추정

CE730 통계와 금융

조남운

목차

- 표본분포
- 추정
    - 추정에 대한 기본 개념
    - 추정방법

표본분포

Namun Cho/ mailto:namun@snu.ac.kr

표본분포
Sampling Distribution

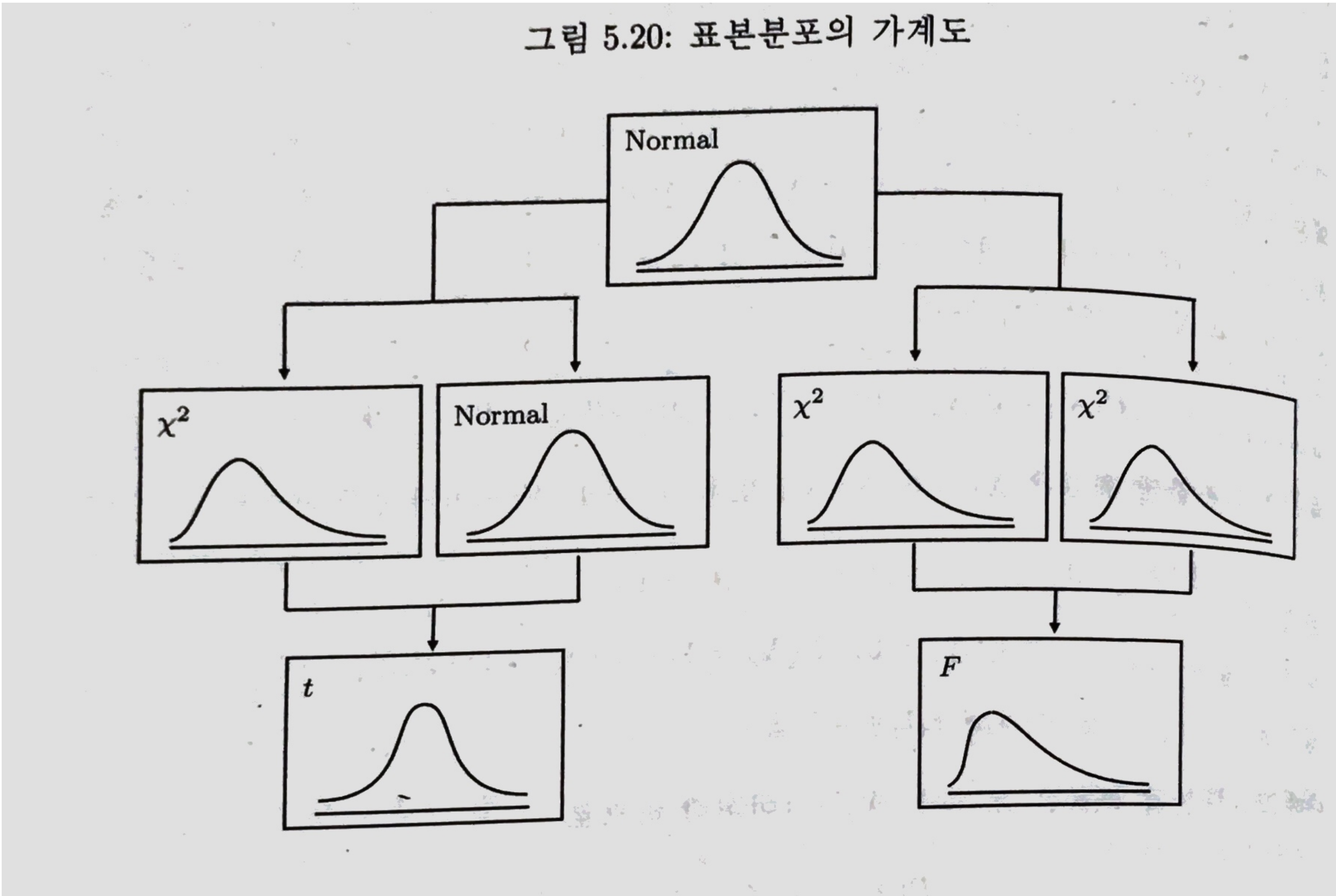
- 표본으로 만들어진 통계량의 분포
  - 표본평균의 분포 $\bar{X} := \sum_i^n X_i/n$
  - 표본분산의 분포 $s^2 := \sum_i^n (X_i - \bar{X})^2/(n-1)$
  - 표본공분산의 분포 $Cov(X, Y) := \sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})/n$
  - 등등
- CLT는 이 중 표본평균의 분포에 대한 이론

표본도 확률분포를 가질 수 있다

- 동전 두 개 던져서 앞면 갯수 맞추기
  - 이 사건의 확률분포는
    - 0 → 1/4, 1 → 1/2, 2 → 1/4
- 표본은 관측 전후에 따라 의미가 달라짐
  - 표본(표집전): 확률분포를 가지는 확률변수
  - 표본(표집후): 정해진 값을 가진 관측치 (확률변수가 아님)

