Introduction to Nonparametric Statistics

Yitae Kwon

SFERS

2023-2

Contents

1. Parameteric Statistics
   - Parametric Statistics
   - Ideas of Statistics
2. Nonparametric Statistics and Robustness
3. Example1: Wilcoxon Test
4. Example2: M-estimator
5. Example3: Bootstrap

Parameteric Statistics

Parametric Statistics

- 우리가 일반적으로 접하는 통계학은 **'모수통계학'**(parametric statistics)이다.
- 추정과 검정의 과정에서 표본의 분포가 정규분포와 같은 특정 분포를 따름을 가정하고, 그 위에서 결론을 얻어내고는 한다.
- 예를 들어, 정규분포 $N(\mu, \sigma^2)$는 두 개의 미지의 **모수**(parameter) μ와 σ^2를 안다면 완벽히 기술할 수 있다.
- 정규분포를 따르는 IID 표본 $X_1, \cdots, X_n$에서 평균의 추정량:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Parametrical Statistics

- **모수**(parameter)는 특정 분포의 특성을 설명/요약할 수 있는 모든 값들을 말한다. 모수는 실수가 될 수도 있고, 가끔은 벡터나 행렬이 되기도 한다. 일반적으로 θ처럼 쓴다.
- **모수모형**(parametric model)은 이러한 모수들 유한 개로써 표현될 수 있는 모형을 의미한다.
- **모수공간**(parametric space)는 모수가 속할 수 있는 공간을 의미한다. θ를 포함한다는 의미에서 Θ로 자주 쓴다.

- 예를 들어, 정규분포 $N(\mu, \sigma^2)$의 모수는 μ와 σ^2이며, 모수공간은 $\mathbb{R} \times [0, \infty)$이다.

Parametrical Statistics

- 어떤 함수가 모수 θ와 상수 α에 의하여 변할 수 있다면, 우리는 해당 함수를
$$f(x_1, \cdots, x_n; \theta, \alpha)$$
처럼 써
(i) f가 θ, α에 의해 함수 그 자체로 변화하면서
(ii) $x_1, \cdots, x_n$이라는 가변적인 값에 의하여 함숫값이 결정됨
을 동시에 표현한다.
- θ와 α가 여러 개여도 되고, 없어도 된다.

Parametrical Statistics

- 모수모형을 이용한 추론에서, 우리는 유한 개의 모수만 얻어내면 된다.
- 많은 경우 우리는 그 모수모형이 참인지, 거짓인지에 관계없이 적당한 추정방식을 통하여 모수를 추정해줄 수 있다.
- 이처럼 모수모형의 분포를 실제 데이터가 따르는 분포에 가장 가깝게 만드는 모수 θ^*를 **pseudo-true parameter**라고 한다.
- 우리는 모수모형이 참인지 모르니, 이 pseudo-true parameter의 추정량 $\hat{\theta}$를 찾는 것을 주로 목표로 한다.
- 만약 진짜로 이 모형이 주어진 모수모형을 따르도록 설계되었다면, pseudo-true parameter는 곧 모수이다. θ^*의 추정량 $\hat{\theta}$는 곧 θ의 추정량이기도 하다.

Estimator and Estimates

- 모수모형에서 많은 경우 우리는 특정한 값 그 자체를 추정하고자 한다.
- 이처럼 특정한 값 자체에 대한 추정량을 **점추정량**(point estimator) 이라고 한다.
- 한편 가끔은 여론조사에서처럼 그 값이 분포할 것으로 예상되는 구간을 추정한다.
- 이처럼 특정한 값 근처의 구간에 대한 추정량을 **구간추정량**(interval estimator)이라고 한다.

- 표본 $X_1, \cdots, X_n$으로 표현되는 추정의 함수(확률변수)를 **추정량**(estimator)이라 하며, 해당 함수에 실제로 얻은 표본값 $x_1, \cdots, x_n$을 대입해 얻은 값을 **추정값**(estimates)이라 한다.
- 추정하려는 대상 자체를 **estimand**라 한다.

$\overline{\text{Bar}}$ & $\widehat{\text{Hat}}$

- 어떠한 값들의 평균을 표현하고자 할 때, 자주 사용하는 것이 bar이다.

$$\overline{X} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_n$$

- bar 안에 포함되는 것이 X같은 것과 달리 너무 길다면

$$E_n[X] = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_n$$

  와 같이 쓰기도 한다.
- 어떠한 값의 점추정량을 표현하고자 한다면 그 위에 hat을 붙인다. $\hat{\theta}$처럼.

Confidence Interval and Confidence Band

- 구간추정량은 우리가 **신뢰구간**(confidence interval)이라는 이름으로 더 잘 알고 있다.
- **신뢰수준**(confidence level) $1-\alpha$에 대한 모수 θ의 신뢰구간이 (L, U)이라고 하자. 이때 L과 U는 표본 $X_1, \cdots, X_n$으로써 얻어지는 일종의 확률변수이다.
- 신뢰수준이 $1-\alpha$란 것은 곧 우리가 표본 추출을 해서 (L, U)를 구하는 과정을 B번 수행한다고 할 때 이 구간 안에 실제 모수가 들어오는 경우가 평균적으로 $(1-\alpha)B$번이라는 것이다.

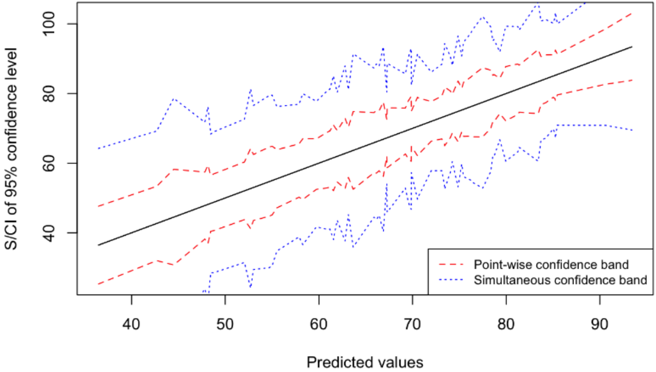
Confidence Interval and Confidence Band

- 어떤 함수 f에 대하여 특정한 a에서의 함숫값 $f(a)$의 신뢰구간을 구하는 상황을 생각하여 보자. 그렇다면 L과 U는 a에 의해서 바뀔 수 있는 값이다. 그러니 a에서의 신뢰구간을 $(L(a), U(a))$처럼 쓰자.
- 그렇다면 이제 추정하고자 하는 게 $f(a)$가 아니라 f 그 자체라고 하자. 이 경우 모든 x에서 $f(x)$를 감싸는 신뢰구간을 만들어주어야 하며, 이를 **신뢰밴드**(confidence band)라 부른다.
- 자연스럽게 생각할 수 있는 f의 신뢰구간은

$$(L(x), U(x))$$

이다. 즉, 모든 x에서 하나하나 신뢰구간을 만들어 그것을 이어붙이자는 것이다. 이렇게 얻은 신뢰구간을 **pointwise confidence band**이라고 부른다.

