

统计计算第6讲作业

Code ▾

苏锦华 2017201620

作业要求

使用数值模拟法构造下题分布的随机变量，取 $n=100$ 的样本量和 $m=100$ 的循环次数来估计其均值与MSE。

又为 $MSE(\hat{\theta}) = E[\hat{\theta} - \theta]^2$.

例3.1.3 (*MSE*的估计) 估计标准正态分布在 $[0,1]$ 截尾后的变量 $\bar{X}_{[0,1]}$ 的均值，以及相应的MSE. n=100

设 X 服从正态分布 $N(\mu, \sigma^2)$ ，如果限制 X 在 (a, b) 区间取值，相应的变量记为 $X_{[a,b]}$ ，则相应的密度函数为

$$f(x; \mu, \sigma, a, b) = \frac{\frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

这里 $\phi(\cdot)$ 和 $\Phi(\cdot)$ 分别表示标准正态分布的密度函数和分布函数。

解答

为了方便验证之后的结果，先计算出该分布的真实均值，根据上述密度函数的计算公式可推得以下密度函数：

$$f(x) = \frac{\phi(x)}{\Phi(1) - \Phi(0)} = \frac{\exp(-0.5 \times x^2)}{2.506 \times 0.3413}$$

$$\mu = \int_0^1 x f(x) dx = -\frac{\exp(-0.5x^2)}{0.8553} \Big|_0^1 = 0.46$$

构造随机变量的样本序列

这里不使用直接变换法生成随机序列，而采用截尾的定义，生成标准正态分布，取落入【0,1】区间的数构成样本。

Hide

```
get_sample = function(n){
  sample = c()
  while (length(sample)<n){
    temp = rnorm(1)
    if (temp>=0 && temp<=1){
      sample = c(sample,temp)
    }
  }
  return(sample)
}
get_sample(100)
```

```
[1] 0.15655191 0.49069826 0.09598239 0.42686795 0.37717786 0.09215945 0.14675666 0.71374004 0.07846630 0.17907611 0.98831677
[12] 0.73087509 0.67288929 0.34315553 0.03306608 0.58482802 0.79757510 0.06761925 0.95183692 0.27949553 0.46615638 0.07588886
[23] 0.86075339 0.66133533 0.80483773 0.05200033 0.66166641 0.45442589 0.88204789 0.64698467 0.15041975 0.08277315 0.58052469
[34] 0.20211064 0.46405253 0.77417886 0.61197947 0.14144616 0.41099614 0.32924661 0.67750776 0.15420901 0.33487795 0.37296594
[45] 0.53813493 0.29391322 0.06362586 0.67191533 0.11570902 0.25028212 0.43058027 0.12894231 0.37683271 0.19928631 0.72322301
[56] 0.73101405 0.38426770 0.56499447 0.79819226 0.17992167 0.21516133 0.66153283 0.01506260 0.82484853 0.68922885 0.60970162
[67] 0.20290319 0.04233170 0.31418894 0.98146514 0.72571281 0.52057093 0.98993316 0.63352656 0.56809614 0.45236388 0.30104302
[78] 0.69360908 0.79464964 0.85828959 0.12789220 0.58832987 0.68537318 0.50560507 0.23063168 0.76751024 0.24178337 0.45162225
[89] 0.37431359 0.15946291 0.14785134 0.65941540 0.74000072 0.60805207 0.79665527 0.76401199 0.39208810 0.17537441 0.71123903
[100] 0.83712448
```

估计均值

通过循环得到m组样本，每组样本计算均值保存，最后求平均可到如下结果，发现误差在0.0036

[Hide](#)

```
m = 100
n=100
mu_group = c()
for(i in 1:m){
  mu_ = mean(get_sample(n))
  mu_group = c(mu_group,mu_)
}
mean(mu_group)
```

```
[1] 0.4564514
```

估计MSE

最后用均值集合计算MSE,结果如下：

[Hide](#)

```
real_mu = 0.46
sum((mu_group - real_mu) * (mu_group - real_mu)) / m
```

```
[1] 0.0007866596
```