표본분포의 가계도

도출



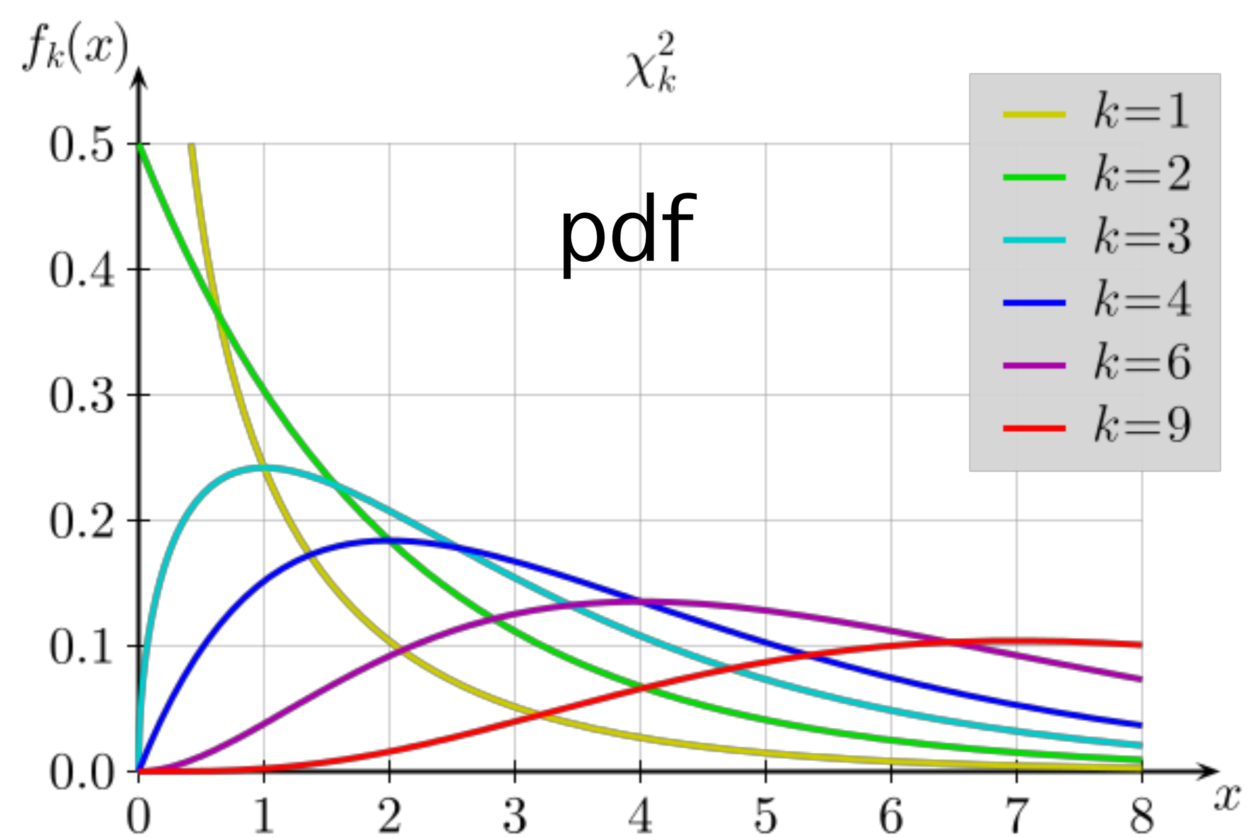
수렴

카이제곱분포

- 표준정규분포를 따르는 확률변수의 제곱합의 분포
- 모수: k
  - df (degree of freedom)
  - 자유도

$$Z_i \sim iidN(0,1^2), \quad i = 1,2,\cdots,k$$

$$Q = \sum_{i}^{k} Z_i^2 \sim \chi^2(k)$$



카이제곱분포: 주요성질

$$Z \sim N(0,1)$$
$$X \sim \chi^2(n_1)$$
$$Y \sim \chi^2(n_2)$$
$$Cov(Z,X) = 0$$
$$Cov(X,Y) = 0$$

C1 $Z^2 \sim \chi^2(1)$

C2 $X + Y \sim \chi^2(n_1 + n_2)$

C3 $\dfrac{Z}{\sqrt{X/n_1}} \sim t(n_1)$

C4 $\dfrac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$

표본분포

C1 $Z^2 \sim \chi^2(1)$

C2 $X + Y \sim \chi^2(n_1 + n_2)$

C3 $\dfrac{Z}{\sqrt{X/n_1}} \sim t(n_1)$

C4 $\dfrac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$

$$X_i \sim iidN(\mu_x, \sigma_x^2), \quad i = 1,2,\cdots,n$$

(1) $$\sum_i^n \left(\frac{X_i - \mu_x}{\sigma_x} \right)^2 \sim \chi^2(n)$$

(2) $$\sum_i^n \left(\frac{X_i - \bar{X}}{\sigma_x} \right)^2 = \frac{(n-1)s_x^2}{\sigma_x^2} \sim \chi^2(n-1)$$

Namun Cho/ mailto:namun@snu.ac.kr

C1 $Z^2 \sim \chi^2(1)$

C2 $X + Y \sim \chi^2(n_1 + n_2)$

C3 $\dfrac{Z}{\sqrt{X/n_1}} \sim t(n_1)$

C4 $\dfrac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$

$$\bar{X} \sim N(\mu_x, \sigma_x^2/n)$$

$$\frac{\bar{X} - \mu_x}{\sqrt{\sigma_x^2/n}} \sim N(0,1)$$

(3) $$\frac{\dfrac{\bar{X} - \mu_x}{\sqrt{\sigma_x^2/n}}}{\sqrt{\dfrac{(n-1)s_x^2}{\sigma_x^2}/(n-1)}} = \frac{\sqrt{n}(\bar{X} - \mu)}{s_x} \sim t(n-1)$$

Namun Cho/ mailto:namun@snu.ac.kr

(4)

$$X_i \sim iidN(\mu_x, \sigma_x^2), \quad i = 1,2,\cdots,n$$

$$Y_i \sim iidN(\mu_y, \sigma_y^2), \quad i = 1,2,\cdots,m$$

C1 $\quad Z^2 \sim \chi^2(1)$

C2 $\quad X + Y \sim \chi^2(n_1 + n_2)$

C3 $\quad \dfrac{Z}{\sqrt{X/n_1}} \sim t(n_1)$

C4 $\quad \dfrac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$

$$s_y^2 := \sum_i^m (Y_i - \bar{Y})^2/(m-1)$$

$$\frac{(m-1)s_y^2}{\sigma_y^2} \sim \chi^2(m-1)$$

(4) $\quad \dfrac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2} \sim F(n-1, m-1)$

자유도
Degree of Freedom

- 값이 변동 가능한 표본의 갯수
- 통상적으로 표본이 n개라면 자유도도 n임
  - n개의 표본이 동시에 달라져도 상관 없기 때문
- 하지만 표본분산의 자유도는 n-1임
  - 모평균 대신 표본평균을 사용 ⇒ 표본평균값이 존재하기 때문에 n개의 표본 중 1개 표본값은 n-1개의 표본값이 정해지면 자동 결정됨 ⇒ df=n-1

일반화된 자유도의 규칙

- 어떤 통계량 계산을 위해 n개의 표본을 사용
- 통계량 계산을 위해 표본으로부터 미리 추정한 모수의 추정량이 k개인 경우
- df = n-k

모집단 분포가 정규분포가 아닐 경우의 표본분포

$$Y \sim B(n, p)$$

$$Y/n \sim ?$$

$$\frac{Y/n - E(Y/n)}{\sqrt{Var(Y/n)}} = \frac{Y/n - p}{\sqrt{p(1-p)/n}} \xrightarrow{d} N(0,1)$$

$$Y/n \sim N(p, p(1-p)/n)$$

- CLT를 이용하여 근사적으로 표본분포 도출
- n이 정규분포에 근사할 정도로 충분히 커야함
- 모집단 분포를 모르고 n도 충분히 크지 않다면 비모수적 방법을 사용해야 함
  - Jackknifing, Bootstrapping 등

