胡晓雪_21377059_6

21377059 胡晓雪

2023-12-02

一、Lab9 理论推导

```r
knitr::include_graphics("D:/桌面/lab9.jpg")
```

理论推导:

由题可知, $Y|\theta \sim Possion(N\theta)$ $\quad \theta \sim Gramma(a,b)$

则 $f(Y|\theta) = \frac{(N\theta)^y e^{-N\theta}}{y!}$ $\quad p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$

似然: $f(Y=y_1,\cdots,Y=y_k|\theta) = \prod_{i=1}^{k} \frac{e^{-N\theta}(N\theta)^y}{y_i!}$

则 $\log f(Y|\theta) = \sum_{i=1}^{k} \log \frac{e^{-N\theta}(N\theta)^y}{y_i!} = \sum_{i=1}^{k} \left[-N\theta + y_i \log(N\theta) - \log y_i!\right]$

$\hat{\theta}_{MAP} = \underset{\theta}{argmax}\, p(\theta|Y) = \underset{\theta}{argmax}\left[\log f(Y|\theta) + \log p(\theta)\right]$

则 $\hat{\theta}_{MAP} = \underset{\theta}{argmax}\left[-NK\theta + \sum_{i=1}^{k} y_i \log \theta N - \sum_{i=1}^{k} \log y_i! + \log\left(\frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}\right)\right]$

$F(\theta) = -NK\theta + \sum_{i=1}^{k} y_i \log\theta + (a-1)\log\theta - b\theta$

$$\frac{\partial F(\theta)}{\partial\theta} = -NK + \frac{\sum_{i=1}^{k} y_i}{\theta} + \frac{a-1}{\theta} - b = 0$$

$$(NK+b)\theta = \sum_{i=1}^{k} y_i + (a-1)$$

$$\theta = \frac{\sum_{i=1}^{k} y_i + (a-1)}{NK+b}$$

即 $\hat{\theta}_{MAP} = \frac{\sum_{i=1}^{k} y_i + a - 1}{NK+b}$

二、Lab10 拒绝采样

（一）定义目标概率密度函数:标准正态分布

```r
# 定义目标概率密度函数（标准正态分布）
target_pdf <- function(x) {
  dnorm(x)
}
```

（二）定义猜想概率密度函数

```r
# 定义猜想概率密度函数
proposal_pdf <- function(x, name) {
  if (name == 'logist') {
    dlogis(x, 0, 1)
  } else if (name == 'Cauchy') {
    dcauchy(x, 0, 1)
  } else if (name == 't') {
    dt(x, 5)
  } else if (name == 'N') {
    dnorm(x, mean = 1, sd = 2)
  }
}
```

（三）定义猜想分布的抽样函数

```r
#定义猜想分布的抽样函数
proposal <- function(name) {
  if (name == 'logist') {
    rlogis(1, 0, 1)
  } else if (name == 'Cauchy') {
    rcauchy(1, 0, 1)
  } else if (name == 't') {
    rt(1, 5)
  } else if (name == 'N') {
    rnorm(1, 1, 2)
  }
}
```

（四）定义拒绝采样函数

```r
# 定义拒绝采样函数
rejection_sampling <- function(n,M,name,al) {
  samples <- numeric(n)
  i <- 1
  j <- 1
  while (i <= n) {
    # 从猜想分布中抽取一个样本
    j<-j+1
    x <- proposal(name)
    # 生成一个在[0, 1]之间的随机数
    u <- runif(1)

    # 判断抽取的样本是否在f(x)之下，只要是满足该条件的都留下，不管u 对应有多大
    if (u <= target_pdf(x) / (M * proposal_pdf(x,name))) {
      samples[i] <- x
      i <- i + 1
    }
  }
```

```r
  # 查看理论与实际接受率
  alpha <- i/j
  al[1,1]<- 1/M
  al[2,1]<- alpha

  return(list(samples = samples, al = al))
}
```

