중심극한정리 (Central Limit Theorem)의 예시

박재성

KTUG Conference 2015

중심극한정리

상자 안에 1에서 999까지 숫자가 표시된 999개의 상태가 균질한 공 (X) 을 넣고 이를 특정한 모집단 (population) 이라고 가정하자. 이 모집단의 평균 μ은 500이다. 모집단의 분산 $\mathrm{var}(X)$는 80,475이다.

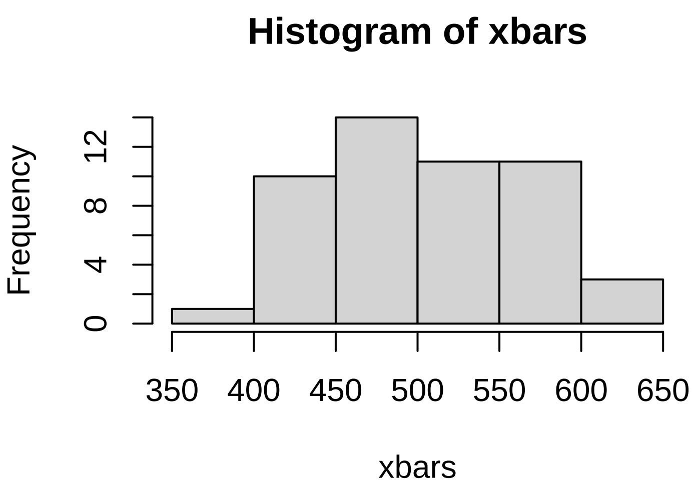
이 중 30개의 공을 50회에 걸쳐 반복 추출한다. 이 경우 표본평균의 분포는 $E(\bar{X}_i) = \mu(X)$ 이고 분산이 $\mathrm{var}(\bar{X}_i) = \frac{\sigma^2}{n}$ 인 정규분포에 근사한다. 즉, $X \sim \mathcal{N}(500,\ 283.7^2)$ 이다.

실험

이를 실험을 통해 살펴보자. 999개의 공이 든 상자에서 30개의 공을 50회에 걸쳐 무작위 반복추출하고 (30개의 공을 뽑은 뒤, 다시 그 공을 상자 안에 집어 놓고 상자를 처음과 같은 상태가 되도록 뒤흔들어서 다시 30개의 공을 뽑는 것을 50회 반복한다), 그 각각의 평균을 기록하면 다음과 같다.

> 50회 반복추출의 평균값 551.1, 441.5, 518.5, 527.5, 449.7, 493.4, 579.7, 557.0, 552.8, 476.3, 581.0, 419.6, 616.1, 483.1, 450.0, 380.8, 447.4, 494.9, 496.6, 578.1, 509.5, 522.6, 457.0, 575.1, 492.6, 604.6, 501.3, 458.0, 441.2, 419.0, 491.2, 443.7, 585.6, 436.2, 526.0, 492.7, 478.0, 533.4, 529.9, 533.9, 439.9, 572.6, 509.9, 561.8, 579.2, 521.4, 603.2, 454.8, 471.4, 456.3

표본평균값의 분포를 히스토그램으로 표현하면 아래 그림과 같다.



그림에서 보듯 $\bar{X}_i$ 는 500을 중심으로 좌우대칭적으로 분포하고 있다. 이들의 평균은 505.9으로 모평균 500과 근사하다. 이들 중 이론적으로 산출한 평균으로부터 약 2 표준편차만큼 떨어진 구간 ($\mu \pm 2\frac{\sigma}{\sqrt{n}}$) 안에 속하는 값, 즉 $[396.4, 603.6]$ 의 범위 안에 있는 값의 개수를 세면 모두 47개이다.

표본평균 분포의 약 95%를 포괄하고 있음을 알 수 있다.