

# 5장. 추정의 Quality를 판단하는 기준

## ◆ Introduction

직장 상사 갑은 치아 보험의 보험료를 책정하기 위해 계리사 을에게 치아 보험의 1인당 평균 보험금 지급액 대한 모델링을 부탁했다. 을은 모수추정을 통해 치아 보험의 1인당 평균 보험금 지급액이  $\hat{\mu} = 25,300$ ,  $\hat{\sigma} = 4,000$  인 정규분포를 따른다고 모델링했다. 여기서 끝이 아니다. 을은 자신이 추정한 모수 값이 정확하다고 확신할 수 있는지 그리고 추정 과정에서 문제는 없었는지 다시 살펴봐야 할 것이다. 추정값이 현실과 크게 다르면 보험회사는 막대한 손실을 입는 것은 물론, 추정한 당사자는 당장 사표를 써야 할 수도 있다. 이번 장은 추정 과정과 결과에 대한 평가를 어떻게 해야할지 알려줄 것이다.

---

Loss Models는 계리적 모델을 주요 내용으로 한다. 통계학이라는 광범위한 분야에서 모델의 '추정'과 '가설검정'은 매우 중요하다. 이번 장에서는 '추정', 그 중에서도 점추정(Point Estimation) 과정에 대해 중점적으로 살펴보겠다.

---

Loss Models에서 모수를 추정하는 방법은 총 3가지가 나온다. 앞서 설명한 예시의  $\mu, \sigma$  와 같은 모수를 추정하는 방법이 3가지 라는 말이다. 그 방법으로 Method of Moments, Percentile Matching, Maximum Likelihood Estimation가 있다. 이번 장에서는 추정 방법보다는 추정 과정과 결과에 대해 어떻게 평가할 수 있는지 살펴볼 것이다. (설명을 제대로 이해하기 위해 밑에 있는 주석을 읽어보자.)

점 추정은 모집단으로부터의 표본을 이용하여 모수에 대한 추정치<sup>1</sup>(estimate)로 하나의 단일값(point)을 산출한다. 어떠한 표본을 이용하느냐에 따라 단일값은 달라질 수 있다. 추정 과정이 합리적이라면 추정치들 중 일부는 모수와 일치하거나 근접할 것이다. 물론 추정치가 모수로부터 많이 벗어나는 경우도 간혹 발생할 수 있다. 잠재적인 추정치들이 참값과 얼마나 일치하는지를 측정(measure)하여 추정 과정의 quality를 알 수 있다.

불편향성(Unbiasedness)은 추정 과정의 quality를 측정해주는 개념이다. 추정량<sup>2</sup>(estimator)의 기댓값  $E[\hat{\theta}|\theta]$  과 알고자 하는 모수  $\theta$ 의 차이를 bias 라고 하며, bias가 0일 때 unbiased라고 한다.

1 확률표본의 특정 관측치를 추출한 후에 계산된 추정량의 값을 추정값(estimate) 또는 추정치라고 한다

2 모집단의 한 모수를 추정하기 위하여 충분히 큰 확률표본을 추출할 때, 이 표본을 기초로 하여 그 모수를 어떻게 추정할 것인지 알려주는 확률표본의 함수를 추정량이라고 한다.

## 5장. 추정의 Quality를 판단하는 기준

Def. An estimator,  $\hat{\theta}$ , is **unbiased** if  $E(\hat{\theta}|\theta) = \theta$  for all  $\theta$ .

The **bias** is  $bias_{\hat{\theta}}(\theta) = E(\hat{\theta}|\theta) - \theta$ .

Asymptotically unbiasedness (점근적 불편향성)는 표본수를 무수히 늘리면 추정량의 기댓값이 점근적으로 모수와 일치한다는 것을 의미한다.

Def. Let  $\hat{\theta}_n$  be an estimator of  $\theta$  based on a sample size of  $n$ .

