非参数统计

Bag of Little Bootstrap

张颖（组长） 2013202023
赖基正 2014201515
冯艺超 2014201503
刘思伽 2014201545
邹艾伶 2014201573

01

PART ONE

Bootstrap研究现状

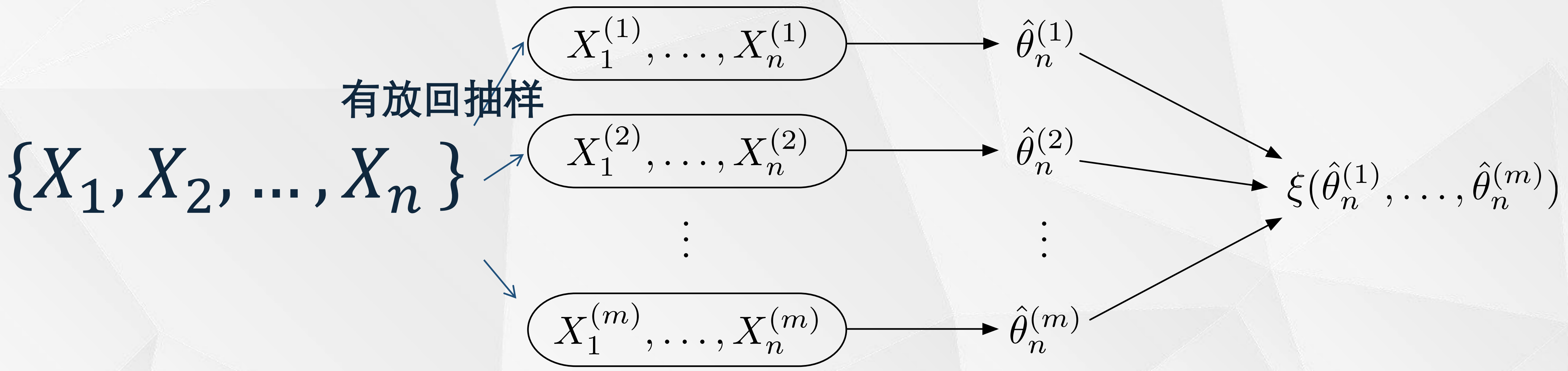
01/ (n of n) Bootstrap

(n of n) Bootstrap

(Efron,1979) 第一次提出使用Bootstrap的非参估计方法

(Bickel and Freedman, 1981; Gin´e and Zinn, 1990; van der Vaart and Wellner, 1996)

证明了Bootstrap估计的收敛性（依概率收敛）



(n of n) Bootstrap 优点

- 开创一种新的枢轴量性质的非参数估计（方差、置信区间 etc.）的方法
- 简单易懂的计算方法，计算机可以自动计算
- Bootstrap估计量具有很好的性质（依概率收敛）比我们使用渐进理论得到的估计量更准确（大体上）
- 每次重抽样过程是独立的（那Bootstrap是否更便于并行与分布式计算呢?？）

01/ 海量数据中Bootstrap的使用

- 在计算机性能得到大大提高后(运行速度，内存容量，成本降低), Bootstrap得到广泛的应用
- 并行式运算（同时运行Bootstrap算法）
- 分布式计算（在多台机器上共同进行Bootstrap算法）
- Bootstrap在**海量**数据中的应用会遇到什么困难？

(n of n) Bootstrap 缺点

当数据量n非常大时，Bootstrap会遇到什么困难？

- 从n个样本里重抽样大小为n的样本，**计算量**会很大，**存储**也占空间（抽出来的样本**种类**的期望=0.632n）
- 假设我们有1TB数据，那么每次重抽样会大约占632GB
- 每次重抽样都需要从这1TB数据里Bootstrap
- 使用分布式计算时，转移这1TB数据的时间就需要很长

(n of n) Bootstrap的改进

(Efron, 1988; Efron & Tibshirani, 1993)

提出减少Bootstrap重抽样次数的方法,加快算法

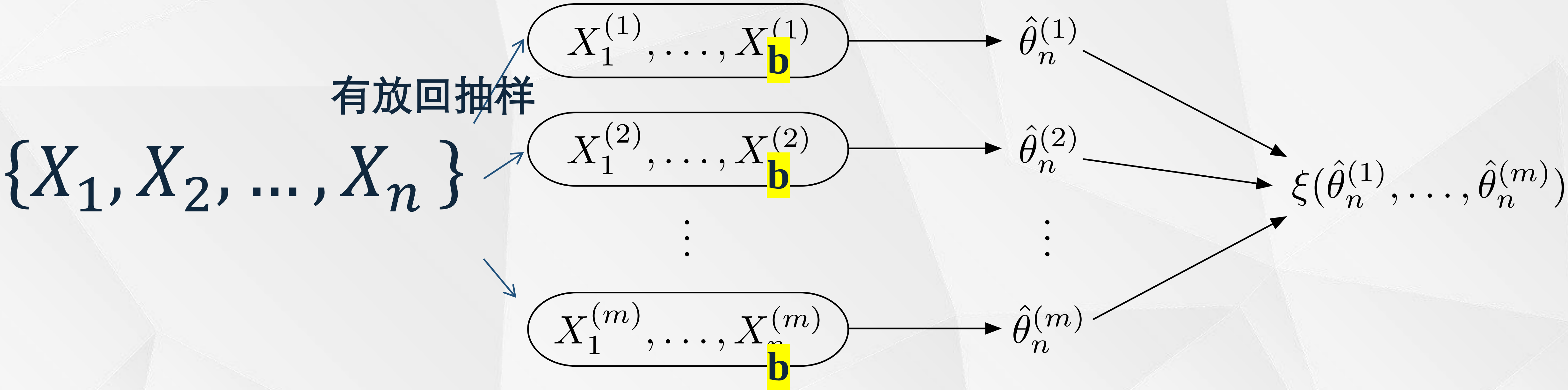
- 但是这种方法又增加了计算复杂程度（怎么确定重抽样的次数的计算）
- 实际上并没有减少很多计算量

