

Chapter six The goodness of fit test

5).The Kolmogorov-Smirnov test:

5.1)The test process.

H_0 : Population distribution is a continuous probability distribution,

H_1 : against H_0

The sample size is n.

The process: i) Finding the cumulative probability under H_0 ,

$F(x) = P(X \leq x), X \sim H_0$: Population distribution is a continuous probability distribution.

The x is sample value after ascending sorting.

ii) Finding the relative frequency according the sample data ,

$G(x) = \text{number}(\text{less than and equal } x)/n$.

The jackknife $\text{Max}|F(x) - G(x)| > J_{\alpha, n}$, reject H_0 .

Note: There are 20 kinds of continuous probability distribution that is can be assigned to null hypothesis.

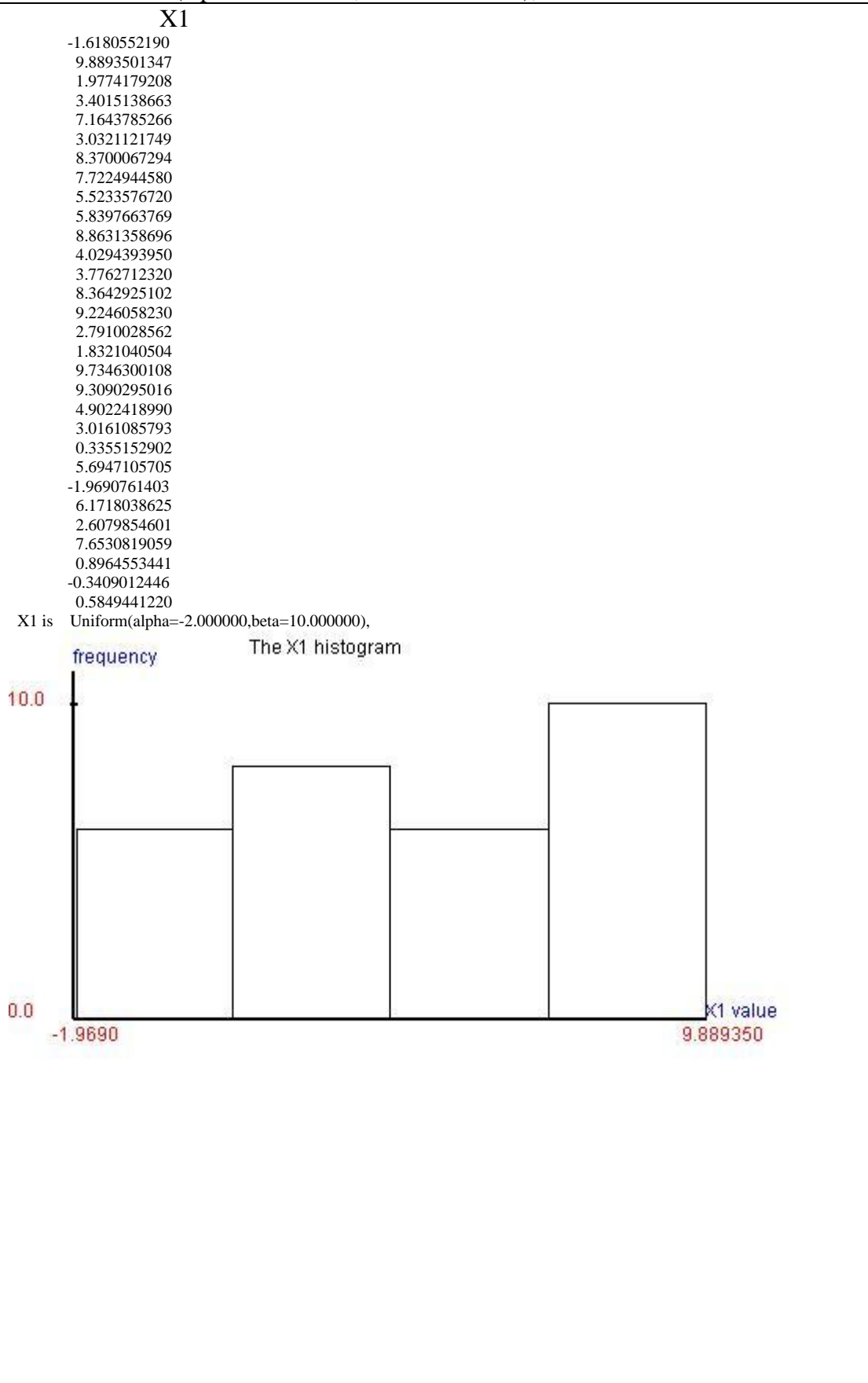
5.2) Example (The simulated sample data and computing the result by the P_S_CCC)

The Kolmogorov Smirnov goodness of fit test

1.H0:Uniform distribution	13.H0:Gumbel distribution
2.H0:Normal distribution	14.H0:Triangular 1 distribution
3.H0:Shifted exponential distribution	15.H0:Trapezoid distribution
4.H0:Pareto 1 distribution	16.H0:U-quadractic distribution
5.H0:Pareto 2 distribution	17.H0:Semi-circle distribution
6.H0:Rayleigh distribution	18.H0:Logistic distribution
7.H0:Double exponential distribution	19.H0:Weibull distribution
8.H0:Log normal distribution	20.H0:Pareto 3 distribution
9.H0:Gamma distribution	** Above Ho population all do once
10.H0:Beta distribution	
11.H0:Cauchy distribution	
12.H0:Arcsin distribution	

5.2.1)The population distribution is uniform distribution.

X1 is Uniform(alpha=-2.000000,beta=10.000000),



H0: $X_1 \sim \text{Uniform}(\alpha, \beta)$, α, β are unknown

alpha point estimated value = -1.969076 (MLE)

beta point estimated value = 9.889350 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X_1)$ is cumulative relative frequency of sample value X_1

$S(X_1)$ is cumulative probability under H0 of sample value X_1

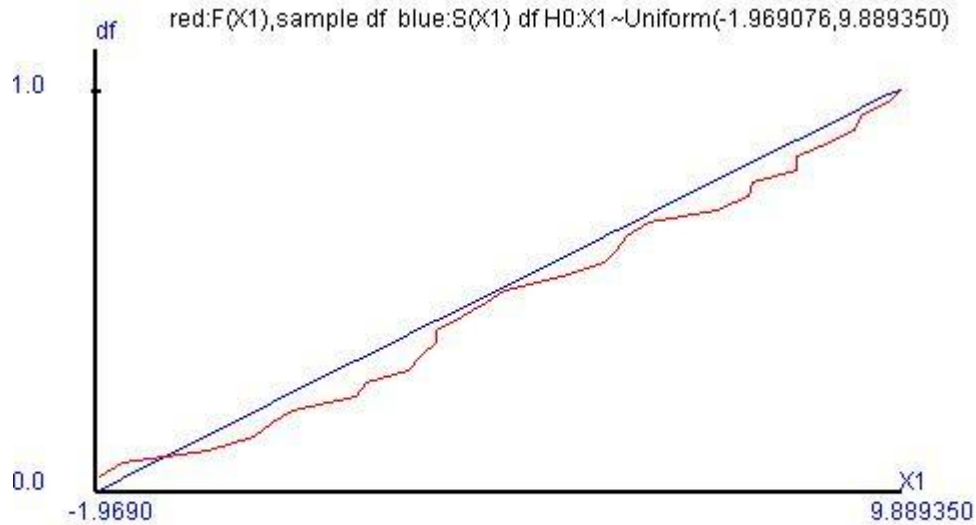
$\text{MAX}|F(X_1) - S(X_1)| = 0.087213$

$0.200000 < \text{p-value} < 1.000000$

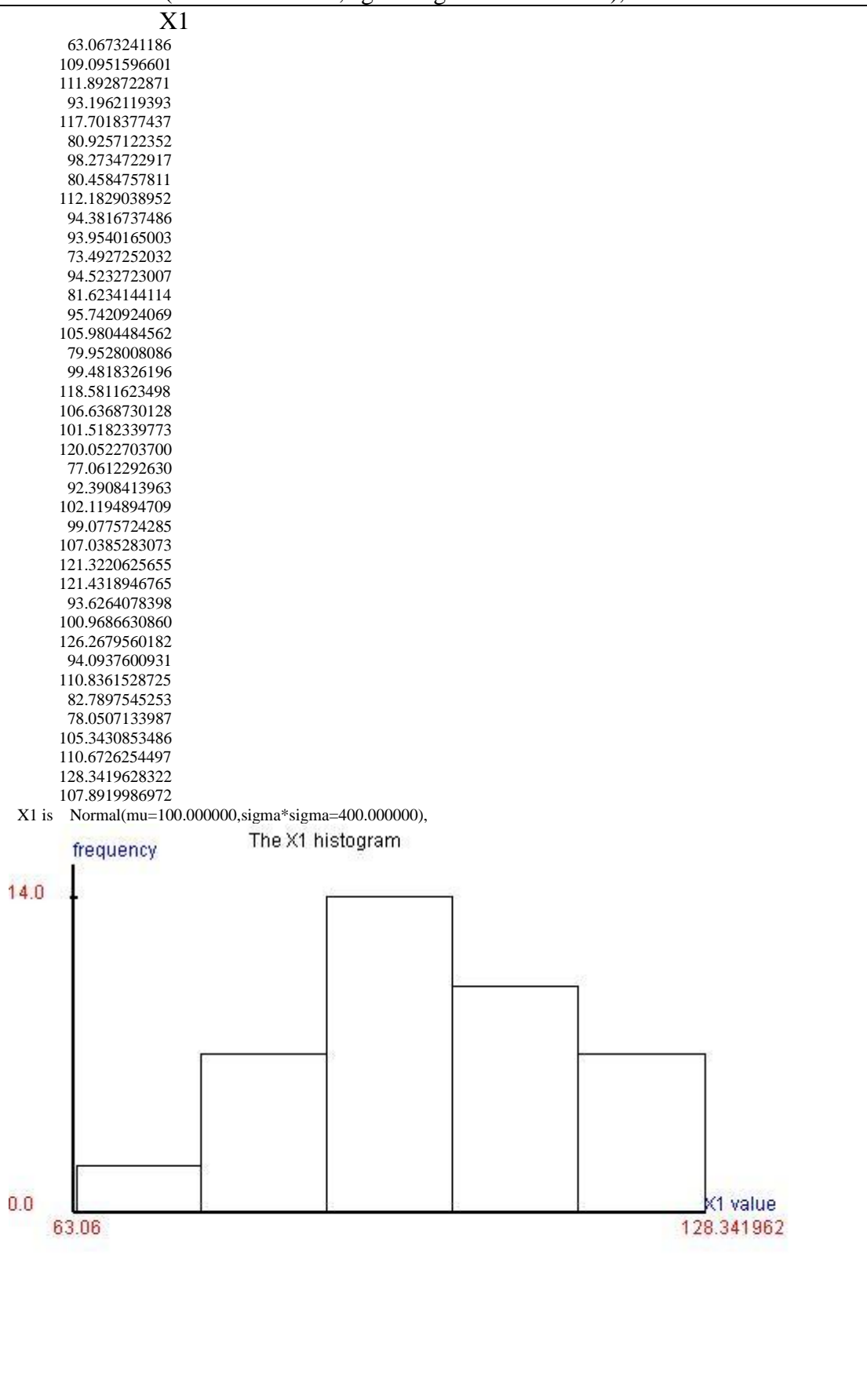
horizontal axis is samples cumulative relative frequency,

vertical axis is cumulative probability under H0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot01_image.jpg



5.2.2)The population distribution is normal distribution.
 X1 is Normal($\mu=100.000000$, $\sigma^2=400.000000$),



H0: $X_1 \sim \text{Normal}(\mu, \sigma^2)$, μ, σ are unknown
population mean(μ) point estimated value=99.800987 (MLE,UMVUE)
population variance(σ^2) which point estimated value=242.418641 (UMVUE)

