X is continuous random variable with pdf $g(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$

$\downarrow$

$f(z)$ $\mu = 0$, $\sigma = 1$

$X = \mu + \sigma Z$ mean $= \mu$, stdev $= \sigma$

$\hookrightarrow$ random variable pdf $f(z)$

Objective = obtain bootstrap interval estimates for μ and σ

Given $X_1, \dots, X_n \rightarrow$ estimates of μ and σ are $\overline{X}$ and S (sample mean/stdev)

t-statistic for $f(z)$ is $t = \frac{\overline{X} - \mu}{S/\sqrt{n}} \rightarrow t$ distribution with n-1 DoF (SLIDE 16)

$t_{.975} = 97.5^{th}$ percentile of this t-distribution

$$P\left(-t_{.975} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{.975}\right) = .95$$

solve the inequality to obtain 95% CI

$$\overline{X} - t_{.975}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{.975}\frac{S}{\sqrt{n}}$$

<u>Note</u>: distribution of t-statistic does not depend on either μ or σ

A function of observations and parameters whose pdf doesn't depend on the parameters is a PIVOT QUANTITY

a pivot quantity doesn't need to be a statistic – function and value can depend on parameters, distribution MUST <u>NOT</u>

if $f(z)$ is not normal, t-statistic is still a pivot quantity, which is referred to as a t-pivot quantity

CANNOT claim distribution is t-distribution with n-1 DoF

<u>assume</u> able to obtain distribution of t-pivot

$t_{.025}$ & $t_{.975}$ are 2.5^{th} & 97.5^{th} percentiles of this distribution

assert (SLIDE 16)

NEXT SLIDE

95% CI is (SLIDE 17) reference to past when normal $t_\alpha = -t_{1-\alpha}$

Can have different percentiles t_a & t_b such that $b-a = .95$

$t_{.025}$, $t_{.975}$ may NOT produce interval symmetric about the mean (like with t-distr.)

for, 100_p% CI ($p > \frac{1}{2}$)

$[(1-p)/2]$th and $[(1+p)/2]$th percentiles

Bootstrap CI for μ

provides a way to approximate the percentiles of distribution of the t-pivot

STEPS (SLIDE 17)

1. Compute $\bar{x}$ and S of original data

2. bootstrap sample size n. mean $\bar{X}_b$ and stdev S_b of bootstrap sample

NEXT SLIDE

3. Compute bootstrap t-pivot quantity (unnecessary *'s)

$$t_b = \frac{\bar{X}_b - \bar{X}}{S_b/\sqrt{n}}$$

repeat 2-3 B times (typically between 1,000 and 5,000) = bootstrap distribution of t_b's

1000 → grab t_{25} and t_{975}

$t_{b,.025}$ and $t_{b,.975}$ 2.5th & 97.5th percentiles of bootstrap distribution

discrete nature of bootstrap distr., may not be percentiles for desired confidence

THEN: interpolate or choose percentiles as close to desired as possible

Can do pivot-quality for variance, χ^2-pivot

bootstrap CI for σ^2

NOTE: bootstrap distribution of a pivot quantity is just an approximation of its true distribution

Thus, level of confidence asserted by bootstrap procedure is an approximation of actual level of confidence

NEXT SLIDE

Comparisons (SHOW table 8.3.2)

if exponential distr., then parametric interval (exact 95% coverage of mean)

may not know distr., use bootstrap interval, even though approximate

Manly (1997) 95% CI for μ exponential $n=20$

percentile = 90.1%, residual = 88.8%, BCA = 92.4%, t-pivot = 95.2%

When intervals miss true value, tend to be on low side

prefer t-pivot (given a pivot quantity) or BCA interval

transformation equivariance = CI for $\hat{\theta}$ (L, U) then $g(\theta)$ $(g(L), g(U))$

Second-order accurate = error E is proportional to step-size to the 2nd power

NEXT SLIDE

Code