01/ b of n Bootstrap

b of n Bootstrap

(Bickel et al., 1997) 提出了b of n Bootstraps 的方法

和Bootstrap不同的地方：每次抽样抽取b个样本（b<n）进行估计

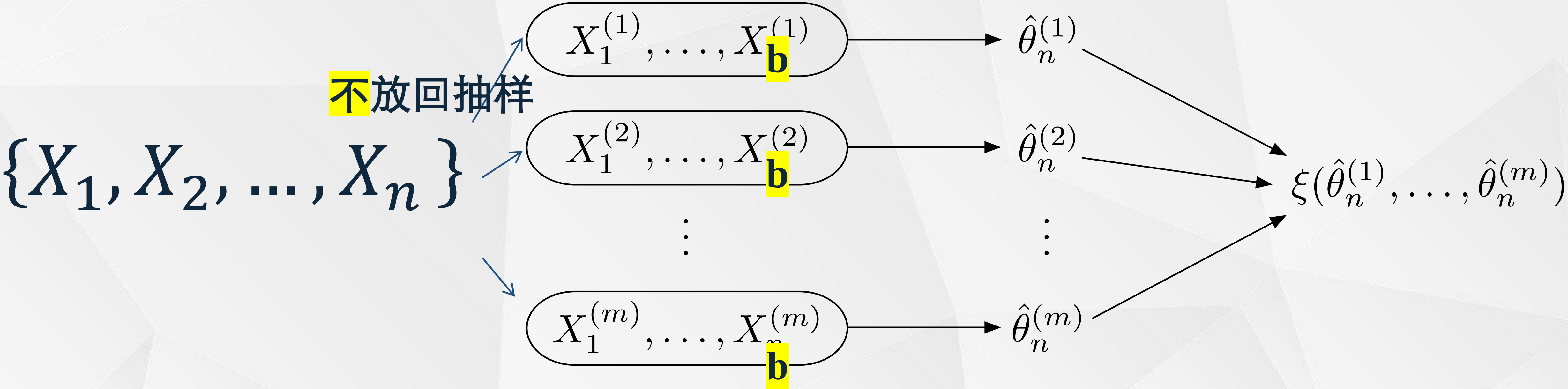


01/ Subsampling $\binom{n}{m}$ Bootstrap

Subsampling $\binom{n}{m}$ Bootstrap

(Politis et al., 1999) 提出了Subsampling的方法

和Bootstrap不同的地方：不放回抽样，抽取b个样本（b<n）进行估计



b of n Bootstrap的优点

- 每次重抽样只抽出b个样本(b<n)，减少了重抽样的计算量
- 每次只需要b个重抽样样本，减少所占内存空间
- 每次计算估计量只需要计算b个重抽样样本(b<n)，减少了估计量的计算量
- 在极值相关的估计方面比n of n bootstrap做得更好（极值估计量的性质的收敛性）

b of n Bootstrap 的缺点

- 估计量对b的取值非常敏感，b取值对估计量影响很大
- 用b个重抽样样本算的估计值的性质（比如方差）和n个重抽样样本计算的估计值的方差是有差别的
- b of n Bootstrap算出的估计量的方差比n of n Bootstrap算出的方差更大
- 需要使用到收敛的阶数的知识来对b of n Bootstrap算出的结果做修正。(计算复杂化，需要理论推导）

b of n Bootstrap 的改进

- (Bickel & Sakov, 2008) 提出了一个用输入数据自动计算出最佳的b的算法。
- 但是最佳b的计算很复杂
- (Bickel and Yahav, 1988; Bickel and Sakov, 2002) 提出了用两个不同的b同时进行估计的方法
- 但同样计算复杂，需要估计量分布的展开等理论知识，同时还需要对不同的b运行多次，耗时长。

01/ 存在问题

- 已有算法都相当复杂
- 并没有充分利用已有的资源来改进算法
- 缺失了(n of n) Bootstrap简单自动的算法优越性
- 例如，没有一种算法充分结合分布式计算以及并行运算

02

PART TWO

BLB 方法的实现、原理和程序

02 BLB 方法的实现、原理和程序

BLB 方法估计过程：

已有数据 X_1， X_2 ，……， X_n

（1）不放回地从 X_1， X_2 ，……， X_n 中抽取b个样本 $X_1^{\#}, X_2^{\#}, \ldots, X_b^{\#}$

（2）有放回地从 $X_1^{\#}, X_2^{\#}, \ldots, X_b^{\#}$ 中抽取n个样本，计算 $\hat{\theta}$

（3）重复步骤（2）共 r 次，得到 r 个估计量 $\hat{\theta}$

（4）计算估计量的评价指标 ξ（ $\hat{\theta}$ ）（方差、置信区间等）

（4）重复步骤（1），（2），（3）共 s 次，得到 s 个 ξ（ $\hat{\theta}$ ）

（5）计算 ξ（ $\hat{\theta}$ ）的均值

$$\xi(\hat{\theta}) = s^{-1} \sum_{i=1}^{s} \xi_i(\hat{\theta})$$

02 BLB方法的实现、原理和程序

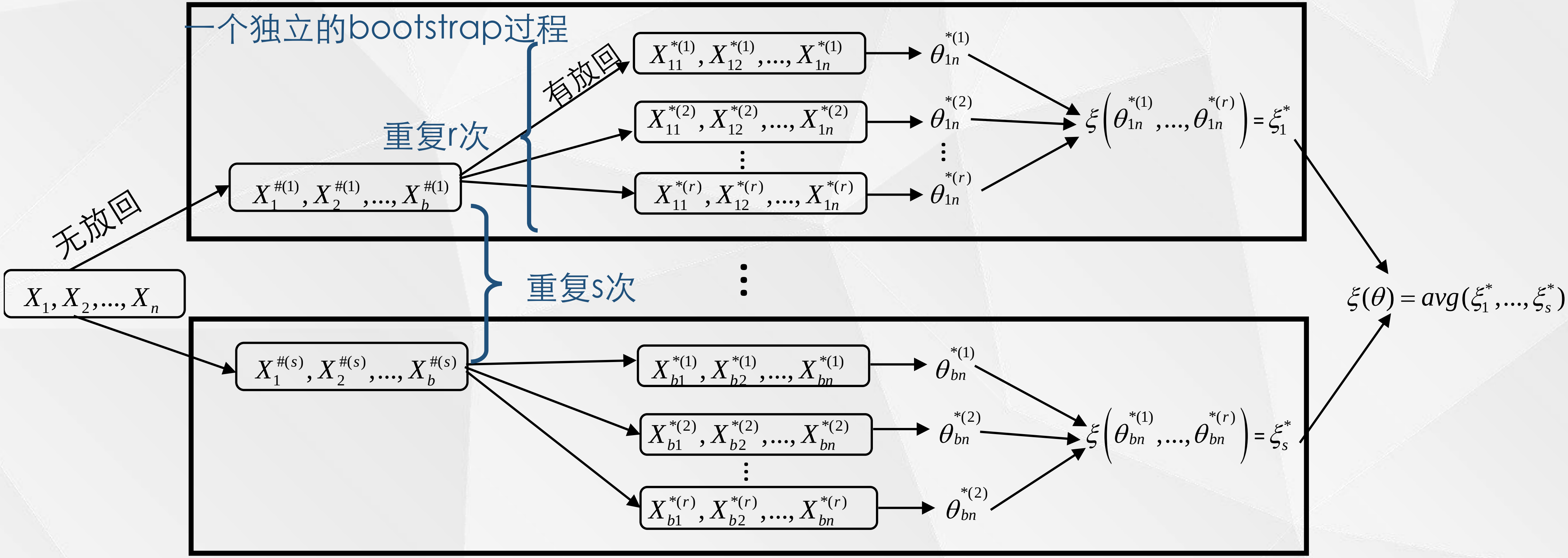
符号解释：

- X1，X2，……，Xn：原始数据
- $\hat{\theta}$：估计量
- b：子样本容量
- s：第一次重抽样的样本组数
- r：第二次重抽样的样本组数
- ξ：评价估计量质量的指标