과제2: CLT 검증

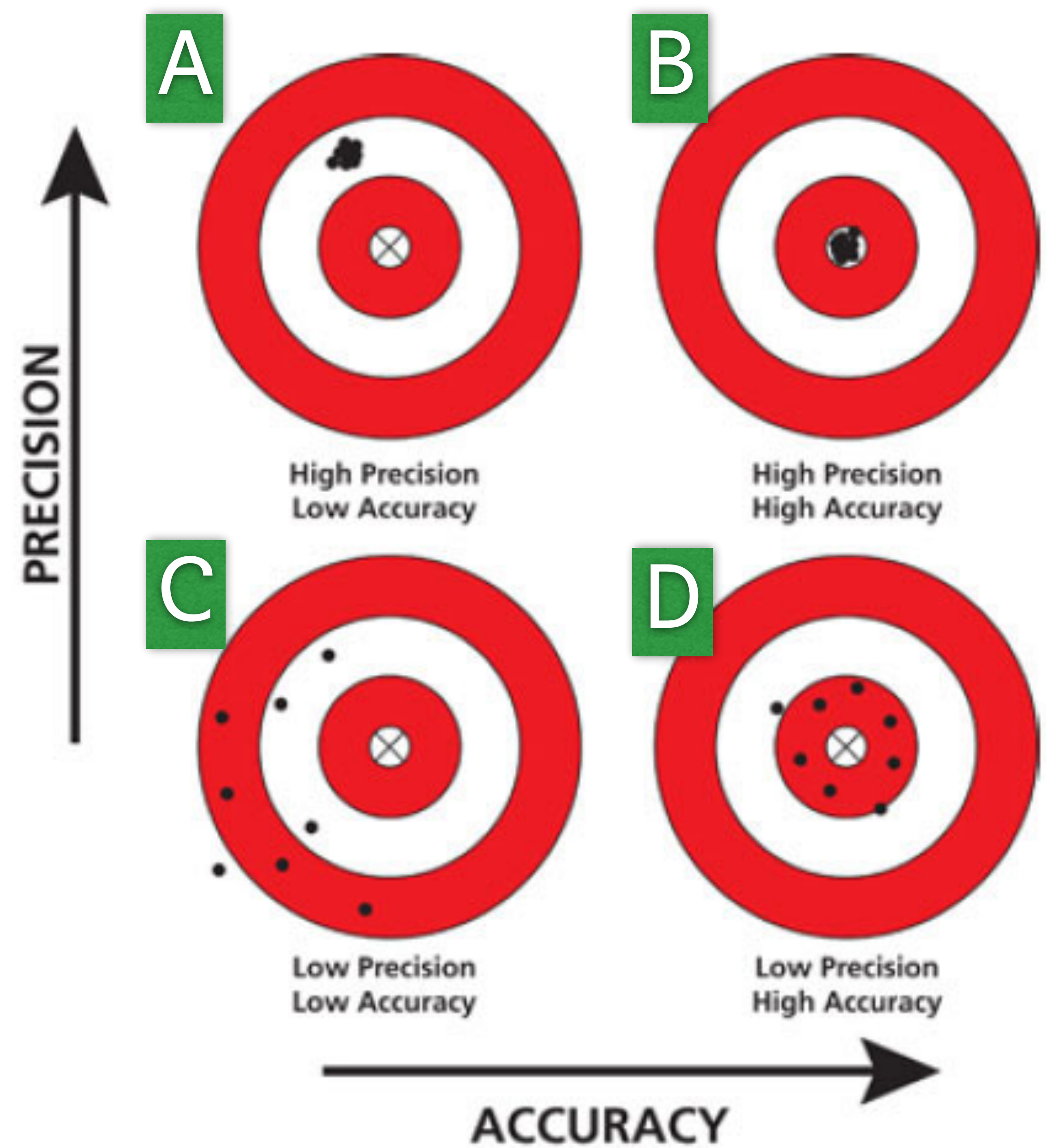
- U(0,100)을 따르는 데이터 10000개 생성: 모집단으로 간주
- 이 모집단에서 1,10,100 의 크기로 이루어진 표본을 반복 표집
  - 각 표본평균 1개 10개, 50개, 500개, 5000개의 분포를 히스토그램으로 그리기
  - CLT가 맞다면 표본평균의 분포는 정규분포에 수렴할 것임
  - Bonus: 표본분산 10, 50, 500, 5000개의 분포도 그려보자
- rmarkdown으로 작성하여 제출할 것
  - 템플릿 파일은 lms에 게시. 이 템플릿 파일의 내용을 채워 제출할 것
  - R markdown은 이 링크에서 다운로드
    - https://rmarkdown.rstudio.com
- 기한: 11/30 (금) 15:00

추정
Estimation

Namun Cho/ mailto:namun@snu.ac.kr

정확도, 정밀도, 분산
Accuracy, Precision, Variance

- 정확도(Accuracy): 참값과 추정값(착탄점의 평균)의 차이
- 정밀도(Precision): 착탄점의 분포된 정도
  - Variance를 여기에 쓰는 경우도 있음
- 좋은 추정량: 정확도와 정밀도가 높은 추정량
- 지표: MSE (Mean Squared Error)



MSE: Definition
Mean Squared Error

- hat theta: 모수 theta에 대한 추정량 (확률변수)
- 모수 theta: 상수
- E(hat theta): 상수
- 좋은 추정량의 가장 중요한 성질은 Bias가 없는 것
  - 불편성 (unbiasedness)
  - 일치성 (consistency)
  - 효율성 (efficiency))
  - 충분성 (sufficiency)

$$MSE(\hat{\theta}) := E(\hat{\theta} - \theta)^2$$

$$= E(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2$$

$$= E(\hat{\theta} - E(\hat{\theta}))^2 + (E(\hat{\theta}) - \theta)^2$$

$$= Var(\hat{\theta}) + Bias^2$$

$$Bias := E(\hat{\theta}) - \theta$$

$$[E(E(X)) = E(X)]$$

불편성
Unbiasedness

- hat theta: 추정량
- theta: 모수
- Bias=0 인 추정량은 불편성을 만족함
  - 정확도를 보장함을 의미
- 불편성을 보장하는 분산이 가장 작은 추정량: UMVUE (Uniformly Minimum Variance Unbiased Estimator)

$$E(\hat{\theta}) = \theta \Rightarrow Bias = 0$$

불편추정량의 예

- 모평균이 mu인 모집단에서 추출된 임의표본의 평균인 bar X는 mu의 불편추정량
- Xi 가 성공확률 p 인 베르누이 시행이라면 hat p 는 p에 대한 불편주정량

$$E(\bar{X}) = \frac{1}{n} E\left(\sum_{i}^{n} X_i \right) = n\mu/n = \mu$$

일치성
Consistency

- hat theta가 n표본으로부터 계산된 경우
- n이 증가함에 따라 hat theta가 모수 theta에 수렴할 경우 hat theta는 theta에 대한 일치추정량이라고 정의
- hat theta의 MSE가 0에 수렴하면 hat theta는 theta의 일치주정량

$\hat{\theta}_n$ is consistent estimator of θ if:

$$\operatorname*{plim}_{n\to\infty} E(\hat{\theta}_n - \theta) = 0$$

$$\forall \epsilon > 0, \quad P(|\hat{\theta}_n - \theta| > \epsilon) \to 0$$

표준오차
Standard Error

$$Var(\hat{\theta}_n) := E(\hat{\theta}_n - E(\hat{\theta}_n))^2$$

$$SE := \sqrt{Var(\hat{\theta}_n)}$$

- 표준오차의 정의
  - 추정량의 표준편차

불편추정량이지만 일치성을 만족하지 못하는 경우

	불편성	일치성
hat theta1	만족	불만족
hat theta2	만족	만족

$$X_i \sim iid?(\mu, \sigma^2)$$

$$\hat{\theta}_1 := (X_1 + X_n)/2$$

$$\hat{\theta}_2 := \bar{X} = \sum_i^n X_i/n$$

(mu에 대한 추정량임)

$$MSE(\hat{\theta}_1) \to \sigma^2/2$$

Namun Cho/ mailto:namun@snu.ac.kr

불편성은 만족하지 않지만 일치추정량인 경우

$$X_i \sim iid?(\mu, \sigma^2)$$

$$\hat{\theta}_3 := \sum_i^n X_i/(n+1)$$

$$E(\hat{\theta}_3) = n\mu/(n+1) \neq \mu$$

Biased!

$$Var(\hat{\theta}_3) = (\sigma^2/n)[n^2/(n+1)^2] = n\sigma^2/(n+1)^2$$

$$MSE(\hat{\theta}_3) = Var(\hat{\theta}_3) + (E(\hat{\theta}_3) - \mu)^2$$

$$= \frac{n}{(n+1)^2}\sigma^2 + \left(\frac{-\mu}{n+1}\right)^2 \to 0$$

as $n \to \infty$

Consistent!

