resampling methods part II

The non-parametric bootstrap is your bestest friend ever!!!!

October 21st 2014

Readings for Thursday and next week (3 lectures): Likelihood

- See syllabus
- Chapter 6 of Bolker is the primary reading. It is a dense and substantial chapter so take your time, and review it!

Goals for today

- Continue with resampling…
- Further develop the np bootstrap
- Using the bootstrap for testing hypotheses.

- Random (pairs) VS fixed (residual) bootstraps.

(non-parametric) Bootstrap

- Sampling **WITH** replacement
- Assumption: Exchangeability of observations within treatments (*independence*).
- No assumptions that different treatments share the same distributional form (i.e. no need to worry about common variance).
- Ho $\mu_a = \mu_b$

Bootstrap

- Bootstrap generates a sampling distribution for quantities of interest (mean, SD, CI).
- Generating random samples from the observed sample of interest.
- Estimates how quantity of interest varies due to random sampling.
- useful for examining *bias* (difference between estimated means for population and sample).

Sampling with replacement

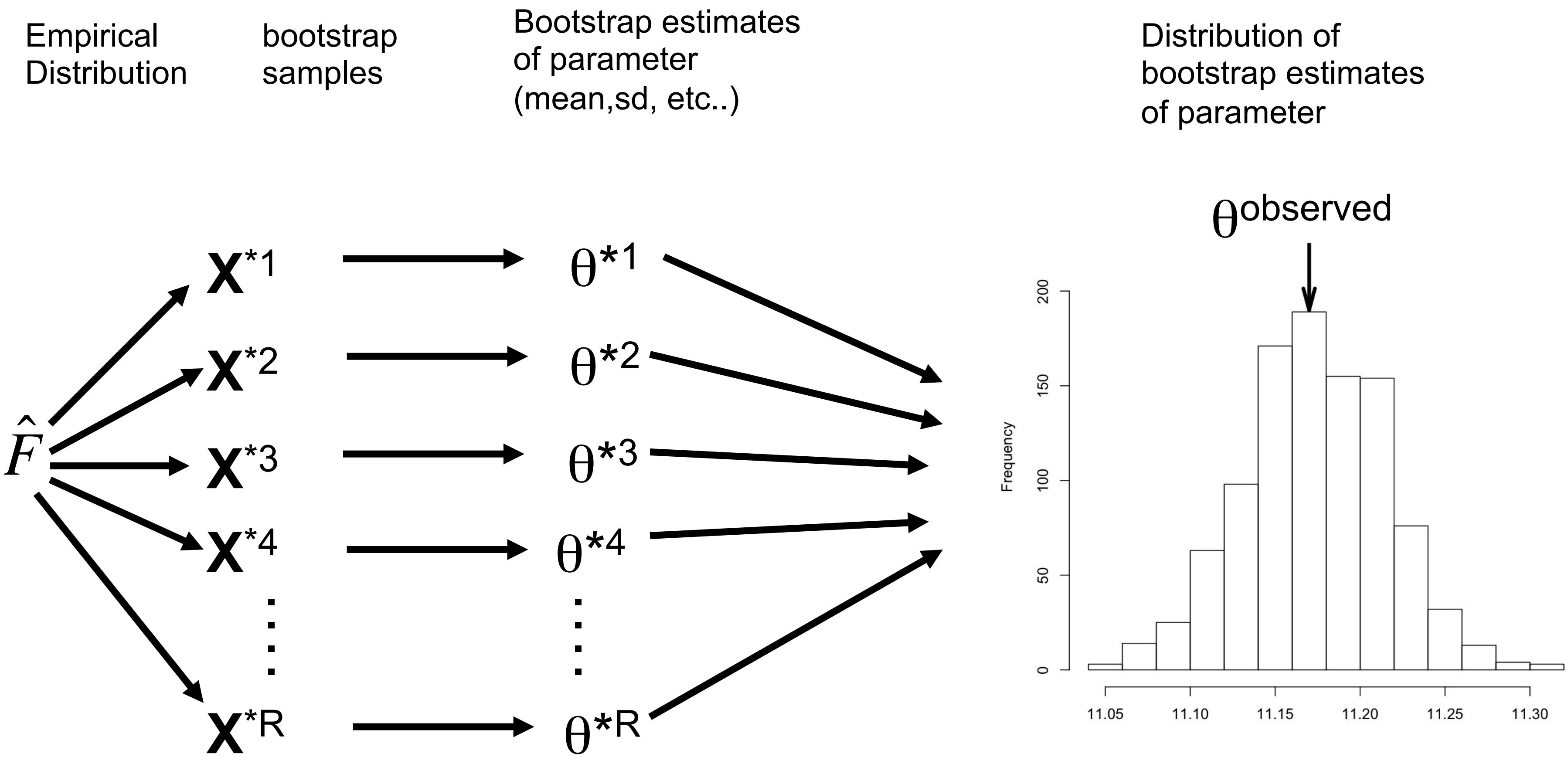
- Each observation from the data vector can potentially be sampled.
- Each observation always has the same (and equal) probability of being "chosen" for each resampling event.

chalkboard

The bootstrap procedure

1. Calculate quantities of interest from the observed data.
2. Resample with replacement lots of times (~10000). Calculate quantities of interest.
3. Look at empirically determined sampling distribution of quantities of interest. Calculate mean, sd, etc..

