

Stat 471: Lecture 16

Bootstrap II.

Moo K. Chung
mchung@stat.wisc.edu

October 15, 2003

1. Based on our repeated sampling,

```
boot=inline('x(unidrnd( length(x),
m, length(x)))', 'x', 'm');
```

let us estimate

$$\mathbb{E} \hat{\theta}(\mathbf{X}) \approx \bar{\theta}^* = \sum_{i=1}^m \hat{\theta}(\mathbf{X}^{*i})/m$$

$$\text{Var} \hat{\theta}(\mathbf{X}) \approx \sum_{i=1}^m \left(\hat{\theta}(\mathbf{X}^{*i}) - \bar{\theta}^* \right)^2 / (m-1)$$

from `strength.data` based on 200 bootstrap replications. For population mean, we use $\hat{\theta}(\mathbf{X}) = \bar{X}$. For population variance, we use $\hat{\theta}(\mathbf{X}) = \sqrt{n}\bar{X}$.

```
bs=boot(arm,200); %200x147 matrix
bsmean=mean(bs,2)
```

```
>>[mean(arm) mean(bsmean)]
```

```
78.7517 78.9025
```

```
%MATLAB built-in function
```

```
>>mean(bootstrp(200,'mean',arm))
```

```
78.7799
```

```
%For matrices, var(X) is a row
%vector containing the variance
%of each column.
```

```
>>[var(arm) 147*var(bsmean)]
```

```
445.6040 423.0257
```

2. $\text{Var} \hat{\theta}(\mathbf{X})$ measures the performance of a bootstrap estimate. To measure the accuracy of an estimate, we compute the *bias*, which is defined as the difference between the expectation of the estimator and the true parameter,

$$\text{bias}(\hat{\theta}) = \mathbb{E} \hat{\theta} - \theta.$$

The bootstrap estimation of the bias is given by

$$\text{bias}^*(\hat{\theta}) = \bar{\theta}^* - \hat{\theta}.$$

Read textbook or Efron (1982) for further detail.

For our example $\hat{\theta} = \bar{X}$,

$$\text{bias}^*(\bar{X}) = \bar{X}^* - \bar{X}.$$

```
>>78.9025-78.7517
0.1508
```

3. Suppose $X \sim \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$. We want to find $100(1-\alpha)\%$ confidence interval (CI) for μ . It depends on finding a pivot of μ . A pivot is a statistic that contains μ but whose distribution does not depend on μ or σ . In the case $f \sim N(\mu, \sigma^2)$, $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ is a pivot for μ . Let $t_{\alpha/2}$ and $t_{1-\alpha/2}$ be $\alpha/2$ and $1-\alpha/2$ quantiles for t_{n-1} distribution. Then the 95% CI is

$$\left(\bar{x} - t_{1-\alpha/2} S / \sqrt{n}, \bar{x} - t_{\alpha/2} S / \sqrt{n} \right).$$

If f is not normal, T will be still a pivot but we do not know its distribution. So we generate 1000 bootstrap repliates of $T^* = \frac{\bar{X}^* - \bar{X}}{S^* / \sqrt{n}}$ of T and get the bootstrap CI

$$\left(\bar{x} - t_{1-\alpha/2}^* S / \sqrt{n}, \bar{x} - t_{\alpha/2}^* S / \sqrt{n} \right).$$

```
t=inline('(mean(x)-u)/(std(x)/
length(x))', 'x', 'u')
```

```
for i=1:1000;
```

```
    trep(i)=t(boot(arm,1),mean(arm))
```

```
end;
```

```
>> quantile(trep,[0.05 0.95])
```

```
-19.8246 22.0677
```

```
>> mean(arm) -22.0677*std(arm)/sqrt(147)
40.3303
```

```
>> mean(arm) +19.8246*std(arm)/sqrt(147)
113.2677
```