Confidence Interval and Confidence Band



- 또다른 것은 **simultaneous(uniform) confidence band**이다.

Confidence Interval and Confidence Band

- 우리는 $f(x)$만이 아니라 전체 f의 신뢰밴드에의 포함확률을 $1-\alpha$ 이상으로 통제하고 싶다.
- pointwise confidence band는 '모든 x에서' $f(x)$의 포함확률을 $1-\alpha$로 통제한다. 그렇다면 한 100개 정도의 x에서 이 신뢰구간을 보았다면, 평균적으로 100α개 만큼의 $f(x_i)$는 신뢰구간에서 벗어나 있을 것이다! (편의상 신뢰구간들은 독립적이라고 가정한다.)
- 그렇다면 전체 f의 신뢰밴드 안에 f가 '온전히' 들어가 있을 확률은 얼마나 될까? 100개에서만 본다고 하더라도, $(1-\alpha)^{100}$의 확률로 이 안에 들어간다.
- 자주 사용하는 $\alpha = 0.05$에서, pointwise confidence band가 실제 f를 '온전히' 포함하고 있을 확률은 $(1-0.05)^{100} \approx 0.0059$밖에 안 된다는 것이다.
- f를 추정하고자 하는 우리의 입장에서 얘는 만족스럽지 못하다.

Confidence Interval and Confidence Band

- 이러한 문제를 통계학에서는 **다중검정**(multiple testing)의 문제라고 말한다. 다양한 모수에 대하여 한번에 추정과 검정을 수행하려다 보니 원하는 신뢰수준에서 검정이 이루어지지 못하는 것이다.
- 대표적인 해법 중 하나는 **Bonferroni correction**이다. 신뢰구간을 만들 때 α 대신 $\alpha/100$을 사용한다고 해보자. 그렇다면 pointwise confidence band를 만들 때 L과 U가

$$L(x;\alpha), U(x;\alpha) \quad \Rightarrow \quad L(x;\alpha/100), U(x;\alpha/100)$$

  로 대체된다.
- 그렇다면 100개 점에서의 '온전한' 포함확률은

$$\left(1-\frac{\alpha}{100}\right)^{100} \approx 1-\frac{\alpha}{100}\cdot 100 = 1-\alpha$$

  로 원하는 바와 같다.

Confidence Interval and Confidence Band

- 이러한 문제를 종합하여 전역 x에서 f가 신뢰밴드 안에 잘 포함되도록, 즉 '온전히' 혹은 '동시에' 포함하는 확률이 $1-\alpha$이도록 만든 신뢰밴드를 **simultaneous confidence band**라고 부른다.
- Bonferroni correction을 포함한 다양한 방법이 이 신뢰밴드의 구축에 사용되며, 뒤에서 다룰 **Bootstrap** 역시 자주 사용되는 방법이다.
- 일반적으로 pointwise confidence band보다 항상 simultaneous confidence band가 더 넓다. 하지만 더 자주 사용된다. 신뢰밴드의 너비가 늘어나는 것은 분명 추정의 효율성을 낮추는 일이지만, 현대 통계학에서는 그 약점을 감수하고 simultaneous confidence band를 더 선호한다.

Normal Distribution

- 많은 경우에 추정량들은 점근적으로(n이 크다면) 정규분포를 따른다.
- 특정 확률변수 X가 어떤 분포를 따른다면, $X \sim ()$처럼 쓴다.
- 일변량 정규분포의 경우 $N(\mu, \sigma^2)$처럼 자주 쓴다.
- 여기서 μ를 추정하는 경우, m 혹은 $\hat{\mu}$를 자주 사용한다.
- 여기서 σ^2을 추정하는 경우, s^2 혹은 $\hat{\sigma^2}$을 자주 사용한다.
- 다변량 정규분포의 경우 $N(\mu, \Sigma)$처럼 자주 쓴다.
- Σ를 추정하는 경우, S를 자주 사용한다.

Normal Distribution

- k차원 다변량 정규분포는 모수로 k차원 벡터 $\mu \in \mathbb{R}^k$와 $k \times k$ 행렬 $\Sigma \in M_{k\times k}(\mathbb{R})$을 가진다.
- 다변량 정규분포는 확률밀도함수로 k차원 벡터 $\mathbf{x} \in \mathbb{R}^k$의 함수

$$f(\mathbf{x}) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

  을 가지는 다변량 분포를 의미한다.
- 만약 $\mathbf{X} \sim N(\mu, \Sigma)$라면,

$$E[\mathbf{X}] = \mu, \quad \mathrm{Var}(\mathbf{X}) = \Sigma$$

  이다.

Probability Space

- 확률변수에 존재하는 '랜덤성'을 수학적 언어로 표현하기 위한 방법 역시 개발되어 있다.
- 특정 시행의 결과로써 나올 수 있는 모든 값들 ω를 모아 둔 집합을 **표본공간**(sample space) Ω로 표시하자.
- 주사위를 굴려 나오는 수들의 집합 $\{1, 2, 3, 4, 5, 6\}$이 대표적인 Σ이다.
- 그 다음, Σ의 부분집합들로 구성된 **event space** $\mathcal{F}$를 고려한다. 이는 특별한 성질을 만족해야 하나, 자세한 것은 생략한다.
- 그 다음 $\mathcal{F}$를 정의역으로 하여 그 원소를 0부터 1까지의 실수에 대응시키는 **probability measure** P를 고려한다.
- $(\Omega, \mathcal{F}, P)$를 **probability space**라 부르며, 사건 $e \in \mathcal{F}$에 대하여 $P(e)$를 이 probability measure 상에서 사건 e가 발생할 확률로 정의한다.

Random Variable

- **확률변수**(Random variable)는 특정 확률로 특정 값을 배출한다. 이는 사실 Ω를 정의역으로 하는 함수

$$X : \Omega \to \mathbb{R}$$

  처럼 해석해줄 수 있다.
- 주사위를 굴려 나온 눈의 제곱을 확률변수 X에 대응시켜 보자. 그렇다면 X는 $1/6$ 확률로 $1, 4, 9, 16, 25, 36$의 값을 내놓는 확률변수이다.
- 혹은, $X(n) = n^2$으로 생각해줄 수도 있다.
- 다른 예시를 들어 보자. X가 균등분포 $U([0, 1])$을 따르는 확률변수다.
- 그렇다면 $\Omega = [0, 1]$이며, $e = [a, b] \subset [0, 1]$에 대해 $P(e) = b - a$이고, $X(\omega) = \omega$인 것이다.