The bootstrap algorithm



Sampling with replacement

In R the function we use is

```
sample(x, n, replace=T)
```

Sampling with replacement

- Observed data (5, 7, 6, 8, 10), mean = 7.2

- “bootstrap”1 (7, 7, 6, 10, 6), mean = 7.2
- “bootstrap”2 (5, 6, 8, 10 ,10), mean = 7.8
- “bootstrap”3 (6, 8, 5, 10 , 8), mean = 7.4
- “bootstrap”4 (7, 5, 7, 10, 5), mean = 6.8
- “bootstrap”5 (6, 5, 7 , 8, 10), mean = 7.2
- bootstrapped mean 7.28

In this instance this is highly biased since we have such a small sample size!!!

R

How would you use the bootstrap to construct confidence intervals?

Interesting point: when using bootstrap to calculate SE, you don't need many replicates. But when you calculate CI, you do.

Why?

Because for CI, we need to know something about the TAILS of the distribution! and we can only learn about the tails with large number of bootstrap replicates.

Bootstrap confidence interval

- basic (percentile) bootstrap 95% CI are just the 2.5%, and 97.5% percentiles from the empirical distribution.

- Bias-corrected and accelerated confidence intervals (BC_a CI) are preferred (but the algorithm is complicated, so we will use a pre-built function).

Useful tidbits

- The bootstrap variance for any estimator ..

$$\hat{V}^*(T^*) = \frac{\sum_{b=1}^{R} (T^* - \overline{T^*})^2}{R-1}$$

where R is # of bootstrap replicates.

- Bootstrap S.E = bootstrap S.D

- Bootstrap estimate of bias = mean of bootstrapped distribution - statistic for original data

Using the bootstrap to make statistical inference

- In addition to providing measures of uncertainty for our point estimates, bootstrapping can also be used for statistical inference.
- Ho $\mu_a = \mu_b$

A parametric t-test

- A test of “means” across treatments
- the “t” test statistic

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}} \text{ where } s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2 + s_2^2}{n}}$$

A bootstrapped based "t-test"

- Test statistic is still the difference in treatment means ($\mu_{a\,-}\ \mu_b$).
- However it is the uncertainty in the estimates that is treated differently.
- By sampling with replacement from the observations under each treatment, we are empirically estimating this uncertainty in the means, and thus the uncertainty in the difference ($\mu_{a\,-}\ \mu_b$).

t-test via bootstrap procedure

1. For each treatment (a & b) calculate test statistic of interest
   (i.e. mean(a) - mean(b))
2. Perform bootstrap resampling events as before for each group (a & b) separately. For each resampling event calculate test statistic.
3. Compare sampling distribution of the test statistic to a pre-determined null hypothesis.
   i.e. mean(a) - mean(b) = 0
4. P-values calculated as
   (# resampled test statistics > HO)/ # resampling events
   (or < depending on hypothesis)

This is a very basic approach.. Now let us move onto more complex models

Ian’s Bias on bootstrap bias

- Those among you who have been paying attention may have noticed something strange about the approach we have used.
- Namely, we tested against a null, but how do we really know what the null distribution should have looked like.

Boostrapping for hypothesis testing

- Some consider the approach we have taken to be a poor one, since we have not evaluated against sampling from a null.
- Indeed it likely only works if our observed and bootstrapped test statistic both follow the same distribution if the null is true.

Bootstrap hypothesis testing

- Thus it is often advocated to test by combining your two samples, resampling from the combined sample randomly. Sort of a combination of permutation and sampling with replacement simultaneously.

BUT…

- In my experience, this rarely gives particularly different results.
- There are some situations where our approach (see references) can go wrong. Usually due to small sample sizes, and small numbers of bootstrap replicates.
- In the approach that is often advocated, assumptions we want avoid start creeping back in….
- So what should you do?

So what to do?

- Some advocate deriving p-values directly from the bootstrap CI.
- Some use permutation tests for p-values, and bootstraps for CI.
- Some ignore P-values as a waste of time.
- Some use one of the many other approaches that have been advocated (See Bootstrap hypothesis testing).

Bootstrapping linear models

- There are two general approaches to using the bootstrap for estimation and inference.
- Pairs resampling.

- Residual Resampling.

$Y = \beta_0 + \beta_1 X + \varepsilon$

"Pairs" bootstrap for linear models.

- Sometimes called "Random" or "Observational" resampling.

- Y and X are treated as random variables (observed with variation and possibly error).

- Resampling across observations (so that y_i and x_i are always held together).

chalkboard

Pairs resampling: Considerations

- Does not assume as much of a “fitted model”.
- Estimates unconditional variation in the co-efficients (β). i.e. Var(β).
- Takes longer to converge (but irrelevant for almost any computer now).
- NOT appropriate for truly “fixed” X designs (I.e. ANOVA models).

Pairs resampling

- Pairs resampling can also be used for estimating SE and CI for correlation co-efficients.

Residual Resampling

- also called "fixed" or "experimental" resampling method.
- X is assumed to be be a "fixed" value. i.e. X is measured without error.
- X is independent of ε.
- model is fit with observed data, residuals are extracted, and ***resampling occurs on residuals*** only.

chalkboard

Residual Resampling: Considerations

- Assumption of X being fixed. Is it appropriate for your data?
- X is independent of ε. This assumes no heterodascticity.
- Residual resampling usually gives a smaller standard error of the co-efficient (as X is not changing). $\mathrm{Var}(\beta|X) \leq \mathrm{Var}(\beta)$

Further reading

- Manly, Brian F. J. 2007. Randomization, bootstrap and Monte Carlo methods in Biology. Chapman & Hall/ CRC.
- Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap. Chapman & Hall.*
- Noreen, E.W. (1989) *Computer Intensive Methods for Testing Hypotheses. John Wiley & Sons.*
- Davison, A.C. and Hinkley, D.V. (1997) *Bootstrap Methods and Their Application. Cambridge University Press.*
- *MacKinnon, James. 2007. Bootstrap Hypothesis testing (on ANGEL).*