The estimator is **asymptotically unbiased** if

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n|\theta) = \theta \text{ for all } \theta$$

추정량이 점근적 불편향성을 만족할 때 관측치의 수가 무수히 많으면 추정량의 기댓값과 모수의 차이는 0이 된다. 좋은 추정량은 그 기댓값이 모수와 같아야 하며, 모수와 오차는 작아야 한다. 이때 그 둘 사이의 오차를 평가하기 위해서 오차의 제곱을 사용하는데 그 값을 MSE(mean squared error) 라고 하며, true value  $\theta$ 하에서 추정량과 모수의 차이를 제곱한 것의 기댓값으로 나타낸다.

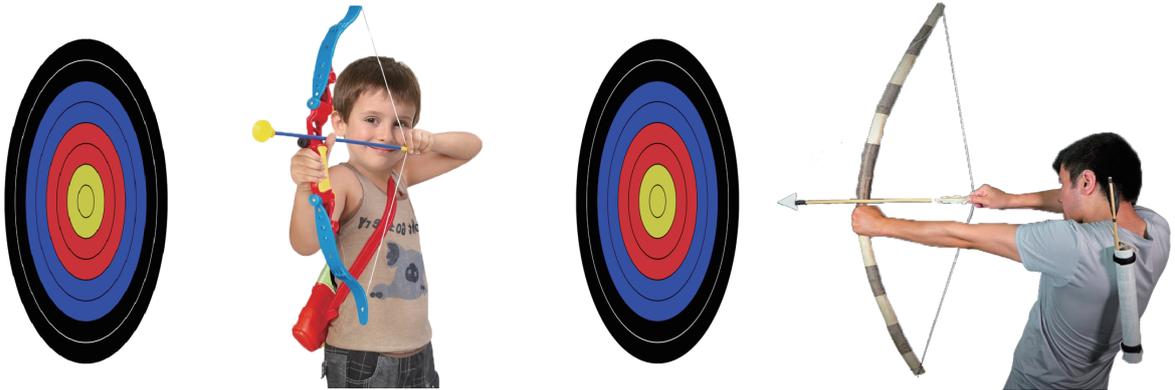
The mean squared error (MSE) of an estimator is  $MSE_{\hat{\theta}}(\theta) = E[(\hat{\theta} - \theta)^2 | \theta]$ .

Unbiased estimator(불편 추정량) 중에서 가장 작은 분산을 가지는(최소 MSE를 갖는) 추정량을 UMVUE(최소 분산 불편추정량, Uniformly Minimum Variance Unbiased Estimator)이라고 한다. 정의에 있는 uniformly는 무슨 뜻일까? 통계학에서 uniformly는 for any true value를 뜻한다. 따라서 UMVUE는 어떤 true value  $\theta$  에 대해서도 성능이 좋은 값을 알 수 있다.

특정 추정치가 모수와 같다고 하여 추정량이 옳다는 판단을 내려서는 안된다. 표본들 중 일부 추정치만 일치하고 대다수의 추정치가 모수와 멀리 떨어져 있을 수도 있기 때문이다. 추정량이 옳다고 결정을 내리려면 위 조건들을 만족해야만 한다. 그때서야 비로소 자신이 만든 추정 함수(추정량)로 부터 나온 추정치가 모수를 대신하여 사용할 수 있다는 정당성을 얻게 된다.



★ 쉬어가는 페이지 ★



## “점 추정과 구간 추정”

어릴 적 가지고 놀던 화살 쏘기 장난감이 기억나는가? 또 양궁선수들이 경기에 사용하는 화살을 본 적이 있는가? 여기서 뾰족한 양궁선수들의 화살은 ‘점추정’으로, 앞이 넓은 장난감 화살은 ‘구간추정’으로 생각 할 수 있다. 우리가 알고자 하는 모수는 정확히 알 수 없지만, 모수를 알기 위해 화살을 쏜다(추정). 그 결과(추정의 결과) 점 추정은 과녁의 딱 한 점에 꽂히게 되고 구간추정은 범위에 걸쳐 꽂히게 된다. 우리는 필요에 따라 점 추정과 구간추정을 사용할 수 있다.