Density Function

- 이러한 probabilty measure P 하에서, 확률변수 X의 **확률밀도함수**(density function) f는

$$\int_a^b f(x)dx = P(X \in (a,b))$$

  로 정의된다.
- 더 넓게는, $\mathcal{F}$에 속하는 모든 사건 e에 대하여 $\int_e f = P(e)$인 함수 f를 의미하기도 한다.
- **다변량 확률변수**(random vector)에 대해서도 동일한 방식으로 density function을 정의할 수 있다. 단, 이 때에는 확률변수 $\mathbf{X}$가 $\mathbf{X} : \Omega \to \mathbb{R}^k$인 벡터함수이다.
- 확률밀도함수 f는 $\mathbb{R}^k$의 부분집합 A에 대하여 아래를 만족한다.

$$\int_A f(\mathbf{x})d\mathbf{x} = P(\mathbf{X} \in A)$$

Expectation

- **기댓값**(expectation)은 확률변수가 평균적으로 가지게 되는 값으로,

$$E[X] = \int_{\mathbb{R}} xf(x)dx$$

처럼 정의된다.

- 다변량 확률변수의 경우에도 마찬가지로

$$E[\mathbf{X}] = \int_{\mathbb{R}^k} \mathbf{x}f(\mathbf{x})d\mathbf{x}$$

로 정의할 수 있다.

- 다행히도, $\mathbf{X} = (X_1, \cdots, X_k)$이라면

$$E[\mathbf{X}] = (E[X_1], \cdots, E[X_n])$$

임이 알려져 있다.

Variance

- **분산**(variance)은 확률변수가 가지는 산포의 척도로,

$$\mathrm{Var}(X) = \int_{\mathbb{R}} (x - E[X])^2 f(x) dx$$

처럼 정의된다.

- 다변량 확률변수의 경우에도 마찬가지로

$$\mathrm{Var}(\mathbf{X}) = \int_{\mathbb{R}^k} (\mathbf{x} - E[\mathbf{X}])(\mathbf{x} - E[\mathbf{X}])^T f(\mathbf{x}) d\mathbf{x}$$

로 정의할 수 있다.

- 다행히도, $\mathbf{X} = (X_1, \cdots, X_k)$이라면

$$\mathrm{Var}(\mathbf{X})_{ij} = (\mathrm{Cov}(X_i, X_j))$$

임이 알려져 있다.

Measure 0 set

- A의 measure가 0라는 것은 event $A \in \mathcal{F}$가 $P(A) = 0$이라는 것이다.
- 예를 들어, 연속형 확률변수 상에서 한 점의 집합(singleton) $\{1\}$ 등은 measure 0 set이다. ($\Omega = \mathbb{R}$, $\mathcal{F} = \{[a,b] | a < b\}$)
- 만약 어떤 두 확률변수가 measure 0인 set에서만 다르다면, 둘이 **almost sure**하게 같다고 한다.
- 예를 들어, 두 연속형 확률변수 X와 Y가 probability space $([0,1], \{[a,b] | 0 \le a < b \le 1\}, P)$에서 정의되고 $P([a,b]) = b - a$라고 하자. 그렇다면

$$X(w) = w, \quad Y(w) = \begin{cases} w & w \neq 0.5 \\ 2023 & w = 0.5 \end{cases}$$

는 almost sure하게 같다. $\{0.5\}$가 measure 0 set이기 때문이다.

Conditional Expectation

- **조건부 기댓값**(conditional expectation)은 일종의 확률변수이다.
- $E[Y|X]$을 구하면, 이는 X에 의해 그 값이 결정되며, X가 확률변수이기 때문에 이 역시 확률에 의해 값이 달라지는 확률변수이다.
- 경제통계학에서 우리는 조건부 밀도함수를 구하고 이를 적분하여 조건부 기댓값을 구하였다.
- 실제 조건부 기댓값은 더 어렵게 정의되며, 여러 개 존재한다.
- 이들은 모두 almost sure하게 같다. 우리가 기존에 구했던 것, 혹은 자연스럽게 생각할 수 있는 것 등은 그 중 하나일 뿐이다.
- 따라서 조건부 기댓값들끼리, 혹은 조건부 기댓값이 특정 확률변수와 같다고 주장하려면 뒤에 a.s.를 붙여주는 것이 좋다.
- **즉 우리 수준에서는 a.s.가 붙은 건 그냥 무시해도 좋다!**

Convergence in Probability

- 확률변수의 수열을 **확률변수열**(seq. of r.v.)이라 한다.
- $$\operatorname*{plim}_{n\to\infty} X_n = X \quad \text{or} \quad X_n \xrightarrow{p} X$$

  로 표시하는 것은 모든 $\epsilon > 0$에 대해

  $$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$$

  임을 의미한다.
- 만약 두 확률변수열 $\{X_n\}$과 $\{Y_n\}$에 대하여,

  $$\frac{X_n}{Y_n} \xrightarrow{p} 0$$

  이라면 $X_n = o_p(Y_n)$처럼 쓸 수 있다.

Convergence in Distribution

- $$X_n \rightsquigarrow X \quad \text{or} \quad X_n \xrightarrow{d} X \quad \text{or} \quad X_n \dot{\sim} X$$
  로 표시하는 것은 모든 집합 A에 대하여
  $$\lim_{n\to\infty} P(X_n \in A) = P(X \in A)$$
  임을 의미한다.
- 더욱 간단하게는, 둘의 density function f_n, f 혹은 distribution function F_n, F에 대하여 그들이 continuous일 때
  $$\lim_{n\to\infty} f_n(x) = f(x), \quad \lim_{n\to\infty} F_n(x) = F(x)$$
  인 것으로 이해해도 된다.
- convergence in probability라면, convergence in distribution이다.

MSE

- 통계학에서 추정량의 평가 기준으로 가장 많이 사용하는 척도는 **MSE**(mean squared error)이다.
- MSE는
$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = \mathrm{Var}(\hat{\theta}) + \mathrm{bias}(\hat{\theta})^2$$
을 의미한다.
- bias는
$$\mathrm{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$$
를 의미한다.
- bias가 0인 추정량을 **불편추정량**(unbiased estimator)이라고 부른다.
- 추정량 평가의 측면에서, variance의 역수를 **효율성**(efficiency)라 부르기도 한다.

