분석적 사고

전종준
서울시립대학교 통계학과

December 7, 2018

Abstract

키워드: 선형대수, 미적분학, 회귀분석, 탐색적 자료분석

1 자료와 집합

2 벡터와 벡터공간

3 행렬과 데이터

4 선형변환과 이차형식

5 미분과 적분

6 벡터미분

<ins>(숙제)</ins> 이차형식의 최소화 문제를 푸는 방법에 대해 기술하여라.

7 확률과 기대값

7.1 확률변수

- (질문) 랜덤하다는 것은?
- (질문) 랜덤한 대상에 대해 알고 있는 것은?
- (질문) 확률의 정의? (실험, 실험결과, 표본공간, 사건, 확률)
- (질문) 확률변수는 왜 생각하게 되었나?
- (질문) 확률변수의 구분: 이산형, 연속형

<u>도수분포표</u>

이산형 변수에 대한 정보

- 도수분포표와 이산형 확률변수의 정보
- 주사위의 예
- 두 개 이상의 이산형 확률변수의 정보 (독립이 아닌 실험의 예)

<u>히스토그램과 확률</u>

연속형 변수에 대해 히스토그램을 그려 자료의 흩어진 상태를 파악한다.

- 히스토그램 그리기
- 어떤 구간에 속하는 자료의 상대비율 구하기
- 자료가 무한한 경우 히스토그램 그리기
- 자료가 무한한 경우 자료의 상대비율 구하기
- 확률분포와 적분

<u>추정대상으로서의 확률분포</u>

- 이산형 변수의 경우에는 $P(X = x)$ (모든 x에 대해서)가 추정대상임.

$$P(a \le X \le b) =$$

같은 의미에서 $F(x) = P(X \le x)$(모든 x에 대해서)에 대한 정보가 있으면 위에서 정의한 확률을 구할 수 있다.

- 연속형 변수의 경우에는 $F(x) = P(X \le x)$ (모든 x에 대해서) 혹은 pdf $f(x)$ 가 추정대상임

$$P(a \le X \le b) =$$

여기서 확률분포라 함은 랜덤한 개체의 모든 정보를 의미하며 확률분포를 안다고 하는 것은 불확실한 것에 대한 불확실성을 모두 기술할 수 있음을 의미한다. 특별히 어떤 확률변수의 확률분포를 기술할 때 $X \sim F$ 라고 쓴다.

- (부가설명) 다변량자료에 대한 확률분포의 예시
- (부가설명) $X, Y \sim F$ 이면 X와 Y가 같은 값을 가지는가?

<u>(토의) 다변량 자료의 히스토그램</u>

- 이변량 자료의 예를 들고 이변량 자료의 히스토그램을 그리는 방법에 대해서 설명하여라.
- 다변량 자료의 예를 들고 다변량 자료의 히스토그램을 그리는 방법에 대해서 설명하여라.
- 두 개 이상의 변량을 가진 자료에 대한 패턴을 기술하는 방법으로서 결합확률분포에 대한 개념을 설명하여라.

7.2 대수의 법칙

<u>대수의 법칙 (일변량)</u>

무한 모집단에서 무작위로 추출된 자료의 평균에 대한 법칙

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\text{random}} \to \underbrace{\text{true average}}_{\text{non-random}}$$

무한 모집단에서 무작위로 추출된 자료의 주어진 함수 g에 대한 변환에 대해서는?

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} g(X_i)}_{\text{random}} \rightarrow \underbrace{\text{true average of } g(X_1)}_{\text{non-random}}$$

$g(X_i)$를 새로운 자료로 보고 그 자료의 모집단에서 평균을 생각한다. 각각의 true mean 들을 $\mathrm{E}X$ 와 $\mathrm{E}g(X)$라고 표시하자.

$g(x) = x^2$ 일 때 예를 생각해보자.

<u>대수의 법칙 (다변량)</u>

무한 모집단에서 무작위로 추출된 다변량 자료의 평균에 대한 법칙

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i\right)^T \rightarrow (\text{true average of } X, \text{true average of } Y)^T$$

무한 모집단에서 무작위로 추출된 자료의 주어진 함수 g에 대한 변환에 대해서는?

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} g(X_i, Y_i)}_{\text{random}} \rightarrow \underbrace{\text{true average of } g(X, Y)}_{\text{non-random}}$$

$g(X_i, Y_i)$를 새로운 자료로 보고 그 자료의 모집단에서 평균을 생각한다. 각각의 true mean 들을 $(\mathrm{E}X, \mathrm{E}Y)$ 와 $\mathrm{E}g(X, Y)$라고 표시하자.

$g(x, y) = xy$ 일 때 예를 생각해보자.

<u>기대값</u>

- 적분을 이용한 일변량 확률변수의 기대값

$$\mathrm{E}X = \int x f(x) dx$$

$$\mathrm{E}X = \int g(x) f(x) dx$$

- 적분을 이용한 다변량 확률변수의 함수의 기대값

$$\mathrm{E}g(X,Y) = \int g(x,y)f(x,y)dxdy$$

(토의) 위 적분을 가중합의 개념을 이용하여 설명하여라.