（五）定义实施抽样函数

```r
# 定义实施抽样函数
result <- function(name,M){

  #查看接受率
  al <-data.frame(matrix(0,nrow = 2,ncol = 1))
  rownames(al)=c("理论接受率","实际接受率")
  colnames(al)=c(name)
  # 进行拒绝采样，生成 1000 个符合标准正态分布的随机样本
  sim<-rejection_sampling(1000,M,name,al)
  samples <- sim$samples
  al<- sim$al
  # 拒绝抽样得到的曲线和直方图
  plot(density(samples),ylim=c(0,0.5),lwd=2,main=name,col=4)
  hist(samples,freq = F,add=TRUE,col = "#80808080")
  # 理论上要得到的密度曲线
  curve(dnorm(x), -6, 6,add=TRUE, lwd=2,xlab = expression(theta), ylab
= "Density", n = 1000)
  legend("topleft", c("proposal density", "Normal density"),
         col = c(4,1), bty = "n", lty = 1,cex=0.8)

  return (al)
}
set.seed(2017-12-4)
```

（六）查看不同猜想分布的结果

（1）logist 分布

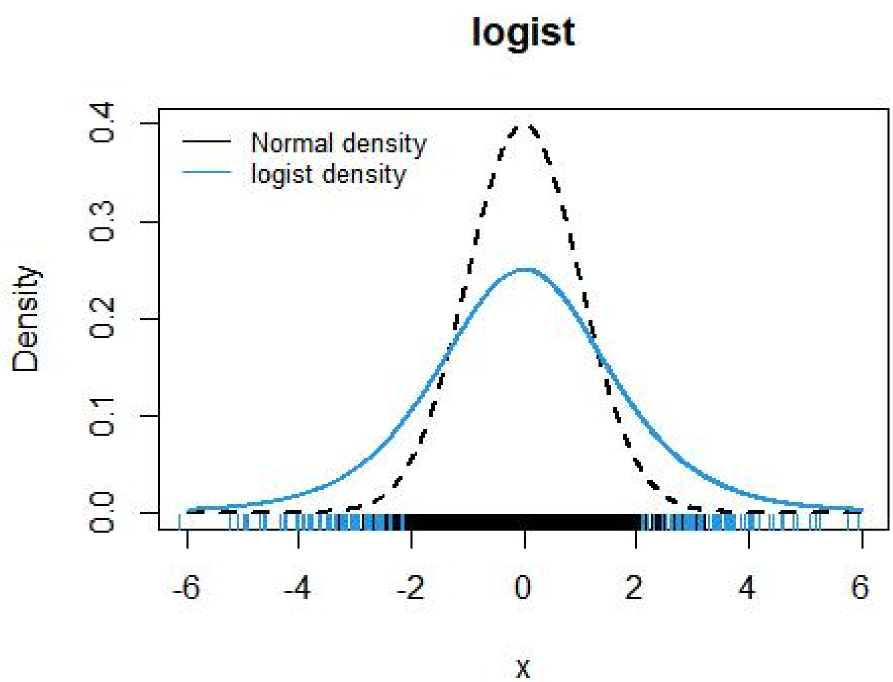
```r
#logist 分布

name="logist"
# 设置上界常数 M
x <- seq(-6, 6, length.out = 10000)
ratio <- target_pdf(x) / proposal_pdf(x,name)
M <- max(ratio, na.rm = TRUE)
#查看猜想分布与目标分布的曲线图
curve(dnorm(x), -6, 6, xlab = "x", ylab = "Density", n = 1000,lwd=2,lty
 = 2,main=name)
curve(dlogis(x, 0, 1), -6, 6, add = TRUE, col = 4, n = 1000,lwd=2)
```

