,

$n_1 <$500+15000$\times$($\hat{\lambda}_1$-0.9),if $\hat{\lambda}_1$>0.9,

or

$\hat{\lambda}_2 = \phi\left(\overline{X}_2\right)$,0.143853919$\le \overline{X}_2 \le$0.856221427 and 0.001$\le \hat{\lambda}_2 \le$0.999.

The large sample is $n_2 <$33+350$\times|\hat{\lambda}_2$-0.5|, if 0.1$\le \hat{\lambda}_2 \le$0.9,

$n_2 <$500+15000$\times$(0.1-$\hat{\lambda}_2$), if $\hat{\lambda}_2 <$0.1,

$n_2 <$500+15000$\times$($\hat{\lambda}_2$-0.9),if $\hat{\lambda}_2$>0.9,

$H_0: \mu_1 = \mu_2$+c, c$\ne$0,

$$W^* = \frac{\overline{X}_1 - \overline{X}_2 - c}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}},$$

the sampling distribution of $W = \frac{\overline{X}_1 - \overline{X}_2 - (\hat{\mu}_1 - \hat{\mu}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ will be simulated using the probability simulator and $\hat{\mu}_1 = G_1\left(\hat{\lambda}_1\right)$ and $\hat{\mu}_2 = G_1\left(\hat{\lambda}_2\right)$,

the simulated data is based on

$$X_{1,1}, X_{\backslash 1,2}, \ldots, X_{1,n_1} \overset{iid}{\sim} CB\left(\hat{\lambda}_1\right), \overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_{1,i}}{n_1}, S_1 = \sqrt{\frac{\sum_{i=1}^{n_1}\left(X_{1,i} - \overline{X}_1\right)^2}{n_1 - 1}},$$

$$X_{2,1}, X_{\backslash 2,2}, \ldots, X_{2,n_2} \overset{iid}{\sim} CB\left(\hat{\lambda}_2\right), \overline{X}_2 = \frac{\sum_{j=1}^{n_2} X_{2,j}}{n_2}, S_2 = \sqrt{\frac{\sum_{j=1}^{n_2}\left(X_{2,j} - \overline{X}_2\right)^2}{n_2 - 1}},$$

p value= $2\times P\left(W \le W^*\right)$, if $P\left(Z \le Z^*\right) < 0.5$

p value= $2\times\left(1 - P\left(W \le W^*\right)\right)$, if $P\left(Z \le Z^*\right) \ge 0.5$

(3)The $\lambda_1$ and $\lambda_2$ estimated value,

(i) $H_0: \mu_1 = \mu_2$+c, c$\ne$0 is rejected,

$\hat{\lambda}_1 = \phi\left(\overline{X}_1\right)$, $\hat{\lambda}_2 = \phi\left(\overline{X}_2\right)$.

(ii) $H_0: \mu_1 = \mu_2$+c, c$\ne$0 is not rejected,

$$\hat{\lambda}_1 = \phi\left(\frac{\sum_{i=1}^{n_1} X_{1,i} + \sum_{j=1}^{n_2}\left(X_{2,j} + c\right)}{n_1 + n_2}\right), \quad \hat{\lambda}_2 = \phi\left(\frac{\sum_{i=1}^{n_1}\left(X_{1,i} - c\right) + \sum_{j=1}^{n_2} X_{2,j}}{n_1 + n_2}\right).$$

The program is
C:\C_Bernoulli\C_Bernoulli_09.cpp,

The input data of C:\C_Bernoulli\C_Bernoulli_09.cpp,
The two data filenames, $H_0: \mu_1 = \mu_2$ +c, c=?

