Stat 2911 Lecture Notes

Class 17, 2016

Uri Keich

© Uri Keich, The University of Sydney

Parametric bootstrap (Rice 8.4, 8.9), Conditional expectation (Rice 4.4)

Parametric Bootstrap

Set up: given a sample $x_1, \dots, x_n$ from a dist. $F(\underline{\theta})$ we estimate $\underline{\theta}$ using $\hat{\underline{\theta}}(x_1, \dots, x_n)$.

Recipe for estimating the bias/var/MSE of $\hat{\underline{\Theta}}$:

- Draw N ind. bootstrap samples of size n from $F(\hat{\underline{\theta}}(x_1, \dots, x_n))$:

$$\underline{x}^{*i} = (x_1^{*i}, \dots, x_n^{*i}) \qquad i = 1, \dots, N$$

- For each, compute $\hat{\underline{\theta}}^{*i} = \hat{\underline{\theta}}(\underline{x}^{*i})$

- Empirically estimate the bias/var/MSE from the sample $\hat{\underline{\theta}}^{*i}$ $i = 1, 2, \dots, N$.

For example,

$$\widehat{\text{bias}}(\hat{\underline{\Theta}}) = \frac{1}{N} \sum_1^N \hat{\underline{\theta}}^{*i} - \hat{\underline{\theta}}(x_1, \dots, x_n),$$

$$\widehat{MSE}(\hat{\underline{\Theta}}_k) = \sum_1^N \left[\hat{\underline{\theta}}_k^{*i} - \hat{\underline{\theta}}_k(x_1, \dots, x_n) \right]^2 \cdot \frac{1}{N} .$$

© Uri Keich, The University of Sydney

Example. Hardy-Weinberg equilibrium:

Model: $\underline{\underline{X}} = (X_1, X_2, X_3) \sim \text{multinomial}\left(n; \overset{AA}{(1-\theta)^2}, \overset{Aa}{2\theta(1-\theta)}, \overset{aa}{\theta^2}\right)$

Data: $n = 1029$ $\quad x_1 = 342 \quad x_2 = 500 \quad x_3 = 187$

Goal: estimate θ, the frequency of the minor allele.

Moment estimator: $\tilde{\Theta} = \sqrt{X_3/n} \underset{\text{case}}{\overset{\text{our}}{=}} \sqrt{187/1029} \simeq 0.4263$

MLE (also moment): $\hat{\Theta} = \dfrac{X_2 + 2X_3}{2n} \underset{\text{case}}{\overset{\text{our}}{\simeq}} 0.4247$

We found: $E_\theta(\hat{\Theta}) = \theta$

$$E(\tilde{\Theta}) = \sum_{k=0}^{n} \sqrt{k/n} \binom{n}{k} (\theta^2)^k (1-\theta^2)^{n-k}$$

$$MSE(\hat{\Theta}_n) = \frac{1}{2n} \theta(1-\theta)$$

$$MSE(\tilde{\Theta}) = 2\theta\left[\theta - E(\tilde{\Theta})\right]$$

We can estimate, say, the bias of $\tilde{\Theta}$ by:

$$\widehat{\text{bias}(\tilde{\Theta})} = \sum_{k=0}^{n} \sqrt{k/n} \binom{n}{k} (\tilde{\theta}^2)^k (1-\tilde{\theta}^2)^{n-k} - \tilde{\theta},$$

or, we can use the δ-method

and similarly, $\quad \widehat{MSE(\hat{\Theta})} = \frac{1}{2n} \hat{\theta}(1-\hat{\theta})$.

But what if we couldn't derive the expressions above?

We can provide estimates using parametric bootstrap:

- Generate bootstrap samples $\underline{x}^{*1}, \underline{x}^{*2}, \dots, \underline{x}^{*N}$, where each $\underline{x}^{*i}$ is sampled using the parameter estimated from the original sample $\underline{x} = (342, 500, 187)$.

That is,

$$\underline{x}^{*i} \sim \text{multinomial}\left[n=1029;\ \underline{p}=\left((1-\tilde{\theta})^2, 2\tilde{\theta}(1-\tilde{\theta}), \tilde{\theta}^2\right)\right],$$

where $\tilde{\theta} = \sqrt{187/1029}$ (or, use $\hat{\theta} = \frac{500+2\cdot 187}{1029}$)

- For each i, compute $\tilde{\theta}^{*i} = \sqrt{x_3^{*i}/n}$

$$\hat{\theta}^{*i} = \frac{x_2^{*i} + 2x_3^{*i}}{n}$$

You can now estimate, say, the bias of $\tilde{\theta}$ by

$$\widehat{E(\tilde{\theta}) - \theta} = \frac{1}{N}\sum_{i=1}^{N} \tilde{\theta}^{*i} - \tilde{\theta}$$

or $\hat{\theta}$ if $\hat{\theta}$ was used to generate the bootstrap samples

Is this estimate very sensitive to small variations in the estimated θ?