MSE

- 불편추정량 중에서 θ의 값에 관계없이 효율성이 가장 좋은 추정량, 즉 분산이 가장 작은 추정량을 **전역최소분산불편추정량**(UMVUE)이라고 부른다.
- 일반적으로 추정량의 bias가 감소하면 variance가 증가하고, variance가 감소하면 bias가 증가하는 경향이 있다. 이를 **편향-분산 상충**(bias-variance tradeoff)이라고 부른다.
- 만약

$$\lim_{n\to\infty} MSE(\hat{\theta}_n) = 0$$

이라면, $\hat{\theta}$는 θ로 **converge in quadratic mean**한다고 말한다.

Almost Sure Convergence

- 확률변수열 $\{X_n\}$이

  $$P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

  을 만족한다면, 즉 X_n이 X로 완벽히 수렴하는 $\omega \in \Omega$의 집합이 measure 0 set일 때, 우리는 X_n이 X로 **almost surely converge** 한다고 말한다.
- 이는 $X_n \overset{a.s.}{\to} X$처럼 쓴다.
- almost surely converge하면 converge in probability이며, 이에 따라 자연스레 converge in distribution이다.

Consistency

- 만약 추정량 $\hat{\theta}_n$이 추정하고자 하는 모수 θ로 converge in probability 한다면, 이는 **일치추정량**(consistent estimator)이다.
- 만약 추정량 $\hat{\theta}_n$이 θ로 almost surely converge한다면, 이는 **강한 일치성**(strong consistency)를 가진다고 말한다.
- 강한 일치성을 가지면 약한 일치성을 가짐은 명확하다!
- 앞서 본 것처럼 n이 무한대로 감에 따라서 원하는 추정량이 가지는 성질을 **점근성질**(asymptotic property)이라고 한다.

Some Analysis

- 어떤 집합 $\Theta \in \mathbb{R}^k$가 **compact**라는 것은, 이것이 **closed**이며 **bounded**임을 의미한다.
- **Neighborhood**는 점 $\mathbf{x} \in \mathbb{R}^k$에 대하여, $\mathbf{x} \in N \subset \mathbb{R}^k$인 **open set** N을 의미한다.
- $\mathbb{R}^k$에서 **open set** A는 A 내부의 어떤 점 a을 고르더라도 상응하는 $\delta(a) \in \mathbb{R}$가 있어 $|x - a| < \delta$인 모든 x가 $x \in A$인 set을 의미한다.
- int(A)는 A 안에 포함되는 모든 open set들의 union이다.
- B가 closed라는 것은 B의 여집합 $\mathbb{R}^k - B$가 open이라는 것이다.
- B가 **bounded**라는 것은 어떠한 $M > 0$이 존재하여 B의 모든 원소 $b \in B$가 $|b| < M$을 만족한다는 것이다.

Some Analysis

- 실수의 bounded인 집합 $A \in \mathbb{R}$에 대하여, 항상 **supremum** $b = \sup A$가 존재한다.
- b는 아래 두 가지 성질을 만족한다.
  (i) 모든 $a \in A$에 대하여, $a \leq b$
  (ii) $c < b$인 c에 대하여, $a' \in A$가 존재하여 $c < a'$이다.
- 함수에 대하여도 이러한 논의가 가능하다.

$$\sup_{\kappa \in \Gamma^*} |\dot{h}^{nev}_{g,t}(W;\kappa)|$$

  는 Γ^*에 있는 모든 κ에 대하여 W에서의 함숫값들을 모아둔 집합의 supremum을 의미한다.
- supp(f)는 sup(f)와는 다르다. supp은 **support**의 약자로, $f(x) \neq 0$인 x의 집합을 의미한다.

Some Notations

- **functional**은 함수의 형태에 따라 다른 값을 내놓는 일종의 함수이다. 즉 정의역을 함수공간으로 하는 함수이다.
- 함수를 $f(\cdot)$처럼 표기하는 것은 그 함수 자체를 의미하기 위한 단순한 표기다.
- transpose를 A^T처럼 쓰기도 하지만, A'이나 A^t도 자주 사용된다.
- **표준오차**(standard error)는 $\sqrt{n}$으로써 표준화한 표본표준편차를 의미한다. 즉
$$\mathrm{se}(\hat{\theta}) = \frac{s}{\sqrt{n}}$$
- 함수가 n번 미분가능하며 n차 도함수가 연속이라면, $f \in C^n$처럼 쓴다.
- 확률변수의 n차 **적률**(moment)는 $E[X^n]$을 의미한다.
- 만약 어떤 함수가 measure 0 set에서만을 제외하고 continuous라면, **a.s. continuous**처럼 자주 쓴다.

In Appendix

- Continuous Mapping Theorem(CMT)
- Weak/Strong Law of Large Numbers(LLN)
- Central Limit Theorem(CLT)
- Mixture Distribution

Nonparametric Statistics and Robustness

t-test for Two Sample

- 두 모집단의 평균을 비교하기 위한 **t-test**를 생각하여 보자.
- t-test는 두 모집단이 모두 정규분포를 따른다는 **모수적 가정** 하에서 이루어진다.
- 경제학부 A반과 B반의 평균 학업성적 차이를 알아보고자 한다.
- 검정에서 모수적 가정은
  1. 경제학부 A반의 학업성적 분포는 $N(\mu_1, \sigma_1^2)$이다.
  2. 경제학부 B반의 학업성적 분포는 $N(\mu_2, \sigma_2^2)$이다.
- 귀무가설과 대립가설은

$$H_0 : \mu_1 = \mu_2, \quad H_1 : \mu_1 \neq \mu_2$$

  으로 주어진다.
- A반 학생들의 학업성적은 $X_1, \cdots, X_{n_1}$이라는 표본으로, B반 학생들의 학업성적은 $Y_1, \cdots, Y_{n_2}$라는 표본으로 얻어졌다.

t-test for Two Sample

- 만약 $\sigma_1^2 = \sigma_2^2$임이 알려져 있다면, 검정통계량

$$t = \frac{\bar{Y} - \bar{X}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

  이 자유도 $n_1 + n_2 - 2$인 t분포를 따름을 이용하여 검정할 수 있다.

- **pooled estimator** s_p는

$$s_p = \sqrt{\frac{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}{n_1 + n_2 - 2}}$$

  인 $\sigma_1^2(\sigma_2^2)$의 UMVUE이다.