The output data,
The test result.

example, the 1st sample data is simulated from $CB(\lambda_1 = 0.8)$ and data size=200, the 2nd sample data is simulated from $CB(\lambda_2 = 0.1)$ and data size=80, two population distributions are independently. $\mu_1 = \phi(\lambda_1 = 0.8)$=0.612, $\mu_2 = \phi(\lambda_2 = 0.1)$=0.33015.

```
The 1st data,......
sample size=200, sample mean=        0.6080133352
sample variance=        0.0752478124,
lamda1 estimated value=        0.7914521738
The 2nd data,......
sample size=80, sample mean=        0.3292495861
sample variance=        0.0679395409,
lamda2 estimated value=        0.0987511464
1st sample size=200, the requirement size=135 for Z distribution
2nd sample size=80, the requirement size=518 for Z distribution,
the small sample, the testing sampling distribution,
please wait a moment to collect the sampling distribution
```

$H_0: \mu_1 = \mu_2$ +0.28185,

```
H0:mu1=mu2+0.281850,
The 1st population estimated mean=        0.6082479697,
estimated population variance=        0.0763912935
The 2nd population estimated mean=        0.3288993307,
estimated population variance=        0.0662856627
test statistic=       -0.0881612084, p value=0.935006,
lamda2 hat=0.770290 and lamd2 hat=0.082790 when H0:mu1=mu2+0.281850.
```

10.Two independent Continuous Bernoulli distribution testing,

$H_0 : \mu_1 = \mu_2$,

$\mu_1$ is the 1st population mean, $\mu_2$ is the 2nd population mean.

There are two independent Continuous Bernoulli populations,

$$X_{1,1}, X_{1,2}, \ldots, X_{1,n_1} \overset{iid}{\sim} CB(\lambda_1), \overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_{1,i}}{n_1}, S_1 = \sqrt{\frac{\sum_{i=1}^{n_1}\left(X_{1,i} - \overline{X}_1\right)^2}{n_1 - 1}},$$

$\mu_1 = G_1(\lambda_1)$,

$$X_{2,1}, X_{2,2}, \ldots, X_{2,n_2} \overset{iid}{\sim} CB(\lambda_2), \overline{X}_2 = \frac{\sum_{j=1}^{n_2} X_{2,j}}{n_2}, S_2 = \sqrt{\frac{\sum_{j=1}^{n_2}\left(X_{2,j} - \overline{X}_2\right)^2}{n_2 - 1}},$$

$\mu_2 = G_1(\lambda_2)$,

but $\lambda_1$ and $\lambda_2$ are unknown, $\hat{\lambda}_1 = \phi(\overline{X}_1), \hat{\lambda}_2 = \phi(\overline{X}_2)$, and $\lambda_1 = \phi(\mu_1), \lambda_2 = \phi(\mu_2)$.

If $\mu_1 = \mu_2$, $\lambda_1 = \lambda_2 = \lambda$,

$$\overline{\overline{X}} = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2}{n_1 + n_2}, S_p^2 = \frac{\sum_{i=1}^{n_1}\left(X_{1,i} - \overline{\overline{X}}\right)^2 + \sum_{j=1}^{n_2}\left(X_{2,j} - \overline{\overline{X}}\right)^2}{n_1 + n_2 - 1},$$

the test statistic= $\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_p^2}{n_1} + \dfrac{S_p^2}{n_2}}}$

(1) The large sample,

$\hat{\lambda} = \phi\left(\overline{\overline{X}}\right)$,$0.143853919 \le \overline{\overline{X}} \le 0.856221427$ and $0.001 \le \hat{\lambda} \le 0.999$.

The large sample is $n_1 + n_2 \ge 33+350\times|\hat{\lambda}$-0.5|, if $0.1 \le \hat{\lambda} \le 0.9$,

$n_1 + n_2 \ge 500+15000\times(0.1-\hat{\lambda}_1)$, if $\hat{\lambda} < 0.1$,

$n_1 + n_2 \ge 500+15000\times(\hat{\lambda}_1-0.9)$,if $\hat{\lambda} > 0.9$,

$H_0 : \mu_1 = \mu_2$,

$$Z^* = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{S_p^2}{n_1} + \dfrac{S_p^2}{n_2}}} \longrightarrow Z \text{ (standard normal distribution)},$$

$Z^* > Z_{\alpha/2}$, $H_0$ is rejected.

p value= $2\times P\left(Z \le Z^*\right)$, if $P\left(Z \le Z^*\right) < 0.5$

p value= $2\times\left(1 - P\left(Z \le Z^*\right)\right)$, if $P\left(Z \le Z^*\right) \ge 0.5$

(2) The small sample,

$\hat{\lambda} = \phi\left(\overline{\overline{X}}\right)$,0.143853919 $\le \overline{\overline{X}} \le$ 0.856221427 and 0.001 $\le \hat{\lambda} \le$ 0.999.