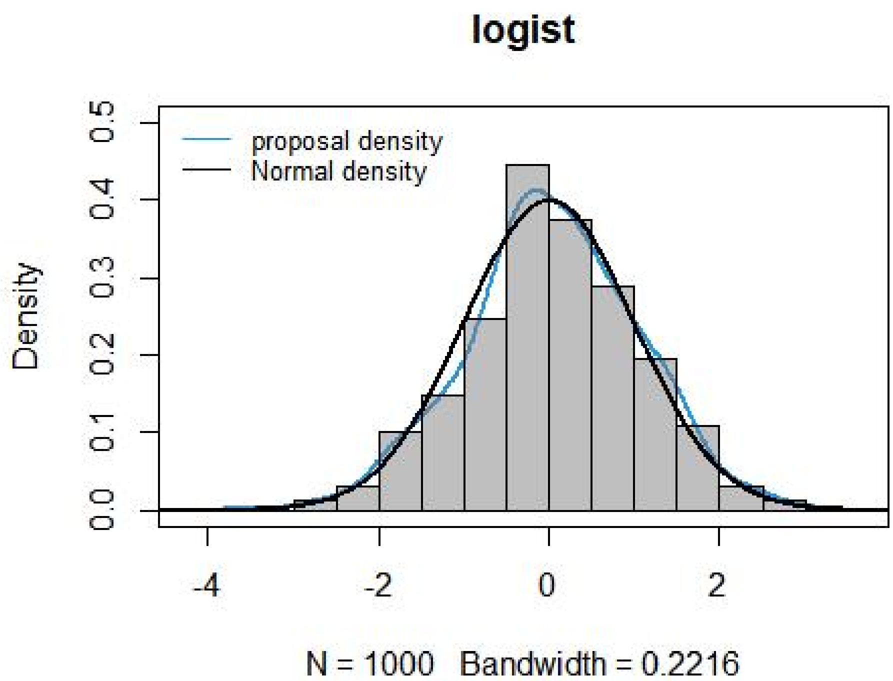
```r
legend("topleft", c("Normal density", "logist density"),
       col = c(1,4), bty = "n", lty = 1,cex=0.8)
#对比样本点的分布
x <- rlogis(1000, 0,1)
rug(x, col = 4)

## Warning in rug(x, col = 4): 有些值不能被修整

y <- rnorm(1000,0,1)
rug(y,col=1)
```




```r
#实施拒绝抽样
result(name,M)
```




```
##              logist
## 理论接受率 0.6266571
## 实际接受率 0.6347495
```

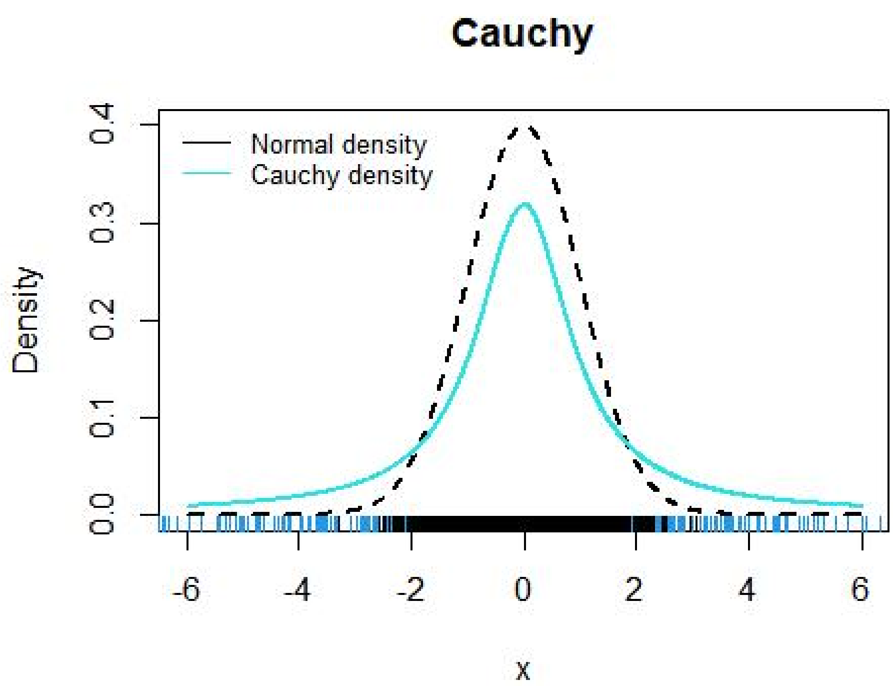
由图可以看出，**logist** 分布的密度函数比标准正态分布函数更加厚尾，可以作为猜想概率密度

