

SimpleCaseEstimatorBehavior

Xiang Ji

9/29/2017

This R markdown file compares the behavior of IGC tract length estimates from MLE and MCLE (maximum composite likelihood estimate).

Simple case: consider a sequence region with three symbols (1, 2, 3):

1. definitely experienced IGC.

2. definitely not experienced IGC.

3. no information.

Case 1, 2 correspond to sites that two paralogs are at different states before the IGC event.

Consider $\eta t \ll 1$, such that there is at most one IGC event in the region and we observe

233...3133...3133...32

which contains substrings of 3s with length a , $L - 2$ and b .

Full Likelihood

The full likelihood is:

$$l_F = \sum_{i=0}^a \sum_{j=0}^b \Pr(L + i + j) = p(1-p)^{L-1} \left[\frac{1 - (1-p)^{a+1}}{p} \right] \left[\frac{1 - (1-p)^{b+1}}{p} \right]$$
$$\ln(l_F) = -\ln(p) + (L-1)\ln(1-p) + \ln[1 - (1-p)^{a+1}] + \ln[1 - (1-p)^{b+1}]$$

when $a = b = 0$, $\hat{p}_{MLE} = \frac{1}{L}$ and $E(\frac{1}{\hat{p}_{MLE}}) = E(L) = p$ is unbiased.

when $a = b = \infty$, $\hat{p}_{MLE} = 0$

Composite Likelihood

The composite likelihood (pair-site) is:

$$l_c = p(p-1)^{L-1} [1 - (1-p)^{a+1}] [1 - (1-p)^{a+L}] [1 - (1-p)^{b+1}] [1 - (1-p)^{b+L}]$$

$$\ln(l_c) = \ln(p) + (L-1)\ln(1-p) + \ln[1 - (1-p)^{a+1}] + \ln[1 - (1-p)^{a+L}] + \ln[1 - (1-p)^{b+1}] + \ln[1 - (1-p)^{b+L}]$$

when $a = b = \infty$, $\hat{p}_{MCLE} = \frac{1}{L}$, again $E(\frac{1}{\hat{p}_{MCLE}}) = E(L) = p$ is unbiased

Relationship between the two estimators

$$\ln(l_c) = \ln(l_F) + 2\ln(p) + \ln[1 - (1-p)^{a+L}] + \ln[1 - (1-p)^{b+L}]$$

$$\frac{d\ln(l_c)}{dp} = \frac{d\ln(l_F)}{dp} + \frac{2}{p} + \frac{(a+L)(1-p)^{a+L-1}}{1 - (1-p)^{a+L}} + \frac{(b+L)(1-p)^{b+L-1}}{1 - (1-p)^{b+L}}$$

And,

$$\frac{2}{p} + \frac{(a+L)(1-p)^{a+L-1}}{1 - (1-p)^{a+L}} + \frac{(b+L)(1-p)^{b+L-1}}{1 - (1-p)^{b+L}} > 0$$

so that we know

$$\hat{p}_{MCL E} \neq \hat{p}_{MLE}$$

and

$$\hat{p}_{MCL E} > \hat{p}_{MLE}, \forall a, b$$

Now, show the estimates and log likelihood surface for several parameter combinations.

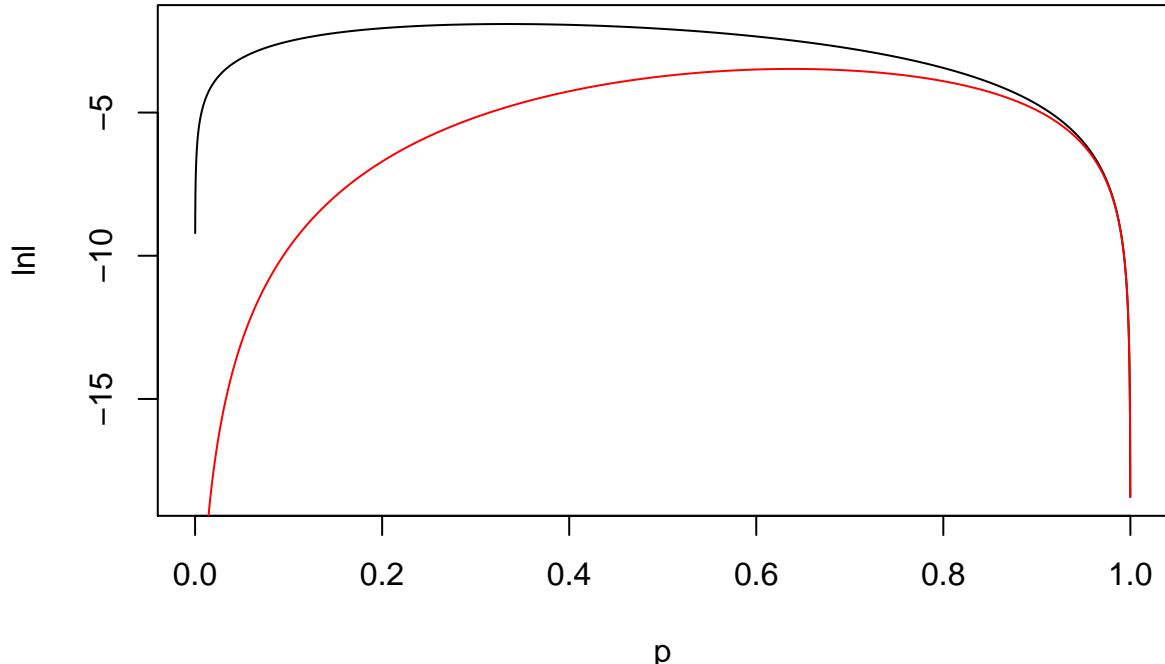
```

rm(list=ls()) # clean up workspace
a.list <- c(0,5,20)
b.list <- c(0,5,20)
L.list <- c(3, 10,50,100, 200, 300, 500)
p <- 1:9999 * 0.0001

for(a in a.list){
  for(b in b.list){
    for(L in L.list){
      lnl = -log(p) + (L-1)*log(1-p) + log(1-(1-p)^(a+1))+ log(1-(1-p)^(b+1))
      lnlc = log(p) + (L-1)*log(1-p) + log(1-(1-p)^(a+1)) + log(1-(1-p)^(a+L)) + log(1-(1-p)^(b+1)) + log(1-(1-p)^(b+L))
      plot(p, lnl, type ="l", main = paste("a=", a, ",b=", b, ",L=", L))
      lines(p, lnlc, type = "l", col = 2)
      print(matrix(c("MLE", which.max(lnl)/10000, "MCLE", which.max(lnlc)/10000,
                    "1/MLE", 10000/which.max(lnl), "1/MCLE", 10000/which.max(lnlc)), 2, 4))
    }
  }
}

```

a = 0 ,b = 0 ,L = 3

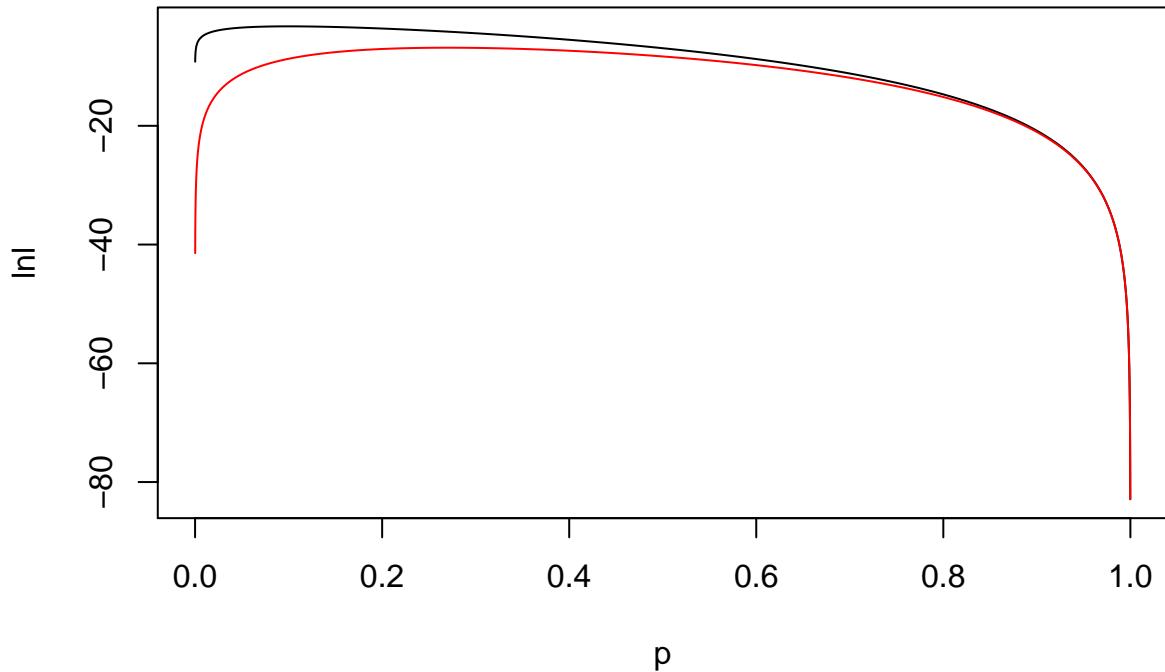


```

##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.3333" "0.6381" "3.000300030003" "1.56715248393669"

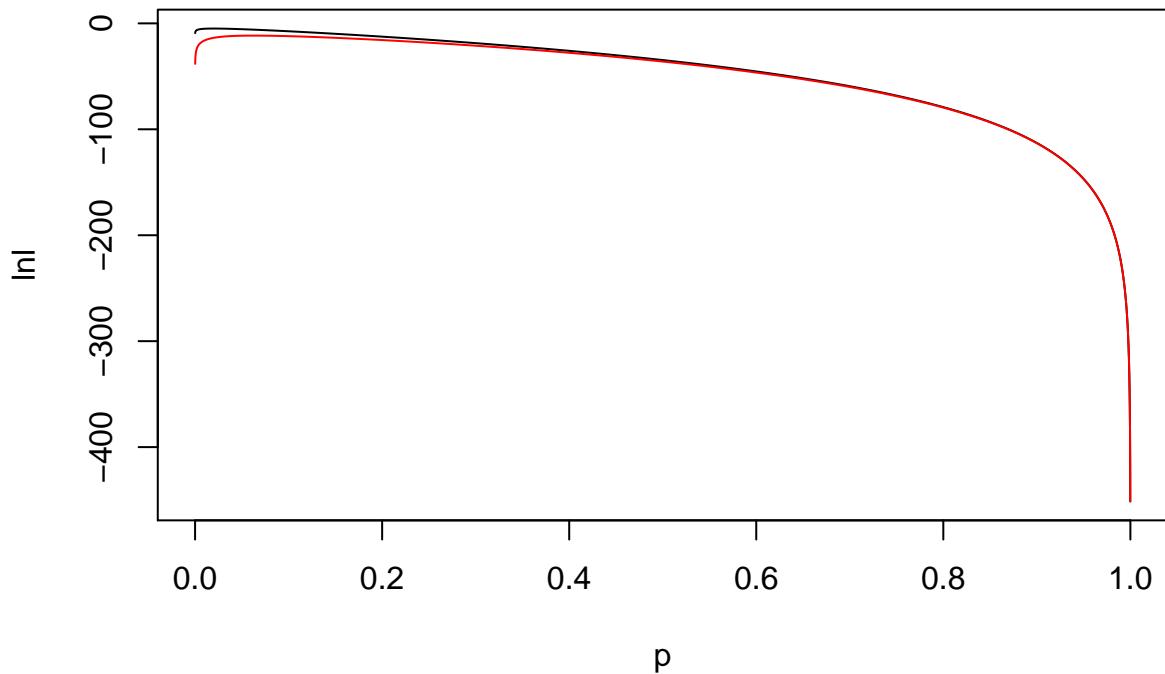
```

a= 0 ,b= 0 ,L= 10



```
##      [,1]   [,2]   [,3]   [,4]
## [1,] "MLE"  "MCLE" "1/MLE" "1/MCLE"
## [2,] "0.1"  "0.2702" "10"    "3.70096225018505"
```

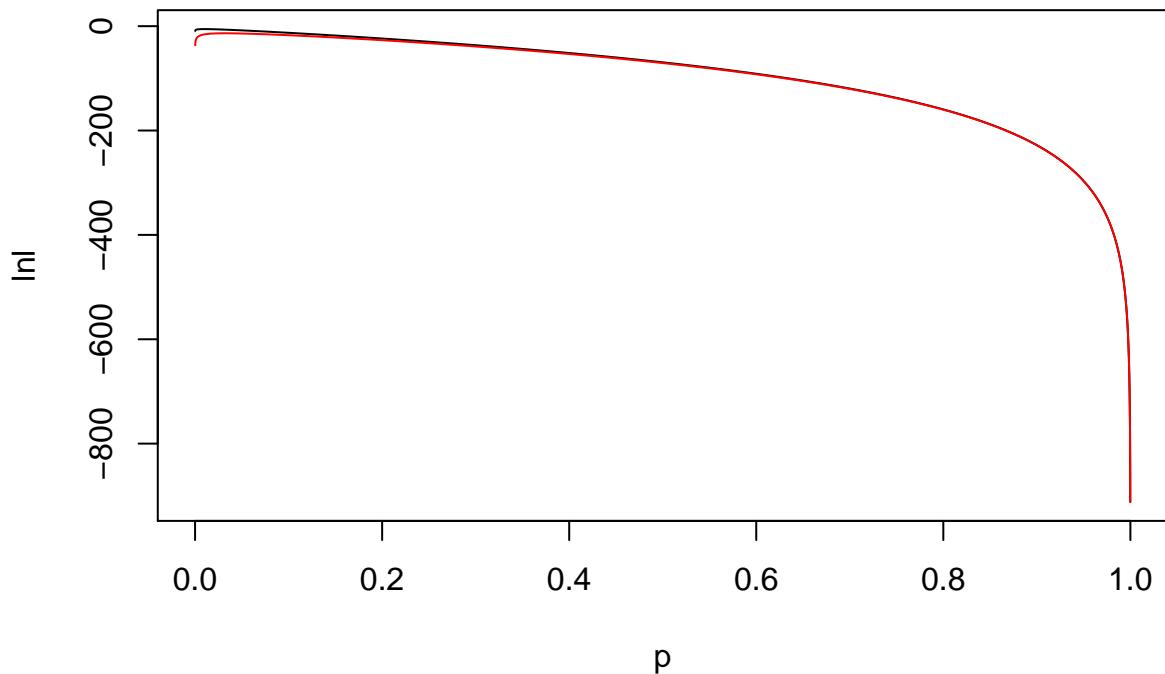
a= 0 ,b= 0 ,L= 50



```
##      [,1]   [,2]   [,3]   [,4]
```

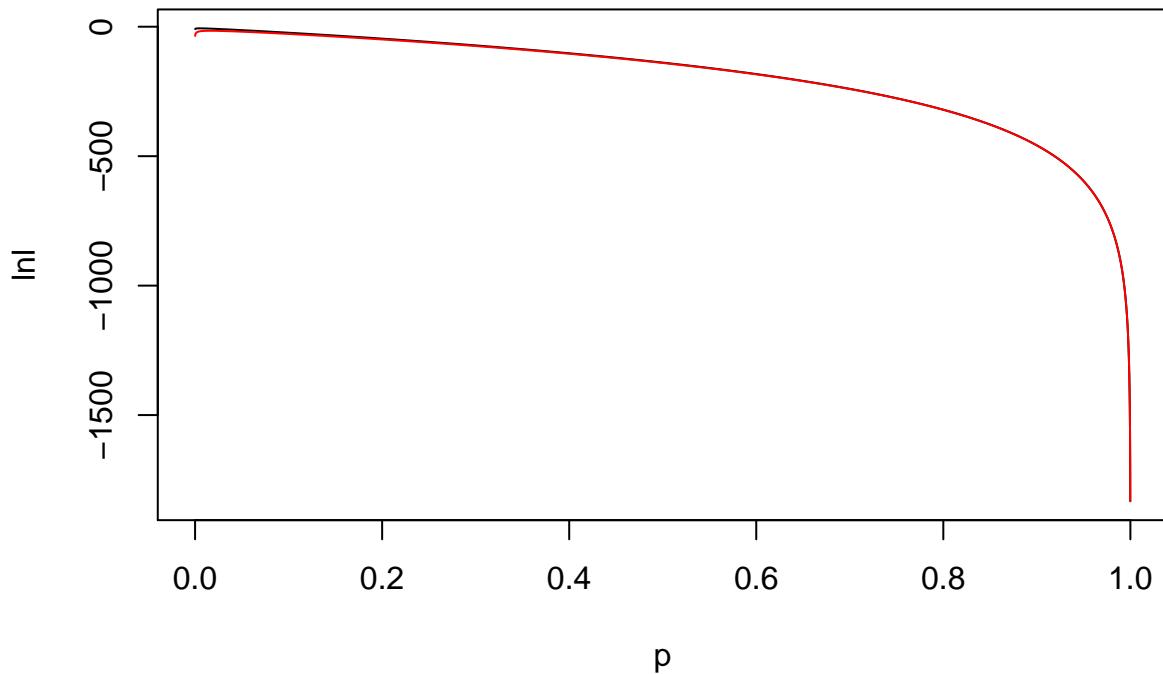
```
## [1,] "MLE"   "MCLE"   "1/MLE"  "1/MCLE"  
## [2,] "0.02"  "0.0626" "50"     "15.9744408945687"
```

a= 0 ,b= 0 ,L= 100



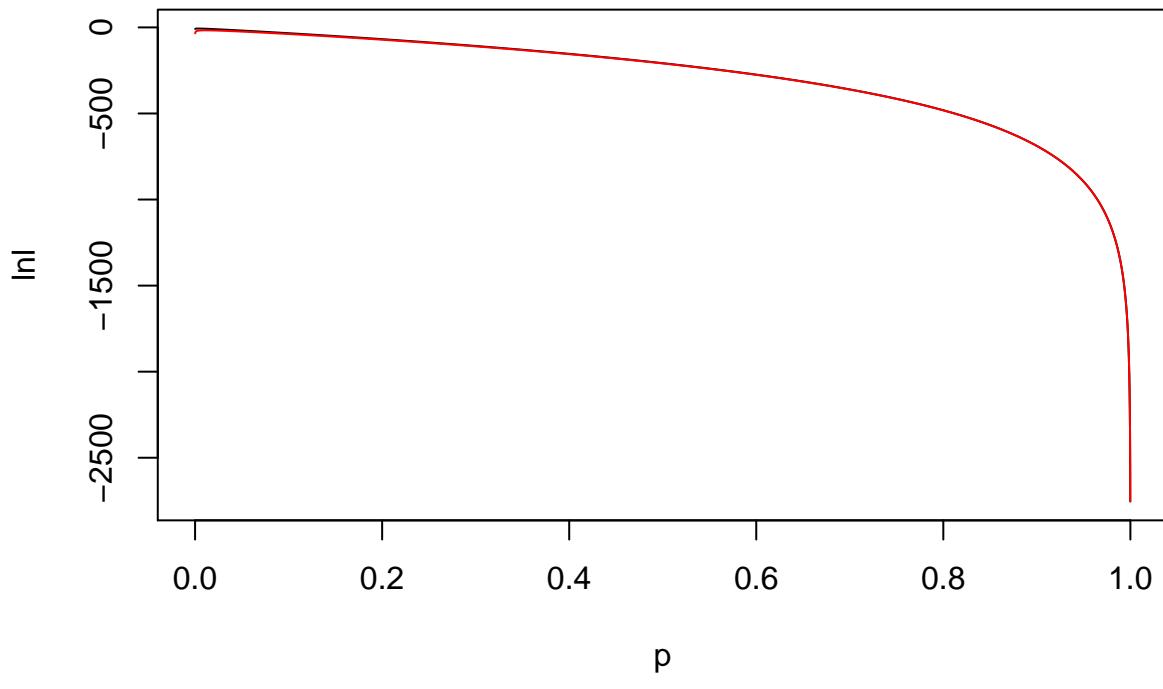
```
##      [,1]    [,2]    [,3]    [,4]  
## [1,] "MLE"   "MCLE"   "1/MLE"  "1/MCLE"  
## [2,] "0.01"  "0.0319" "100"    "31.3479623824451"
```

a= 0 ,b= 0 ,L= 200



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"    "1/MLE"   "1/MCLE"
## [2,] "0.005"  "0.0161"  "200"     "62.111801242236"
```

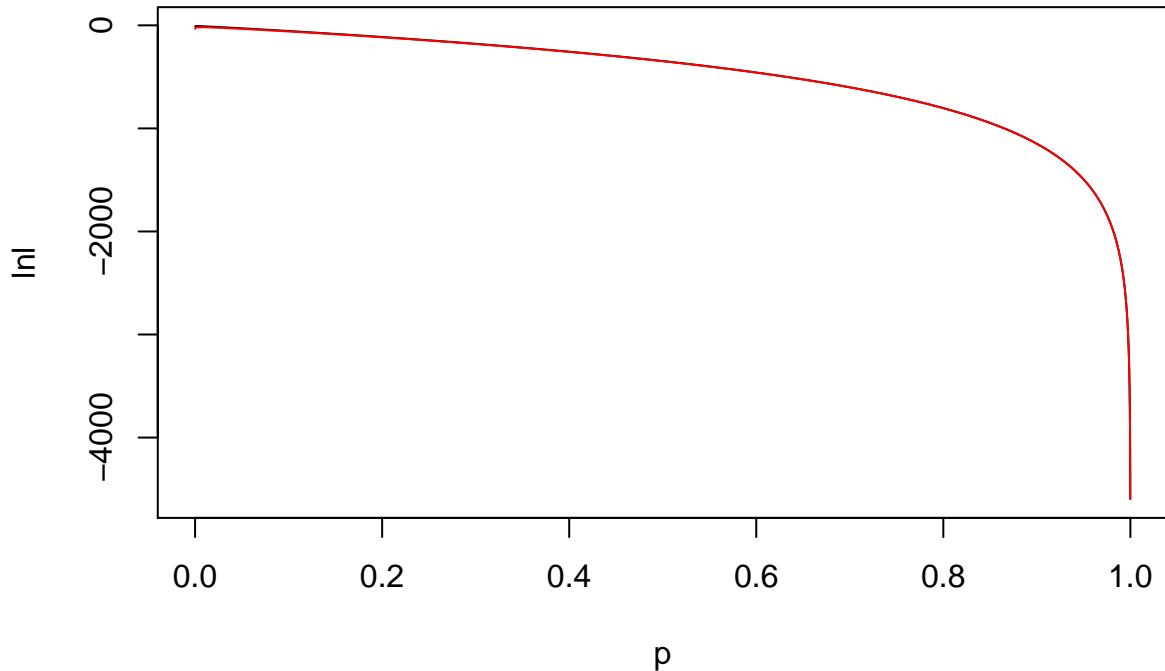
a= 0 ,b= 0 ,L= 300