The large sample is $n_1 + n_2 <$ 33+350 $\times |\hat{\lambda}$ -0.5|, if 0.1 $\le \hat{\lambda} \le$ 0.9,

$n_1 + n_2 <$ 500+15000 $\times$ (0.1- $\hat{\lambda}_1$ ), if $\hat{\lambda} <$ 0.1,

$n_1 + n_2 <$ 500+15000 $\times$ ( $\hat{\lambda}_1$ -0.9),if $\hat{\lambda}$ >0.9,

$H_0 : \mu_1 = \mu_2$,

$$W^* = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}},$$

the sampling distribution of $W = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$ will be simulated using the probability simulator and $\hat{\lambda} = \phi\left(\overline{\overline{X}}\right)$,

the simulated data is based on

$$X_{1,1}, X_{1,2}, \ldots, X_{1,n_1} \overset{iid}{\sim} CB\left(\hat{\lambda}\right), X_{2,1}, X_{2,2}, \ldots, X_{2,n_2} \overset{iid}{\sim} CB\left(\hat{\lambda}\right),$$

$$\overline{\overline{X}} = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2}{n_1 + n_2}, S_p^2 = \frac{\sum_{i=1}^{n_1}\left(X_{1,i} - \overline{\overline{X}}\right)^2 + \sum_{j=1}^{n_2}\left(X_{2,j} - \overline{\overline{X}}\right)^2}{n_1 + n_2 - 1},$$

p value= $2 \times P\left(W \le W^*\right)$, if $P\left(Z \le Z^*\right) < 0.5$

p value= $2 \times \left(1 - P\left(W \le W^*\right)\right)$, if $P\left(Z \le Z^*\right) \ge 0.5$

(3)The $\lambda_1$ and $\lambda_2$ estimated value,

(i) $H_0 : \mu_1 = \mu_2$ is rejected,

$\hat{\lambda}_1 = \phi\left(\overline{X}_1\right)$, $\hat{\lambda}_2 = \phi\left(\overline{X}_2\right)$.

(ii) $H_0 : \mu_1 = \mu_2 \neq$ is not rejected,

$\hat{\lambda}_1 = \hat{\lambda}_2 = \hat{\lambda} = \phi\left(\overline{\overline{X}}\right)$,.

The program is
C:\C_Bernoulli\C_Bernoulli_10.cpp,

The input data of C:\C_Bernoulli\C_Bernoulli_10.cpp,
The two data filenames,

The output data,
The test result.

example, the 1st sample data is simulated from $CB(\lambda_1 = 0.8)$ and data size=200, the 2nd sample data is simulated from $CB(\lambda_2 = 0.8)$ and data size=80, two population distributions are independently.

$H_0 : \mu_1 = \mu_2$

```
The 1st data,......
sample size=200, sample mean=       0.6080133352
sample variance=       0.0752478124,
lamda1 estimated value=       0.7914521738
The 2nd data,......
sample size=80, sample mean=       0.5788463517
sample variance=       0.0687515636,
lamda2 estimated value=       0.7231317654
Under H0:lamda1=lamda2,.....
sample size=280, sample mean=       0.5996799113
sample variance=       0.0733129045,
lamda1 estimated value=       0.7729998259
sample size=280, the requirement size=128 for Z distribution,
H0:lamda1=lamda2,
The population estimated mean=       0.5997897469,
population estimated variance=       0.0774275538 under H0
test statistic=       1.2745747796, p value=0.203006,
lamda1 hat=lamda2 hat=lamda hat=0.773000 hen H0:lamda1=lamda2 not rejected.
```

example, the 1st sample data is simulated from $CB(\lambda_1 = 0.8)$ and data size=200, the 2nd sample data is simulated from $CB(\lambda_2 = 0.1)$ and data size=80, two population distributions are independently.