（2）Cauchy 分布

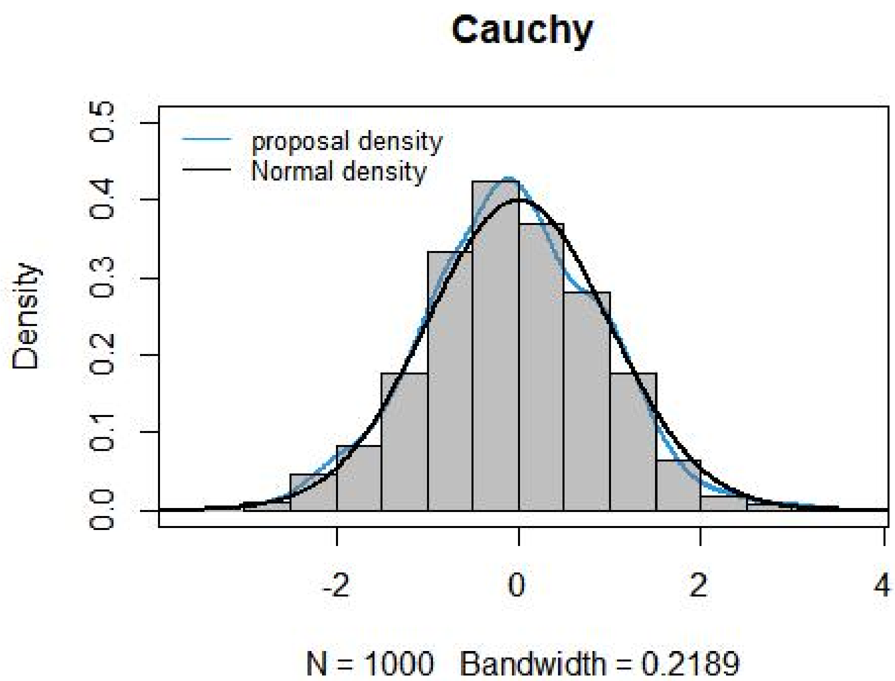
```
#Cauchy 分布
name="Cauchy"
# 设置上界常数M
x <- seq(-6, 6, length.out = 10000)
ratio <- target_pdf(x) / proposal_pdf(x,name)
M <- max(ratio, na.rm = TRUE)
#查看猜想分布与目标分布的曲线图
curve(dnorm(x), -6, 6, xlab = "x", ylab = "Density", n = 1000,lwd=2,lty
 = 2,main=name)
curve(dcauchy(x, 0, 1), -6, 6, add = TRUE, col = 5, n = 1000,lwd=2)
legend("topleft", c("Normal density", "Cauchy density"),
       col = c(1,5), bty = "n", lty = 1,cex=0.8)
#对比样本点的分布
x <- rcauchy(1000, 0,1)
rug(x, col = 4)

## Warning in rug(x, col = 4): 有些值不能被修整
```

```r
y <- rnorm(1000,0,1)
rug(y,col=1)
```




```r
#实施拒绝抽样
result(name,M)
```

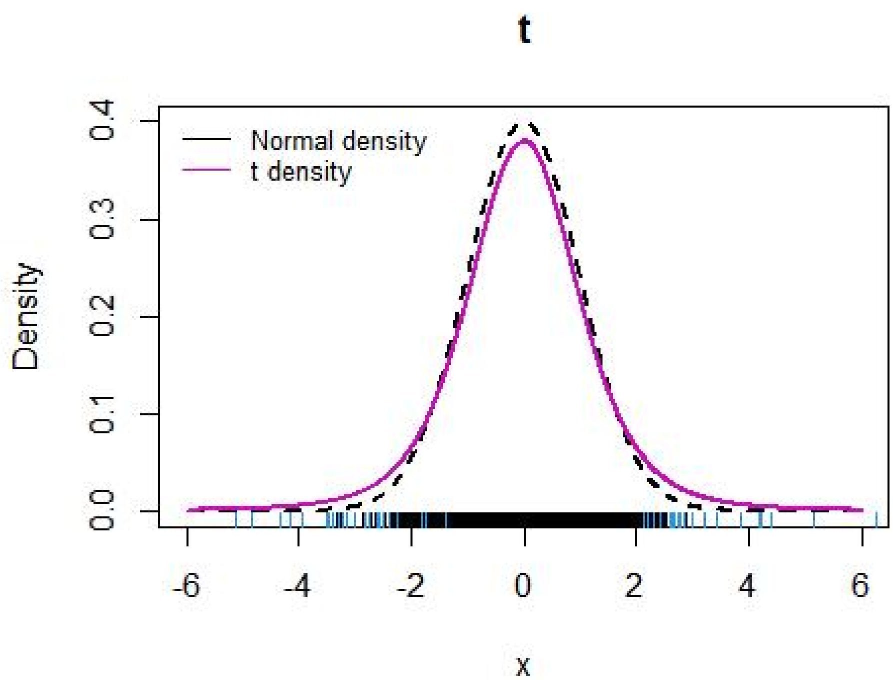



```
##                 Cauchy
## 理论接受率 0.6577447
## 实际接受率 0.6568241
```

