추정

- 목표: 미지의 모수 θ를 추정하고 싶음.
- 추정량 $\hat{\theta}$을 통하여 미지의 모수 θ를 추정.

example: 평균이 μ인 정규분포에서 10개의 샘플을 뽑았다고 하자.

$$x_1 = 26.3, x_2 = 29.2, \dots$$

우리는 μ를 알고싶다.

- 사람1: 아래와 같이 평균을 추정하자.

$$\hat{\mu} = \frac{26.3 + \cdots + 29.2}{10} = 28.5$$

- 사람2: 아래와 같이 평균을 추정하자.

$$\hat{\mu} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{28.0 + 28.4}{2} = 28.2$$

- 하나의 모수를 추정하는데 여러가지 방법이 있을 수 있다.
- 어떤것이 좋은 추정량인지 판단하는 기준이 필요하다.
- 또한 좋은 추정량들을 구하는 방법도 연구할 필요가 있다.
- 좋은 추정량 구하는 방법: (1) 적률추정법 (2) 최대가능도추정법 (3) 최소분산불편추정법
- 좋은 추정량을 판단하는 기준: (1) 불편성 (2) 효율성 (3) 최소분산성 (4) 일치성 (5) 충분성

적률추정법

- 적률추정법은 말그대로 적률을 활용하여 모수를 추정하는 방법을 말함. 적률은 아래와 같은것을 말함.

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

example: $X_1 \dots X_n \overset{iid}{\sim} Gamma(\alpha, \beta)$ 라고 하자. α, β를 추론해보자. 아래의 관계를 관찰하자.

$$E(X) = \alpha\beta$$
$$V(X) = \alpha\beta^2 + (\alpha\beta)^2$$

그런데 $E(X)$는 m_1으로 추론할 수 있고 $V(X)$는 $m_2 - m_1^2$으로 추론할 수 있다. 따라서 아래를 연립하여 풀면 α,β를 추론할 수 있다.

$$\begin{cases} m_1 = \alpha\beta \\ m_2 - m_1^2 = \alpha\beta^2 + (\alpha\beta)^2 \end{cases}$$

- 장점: 직관적이고 쉽다. WLLN에 의해서 일치성을 보이기 쉽다.

<u>note:</u> 일치성: $\hat{\theta} \xrightarrow{p} \theta \; as \; n \to \infty$.

- 단점: 유일하지 않을 수 있다. 효율이 좋지 않은 경우가 많다.

최대가능도 추정법

example: $X_1, X_2, \dots, X_6 \overset{iid}{\sim} Bernoulli(p)$ 이라고 하자. 아래의 샘플을 관찰했다고 하자.

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = 0, \quad x_4 = 0, \quad x_5 = 0, \quad x_6 = 0.$$

p를 추정하고 싶다고 하자.

- 사람1: $p = 0.5$ 라고 하자.

- 사람2: $p = 0.5$ 보다 $p = 0.3$ 이라고 추정하는것이 더 합리적일것 같다. 왜냐하면 $p = 0.5$라고 가정하였을때

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = 0, \quad x_4 = 0, \quad x_5 = 0, \quad x_6 = 0.$$

와 같은 샘플을 얻을 확률은

$$\frac{5}{10} \times \frac{5}{10} \times \cdots \times \frac{5}{10} = 0.015625$$

이지만 $p = 0.3$이라고 가정하였을 경우는

$$\frac{7}{10} \times \frac{3}{10} \times \cdots \times \frac{7}{10} = 0.050421$$

가 된다. 따라서 $p = 0.3$ 이라고 추정하는 것이 더 합리적이다.

