Parametric Bootstrapping
18.05 Spring 2017
Parametric bootstrapping

Use the estimated parameter to estimate the variation of estimates of the parameter!

- **Data**: $x_1, \ldots, x_n$ drawn from a parametric distribution $F(\theta)$.
- Estimate $\theta$ by a statistic $\hat{\theta}$.
- **Generate many bootstrap samples from $F(\hat{\theta})$**.
- Compute the statistic $\theta^*$ for each bootstrap sample.
- Compute the bootstrap difference $\delta^* = \theta^* - \hat{\theta}$.
- Use the quantiles of $\delta^*$ to approximate quantiles of $\delta = \hat{\theta} - \theta$.
- Set a confidence interval $[\hat{\theta} - \delta^*_{1-\alpha/2}, \hat{\theta} - \delta^*_\alpha/2]$
Parametric sampling in R

# Data from binomial(15, \(\theta\)) for an unknown \(\theta\)
\[
x = c(3, 5, 7, 9, 11, 13)
\]
\[
\text{binomSize} = 15 \quad \# \text{known size of binomial}
\]
\[
n = \text{length}(x) \quad \# \text{sample size}
\]
\[
\text{thetahat} = \text{mean}(x)/\text{binomSize} \quad \# \text{MLE for } \theta
\]
\[
nboot = 5000 \quad \# \text{number of bootstrap samples to use}
\]

# nboot parametric samples of size n; organize in a matrix
\[
\text{tmpdata} = \text{rbinom}(n\times nboot, \text{binomSize}, \text{thetahat})
\]
\[
\text{bootstrapsample} = \text{matrix}(\text{tmpdata}, \text{nrow}=n, \text{ncol}=nboot)
\]

# Compute bootstrap means \(\text{thetahat}^*\) and differences \(\text{delta}^*\)
\[
\text{thetahatstar} = \text{colMeans}(\text{bootstrapsample})/\text{binomSize}
\]
\[
\text{deltastar} = \text{thetahatstar} - \text{thetahat}
\]

# Find quantiles and make the bootstrap confidence interval
\[
\text{d} = \text{quantile}(\text{deltastar}, c(.1,.9))
\]
\[
\text{ci} = \text{thetahat} - c(\text{d}[2], \text{d}[1])
\]
Board question

Data: 6 5 5 5 7 4 \sim \text{binomial}(8, \theta)

1. Estimate $\theta$.

2. Write out the R code to generate data of 100 parametric bootstrap samples and compute an 80% confidence interval for $\theta$.

(Try this without looking at your notes. We’ll show the previous slide at the end)
Preview of linear regression

- Fit lines or polynomials to bivariate data

- Model: $y = f(x) + E$
  - $f(x)$ function, $E$ random error.

- Example: $y = ax + b + E$

- Example: $y = ax^2 + bx + c + E$

- Example: $y = e^{ax+b+E}$ (Compute with $\ln(y) = ax + b + E$.)