```
##      [,1]      [,2]      [,3]      [,4]
```

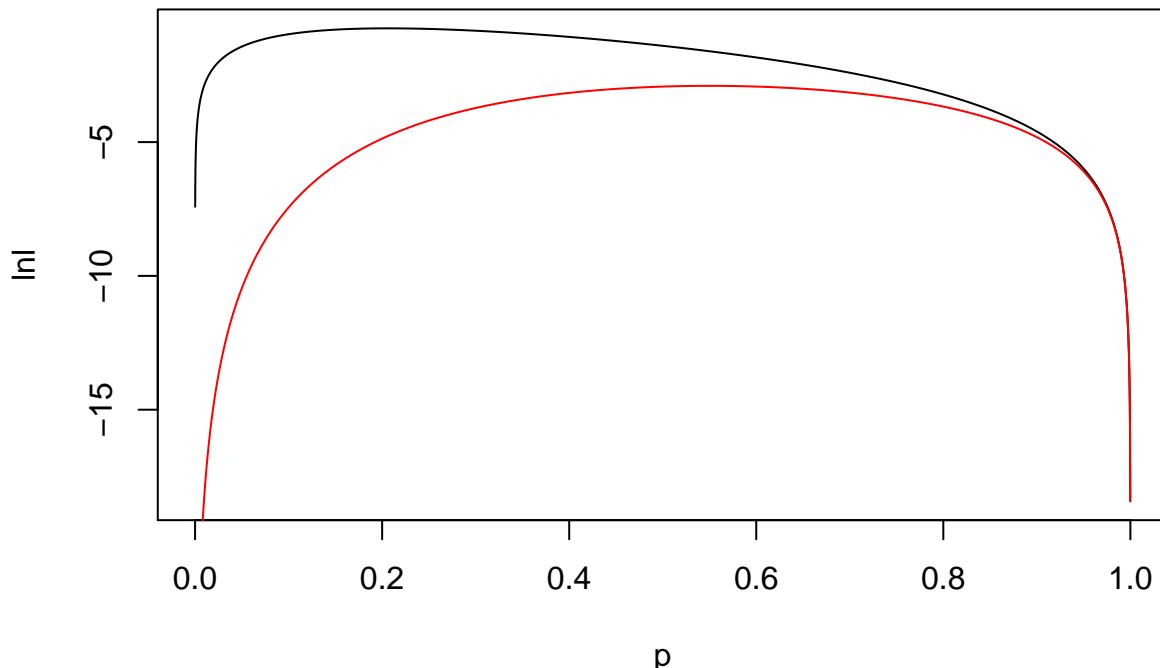
```
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0033"   "0.0108"   "303.030303030303" "92.5925925925926"
```

a= 0 ,b= 0 ,L= 500



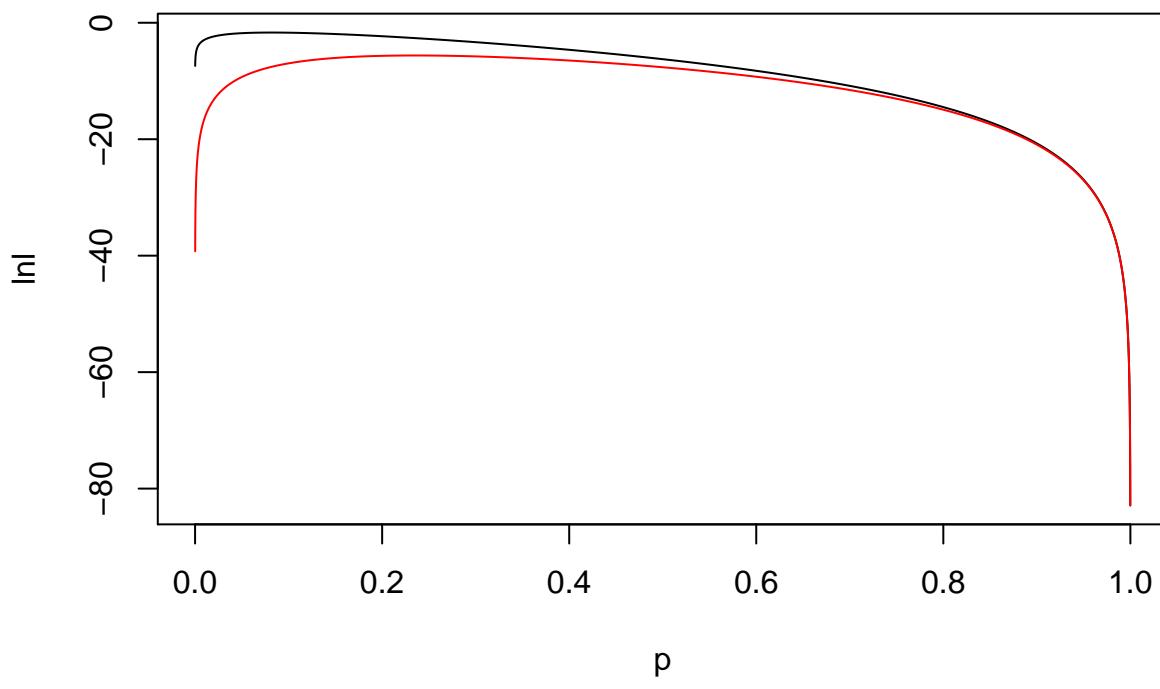
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.002"    "0.0065"   "500"       "153.846153846154"
```

a= 0 ,b= 5 ,L= 3



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"    "1/MLE"   "1/MCLE"
## [2,] "0.2063" "0.5501" "4.84730974309258" "1.81785129976368"
```

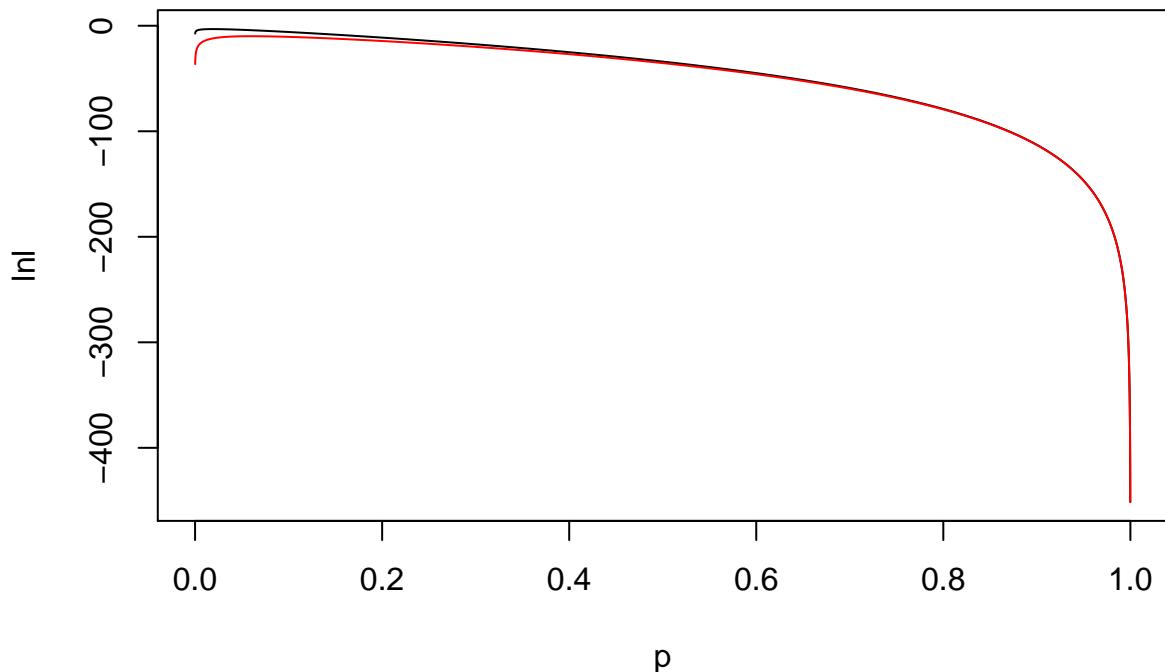
a= 0 ,b= 5 ,L= 10



```
##      [,1]      [,2]      [,3]      [,4]
```

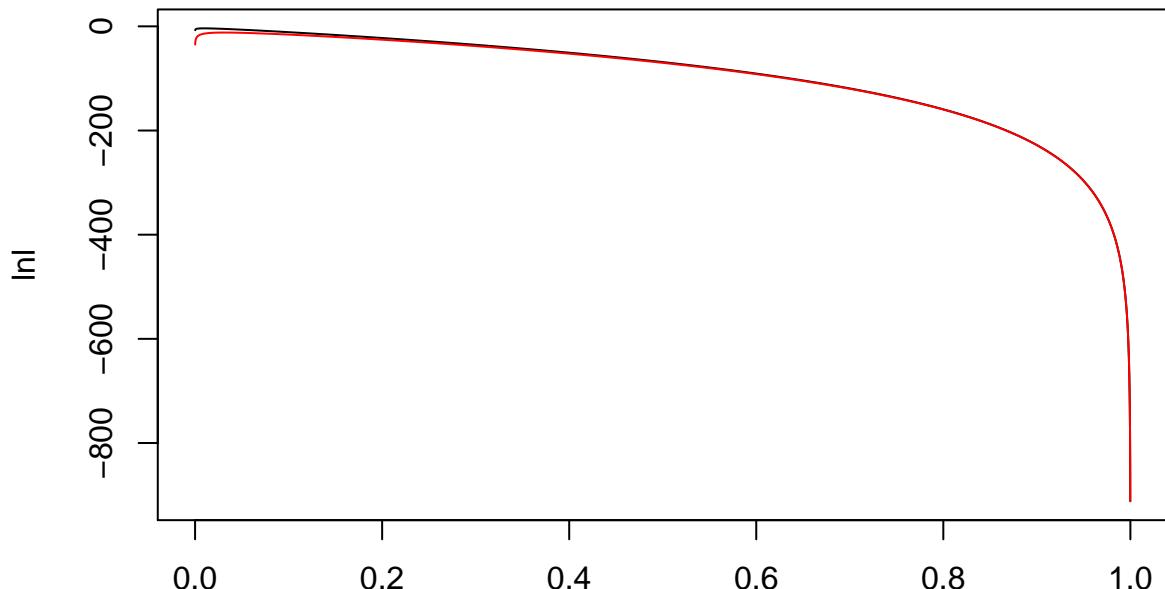
```
## [1,] "MLE"      "MCLE"      "1/MLE"          "1/MCLE"  
## [2,] "0.0816"   "0.2354"   "12.2549019607843" "4.24808836023789"
```

a= 0 ,b= 5 ,L= 50



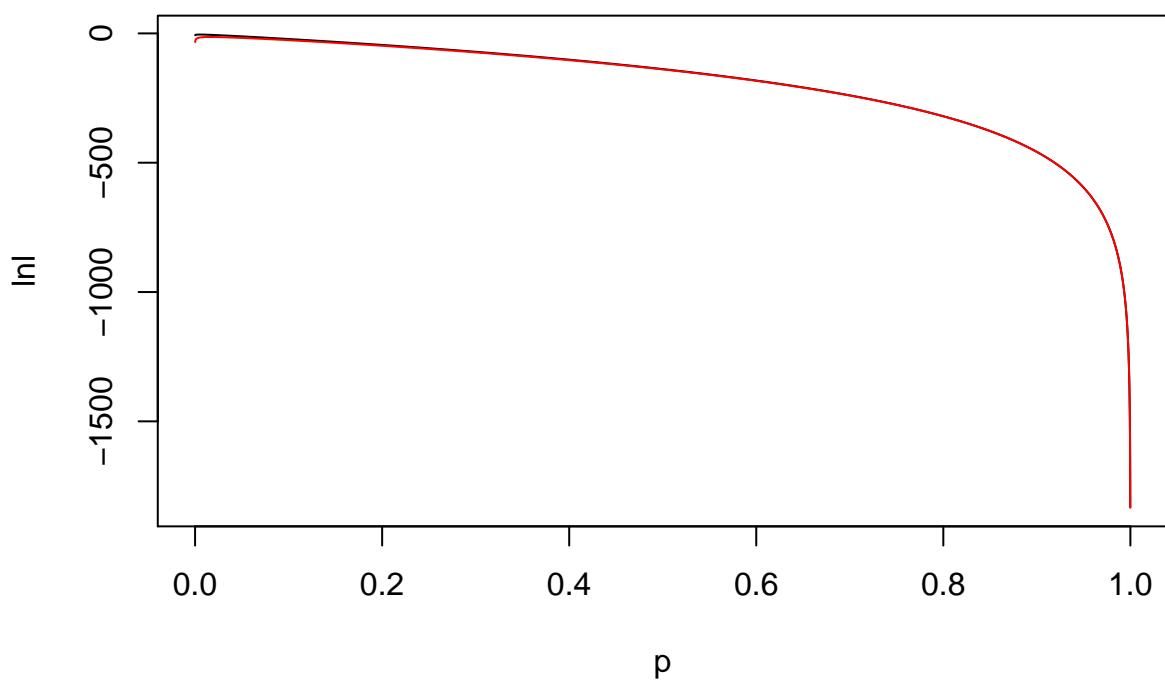
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0191"   "0.06"     "52.3560209424084" "16.6666666666667"
```

a= 0 ,b= 5 ,L= 100



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.0098" "0.0312" "102.040816326531" "32.0512820512821"
```

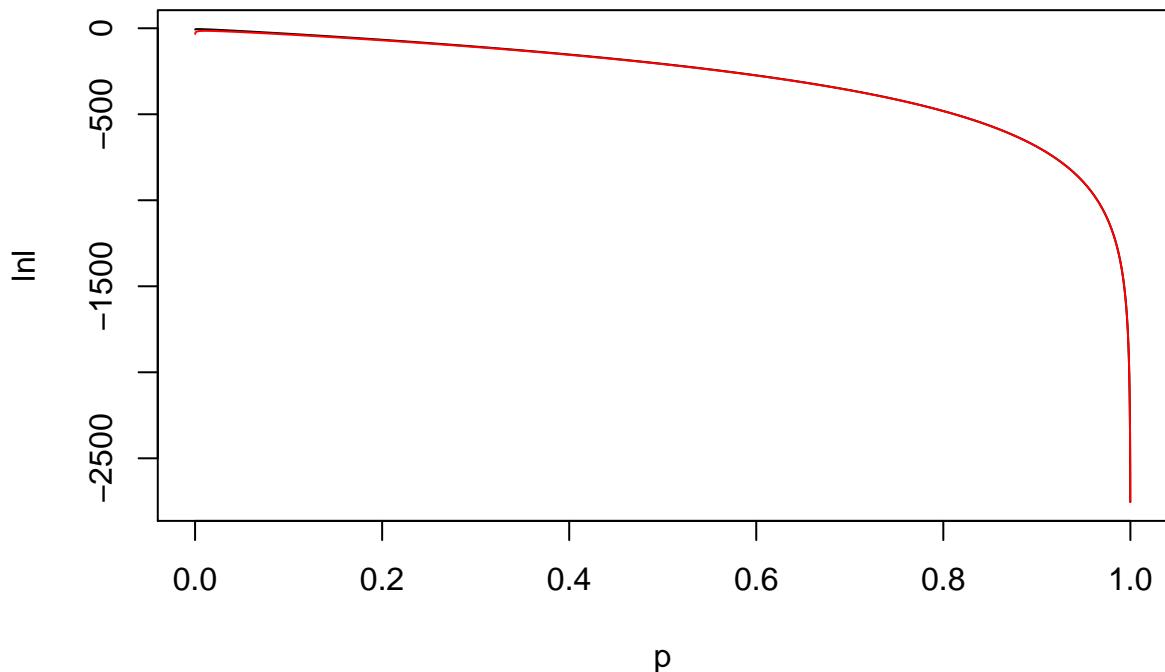
a= 0 ,b= 5 ,L= 200



```
##      [,1]      [,2]      [,3]      [,4]
```

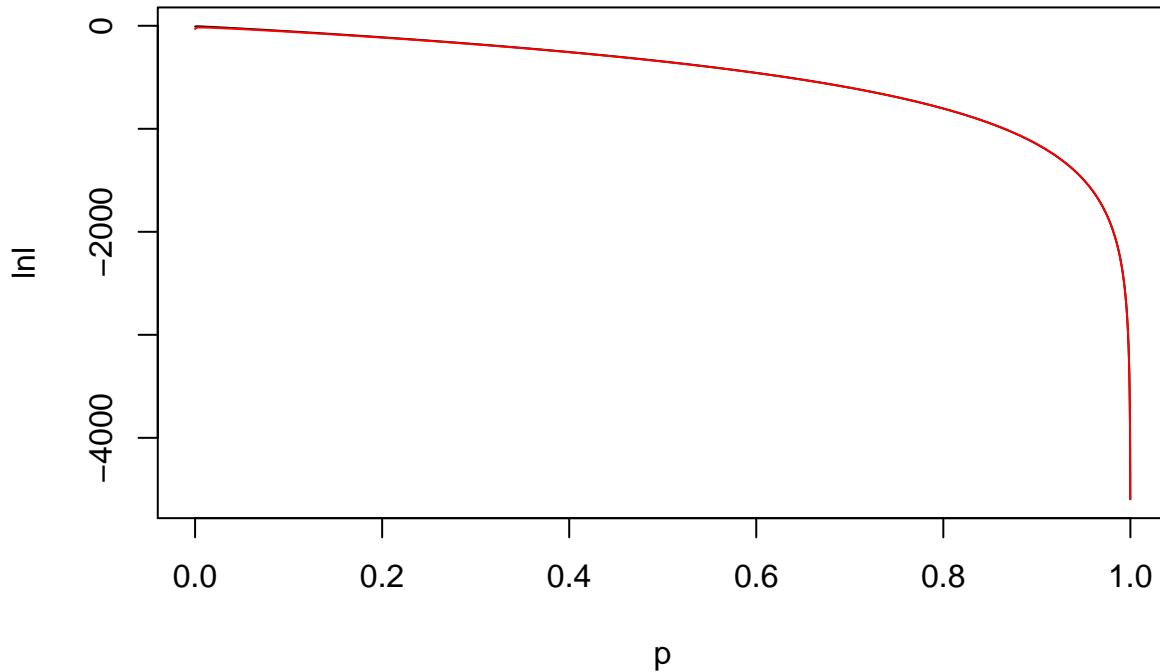
```
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0049"   "0.0159"   "204.081632653061" "62.8930817610063"
```

a= 0 ,b= 5 ,L= 300



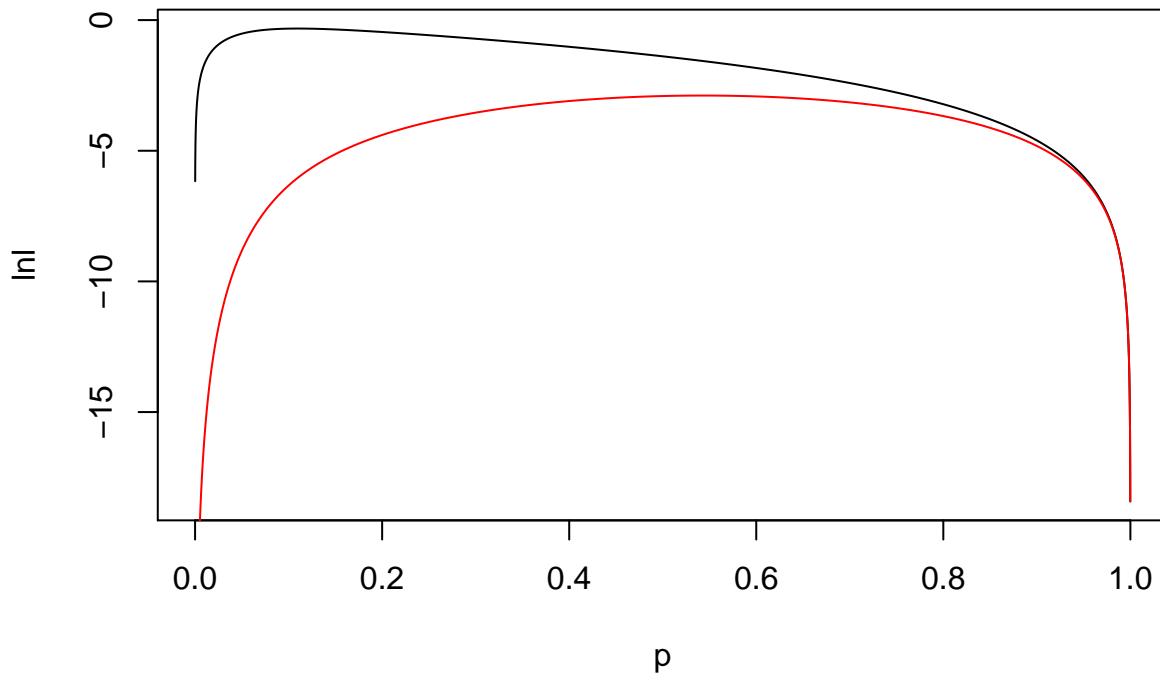
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0033"   "0.0107"   "303.030303030303" "93.4579439252336"
```

a= 0 ,b= 5 ,L= 500



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"    "1/MLE"   "1/MCLE"
## [2,] "0.002"  "0.0065" "500"     "153.846153846154"
```

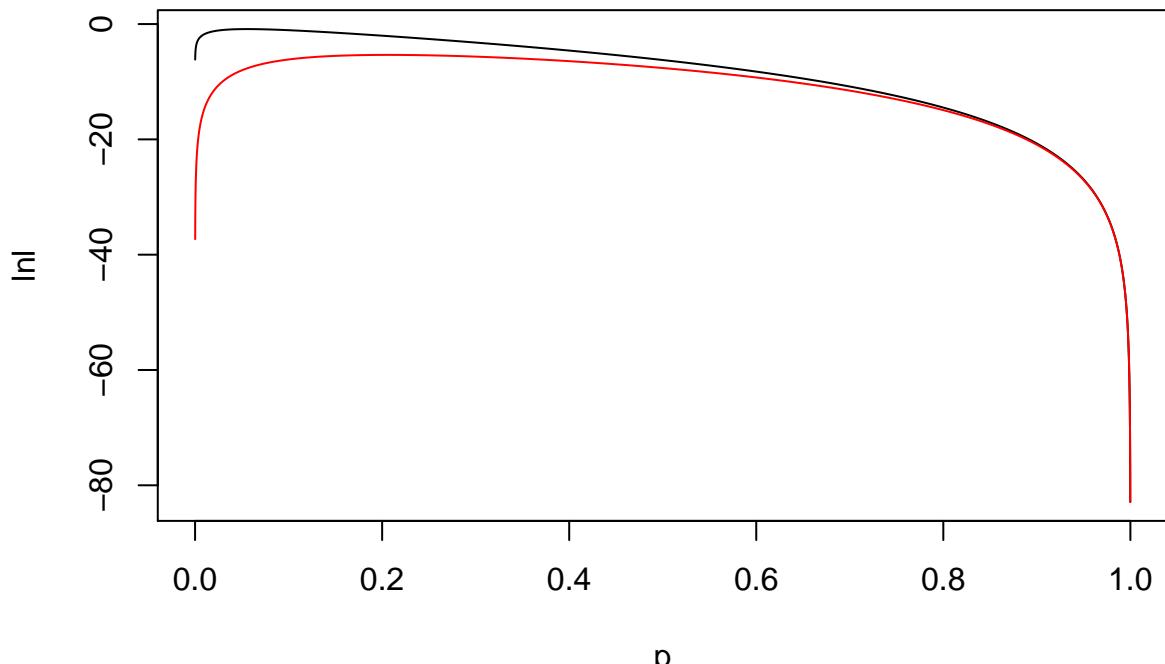
a= 0 ,b= 20 ,L= 3