The Kolmogorov Smirnov goodness of fit test

$F(X_1)$ is cumulative relative frequency of sample value X_1

$S(X_1)$ is cumulative probability under H_0 of sample value X_1

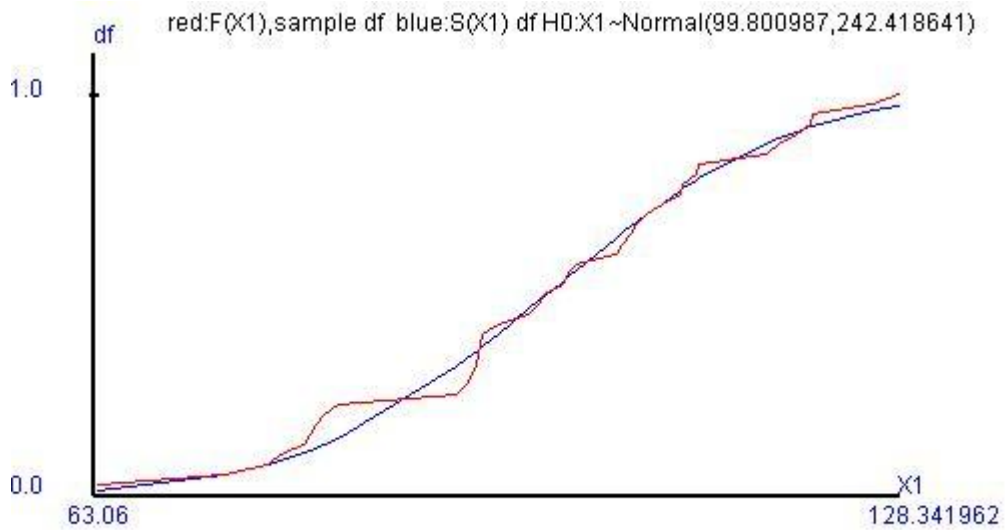
$\text{MAX}|F(X_1)-S(X_1)|=0.087700$

$0.200000 < \text{p-value} < 1.000000$

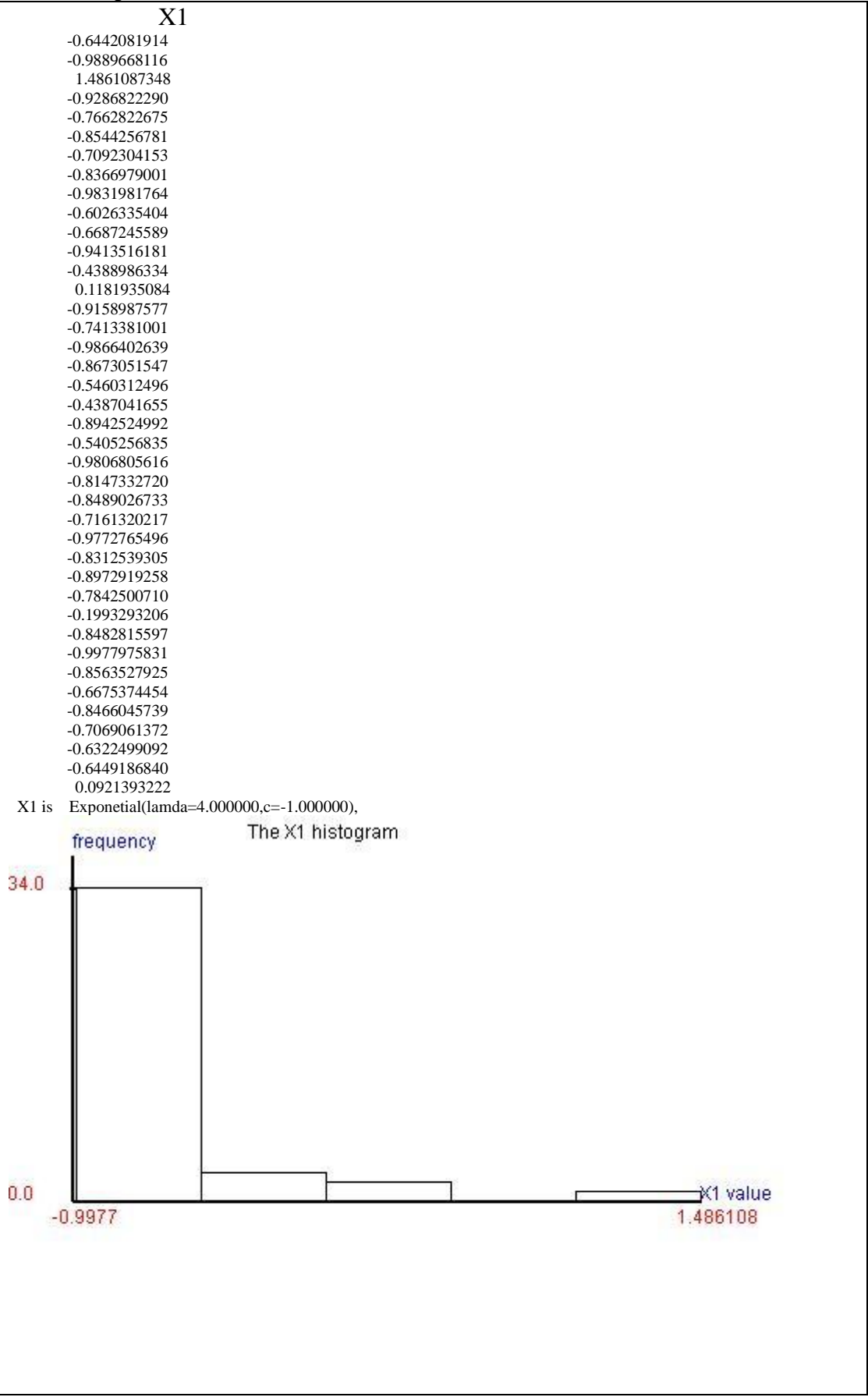
horizontal axis is samples cumulative relative frequency,

vertical axis is cumulative probability under H_0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot02_image.jpg



5.2.3) The population distribution is shifted exponential distribution.
 X_1 is Exponential($\lambda=4.000000, c=-1.000000$),



H0: $X_1 \sim \text{Shifted exponential}(\lambda, c)$, λ, c are unknown

λ point estimated value = 3.061885 (MLE)

c point estimated value = -0.997798 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X_1)$ is cumulative relative frequency of sample value X_1

$S(X_1)$ is cumulative probability under H_0 of sample value X_1

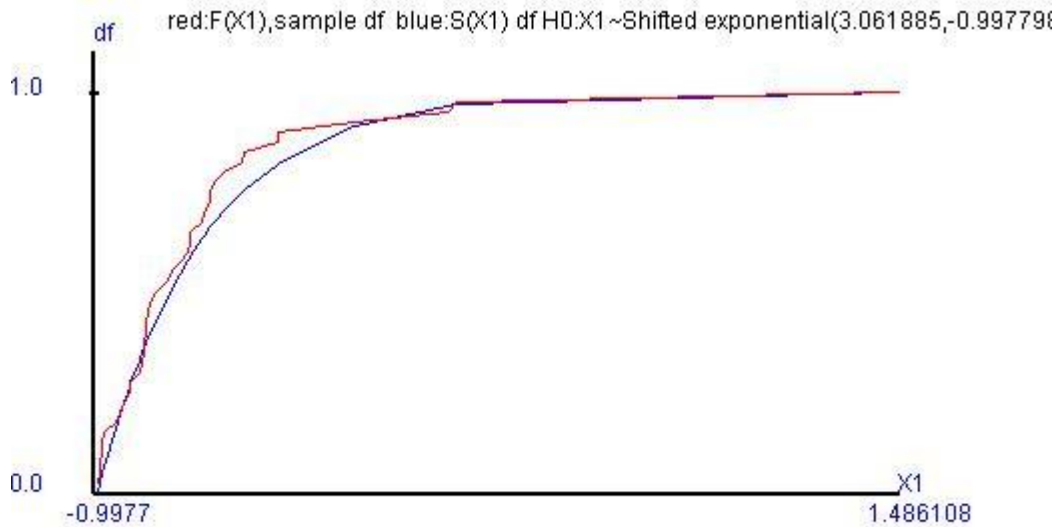
$\text{MAX}|F(X_1) - S(X_1)| = 0.101520$

$0.200000 < p\text{-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

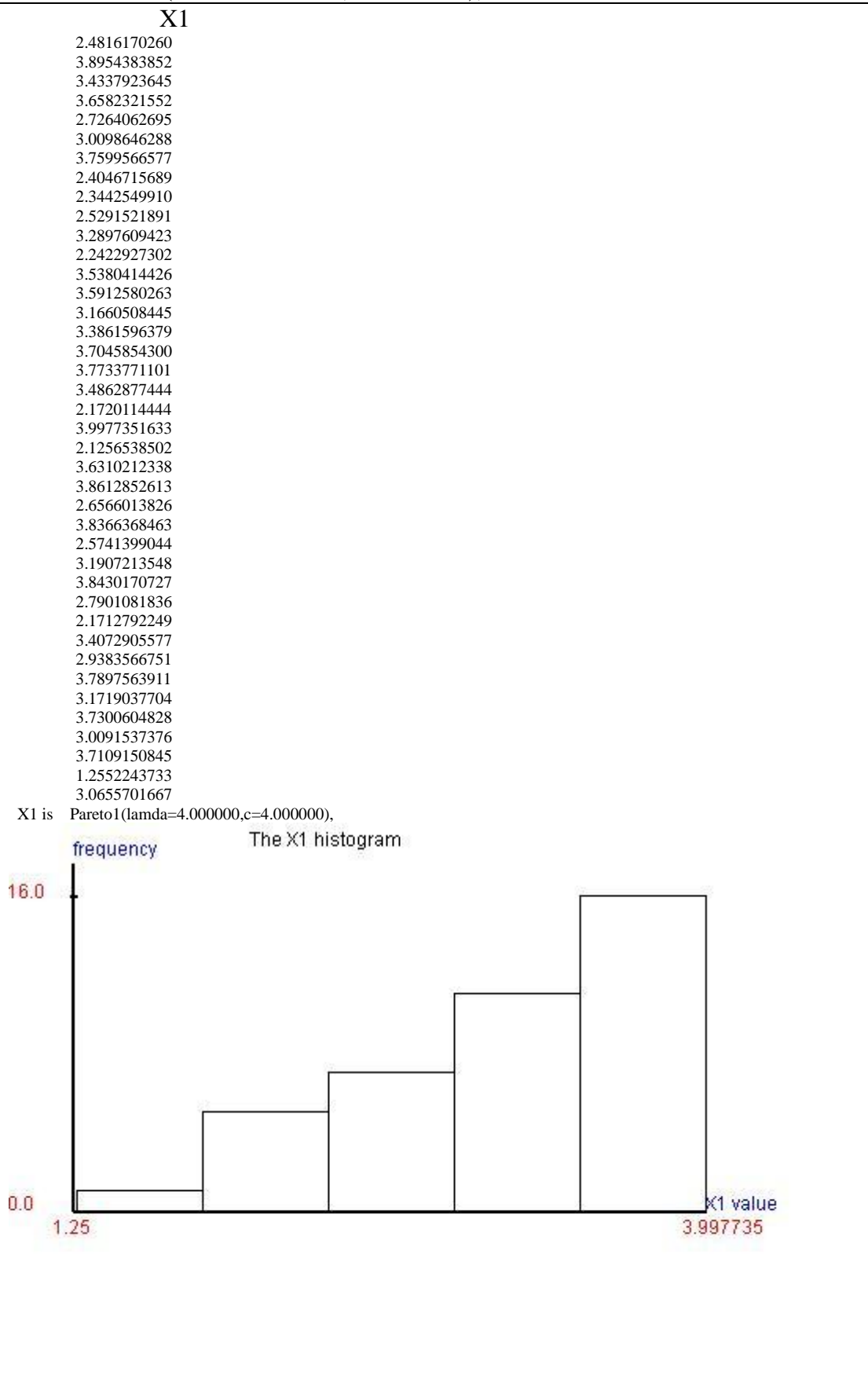
vertical axis is cumulative probability under H_0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot03_image.jpg



5.2.4)The population distribution is pareto1 distribution.

X1 is Pareto1(lamda=4.000000,c=4.000000),



H0: $X_1 \sim \text{Pareto 1}(\lambda, c)$, λ, c are unknown

λ point estimated value = 3.722347 (MLE)

c point estimated value = 3.997735 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X_1)$ is cumulative relative frequency of sample value X_1

$S(X_1)$ is cumulative probability under H_0 of sample value X_1

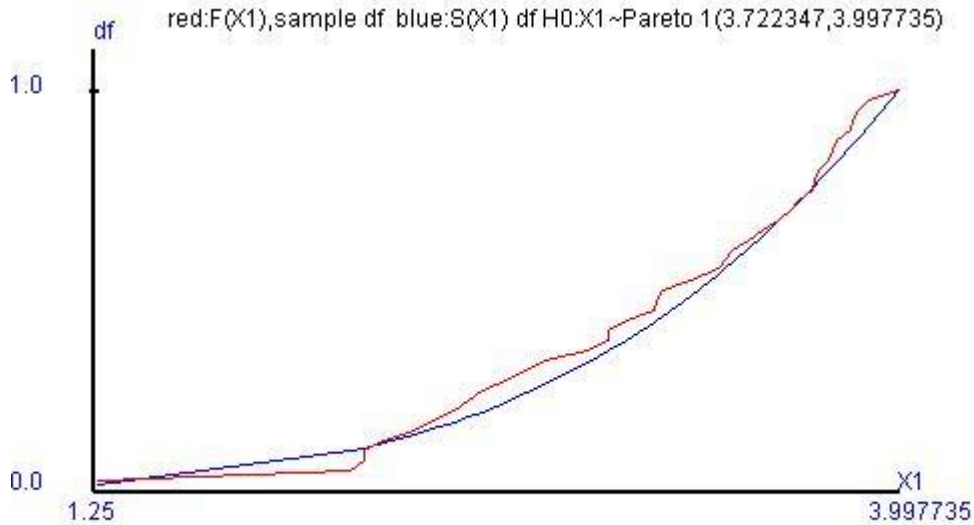
$\text{MAX}|F(X_1) - S(X_1)| = 0.071263$

$0.200000 < \text{p-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

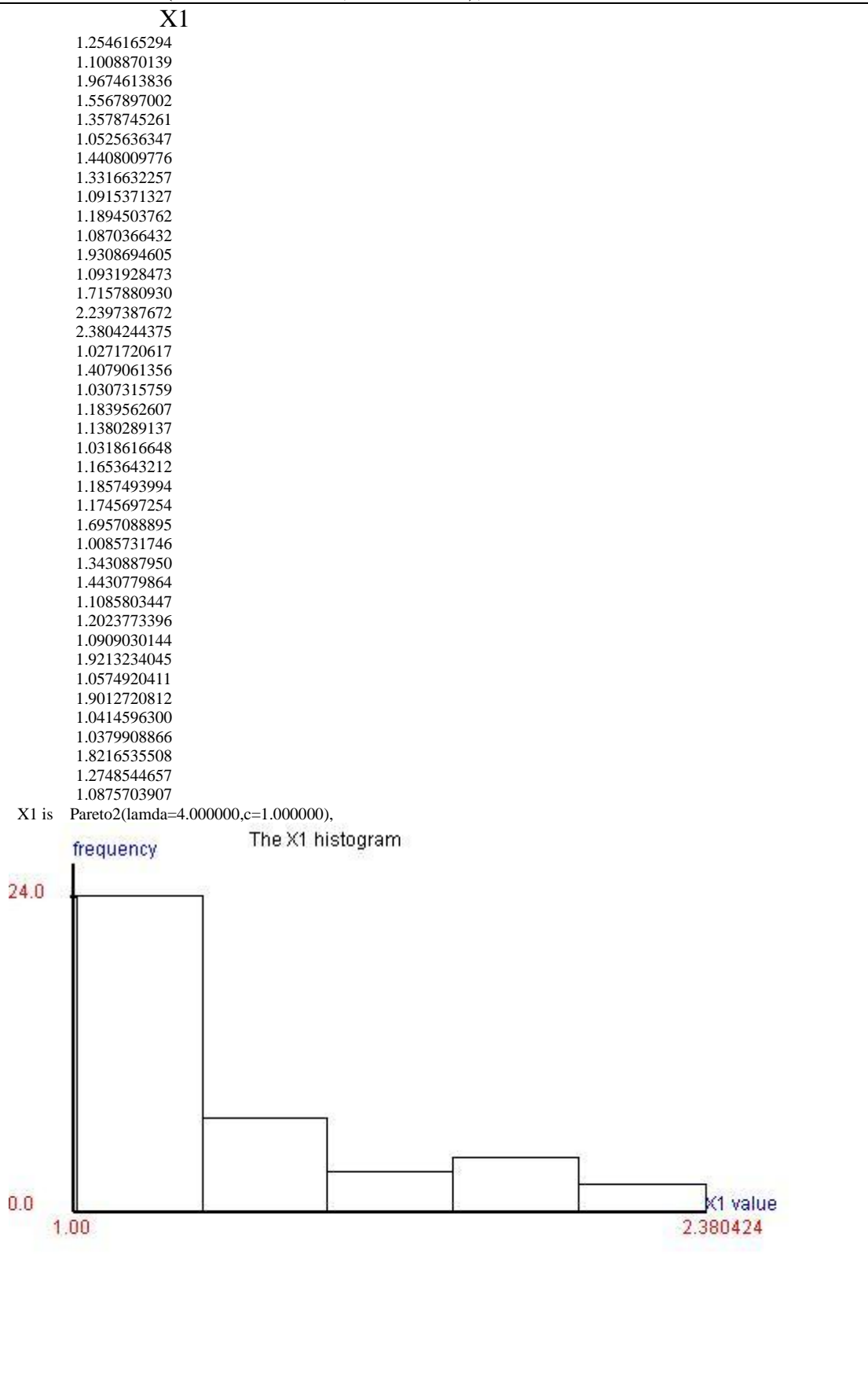
vertical axis is cumulative probability under H_0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot04_image.jpg



5.2.5)The population distribution is pareto2 distribution.

