<뉴튼-랩슨 레포트>

수리통계학 제 1분반

2019580027 정우진

과제는 PYTHON의 NUMPY 패키지를 이용하여 수행하였다.

1. 100개 난수 추출

import numpy as np  
# 로지스틱 분포로부터 랜덤 수 100개 추출  
  
a = np.random.uniform(0,1,100)  
def logistic\_cmf\_inverse(d, u):  
 return np.log(d / (1-d)) + u  
mean = 4  
dat = logistic\_cmf\_inverse(a, mean)

균등분포를 따르는 확률변수 U ~ Uniform(0,1)와 어떤분포의 분포함수 F(x)에 대하여,

F\_inverse(U)는 X가 따르는 분포를 따르므로 이를 이용하여 Numpy 패키지를 이용해 로지스틱분포를 따르는 난수 100개를 추출하였다. 이때 로지스틱분포의 평균값은 4로 정하였다.

1. 뉴턴-랩슨 메서드 구현 코드
2. # 로지스틱로그가능도함수 미분값 산출  
   init = 5  
   def lld(x, k = init): # logistic\_likelihood\_derivative  
    arr = np.array(x)  
    return len(arr) - 2\*np.sum(1 / (1 + np.exp(-k+arr)))  
     
   def diff\_lld(x, k):  
    eps = 0.000001  
    return (lld(x, k + eps) - lld(x, k)) / eps  
     
   theta = init  
     
   adj = 0.000001  
   iter = 0  
   while True:  
    newtheta = theta - lld(dat, theta) / (diff\_lld(dat,theta) + adj)  
    if np.abs(newtheta - theta) < 0.000001:  
    break  
    theta = newtheta  
    print(newtheta, lld(dat, newtheta))  
    iter += 1

Theta의 초기값은 임의로 5로 설정하였다.

기본적은 틀은 다음과 같다.

* newtheta = theta - lld(dat, theta) / (diff\_lld(dat,theta) + adj)

과정을 newtheta 와 theta의 차가 0.000001보다 작을 때까지 계속한다.

* 위 식의 lld(dat, theta)에서 dat는 데이터값. 즉 1단계에서 추출한 난수 100개를 의미하며, theta는 로지스틱 로그가능도함수의 세타값이며, lld함수는 두 매개변수를 이용하여 로지스틱로그가능도함수의 미분값을 구한다.
* Diff\_lld함수는 dat과 theta를 이용하여 lld함수의 theta에서의 미분값을 산출한다. 이 함수를 구현할 때는 미분의 정의를 이용하였다.
* While 반복문이 재시작 될 때마다 그 단계에서 구한 새로운 theta값과, 그 때의 로지스틱 로그가능도함수의 미분값을 출력한다.
* While 반복문이 종료되면, 새로운 최종 theta 추정값과 총 반복횟수를 출력한다.

3. 갱신과정.

위의 코드를 돌리면,

텍스트이(가) 표시된 사진

자동 생성된 설명

다음과 같이 출력된다.

즉, theta의 추정량을 구하기 위해 N-R메서드를 3회 반복하였으며,

최종 구한 추정량은 약 3.89이고, 이는 참값 4와 가깝다.