

전산통계

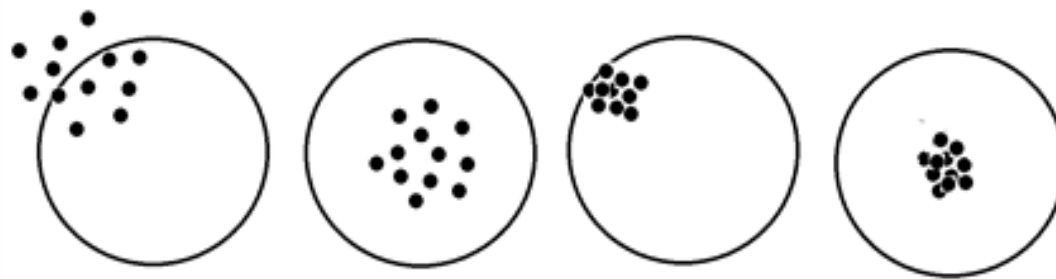
Chapter 5. 붓스트랩 구간추정

구간추정

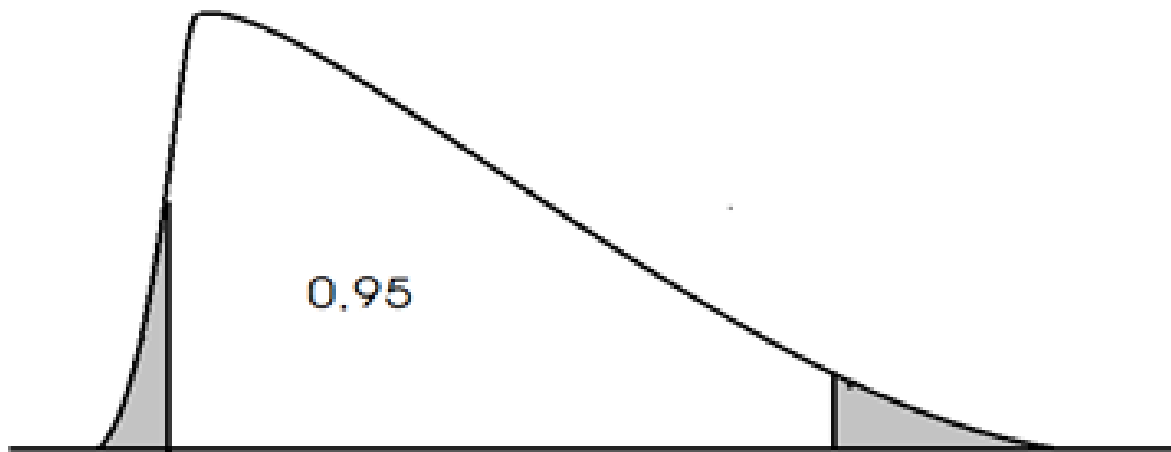
- 모수 : 모집단의 특성을 나타내는 수치 θ
- 모수 추정을 위해서는 표본이 필요 $\{x_1, x_2, \dots, x_n\}$
- 표본은 모집에 대하여 대표성을 지녀야 한다.
- 추정량 (estimator) : 모수 추정에 사용되는 함수 $\hat{\theta} = T(x_1, x_2, \dots, x_n)$

모수	추정량
평균	표본평균
비율	표본비율
표준편차	표본표준편차
분위수	표본분위수
⋮	⋮

-
- 표본에 따라 추정량이 변하므로 추정량은 분포를 가진다.
 - 바람직한 추정량의 분포 특성 : 추정치가 모수 근처에 몰려 있어야 한다.
 - $E(\hat{\theta}) = \theta$
 - 작은 표준편차
 - 점추정 (point estimation) : 추정치



- 신뢰구간 (confidence interval)
- 추정치와 추정량의 분포로 만들어진다.
- 신뢰수준에 따라 신뢰구간이 모수를 포함할지 믿음의 문제
- 점추정보다 구간추정이 신뢰성 있는 정보를 준다.
- 동일한 신뢰수준에서 신뢰구간의 폭이 작아질수록 정밀도가 높아진다.



-
- 추정량의 분포를 알고 있는 경우
 - $\hat{\theta} \sim N(\theta, Var(\hat{\theta}))$ 일 때, θ 에 대한 95% 신뢰구간은 $(\hat{\theta} - 1.96\sqrt{Var(\hat{\theta})}, \hat{\theta} + 1.96\sqrt{Var(\hat{\theta})})$
 - $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ 이면 μ 에 대한 95%신뢰구간은
 - σ^2 을 알 때, $(\bar{x} - 1.96\sqrt{\frac{\sigma^2}{n}}, \bar{x} + 1.96\sqrt{\frac{\sigma^2}{n}})$
 - σ^2 을 모를 때, $(\bar{x} - t_{0.025}(n-1)\sqrt{\frac{s^2}{n}}, \bar{x} + t_{0.025}(n-1)\sqrt{\frac{s^2}{n}})$

-
- t-분포를 이용한 평균에 대한 95% 신뢰구간
 - χ^2 분포를 이용한 분산에 대한 95% 신뢰구간

붓스트랩 구간추정

- 독립적으로 추출된 표본 $\{x_1, x_2, \dots, x_n\}$ 으로 모수 θ 를 구간추정
- 추정량 $\hat{\theta}$ 로 θ 를 점추정 할 수 있으나, $\hat{\theta}$ 의 분포를 모른다.
- 붓스트랩은 자료만 사용하는 재표본추출방법 (resampling method)으로 문제를 해결
- 붓스트랩 추정은 추정량이 아니라 통계적 모의실험의 형태로 표현된다.

-
- 단계1) $x = (x_1, x_2, \dots, x_n)$
 - 단계2) 붓스트랩 표본 반복 생성단계 : $b = 1, 2, \dots, B$
 - $x_1, x_2, \dots, x_n \rightarrow (x_1^{*b}, \dots, x_n^{*b})$
 - 붓스트랩 표본을 사용하여 통계치를 산출한다. $(x_1^{*b}, \dots, x_n^{*b}) \rightarrow \widehat{\theta}^{*b}$
 - 단계3) 구간추정단계 : 산출된 $\widehat{\theta}^{*b}$, $b = 1, 2, \dots, B$ 을 이용하여 θ 에 대한 구간추정을 한다.