- t-test는 모수적 검정이 맞다는 가정 하에 **검정력**(power)이 가장 높은 검정이다.

t-test for Two Sample

- 만약 $\sigma_1^2 \neq \sigma_2^2$임이 알려져 있다면, **Welch's t-test**를 이용한다.
- 만약 σ_1^2과 σ_2^2을 모른다면, 개인의 선택에 의존한다. 위키백과는 하나가 다른 하나의 2배 이상일 때 Welch's t-test를 이용할 것을 권고한다.
- 다른 선택은 모수적 가정 하에서 **F-test**를 이용하여 equal variance를 검정한 뒤, 각 상황에 맞는 검정을 이용하는 것이다.

Evaluating Tests

- 통계적 검정을 평가할 때에는 두 가지 척도를 들여다본다.
- **제1종의 오류율**: 귀무가설이 참일 때 검정이 귀무가설을 기각할 확률
- **유의수준** α에서 검정을 수행하라는 것은 곧 해당 검정의 제1종의 오류율이 α 이하이도록 설계하라는 것이다.

- **제2종의 오류율**: 대립가설이 참일 때 검정이 귀무가설을 기각하지 않을 확률
- **검정력**(power)은 실제 모수가 θ일 때 1-(제2종의 오류율)을 의미한다. 즉, θ의 함수로 주어진다.

Evaluating Tests

- 통계적 검정은 아래의 방식으로 평가한다.
- 1. 검정이 제1종의 오류율을 원하는 수준에서 통제하는가?
  2. 검정의 검정력은 어떠한가?
- 만약 첫째 평가 기준을 통과하지 못했더라면, 검정의 **기각역**(critical region)을 바꾸어 유의수준을 통제한다.
- 첫째 평가 기준을 통과한 검정들 사이에서는 검정력을 비교한다. 자주 사용되는 것은 **점근상대효율**(Pitman's asymptotic relative efficiency)이다.
- 이는 주어진 대립가설 $\theta = \theta_0$와 주어진 검정력 β에 도달하기 위하여 필요한 표본의 크기 n을 비교하는 방법이다.
  (일반적으로 표본의 크기 n이 증가할수록 검정력이 증가한다)

Nonparametric Statistics

- t-test는 모수적 가정(정규분포 가정)이 참일 때 검정력이 가장 좋은 검정이다.
- 만약 경제학부 학생들의 학업성적 분포가 정규분포가 아니라면 어떻게 될까?
- 그 분포의 모양조차 알 수 없는 상황이라면?
- **비모수통계**(nonparametric statistics) 혹은 **분포무관 방법**은 그 해법을 제공한다.
- 이들은 모수적 가정을 최대한 줄인 채 통계적 추정과 검정을 수행한다.

Nonparametric Statistics

- 비모수통계학에서의 가정은 매우 제한적이다.
- 정규분포로 분포의 모양을 한정시켰던 t-test와 달리, 비모수적 검정에서는 분포족을

$$\mathcal{F}_1 = \{f : \int x^2 f(x)dx < \infty\} \quad \text{(평균과 분산이 존재하는 모든 분포)}$$

$$\mathcal{F}_2 = \{f : f(\mu + x) = f(\mu - x)\} \quad \text{(중앙값 } \mu \text{를 기준으로 대칭인 분포)}$$

와 같이 매우 넓게 설정할 수 있다.

Nonparametric Statistics

- t-test는 사실 정규분포보다 꼬리가 더 짧은 균등분포 등에서는 유의수준을 제대로 통제하지 못한다.
- 반대로 꼬리가 긴 Slash 분포 등에서는 검정력이 매우 감소하는 약점이 있다.
- 반면, 비모수적 검정방법은 정규분포 가정을 하지 않기에 균등분포부터 Slash 분포까지 모든 분포 형태에서 유의수준을 α로 통제할 수 있다.
- 심지어 꼬리가 긴 분포에서는 검정력이 t-test보다 높다.
- 따라서 정규분포와 유사하지 않은 모집단 형태가 나타난다면, 비모수적 방법을 사용하는 것이 안전하다.

Robustness

- 표본 $X_1, \cdots, X_n$을 관측했을 때, 그 관측값이 다른 관측값들과는 매우 이질적인 관측값을 **이상점**(outlier)라 부른다.
- 만약 모집단의 분포가 꼬리가 긴 분포라면, 극단값이 나올 확률이 크므로, 이상점이 나올 확률 또한 높다.
- 비모수적 방법은 이러한 꼬리가 긴 분포에 대해서도 유의수준을 잘 통제하므로, 이상점에도 **강건**(robust)하게 추정과 검정을 수행할 수 있다.
- 중앙값은 관측값의 일부가 오염(이상점)되더라도 추정값이 크게 변하지 않으므로 평균보다 로버스트하다.

Breakdown Point

- **붕괴점**(breakdown point)는 표본으로부터 계산되는 통계량을 붕괴시키기 위한 표본의 비율을 의미한다.
- 표본 $X_1, \cdots, X_n$을 표집하여 통계량 $T = T(X_1, \cdots, X_n)$을 계산하였다.
- 어떤 $b-1$개의 관측값 $X_{i_1}, \cdots, X_{i_{b-1}}$에 대해서도

$$\lim_{X_{i_1}, \cdots, X_{i_{b-1}} \to \infty} |T| < \infty$$

이면서, 적절한 b개의 관측값 $X_{i_1}, \cdots, X_{i_b}$를 골라주면

$$\lim_{X_{i_1}, \cdots, X_{i_b} \to \infty} |T| = \infty$$

라면 그 붕괴점은 $BP(T) = \frac{b}{n}$인 것이다.

Breakdown Point

- 표본평균의 붕괴점은 $\frac{1}{n} \approx 0$이다.
- 표본중앙값의 붕괴점은 대략 $\frac{1}{2}$이다.
- 붕괴점이 높을수록 더 로버스트한 추정량이다.
- 로버스트함을 평가하기 위한 붕괴점의 절대적인 기준은 없다.
- 비모수적인 추정량은 분포에 무관하므로, 모수적인 가정 하에 만든 추정량에 비해 붕괴점이 큰 경향이 있다.

Example1: Wilcoxon Test

Randimization Framework

- 실험 설계가 아주 잘 되어 있어 완전히 동질한 두 집단 A와 B가 있다고 하자.
- 두 집단의 차이는 오직 처리 여부이다.
- coin toss를 이용하여 임의로 A그룹과 B그룹을 나누었고, 처리 효과를 보고자 한다.
- 이처럼 두 그룹의 처리 여부가 랜덤하게 결정되어 각 그룹의 동질성을 보존할 수 있는 실험계획을 **랜덤화 실험**(Randomization Experiment)이라고 부른다.