© Uri Keich, The University of Sydney

17.4

Parametric bootstrap for Hardy-Weinberg

Uri Keich

April 12, 2016

If the gene frequencies are in equilibrium, the genotypes AA, Aa, aa occur with probabilities $(1-\theta)^2, 2\theta(1-\theta), \theta^2$. The maximum likelihood estimate of θ is

$$\hat{\theta} = \frac{2x_3 + x_2}{2n},$$

where $n = x_1 + x_2 + x_3$. The moment estimator is

$$\tilde{\theta} = \sqrt{\frac{x_3}{n}}$$

The observerved values in the Hong Kong study were 342,500,187. First we find the values of the estimators for our particular sample.

```
> x = c(342,500,187)
> n = sum(x)
> hatth = (2*x[3]+x[2])/(2*n)
> hatth
```

$\hat{\theta}$

```
[1] 0.4246842
```

```
> tilth = sqrt(x[3]/n)
> tilth
```

$\tilde{\theta}$

```
[1] 0.4262978
```

Next we obtain $N = 10,000$ bootstrap samples from a multinomial distribution with $n = 1029$ and probabilities $(1-\hat{\theta})^2, 2\hat{\theta}(1-\hat{\theta}), \hat{\theta}^2$. Then, for each sample we obtain bootstrap estimates $\hat{\theta}^*$ and $\tilde{\theta}^*$

```
> set.seed(17)
> N = 10000
> xst = rmultinom(N, n, c((1-hatth)^2,2*hatth*(1-hatth),hatth^2))
> thsthat = (2*xst[3,]+xst[2,]) / (2*n)
> thsttil = sqrt(xst[3,]/n)
```

} bootstrap estimates

generating the bootstrap samples using $\hat{\theta}$ first.

17.5

We next estimate the mean, variance, bias and MSE of $\hat{\Theta}$ and $\tilde{\Theta}$ from the bootstrap samples:

```
> mean(thsthat)

[1] 0.4246392

> var(thsthat)

[1] 0.0001201425

> bootbias = mean(thsthat) - hatth
> bootbias

[1] -4.494655e-05

> mean((thsthat-hatth)^2)

[1] 0.0001201326

> mean(thsttil)

[1] 0.424283

> var(thsttil)

[1] 0.0002021445

> bootbias = mean(thsttil) - hatth
> bootbias

[1] -0.0004011775

> mean((thsttil-hatth)^2)

[1] 0.0002022852
```

It is also instructive to view the histograms of the boostrapped estimators:

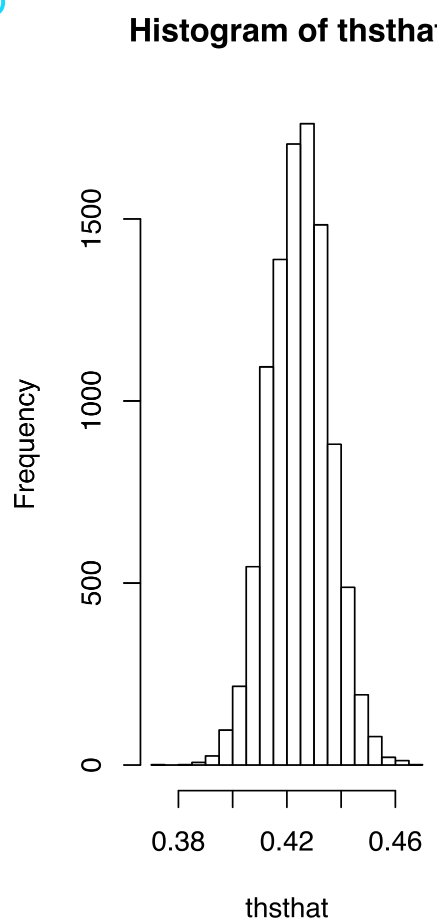
```
> par(mfrow=c(1,2))
> hist(thsthat)
> boxplot(thsthat,thsttil)
```

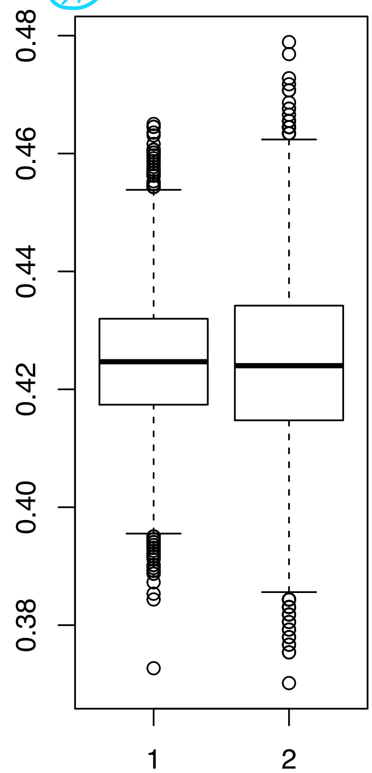
© Uri Keich, The University of Sydney

17.6

does the shape of this histogram look familiar?

which of these boxplots is of $\tilde{\Theta}$ and which is of $\hat{\Theta}$?