효율성
Efficiency

- 모수를 추정하기 위해 동일한 크기의 표본을 이용한 추정량들 중 가장 작은 MSE를 가진 추정량
- 효율적 추정량은 MVUE임
  - Minimum Variance Unbiased Estimator

$$\hat{\theta}_1 := \frac{X_1 + X_2 + \cdots + X_{n-1} + 2X_n}{n+1},$$

$$\hat{\theta}_2 := \bar{X}$$

$$E(\hat{\theta}_1) = E(\hat{\theta}_2) = \mu$$

Unbiased!

$$MSE(\hat{\theta}_1) = \frac{\sigma^2(n-1+4)}{(n+1)^2}$$

Consistent!

$$MSE(\hat{\theta}_2) = \sigma^2/n$$

Consistent!

$$MSE(\hat{\theta}_1) - MSE(\hat{\theta}_2) = \sigma^2 \frac{n-1}{n(n-1)^2} > 0$$

theta2 is more efficient

충분성
Sufficiency

- T가 θ의 충분통계량일 조건
  - T=t로 주어져 있을 때 표본의 조건부 분포함수가 모수 θ 에 의존하지 않을 것
- T가 모수 θ 에 대한 정보를 모두 가지고 있다는 의미

$f(x_1, \cdots, x_n; \theta)$ 결합분포함수

$X_1, \cdots, X_n$

f의 분포로부터 구한 표본

θ 모수

$f(X_1, \cdots, X_n \,|\, T = t)$

통계량 T = t로 주어졌을 때 표본의 조건부 분포함수

좋은 통계량의 조건
Good Estimator

- 불편성 Unbiasedness
- 일치성 Consistency
- 효율성 Efficiency
- 충분성 Sufficiency

추정 방법

Namun Cho/ mailto:namun@snu.ac.kr

추정 방법의 종류

- 점 추정법 Point Estimation
  - 적률 추정 Moment Estimation
  - 최우추정 Maximimum Likelihood Estimation
  - 최소제곱추정 Least Squared Estimation
- 구간 추정법 Interval Estimation

적률추정
Moment Estimation

- k차 적률
  - kth order moment
  - 평균은 1차적률
  - 분산은 2차적률
  - 왜도는 3차적률
  - 첨도는 4차적률
- 모수인 k차 적률에 대한 추정량인 k차 표본적률은 대체로 좋은 추정량임

k차 적률:= $E(X^k)$ (모수)

k차 표본적률:= $\frac{1}{n}\sum_{i}^{n} X_i^k$

(추정량)

Namun Cho/ mailto:namun@snu.ac.kr

최우추정량 MLE: Maximum Likelihood Estimation

- 우도함수(Likelihood function)를 극대화하는 추정량
- 관측된 표본값을 모수로 가지는 θ(vector)의 함수
- 여전히 결합분포함수이므로 이 함수값이 높다는 것은 발생확률이 높다는 것을 의미함
- MLE의 아이디어: 우도함수를 극대화하는 모수를 찾는 방식으로 추정량 결정
  - Xi = xi 로 관측될 가능성을 가장 크게 만드는 모수 θ 를 찾는 것

$X_1, X_2, \cdots, X_n$ 표본(관측전: 확률변수)

$f(x_1, \cdots, x_n; \theta)$ 표본의 결합분포함수

$x_1, x_2, \cdots, x_n$ 표본값(관측후: 일반변수)

$L(\theta; x_1, \cdots, x_n)$ 우도함수

$\arg\max_{\theta} L(\theta; x_1, \cdots, x_n)$ MLE

f(x,y;a,b,c) and g(a,b,c;x,y): Exercise

$$f(x, y; a, b, c) := ax + bxy + cy + 1$$

x,y의 함수. a,b,c는 모수(parameter)

$$g(a, b, c; x, y) := ax + bxy + cy + 1$$

a,b,c의 함수. x,y는 모수(parameter)

다변수 함수의 극대화

Namun Cho/ mailto:namun@snu.ac.kr

다변수 함수의 미분

- MLE: 극대화문제의 일종
- 우도함수는 다변수함수이므로 MLE는 다변수함수의 극대화문제를 풀 수 있어야 함
- 다변수함수: input, 또는 output이 벡터인 함수
  - MLE에서는 input이 다변수인 경우만 다룸
  - Maximum을 찾는 문제는 다변수의 경우 불가능

전체구조

- 편미분
- 전미분
- 극대화문제의 1계조건
- 극대화문제의 2계조건

Partial Derivative

Let $f : D \in \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{e_i}$ be a vector whose i th element is 1 and others are 0.

$$\mathbf{e_i} := (\overbrace{0, 0, \cdots, 0, 1}^{i}, 0, \cdots, 0)$$

Definition (Partial Derivative)

<u>Partial derivative</u> at $\bar{\mathbf{x_0}} \in D$ is

$$\frac{\partial f}{\partial x_i} := \lim_{h \to 0} \frac{f(\overline{\mathbf{x_0}} + h\mathbf{e_i}) - f(\overline{\mathbf{x_0}})}{h}$$

When $n = 1$, partial derivative is equivalent to derivative of one variable function.