  （方差、置信区间……）

02/ BLB 方法的实现、原理和程序

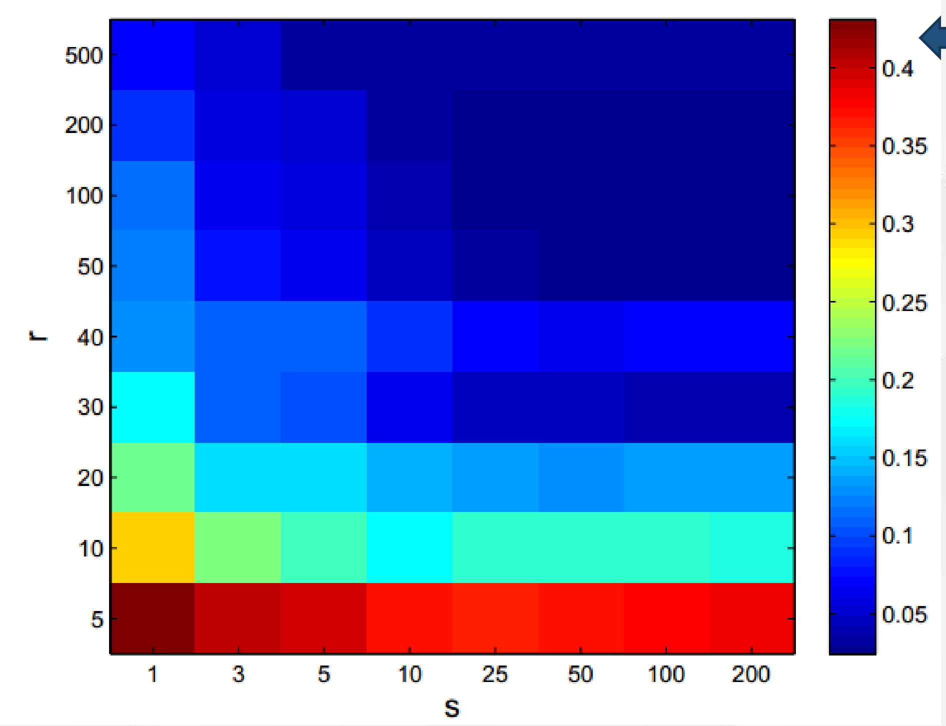
流程：



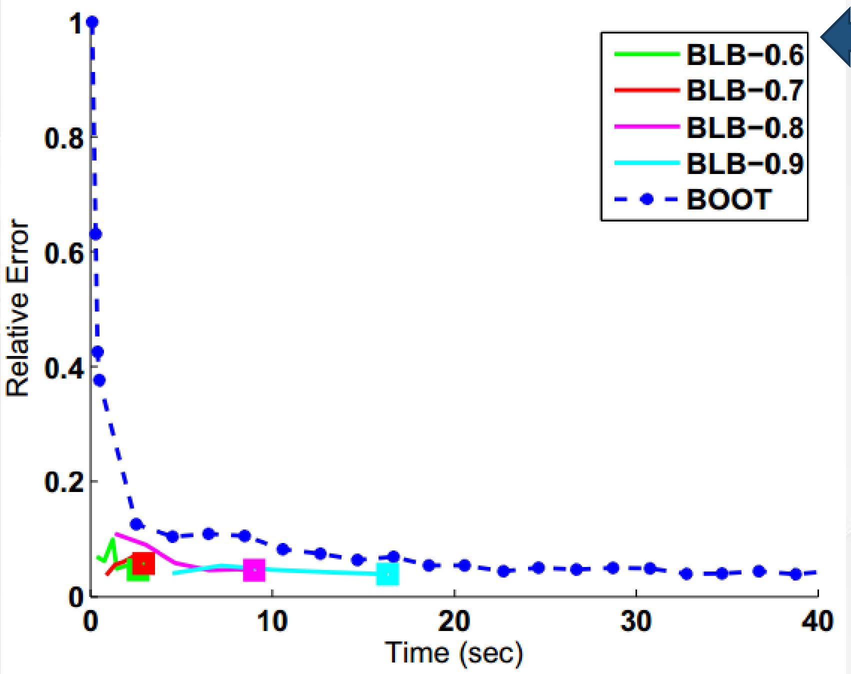
02 BLB 方法的实现、原理和程序

b、s、r的选择：

（b：第一次重抽样样本量　s：第一次重抽样的样本组数　r：第二次重抽样的样本组数）



相对误差



b=n^{γ}

根据需要的精度，选择合适的 r 、s 、b

02 BLB 方法的实现、原理和程序

BLB 中位数方差估计：

给定数据X=（$X_1, X_2, \cdots, X_n$）

for (i in 1 to s)

$X_b^{\#(i)}$=样本量为b，对X进行无放回简单随机抽样得到的样本；

for (j in 1 to r)

$X_{in}^{*(j)}$ =样本量为n，对 $X_b^{\#(i)}$进行有放回简单随机抽样得到的样本；

$M_{in}^{*(j)}$ = $X_{in}^{*(j)}$ 的中位数；

$M_{in}^{*(j)} = n^{-1}\sum_{k=1}^{n} M_{ik}^{*(j)}$

end

$M_{in}^{\#} = r^{-1}\sum_{j=1}^{r} M_{in}^{*(j)}$

end

$M = s^{-1}\sum_{i=1}^{s} M_{in}^{\#}$

02/ BLB 方法的实现、原理和程序

BLB 与传统 Bootstrap 的流程比较：

同：传统bootstrap相当于BLB方法流程中的一部分

BLB方法

抽取样本量为b的小样本 >> 从小样本中抽取样本量为n的大样本 >> 计算估计量 >> 计算方差 >> 计算方差均值

传统bootstrap方法

抽取样本量为n的大样本 >> 计算估计量 >> 计算方差

异：BLB从原始数据中抽取的是小样本

传统bootstrap抽取的是大样本

异：BLB计算出了s个方差取均值

传统bootstrap只计算了一个方差

02 BLB 方法的实现、原理和程序

BLB 方法的性质：

- 性质1：（一致性）

  n→ +∞ 时，$s^{-1}\sum_{i=1}^{s}\xi_i(\widehat{\theta})$ 依概率收敛于ξ（θ）

- 性质2：（高阶正确性）

  $\left|s^{-1}\sum_{i=1}^{s}\xi_i(\widehat{\theta})-\xi（\theta）\right|=O(\frac{1}{n})$，

即 BLB 方法和传统bootstrap有相同的高阶正确性

03

PART THREE

回归系数置信区间的Bootstrap

03 回归系数置信区间的估计——两个应用

Regression(线性回归系数估计)

Classification (Logistic分类回归系数估计)

03 Regression

$$Y = X\beta + \epsilon$$

X矩阵

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1d} \\ X_{21} & X_{22} & \cdots & X_{2d} \\ X_{31} & X_{32} & \cdots & X_{3d} \\ & & \vdots & \\ X_{n1} & X_{n2} & \cdots & X_{nd} \end{pmatrix} \quad \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix}$$