Finally, we repeat the bootstrap procedure only now we plug in $\tilde{\theta}$ instead of $\hat{\theta}$.

```
> set.seed(17)
> xst = rmultinom(N, n, c((1-tilth)^2, 2*tilth*(1-tilth),tilth^2))
> thsthat = (2*xst[3,]+xst[2,]) / (2*n)
> thsttil = sqrt(xst[3,]/n)
> mean(thsthat)

[1] 0.4261767

> var(thsthat)

[1] 0.0001189849

> mean(thsthat) - tilth

[1] -0.0001211477
```

17.7

```
> mean((thsthat-tilth)^2)
[1] 0.0001189877
> mean(thsttil)
[1] 0.4258385
> var(thsttil)
[1] 0.0002012665
> mean(thsttil) - tilth
[1] -0.0004593526
> mean((thsttil-tilth)^2)
[1] 0.0002014574
```

good agreement between the results sampled using $\hat{\theta}(\underline{x})$ and $\check{\theta}(\underline{x})$ is good news.

© Uri Keich, The University of Sydney

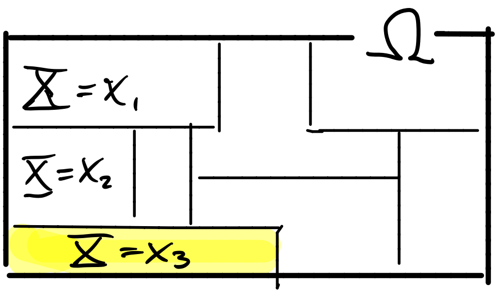
Conditional Expectation

If X and Y are jointly distributed RVs, then $\forall x$ with $P(X=x)>0$, we define the **conditional expectation of Y given $X=x$** as

$$E(Y|X=x) = \sum_y{}' \, y \, P(Y=y|X=x)$$

provided $\sum_y{}' |y| P(Y=y|X=x) < \infty$ (*)

Note that if (*) holds then $E(Y|X=x) \in \mathbb{R}$. (why?)



$E(Y|X=x_3)$ = the weighted average of Y over $\{\omega : X(\omega) = x_3\}$

Example. Roll a fair die and let Y = outcome

$X = 1_{\{\text{even outcome}\}}$

Ω:

1	3	5
2	4	6

$X =$

$X =$

$E(Y|X=0) =$

$E(Y|X=1) =$

How can we guarantee that condition (*) holds?

Claim. If $Y \in L^1$ then $E(Y \mid X = x)$ is defined and $\in \mathbb{R}$ $\forall x$ with $P(X = x) > 0$.

Proof. We need to verify that $(*)$ holds:

$$\sum_y{}' |y| \, P(Y = y \mid X = x) = \sum_y{}' |y| \frac{P(Y = y, X = x)}{P(X = x)}$$

why?

$$\le \frac{1}{P(X = x)} \sum_y{}' |y| \cdot P(Y = y) < \infty$$

$\square$

Random Sums

Let X_i be iid RVs and let N be an ind. $\mathbb{N}$-valued RV. Define the random sum $T = \sum_1^N{}' X_i$.

The RV N determines the number of summands, or elements X_i, that are summed.
If $N = 0$ then $T = 0$.

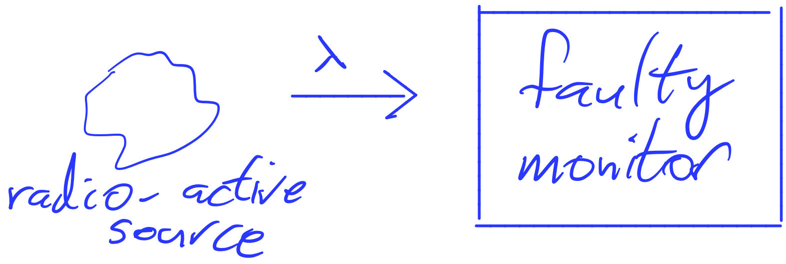
Examples.