X1 is Pareto2(lamda=4.000000,c=1.000000),



H0: $X_1 \sim \text{Pareto 2}(\lambda, c)$, λ, c are unknown

λ point estimated value = 3.789168 (MLE)

c point estimated value = 1.008573 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X_1)$ is cumulative relative frequency of sample value X_1

$S(X_1)$ is cumulative probability under H_0 of sample value X_1

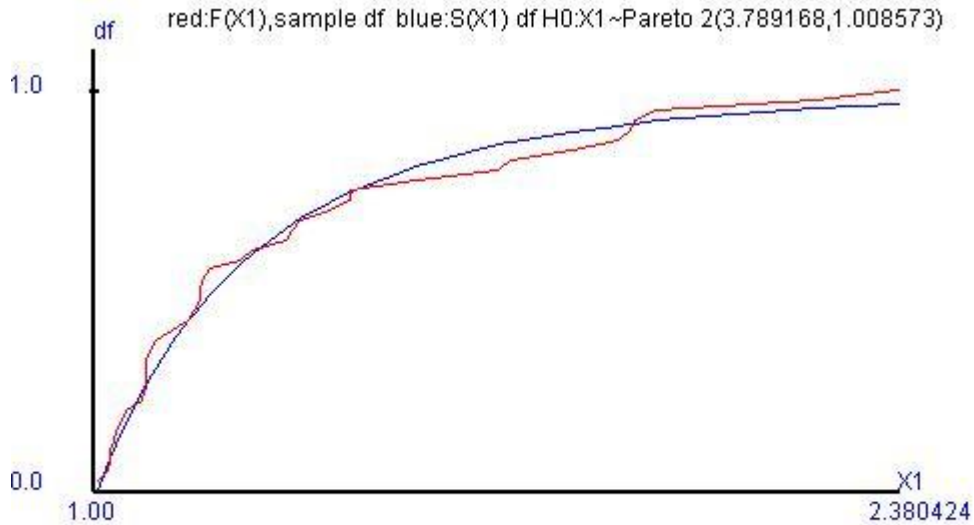
$\text{MAX}|F(X_1) - S(X_1)| = 0.073904$

$0.200000 < \text{p-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

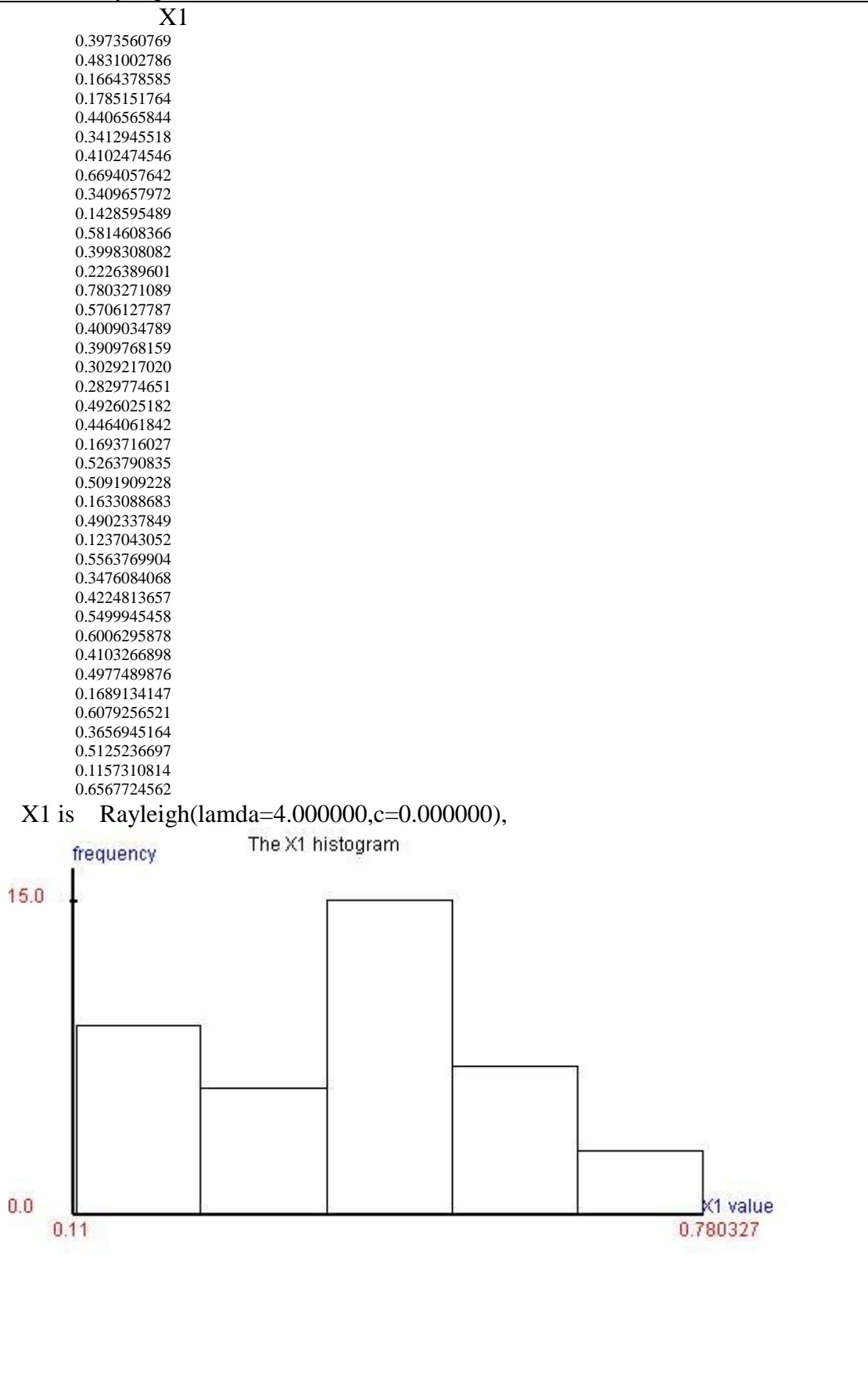
vertical axis is cumulative probability under H_0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot05_image.jpg



5.2.6)The population distribution is rayleigh distribution.

X1 is Rayleigh(lamda=4.000000,c=0.000000),



H0: $X1 \sim \text{Rayleigh}(\lambda, c)$, λ, c are unknown

λ point estimated value = 8.936091 (MLE)

c point estimated value = 0.115731 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under $H0$ of sample value $X1$

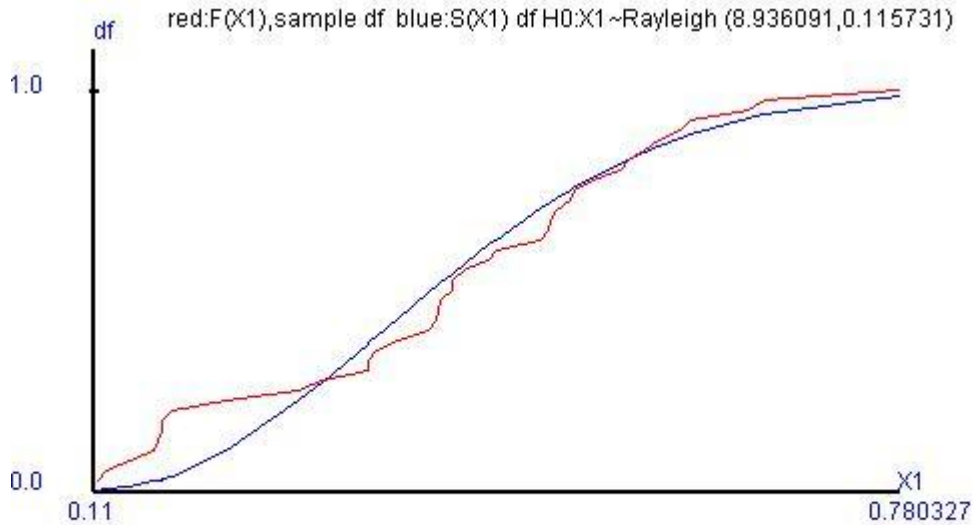
$\text{MAX}|F(X1) - S(X1)| = 0.165389$

$0.200000 < \text{p-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

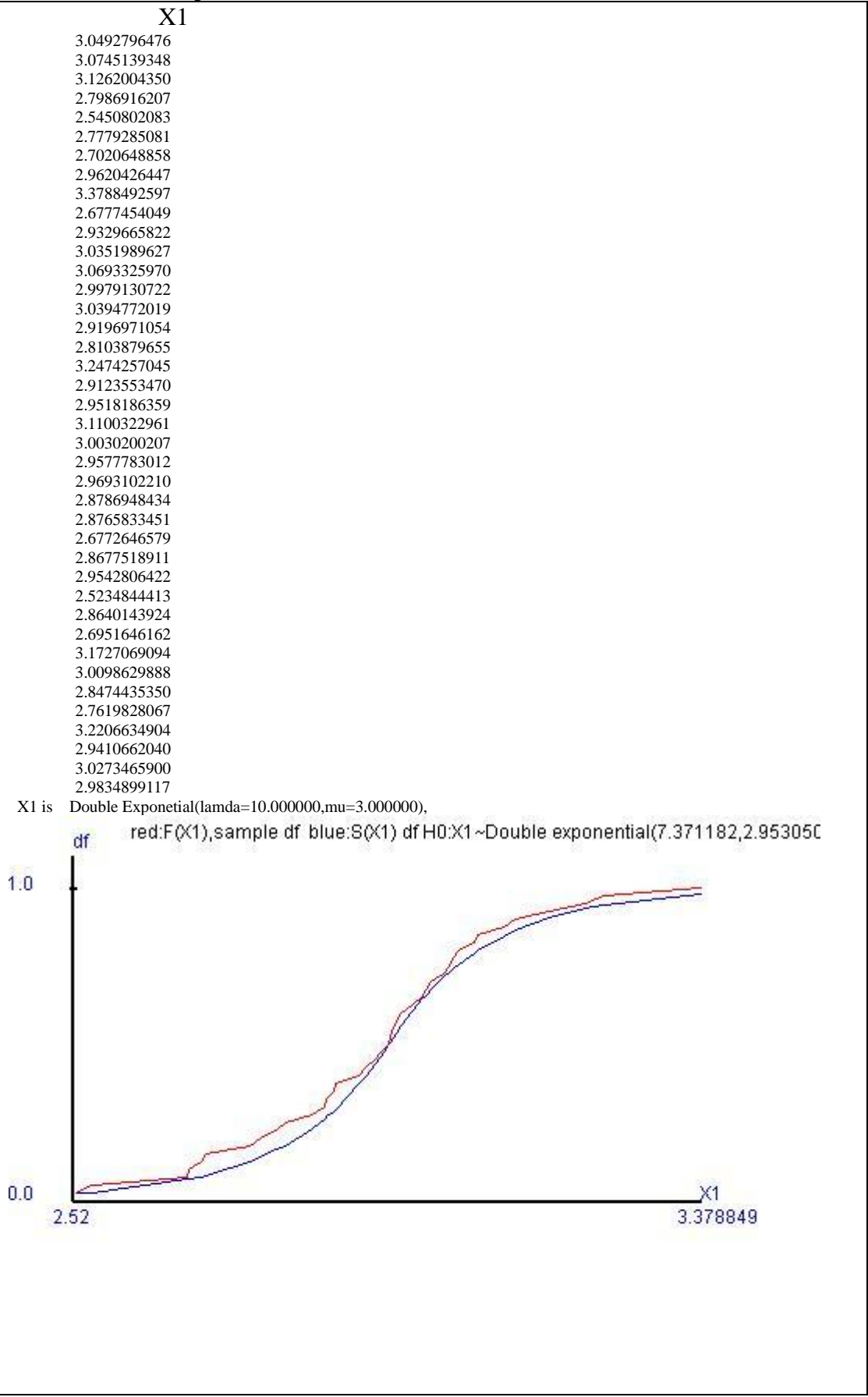
vertical axis is cumulative probability under $H0$ from sample values.

The df comparison plot images is stored in `c:\book_01\K_S_plot06_image.jpg`



5.2.7) The population distribution is double exponential distribution.

X1 is Double Exponential($\lambda=10.000000, \mu=3.000000$),



H0: $X_1 \sim \text{Double exponential}(\lambda, \mu)$, λ, μ are unknown

λ point estimated value = 7.371182 (MLE)

μ point estimated value = 2.953050 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X_1)$ is cumulative relative frequency of sample value X_1

$S(X_1)$ is cumulative probability under H_0 of sample value X_1

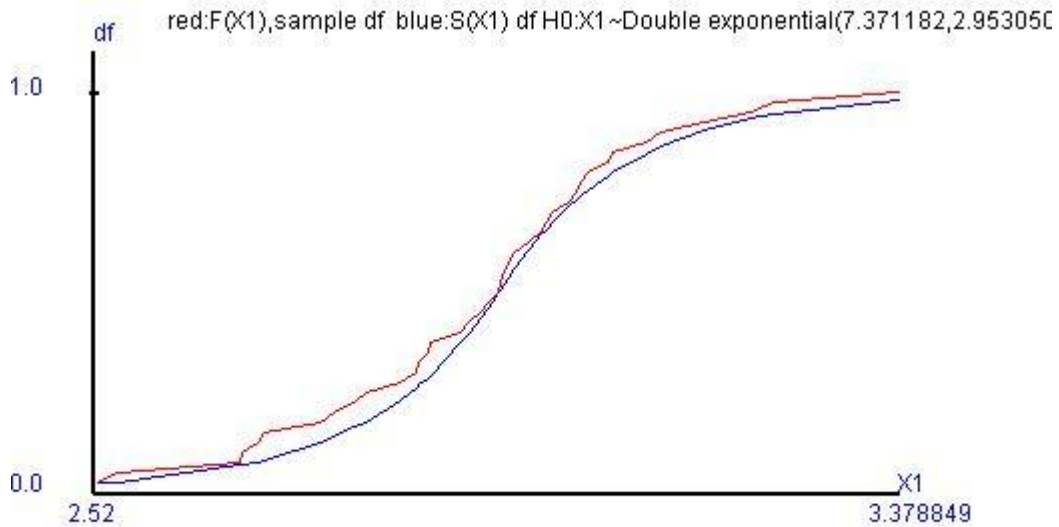
$\text{MAX}|F(X_1) - S(X_1)| = 0.085971$