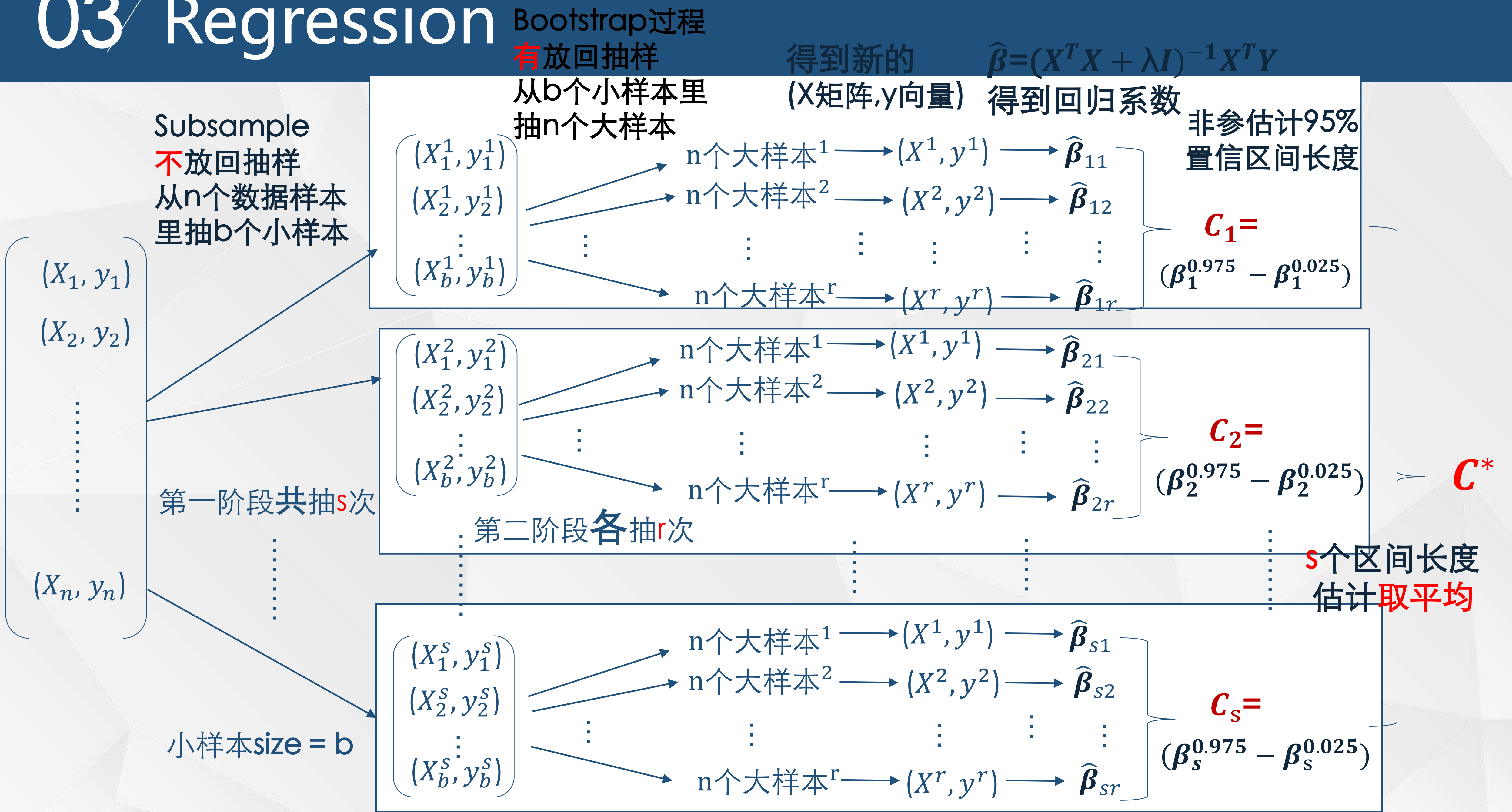
$\hat{\beta} = (X^T X)^{-1} X^T Y$

最小二乘法得到的回归系数的表达式

$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T Y$

为了让系数估计更稳定一些,论文里使用岭回归估计系数，其中$\lambda = 0.0001$。

03 Regression



03 Regression

- $n = 20000$ #总共生成20000个样本
- $d = 100$ #多元回归,维数为100
- $\boldsymbol{\beta} = (1,1,1,\dots,1)^T$ #长度为d=100的向量，元素取值均为1
- 生成模拟数据

For (i in 1:n) {

#从正态分布里面抽取（i.i.d.）

$\boldsymbol{X_i} = (X_{i1}, X_{i2}, \dots, X_{id})^T \sim \boldsymbol{Normal(0, I_d)}$ length($\boldsymbol{X_i}$)=d

#我们建的线性模型$Y_i = X_i\beta + \epsilon_i$；$\epsilon_i \sim Normal(0, 10)$

$Y_i / X_i \sim Normal(X_i\boldsymbol{\beta}, 10) = Normal(X_{i1} + X_{i2} + \cdots + X_{id}, 10)$

#得到($\boldsymbol{X_i}, Y_i$)的第i个样本

}#重复了n=20000次循环

03 Regression

- *如何衡量估计的好坏程度?*
- $\hat{\beta}=(X^TX+\lambda I)^{-1}X^TY$
- 我们知道X的分布（$\boldsymbol{X_i}=(X_{i1},X_{i2},\ldots,X_{id})^T \sim \boldsymbol{Normal(0,I_d)}$）
- 我们知道Y的分布（$\boldsymbol{Y_i/X_i} \sim \boldsymbol{Normal(}X_{i1}+X_{i2}+\cdots+X_{id}\boldsymbol{,10)}$）
- 我们知道λ的大小（λ=0.00001 实际中可以忽略不计）
- 那么我们就可以知道

$\hat{\beta}=(X^TX+\lambda I)^{-1}X^TY$的真实分布！

03 Regression

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T Y$$

C_0在论文里*用多次计算机*

模拟实验计算所得 $C_0 = 0.1$

已知$\hat{\beta}$真实分布

推导出$\hat{\beta}$**真实的**
95%置信区间长度C_0

BLB方法得到的
95%置信区间长度估计C^*

相对误差

$$= \frac{(C^* - C_0)}{C_0}$$

03 Regression

生成模拟数据

```
d = 100
n = 20000
set.seed(1247)
x <- matrix(NA,nrow = n, ncol = d)
y <- vector()
x <- matrix(rnorm(n*d,0,1),ncol = d)
for (i in 1:n) {
  y[i] <- rnorm(1,sum(x[i,]),sqrt(10))
}
```

03 Regression

```r
blb.ci <- function(alpha=0.05,x,y,s=20,bsize=0.7,r=60,lambda = 0.00001) {
  n = length(y)
  b = round(n^{bsize})
  ci.true <- 0.1
  ci <- matrix(NA,nrow = s,ncol = d)
  ci.length <- function (m) { #计算非参数置信区间长度的函数
    ci.length <-quantile(m,1-alpha/2) - quantile(m,alpha/2)
    return(ci.length[[1]])
  }
  for (i in 1:s)  {
    bsample <- sample(1:n,b,replace = FALSE) #subsample不放回做s次
    beta <- matrix(NA,ncol = d,nrow = r)
    for (j in 1:r) {
      nsample <- sample(bsample,n,replace = TRUE) #bootstrap sample放回 每个subsample做r次
      beta[j,] <- solve(t(x[nsample,])%*%x[nsample,] + lambda*diag(d))%*%t(x[nsample,])%*%y[nsample]
    }  #Ridge regression的beta估计
    ci[i,] <- apply(beta,2,ci.length)
  }
  ci.all <- colMeans(ci) #BLB置信区间
  relative.error <- abs((mean(ci.all) - ci.true))/ci.true #相对误差
  return(c(relative.error))
}
```

BLB估计置信区间长度函数

BLB估计置信区间长度函数

核心算法部分

```
for (i in 1:s)  {
  bsample <- sample(1:n,b,replace = FALSE) #subsample不放回做s次
  beta <- matrix(NA,ncol = d,nrow = r)
  for (j in 1:r) {
    nsample <- sample(bsample,n,replace = TRUE) #bootstrap sample放回 每个subsample做r次
    beta[j,] <- solve(t(x[nsample,])%*%x[nsample,] + lambda*diag(d))%*%t(x[nsample,])%*%y[nsample]
  }  #Ridge regression的beta估计
  ci[i,] <- apply(beta,2,ci.length)
}
ci.all <- colMeans(ci) #BLB置信区间
```