$H_0 : \mu_1 = \mu_2$

```
The 1st data,......
sample size=200, sample mean=       0.6080133352
sample variance=       0.0752478124,
lamda1 estimated value=       0.7914521738
The 2nd data,......
sample size=80, sample mean=       0.3292495861
sample variance=       0.0679395409,
lamda2 estimated value=       0.0987511464
Under H0:lamda1=lamda2,.....
sample size=280, sample mean=       0.5283665497
sample variance=       0.0888246082,
lamda1 estimated value=       0.5842938463
sample size=280, the requirement size=62 for Z distribution,
H0:lamda1=lamda2,
The population estimated mean=       0.5282514093,
population estimated variance=       0.0828701866 under H0
test statistic=       11.0670956682, p value=0.000000,
lamda1 hat=0.791452, lamda2 hat=0.098751 hen H0:lamda1=lamda2 rejected.
```

11.Two independent Continuous Bernoulli distribution confidence interval of $\mu_1-\mu_2$ and $\lambda_1-\lambda_2$

$\mu_1$ is the 1st population mean, $\mu_2$ is the 2nd population mean.
There are two independent Continuous Bernoulli populations,

$$X_{1,1},X_{\backslash 1,2},...,X_{1,n_1}\overset{iid}{\sim}CB(\lambda_1),\overline{X}_1=\frac{\sum_{i=1}^{n_1}X_{1,i}}{n_1},S_1=\sqrt{\frac{\sum_{i=1}^{n_1}\left(X_{1,i}-\overline{X}_1\right)^2}{n_1-1}},$$

$\mu_1=G_1(\lambda_1)$,

$$X_{2,1},X_{\backslash 2,2},...,X_{2,n_2}\overset{iid}{\sim}CB(\lambda_2),\overline{X}_2=\frac{\sum_{j=1}^{n_2}X_{2,j}}{n_2},S_2=\sqrt{\frac{\sum_{j=1}^{n_2}\left(X_{2,j}-\overline{X}_2\right)^2}{n_2-1}},$$

$\mu_2=G_1(\lambda_2)$,
but $\lambda_1$ and $\lambda_2$ are unknown, $\hat{\lambda}_1=\phi(\overline{X}_1),\hat{\lambda}_2=\phi(\overline{X}_2)$, and $\lambda_1=\phi(\mu_1),\lambda_2=\phi(\mu_2)$.

If $\mu_1\neq\mu_2$, $\lambda_1\neq\lambda_2$,
the statistic= $\dfrac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$

(1) The large sample,
$\hat{\lambda}_1=\phi(\overline{X}_1)$,$0.143853919\leq\overline{X}_1\leq 0.856221427$ and $0.001\leq\hat{\lambda}_1\leq 0.999$.
The large sample is $n_1\geq 33+350\times|\hat{\lambda}_1-0.5|$, if $0.1\leq\hat{\lambda}_1\leq 0.9$,
$n_1\geq 500+15000\times(0.1-\hat{\lambda}_1)$, if $\hat{\lambda}_1<0.1$,
$n_1\geq 500+15000\times(\hat{\lambda}_1-0.9)$,if $\hat{\lambda}_1>0.9$,
and
$\hat{\lambda}_2=\phi(\overline{X}_2)$,$0.143853919\leq\overline{X}_2\leq 0.856221427$ and $0.001\leq\hat{\lambda}_2\leq 0.999$.
The large sample is $n_2\geq 33+350\times|\hat{\lambda}_2-0.5|$, if $0.1\leq\hat{\lambda}_2\leq 0.9$,
$n_2\geq 500+15000\times(0.1-\hat{\lambda}_2)$, if $\hat{\lambda}_2<0.1$,
$n_2\geq 500+15000\times(\hat{\lambda}_2-0.9)$,if $\hat{\lambda}_2>0.9$,

$$\frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}\longrightarrow Z \text{ (standard normal distribution),}$$