```
##      [,1]      [,2]      [,3]      [,4]
```

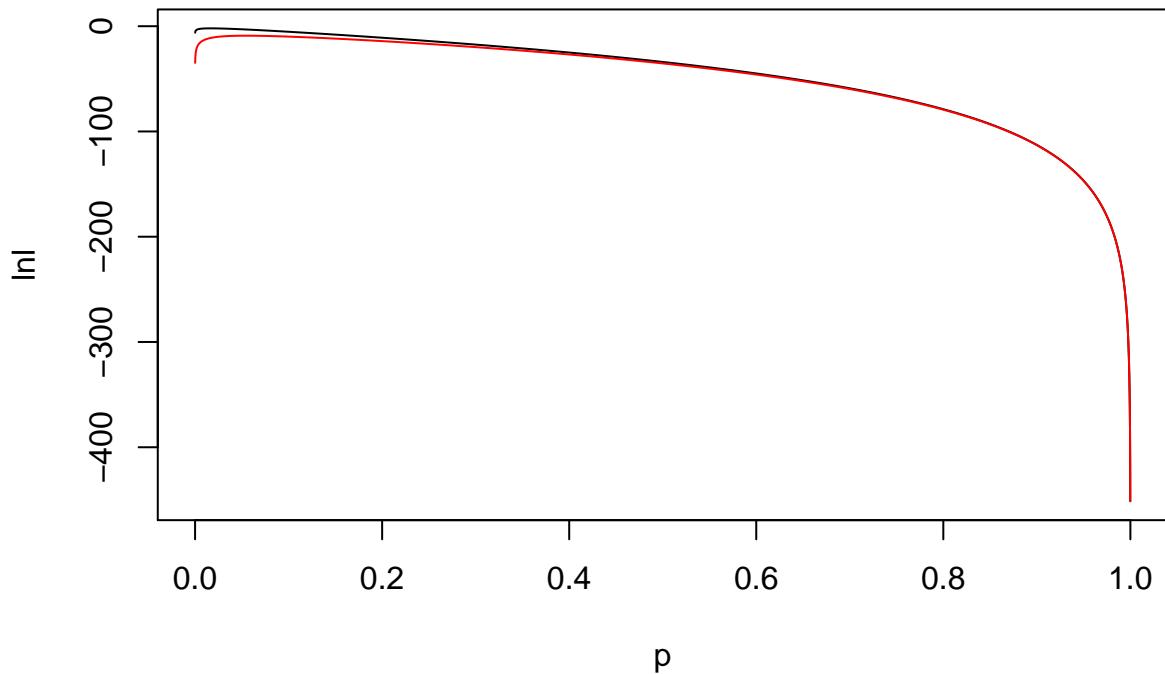
```
## [1,] "MLE"      "MCLE"     "1/MLE"          "1/MCLE"  
## [2,] "0.1098"   "0.543"    "9.10746812386157" "1.84162062615101"
```

a= 0 ,b= 20 ,L= 10



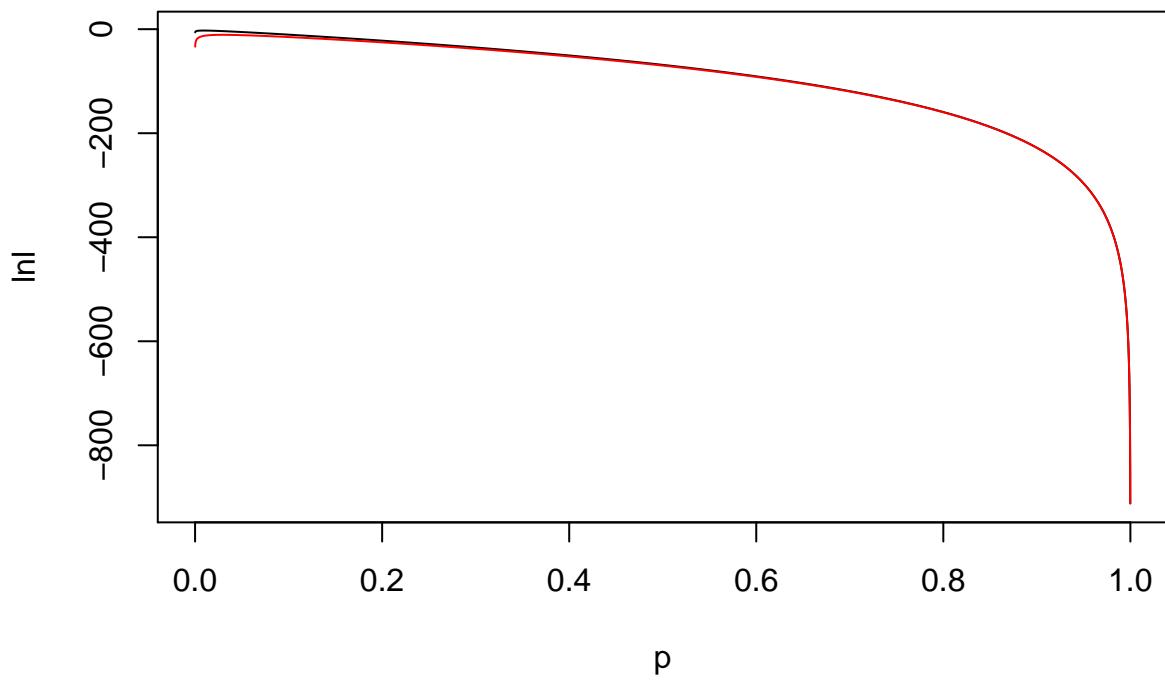
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"     "1/MLE"     "1/MCLE"  
## [2,] "0.0557"   "0.2062"   "17.9533213644524" "4.84966052376334"
```

a= 0 ,b= 20 ,L= 50



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"    "1/MLE"   "1/MCLE"
## [2,] "0.0168" "0.0543" "59.5238095238095" "18.4162062615101"
```

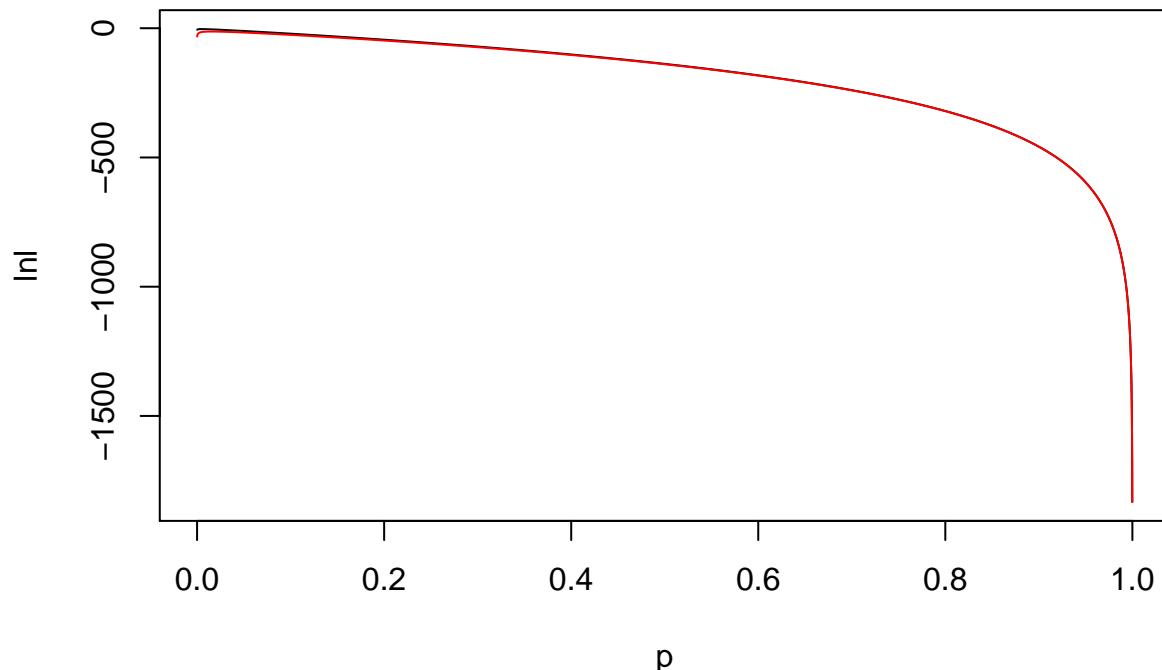
a= 0 ,b= 20 ,L= 100



```
##      [,1]      [,2]      [,3]      [,4]
```

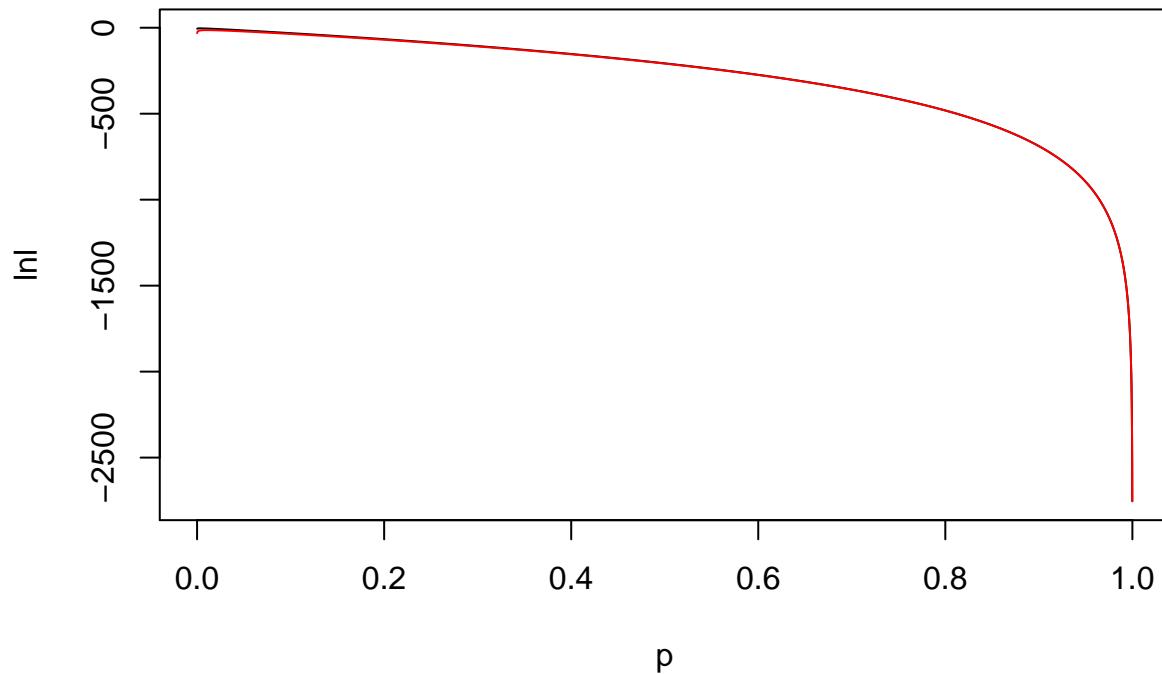
```
## [1,] "MLE"      "MCLE"      "1/MLE"          "1/MCLE"  
## [2,] "0.0091"   "0.0294"   "109.89010989011" "34.0136054421769"
```

a= 0 ,b= 20 ,L= 200



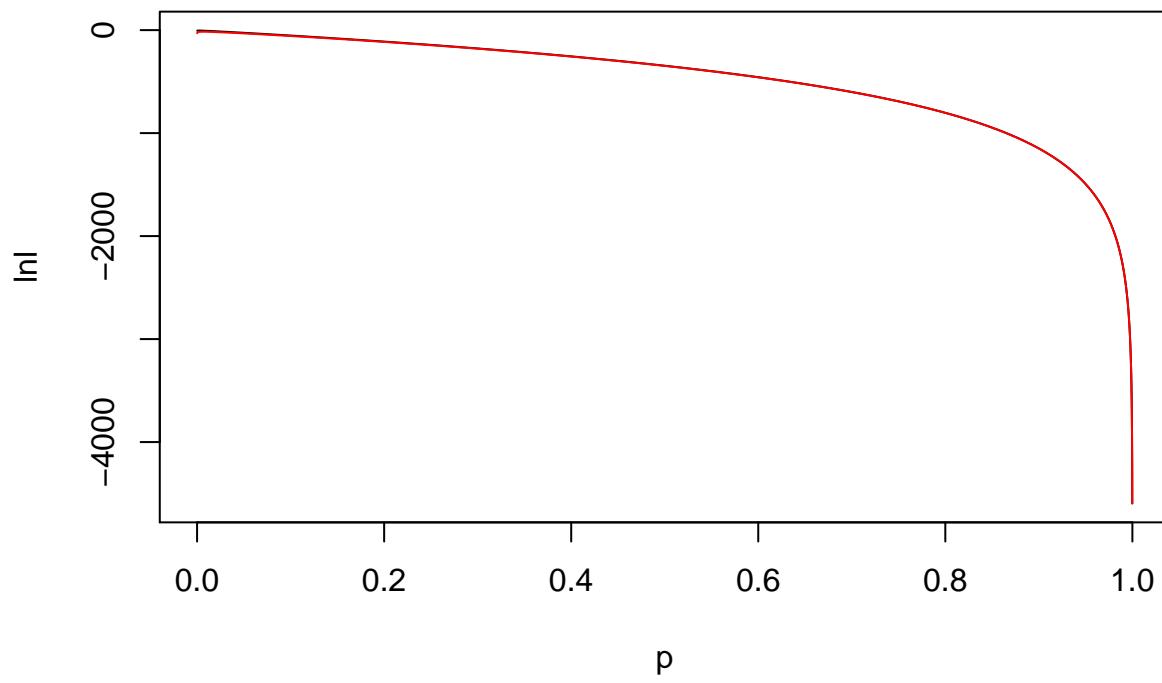
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0048"   "0.0154"   "208.33333333333" "64.9350649350649"
```

a = 0 , b = 20 , L = 300



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"   "1/MCLE"
## [2,] "0.0032" "0.0105" "312.5"   "95.2380952380952"
```

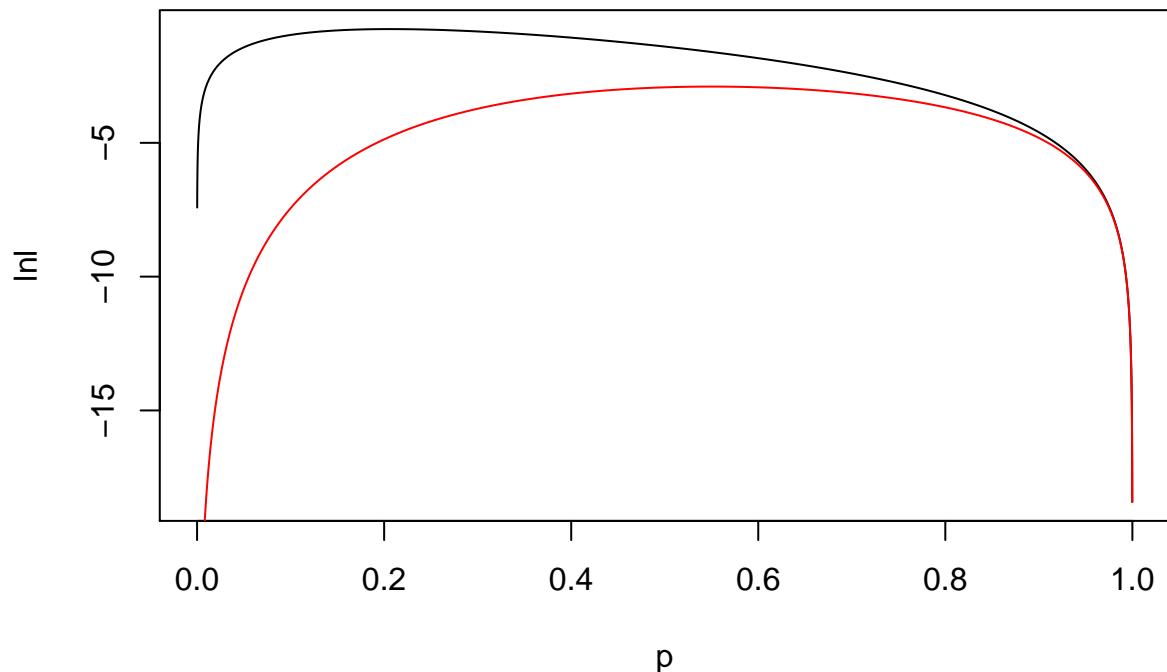
a = 0 , b = 20 , L = 500