기대값이 E에 대한 식이 있는 경우는 다음을 확인하여라

- 항상 E은 적분으로 표시할 수 있다.
- 적분으로 표시할 때 $\int$ 안쪽에 들어갈 pdf를 정확하게 기술하여야 한다.

$$\mathrm{E}(Y - X^T\boldsymbol{\beta})^2 =$$

예

- Uniform distribution

<u>근사이론으로서 대수의 법칙</u>

- 궁극적으로 알고자 하는 것이 EX라고 하자. EX에 계산에 필요한 것이 무엇인가?
- 유한한 자료 하에서 EX를 구하고자 할 때 가능한 방법은?
- 대수의 법칙이 말해주는 것은?
  - 자료가 가질 수 있는 값이 유한한 경우에 기대값과 대수의 법칙의 관계를 생각해보자.
  - 자료가 가질 수 있는 값이 무한한 경우 (연속형 변수)인 경우에 기대값과 대수의 법칙의 관계를 생각해보자.

앞서 살펴본 바와 같이 확률분포는 자료의 패턴을 설명해주는 정보로 볼 수 있다. 만약 다음과 같은 값에 관심이 있다고 가정하자.

$$f(\beta) = \mathrm{E}(Y - \beta X)^2$$

여기서 (X, Y)는 확률변수다. 이 모형은 모집단에서의 Y와 βX의 거리 제곱을 나타내는 값이다. 여기서 β 는 상수로서 우리가 선택해야할 숫자라고 하자. $f(\beta)$를 구하기 위해서는 적분을 통해 이 함수의 값을 구해야 하는데, (X, Y)의 패턴 즉, 확률밀도함수는 일반적으로 우리가 알지 못하는 값이다.

만약 우리가 자료들을 가지고 있다면 위 값을 대수의 법칙에 의해 근사할 수 있을 것이다.

$$f(\beta) = \mathrm{E}(Y - \beta X)^2 \simeq \frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta X_i)^2$$

7.3 조건부 기대값

Marginalization: 어떤 확률분포 $\Pr(X = x, Y = y) = p(x, y)$ 가 주어져 있다고 하자. 다음 식은 어떤 의미를 가지나? 예를 들어서 해석해보아라.

$$\sum_{\forall x} p(x, y)$$

확률분포가 확률밀도함수로 표현되어 있을 경우 위 식은 다음과 같이 표현된다.

$$\int f(x, y)dx$$

여러개의 변수가 있는 확률분포를 생각해보자.

$$\Pr(X_1 = x_1, \cdots, X_p = x_p)$$

여기서 다음 식을 해석해보자 (입력과 출력).

$$\sum_{\forall x_1, x_2} \Pr(X_1 = x_1, \cdots, X_p = x_p)$$

$f(x_1, \cdots, x_p)$를 확률변수 $(X_1, \cdots, X_p)$에 대응되는 확률밀도 함수라고 하자.

$$\int f(x_1, \cdots, x_p)dx_3 dx_4 \cdots dx_p$$

를 해석해보자.

Conditional probability/expectation: 다음 식을 해석해보아라.

$$\frac{\Pr(X = x, Y = y)}{P(X = x)} =$$

$$\frac{\Pr(X \in B, Y y)}{P(X \in B)} =$$

x라는 값이 주어져 있다고 생각하고, 다음 함수를 생각해보자.

$$p(y|x) = \frac{\Pr(X = x, Y = y)}{P(X = x)}.$$

- $p(y|x)$는 어떤 변수에 대한 함수인가?
- $p(y) := \Pr(Y = y)$와 $p(y|x)$를 비교해아라.
- 만약 $p(y)$와 $p(y|x)$가 모든 y에 대해 같다면 그 의미를 해석해보아라.
- 어떠한 x가 주어지더라고 우리는 그 x에 대응 되는 $p(y|x)$를 생각할 수 있다. 이 때 임의의 x에 대해서 $p(y) = p(y|x)$ for all y라면, 우리는 이 상황을 어떻게 설명할 수 있을까?

다음은 두가지 다른 종류의 기대값을 나타낸 것이다.

$$E(Y) = \sum yp(y)$$

$$E(Y|x) = \sum yp(y|x)$$

두 기대값의 의미를 비교해보고, 실제 사례를 들어보자.

$E(Y|X)$와 $E(Y|x)$의 차이를 비교해보자.

<u>기호의 모호함과 해결책</u> $p(x)$, $p(y|x)$, $f(x)$, $f(y|x)$와 같은 기호에서 p와 f의미를 해석해보고, 의미가 모두 다름을 확인하여라. 그리고 이를 구분하기 위해 우리가 도입해야 하는 기호들을 생각해보자.

<u>Conditional modeling</u>

- 남자와 여자의 키의 분포: 각각의 분포를 모형화 하는 방법과 평균을 모형화 하는 방법을 비교
- 교육수준에 따른 소득의 분포

- 지하철역의 거리에 따른 집값의 분포

다음은 기본적인 조건부 모형의 식 중 하나다.

$$Y = x^\top \boldsymbol{\beta} + \epsilon,$$

where $\epsilon \sim (0, \sigma^2)$, $x \perp \epsilon$