Calculation Procedure

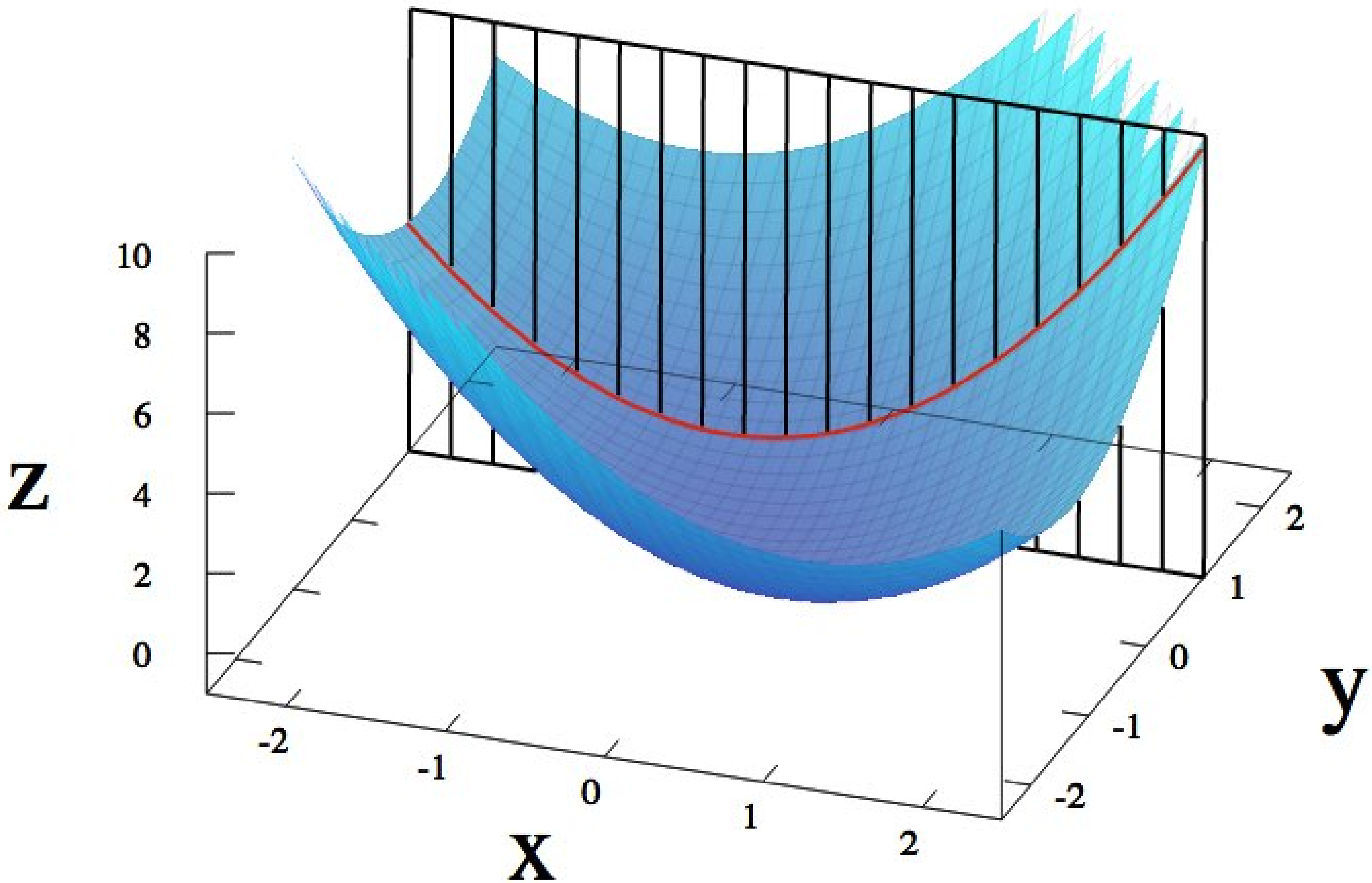
- Treat x_i as the only variable in f
- Treat x_{-i} as constant

Partial Derivative: Geometric Interpretation

$f : \mathbb{R}^2 \to \mathbb{R}$

- Think of $f(\mathbf{x}) = x_1^2 + x_2^2$.
- If $x_2 = \bar{x}_2$, $f(x_1, \bar{x}_2)$ is equivalent to one variable function $\tilde{f}(x_1) = x_1^2 + \bar{x}_2^2$.
- Graph of $\tilde{f}$ is intersection of the graph of f with the slice $x_2 = \bar{x}_2$.
- $\frac{\partial f}{\partial x_1}(\bar{x}_1, \bar{x}_2) = \frac{\partial \tilde{f}}{\partial x_1}(\bar{x}_1)$ is the slope of $\tilde{f}$ on $\bar{x}_1$, slope of the tangent line to the curve $\tilde{f}$ (on the plane $x_2 = \bar{x}_2$)

Example



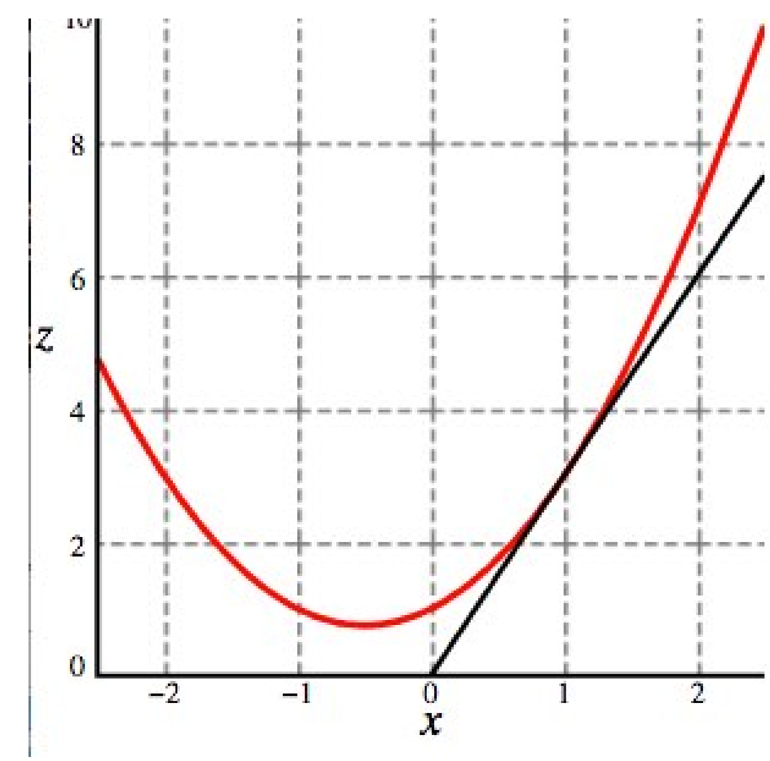


Figure: Graph of $z = x^2 + xy + y^2$ with intersection $y = 1$

Finding Tangent Plane

Let $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable. When finding a tangent plane on $(\bar{x}_1, \bar{x}_2)$, we need to get at least two independent vectors: $\left(1, 0, \frac{\partial f}{\partial x_1}(\bar{x}_1, \bar{x}_2)\right)$ (slice $x_2 = \bar{x}_2$), and $\left(0, 1, \frac{\partial f}{\partial x_2}(\bar{x}_1, \bar{x}_2)\right)$ (slice $x_1 = \bar{x}_1$)
Then the tangent plane with two parameters $\Delta x_1, \Delta x_2$ is:

$$(\bar{x_1}, \bar{x_2}, f(\bar{\mathbf{x}})) + \Delta x_1 \left(1, 0, \frac{\partial f}{\partial x_1}(\bar{x}_1, \bar{x}_2)\right) + \Delta x_2 \left(0, 1, \frac{\partial f}{\partial x_2}(\bar{x}_1, \bar{x}_2)\right)$$

$$= \left(\bar{x_1} + \Delta x_1, \bar{x_2} + \Delta x_2, f(\bar{\mathbf{x}}) + \frac{\partial f}{\partial x_1}(\bar{x}_1, \bar{x}_2)\Delta x_1 + \frac{\partial f}{\partial x_2}(\bar{x}_1, \bar{x}_2)\Delta x_2\right)$$

This interpretation can be extended to n dimension.