Wilcoxon Test

- 만약 A그룹과 B그룹의 관측값이 정규분포를 따름이 알려져 있다면, t-test를 사용하면 된다.
- 만약 A그룹과 B그룹의 관측값 분포를 그려봤을 때 정규분포와 거리가 멀다면, t-test를 사용할 수 없다.
- **Shapiro-Wilk test**, **Jarque-Bera test** 등을 사용하거나 **q-q plot**을 그려봐 판단해줄 수 있다.
- 이 경우 평균(중앙값) 차이에 대한 비모수적인 추정과 검정이 필요하다.
- 대표적인 방법이 부호와 순위에 근거한 **Wilcoxon Test**이다.

Wilcoxon Test

- A그룹의 관측값: $X_1, \cdots, X_{n_1}$
- B그룹의 관측값: $Y_1, \cdots, Y_{n_2}$
- 가정: A그룹과 B그룹은 중앙값만 다르며, 분포의 모양은 동일하다.
- $N = n_1 + n_2$개의 관측값들을 순서대로 나열해 순위를 작은 것부터 $1 \sim N$으로 매긴다. 이제 순위 변환된 관측값들을 얻는다.
- A그룹의 순위변환된 관측값: $S_1, \cdots, S_{n_1}$
- B그룹의 순위변환된 관측값: $R_1, \cdots, R_n$
- $(X_1, \cdots, X_{n_1}, Y_1, \cdots, Y_{n_2})$가 관측값의 분포에 의존하는 것과는 달리, $(S_1, \cdots, S_{n_1}, R_1, \cdots, R_{n_2})$는 분포에 무관하다.

Wilcoxon Test

- 윌콕슨의 검정통계량은 B그룹의 순위변환된 관측값들의 총합

$$W = \sum_{i=1}^{n_2} R_i$$

  로 정의된다.
- 만약 B그룹 관측값의 중앙값이 크다면, Y_j가 X_i에 비하여 큰 값일 경향성이 존재하며, R_j가 S_i보다 클 것이므로, W가 커지는 경향성을 가진다.
- 만약 B그룹 관측값의 중앙값이 작다면, Y_j가 X_i에 비하여 작은 값일 경향성이 존재하며, R_j가 S_i보다 작을 것이므로, W가 작아지는 경향성을 가진다.

Wilcoxon Test

- 따라서

$$H_0 : \text{처리효과가 존재하지 않는다}, \quad H_1 : \text{처리효과가 존재한다}$$

  에 대하여, 관측된 윌콕슨의 검정통계량 W_{obs}가 극단적인 값이라면 가설을 기각하는 것이 옳다.
- W는 순위변환된 관측값들에 의해 구성되므로, 이 역시 분포에 무관한 검정통계량이다.
- 같은 이유로 W의 분포 역시 (관측값의)분포에 무관하다. 따라서 Wilcoxon test는 분포무관한 검정법이다.

Wilcoxon Test

- 만약 두 집단이 별개의 대상이 아니라 같은 집단의 처리전후라면, X_i와 Y_i는 독립적이지 않다. 또한 $n = n_1 = n_2$여야만 한다.
- 이 경우 $D_i = Y_i - X_i$를 새롭게 정의하고, D_i를 i번째 개체가 경험하는 처리효과라고 생각할 수 있다.
- D_i의 중앙값을 Δ라 할 때, 처리효과의 유무를 검정하는 것은 곧 $\Delta = 0$를 검정하는 것과 동일하다.
- 이때는 $|D_i|$의 대소를 바탕으로 $1 \sim n$으로 순위를 매기고 $R_1, \cdots, R_n$으로 변환한다.
- 그 다음

$$W = \sum_{i=1}^{n} R_i I(D_i > 0)$$

을 검정통계량으로 이용하여 검정할 수 있다.

Wilcoxon Test

- Wilcoxon test는 꼬리가 두터운 분포일 때 검정력이 높음이 알려져 있다.
- 라플라스분포, 코시분포와 같이 정규분포보다 꼬리가 두터울 때 표본의 수에 무관하게 t-test보다 검정력이 높다.
- 꼬리가 얇은 경우 t-test에 비해 검정력은 낮지만, 유의수준을 항상 통제할 수 있다는 장점이 있다.

Example

- 경제학부 A반과 B반에서 10명씩을 뽑아 학생들의 학업성적을 확인하였다.

소속 반		관측값									
A반	순위변환 전	3.5	3.7	4.3	2.2	3.6	4.2	2.3	1.5	2.6	3.1
	순위변환 후	12	14	20	2	13	19	3	1	4	8
B반	순위변환 전	4.1	4.0	3.9	3.8	3.4	3.3	3.2	2.9	2.8	2.7
	순위변환 후	18	17	16	15	11	10	9	7	6	5

- 윌콕슨의 검정통계량은 114이다.
- 영분포 상에서 114의 p값은 0.5288으로, 유의수준보다 크다.
- 따라서 A반과 B반의 성적 차이는 존재하지 않는다.

Example2: M-estimator

Robust Estimator of Mean

- 올림픽 다이빙 경기에서 예술점수는 최고점과 최저점을 제외한 점수들의 평균으로 매겨진다.
- 평가의 주관성은 그 연기가 응당 받아야 할 점수의 범위에 비해 너무 낮거나 높은 점수가 부여될 수 있게 한다.
- 이러한 이상점의 오염으로부터 쉽게 붕괴되지 않는 평균의 추정량이 없을까?
- 다이빙 점수 체계처럼 '중심'으로부터 적당히 먼 관측값들을 적당히 무시하는 방법이 대표적이다.
- 다이빙 경기에서 매기는 점수는 **절사평균**(trimmed mean)으로, 양 극단의 일정 퍼센트에 해당하는 관측값들을 제외하고 매긴 관측값이다.

M-estimator

- 인과추론 연구에서도 마찬가지다. 이상점들이 결과를 완전히 뒤바꾸어 놓을 수 있다.
- 따라서 꼬리가 긴 분포를 자주 다루는 실제 응용상황에서는 처리효과의 추정량으로 **M-추정량**(M-estimator)가 선호된다.
- M-추정량은 어떤 함수 $y = \Psi(x)$에 대하여

$$\hat{\mu}^{\Psi} = \arg\min_{\mu} \sum_{i=1}^{n} \Psi\left(\frac{x_i - \mu}{\tau}\right)$$

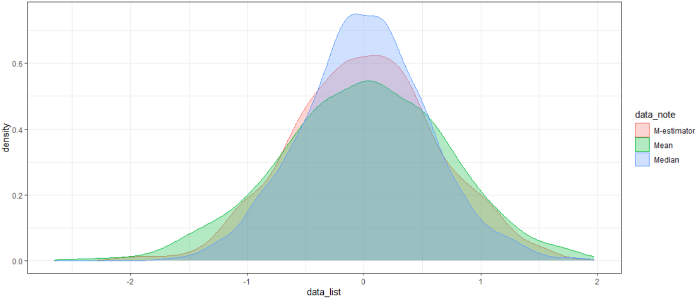
  으로 정의된다.
- $\frac{x_i - \mu}{\tau}$는 표준화된 오차이다.