```
##      [,1]      [,2]      [,3]      [,4]
```

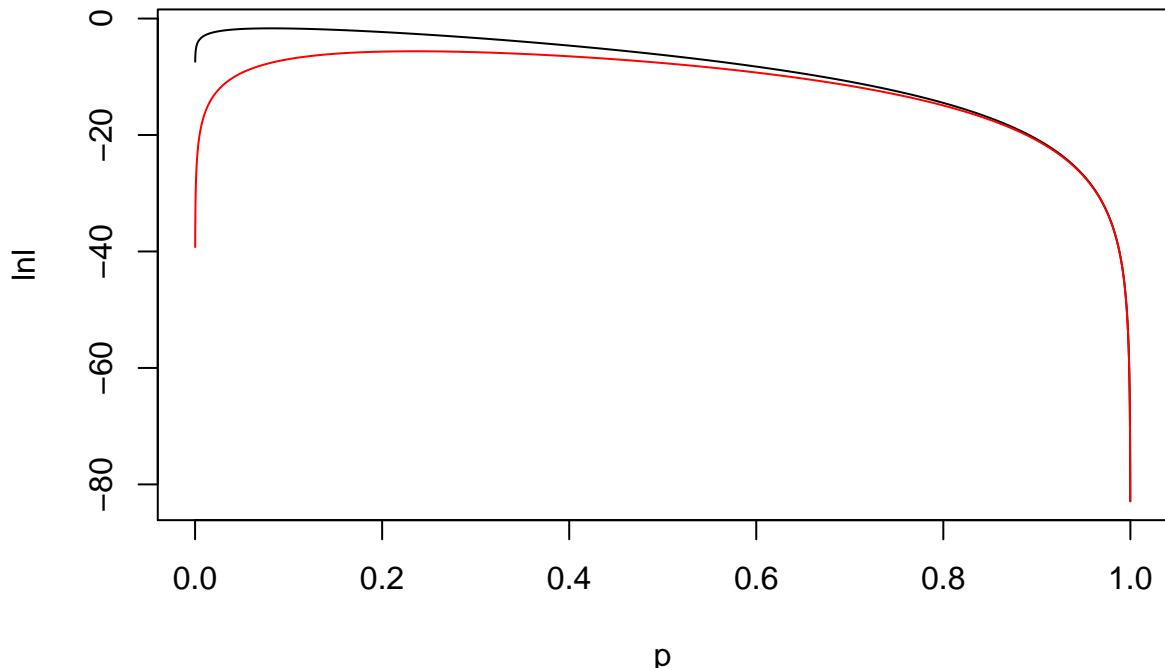
```
## [1,] "MLE"      "MCLE"      "1/MLE"     "1/MCLE"  
## [2,] "0.002"    "0.0064"    "500"       "156.25"
```

$$a = 5, b = 0, L = 3$$



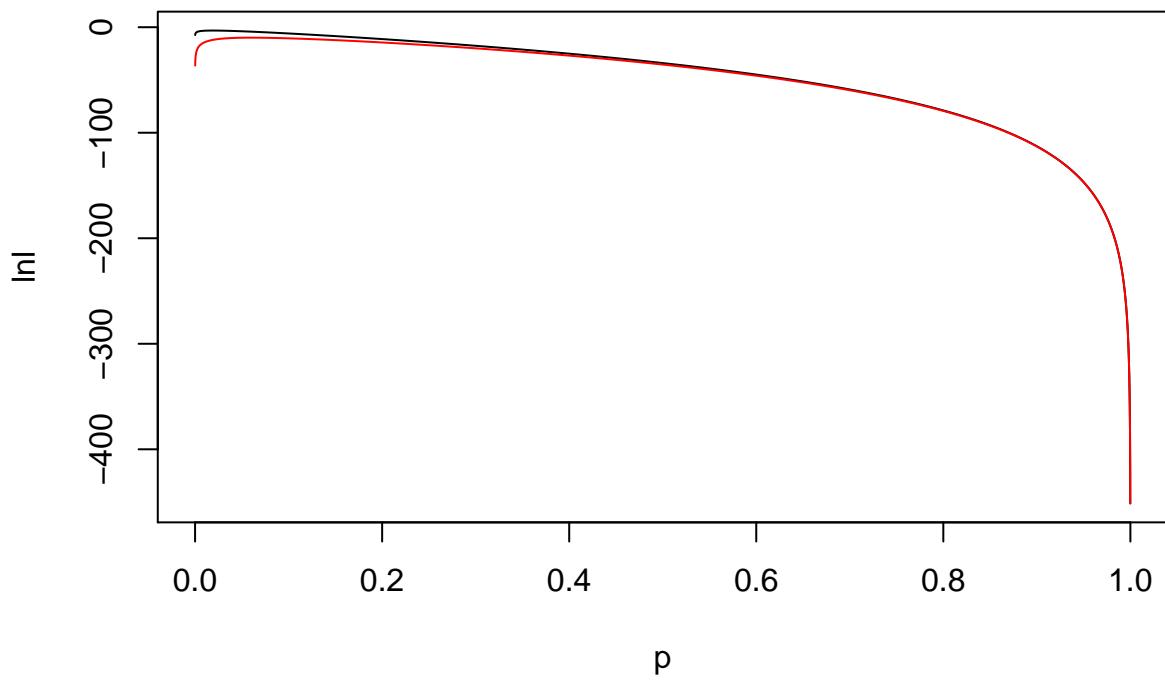
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"     "1/MCLE"  
## [2,] "0.2063"   "0.5501"   "4.84730974309258" "1.81785129976368"
```

a= 5 ,b= 0 ,L= 10



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"    "1/MLE"   "1/MCLE"
## [2,] "0.0816" "0.2354" "12.2549019607843" "4.24808836023789"
```

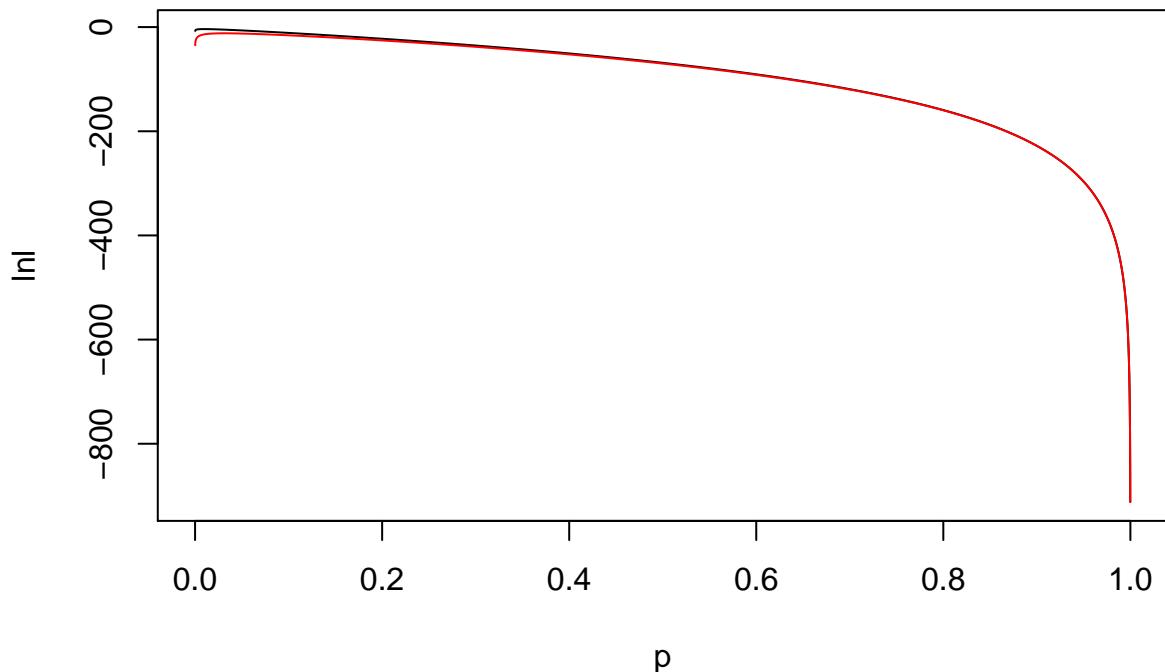
a= 5 ,b= 0 ,L= 50



```
##      [,1]      [,2]      [,3]      [,4]
```

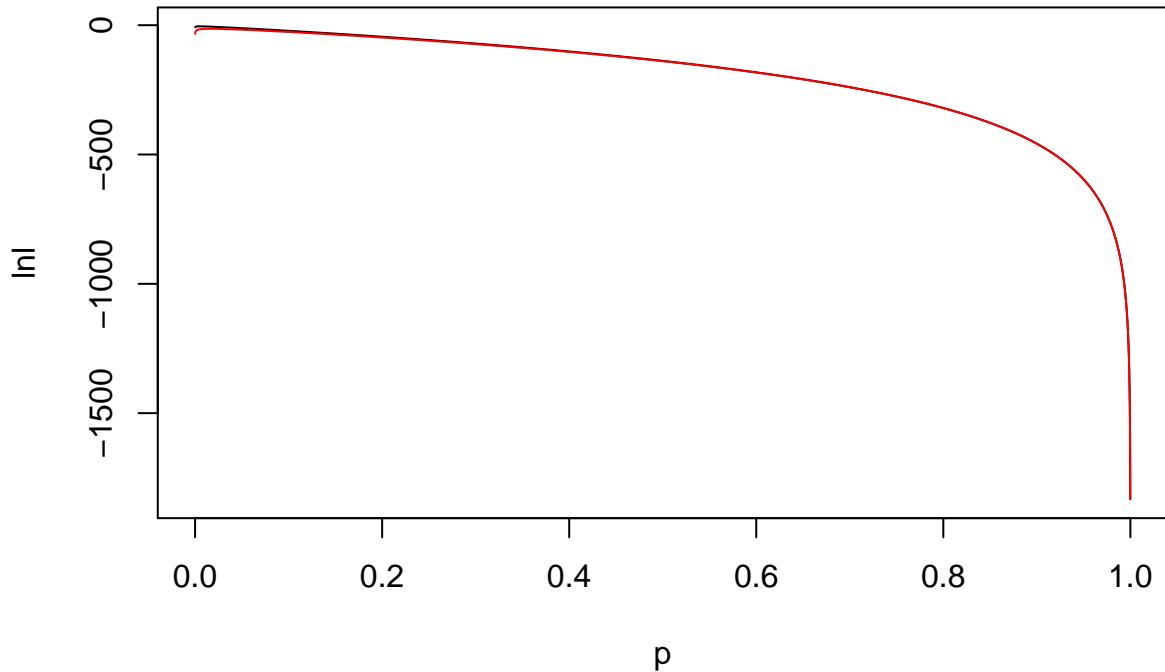
```
## [1,] "MLE"      "MCLE"     "1/MLE"          "1/MCLE"  
## [2,] "0.0191"   "0.06"    "52.3560209424084" "16.666666666667"
```

a= 5 ,b= 0 ,L= 100



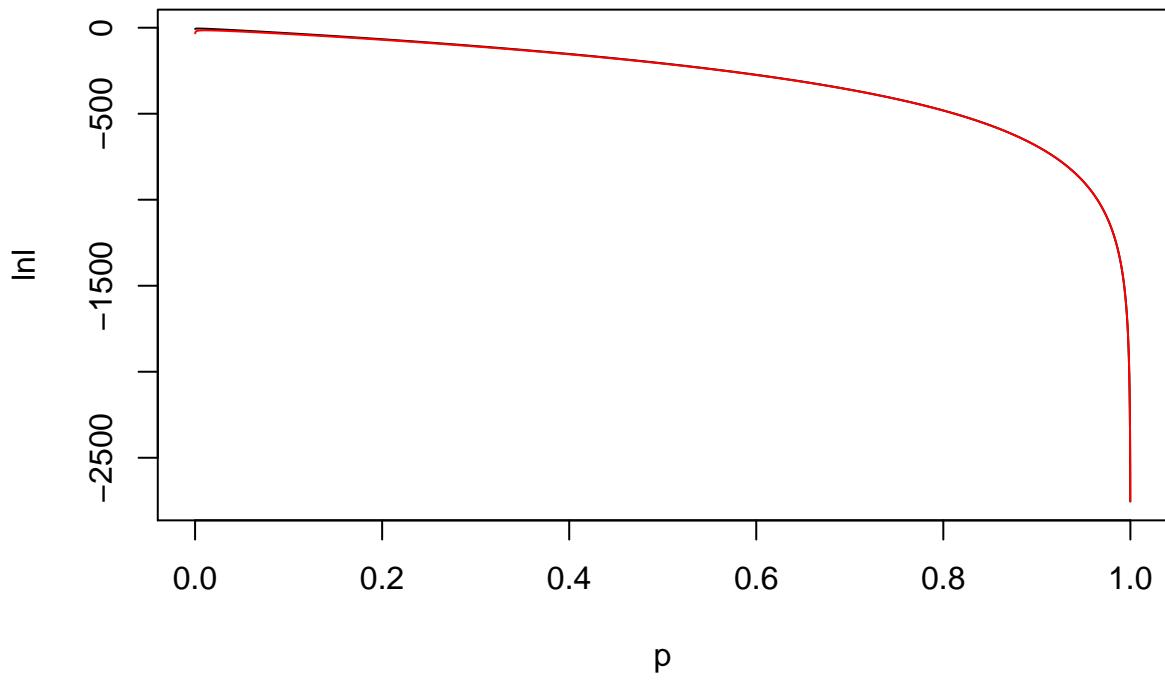
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"     "MCLE"     "1/MLE"     "1/MCLE"  
## [2,] "0.0098"  "0.0312"  "102.040816326531" "32.0512820512821"
```

a= 5 ,b= 0 ,L= 200



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.0049" "0.0159" "204.081632653061" "62.8930817610063"
```

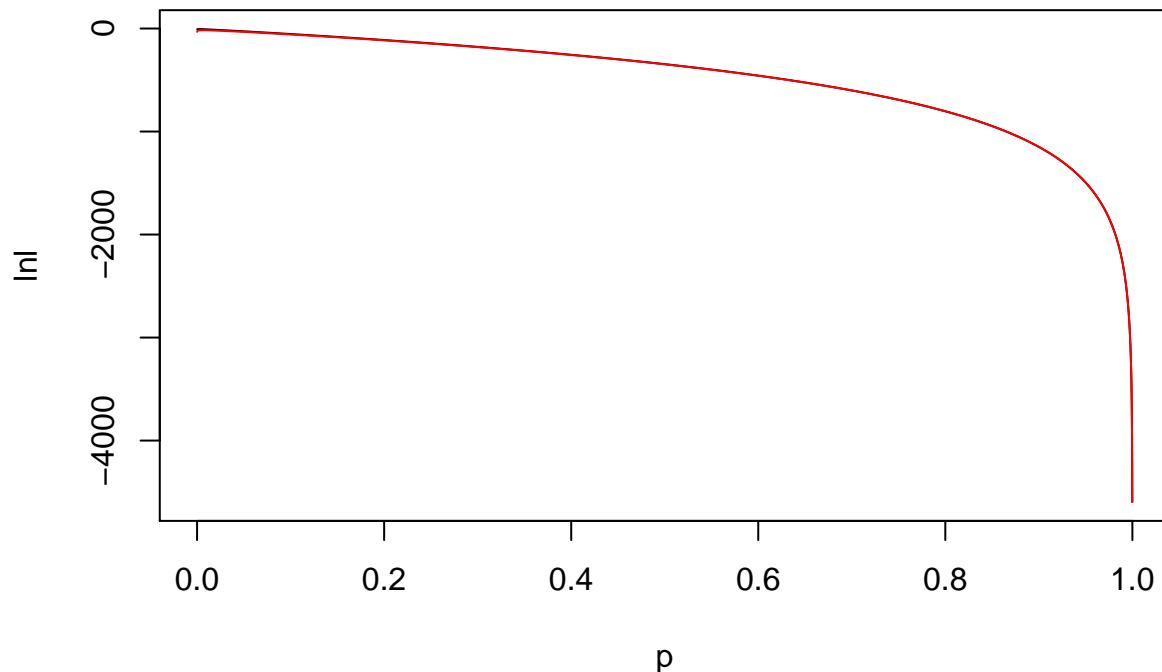
a= 5 ,b= 0 ,L= 300



```
##      [,1]      [,2]      [,3]      [,4]
```

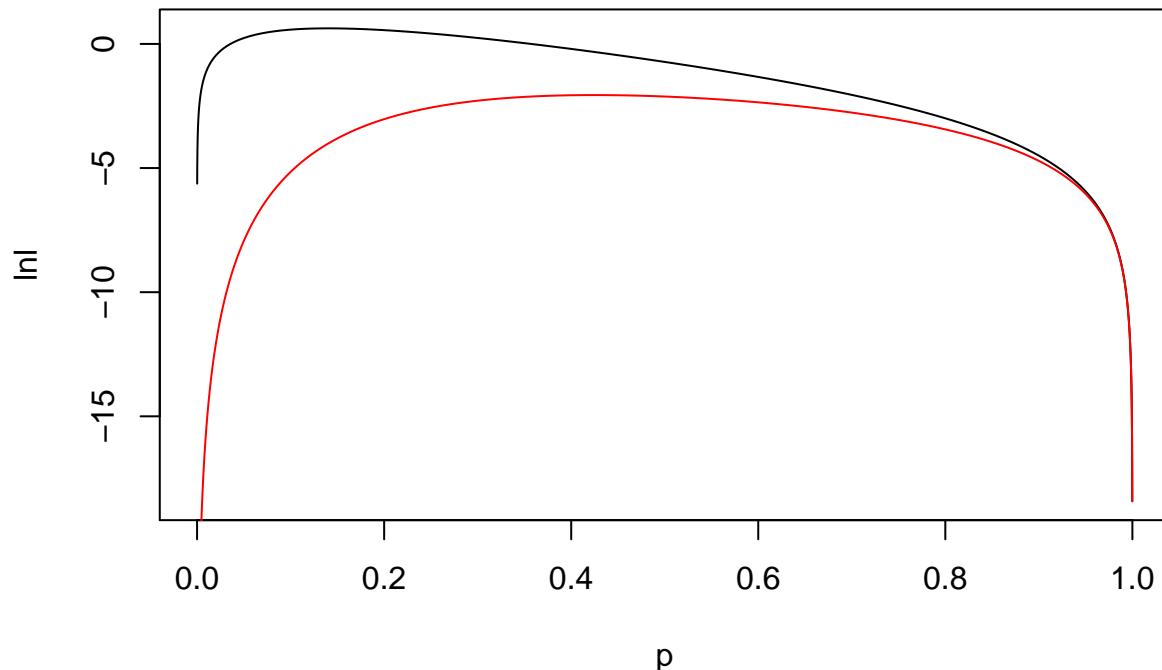
```
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0033"   "0.0107"   "303.030303030303" "93.4579439252336"
```

a= 5 ,b= 0 ,L= 500



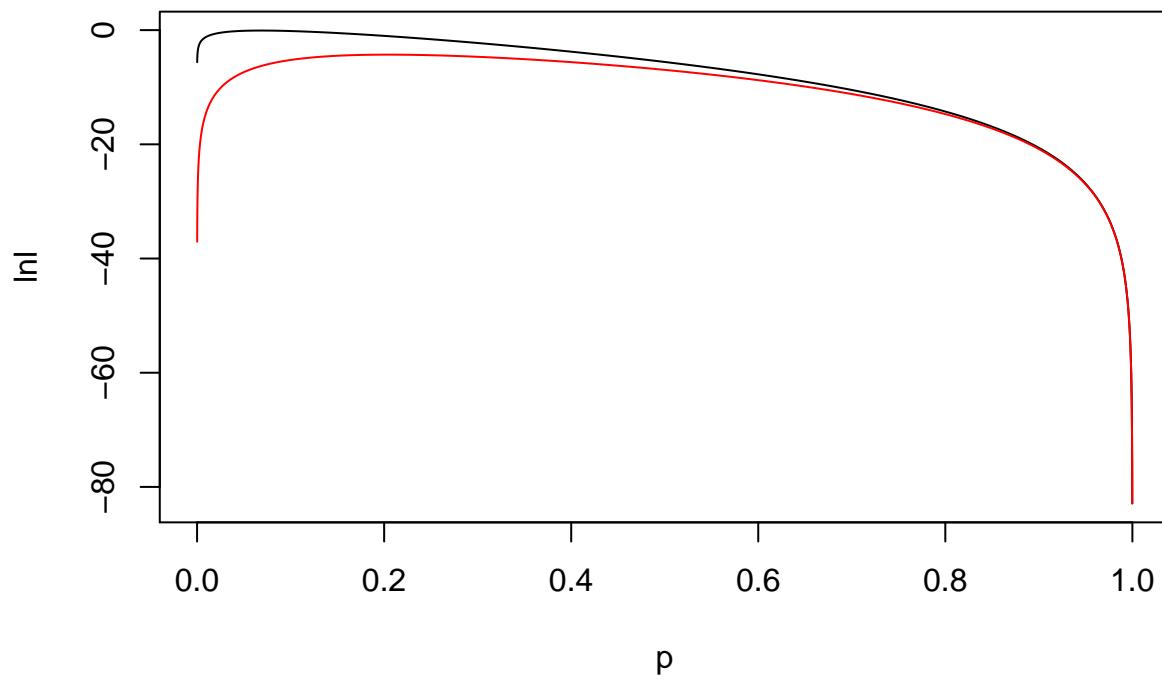
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"     "1/MCLE"  
## [2,] "0.002"    "0.0065"   "500"       "153.846153846154"
```

a= 5 ,b= 5 ,L= 3



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.1402" "0.4247" "7.13266761768902" "2.35460324935248"
```

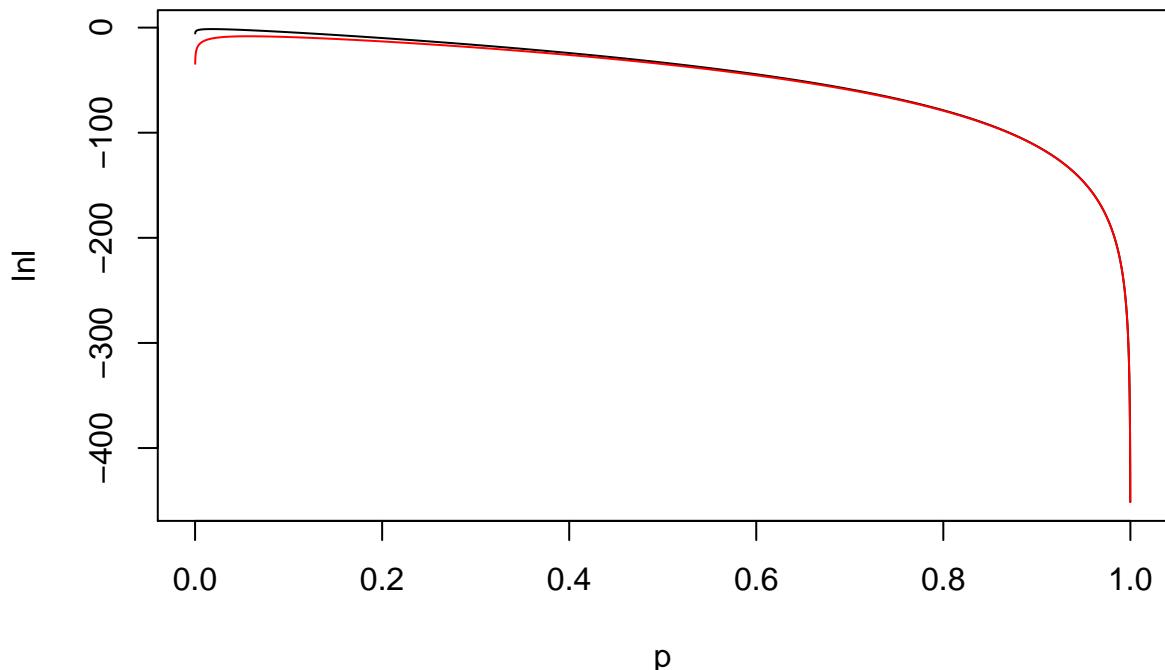
a= 5 ,b= 5 ,L= 10