The Total Derivative

Changes in All Direction: $f : \mathbb{R}^n \to \mathbb{R}^1$

Let $d\mathbf{x} = (dx_1, \cdots, dx_n)$ and $f : \mathbb{R}^n \to \mathbb{R}$, differentiable. Then small change of $d\mathbf{x}$ will cause small change of $df = f(\bar{\mathbf{x}} + d\bar{\mathbf{x}}) - f(\bar{\mathbf{x}}) \in \mathbb{R}$ and

$$df = f(\bar{\mathbf{x}} + d\bar{\mathbf{x}}) - f(\bar{\mathbf{x}}) = \frac{\partial f}{\partial x_1}(\bar{\mathbf{x}})dx_1 + \cdots + \frac{\partial f}{\partial x_n}(\bar{\mathbf{x}})dx_n = Df_{\mathbf{x}}d\mathbf{x}$$

And $Df_{\mathbf{x}} := \begin{pmatrix} \frac{\partial f}{\partial x_1}(\bar{\mathbf{x}}) & \frac{\partial f}{\partial x_2}(\bar{\mathbf{x}}) & \cdots & \frac{\partial f}{\partial x_n}(\bar{\mathbf{x}}) \end{pmatrix}$: (Jacobian) derivative of f at $\bar{\mathbf{x}}$ or The linear approximation of f at $\bar{\mathbf{x}}$, or Gradient vector ∇f

Note: In this case, $Df_{\mathbf{x}}$ is a vector or $1 \times n$ matrix.

다변수 함수의 미분: 결론

$$Df_{\mathbf{x}} := \left(\frac{\partial f}{\partial x_1}(\bar{\mathbf{x}}) \quad \frac{\partial f}{\partial x_2}(\bar{\mathbf{x}}) \quad \cdots \quad \frac{\partial f}{\partial x_n}(\bar{\mathbf{x}}) \right)$$

2계미분은 n×n 행렬이 됨

Hessian

Definition

Hessian matrix

$$D^2 f_{\mathbf{x}} = D(Df)_{\mathbf{x}} := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

Theorem (14.5:Young's theorem)

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \quad \forall i, j$$

This means hessian is symmetric.

다변수함수의 극대화문제: 극대화문제의 정의

Definitions

Definition ((strict) max/min, (strict) local max/min)

Let $f : U \in \mathbb{R}^n \rightarrow \mathbb{R}$

1. *A point* $\mathbf{x}^*$ *is a (global, or absolute) max, maximizer, maximum point of* f *on* U *if* $f(\mathbf{x}^*) \geq f(\mathbf{x}) \quad \forall \mathbf{x} \in U$
2. $\mathbf{x}^* \in U$ *is a strict (global, or absolute) max if* $\mathbf{x}^*$ *is a max and* $f(\mathbf{x}^*) > f(\mathbf{x}) \quad \forall \mathbf{x} \in U - \{\mathbf{x}^*\}$
3. $\mathbf{x}^* \in U$ *is a local (relative) max of* f *if* $\exists \epsilon > 0 \quad s.t.$ $f(\mathbf{x}^*) \geq f(\mathbf{x}) \quad \forall \mathbf{x} \in B_\epsilon(\mathbf{x}^*) \cap U$
4. $\mathbf{x}^* \in U$ *is a strict local (relative) max of* f *if* $\exists \epsilon > 0 \quad s.t.$ $f(\mathbf{x}^*) > f(\mathbf{x}) \quad \forall \mathbf{x} \in B_\epsilon(\mathbf{x}^*) \cap U - \{\mathbf{x}^*\}$

- Definition of min: $>, \geq \rightarrow <, \leq$

FOC

Theorem (17.1)

Let $f : U \in \mathbb{R}^n \to \mathbb{R}$ be a C^1 function. If $\mathbf{x}^*$ is a local max or min of f and $\mathbf{x}^*$ is an interior point of U, then

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0 \quad \forall i$$

In short,

$$Df_{\mathbf{x}}(\mathbf{x}^*) = \mathbf{0}$$

$\mathbf{x}^*$ is a critical point of f

Note: Compare with one-var version FOC (Theorem 3.3)

Theorem (3.3: First Order Condition (FOC))

x_0 is an interior max or min of $f \Rightarrow x_0$ is a critical point of f. $i.e.$, $f'(x_0) = 0$ (Inverse is not always true)

SOC (Sufficient Conditions)

Theorem (17.2)

Let $f : U \in \mathbb{R}^n \to \mathbb{R}$ be a C^2 function and U is open. Suppose $\mathbf{x}^*$ is a critical point of f. $(i.e., Df_{\mathbf{x}}(\mathbf{x}^*) = \mathbf{0})$ Then,

1. *If Hessian $(D^2 f_{\mathbf{x}}(\mathbf{x}^*))$ is ND, then $\mathbf{x}^*$ is a strict local max of f*
2. *If Hessian $(D^2 f_{\mathbf{x}}(\mathbf{x}^*))$ is PD, then $\mathbf{x}^*$ is a strict local min of f*
3. *If Hessian is ID, $\mathbf{x}^*$ is neither a local max nor local min of f. (saddle point)*

Note: one-var version: (Theorem 3.4)

$$f'(x^*) = 0 \quad \wedge \quad f'' < 0 \quad \Rightarrow \quad x^* \text{ is a local max}$$

Definiteness: Overview

When $Q = \mathbf{x}^T A \mathbf{x}$ and A is a diagonal matrix

$$A = \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

- Positive Definite (PD): $a_{ii} > 0 \quad \forall i$
- Positive Semi Definite (PSD): $a_{ii} \geq 0 \quad \forall i$
- Negative Definite (ND): $a_{ii} < 0 \quad \forall i$
- Negative Semi Definite (NSD): $a_{ii} \leq 0 \quad \forall i$
- Indefinite (ID): $a_{ii} < 0$ for some i, and $a_{ii} > 0$ for some i

SOC (Necessary Conditions)

Theorem (17.6)

Let $f : U \in \mathbb{R}^n \to \mathbb{R}$ be a C^2 function and U is open. Then,

1. $\mathbf{x}^*$ *is a local min of* $f \Rightarrow Df(\mathbf{x}^*) = \mathbf{0} \quad \wedge \quad D^2 f(\mathbf{x}^*)$ *is PSD*
2. $\mathbf{x}^*$ *is a local max of* $f \Rightarrow Df(\mathbf{x}^*) = \mathbf{0} \quad \wedge \quad D^2 f(\mathbf{x}^*)$ *is NSD*

Note: one-var version:

$$x^* \text{ is local max} \quad \Rightarrow \quad x' = 0 \quad \wedge \quad f'' \leq 0$$

Finding Global Max/Min

Different from one-var function, condition 1 (below) is not true when f is multi-var function

Sufficient Conditions for Global Max/Min ($f : I \in \mathbb{R} \to \mathbb{R}$)

1. x^* is a local max/min and x^* is the only critical point of f in I
2. $f'' \leq 0 \quad \forall I.\ i.e.,\ f$ is concave on I (max)
   - $f'' \leq 0 \quad \forall I$ (max)
   - $f'' \geq 0 \quad \forall I$ (min)

However, condition 2 is true even when f is multi-var function!

