statComp_hw4

凌浩东

10/28/2021

1

For $x \in [a_{i-1}, a_i]$, $i \in 1, \cdots, j$,

$$\begin{aligned}
P(X^* \le x | J = i) &= \int_{-\infty}^{x} f_i(x) dx \\
&= k \int_{-a_{i-1}}^{x} f(x) dx \\
&= k(F(x) - \frac{i-1}{k}) \\
&= kF(x) - i + 1
\end{aligned}$$

Hence,

$$P(X^* \le x | J = i) = \begin{cases} kF(x) - i + 1 & x \in [a_{i-1}, a_i] \\ 0 & x \in (-\infty, a_{i-1}) \\ 1 & x \in (a_i, \infty) \end{cases}$$

Hence,

$$
\begin{aligned}
P(X^* \leq x) &= \sum_{j=1}^{n} P(X^* \leq x | J = j) P(J = j) \\
&= \frac{1}{k} \sum_{j=1}^{n} P(X^* \leq x | J = j) \\
&= \frac{1}{k} \left(\sum_{j=1}^{i-1} 1 + P(X^* \leq x | J = i) + \sum_{j=i+1}^{n} 0 \right) \\
&= \frac{1}{k} \left(i - 1 + [kF(x) - i + 1] \right) \\
&= F(x)
\end{aligned}
$$

Hence, $X^* \sim X$.

For Y^*,

$$
Y^* = \frac{g_j(X^*)}{f_j(X^*)} = \frac{g_j(X^*)}{kf(X^*)} \sim \frac{g(X)}{kf(X)} = \frac{Y}{k}
$$

with $X^* \sim X$.

2

exercise 6.1

```r
n <- 20
m <- 1000
mse <- numeric(9)
for (k in 1:9) {
  tmean <- numeric(m)
  for (i in 1:m) {
    x <- sort((rcauchy(n)))
    tmean[i] <- sum(x[(k+1):(n-k)])/(n-2*k)
  }
  target <- median(tmean)
  mse[k] <- mean((tmean - target)^2)
}
```

```r
k <- c(1:9)
rbind(k, mse)
```

```
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## k   1.000000 2.0000000 3.0000000 4.0000000 5.0000000 6.0000000 7.0000000
## mse 1.735897 0.4097611 0.2216231 0.1745251 0.1683093 0.1426654 0.1330567
##           [,8]      [,9]
## k   8.0000000 9.0000000
## mse 0.1290898 0.1357178
```

exercise 6.9

```r
n <- 20
m <- 1000
gini.ratio <- numeric(m)
for (j in 1:m) {
  x <- sort(rlnorm(n))
  mu <- mean(x)
  G <- 0
  for (i in 1:n) {
    G <- G + (2*i - n - 1) * x[i]
  }
  gini.ratio[j] <- G / (n^2 * mu)
}
mean <- mean(gini.ratio)
median <- median(gini.ratio)
c(mean, median)
```

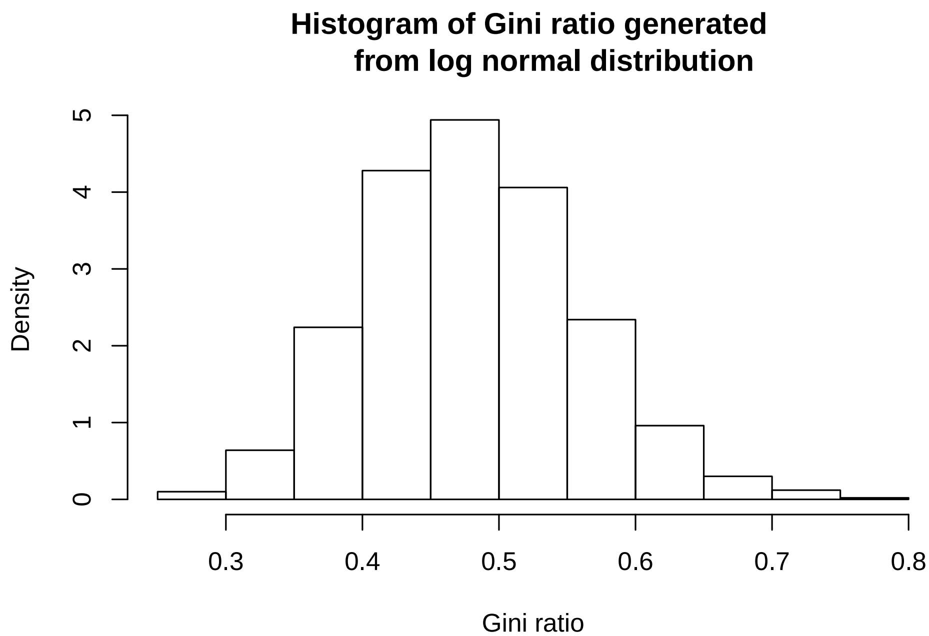
```
## [1] 0.4811732 0.4759992
```

```r
quantile(gini.ratio, probs = seq(.1, .9, by = .1))
```

```
##       10%       20%       30%       40%       50%       60%       70%       80%
## 0.3851137 0.4159509 0.4385938 0.4559212 0.4759992 0.4978336 0.5189110 0.5455270
```

```
##       90%
## 0.5847519
```

```
hist(gini.ratio, probability = T,
     main = 'Histogram of Gini ratio generated
     from log normal distribution',
     xlab = 'Gini ratio')
```




```
n <- 20
m <- 1000
gini.ratio <- numeric(m)
for (j in 1:m) {
  x <- sort(runif(n))
  mu <- mean(x)
  G <- 0
  for (i in 1:n) {
    G <- G + (2*i - n - 1) * x[i]
  }
  gini.ratio[j] <- G / (n^2 * mu)
```