```
##      [,1]      [,2]      [,3]      [,4]
```

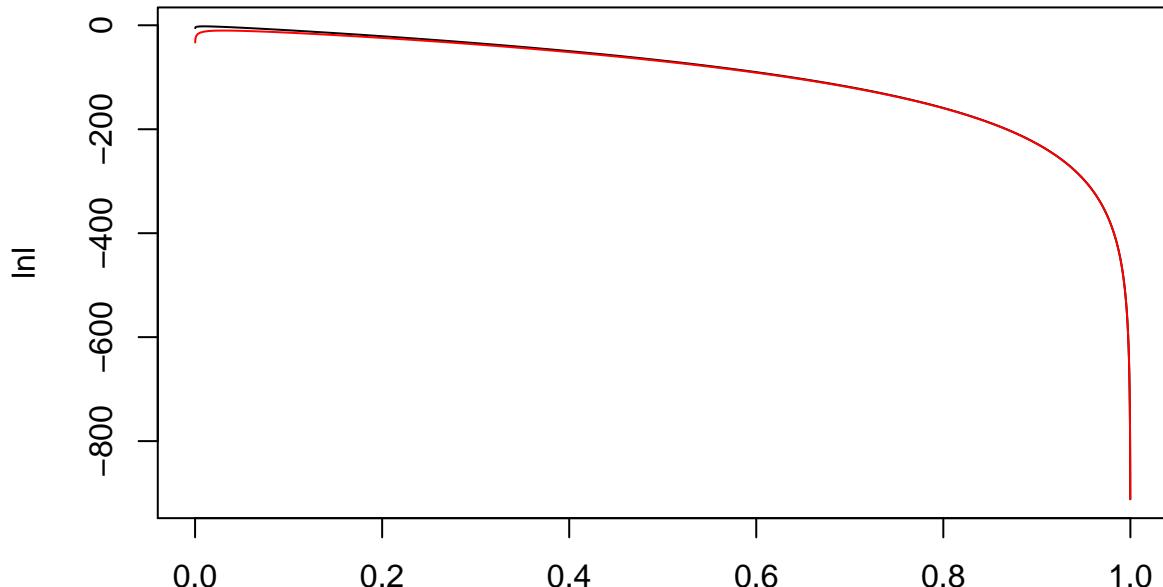
```
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0686"   "0.2041"   "14.5772594752187" "4.89955903968643"
```

a= 5 ,b= 5 ,L= 50



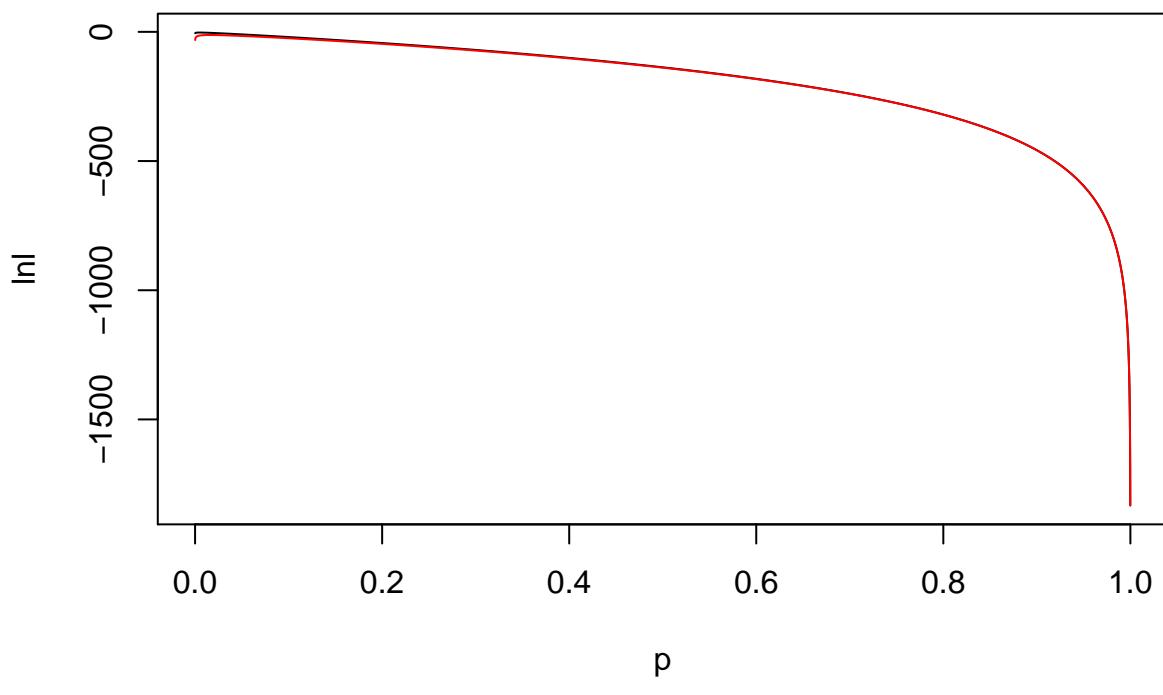
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0182"   "0.0574"   "54.9450549450549" "17.4216027874564"
```

a= 5 ,b= 5 ,L= 100



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.0095" "0.0305" "105.263157894737" "32.7868852459016"
```

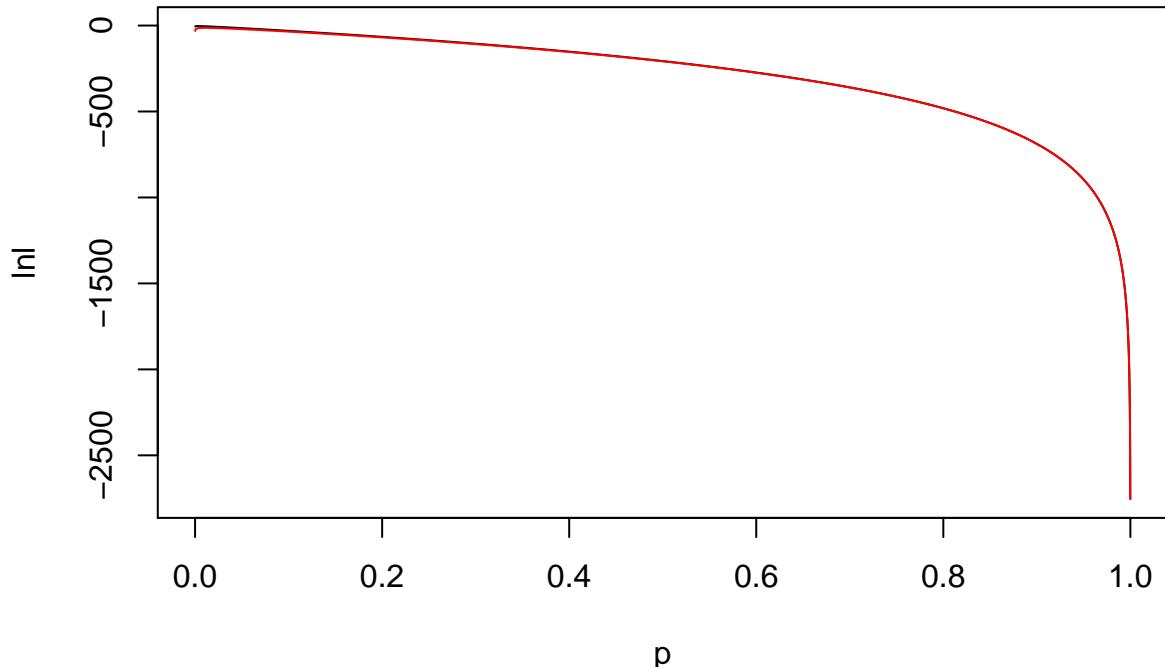
a= 5 ,b= 5 ,L= 200



```
##      [,1]      [,2]      [,3]      [,4]
```

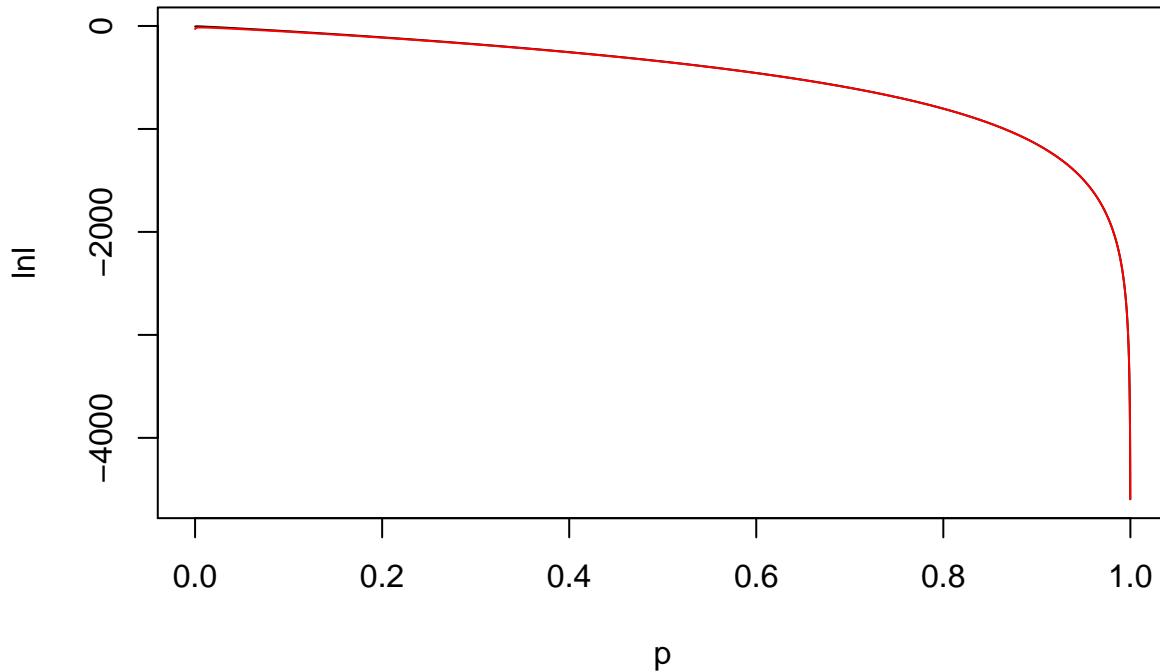
```
## [1,] "MLE"      "MCLE"      "1/MLE"          "1/MCLE"  
## [2,] "0.0049"   "0.0158"   "204.081632653061" "63.2911392405063"
```

a= 5 ,b= 5 ,L= 300



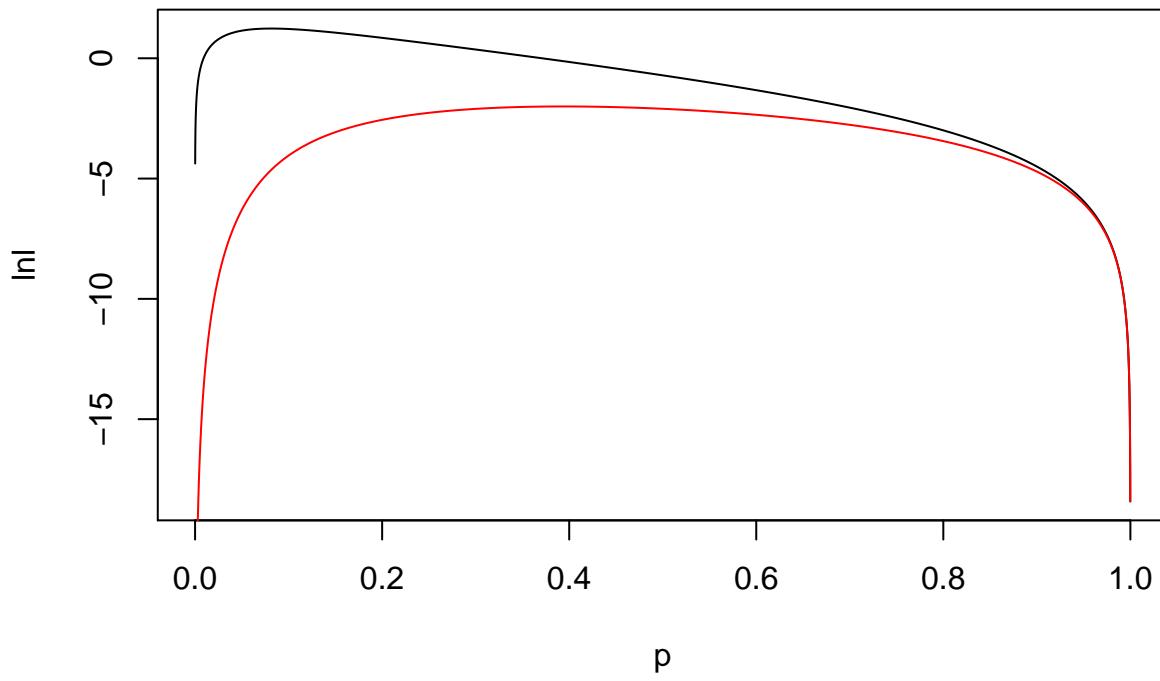
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0033"   "0.0106"   "303.0303030303" "94.3396226415094"
```

a= 5 ,b= 5 ,L= 500



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.002"  "0.0064" "500"    "156.25"
```

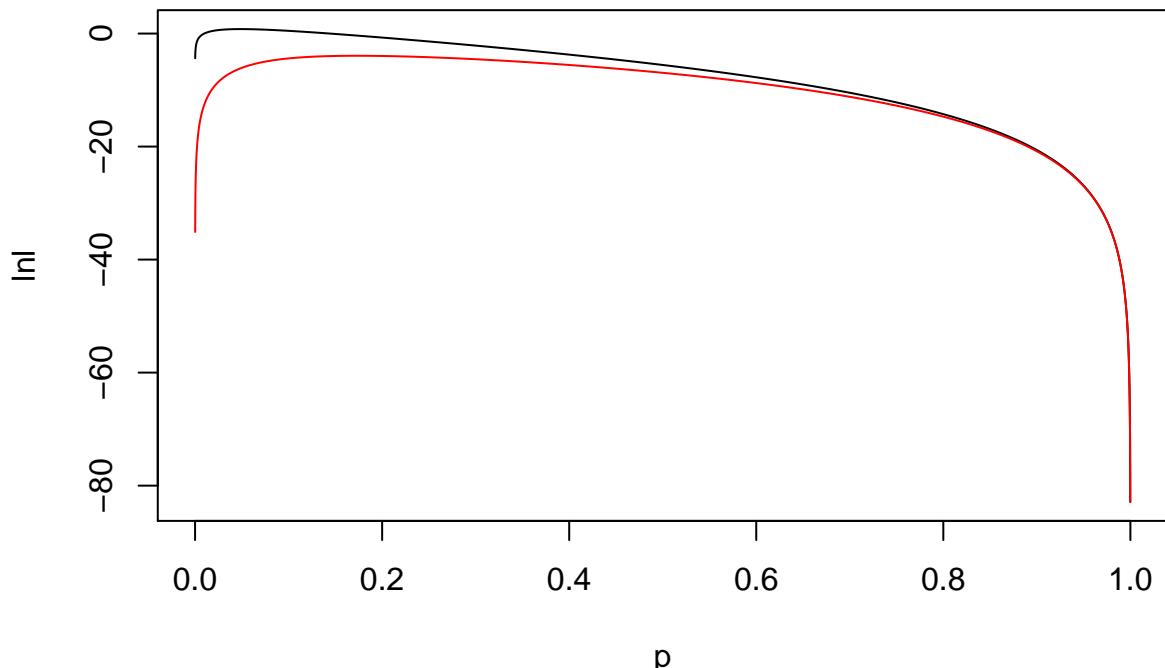
a= 5 ,b= 20 ,L= 3



```
##      [,1]      [,2]      [,3]      [,4]
```

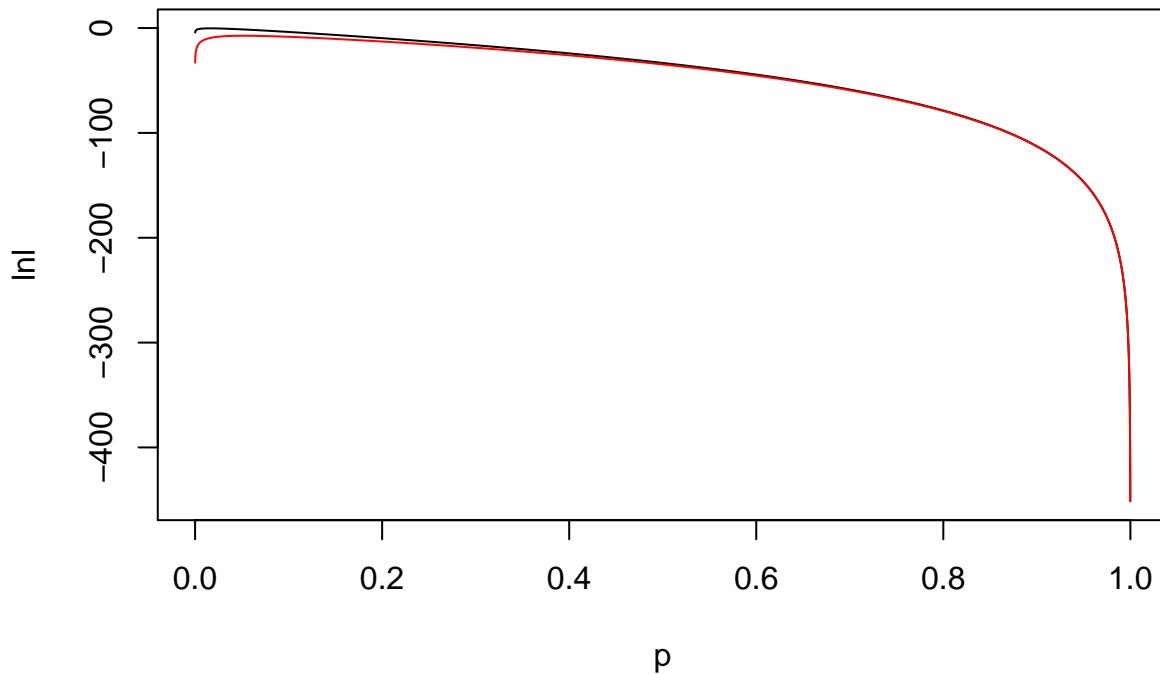
```
## [1,] "MLE"      "MCLE"     "1/MLE"          "1/MCLE"  
## [2,] "0.0813"   "0.394"    "12.30012300123" "2.53807106598985"
```

a= 5 ,b= 20 ,L= 10



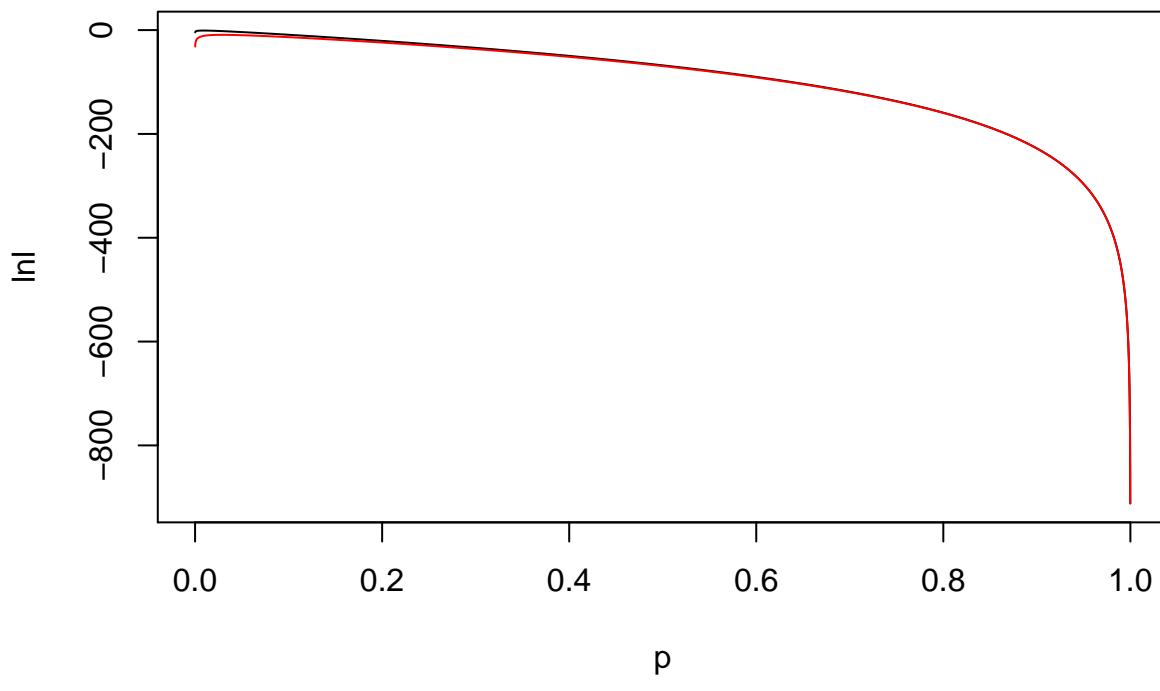
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"     "1/MLE"          "1/MCLE"  
## [2,] "0.0486"   "0.1732"   "20.5761316872428" "5.77367205542725"
```

a= 5 ,b= 20 ,L= 50



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"    "1/MLE"   "1/MCLE"
## [2,] "0.0162" "0.052"   "61.7283950617284" "19.2307692307692"
```

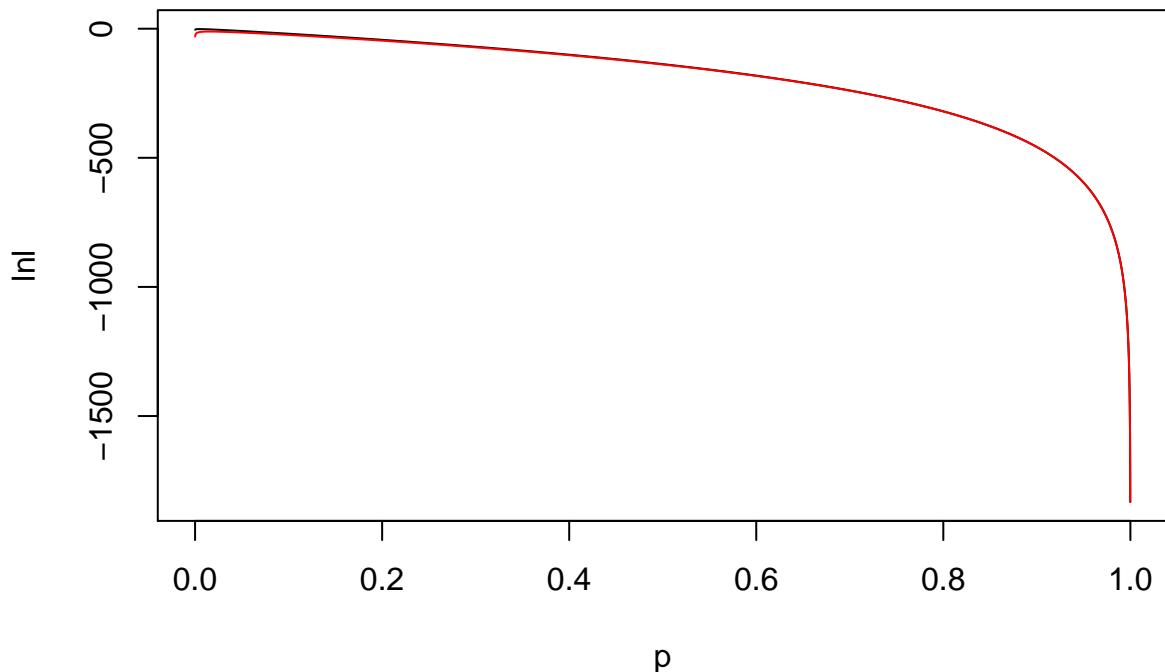
a= 5 ,b= 20 ,L= 100