$0.200000 < p\text{-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

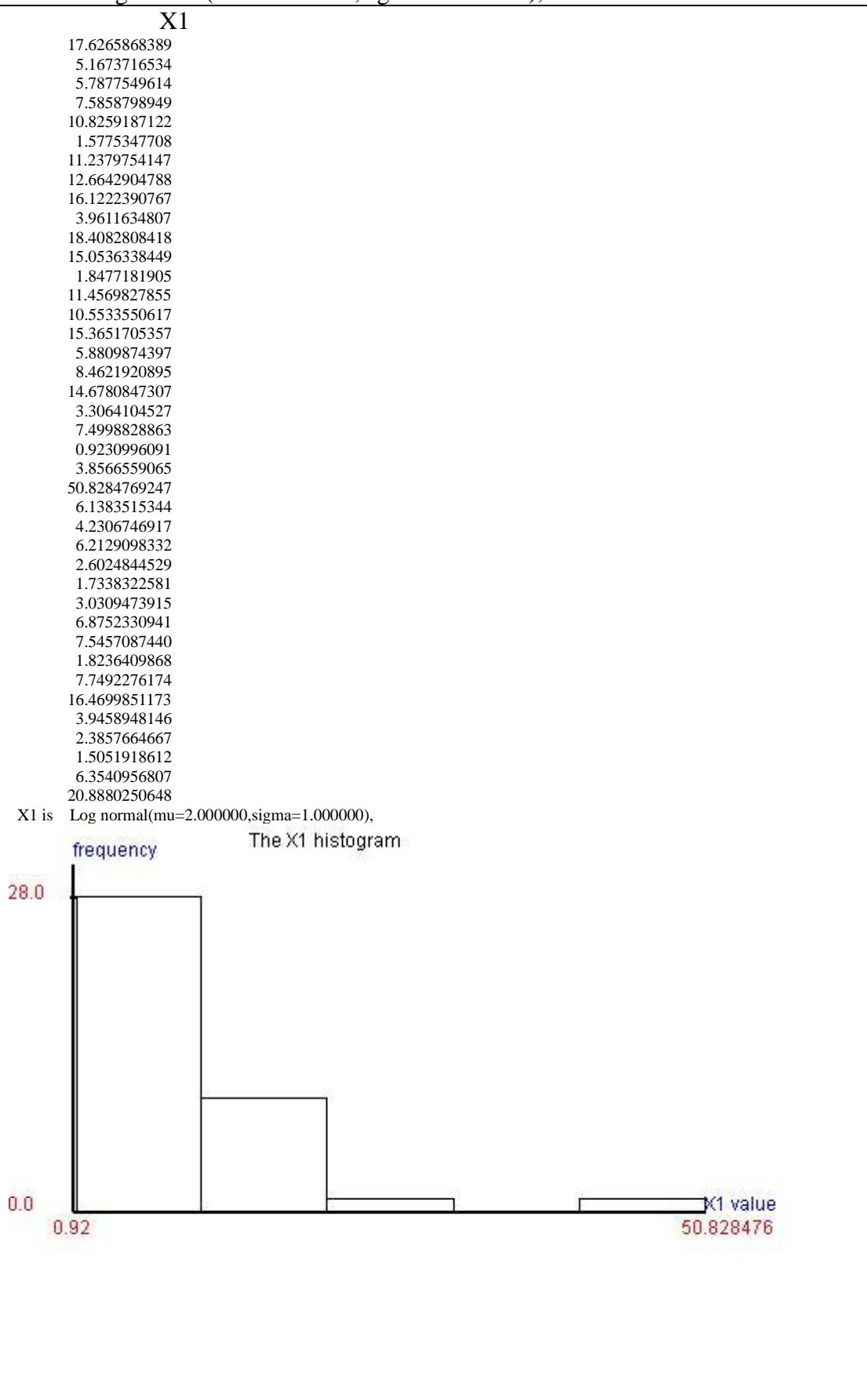
vertical axis is cumulative probability under H_0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot07_image.jpg



5.2.8)The population distribution is lognormal distribution.

X1 is Log normal($\mu=2.000000,\sigma=1.000000$),



H0: $X1 \sim \text{Log_Normal}(\mu, \sigma^2)$, μ, σ are unknown
 population mean(μ) point estimated value=1.841303 (MLE,UMVUE)
 population variance(σ^2) which point estimated value=0.768641 (UMVUE)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under $H0$ of sample value $X1$

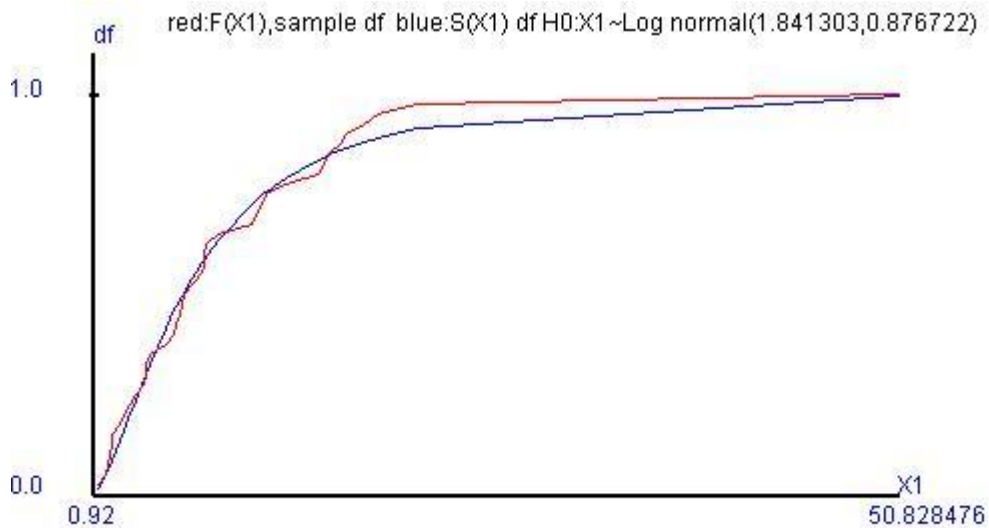
$\text{MAX}|F(X1)-S(X1)|=0.069200$

$0.200000 < p\text{-value} < 1.000000$

horizon axis is samples cumulative relative frequency,

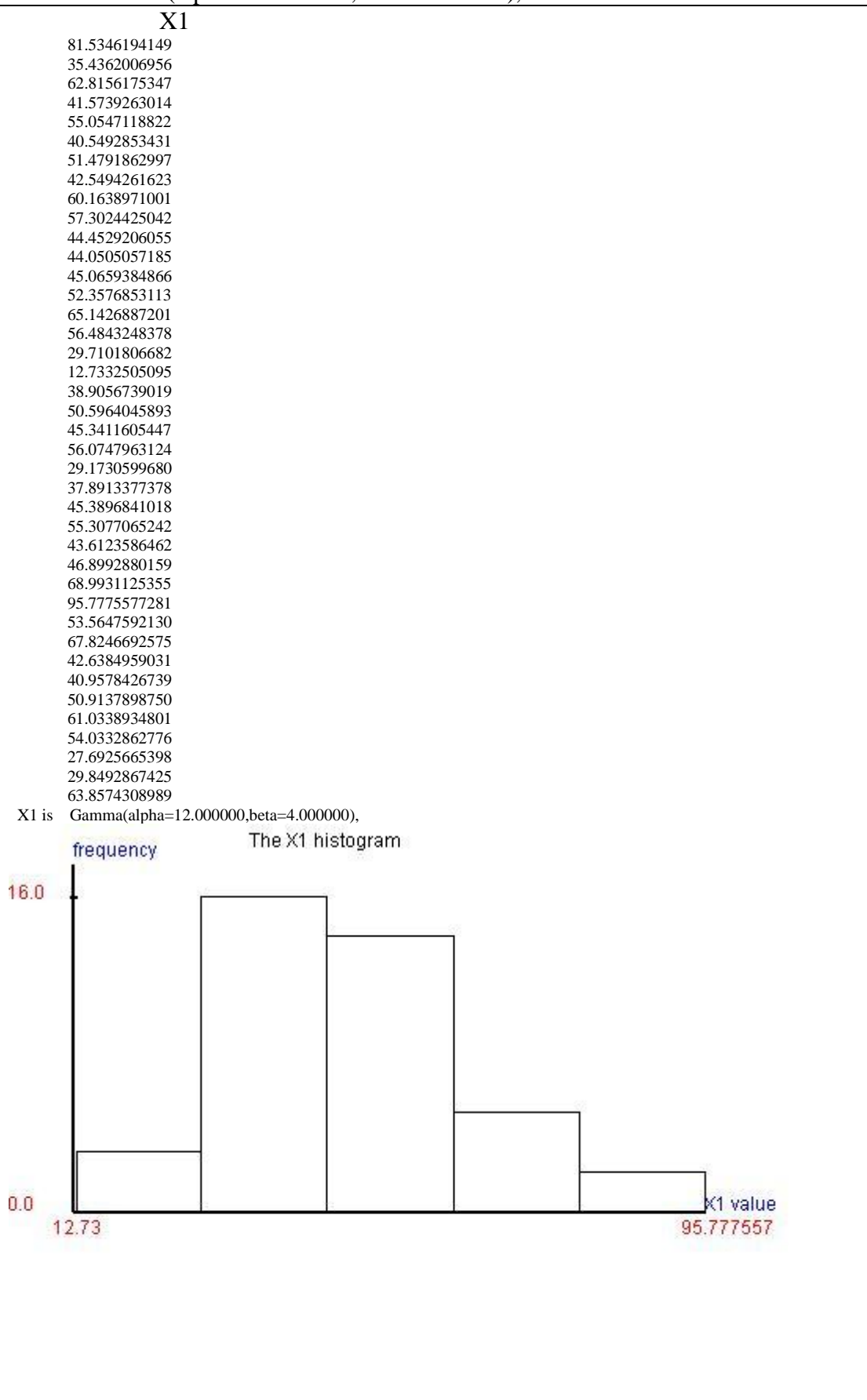
vertical axis is cumulative probability under $H0$ from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot08_image.jpg



5.2.9)The population distribution is gamma distribution.

X1 is Gamma(alpha=12.000000,beta=4.000000),



H0: $X1 \sim \text{Gamma}(\alpha, \beta)$, α, β are unknown

α point estimated value=11.000000 (MME)

β point estimated value=4.626278 (MME)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under $H0$ of sample value $X1$

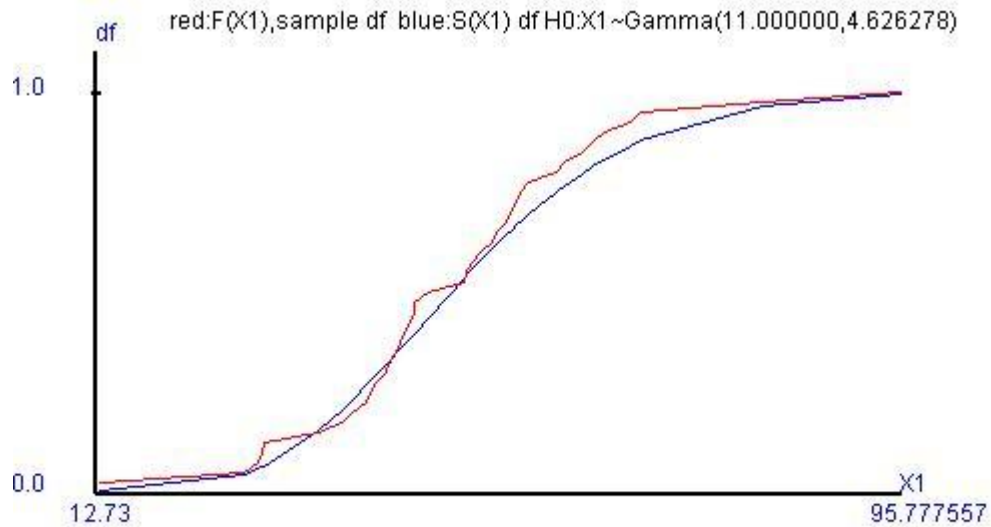
$\text{MAX}|F(X1)-S(X1)|=0.083000$

$0.200000 < \text{p-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

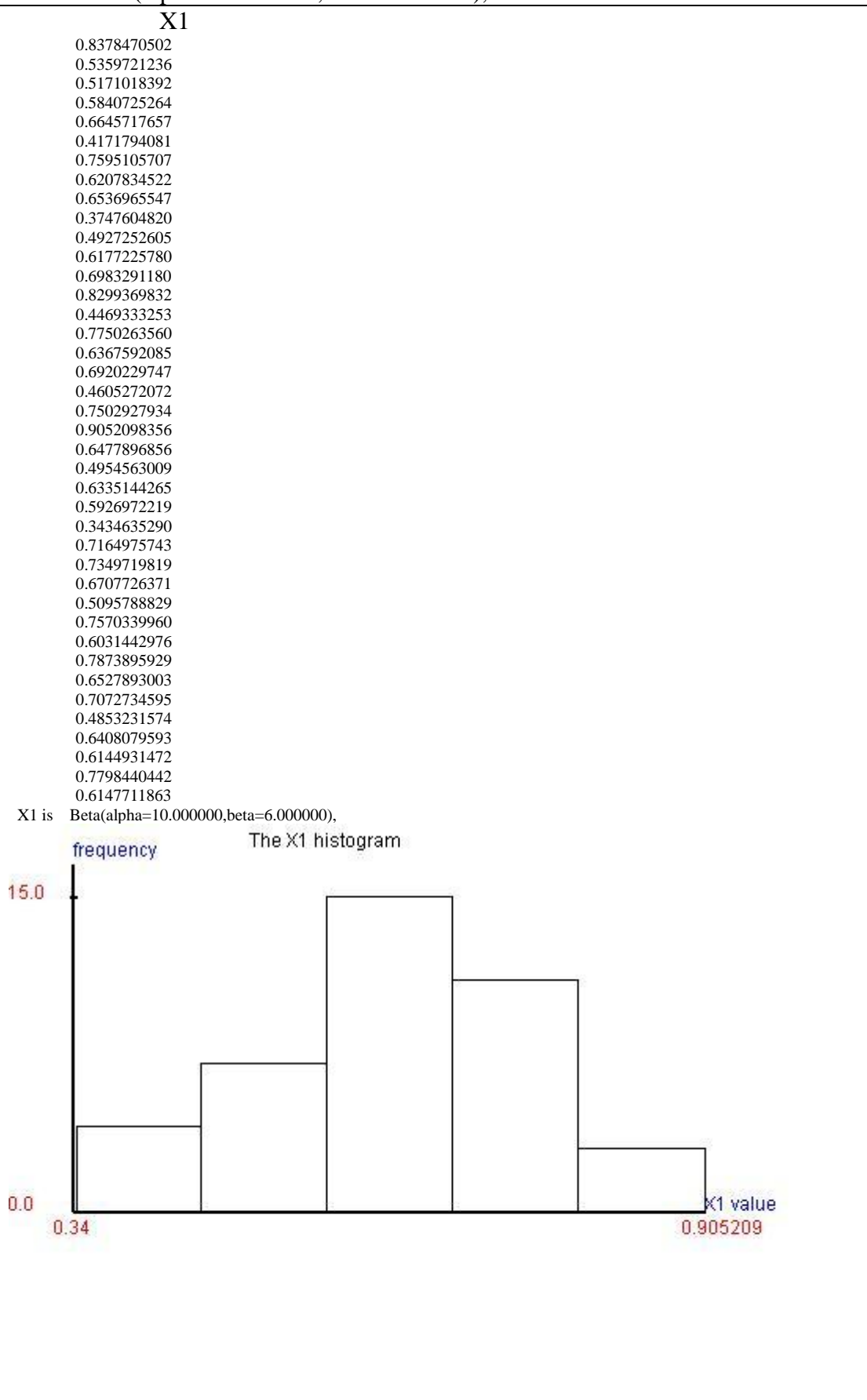
vertical axis is cumulative probability under $H0$ from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot02_image.jpg



5.2.10)The population distribution is beta distribution.