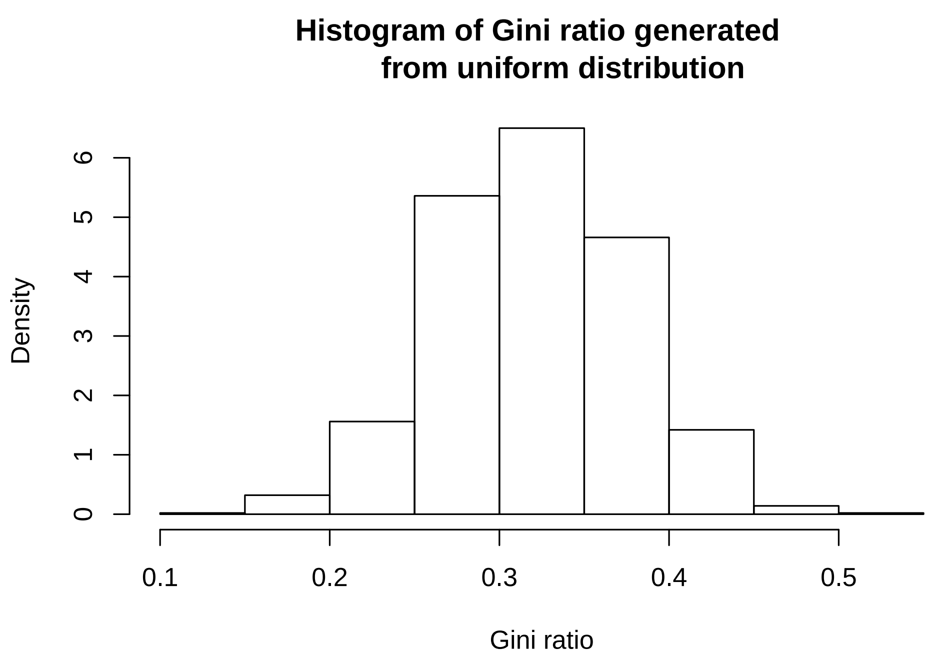
```
}
mean <- mean(gini.ratio)
median <- median(gini.ratio)
c(mean, median)
```

```
## [1] 0.3214615 0.3202069
```

```
quantile(gini.ratio, probs = seq(.1, .9, by = .1))
```

```
##       10%       20%       30%       40%       50%       60%       70%       80%
## 0.2531085 0.2730138 0.2914422 0.3059615 0.3202069 0.3356224 0.3515961 0.3691293
##       90%
## 0.3937447
```

```
hist(gini.ratio, probability = T,
     main = 'Histogram of Gini ratio generated
     from uniform distribution',
     xlab = 'Gini ratio')
```



