Unit 9: Bootstrap Confidence Intervals

Unit 9 Objectives

• Explain basic ideas behind the use of computer simulation to obtain bootstrap confidence intervals
• Explain different methods for generating bootstrap samples
• Obtain and interpret simulation-based pointwise parametric bootstrap confidence intervals
Parametric Bootstrap Sampling for Parametric Inference
Nonparametric Bootstrap Sampling for Parametric Inference
Scatterplot of 1,000 (Out of $B = 10,000$) Bootstrap Estimates \( \hat{\mu}^* \) and \( \hat{\sigma}^* \) for Shock Absorber

Weibull Model
Weibull Plot of $F(t; \hat{\mu}, \hat{\sigma})$ from the Original Sample (dark line) and 50 (Out of $B=10,000$) $F(t; \hat{\mu}^*, \hat{\sigma}^*)$ Computed from Bootstrap Samples for the Shock Absorber
Bootstrap Sampling and Bootstrap Confidence Intervals

- Instead of assuming $Z_{\hat{\mu}} = (\hat{\mu} - \mu)/\widehat{se}_{\hat{\mu}} \sim N(0,1)$, use Monte Carlo simulation to approximate the distribution of $Z_{\hat{\mu}}$.

- Simulate $B = 4000$ values of $Z_{\hat{\mu}^*} = (\hat{\mu}^* - \hat{\mu})/\widehat{se}_{\hat{\mu}^*}$.

- Some bootstrap approximations:
  
  $\Rightarrow Z_{\hat{\mu}} \sim Z_{\hat{\mu}^*}$
  
  $\Rightarrow Z_{\log(\hat{\sigma})} \sim Z_{\log(\hat{\sigma}^*)}$
  
  $\Rightarrow Z_{\logit[\hat{F}(t)]} \sim Z_{\logit[\hat{F}^*(t)]}$

  when computing confidence intervals for $\mu$, $\sigma$, and $F$. 

Bootstrap Confidence Interval for $\mu$

- With complete data or Type II censoring,
  \[ Z_{\hat{\mu}_j^*} = \frac{\hat{\mu}_j^* - \hat{\mu}}{\widehat{se}_{\hat{\mu}_j^*}} \]
  has a distribution that does not depend on any unknown parameters. Such a statistic is called a \textit{pivotal} statistic.

- By the definition of quantiles, then
  \[ \Pr \left( z_{\hat{\mu}^*_{(\alpha/2)}} < Z_{\hat{\mu}_j^*} \leq z_{\hat{\mu}^*_{(1-\alpha/2)}} \right) = 1 - \alpha \]

- Simple algebra shows that
  \[ [\hat{\mu}^*, \hat{\mu}] = [\hat{\mu} - z_{\hat{\mu}^*_{(1-\alpha/2)}} \widehat{se}_{\hat{\mu}}, \hat{\mu} - z_{\hat{\mu}^*_{(\alpha/2)}} \widehat{se}_{\hat{\mu}}] \]
  provides an exact 95% confidence interval for $\mu$. With other kinds of censoring, the interval is, in general, only \textit{approximate}. (Assuming location scale or log location scale distribution.)
Bootstrap Distributions of Weibull $\hat{\mu}^*$ and $Z\hat{\mu}^*$ Based on $B=10,000$ Bootstrap Samples for the Shock Absorber

Bootstrap Estimates

$\muhat^*$

Bootstrap-\textit{t} Untransformed

$Z-muhat^*$

Bootstrap-\textit{t} Untransformed

Bootstrap cdf

$Z-muhat^*$
Bootstrap Confidence Interval for $\sigma$

- With complete data or Type II censoring,
  \[
  Z_{\log(\hat{\sigma}^*)} = \frac{\log(\hat{\sigma}^*) - \log(\hat{\sigma})}{\hat{\text{se}}_{\log(\hat{\sigma}^*)}}
  \]
  has a distribution that does not depend on any unknown parameters. Such a statistics is called a **pivotal** statistic.