X1 is $\text{Beta}(\alpha=10.000000,\beta=6.000000)$,



H0: $X1 \sim \text{Beta}(\alpha, \beta)$, α, β are unknown

alpha point estimated value=7.500000 (MME)

beta point estimated value=4.500000 (MME)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under H0 of sample value $X1$

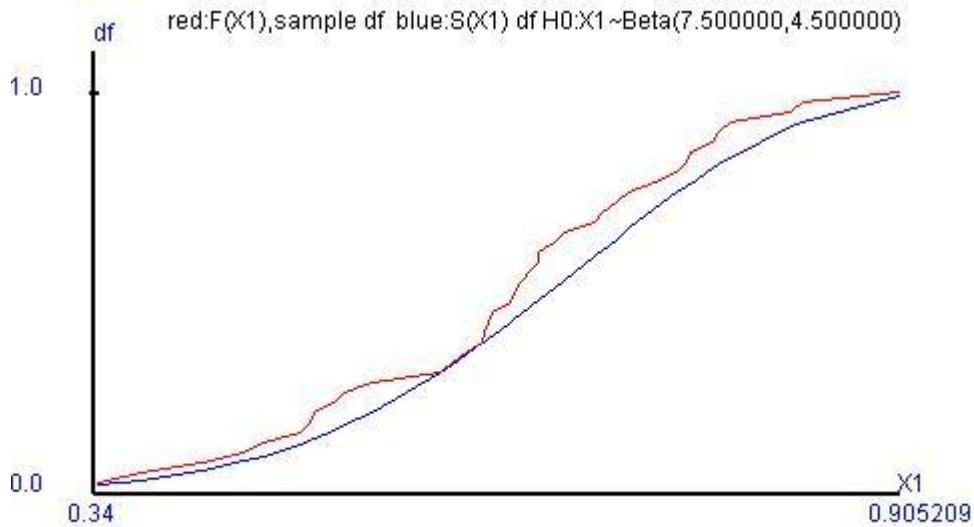
$\text{MAX}|F(X1)-S(X1)|=0.123367$

$0.200000 < \text{p-value} < 1.000000$

horizon axis is samples cumulative relative frequency,

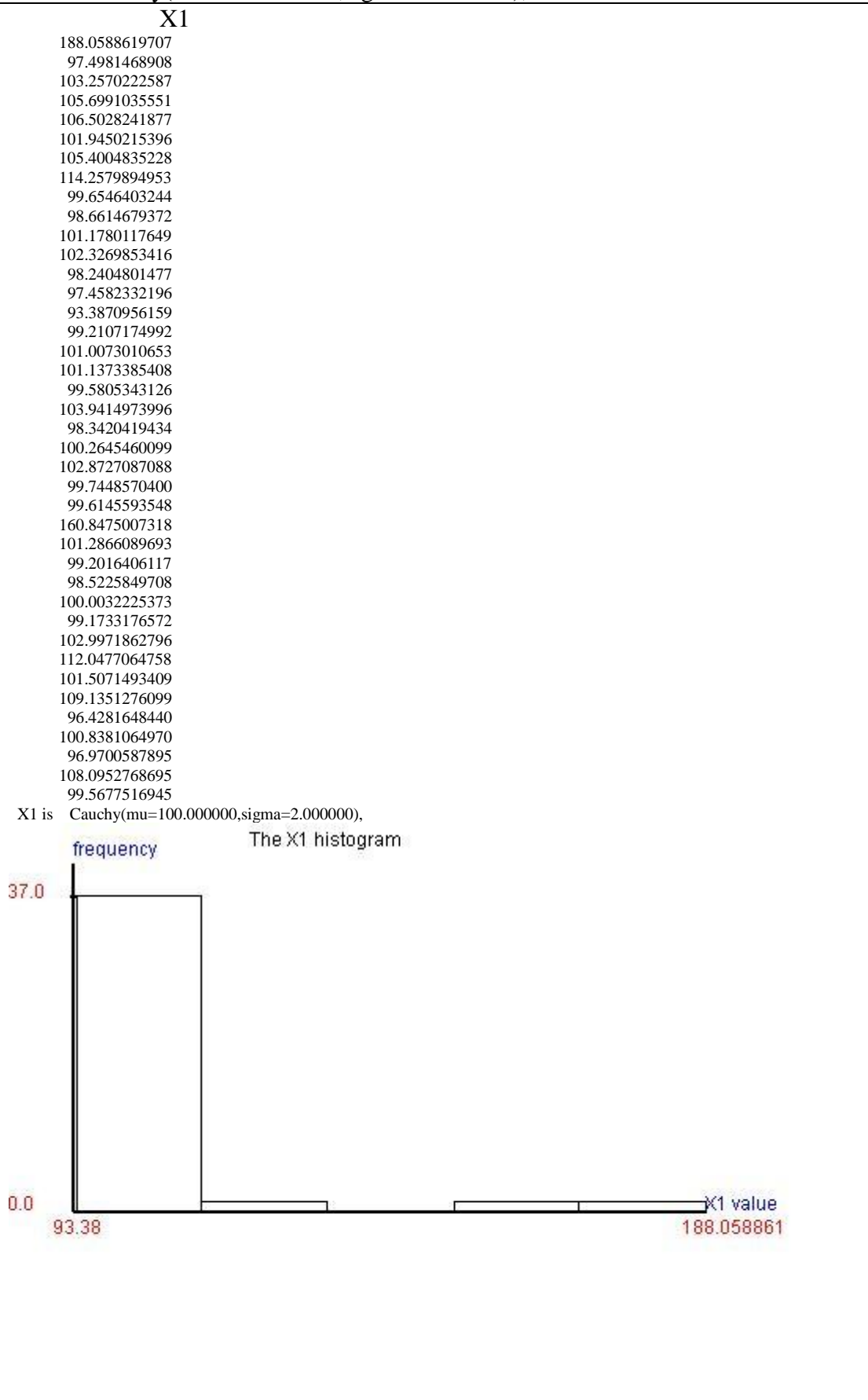
vertical axis is cumulative probability under H0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot10_image.jpg



5.2.11)The population distribution is cauchy distribution.

X1 is Cauchy($\mu=100.000000,\sigma=2.000000$),



H0: $X_1 \sim \text{Cauchy}(\mu, \sigma)$, μ, σ are unknown

μ point estimated value=100.922704

σ point estimated value=2.377009

The Kolmogorov Smirnov goodness of fit test

$F(X_1)$ is cumulative relative frequency of sample value X_1

$S(X_1)$ is cumulative probability under H0 of sample value X_1

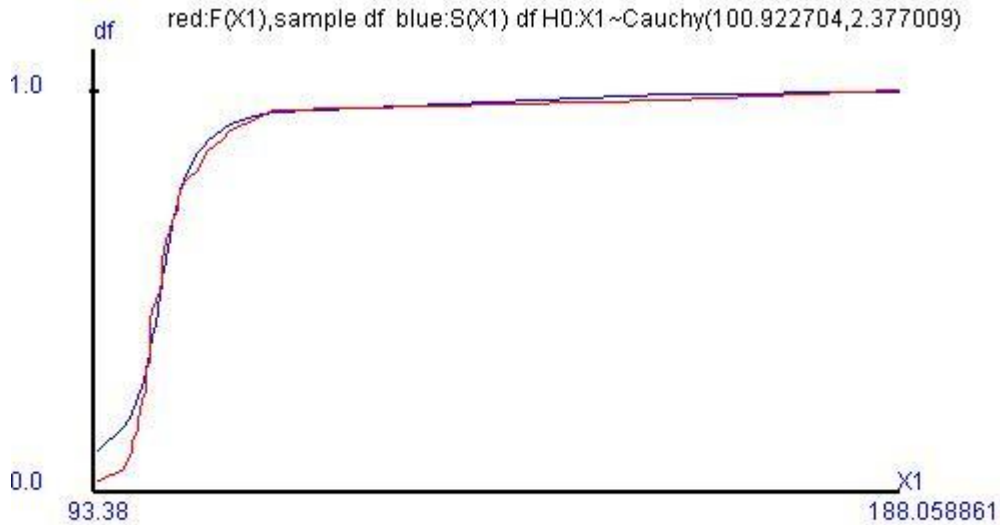
$\text{MAX}|F(X_1)-S(X_1)|=0.104849$

$0.200000 < \text{p-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

vertical axis is cumulative probability under H0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot11_image.jpg

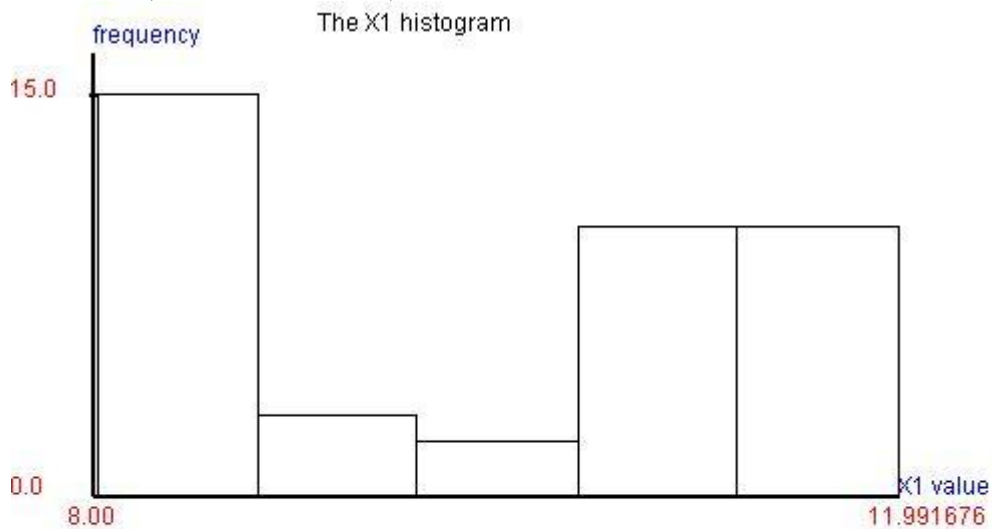


5.2.12) The population distribution is arcsin distribution.

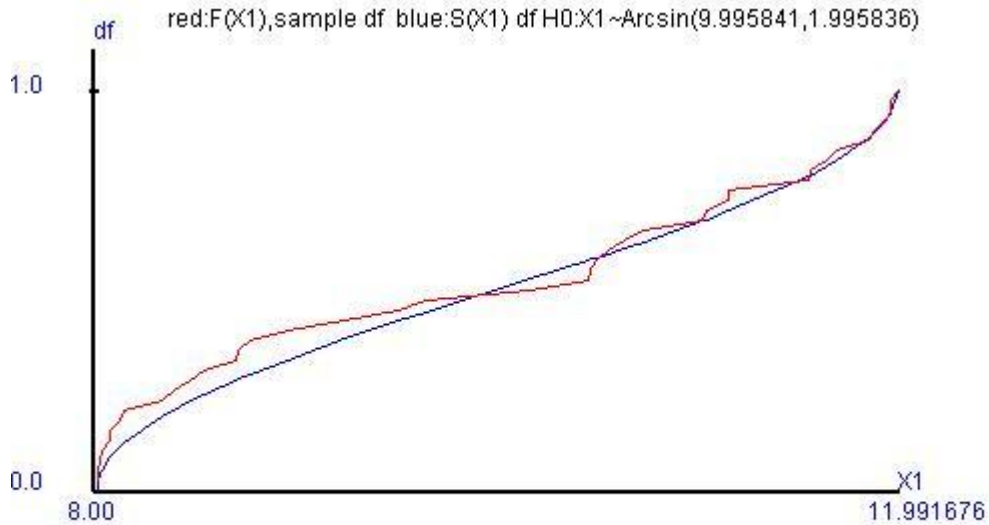
X1 is $\text{Arcsin}(\mu=10.000000, c=2.000000)$,

X1
 10.4919889868
 11.5534238842
 11.1478798539
 8.4632706564
 11.0218137294
 8.9482679362
 10.1597462766
 8.0056432740
 10.5460398864
 11.1507638781
 11.9382901817
 8.3809882974
 8.0000048419
 8.7000421739
 9.6426942541
 8.7593317000
 8.0196915846
 11.8779547680
 11.9586586661
 10.6314953982
 8.1132564180
 11.8481581323
 8.1393025510
 10.7296484673
 8.5367165410
 11.5452804003
 11.6853387951
 11.9916767994
 8.3261590080
 11.0369281615
 11.6381664787
 8.0665328298
 8.0223709200
 8.0683180942
 8.6889166962
 9.5046773603
 10.4619421388
 10.4481907637
 11.9551542726
 9.2339216097

X1 is $\text{Arcsin}(\mu=10.000000, c=2.000000)$,

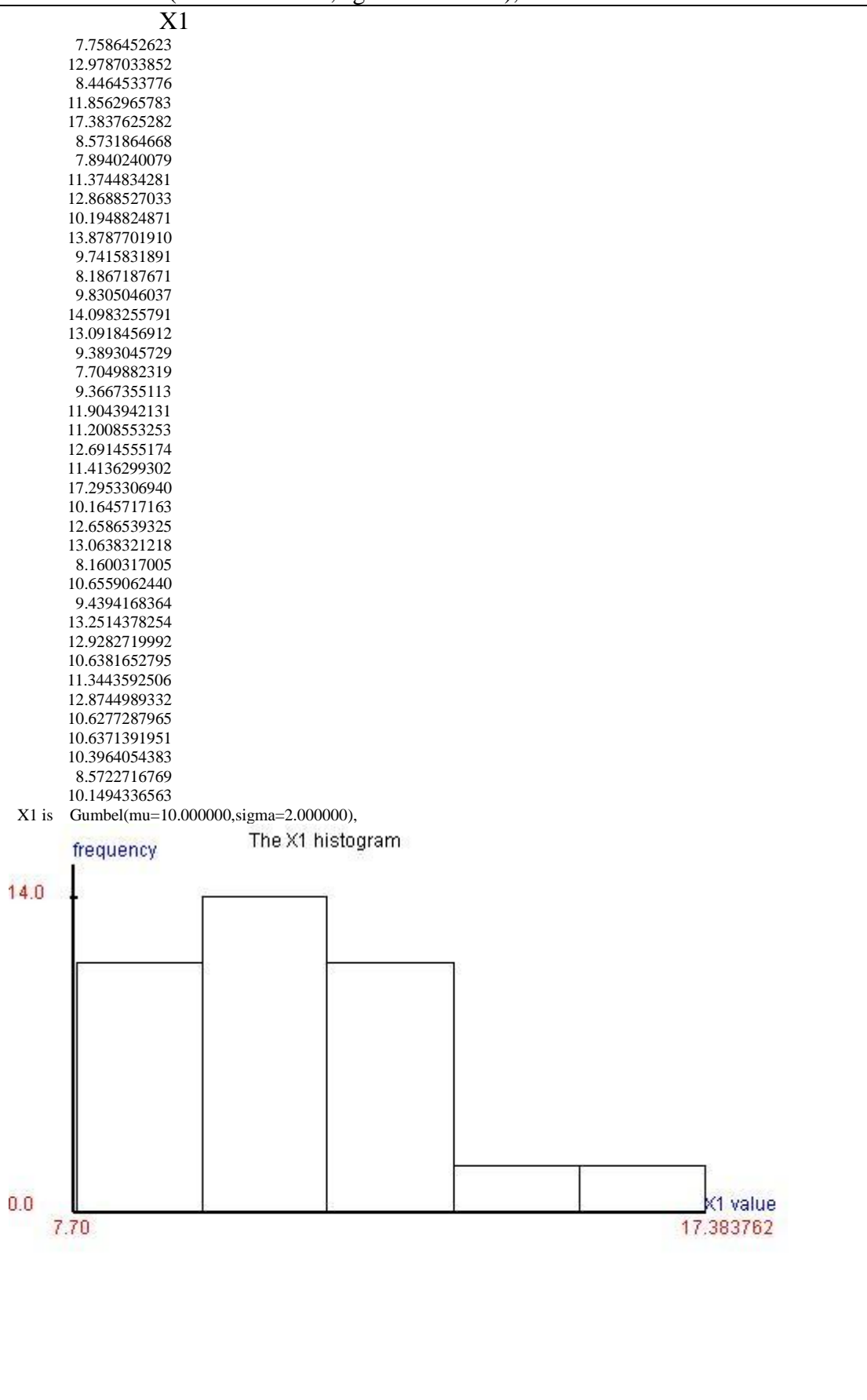


$H_0: X_1 \sim \text{Arcsin}(\mu, c)$, μ, c are unknown
 μ point estimated value = 9.995841 (MLE)
 c point estimated value = 1.995836 (MLE)
 The Kolmogorov Smirnov goodness of fit test
 $F(X_1)$ is cumulative relative frequency of sample value X_1
 $S(X_1)$ is cumulative probability under H_0 of sample value X_1
 $\text{MAX}|F(X_1) - S(X_1)| = 0.087683$
 $0.200000 < \text{p-value} < 1.000000$
 horizon axis is samples cumulative relative frequency,
 vertical axis is cumulative probability under H_0 from sample values.
 The df comparison plot images is stored in c:\book_01\K_S_plot12_image.jpg



5.2.13) The population distribution is gumbel distribution.