M-estimator

- 만약 이상점이 등장하면 $\frac{x_i - \mu}{\tau}$의 절댓값이 커지기 마련이다.
- 따라서 Ψ가 극단값에서 작아지게 만든다면 이상점의 영향을 없애줄 수 있다.
- 표본평균은 Ψ를 제곱으로 결정하여 $\hat{\mu}$를 얻어내는 것과 동일하다. 이는 이상점의 영향을 극대화하기 때문에, 로버스트하지 않다.
- 표본중앙값은 Ψ가 절댓값 함수일 때의 평균의 추정량이다. 이상점의 영향이 Ψ가 제곱일 때보다 작기에, 더 로버스트하다.
- 인과추론 연구에서 자주 등장하는 **weighting** 기반의 처리효과 추정 역시 일종의 M-estimator로 볼 수 있다.

Example

- 라플라스 분포(꼬리가 두터운 분포)를 따르는 표본 100개를 생성하여 각 추정량이 얼마나 로버스트한지 확인하였다.
- 여기서 M-estimator는 ψ를 적절히 조정한 Huber의 추정량이다.



- 중앙값이나 M-estimator가 평균보다 실제 평균인 0에 더 가까운 추정량을 내놓음을 확인할 수 있다.

Example3: Bootstrap

Bootstrap

- 우리는 통계량의 표본분포를 알아야 할 때가 있다.
- 예를 들어 정규분포를 따르는 모집단에서 표본을 뽑아 표본평균을 본다.
- 표본평균으로써 모평균의 95퍼센트 신뢰구간을 구하려면, 표본평균의 표본분포를 알고 있어야만 한다.
- 모수적 가정 하에서는 이것이 정규분포를 따르며, 평균은 모평균과 같고, 분산은 모분산을 표본의 크리고 나눈 것임을 안다.
- 비모수적 상황에서는 어떻게 해야 할까?

Bootstrap

- 어떠한 미지의 분포 F에 대하여

$$X_1, \cdots, X_n \sim_{i.i.d} F$$

  를 표본으로 얻었다.
- $\hat{\theta}_n = T(X_1, \cdots, X_n)$은 표본으로부터 얻은 θ의 추정량이다.
- 만약 우리가 시간과 돈을 더 많이 들여서 크기가 n인 표본을 더 많이 만들어낼 수 있다면, 그들의 분포를 관찰하여 표본분포를 알아낼 수 있다.
- 단 실제 상황에서는 접근 가능한 자료가 많지 않다. 표본을 좀 복제할 수 있는 방법이 없을까?

Bootstrap

- 표본은 미지의 모집단 F에서 뽑았다.
- 미지의 모집단 F는 표본으로부터 얻은 **경험적 분포함수**(empirical CDF)

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$$

로써 추정가능하다.
- F_n은 F와 $n^{-1/2}$ 수준의 오차만을 갖는 분포함수의 좋은 추정량이다.
- 즉 모집단 F에서 크기가 n인 표본을 뽑는 것과, 근사된 분포 F_n을 모집단으로 하여 크기가 n인 표본을 뽑는 것은 매우 유사하다.

Bootstrap

- F_n을 잘 들여다보면, 표본에 존재하는 하나의 X_i를 지날 때마다 이산적으로 $1/n$씩 증가하는 계단함수이다.
- 즉 F_n을 모집단으로 하여 얻어낸 **붓스트랩 표본**은 표본 $X_1, \cdots, X_n$에서 n번의 **단순복원추출**을 하여 얻어낸 것과 동일하다.
- 얻어낸 붓스트랩 표본을

$$(X_1^{*(i)}, X_2^{*(i)}, \cdots, X_n^{*(i)}) \quad i = 1, \cdots, B$$

으로 쓰며, 이로써 다시 계산한 추정량이

$$\hat{\theta}_n^{*(i)} = T(X_1^{*(i)}, \cdots, X_n^{*(i)})$$

이다.

Bootstrap

- n번의 단순복원추출을 수행하는 것은 우리가 표본분포를 보고자 하는 추정량 $\hat{\theta}_n$ 역시도 n개의 관측값으로부터 계산된 것이기 때문이다.
- B개의 $\hat{\theta}_n^{*(i)}$들을 얻은 뒤, 그들로써 다시 경험적 분포함수를 구하면 표본분포를 근사할 수 있다. 이를 **붓스트랩 표본분포**라고 한다.
- 마찬가지로, 이 과정에서 $B^{-1/2}$에 비례하는 오차가 발생한다.

Bootstrap

- 붓스트랩을 이용하면...
  1. 편향과 표준오차의 추정
  2. 추정량의 신뢰구간 생성
  3. 붓스트랩 검정
  4. 회귀계수의 붓스트랩 검정
  5. 비모수적 회귀분석에서의 신뢰밴드 생성

  등을 수행할 수 있다.
- 우리에게는 24페이지에서 나온 Algorithm 1과 같은 방식의 bootstrap 을 이해하는 것이 필요하다.
- 이는 limiting distribution을 구하기 위한 **multiplier bootstrap**이다.
- 1 ∼5를 모두 알아보기엔 욕심이다. Algorithm 1을 이해하기 위한 것들만 살짝씩 알아보자!

Some Ideas

- 4)에서 IQR의 비율을 봄으로써 $\hat{\Sigma}^{1/2}$를 얻는다. 분포의 척도모수(산포 or 퍼져 있는 척도)를 추정하고자 할 때, 척도모수가 1인 정규분포의 IQR에 대한 비율을 봄으로써 로버스트한 추정량을 얻는다.
- 5)에서 t-test를 계산하는데, 이는 $\hat{\Sigma}$로써 표준화된 $\hat{R}^*$의 최댓값이라 볼 수 있다. 이는 simultaneous confidence band를 만들기 위하여 계산해준 값이다.(뒤에 설명함)
- 여기에서 사용한 붓스트랩 신뢰구간은 **절대편차 붓스트랩 신뢰구간**의 일종이다.