- By the definition of quantiles, then
  \[
  \Pr\left( z_{\log(\hat{\sigma}^*) (\alpha/2)} < Z_{\log(\hat{\sigma}^*)} \leq z_{\log(\hat{\sigma}^*) (1-\alpha/2)} \right) = 1 - \alpha
  \]

- Simple algebra shows that
  \[
  [\sigma, \tilde{\sigma}] = [\hat{\sigma}/\psi, \hat{\sigma}/\tilde{\psi}]
  \]
  provides an exact 95% confidence interval for $\sigma$, where $\psi = \exp\left[ z_{\log(\hat{\sigma}^*) (1-\alpha/2)} \hat{\text{se}}_{\log(\hat{\sigma})} \right]$ and $\tilde{\psi} = \exp\left[ z_{\log(\hat{\sigma}^*) (\alpha/2)} \hat{\text{se}}_{\log(\hat{\sigma})} \right]$

With other kinds of censoring, the interval is, in general, only **approximate**.
Bootstrap Distributions of $\hat{\sigma}^*$, $Z_{\hat{\sigma}^*}$, and $Z_{\log(\hat{\sigma}^*)}$ Based on $B=10,000$ Bootstrap Samples
Bootstrap Confidence Interval for $F(t_e)$

- With complete data or Type II censoring [using $F = F(t_e)$],

$$Z_{\logit(\hat{F}^*)} = \frac{\logit(\hat{F}^*) - \logit(\hat{F})}{\hat{se}_{\logit(\hat{F}^*)}}$$

has a distribution that does not depend on any unknown parameters. Such a statistics is called a **pivotal** statistic.

- By the definition of quantiles, then

$$\Pr\left(z_{\logit(\hat{F}^*);(\alpha/2)} < Z_{\logit(\hat{F}^*)} < z_{\logit(\hat{F}^*);(1-\alpha/2)}\right) = 1 - \alpha$$

- Simple algebra shows that

$$[\hat{F}, \hat{F}] = [\hat{F}/\gamma, \hat{F}/\tilde{\gamma}]$$

provides an exact 95% confidence interval for $F$, where $\gamma = \exp\left[z_{\logit(\hat{F}^*);(1-\alpha/2)} \hat{se}_{\logit(\hat{F})}\right]$ and $\tilde{\gamma} = \exp\left[z_{\logit(\hat{F}^*);(\alpha/2)} \hat{se}_{\logit(\hat{F})}\right]$

With other kinds of censoring, the interval is, in general, only **approximate**.
Bootstrap Distributions of $\hat{F}(t_e)^*$, $Z_{\logit[\hat{F}(t_e)^*]}$, and $Z_{\logit[\hat{F}(t_e)^*]}$ for $t_e=10,000$ km Based on $B=10,000$ Bootstrap Samples
Bootstrap Confidence Interval for $t_p$

- With complete data or Type II censoring,

$$Z_{\log(t_p^*)} = \frac{\log(t_p^*) - \log(t_p)}{\hat{se}_{\log(t_p^*)}}$$

has a distribution that does not depend on any unknown parameters. Such a statistic is called a **pivotal** statistic.

- By the definition of quantiles, then

$$\Pr \left( z_{\log(t_p^*)(\alpha/2)} < Z_{\log(t_p^*)} < z_{\log(t_p^*)(1-\alpha/2)} \right) = 1 - \alpha$$

- Simple algebra shows that

$$[t_p, \; \tilde{t}_p] = \left[ \frac{\hat{t}_p}{\hat{w}}, \; \hat{t}_p/\tilde{w} \right]$$

provides an exact 95% confidence interval for $t_p$, where $w = \exp \left( z_{\log(t_p^*)(1-\alpha/2)} \hat{se}_{\log(t_p)} \right)$ and $\tilde{w} = \exp \left( z_{\log(t_p^*)(\alpha/2)} \hat{se}_{\log(t_p)} \right)$

With other kinds of censoring, the interval is, in general, only approximate.
Bootstrap Distributions of $\hat{t}_p^*$, $Z_{\hat{t}_p^*}$, and $Z_{\log[\hat{t}_p^*]}$ for $t_e = 10,000$ km Based on $B = 10,000$ Bootstrap Samples
Other Topics in Chapter 9

• Nonparametric bootstrap confidence Intervals (with nonparametric resampling)
• Bootstrap percentile method