- 이것이 최대가능도 추정의 모티브이다. 위의 상황을 수식으로 표현하여 보자. 결국 사람2는 아래의 함수를 더 크게 만드는 p가 좋은 추정량임을 주장하는 것이다.

$$L(p) = \prod_{i=1}^{6} pdf(x_i; p)$$

여기에서 $pdf(x_i; p)$는 모수가 p라고 가정하였을 경우 X_i의 pdf이다.

note: 위의 예제의 경우 사람1은 $p = 0.5$라고 믿고 있으므로

$$pdf(x_i; p = 0.5) = p^{x_i}(1-p)^{1-x_i}$$

이다. 따라서 $pdf(x_1; p = 0.5) = \cdots = pdf(x_1; p = 0.5) = 0.5$이다. 따라서

$$L(0.5) = (0.5)^6 = 0.015625$$

라고 쓸 수 있다.

note: 사람2는 $p = 0.3$이라고 믿고 있으므로

$$\begin{aligned} pdf(x_1; p = 0.3) =& 0.3^0 \times 0.7^1 = 0.7 \\ pdf(x_2; p = 0.3) =& 0.3^1 \times 0.7^0 = 0.3 \\ & \dots \\ pdf(x_6; p = 0.3) =& 0.3^0 \times 0.7^1 = 0.7 \end{aligned}$$

와 같이 된다. 따라서

$$L(0.3) = (0.7)^5 \times (0.3)^1 = 0.050421$$

이 된다.

- 그런데 사람3이 $p = 0.2$라고 주장하였다. 이 주장이 더 합리적인지 판단하기 위해서 $L(p)$를 조사해보자. 조사결과

$$L(0.2) = 0.8^5 \times (0.2)^1 = 0.065536$$

따라서 사람3의 주장이 더 합리적이다.

- 전략: 어차피 $p \in (0,1)$이므로 모든 p에 대하여 아래의 값을 조사하고 이것을 최소화하는 p를 구하자.

$$L(p) = \prod_{i=1}^{n} pdf(x_i; p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

로그를 취하면

$$\log L(p) = \sum_{i=1}^{n} x_i \log p + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-p).$$

미분을 하면

$$\frac{\partial}{\partial p}\log L(p)=\frac{\sum_{i=1}^n x_i}{p}-\frac{n-\sum_{i=1}^n x_i}{1-p}.$$

따라서 $\frac{\partial}{\partial p}\log L(p)=0$를 풀면

$$\frac{\sum_{i=1}^n x_i}{p}=\frac{n-\sum_{i=1}^n x_i}{1-p}$$

정리하면

$$p=\frac{1}{n}\sum_{i=1}^n x_i=\bar{x}$$

이다.

- 따라서 $p=\bar{x}$로 추정하는게 좋겠다. 즉

$$\hat{p}=\bar{x}$$

따라서 문제의 경우

$$\bar{x}=\frac{0+1+0+0+0+0}{6}=\frac{1}{6}$$

로 p를 추정하는 것이 가장 합리적이다. 이 결과는 놀라울정도로 직관적이고 당연하다. 6번던져서 한번 성공했다면, 성공확률은 대충 1/6로 보는것이 타당할테니까.

- 여기에서 $L(p)$를 p에 대한 likelihood function이라고 한다.

note: 한글로는 우도함수라고 변역하기도 하고 가능도함수라고 변역하기도 한다.

- $\hat{p}$는 likelihood function을 최대화하여 얻은 p의 추정량인데 이러한 이유로 maximum likelihood estimator라고 부르고 줄여서 MLE라고 부른다.

note: 한글로는 최대가능도추정량, 혹은 최대우도추정량이라고 말한다.

- 따라서 $\hat{p}$이 어떠한 방식으로 얻은 추정량인지 더 명확하게 하기위해서 아래와 같이 표기하기도 한다.

$$\hat{p}^{MLE}$$

- 참고로 p를 적률추정법 (method of moments estimator) 으로 추정한다고 하자. 베르누이 분포의 경우

$$EX=p$$

이고 EX는 m_1 즉 $\bar{x}$로 추정할 수 있으므로

$$\hat{p}^{MME} = \bar{x}$$

가 된다. 따라서 베르누이 분포의 경우 적률추정법과 최대가능도 추정법은 같다.