X1 is Gumbel($\mu=10.000000, \sigma=2.000000$),



H0: $X_1 \sim \text{Gumbel}(\mu, \sigma)$, μ, σ are unknown

μ point estimated value=10.066199 (MME)

σ point estimated value=1.820452 (MME)

The Kolmogorov Smirnov goodness of fit test

$F(X_1)$ is cumulative relative frequency of sample value X_1

$S(X_1)$ is cumulative probability under H_0 of sample value X_1

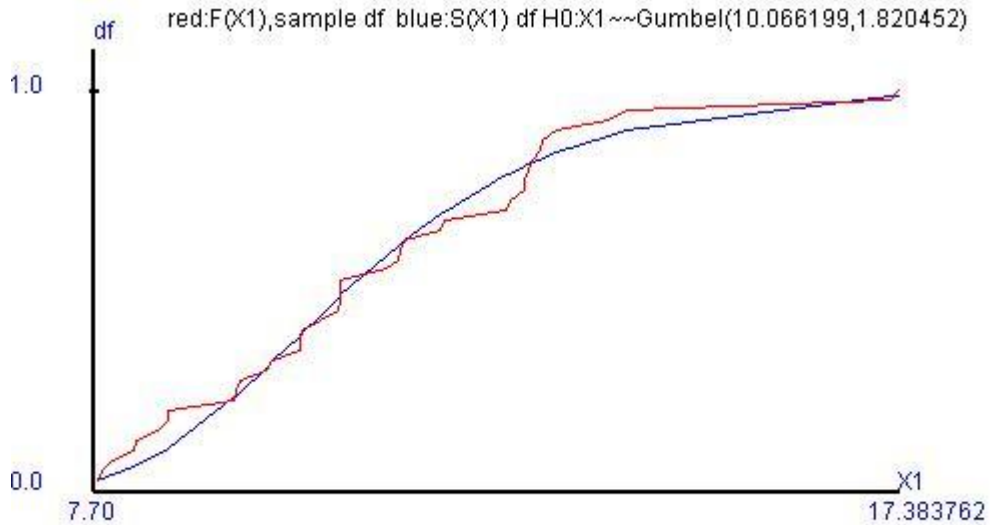
$\text{MAX}|F(X_1)-S(X_1)|=0.096771$

$0.200000 < \text{p-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

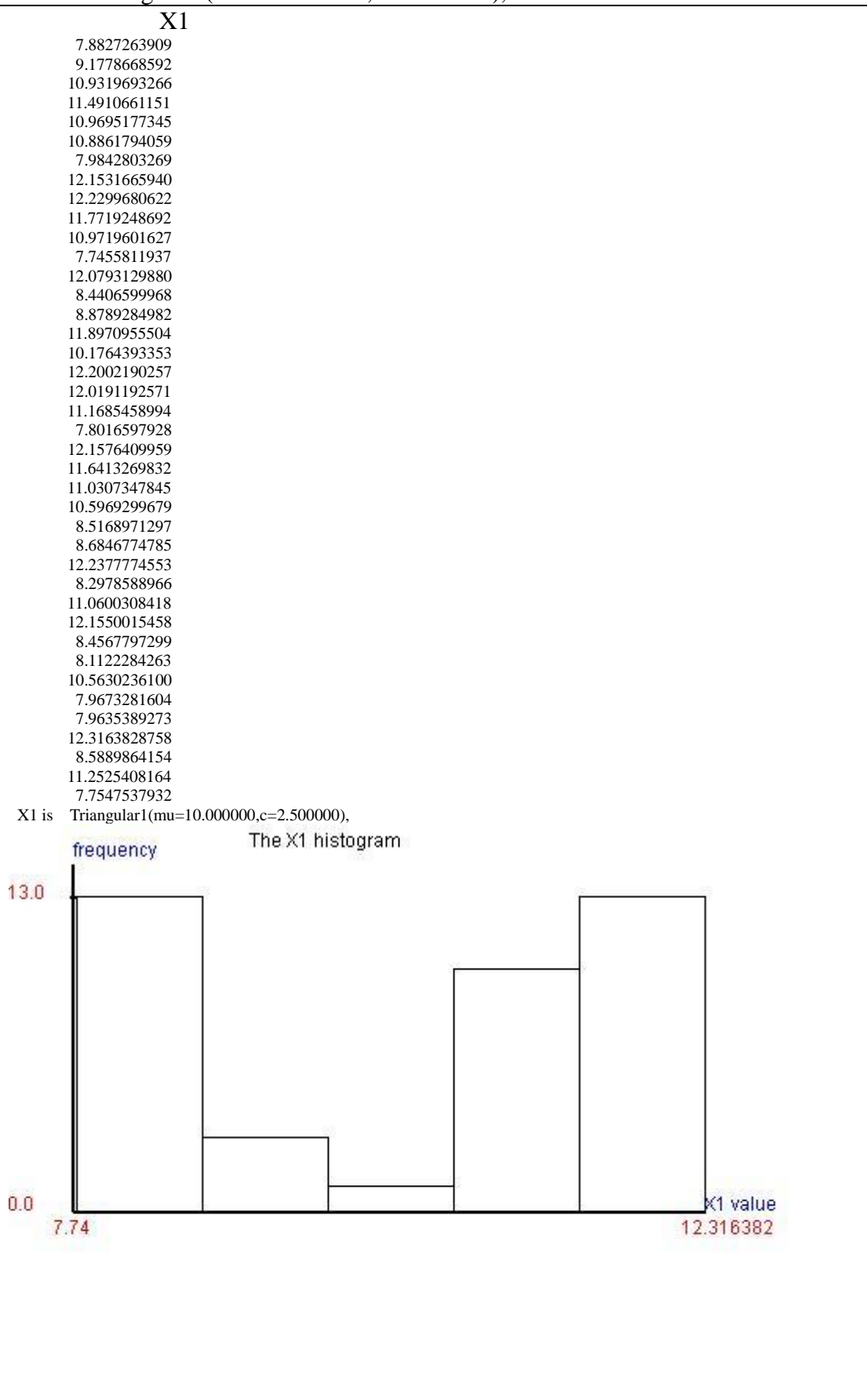
vertical axis is cumulative probability under H_0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot13_image.jpg



5.2.14) The population distribution is triangular 1 distribution.

X1 is $\text{Triangular1}(\mu=10.000000, c=2.500000)$,



H0: $X1 \sim \text{Triangular } 1(\mu, c)$, μ, c are unknown

μ point estimated value=10.030982 (MLE)

c point estimated value=2.285401 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under $H0$ of sample value $X1$

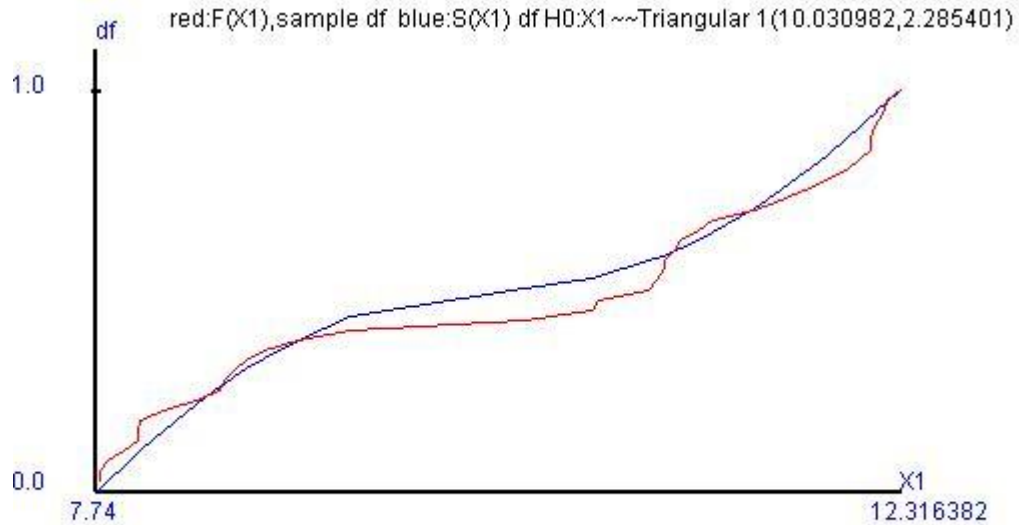
$\text{MAX}|F(X1)-S(X1)|=0.081133$

$0.200000 < \text{p-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

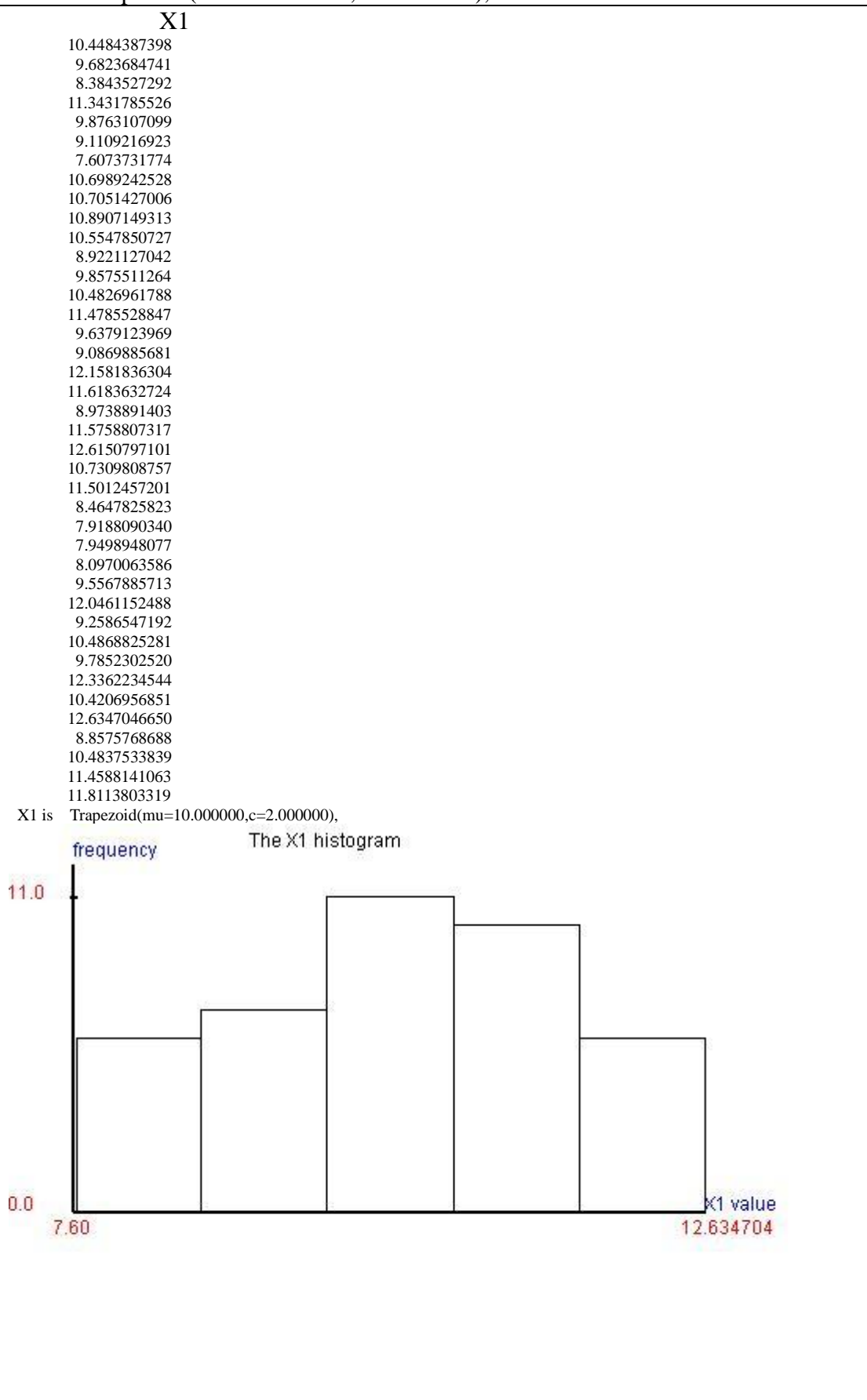
vertical axis is cumulative probability under $H0$ from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot14_image.jpg

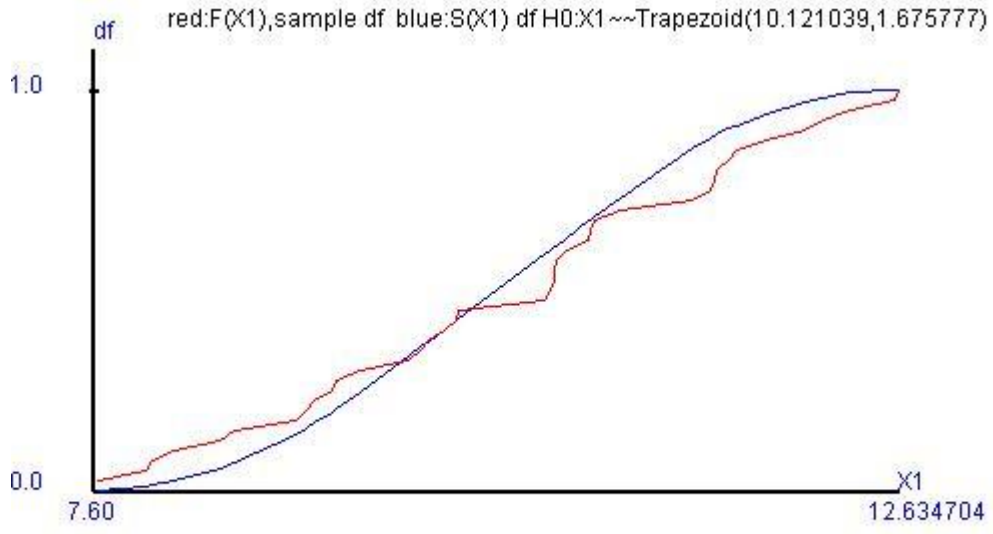


5.2.15) The population distribution is trapezoid distribution.