Bootstrap Confidence Interval

- 우리는 붓스트랩을 통해 추정하려는 모수 θ의 표본분포를 구하려 한다.
- 첫 표본에서 우리는 모수의 추정량 $\hat{\theta}$를 얻는다.
- 붓스트랩 표본에서 얻은 추정량 $\hat{\theta}^{*(i)}$는 실제로 모수가 $\hat{\theta}$인 근사된 모집단 F_n에서 뽑은 표본으로 만들어진다.
- 그렇다면

  $$(\theta, \hat{\theta}) \overset{\text{analogue}}{?} (\hat{\theta}, \hat{\theta}^{*(i)})$$

  이라 생각할 수 있을 것이며, 이에 따라...

  $$\hat{\theta} - \theta \overset{d}{\approx} \hat{\theta}^{*(i)} - \hat{\theta}$$

  임을 추론할 수 있다.

Bootstrap Confidence Interval

- 먼저 편향을 추정해보자.
- $(\hat{\theta}_n^{*(1)}, \cdots, \hat{\theta}_n^{*(B)})$를 얻었다고 하자.
- 이를 크기가 B인 표본이라고 보고 그 평균 $\mu^*_{\hat{\theta}_n}$를 구한다면, 이를 평균의 추정량으로 사용할 수 있다.
- 다르게 쓰면,

$$\begin{aligned}
\text{bias}(\hat{\theta}_n) &= E(\hat{\theta}_n) - \theta \\
&= \mu_{\hat{\theta}_n} - \theta \\
&\approx \mu_{\hat{\theta}_n^*} - \hat{\theta}_n \\
&= \mu_{\hat{\theta}_n^*} - \hat{\theta}_n = \widehat{\text{bias}}_{\text{Boot}}(\hat{\theta}_n)
\end{aligned}$$

Bootstrap Confidence Interval

- 표준오차를 추정해보자.
- $(\hat{\theta}_n^{*(1)}, \cdots, \hat{\theta}_n^{*(B)})$를 얻었다고 하자.
- 이를 크기가 B인 표본이라고 보고 그 표본표준편차 s^*를 구한다면, 이를 표준오차의 추정량으로 사용할 수 있다.
- 즉

$$\hat{\mathrm{se}}_{\mathrm{Boot}}(\hat{\theta}_n) = \left(\frac{1}{B-1} \sum_{i=1}^{B} (\hat{\theta}_n^{*(i)} - \mu_{\hat{\theta}_n^*})^2 \right)^{\frac{1}{2}} \left(\approx \left(\int (\hat{\theta}_n - \mu_{\hat{\theta}_n})^2 dG \right)^{\frac{1}{2}} \right)$$

를 사용한다.

Bootstrap Confidence Interval

- 가장 기초적인 붓스트랩 신뢰구간은 **정규근사 붓스트랩 신뢰구간**(normal-approximated bootstrap C.I)이다.
- n이 충분히 크거나 분포가 대칭이면 중심극한정리에 의하여 많은 추정량이 정규분포로 수렴한다.
- 따라서 n이 크다면

$$\frac{\hat{\theta}_n - \text{bias}(\hat{\theta}_n) - \theta}{\text{se}(\hat{\theta}_n)} \dot{\sim} N(0,1)$$

  이다.
- 편향과 표준오차를 추정량으로 대체하면 아래의 신뢰구간을 얻는다.

$$\theta \in \hat{\theta}_n - \widehat{\text{bias}}_{\text{Boot}}(\hat{\theta}_n) \pm z_{\alpha/2}\widehat{\text{se}}_{\text{Boot}}(\hat{\theta}_n)$$

Bootstrap Confidence Interval

- 가장 간단한 붓스트랩 신뢰구간은 **bootstrap percentile C.I**이다.
- 붓스트랩 표본분포 G^*에 대하여, 이는 그 분위수함수 $Q_{G^*}(p)$에 대하여

$$\theta \in \left(Q_{G^*}\left(\frac{\alpha}{2}\right), Q_{G^*}\left(1 - \frac{\alpha}{2}\right)\right)$$

  으로 주어진다.
- 즉 추정량의 붓스트랩 표본분포 하에서 위와 아래의 $\alpha/2$만큼을 잘라낸 것과 같다.
- 이는 근사적으로 포함확률을 $1 - \alpha$ 이상으로 한다. 포함확률의 오차는 $o(n^{-1})$ 수준임이 알려져 있다.

Bootstrap Confidence Interval

- 우리가 구하려는 ATT는 실수가 아니라 함수이다.
- 따라서 일반적인 신뢰구간이 아니라 신뢰밴드를 구성해주는 것이 좋다.
- 함수 f에 대해서도 유사하게,

$$\hat{f}(x) - f(x) \overset{d}{\approx} \hat{f}^{*(i)}(x) - \hat{f}(x)$$

  임을 이용해줄 수 있다.
- 신뢰밴드에서는 전체 f의 포함확률이 $1 - \alpha$ 이상으로 근사되어야 하므로,

$$\Delta = \sup_x |\hat{f}(x) - f(x)| \overset{d}{\approx} \sup_x |\hat{f}^{*(i)}(x) - \hat{f}(x)| = \Delta^{*(i)}$$

  임을 이용할 것이다.
- 우리가 5)에서 구하는 t-test가 표준화된 이녀석이라고 생각하면 된다.

Bootstrap Confidence Interval

- 그렇다면 $\Delta \geq 0$이 항상 성립하므로, bootstrap percentile C.I를 모방하면
$$\Delta \in (0, Q_{\Delta^*}(1-\alpha))$$
를 신뢰구간으로 얻을 수 있다. Δ^*는 $\Delta^{*(i)}$의 경험분포이다.
- 그렇다면 결과적으로 얻는 신뢰밴드는
$$f(x) \in \hat{f}(x) \pm Q_{\Delta^*}(1-\alpha)$$
의 형태이다.
- 6)에서 최종적으로 구하는 신뢰밴드도 이 과정을 그들의 언어로... 잘 표현해둔 것이나 마찬가지다.
- Cor1은 이렇게 얻은 붓스트랩 신뢰구간의 포함확률이 근사적으로 $1-\alpha$로 통제된다고 주장한다.

Bootstrap Confidence Interval

- 이외에도 다양한 형태의 붓스트랩 신뢰구간이 존재한다.
- 정규근사 없이도 **피벗**의 존재 가정 하에 구할 수 있는 **표준 붓스트랩 신뢰구간**
- 스튜던트화된 피벗을 이용하는 **붓스트랩-t 신뢰구간**
- **편향 조정 및 가속 신뢰구간**(BCa C.I)
- 밀도함수 추정과 비모수적 회귀분석에서 사용하는 **붓스트랩-플러그인 신뢰구간**