03 Regression

BLB

n of n Bootstrap

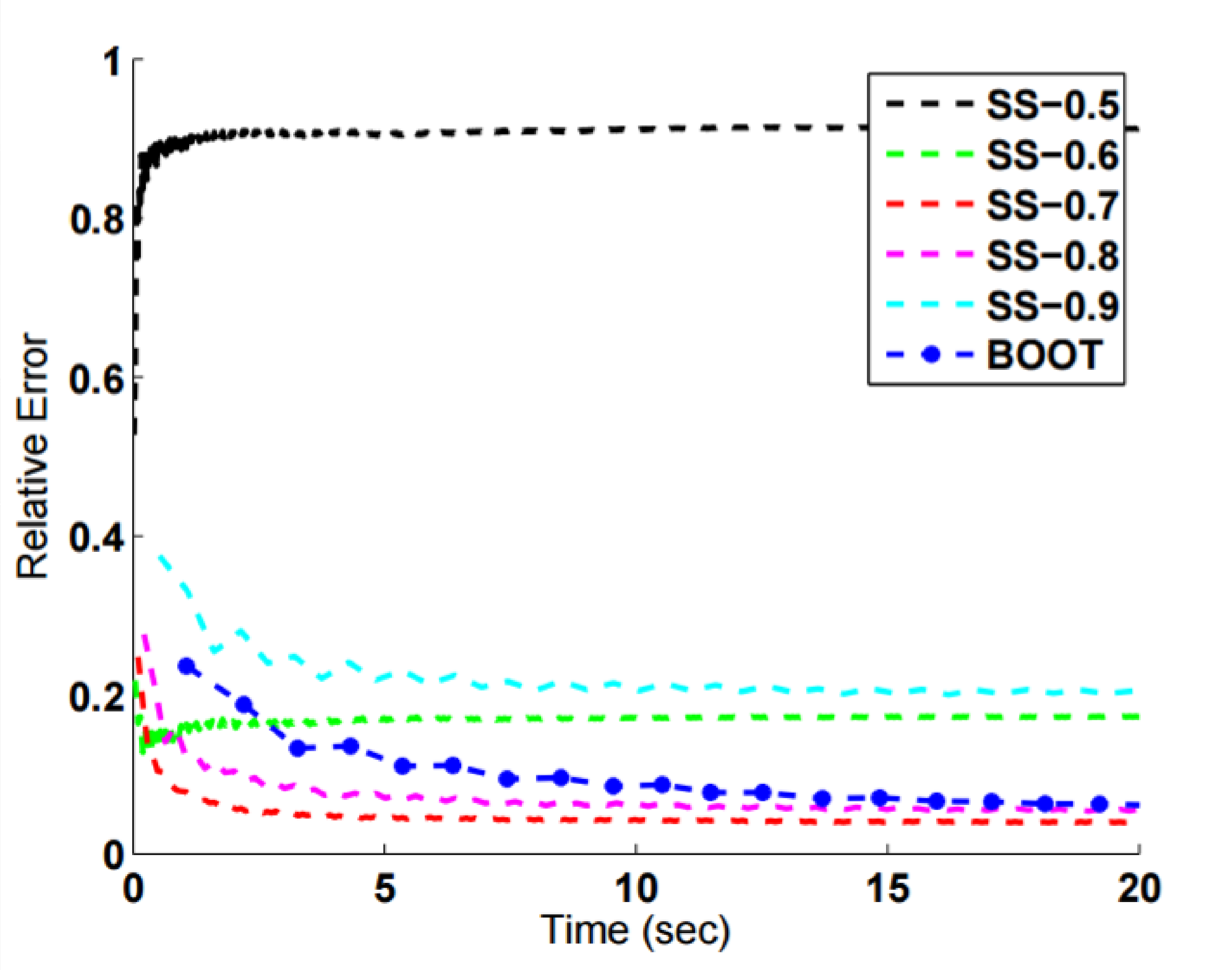
b of n Subsampling

b of n Bootstrap

到底哪个又快又准呢?

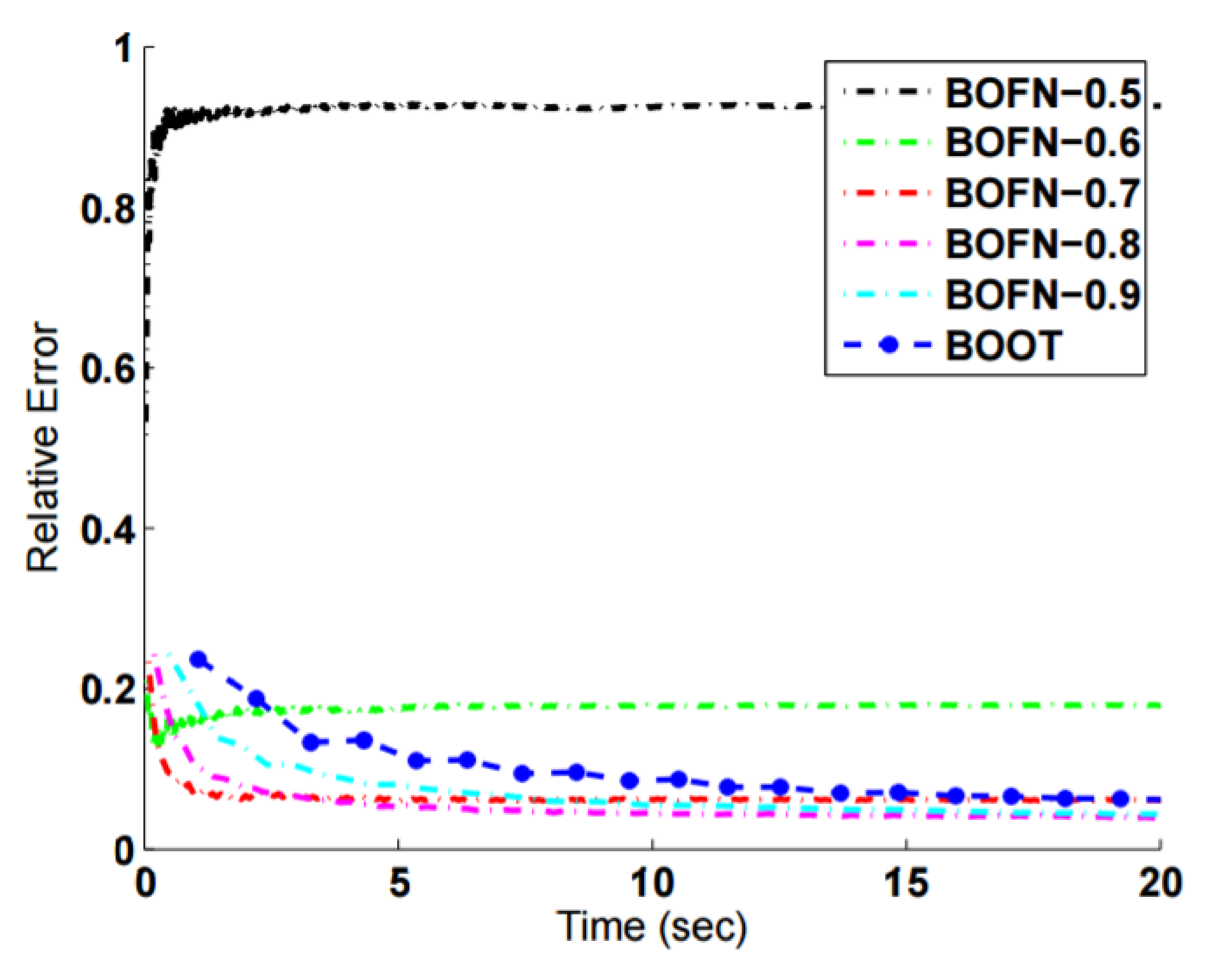
03 Regression

- subsampling的估计95%置信区间长度的相对误差
- $b = n^{\gamma}$
- $\gamma = 0.5, 0.6, 0.7, 0.8, 0.9$
- 它的估计的好坏很大程度上依据b的大小而决定
- $\gamma = 0.5$相对误差太大了，而且γ=0.6,0.9的时候比n of n bootstrap差很多）