```
##      [,1]      [,2]      [,3]      [,4]
```

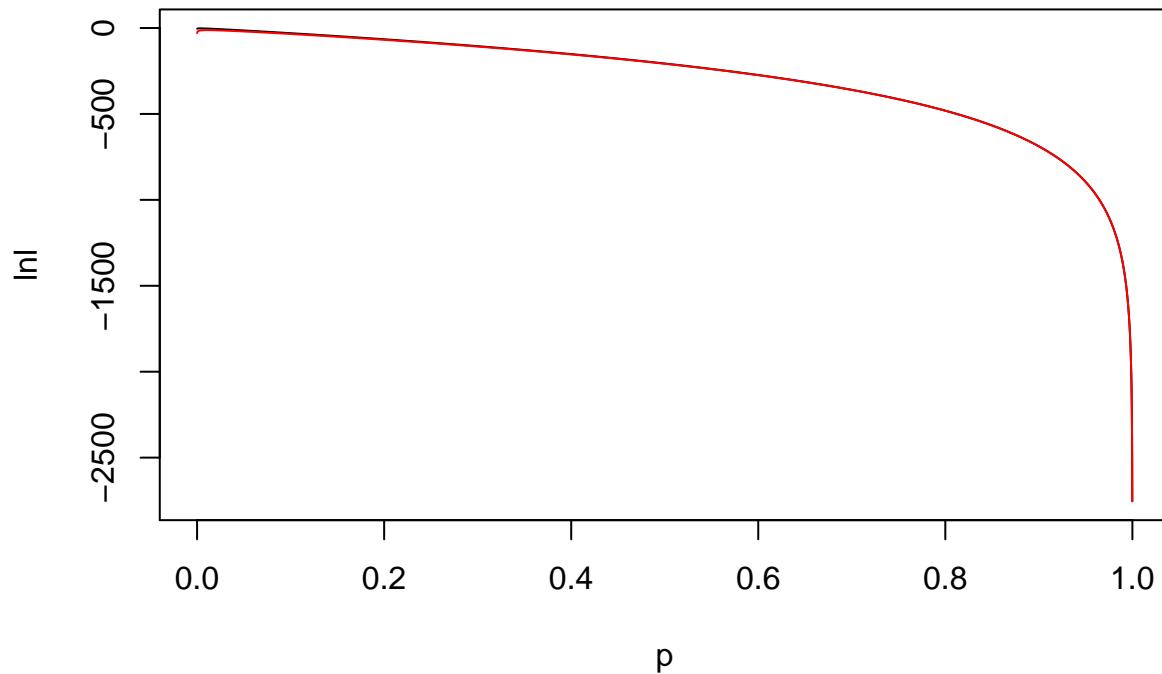
```
## [1,] "MLE"      "MCLE"      "1/MLE"          "1/MCLE"  
## [2,] "0.0089"   "0.0287"   "112.359550561798" "34.8432055749129"
```

a= 5 ,b= 20 ,L= 200



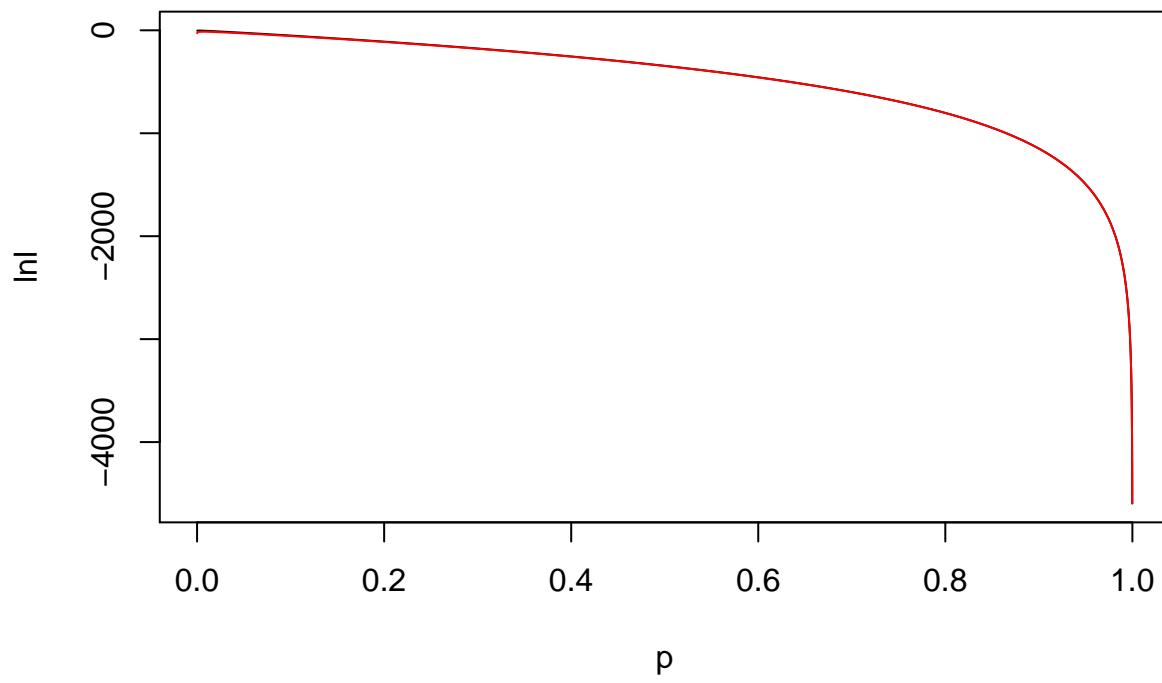
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0047"   "0.0152"   "212.765957446808" "65.7894736842105"
```

a = 5 , b = 20 , L = 300



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.0032" "0.0104" "312.5"  "96.1538461538462"
```

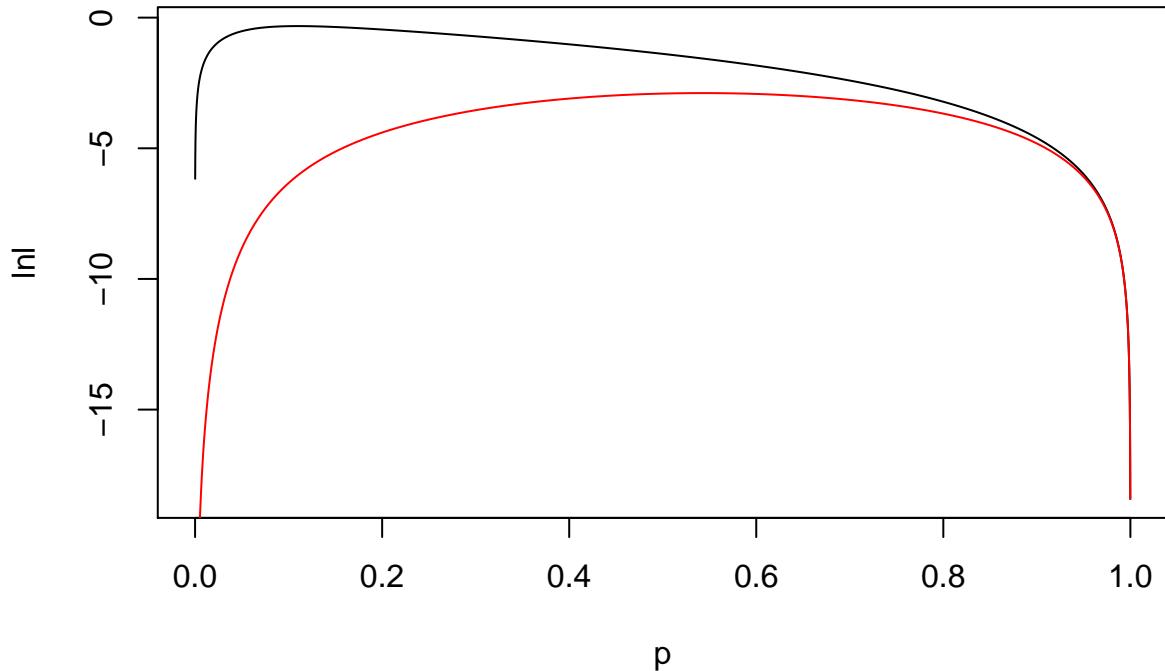
a = 5 , b = 20 , L = 500



```
##      [,1]      [,2]      [,3]      [,4]
```

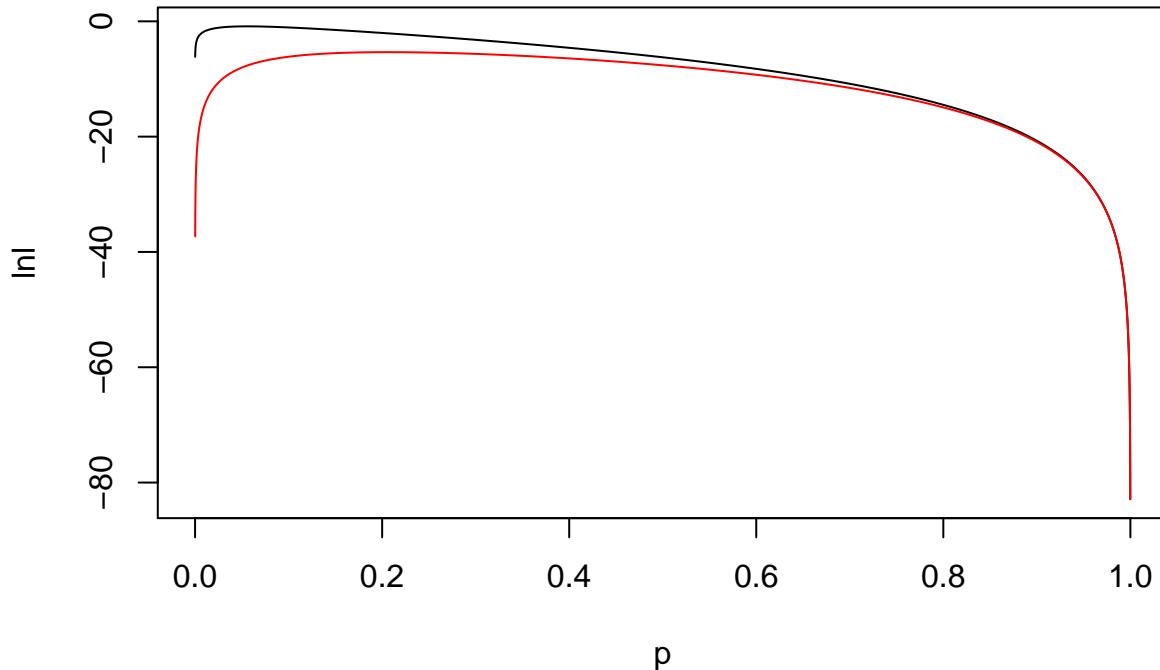
```
## [1,] "MLE"      "MCLE"      "1/MLE"     "1/MCLE"  
## [2,] "0.002"    "0.0063"    "500"       "158.730158730159"
```

a= 20 ,b= 0 ,L= 3



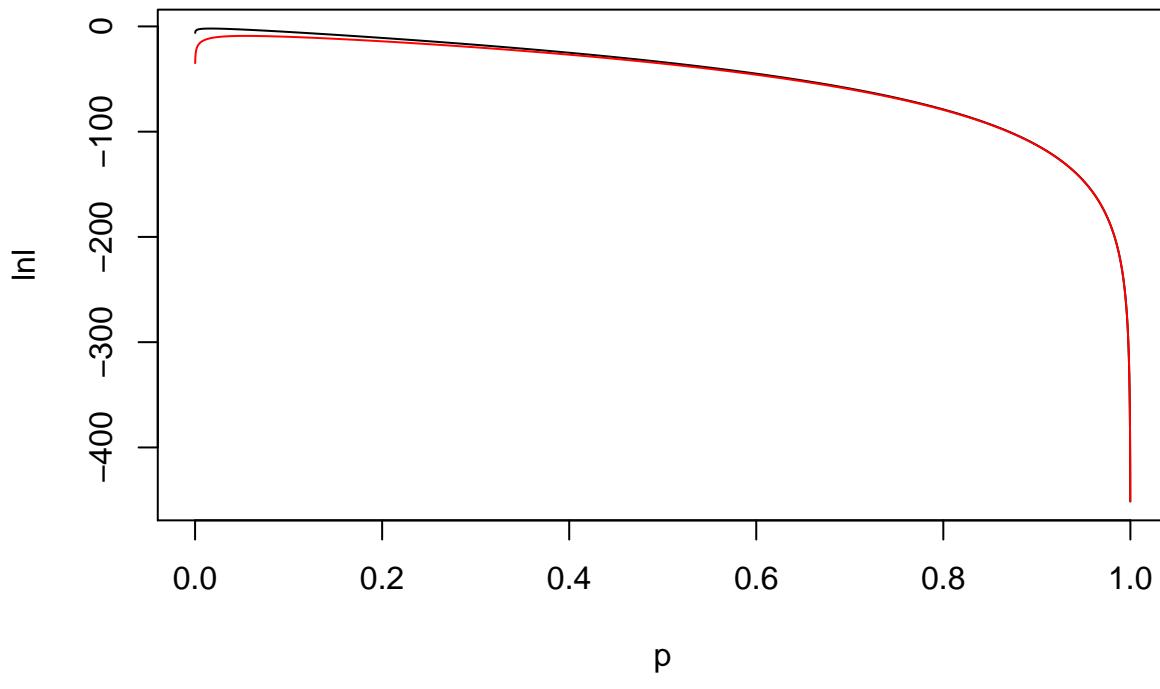
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"     "1/MCLE"  
## [2,] "0.1098"   "0.543"    "9.10746812386157" "1.84162062615101"
```

a= 20 ,b= 0 ,L= 10



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.0557" "0.2062" "17.9533213644524" "4.84966052376334"
```

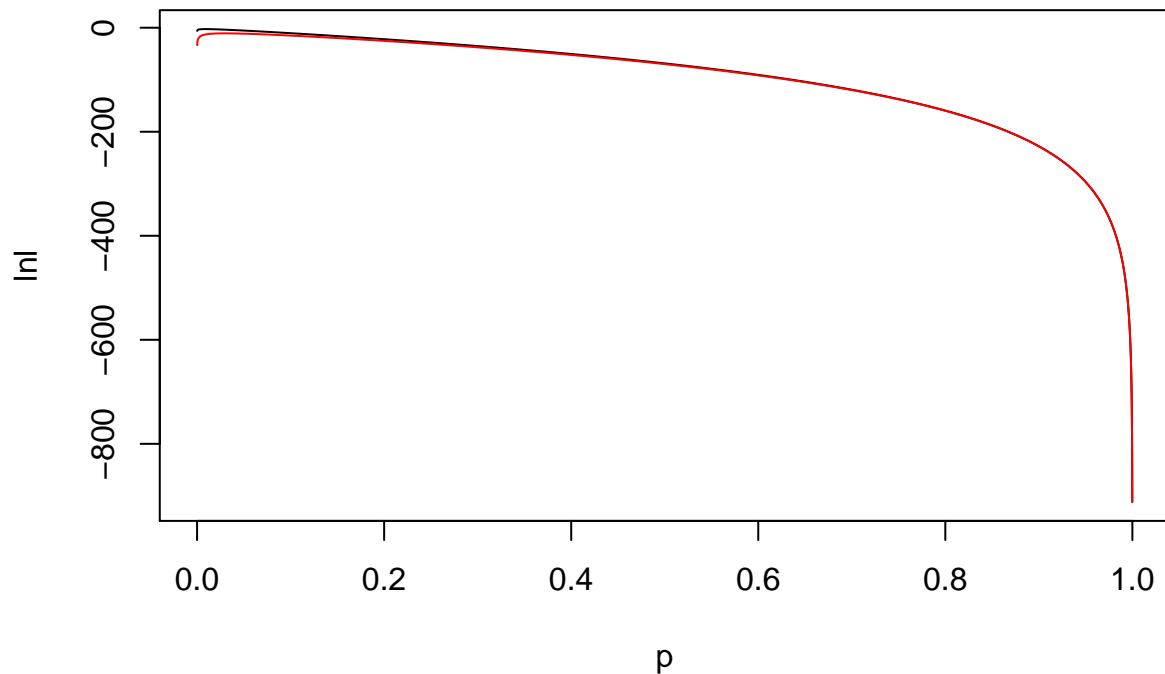
a= 20 ,b= 0 ,L= 50