$(1-\alpha)\times 100\%$ C.I. of $\mu_1-\mu_2$

$$\overline{X}_1-\overline{X}_2-Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\leq\mu_1-\mu_2\leq\overline{X}_1-\overline{X}_2+Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

(2) The small sample,

$\hat{\lambda}_1 = \phi(\overline{X}_1)$,0.143853919 $\le \overline{X}_1 \le$ 0.856221427 and 0.001 $\le \hat{\lambda}_1 \le$ 0.999.

The large sample is $n_1 <$ 33+350 $\times$ | $\hat{\lambda}_1$ -0.5|, if 0.1 $\le \hat{\lambda}_1 \le$ 0.9,

$n_1 <$ 500+15000 $\times$ (0.1- $\hat{\lambda}_1$), if $\hat{\lambda}_1 <$ 0.1,

$n_1 <$ 500+15000 $\times$ ( $\hat{\lambda}_1$ -0.9),if $\hat{\lambda}_1$ >0.9,

or

$\hat{\lambda}_2 = \phi(\overline{X}_2)$,0.143853919 $\le \overline{X}_2 \le$ 0.856221427 and 0.001 $\le \hat{\lambda}_2 \le$ 0.999.

The large sample is $n_2 <$ 33+350 $\times$ | $\hat{\lambda}_2$ -0.5|, if 0.1 $\le \hat{\lambda}_2 \le$ 0.9,

$n_2 <$ 500+15000 $\times$ (0.1- $\hat{\lambda}_2$), if $\hat{\lambda}_2 <$ 0.1,

$n_2 <$ 500+15000 $\times$ ( $\hat{\lambda}_2$ -0.9),if $\hat{\lambda}_2$ >0.9,

the statistic= $\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ ,

the sampling distribution of $W = \dfrac{\overline{X}_1 - \overline{X}_2 - (\hat{\mu}_1 - \hat{\mu}_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ will be simulated using the probability simulator and $\hat{\mu}_1 = G_1(\hat{\lambda}_1)$ and $\hat{\mu}_2 = G_1(\hat{\lambda}_2)$,

the simulated data is based on

$$X_{1,1}, X_{\backslash 1,2}, \ldots, X_{1,n_1} \overset{iid}{\sim} CB(\hat{\lambda}_1), \overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_{1,i}}{n_1}, S_1 = \sqrt{\frac{\sum_{i=1}^{n_1}(X_{1,i} - \overline{X}_1)^2}{n_1 - 1}},$$

$$X_{2,1}, X_{\backslash 2,2}, \ldots, X_{2,n_2} \overset{iid}{\sim} CB(\hat{\lambda}_2), \overline{X}_2 = \frac{\sum_{j=1}^{n_2} X_{2,j}}{n_2}, S_2 = \sqrt{\frac{\sum_{j=1}^{n_2}(X_{2,j} - \overline{X}_2)^2}{n_2 - 1}},$$

$(1-\alpha)\times 100\%$ C.I. of $\mu_1 - \mu_2$

$$\overline{X}_1 - \overline{X}_2 + W_{1-\alpha}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \le \mu_1 - \mu_2 \le \overline{X}_1 - \overline{X}_2 + W_{\alpha}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$P(W > W_\alpha) = \alpha,$

Note: $(1-\alpha)\times 100\%$ C.I. of $\mu_1 - \mu_2$ cannot convert to
$(1-\alpha)\times 100\%$ C.I. of $\lambda_1 - \lambda_2$.