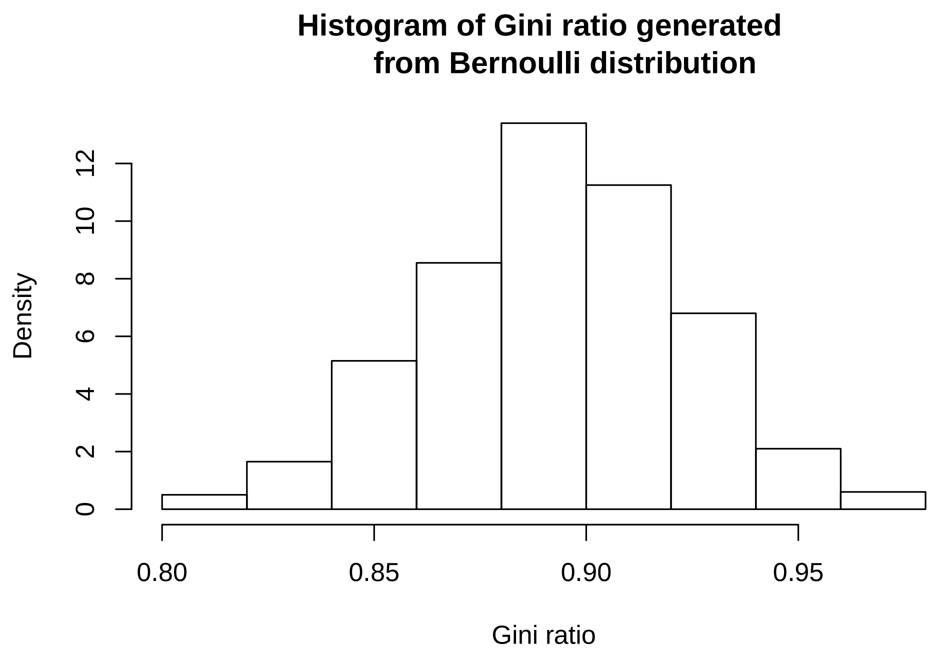
```
n <- 100
m <- 1000
gini.ratio <- numeric(m)
for (j in 1:m) {
  x <- sort(rbinom(n, size = 1, prob = .1))
  mu <- mean(x)
  G <- 0
  for (i in 1:n) {
    G <- G + (2*i - n - 1) * x[i]
  }
  gini.ratio[j] <- G / (n^2 * mu)
}
mean <- mean(gini.ratio)
median <- median(gini.ratio)
c(mean, median)
```

```
## [1] 0.89808 0.90000
```

```
quantile(gini.ratio, probs = seq(.1, .9, by = .1))
```

```
##  10%  20%  30%  40%  50%  60%  70%  80%  90%
## 0.86 0.87 0.88 0.89 0.90 0.91 0.91 0.92 0.94
```

```
hist(gini.ratio, probability = T,
     main = 'Histogram of Gini ratio generated
     from Bernoulli distribution',
     xlab = 'Gini ratio')
```



3

As $X \sim N(0,1)$, denote the pdf and cdf of standard normal as $f(x)$ and $F(x)$ respectively, then

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} F^{i-1}(x)(1-F(x))^{n-i} f(x)$$
$$f_{X_{(n-i+1)}}(x) = \frac{n!}{(i-1)!(n-i)!} F^{n-i}(x)(1-F(x))^{i-1} f(x)$$

Known $F(-x)=1-F(x)$, $f(x)=f(-x)$, we know that

$$\begin{aligned}
E\left[X_{(i)}+X_{(n-i+1)}\right]=&\int_{-\infty}^{\infty}\frac{n!}{(i-1)!(n-i)!}F^{i-1}(x)(1-F(x))^{n-i}f(x)dx+\\
&\int_{-\infty}^{\infty}\frac{n!}{(i-1)!(n-i)!}F^{n-i}(x)(1-F(x))^{i-1}f(x)dx\\
=&\frac{n!}{(i-1)!(n-i)!}\left(\int_{-\infty}^{\infty}F^{i-1}(x)(1-F(x))^{n-i}f(x)dx+\right.\\
&\left.\int_{-\infty}^{\infty}F^{n-i}(-x)(1-F(-x))^{i-1}f(-x)dx\right)\\
=&\frac{n!}{(i-1)!(n-i)!}\left(\int_{-\infty}^{\infty}F^{i-1}(x)(1-F(x))^{n-i}f(x)dx+\right.\\
&\left.\int_{-\infty}^{\infty}(1-F(x))^{n-i}F^{i-1}(x)f(x)dx\right)\\
=&0
\end{aligned}$$

Hence,

$$\begin{aligned}
E\left[\bar{X}_{[-k]}\right]&=E\left[\frac{1}{n-2k}\sum_{i=k+1}^{n-k}X_{(i)}\right]\\
&=\frac{1}{n-2k}\sum_{i=k+1}^{n-k}E\left[X_{(i)}\right]\\
&=\frac{1}{2(n-2k)}\sum_{i=k+1}^{n-k}E\left[X_{(i)}+X_{(n-i+1)}\right]\\
&=0
\end{aligned}$$

4

Denote $\frac{X_i-\mu}{\sigma}$ as Z_i, $Y=AZ$, where A is an orthogonal matrix, and the first row of A is $(\frac{1}{\sqrt{n}},\cdots,\frac{1}{\sqrt{n}})$, then $Y\sim N(0,I)$,

$$\|Y\|^2=\sum_i Y_i^2\sim\chi^2(n)$$

and

$$Y_1=\frac{\sum_i(X_i-\mu)}{\sqrt{n}\sigma}=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1)$$

where Y_1 is the first value of vector Y.

Hence

$$\begin{aligned}\frac{\sum_i (X_i - \bar{X})^2}{\sigma^2} &= \frac{\sum_i (X_i - \mu)^2}{\sigma^2} - \frac{n(\bar{X} - \mu)^2}{\sigma^2} \\ &= \|Z\|^2 - Y_1^2 \\ &= \|Y\|^2 - Y_1^2 \\ &\sim \chi^2(n-1)\end{aligned}$$

and $\frac{\sum_i (X_i - \bar{X})^2}{\sigma^2}$ is independent with $Y_1 = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$, i.e., $\frac{\sum_i (X_i - \bar{X})^2}{\sigma^2}$ is independent with $\bar{X}$