```
##      [,1]      [,2]      [,3]      [,4]
```

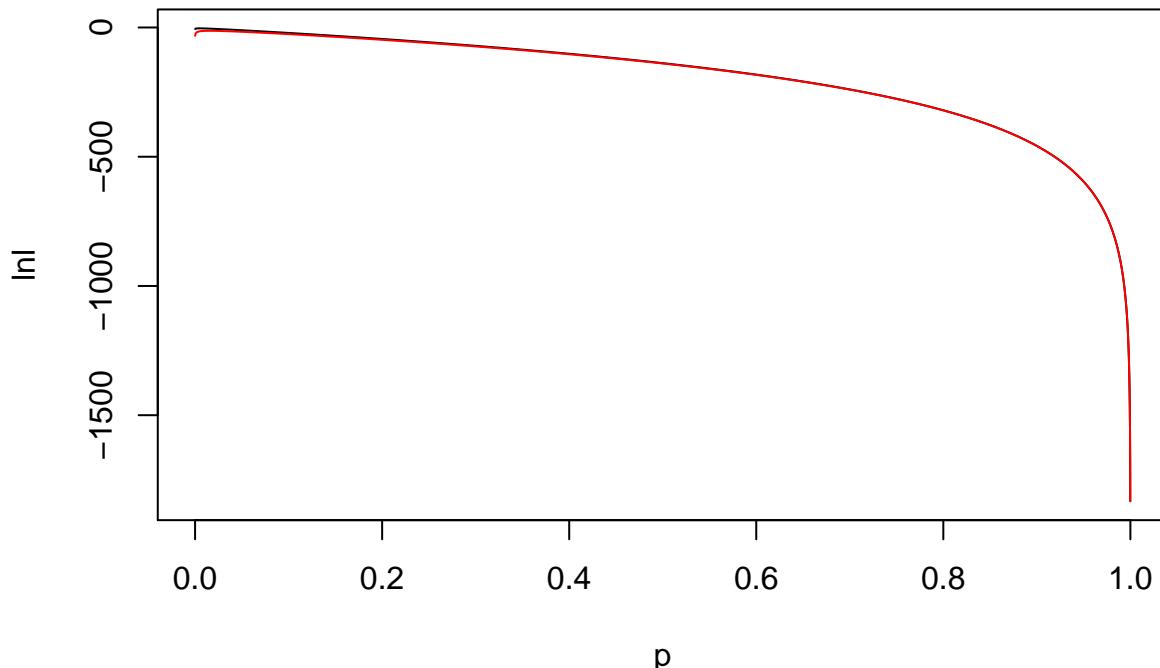
```
## [1,] "MLE"      "MCLE"      "1/MLE"          "1/MCLE"  
## [2,] "0.0168"   "0.0543"   "59.5238095238095" "18.4162062615101"
```

a= 20 ,b= 0 ,L= 100



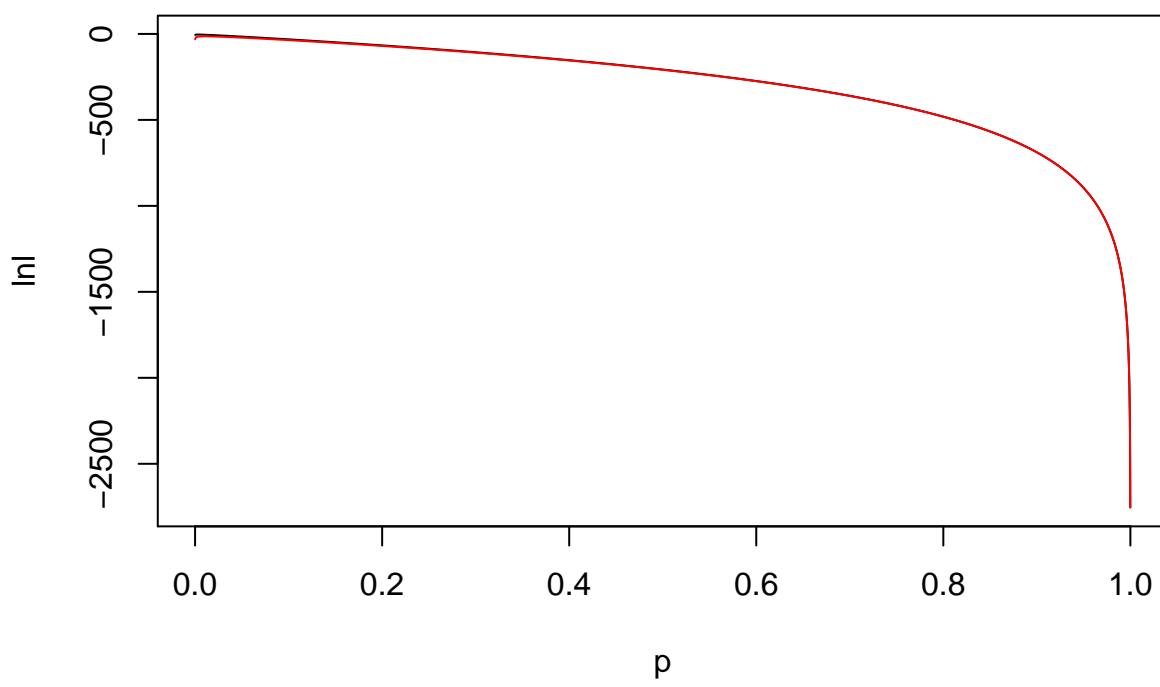
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0091"   "0.0294"   "109.89010989011" "34.0136054421769"
```

a = 20 ,b = 0 ,L = 200



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.0048" "0.0154" "208.33333333333" "64.9350649350649"
```

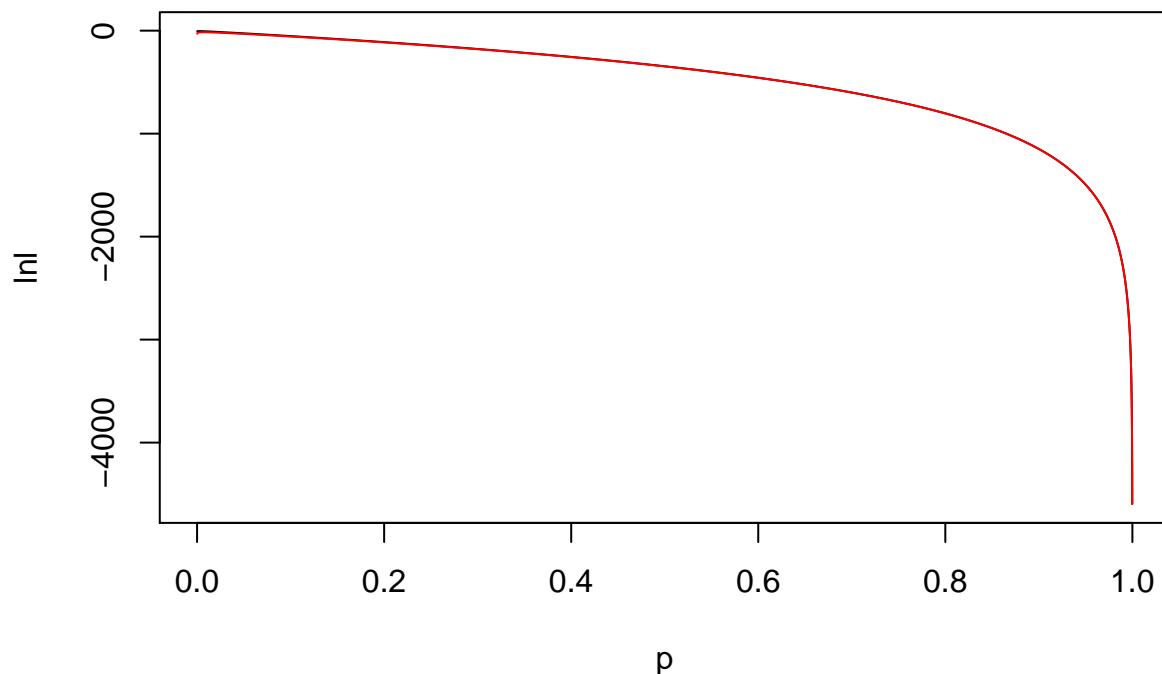
a = 20 ,b = 0 ,L = 300



```
##      [,1]      [,2]      [,3]      [,4]
```

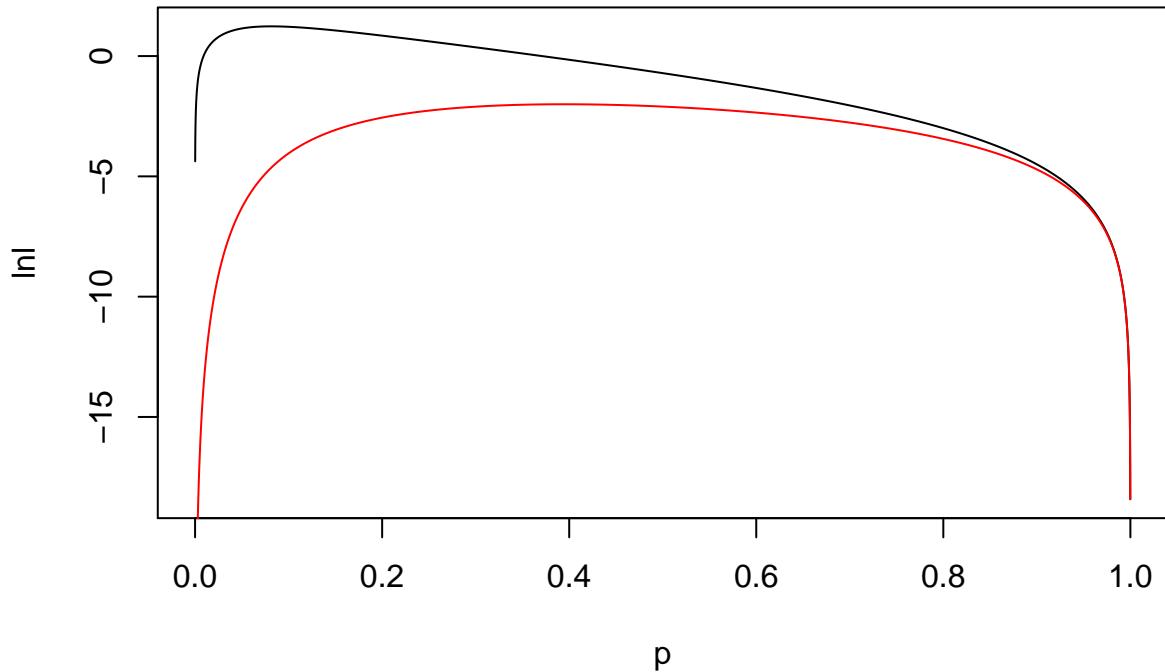
```
## [1,] "MLE"      "MCLE"      "1/MLE"     "1/MCLE"  
## [2,] "0.0032"   "0.0105"   "312.5"    "95.2380952380952"
```

a= 20 ,b= 0 ,L= 500



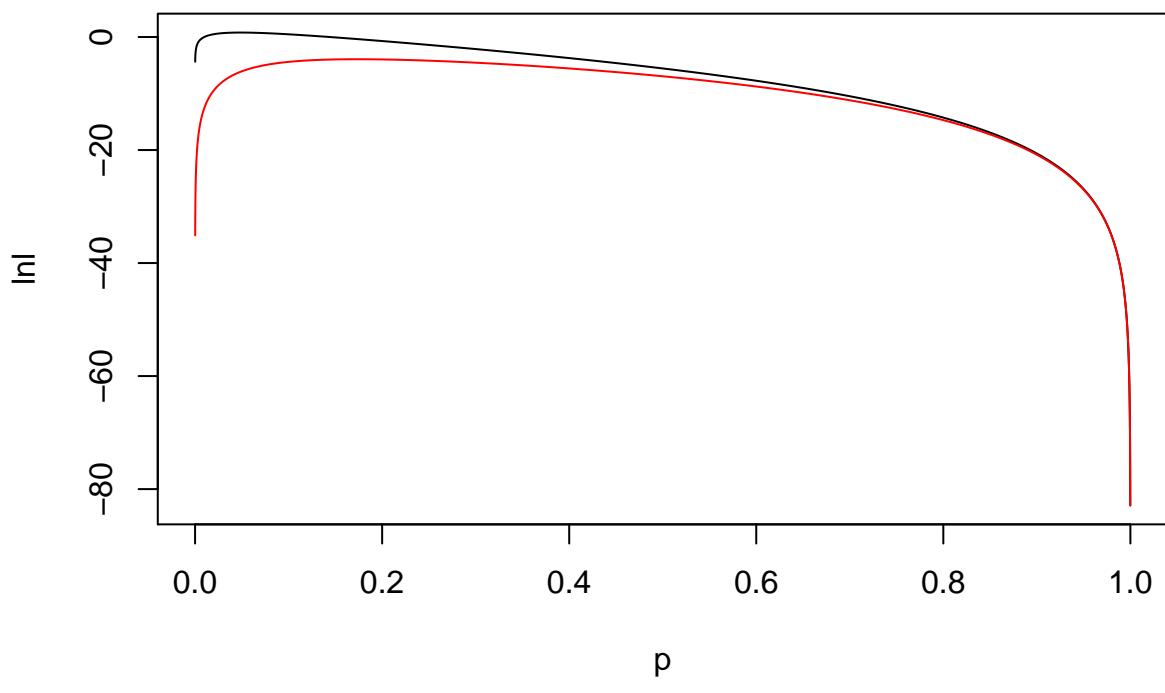
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"     "1/MCLE"  
## [2,] "0.002"    "0.0064"   "500"       "156.25"
```

a= 20 ,b= 5 ,L= 3



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"   "1/MCLE"
## [2,] "0.0813" "0.394"  "12.30012300123" "2.53807106598985"
```

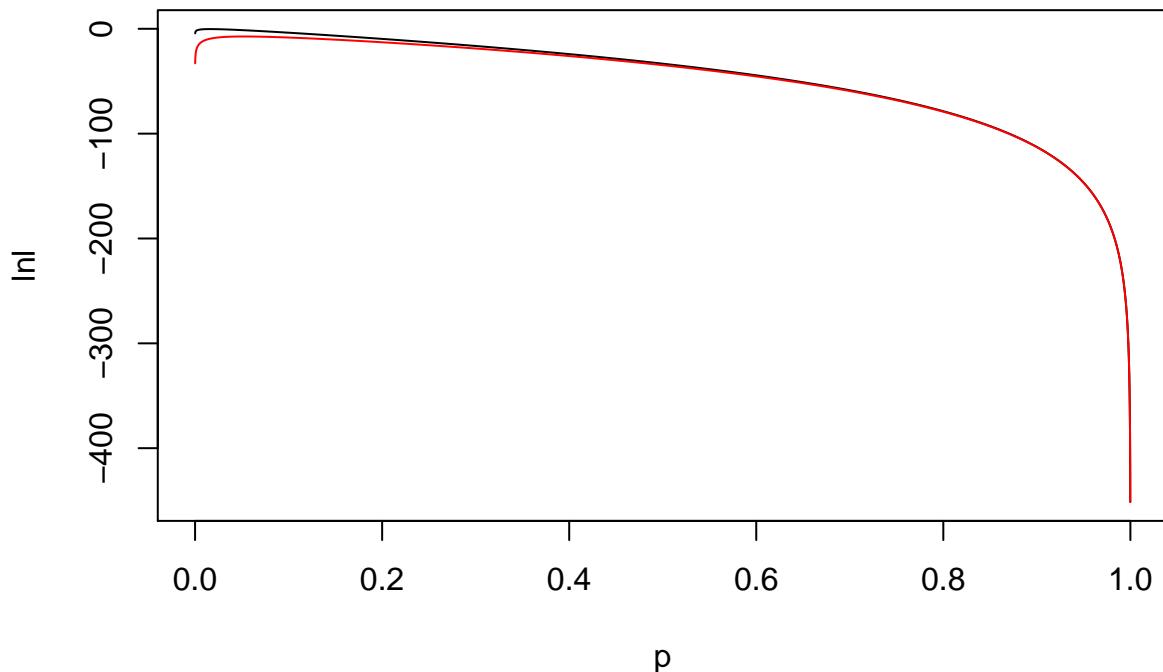
a = 20 ,b = 5 ,L = 10



```
##      [,1]      [,2]      [,3]      [,4]
```

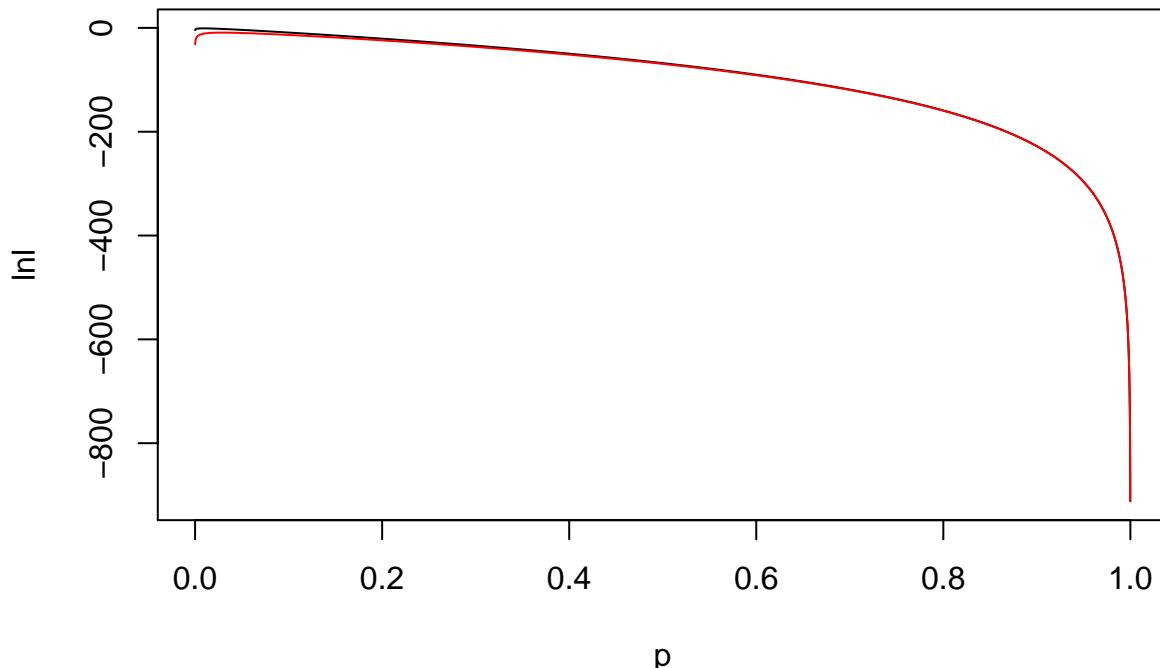
```
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0486"   "0.1732"   "20.5761316872428" "5.77367205542725"
```

a= 20 ,b= 5 ,L= 50



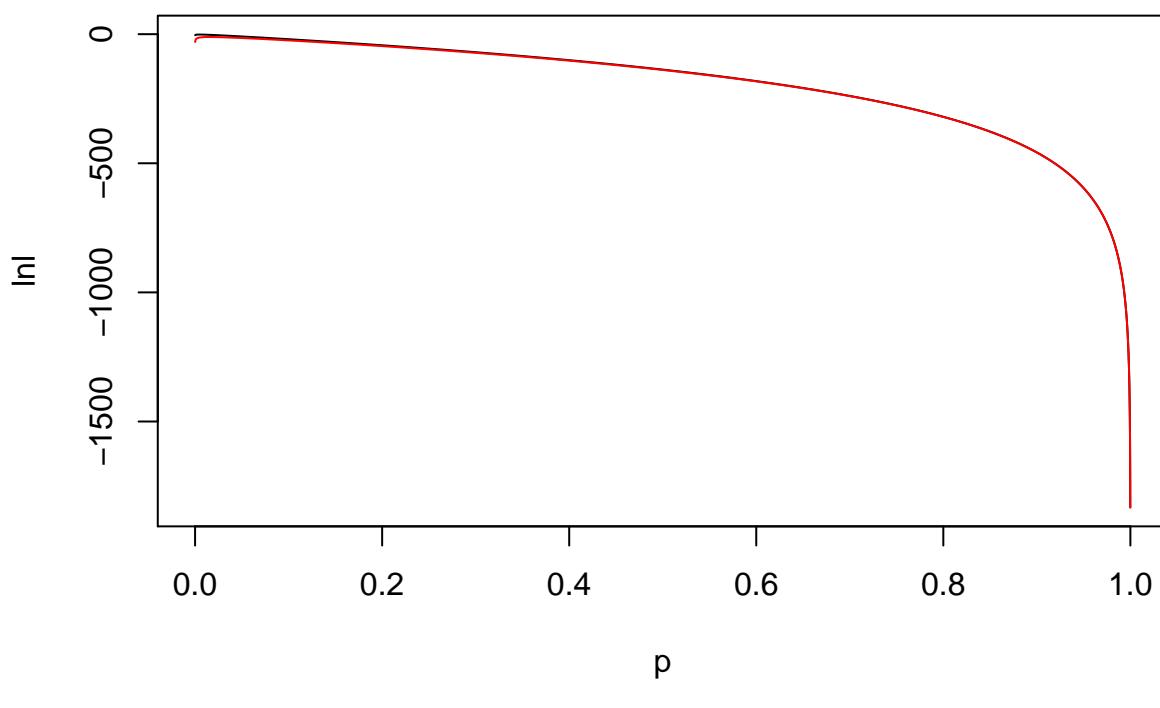
```
##      [,1]     [,2]     [,3]      [,4]  
## [1,] "MLE"    "MCLE"   "1/MLE"   "1/MCLE"  
## [2,] "0.0162" "0.052"  "61.7283950617284" "19.2307692307692"
```

a = 20 , b = 5 , L = 100



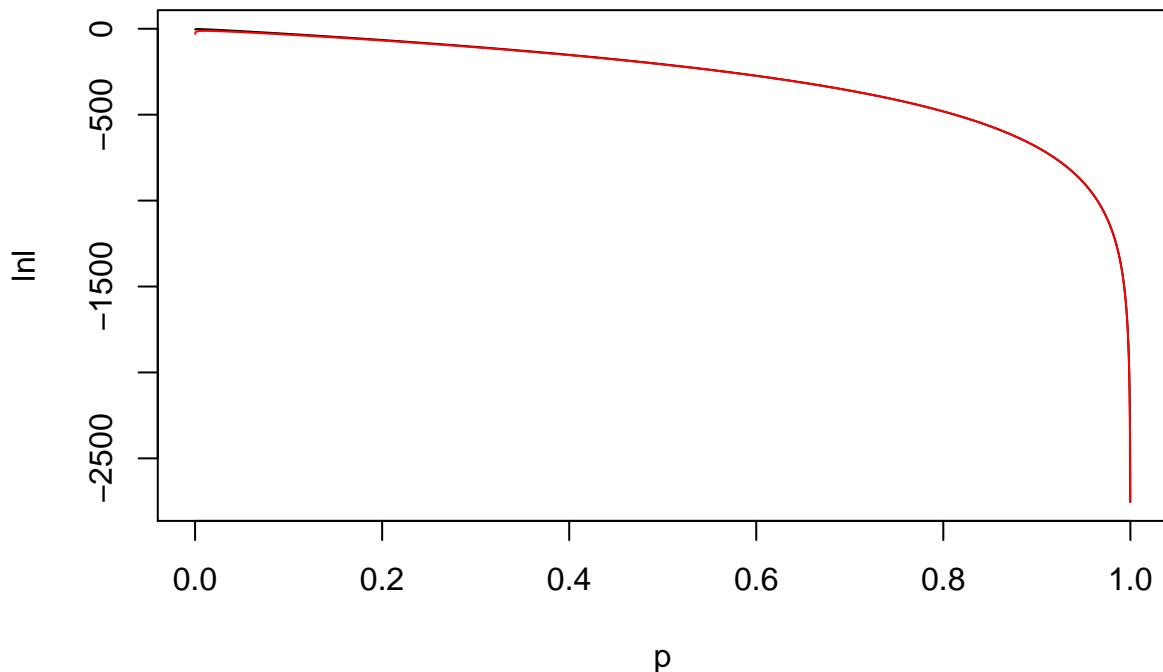
a = 20 , b = 5 , L = 200