Let $\hat{\lambda}_2 = \phi(\overline{X}_2), \hat{\lambda}_{L,1} = \phi\left(\overline{X}_1 + W_{1-\alpha}\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}\right), \hat{\lambda}_{U,1} = \phi\left(\overline{X}_1 + W_{\alpha}\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}\right),$

$(1-\alpha)\times 100\%$ C.I. of $\lambda_1 - \lambda_2$

$\hat{\lambda}_{L,1} - \hat{\lambda}_2 \le \lambda_1 - \lambda_2 \le \hat{\lambda}_{U,1} - \hat{\lambda}_2$

The program is
C:\C_Bernoulli\C_Bernoulli_11.cpp,

The input data of C:\C_Bernoulli\C_Bernoulli_11.cpp,
The two data filenames,

The output data,
The confidence intervals of $\mu_1 - \mu_2$ and $\lambda_1 - \lambda_2$.

example, the 1st sample data is simulated from $CB(\lambda_1 = 0.8)$ and data size=200,
the 2nd sample data is simulated from $CB(\lambda_2 = 0.8)$ and data size=80, two population distributions are independently.
$\mu_1$-$\mu_2$=0, $\lambda_1 - \lambda_2 = 0$.

```
The 1st data,......
sample size=200, sample mean=        0.6080133352
sample variance=        0.0752478124,
lamda estimated value=        0.7914521738
The 2nd data,......
sample size=80, sample mean=        0.5788463517
sample variance=        0.0687515636,
lamda estimated value=        0.7231317654
1st sample size=200, the requirement size=135 for Z distribution,
2nd sample size=80, the requirement size=111 for Z distribution,
the small sample, the testing sampling distribution,
please wait a moment to collect the sampling distribution
90% C.I. for mu1-mu2
       -0.0301412341<= mu1- mu2 <=        0.0571767441
95% C.I. for mu1-mu2
       -0.0417022122<= mu1- mu2 <=        0.0973708615
99% C.I. for mu1-mu2
       -0.0646936007<= mu1- mu2 <=        0.1192814401
90% C.I. for lamda1-lamda2
       -0.0804110874<=  lamda1-lamda2  <=        0.1703462577
95% C.I. for lamda1-lamda2
       -0.1133054925<=  lamda1-lamda2  <=        0.1851613740
99% C.I. for lamda1-lamda2
       -0.1809132397<=  lamda1-lamda2  <=        0.2103092503
```

example, the 1st sample data is simulated from $CB(\lambda_1 = 0.8)$ and data size=200, the 2nd sample data is simulated from $CB(\lambda_2 = 0.1)$ and data size=80, two population distributions are independently.

$\mu_1$ - $\mu_2$ =0.28185, $\lambda_1 - \lambda_2 = 0.7$ .

```
The 1st data,......
sample size=200, sample mean=        0.6080133352
sample variance=        0.0752478124,
lamda estimated value=        0.7914521738
The 2nd data,......
sample size=80, sample mean=        0.3292495861
sample variance=        0.0679395409,
lamda estimated value=        0.0987511464
1st sample size=200, the requirement size=135 for Z distribution
2nd sample size=80, the requirement size=518 for Z distribution,
the small sample, the testing sampling distribution,
please wait a moment to collect the sampling distribution
90% C.I. for mu1-mu2
        0.2221816330<= mu1- mu2 <=        0.0594825475
95% C.I. for mu1-mu2
        0.2116273140<= mu1- mu2 <=        0.3503318726
99% C.I. for mu1-mu2
        0.1912495534<= mu1- mu2 <=        0.3747772556
90% C.I. for lamda1-lamda2
        0.5515828805<=  lamda1-lamda2  <=        0.7979491137
95% C.I. for lamda1-lamda2
        0.5217961783<=  lamda1-lamda2  <=        0.8137668442
99% C.I. for lamda1-lamda2
        0.4623429739<=  lamda1-lamda2  <=        0.8405402287
```

12. The one way analysis and populations are Continuous Bernoulli,
There are k independent Continuous Bernoulli distributions, the random samples from each population and the same size.