1) $X_i \equiv 1$ and $N \sim \text{Poisson}(\lambda) \Rightarrow T = \sum_1^N 1 = N$

© Uri Keich, The University of Sydney

2) $X_i \sim \text{Bernoulli}(p)$ and $N \sim \text{Poisson}(\lambda)$

$$T = \sum_{i=1}^{N} X_i$$



The monitor registers an incoming particle with prob. p independently of anything else.

Recall: $T \sim \text{Poisson}(\lambda p)$

$$\Rightarrow E(T) = \lambda p$$

Our goal is to be able to compute $E(T)$ directly i.e., without the need to first find the dist. of T.

Expectation is linear:

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n\,E(X_i)$$

$$E(T) = E\left(\sum_{i=1}^{N} X_i\right) \neq N\,E(X_i) \quad (\text{why?})$$

But we will show $E(T) = E(N)\,E(X_i)$.

How?

Recall the total probability law:

$$P(T=s) = \sum_{n} P(T=s \mid N=n)\,P(N=n).$$

We will establish an analogous "total expectation law":

$$E(T) = \sum_{n} E(T \mid N=n)\,P(N=n).$$

Again, X_i are iid RVs and N is an independent $\mathbb{N}$-valued RV. Let $T = \sum_1^N X_i$ and let $S_n = \sum_1^n X_i$, so $T = S_N$.

Finally, let $\{s_k\} = T(\Omega)$, then for n with $P(N=n) > 0$ we have:

$$P(T = s_k \mid N=n) = \frac{P(T=s_k, N=n)}{P(N=n)}$$

$$= \frac{P(\sum_1^N X_i = s_k, N=n)}{P(N=n)}$$

$$= \frac{P(\sum_1^n X_i = s_k, N=n)}{P(N=n)}$$

(exercise: $\sum_1^n X_i$ is ind. of N)

$$= \frac{P(\sum_1^n X_i = s_k)\, P(N=n)}{P(N=n)}$$

$$= P(S_n = s_k)$$

Claim. If $P(N=n) > 0$ then the dist. of T given that $N=n$ is the same as the (unconditional) dist. of $S_n = \sum_1^n X_i$, and in particular $S_n(\Omega) \subset T(\Omega)$.

Proof. $P(T = s_k \mid N=n) = P(S_n = s_k)$

© Uri Keich, The University of Sydney

Cor. If $X_i \in L^1$ then $E(T \mid N=n) = n\,E(X_i)$ and, in particular, it is well defined and in $\mathbb{R}$.

Proof.

$$\sum_k |s_k| \, P(T=s_k \mid N=n) = \sum_k |s_k| \, P(S_n = s_k)$$

$$\overset{\text{why?}}{=} E(|S_n|) < \infty \;\text{why?}$$

tricky!

$\Rightarrow E(T \mid N=n)$ is well defined. Moreover,

$$\sum_k s_k \, P(T=s_k \mid N=n) = \sum_k s_k \, P(S_n = s_k)$$

$$\overset{?}{=} E\left(\sum_{i=1}^{n} X_i\right)$$

$$= \sum_{i=1}^{n} E(X_i)$$

$$= n\,E(X_i).$$

up to here we didn't need X_i to be identically distributed

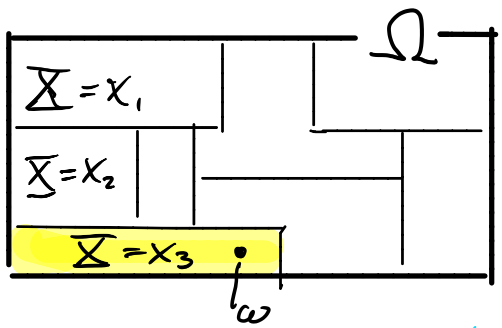
Do we need X_i to be ind. of one another?

Conditional Expectation (the RV)

Recall: If $Y \in L^1$ then $E(Y | X = x)$ is defined $\forall x$ with $P(X = x) > 0$.

<u>Def</u>. For $Y \in L^1$ we define the <u>RV</u> conditional expectation of Y given X as

$$\underbrace{E(Y | X)}_{\text{notation}}(\omega) = E[Y | X = X(\omega)]$$



Suppose $\omega \in \Omega$ satisfies $X(\omega) = x_3$.
Then $E(Y|X)(\omega) = E(Y | X = x_3)$

Note that $E(Y|X)$ is const. on the event $X = x_k$, i.e., $\forall \omega \in \{X = x_k\}$ or $X(\omega) = x_k$, $E(Y|X)(\omega) = E(Y | X = x_k)$.
That is, both X and $E(Y|X)$ are constant on the event $\{X = x_k\}$, the constant being the average of X or Y on that event.

© Uri Keich, The University of Sydney