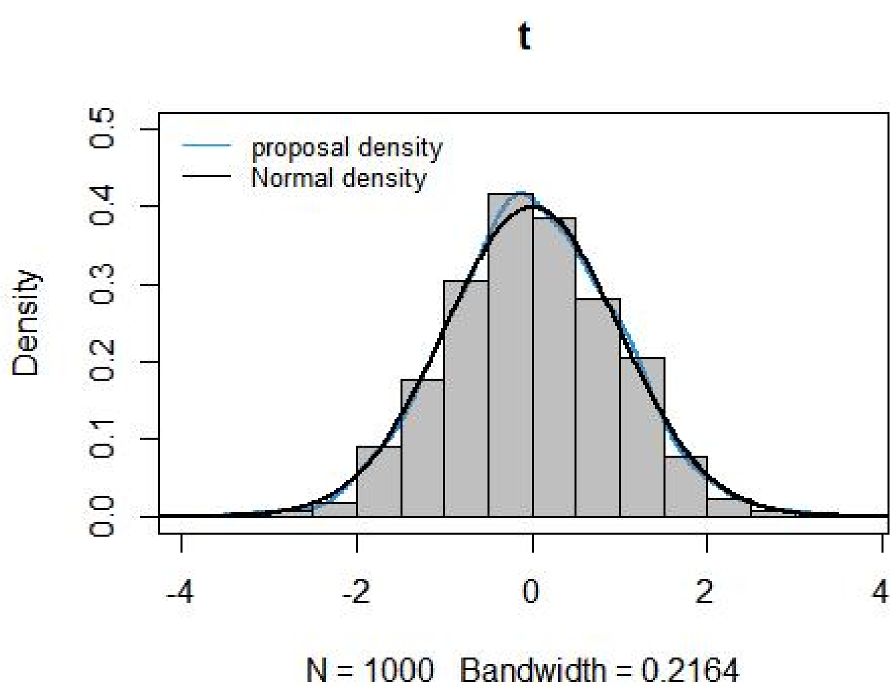
由图可以看出，Cauchy 分布的密度函数比标准正态分布函数更加厚尾，可以作为猜想概率密度

（3）t 分布（自由度为 5）

```
#t 分布
name="t"
# 设置上界常数M
x <- seq(-6, 6, length.out = 10000)
ratio <- target_pdf(x) / proposal_pdf(x,name)
M <- max(ratio, na.rm = TRUE)
#查看猜想分布与目标分布的曲线图
curve(dnorm(x), -6, 6, xlab = "x", ylab = "Density", n = 1000,lwd=2,lty
 = 2,main=name)
curve(dt(x, 5), -6, 6, add = TRUE, col = 6, n = 1000,lwd=2)
legend("topleft", c("Normal density","t density"),
       col = c(1,6), bty = "n", lty = 1,cex=0.8)
#对比样本点的分布
x <- rt(1000,5)
rug(x, col = 4)
y <- rnorm(1000,0,1)
rug(y,col=1)
```




```
#实施拒绝抽样
result(name,M)
```



```
##                    t
## 理论接受率 0.9078776
## 实际接受率 0.9124886
```

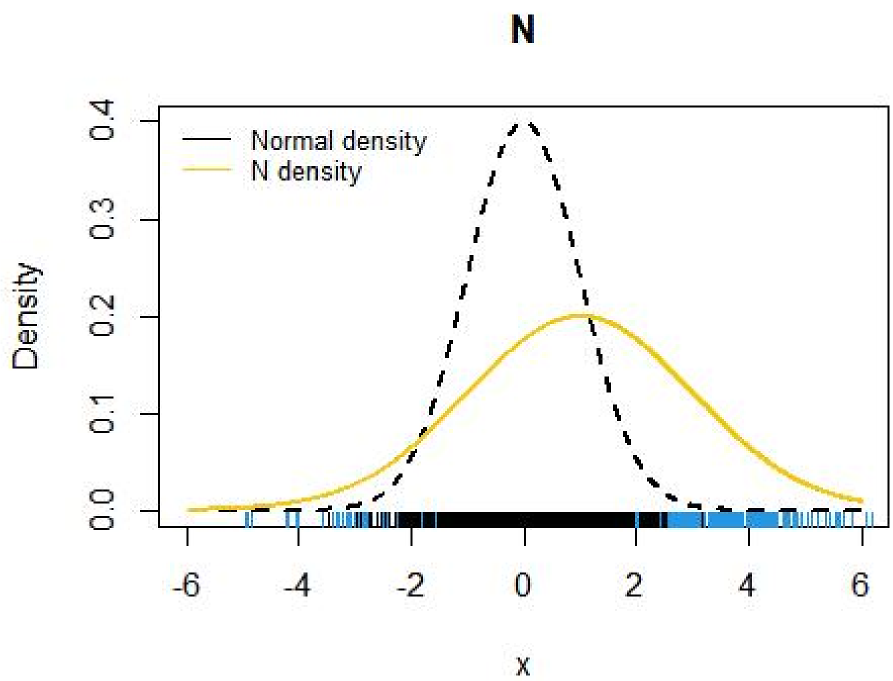
由图可以看出，自由度为 5 的 t 分布的密度函数比标准正态分布函数更加厚尾，可以作为猜想概率密度