```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"    "1/MLE"   "1/MCLE"
## [2,] "0.0089" "0.0287" "112.359550561798" "34.8432055749129"
```



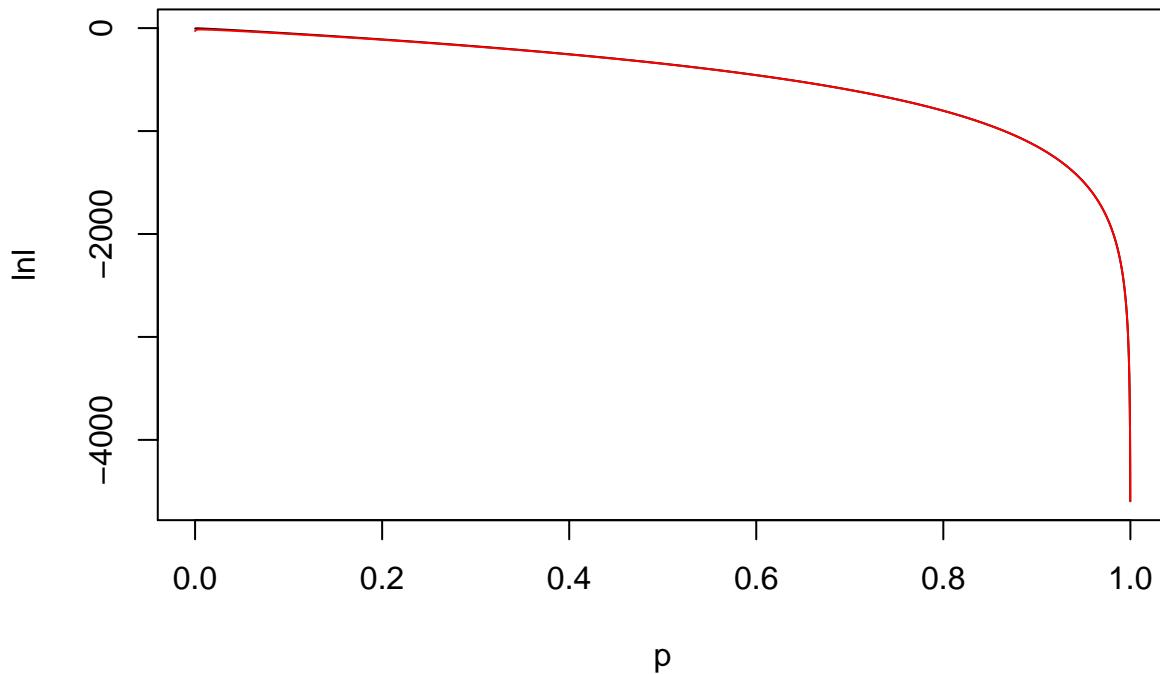
```
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0047"   "0.0152"   "212.765957446808" "65.7894736842105"
```

a= 20 ,b= 5 ,L= 300



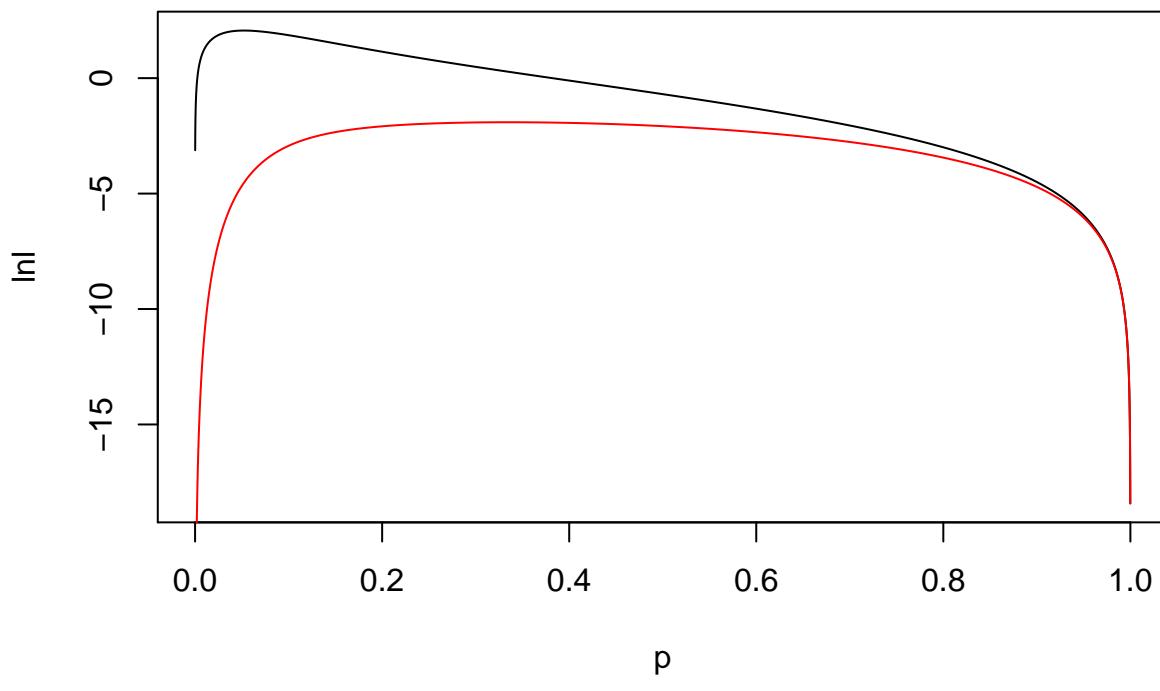
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0032"   "0.0104"   "312.5"     "96.1538461538462"
```

a= 20 ,b= 5 ,L= 500



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"    "1/MLE"   "1/MCLE"
## [2,] "0.002"  "0.0063" "500"     "158.730158730159"
```

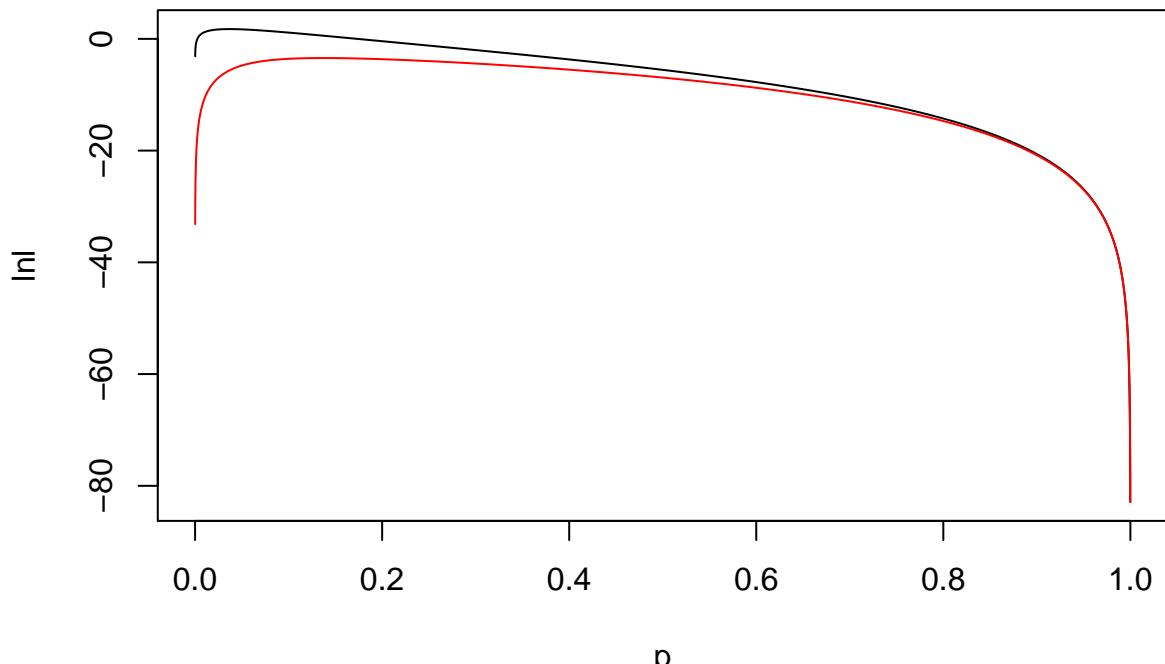
a= 20 ,b= 20 ,L= 3



```
##      [,1]      [,2]      [,3]      [,4]
```

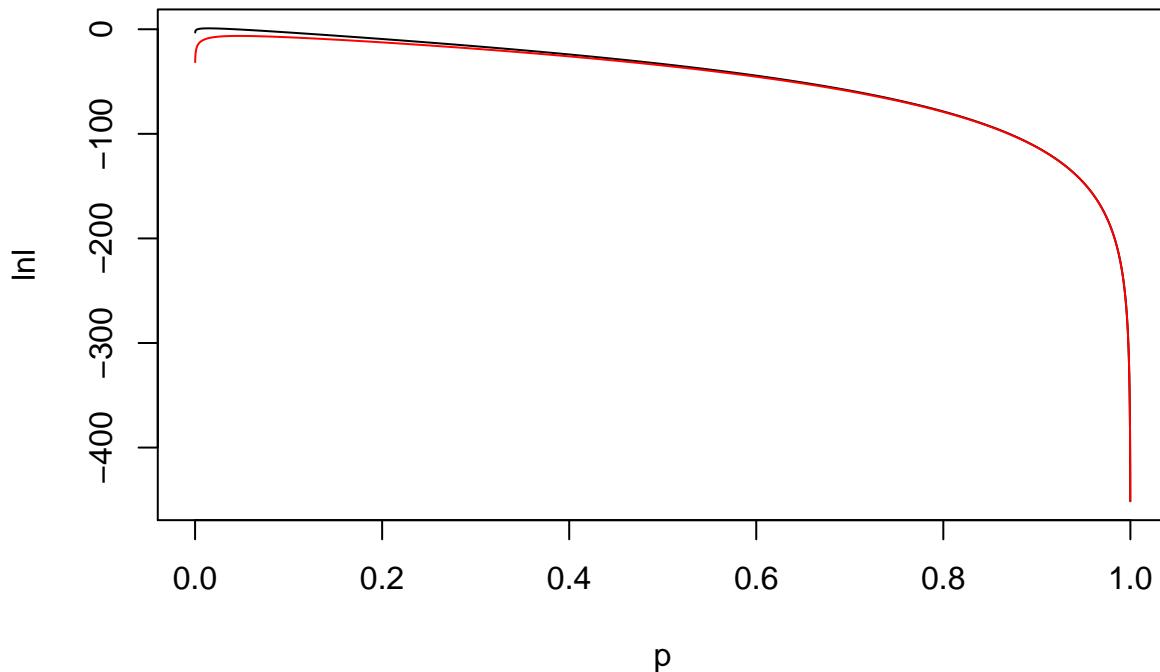
```
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0522"   "0.3347"   "19.1570881226054" "2.98775022408127"
```

a= 20 ,b= 20 ,L= 10



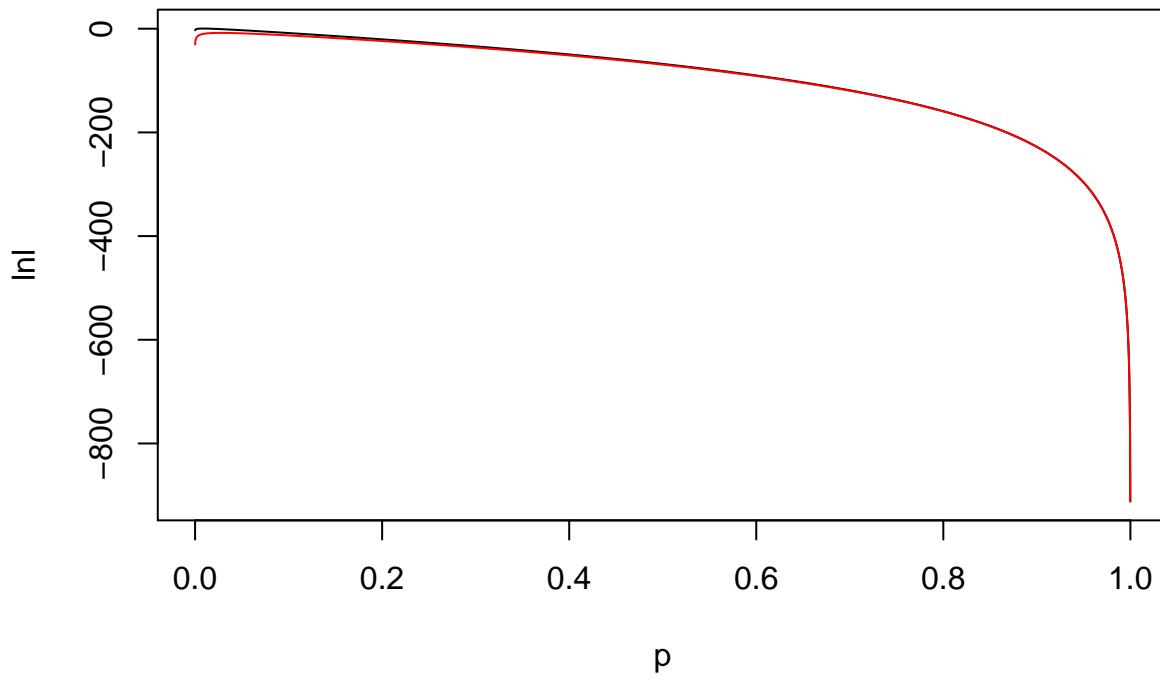
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0366"   "0.1372"   "27.3224043715847" "7.28862973760933"
```

a= 20 ,b= 20 ,L= 50



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"  "1/MCLE"
## [2,] "0.0145" "0.0472" "68.9655172413793" "21.1864406779661"
```

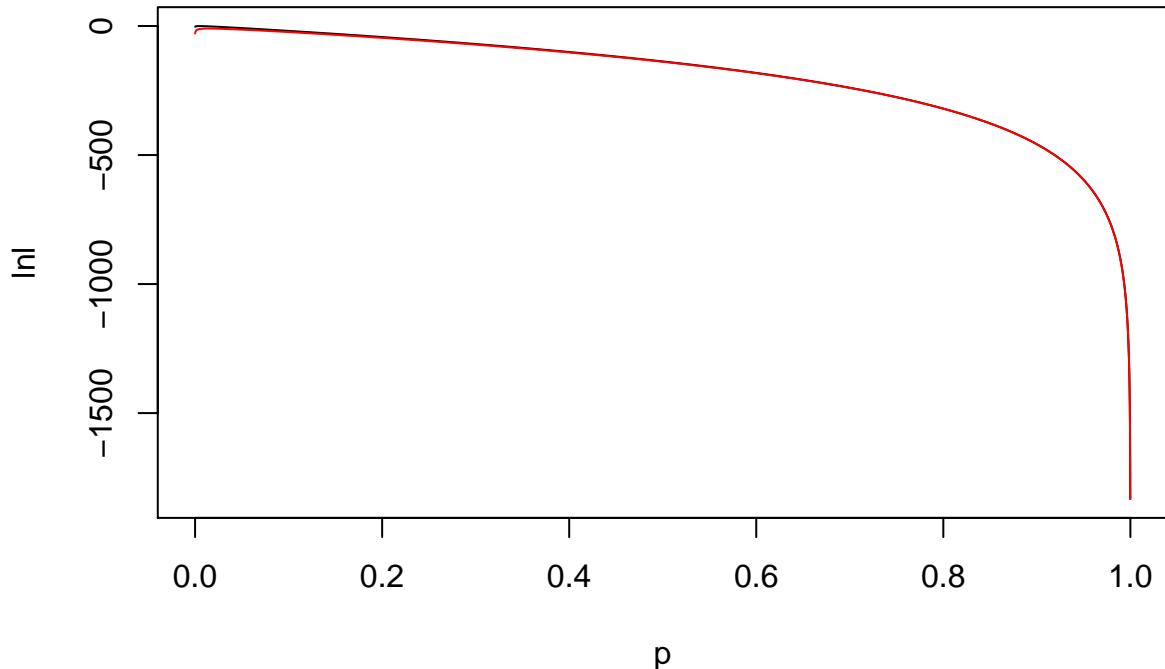
a= 20 ,b= 20 ,L= 100



```
##      [,1]      [,2]      [,3]      [,4]
```

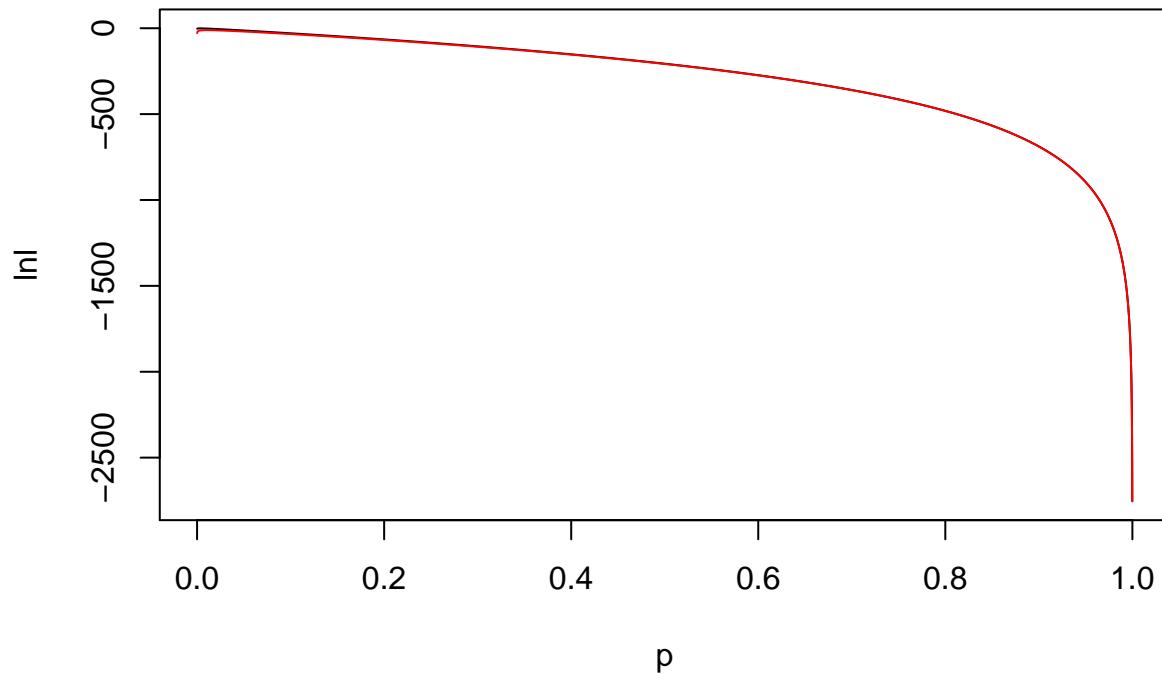
```
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0084"   "0.0271"   "119.047619047619" "36.90036900369"
```

a= 20 ,b= 20 ,L= 200



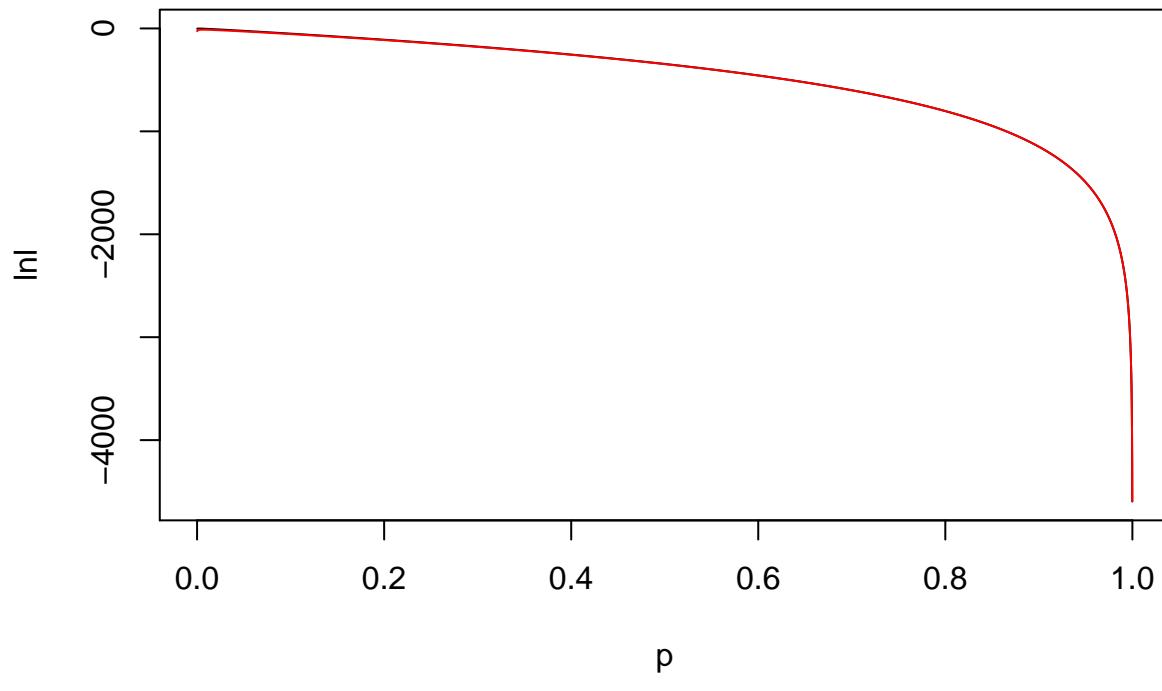
```
##      [,1]      [,2]      [,3]      [,4]  
## [1,] "MLE"      "MCLE"      "1/MLE"      "1/MCLE"  
## [2,] "0.0046"   "0.0147"   "217.391304347826" "68.0272108843537"
```

a= 20 ,b= 20 ,L= 300



```
##      [,1]      [,2]      [,3]      [,4]
## [1,] "MLE"    "MCLE"   "1/MLE"   "1/MCLE"
## [2,] "0.0031" "0.0101" "322.58064516129" "99.009900990099"
```

a= 20 ,b= 20 ,L= 500



```
##      [,1]      [,2]      [,3]      [,4]
```

```
## [1,] "MLE"      "MCLE"      "1/MLE"          "1/MCLE"  
## [2,] "0.0019"   "0.0062"   "526.315789473684" "161.290322580645"
```