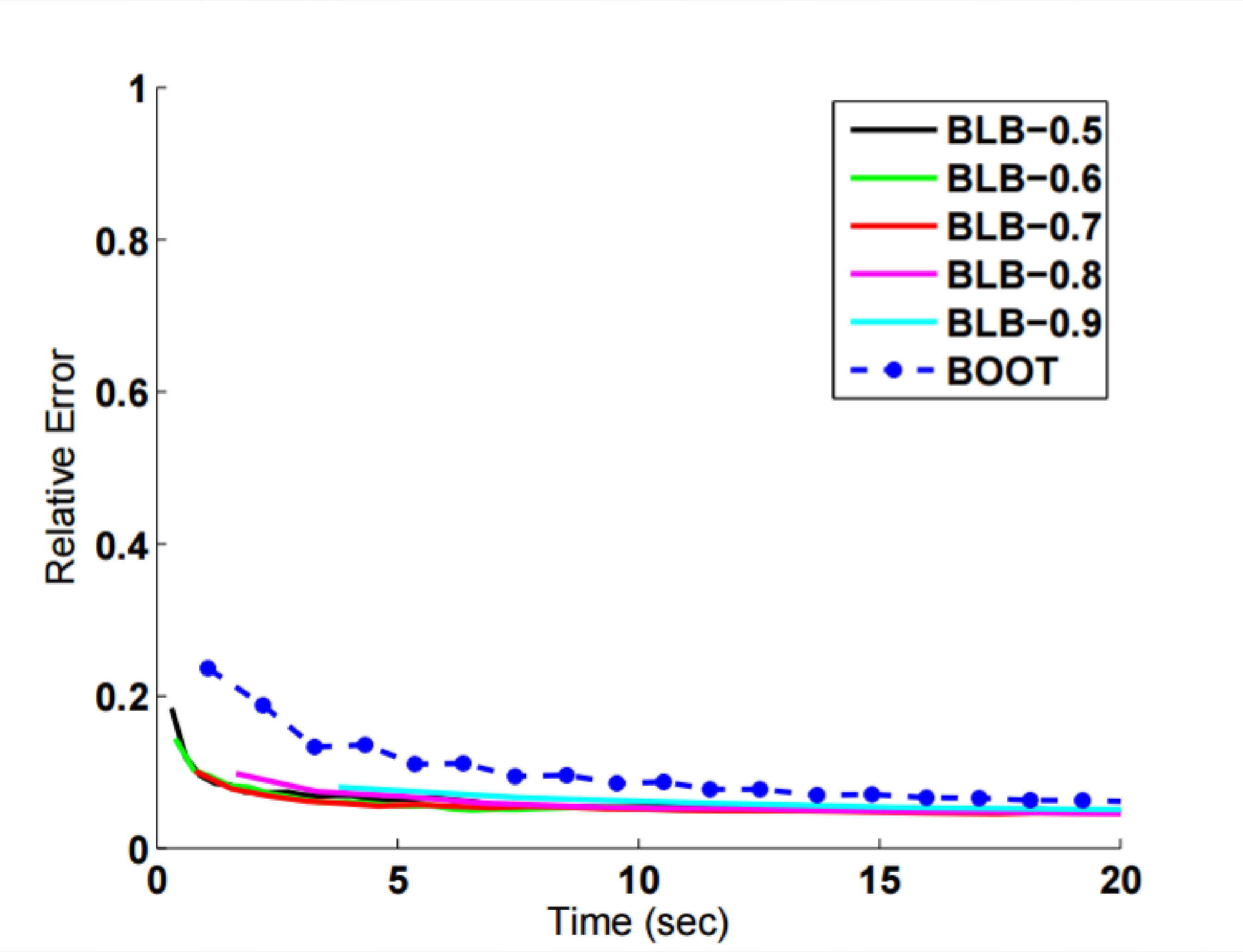
03 Regression

- b of n bootstrap的估计95%置信区间长度的相对误差
- $b = n^{\gamma}$
- $\gamma = 0.5, 0.6, 0.7, 0.8, 0.9$
- 它的估计的好坏很大程度上依据b的大小而决定（$\gamma = 0.5$ 的时候相对误差随着时间居然还递增）



03 Regression

- BLB估计95%置信区间长度的相对误差（随着运行时间的增加而减小）
- $b = n^{\gamma}$
- $\gamma = 0.5, 0.6, 0.7, 0.8, 0.9$
- 同样的运行时间下，BLB表现都比n of n Bootstrap更好，并且很稳定，对γ选择不敏感



03 Regression

- 在论文里还做了二次模型的模拟实验
- X与Y生成的模式和之前的一样，只是模型变成了关于X的二次回归方程

$$Y = X\beta + X^T X + \epsilon$$

- 论文里还包含了X与Y服从Gamma分布、t分布的模拟实验

02 Classification(Logistic Regression)

- $n = 20000$ #总共生成20000个样本
- $d = 100$ #多元回归,维数为100
- $\boldsymbol{\beta} = (1,1,1,\dots,1)^T$ #长度为d=100的向量，元素取值均为1
- 生成模拟数据

For (i in 1:n) {

从正态分布里面抽取（i.i.d.）

$\boldsymbol{X_i} = (X_{i1}, X_{i2}, \dots, X_{id})^T \sim \boldsymbol{Normal(0, I_d)}$ length($\boldsymbol{X_i}$)=d

#我们建的线性模型是 $\ln(\frac{p}{1-p}) = X_i\beta + \epsilon_i$；$\epsilon_i \sim Normal(0, 10)$

$$\mathbf{logit(p)} = \boldsymbol{ln}\left(\frac{p}{1-p}\right) / \boldsymbol{X_i} \sim \boldsymbol{Normal(X_i\beta, 10) = Normal(X_{i1} + X_{i2} + \cdots + X_{id}, 10)}$$