X1 is Trapezoid($\mu=10.000000, c=2.000000$),

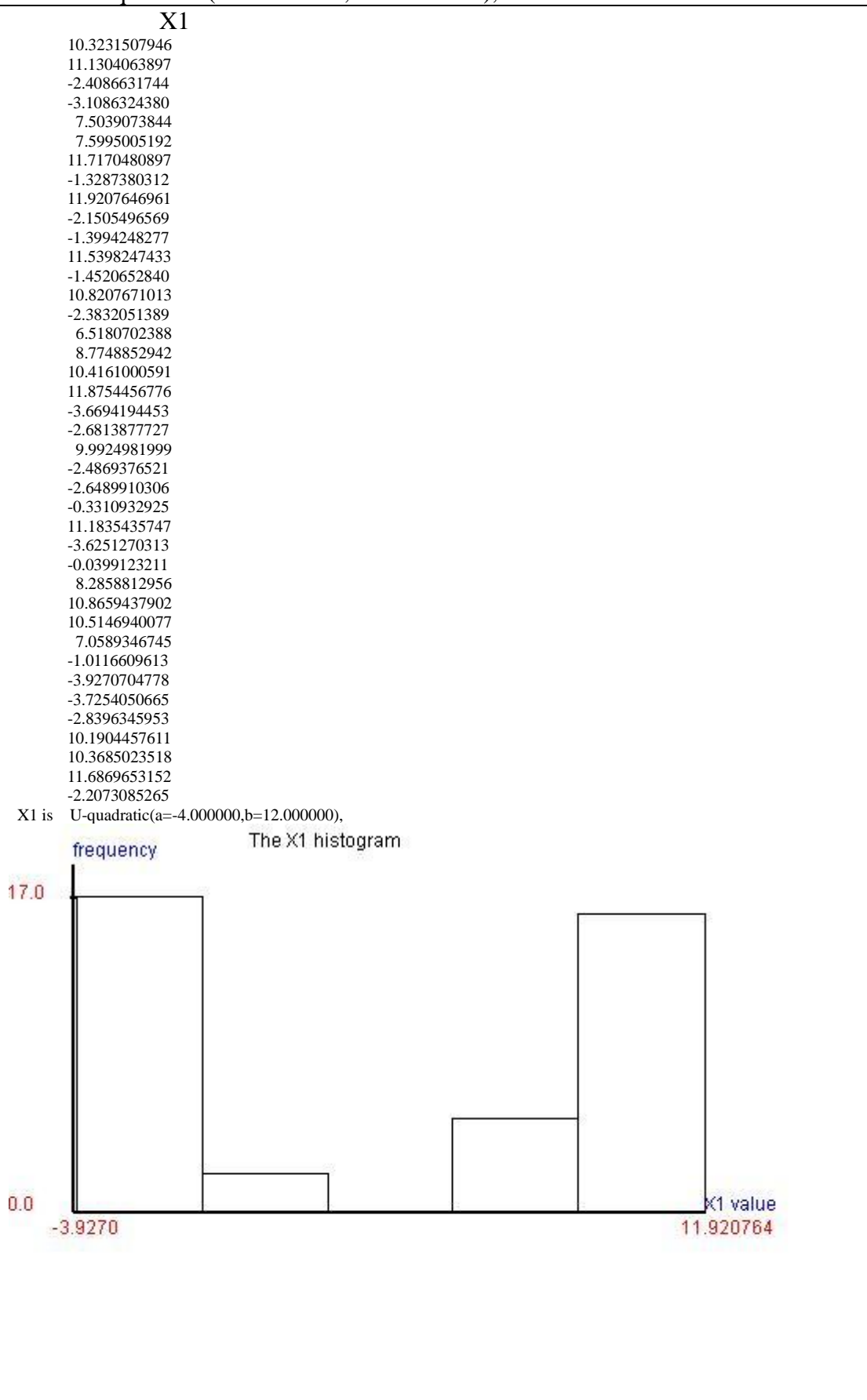


$H_0: X_1 \sim \text{Trapezoid}(\mu, c)$, μ, c are unknown
 μ point estimated value = 10.121039 (MLE)
 c point estimated value = 1.675777 (MLE)
 The Kolmogorov Smirnov goodness of fit test
 $F(X_1)$ is cumulative relative frequency of sample value X_1
 $S(X_1)$ is cumulative probability under H_0 of sample value X_1
 $\text{MAX}|F(X_1) - S(X_1)| = 0.126905$
 $0.200000 < \text{p-value} < 1.000000$
 horizon axis is samples cumulative relative frequency,
 vertical axis is cumulative probability under H_0 from sample values.
 The df comparison plot images is stored in c:\book_01\K_S_plot15_image.jpg



5.2.16) The population distribution is U quadratic distribution.

X1 is U-quadratic(a=-4.000000,b=12.000000),



H0: $X1 \sim U_quadratic(a,b)$, a,b are unknown

a point estimated value=-3.927070 (MLE)

b point estimated value=11.920765 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under $H0$ of sample value $X1$

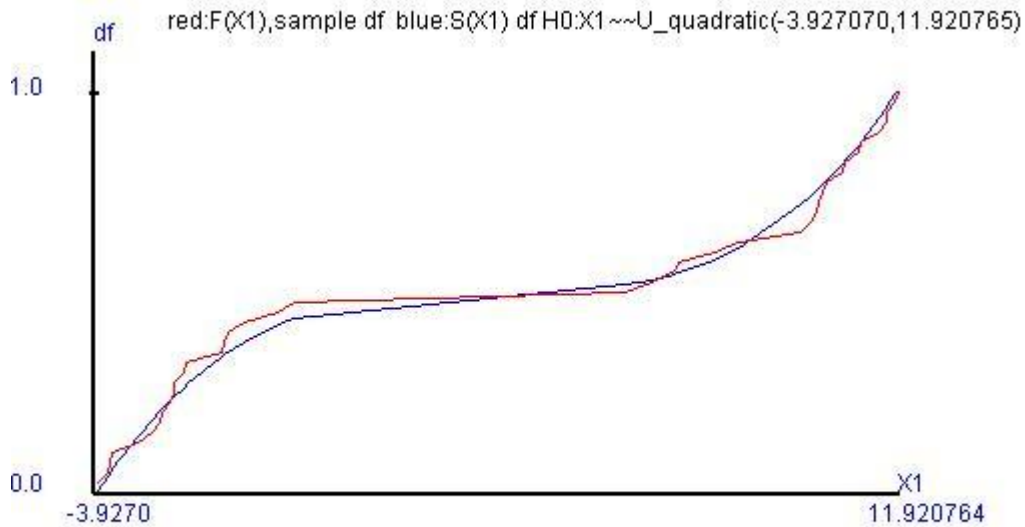
$MAX|F(X1)-S(X1)|=0.066600$

$0.200000 < p\text{-value} < 1.000000$

horizon axis is samples cumulative relative frequency,

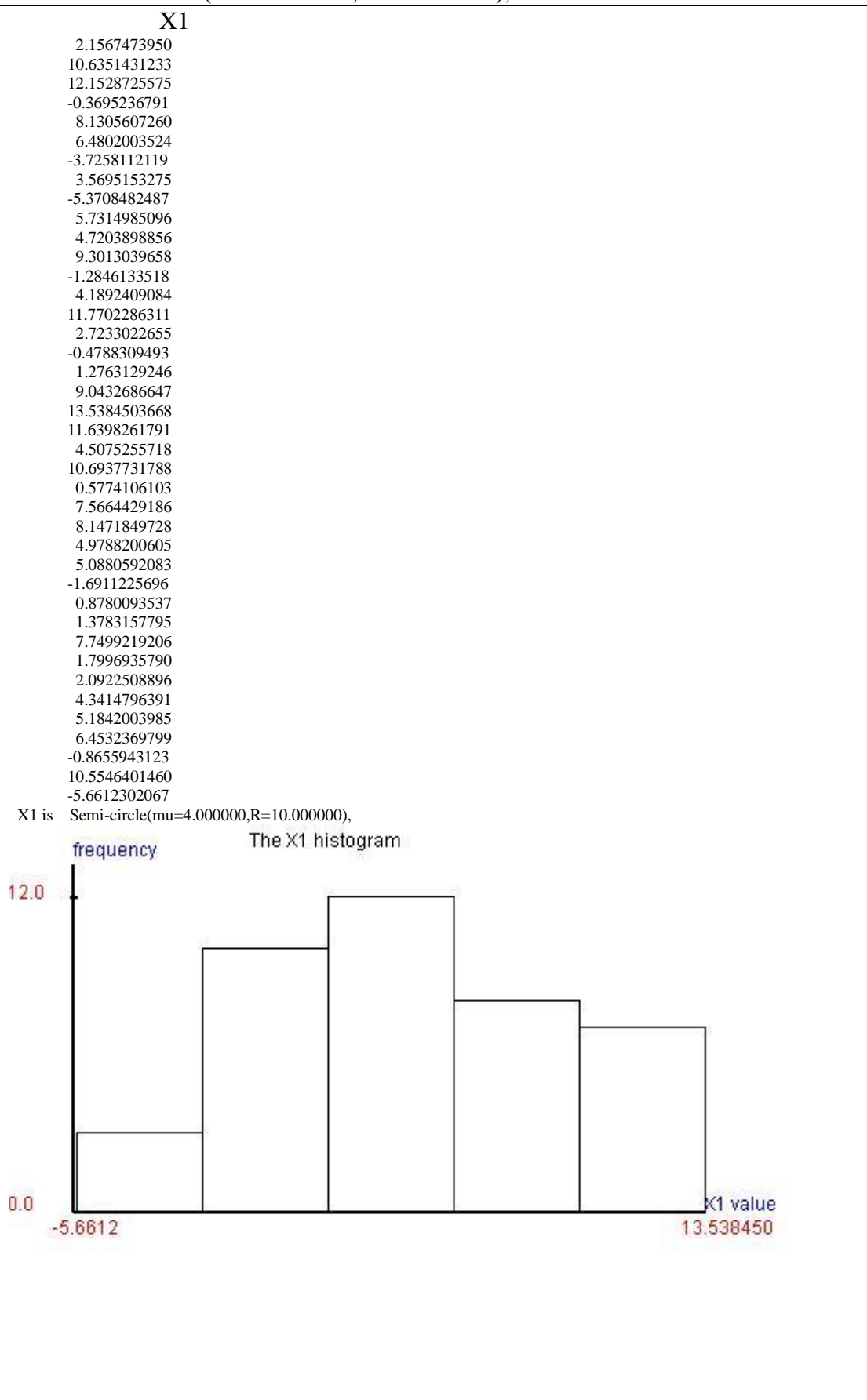
vertical axis is cumulative probability under $H0$ from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot16_image.jpg



5.2.17)The population distribution is semi circle distribution.

X1 is Semi-circle($\mu=4.000000,R=10.000000$),



H0: $X1 \sim \text{Semi-circle}(\mu, R)$, μ, R are unknown

μ point estimated value=4.490056 (MLE)

R point estimated value=9.599840 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under $H0$ of sample value $X1$

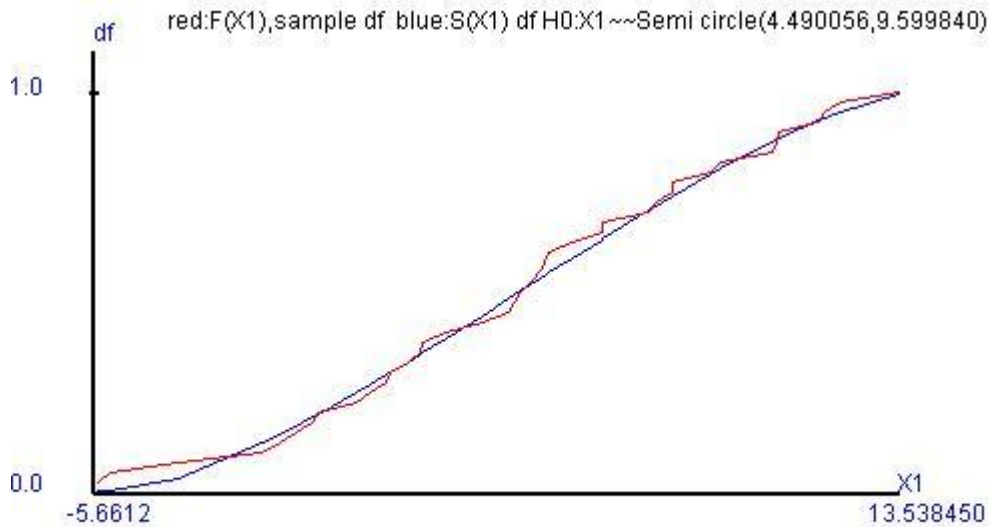
$\text{MAX}|F(X1)-S(X1)|=0.054000$

$0.200000 < \text{p-value} < 1.000000$

horizon axis is samples cumulative relative frequency,

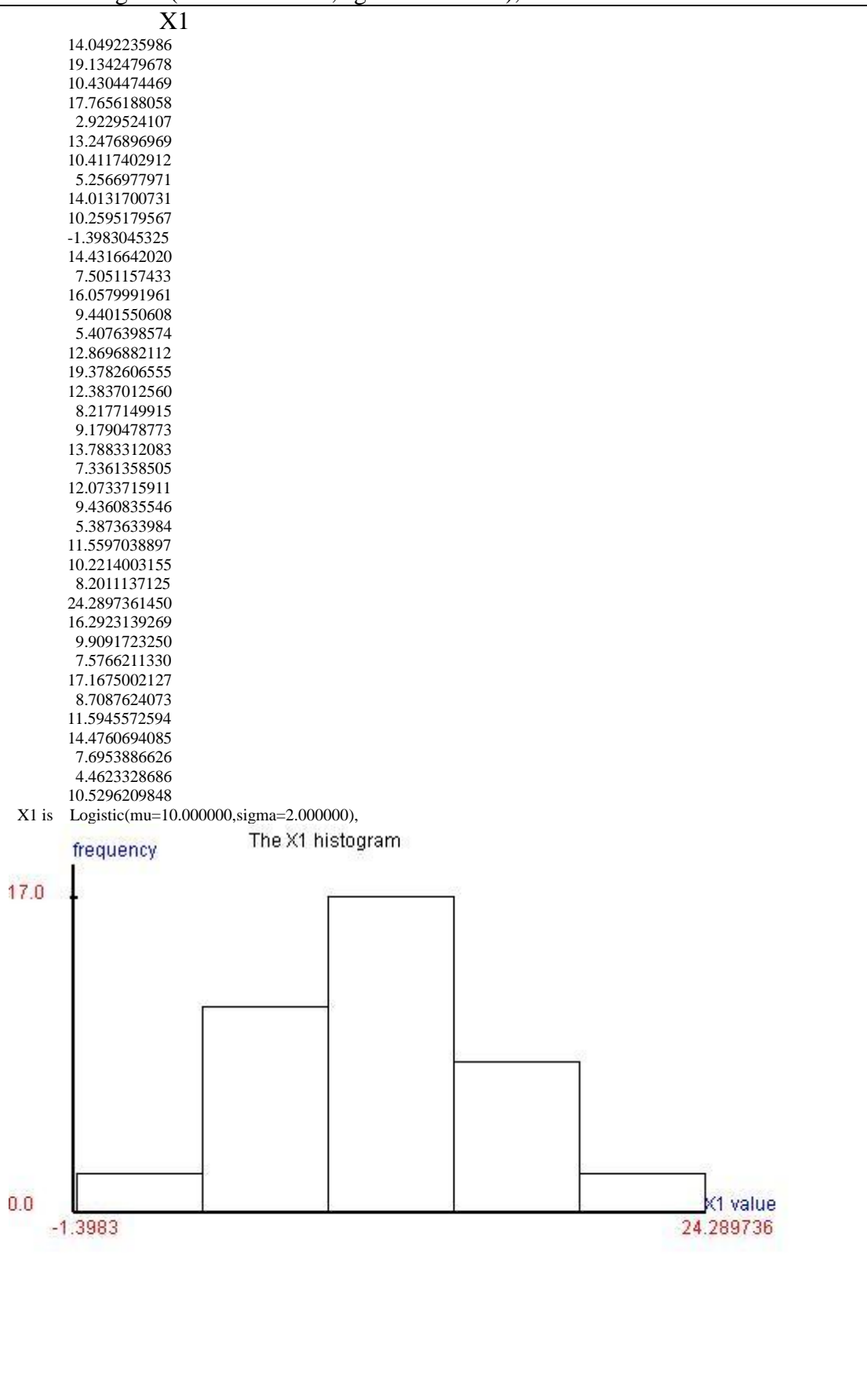
vertical axis is cumulative probability under $H0$ from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot17_image.jpg



5.2.18) The population distribution is logistic distribution.