Theorem (17.8)

Let $f : U \in \mathbb{R}^n \to \mathbb{R}$ be a C^2 function with convex open domain U.

1. $DF(\mathbf{x}^*) = \mathbf{0}$ *and* $D^2 f_{\mathbf{x}}$ *is PSD on* $U \Rightarrow \mathbf{x}^*$ *is a global min of* f *on* U
2. $DF(\mathbf{x}^*) = \mathbf{0}$ *and* $D^2 f_{\mathbf{x}}$ *is NSD on* $U \Rightarrow \mathbf{x}^*$ *is a global max of* f *on* U

로그 우도 함수
Log Likelihood Function

- 우도함수가 지수형일 때에는 로그를 취할 경우 극대화문제를 풀기 수월해질 수 있음
- 이러한 경우 로그우도함수를 사용
- 로그함수는 단조증가함수이므로 로그우도함수의 극대화문제의 해는 우도함수의 극대화문제의 해와 같음

$$\arg\max_{\theta} L(\theta; x_1, \cdots, x_n)$$

$$\Leftrightarrow \arg\max_{\theta} \ln L(\theta; x_1, \cdots, x_n)$$

MLE 연습

- 분포의 형태 (분포함수)는 알고 있지만 모수 (모평균, 모분산)는 모르는 상황
- 모수를 추정해야 함
  - 본 연습에서는 MLE의 방법으로 추정 시도

$$X_1, \cdots X_n \sim iid N(\mu, \sigma^2)$$

$$(X_1, \cdots, X_n) = (x_1, \cdots x_n)$$

$$\Rightarrow \begin{array}{l} \hat{\mu} = ? \\ \hat{\sigma}^2 = ? \end{array}$$

우도함수 유도하기

$$f(x_i;\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$f(x_1,\cdots,x_n;\mu,\sigma^2) = \prod_i^n f(x_i;\mu,\sigma^2)$$

(by iid assumption)

$$= (2\pi\sigma^2)^{-n/2}\exp\left(-\frac{\sum_i^n (x_i-\mu)^2}{2\sigma^2}\right)$$

$$L(\mu,\sigma^2;x_1,\cdots,x_n) = (2\pi\sigma^2)^{-n/2}\exp\left(-\frac{\sum_i^n (x_i-\mu)^2}{2\sigma^2}\right)$$

(Likelihood Function)

로그 우도함수 계산하기

$$L(\mu, \sigma^2; x_1, \cdots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum_i^n (x_i - \mu)^2}{2\sigma^2}\right)$$

(Likelihood Function)

$$\ln L(\mu, \sigma^2; x_1, \cdots, x_n) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_i^n (x_i - \mu)^2$$

(Log-Likelihood Function)

MLE

$$\arg\max_{\mu,\sigma^2} \ln L$$

로그우도함수의 극대화문제

$$D_{\mu,\sigma^2} \ln L = \begin{pmatrix} \dfrac{\partial \ln L}{\partial \mu} & \dfrac{\partial \ln L}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

FOC

$$D^2_{\mu,\sigma^2} \ln L = \begin{pmatrix} \dfrac{\partial^2 \ln L}{\partial^2 \mu} & \dfrac{\partial^2 \ln L}{\partial \mu \partial \sigma^2} \\ \dfrac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} & \dfrac{\partial^2 \ln L}{\partial^2 \sigma^2} \end{pmatrix} \text{ is ND or NSD}$$

SOC

FOC

$$D_{\mu,\sigma^2} \ln L = \begin{pmatrix} \frac{\partial \ln L}{\partial \mu} & \frac{\partial \ln L}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

$$\ln L(\mu, \sigma^2; x_1, \cdots, x_n) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i}^{n}(x_i - \mu)^2$$

(Log-Likelihood Function)

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i}^{n}(x_i - \mu) = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{n}\sum_{i}^{n} x_i$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i}^{n}(x_i - \mu)^2 = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i}^{n}(x_i - \hat{\mu})^2$$

SOC: Passed

$$D^2_{\mu,\sigma^2} \ln L = \begin{pmatrix} \frac{\partial^2 \ln L}{\partial^2 \mu} & \frac{\partial^2 \ln L}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 \ln L}{\partial^2 \sigma^2} \end{pmatrix} \text{ is ND or NSD}$$

SOC

$$\hat{\mu} = \frac{1}{n}\sum_i^n x_i \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_i^n (x_i - \hat{\mu})^2$$

ND

$$D^2_{\mu,\sigma^2} \ln L(\hat{\mu}, \hat{\sigma}^2) = \begin{pmatrix} -\frac{n\hat{\mu}}{\hat{\sigma}^2} & -\frac{1}{\hat{\sigma}^4}\sum_i^n (x_i - \hat{\mu}) \\ -\frac{1}{\hat{\sigma}^4}\sum_i^n (x_i - \hat{\mu}) & \frac{n}{2\hat{\sigma}^4} - \frac{1}{(\hat{\sigma}^2)^3}\left(\sum_i^n (x_i - \hat{\mu})^2\right) \end{pmatrix}$$

$$-\frac{1}{\hat{\sigma}^4}\sum_i^n (x_i - \hat{\mu}) = 0 \qquad \frac{n}{2\hat{\sigma}^4} - \frac{1}{(\hat{\sigma}^2)^3}\left(\sum_i^n (x_i - \hat{\mu})^2\right) = \frac{n}{2\hat{\sigma}^4} - \frac{n\hat{\sigma}^2}{(\hat{\sigma}^2)^3} = -\frac{n}{2\hat{\sigma}^4} < 0$$

구간추정
Interval Estimation

Namun Cho/ mailto:namun@snu.ac.kr

신뢰구간
CI: Confidence Interval

- d% 신뢰구간: 모수를 포함할 확률이 d%인 구간
  - 구간의 크기는 작을수록 유용함
- d: 신뢰도
  - 1-d: 유의수준
  - 높을 수록 좋음
  - Trade-off: 신뢰도가 높아지면 구간이 커짐
- 분야마다 통용되는 신뢰도는 다름
  - 대체로 사회과학은 95% 정도
  - 일부 자연과학분야에서는 1-10e-6 (99.999%)을 사용하기도 함

CI 구하기

- 표본평균에 대한 신뢰구간
  - 모평균, 모분산을 아는 경우
  - 모분산을 모르는 경우
- 평균차에 대한 신뢰구간
  - 분산을 아는 경우
  - 분산은 모르지만 두 분산이 같을 경우
- 분산에 관한 신뢰구간
- 분산비에 관한 신뢰구간