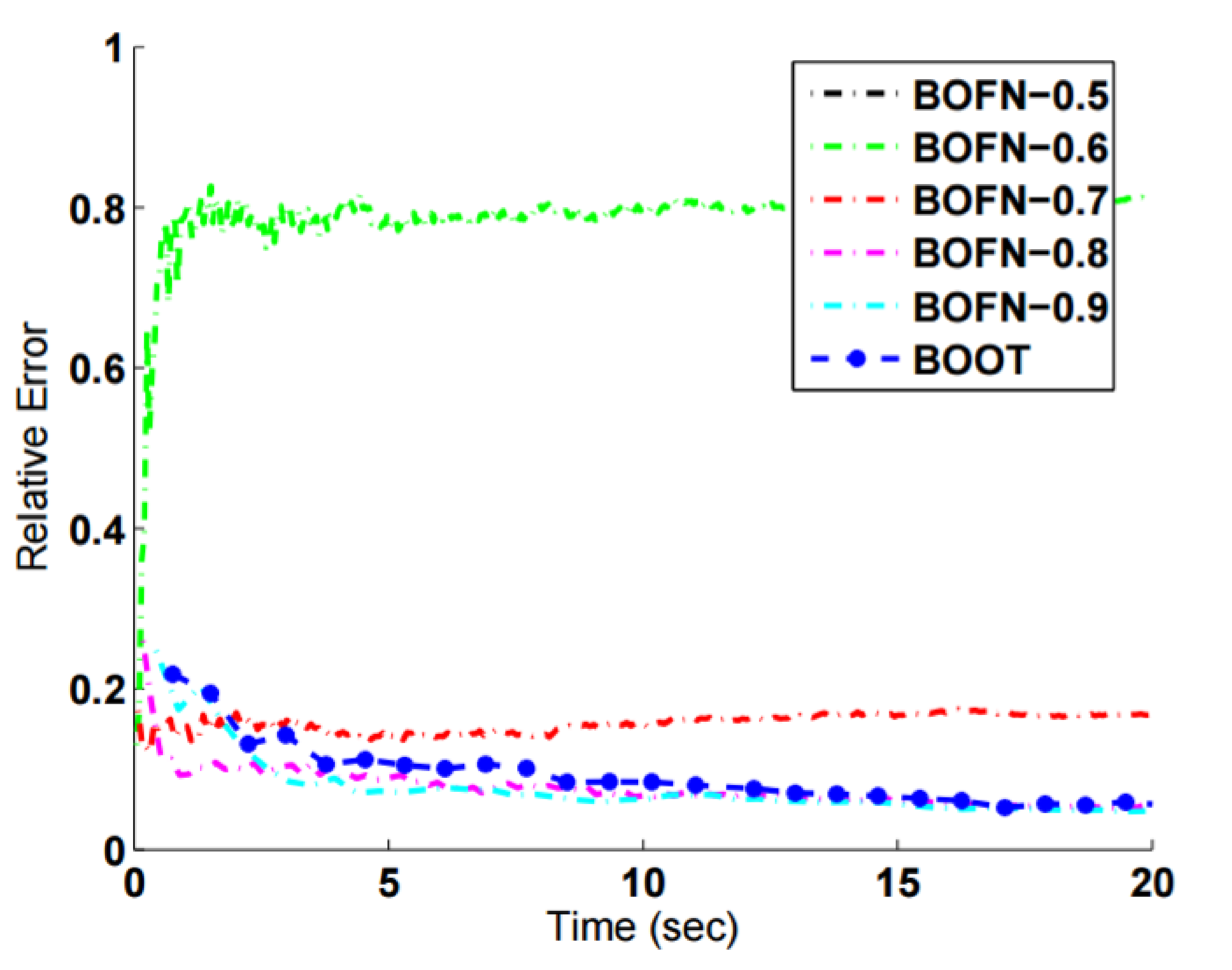
$$\boldsymbol{p} = \frac{e^{X_i\beta+\epsilon_i}}{e^{X_i\beta+\epsilon_i}+1} \in (0,1)$$

$\boldsymbol{Y_i} \sim Bernoulli(p)$ #得到($\boldsymbol{X_i}$, Y_i)的第i个样本

}#重复了n=20000次循环

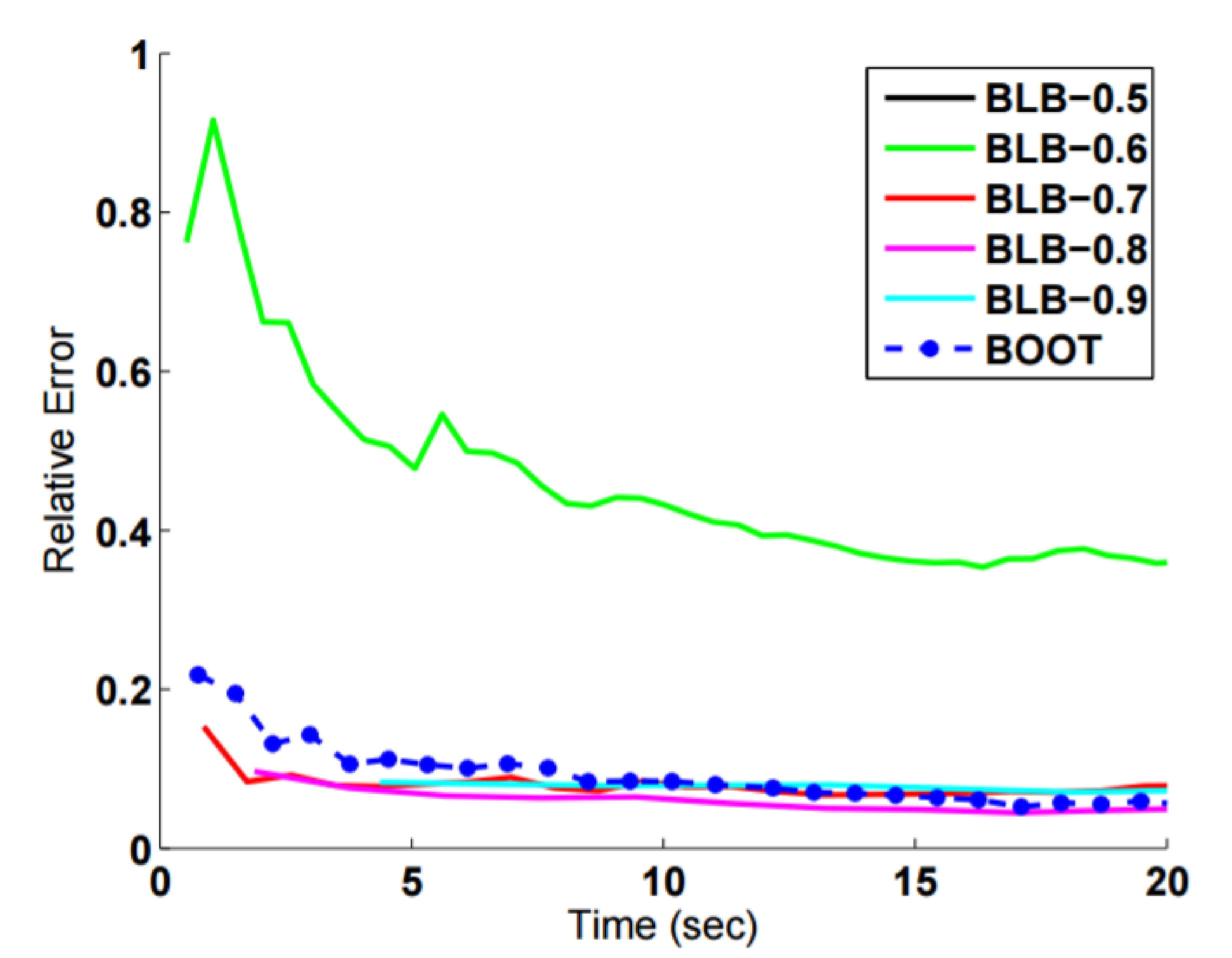
03 Classification

- b of n bootstrap的估计95%置信区间长度的相对误差
- $b = n^{\gamma}$
- $\gamma = 0.5, 0.6, 0.7, 0.8, 0.9$
- $\gamma = 0.6$的时候相对误差随着时间递增，$\gamma = 0.7$表现随时间增加比n of n bootstrap差



03 Classification

- BLB估计95%置信区间长度的相对误差（随着运行时间的增加而减小）
- $b = n^{\gamma}$
- γ = 0.5,0.6,0.7,0.8,0.9
- γ =0.6的时候同样表现不佳，但是至少随着时间增加，相对误差是在减小的（只是收敛速度很慢）
- BLB在γ >0.6时表现都比n of n Bootstrap更好些或者相持，并且很稳定。