X1 is Logistic($\mu=10.000000, \sigma=2.000000$),



H0: $X1 \sim \text{Logistic}(\mu, \sigma)$, μ, σ are unknown

μ point estimated value=11.041739 (MME)

σ point estimated value=2.721875 (MME)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under H0 of sample value $X1$

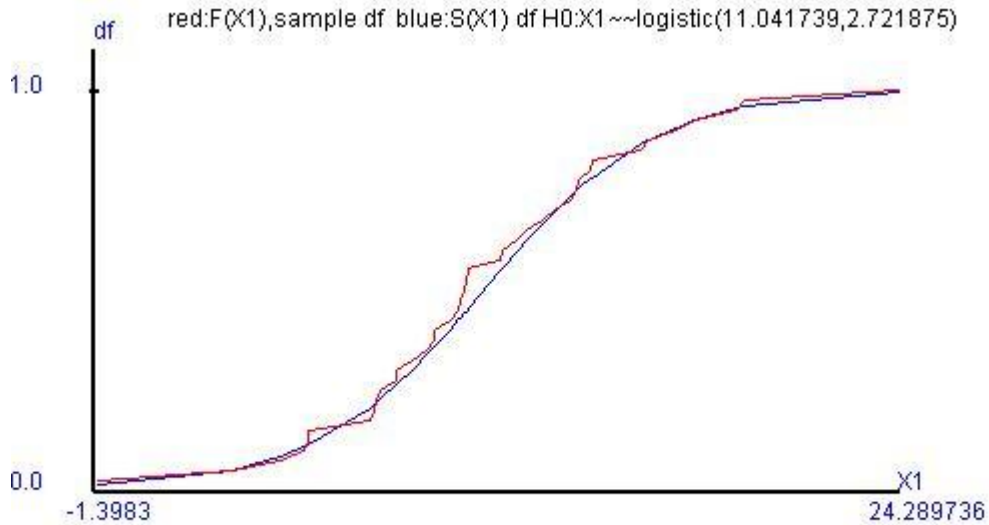
$\text{MAX}|F(X1)-S(X1)|=0.096899$

$0.200000 < \text{p-value} < 1.000000$

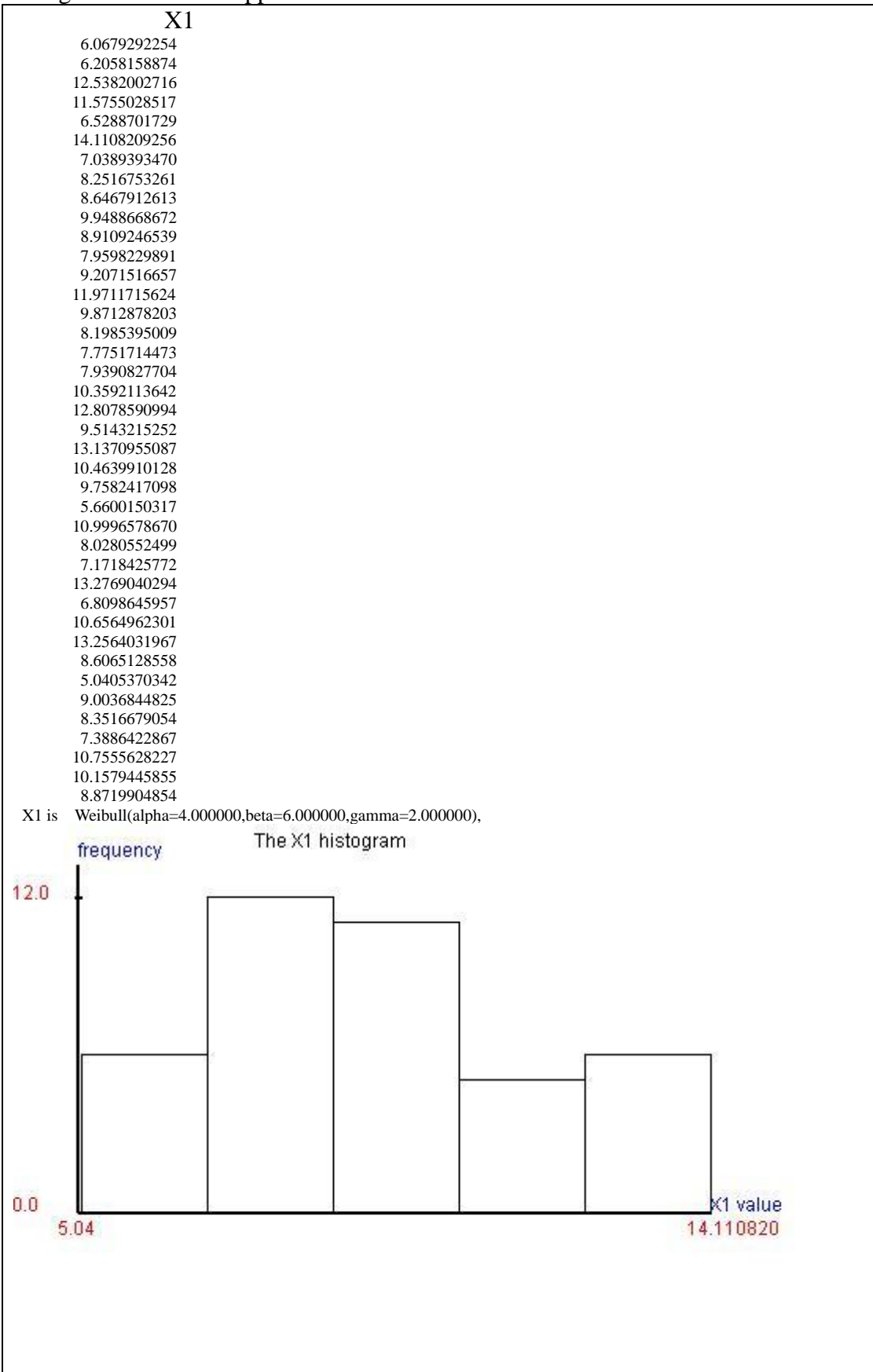
horizontal axis is samples cumulative relative frequency,

vertical axis is cumulative probability under H0 from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot18_image.jpg



5.2.19)The population distribution is weibull distribution.
 X1 is Weibull(alpha=4.000000,beta=6.000000,gamma=2.000000),
 The gamma value is supposed to 2.



H0: $X1 \sim \text{Weibull}(\alpha, \beta, \gamma=2.000000)$, α, β are unknown

α point estimated value=5.040537 (MLE)

β point estimated value=4.845011 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under $H0$ of sample value $X1$

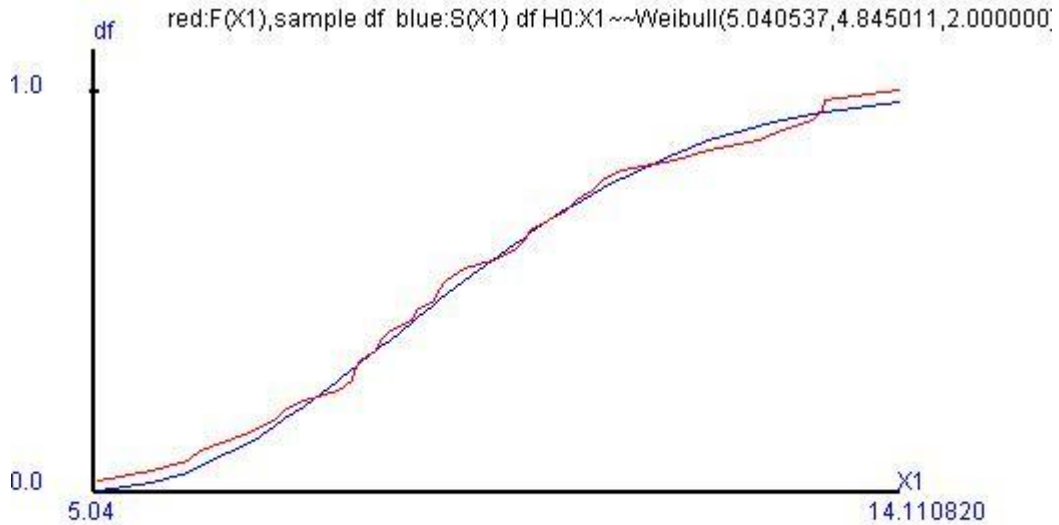
$\text{MAX}|F(X1)-S(X1)|=0.043796$

$0.200000 < \text{p-value} < 1.000000$

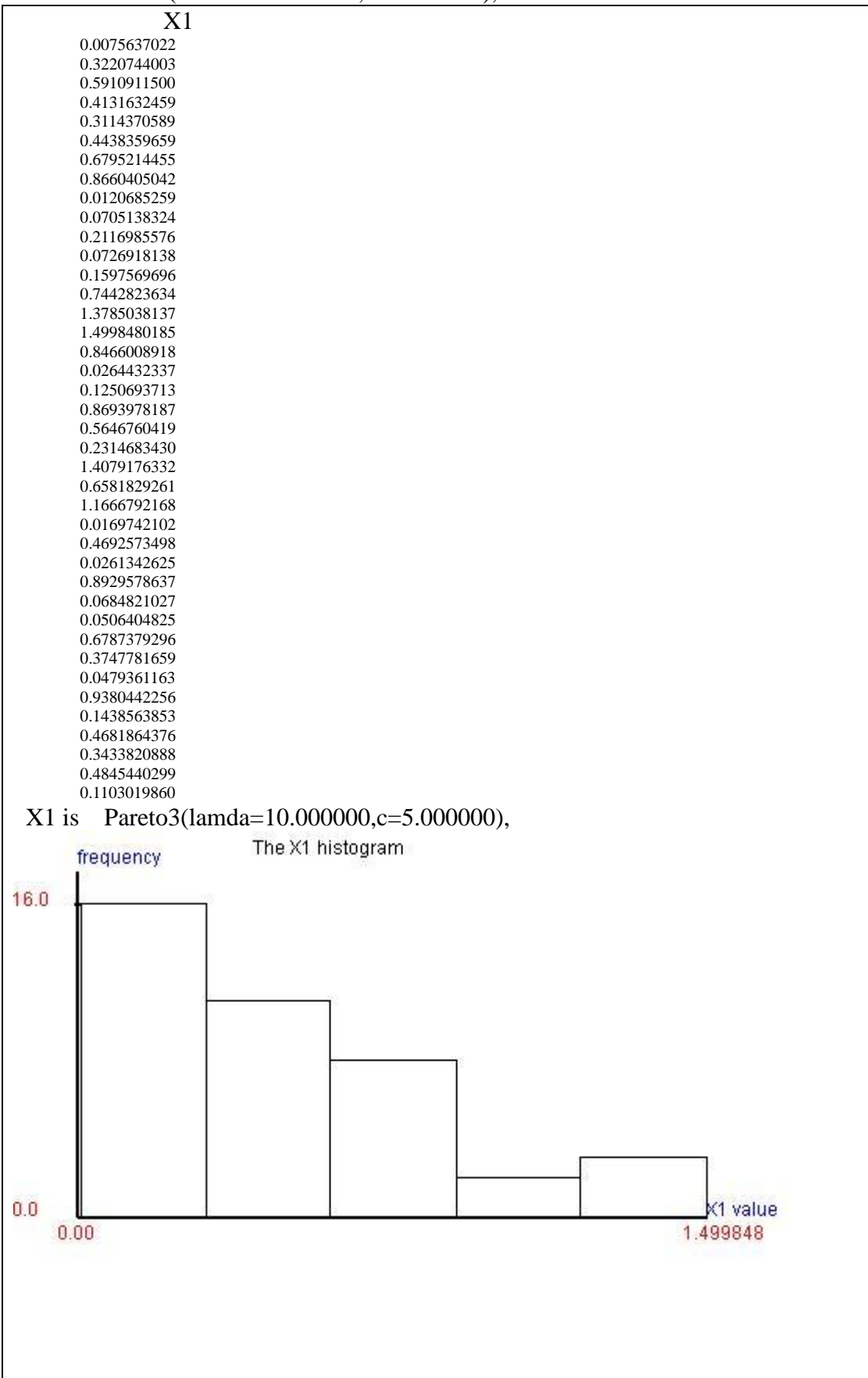
horizontal axis is samples cumulative relative frequency,

vertical axis is cumulative probability under $H0$ from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot19_image.jpg



5.2.20)The population distribution is pareto 3 distribution
 X1 is Pareto3(lamda=10.000000,c=5.000000),



H0: $X1 \sim \text{Pareto } 3(\lambda, c)$, λ, c are unknown

λ point estimated value = 2.161885 (MLE)

c point estimated value = 1.499848 (MLE)

The Kolmogorov Smirnov goodness of fit test

$F(X1)$ is cumulative relative frequency of sample value $X1$

$S(X1)$ is cumulative probability under $H0$ of sample value $X1$

$\text{MAX}|F(X1) - S(X1)| = 0.148164$

$0.200000 < \text{p-value} < 1.000000$

horizontal axis is samples cumulative relative frequency,

vertical axis is cumulative probability under $H0$ from sample values.

The df comparison plot images is stored in c:\book_01\K_S_plot22_image.jpg