$$X_{1,1}, X_{1,2}, \ldots, X_{1,n} \overset{iid}{\sim} CB(\lambda_1),$$

$$X_{2,1}, X_{2,2}, \ldots, X_{2,n} \overset{iid}{\sim} CB(\lambda_2),$$

……………………,

$$X_{k,1}, X_{k,2}, \ldots, X_{k,n} \overset{iid}{\sim} CB(\lambda_k),$$

$$X_{i,j} = \mu + \alpha_i + \varepsilon_{ij}, i = 1,2,\ldots,k, j = 1,2,\ldots., n,$$

$$X_{i,j} \overset{iid}{\sim} CB(E(X_{ij})), \quad E(X_{ij}) = \mu_i = \mu + \alpha_i = \mathrm{G}_1(\lambda_i), i = 1,2,\ldots,k,$$

$$H_0 : \lambda_1 = \lambda_2 = \ldots = \lambda_k, (\mu_1 = \mu_2 = \ldots = \mu_k = \mu), (\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0),$$

$$\overline{X}_i = \frac{\sum_{j=1}^{n} X_{i,j}}{n}, S_1^2 = \frac{\sum_{j=1}^{n} \left(X_{i,j} - \overline{X}_i\right)^2}{n-1}, i = 1,2,\ldots,k,$$

The grand mean $\overline{\overline{X}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} X_{i,j}}{n_{\mathrm{T}}}, n_{\mathrm{T}} = n \times k,$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{i,j} - \overline{\overline{X}}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{i,j} - \overline{X}_i + \overline{X}_i - \overline{\overline{X}}\right)^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{i,j} - \overline{X}_i\right)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}\left(\overline{X}_i - \overline{\overline{X}}\right)^2,$$

$$SSTR = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(\overline{X}_i - \overline{\overline{X}}\right)^2, SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{i,j} - \overline{X}_i\right)^2,$$

SST degree of freedom = $n_{\mathrm{T}}$-1, SSTR degree of freedom=k-1,
SSE degree of free= $n_{\mathrm{T}}$-k, MSTR=SSTR/(k-1), MSE=SSE/( $n_{\mathrm{T}}$ -k).
The test statistic=MSTR/MSE and the rejected region is the right region.
The p vlaue=P(MSTR/MSE>W), p vlaue< $\alpha$ , rejected H0.
W~MSTR/MSE probability distribution.

the sampling distribution of $W$ will be simulated using the probability simulator and the simulated data is based on

$$X_{2,1}, X_{2,2}, \ldots, X_{2,n} \overset{iid}{\sim} CB(\hat{\lambda}), \hat{\lambda} = \phi\left(\overline{\overline{X}}\right),$$

$$X_{2,1}, X_{2,2}, \ldots, X_{2,n} \overset{iid}{\sim} CB(\hat{\lambda}),$$ ……………………,

$$X_{k,1}, X_{k,2}, \ldots, X_{k,n} \overset{iid}{\sim} CB(\hat{\lambda}),$$

$$SST=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{i,j}-\overline{\overline{X}}\right)^2=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{i,j}-\overline{X}_i+\overline{X}_i-\overline{\overline{X}}\right)^2$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{i,j}-\overline{X}_i\right)^2+\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\overline{X}_i-\overline{\overline{X}}\right)^2,$$

$$SSTR=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\overline{X}_i-\overline{\overline{X}}\right)^2,SSE=\sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{i,j}-\overline{X}_i\right)^2,W=\frac{MSTR}{MSE}$$

The program is
C:\C_Bernoulli\C_Bernoulli_12.cpp,

The input data of C:\C_Bernoulli\C_Bernoulli_12.cpp,
The treatment number, the lamda of each treatment and the sample size.
The sample data is the simulated data.