04

PART FORE

枢轴量的置信区间

04 枢轴量的置信区间

针对之前所述的Bootstrap方法，用$\hat{\theta}$估计参数θ，置信水平为1-a，构造枢轴量的置信区间

参数方法 >> 枢轴量法 >> 实际应用

04 枢轴量的置信区间

参数方法：

- 用n个样本量计算得到枢轴量$\hat{\boldsymbol{\theta}}$
- BLB估计枢轴量方差$\boldsymbol{\sigma}_*^2$
    - 从样本量为n的样本中subsample抽取s个样本容量为b的小样本
    - 从每个小样本里bootstrap抽取r个容量为n的大样本
    - 利用每个大样本都计算得到枢轴量得到s组bootstrap结果

$$(\hat{\boldsymbol{\theta}}_{11}, \hat{\boldsymbol{\theta}}_{12}, \cdots, \hat{\boldsymbol{\theta}}_{1r})$$
$$(\hat{\boldsymbol{\theta}}_{21}, \hat{\boldsymbol{\theta}}_{22}, \cdots, \hat{\boldsymbol{\theta}}_{2r})$$
$$\cdots$$
$$(\hat{\boldsymbol{\theta}}_{s1}, \hat{\boldsymbol{\theta}}_{s2}, \cdots, \hat{\boldsymbol{\theta}}_{sr})$$

    - 对每组bootstrap结果算方差，得到$(\sigma_1^2, \sigma_2^2, \dots, \sigma_s^2)$
    - 对s个方差取平均得到方差的BLB的估计$\boldsymbol{\sigma}_*^2$

枢轴量置信区间为$(\hat{\boldsymbol{\theta}} - \boldsymbol{z}_{\alpha/2}\boldsymbol{\sigma}_*, \hat{\boldsymbol{\theta}} + \boldsymbol{z}_{\alpha/2}\boldsymbol{\sigma}_*)$

04 枢轴量置信区间

枢轴量法：

$\hat{\theta}^*_{1-\alpha/2}$ 为用 Bootstrap 方法得到样本参数的$1-\alpha/2$分位点，$\hat{\theta}^*_{\alpha/2}$ 为用 Bootstrap 方法得到样本参数的$\alpha/2$分位点

对θ的估计值$\hat{\theta}$的求解同上，但此时为非正态分布

（$\hat{\theta}$-θ）的置信区间为（$(\hat{\theta}-\theta)_{\alpha/2}$，$(\hat{\theta}-\theta)_{1-\alpha/2}$）

$\Rightarrow(-1)\times(\hat{\theta}-\theta)\Rightarrow(\theta-\hat{\theta})$的置信区间为（-$(\hat{\theta}-\theta)_{\alpha/2}$，-$(\hat{\theta}-\theta)_{1-\alpha/2}$）

$\Rightarrow\theta$置信区间（$\hat{\theta}$-$(\hat{\theta}-\theta)_{\alpha/2}$，$\hat{\theta}$-$(\hat{\theta}-\theta)_{1-\alpha/2}$）

04 枢轴量的置信区间

非参数方法：

其中$(\hat{\theta}-\theta)_{\alpha/2}$与$(\hat{\theta}-\theta)_{1-\alpha/2}$未知，用（$\hat{\theta}^*_{1-\alpha/2}$-$\hat{\theta}$）与（$\hat{\theta}^*_{\alpha/2}-\hat{\theta}$）代替

则θ置信区间（$2\hat{\theta}$-$\hat{\theta}^*_{1-\alpha/2}$，$2\hat{\theta}$-$\hat{\theta}^*_{\alpha/2}$）

将$\hat{\theta}$，$\hat{\theta}^*_{1-\alpha/2}$，$\hat{\theta}^*_{\alpha/2}$带入求解即可得$\theta$的置信区间

04 枢轴量的置信区间

枢轴量法：

- 用n个样本量计算得到枢轴量$\hat{\theta}$
- BLB估计$(\hat{\theta}-\theta)$的$({}^{\alpha}/_{2})$分位数$(\hat{\theta}-\theta)_{\alpha/2}$以及$(1-{}^{\alpha}/_{2})$分位数$(\hat{\theta}-\theta)_{1-\alpha/2}$
  - 从样本量为n的样本中subsample抽取s个样本容量为b的小样本
  - 从每个小样本里bootstrap抽取r个容量为n的大样本
  - 利用每个大样本都计算得到枢轴量得到s组bootstrap结果

$$
\begin{gathered}
(\hat{\theta}_{11},\hat{\theta}_{12},\cdots,\hat{\theta}_{1r})\\
(\hat{\theta}_{21},\hat{\theta}_{22},\cdots,\hat{\theta}_{2r})\\
\cdots\\
(\hat{\theta}_{s1},\hat{\theta}_{s2},\cdots,\hat{\theta}_{sr})
\end{gathered}
$$

  - 对每组bootstrap结果算$({}^{\alpha}/_{2})$分位数、$(1-{}^{\alpha}/_{2})$分位数
  - 对s对（$({}^{\alpha}/_{2})$分位数、$(1-{}^{\alpha}/_{2})$分位数）各自取平均得到（$\hat{\theta}_{*\alpha/2},\hat{\theta}_{*1-\alpha/2}$）

枢轴量置信区间为$\left(2\hat{\theta}-\hat{\theta}_{*1-\alpha/2},\widehat{2\theta}-\hat{\theta}_{*\alpha/2}\right)$

04 枢轴量的置信区间

实际应用

均值

其中“利用每个样本对参数θ进行估计，得到$\hat{\theta}_{11}$，$\hat{\theta}_{12}$，……$\hat{\theta}_{1m_{(2)}}$，……，$\hat{\theta}_{m_{(1)}1}$,$\hat{\theta}_{m_{(1)}2}$,……，$\hat{\theta}_{m_{(1)}m_{(2)}}$”中的$\hat{\theta}_{ij}=\frac{1}{n}\sum_{k=1}^{n}X_k$，之后依步骤求解即可。

方差

其中“利用每个样本对参数θ进行估计，得到$\hat{\theta}_{11}$，$\hat{\theta}_{12}$，……$\hat{\theta}_{1m_{(2)}}$，……，$\hat{\theta}_{m_{(1)}1}$,$\hat{\theta}_{m_{(1)}2}$,……，$\hat{\theta}_{m_{(1)}m_{(2)}}$”中的$\hat{\theta}_{ij}=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X})^2$，之后依步骤求解即可。

05 参考文献

- Efron, B. "Bootstrap methods: another look at the jackknife Annals of Statistics 7: 1–26." *View Article PubMed/NCBI Google Scholar* (1979).
- Bickel, Peter J., and David A. Freedman. "Some asymptotic theory for the bootstrap." The Annals of Statistics (1981): 1196-1217.
- Giné, Evarist, and Joel Zinn. "Bootstrapping general empirical measures." The Annals of Probability (1990): 851-869.
- Van Der Vaart, Aad W., and Jon A. Wellner. "Weak Convergence." Weak Convergence and Empirical Processes. Springer New York, 1996. 16-28.
- Rasmussen, Jeffrey L. "" Bootstrap confidence intervals: Good or bad": Comments on Efron (1988) and Strube (1988) and further evaluation." (1988): 297.
- Bickel, P., and J. Ren. "On choice of m for the m out of n bootstrap in hypothesis testing." Preprint, Department of Statistics, University of California, Berkeley (1997).
- Politis, Dimitris, Joseph P. Romano, and Michael Wolf. "Weak convergence of dependent empirical measures with application to subsampling in function spaces." Journal of statistical planning and inference 79.2 (1999): 179-190.
- Bickel, Peter J., and Anat Sakov. "On the choice of m in the m out of n bootstrap and confidence bounds for extrema." Statistica Sinica (2008): 967-985.
- Bickel, Peter J., and Anat Sakov. "Equality of types for the distribution of the maximum for two values of n implies extreme value type." Extremes 5.1 (2002): 45-53.
- Bickel, Peter J., and Joseph A. Yahav. "Richardson extrapolation and the bootstrap." Journal of the American Statistical Association 83.402 (1988): 387-393.
- Kleiner, Ariel, et al. "A scalable bootstrap for massive data." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76.4 (2014): 795-816.