부트스트랩 절차

정의

- 백분위 부트스트랩 신뢰구간

1. X가 $f(x;\theta)$를 갖는 연속확률변수라 하자.

1) $X_1, \ldots, X_n$이 X의 확률표본이고, $\hat{\theta}$를 θ의 점추정량이라고 하자.

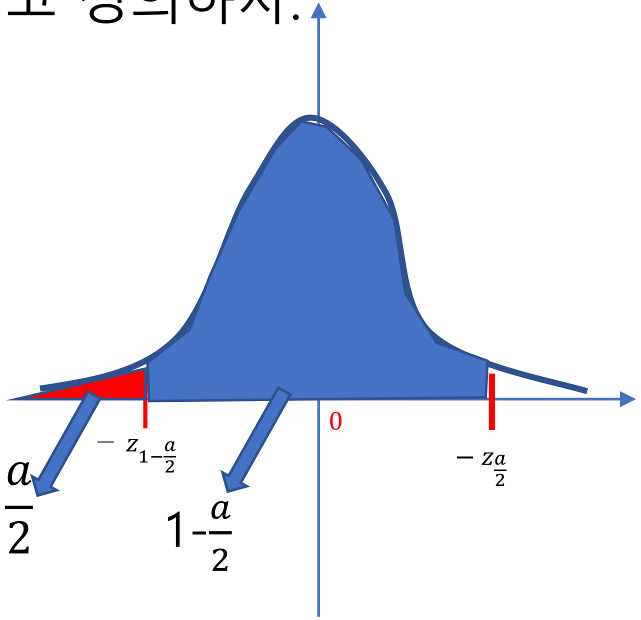
2) $\hat{\theta}$ ~ N(θ, $\sigma_{\hat{\theta}}^2$) 이라면, θ에 대한 신뢰구간은 $\hat{\theta}- z_{1-\frac{a}{2}} \cdot \sigma < \theta < \hat{\theta} - z_{\frac{a}{2}} \cdot \sigma$

(1) 이 때, $\hat{\theta}_L = \hat{\theta}- z_{1-\frac{a}{2}} \cdot \sigma$ 라고 정의하고, $\hat{\theta}_U = \hat{\theta}- z_{\frac{a}{2}} \cdot \sigma$ 라고 정의하자.

3) $\hat{\theta}_*$ ~ N($\hat{\theta}$, $\sigma_{\hat{\theta}}^2$) 이라고 할 때, [$\hat{\theta}_L$, $\hat{\theta}_U$] 은 $\hat{\theta}_*$의 분포에서

(1) $\mathrm{p}\left(\hat{\theta}_* \leq \hat{\theta}_L\right) = \mathrm{p}\left(\frac{(\hat{\theta}_* - \hat{\theta})}{\sigma} \leq - z_{1-\frac{a}{2}}\right) = \frac{a}{2}$ 이다.

(2) $\mathrm{p}\left(\hat{\theta}_* \leq \hat{\theta}_U\right) = \mathrm{p}\left(\frac{(\hat{\theta}_* - \hat{\theta})}{\sigma} \leq - z_{\frac{a}{2}}\right) = 1 - \frac{a}{2}$ 이다.



정의

- 백분위 부트스트랩 신뢰구간

2. 이 때, 무한한 수의 확률표본 $X_* = \{X_1, \dots X_\infty\}$ 을 추출하고,

1) 각각의 표본 X_* 에 대하여 그 통계량 $\hat{\theta}_*$ = $\hat{\theta}(X_*)$를 구하면 그 통계량의 추정량과 신뢰구간을 추정할 수 있다.
2) 하지만, 현실의 경우 단 하나의 $\hat{\theta}_*$만 구할 수 있기 때문에, 표본과 동일한 크기인 n만큼의 샘플을 X에서 복원 추출로 추출하여 통계량을 추정한다.

정의

- 부트스트랩이 타당성을 갖는 증명

1. H가 $\hat{\theta}$에 대한 CDF($\hat{\theta};\theta$) 이고 , $F_c(x)$ 가 N(∅, σ^2)인 분포의 cdf라고 할 때

1) ∅= m(θ)= $F_c^{-1}[H(\theta)]$ ~ N(∅, σ^2)

2) 이 때, N(∅, σ^2)에서 ∅의 (1-a)% 신뢰구간은

(1) (1-a) = p[$\hat{\emptyset}$ - $z_{1-\frac{a}{2}} \cdot \sigma^2$< ∅ < $\hat{\emptyset}$ - $z_{\frac{a}{2}} \cdot \sigma^2$]

(2) 이 때, ∅ = m(θ) 이므로, θ = $m^{-1}(\theta)$ =$[F_c^{-1}[H(\theta)]]^{-1}$= $H^{-1}[F_c$ (∅)]

(3) 신뢰구간을 정의하면 p[m^{-1}($\hat{\emptyset}$ - $z_{1-\frac{a}{2}} \cdot \sigma^2$) < θ < m^{-1}($\hat{\emptyset}$ - $z_{\frac{a}{2}} \cdot \sigma^2$)] 이다.

3)이 때, $\hat{H}$ = CDF($\hat{a};\hat{\theta}$), 즉 $\hat{\theta}$에 의존하는 cdf로 변환하고$\hat{\theta}_*$가 CDF($\hat{\theta}_*;\hat{\theta}$) 로 정의하면

(1) $\mathrm{p}\left(\hat{\theta}_* \le \hat{\theta}_L\right) = \mathrm{p}\left(\hat{\theta}_* \le m^{-1}(\hat{\theta} - z_{1-\frac{a}{2}} \cdot \sigma)\right) = \mathrm{p}\left(\hat{\emptyset}_* \le \hat{\emptyset} - z_{1-\frac{a}{2}} \cdot \sigma)\right)$

$= \mathrm{p}\left(\frac{\widehat{(\emptyset_* - \emptyset)}}{\sigma} \le - z_{1-\frac{a}{2}})\right) = \frac{a}{2}$

(2) 즉, 여전히 동일한 신뢰구간을 형성하는것을 확인할 수 있다.

예제

- 백분위 부트스트랩 신뢰구간

1. $\Gamma(1,100)$ 분포에서 20개의 확률표본 $X_1, \dots, X_{20}$을 추출한다.

1) $X_1, \dots, X_{20}$의 실현값 $x_1, \dots, x_{20}$을 n=20개에 대하여 재복원을 허용하여 복원추출한다.

2) 추출한 j-iter의 $X_{j*} = \{x_{1*}, \dots, x_{20*}\}$에 대하여 $\overline{X_{j*}}$를 계산한다.

3) J를 3000회 순회하고, 이 때 저장된 $\overline{X_{1*}} \cdots \overline{X_{j*}}$ 의 평균, 신뢰구간을 구한다.