The output data,
ANOVA and p value.

example, k=4(treatment number), lamda1=lamda2=lamda3=lamda4=0.2, sample size of each treatment=20,

```
 sample size of each treatment>20
X1~Continuous Bernoulli(lamda1=0.200000),
X2~Continuous Bernoulli(lamda2=0.200000),
X3~Continuous Bernoulli(lamda3=0.200000),
X4~Continuous Bernoulli(lamda4=0.200000),
 The treatment number=4,
 The sample size of each treatment=20,
 The sample mean of each treatment,
 Treatment 1 sample mean=       0.3530010709,
 Treatment 2 sample mean=       0.3311762079,
 Treatment 3 sample mean=       0.3055671484,
 Treatment 4 sample mean=       0.3973718211,
 grand mean=        0.3467790620, the grand lamda estimated=       0.1242176794

 The ANOVA,
 Treatment:SSTR=        0.0908042309, df=3, MSTR=        0.0302680770,
 Error:    SSE =        5.5888935086, df=76, MSE=        0.0735380725,
 Total:    SST =        5.6796977395, df=79,
 Please wait a moment to get the test statistic sampling distribution,..
 H0:lamda1=lamda2=......lamda4,
 R2=        0.0159875111, F test=MSTR/MSE=        0.4115973666
 p value=0.745496,
```

example, k=5(treatment number), lamda1=0.1,lamda2=0.2,lamda3=03,lamda4=0.4, lamda5=0.5,sample size of each treatment=50,

```
sample size of each treatment>50
X1~Continuous Bernoulli(lamda1=0.100000),
X2~Continuous Bernoulli(lamda2=0.200000),
X3~Continuous Bernoulli(lamda3=0.300000),
X4~Continuous Bernoulli(lamda4=0.400000),
X5~Continuous Bernoulli(lamda5=0.500000),
The treatment number=5,
The sample size of each treatment=50,
The sample mean of each treatment,
Treatment 1 sample mean=        0.2816803809,
Treatment 2 sample mean=        0.3785473072,
Treatment 3 sample mean=        0.4090435329,
Treatment 4 sample mean=        0.4559416525,
Treatment 5 sample mean=        0.4945801495,
grand mean=        0.4039586046, the grand lamda estimated=        0.235142020

The ANOVA,
Treatment:SSTR=        1.3269028100, df=4, MSTR=        0.3317257025,
Error:    SSE =       19.3308477432, df=245, MSE=        0.0789014194,
Total:    SST =       20.6577505533, df=249,
Please wait a moment to get the test statistic sampling distribution,..
H0:lamda1=lamda2=......lamda5,
R2=        0.0642326863, F test=MSTR/MSE=        4.2043058946
p value=0.002670,
```

13.The order statistic sampling distributin of several random variables which are Continuous Bernoulli distribution,

$X_1, X_2, ..., X_n \overset{iid}{\sim} CB(\lambda)$

$X_{(1)}$ is the 1st order statistic of $(X_1, X_2, ..., X_n)$,

$X_{(2)}$ is the 2nd order statistic of $(X_1, X_2, ..., X_n)$,

,…….,

$X_{(n)}$ is the n-th order statistic of $(X_1, X_2, ..., X_n)$,

$X_{(1)} \le X_{(2)} \le X_{(3)} \le ..... \le X_{(n)}$,

The $f(x_j), j = 1,2....,n$ is the marginal probability density function of $j$ -th order statistic of $(X_1, X_2, ..., X_n)$.

The program is
C:\C_Bernoulli\C_Bernoulli_13.cpp,

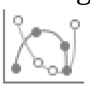
The input data of C:\C_Bernoulli\C_Bernoulli_13.cpp,
Please input the random variable number and lamda, the order index of the order statistic,

The output data,
ANOVA and p value.

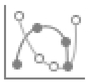
The output data,
computing the sample mean, sample variance, skewed coefficient and kurtosis coefficient of order statistic and the simulated data size=100,000,000, the sample coefficient is approached to the population coefficient.
The frequency table of 100,000,000 data filename is c:\C_Bernoulli\tep\frequency_table.txt which can use the excel.exe to plot the probability distribution and distribution function.

Continuous Bernoulli( $\lambda$ =0.3) and sample size=10 and order number=1, running this program to get frequency_table.txt.
Please use excel.exe plot the probability distribution and the distribution function,
Running excel.exe
Reading the c:\C_Bernoulli\tep\frequency_table.txt

Marking A1 B1 and



Marking C1 D1 and



Note:
frequency_table.txt which class number=200,