（四）正态分布（均值为 1，方差为 2）

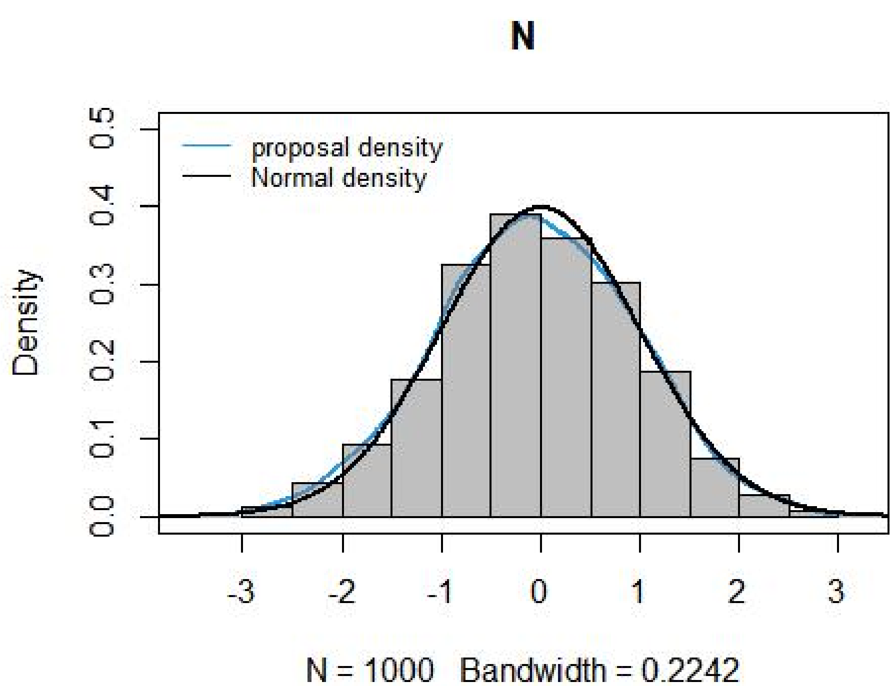
```
#正态分布
name="N"
# 设置上界常数M
x <- seq(-6, 6, length.out = 10000)
ratio <- target_pdf(x) / proposal_pdf(x,name)
M <- max(ratio, na.rm = TRUE)
#查看猜想分布与目标分布的曲线图
curve(dnorm(x), -6, 6, xlab = "x", ylab = "Density", n = 1000,lwd=2,lty
 = 2,main=name)
curve(dnorm(x, mean = 1, sd = 2), -6, 6, add = TRUE, col = 7, n = 1000,
lwd=2)
legend("topleft", c("Normal density","N density"),
       col = c(1,7), bty = "n", lty = 1,cex=0.8)
#对比样本点的分布
x <- rnorm(1000,1,2)
rug(x, col = 4)

## Warning in rug(x, col = 4): 有些值不能被修整

y <- rnorm(1000,0,1)
rug(y,col=1)
```




```
#实施拒绝抽样
result(name,M)
```



```
##                     N
## 理论接受率 0.4232409
## 实际接受率 0.4292453
```

由图可以看出，均值为 **1**，方差为 **2** 的正态分布的密度函数比标准正态分布函数更加厚尾，可以作为猜想概率密度

总结

由图对比我们可以看出，自由度为 **5** 的 **t** 分布作为猜想概率密度时通过拒绝抽样得到的曲线与标准正态分布曲线拟合得较好，并且对于实际接受率可以看出，自由度为 **5** 的 **t** 分布的实际接受率比较高，对于样本的浪费较少，因此我认为选用自由度为 **5** 的 **t** 分布作为猜想概率密度更好