표본평균의CI:
모평균, 모분산을 아는 경우

$$X_1, \cdots X_n \sim iid(\mu, \sigma^2)$$

$$\bar{X} \sim (\mu, \sigma^2/n)$$

CLT

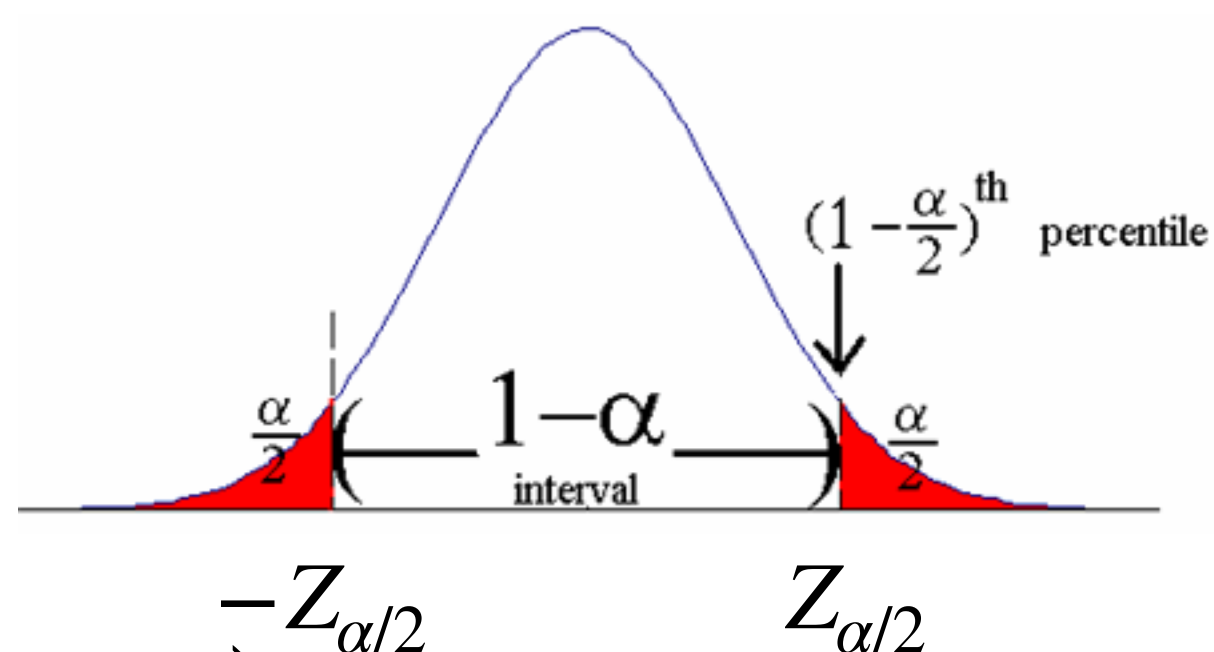
$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Z Transformation

1-α 신뢰구간

$$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < Z_{\alpha/2}\right) = 1-\alpha$$

$$= P\left(-Z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}\right)$$



$$= P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

1-α CI

μ 가 CI 안에 있을 확률: 1-α

Namun Cho/ mailto:namun@snu.ac.kr

모분산을 모르는 경우

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- 모분산 σ^2 대신 표본분산 s^2 를 사용
- 이때 표본평균의 분포는 자유도 n-1의 t분포를 따름

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

평균차 신뢰구간: 모분산이 모두 알려져 있는 경우

서로 독립

$$X_1, \cdots, X_n \sim iid(\mu_1, \sigma_1^2) \qquad Y_1, \cdots, Y_m \sim iid(\mu_2, \sigma_2^2)$$

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1)$$

$(\mu_1 - \mu_2)$ 의 1-α CI $\qquad \bar{X} - \bar{Y} \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$

평균차 신뢰구간: 분산은 모르지만 동분산일 경우

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

서로 독립

$$X_1, \cdots, X_n \sim iid(\mu_1, \sigma^2) \qquad Y_1, \cdots, Y_m \sim iid(\mu_2, \sigma^2)$$

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim N(0,1)$$

$$s_x^2 := \sum_i^n (X_i - \bar{X})^2/(n-1) \qquad s_y^2 := \sum_i^m (Y_i - \bar{Y})^2/(m-1)$$

$$\frac{(n-1)}{\sigma^2} s_x^2 \sim \chi^2(n-1) \qquad \frac{(m-1)}{\sigma^2} s_y^2 \sim \chi^2(m-1)$$

동분산 평균차 신뢰구간

$$\frac{(n-1)}{\sigma^2}s_x^2 \sim \chi^2(n-1) \qquad \frac{(m-1)}{\sigma^2}s_y^2 \sim \chi^2(m-1)$$

$$\frac{(n-1)}{\sigma^2}s_x^2 + \frac{(m-1)}{\sigma^2}s_y^2 \sim \chi^2(n-1+m-1)$$

C2 $X + Y \sim \chi^2(n_1 + n_2)$

$$s_p^2 := \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

Pooled Variance (합동 표본분산)

동분산 평균차 신뢰구간

C3 $\dfrac{Z}{\sqrt{X/n_1}} \sim t(n_1)$

$$\frac{\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}}}{\sqrt{\dfrac{s_p^2}{\sigma^2}}} = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{s_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim t(n+m-2)$$

$\mu_1 - \mu_2$의 1-α CI

$$\bar{X}-\bar{Y} \pm t_{\alpha/2}(n+m-2)s_p\sqrt{\frac{1}{n}+\frac{1}{m}}$$

분산 신뢰구간

σ^2 의 1-α CI

$$\frac{(n-1)}{\sigma^2}s_x^2 \sim \chi^2(n-1)$$

$$1-\alpha = P\left(\chi_{\alpha/2}(n-1) < \frac{(n-1)}{\sigma^2}s_x^2 < \chi_{1-\alpha/2}(n-1)\right)$$

$$= P\left(\frac{(n-1)s_x^2}{\chi_{1-\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)s_x^2}{\chi_{\alpha/2}(n-1)}\right)$$

분산비 신뢰구간

σ_2^2/σ_1^2 의 1-α CI

C4 $\dfrac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$

$$\frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2} \sim F(n-1, m-1)$$

$$1-\alpha = P\left(F_{\alpha/2}(n-1, m-1) < \frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2} < F1-\alpha/2(n-1, m-1)\right)$$

$$= P\left(F_{\alpha/2}(n-1, m-1)\frac{s_y^2}{s_x^2} < \frac{\sigma_2^2}{\sigma_1^2} < F_{1-\alpha/2}(n-1, m-1)\frac{s_y^2}{s_x^2}\right)$$

Next Topics

- 검정 test

수고하셨습니다!

Namun Cho/ mailto:namun@snu.ac.kr

수고하셨습니다!

Namun Cho/ mailto:namun@snu.ac.kr