■ 확률표본과 통계량

- 확률표본(random sample): 서로 독립적이고 동일한 분포에서 추출된 표본 ⇨ independent and identically distributed (iid)
- 통계학적 관점에서 표본을 뽑는 이유
  ⇒ 모집단(모수)에 대한 추론
- 통계량(statistic): 관측가능한 표본의 함수
  - 추정량(estimator): 모수의 추정에 사용되는 통계량 ⇨ 확률변수
  - 추정값(estimate, 추정치): 추정량의 관측값
- 통계량의 확률분포 ⇒ 표집분포(sampling distribution)

■ 표집분포(sampling distribution)

- 통계량의 확률분포
  - 주요관심 통계량
    - 표본평균: $\overline{X}$ (표본비율 포함)
    - 표본평균: S^2 (표본표준편차)
- 모집단이 정규분포일때 표본평균 $\overline{X}$ 의 분포는?

$$\overline{X}=\frac{X_1+\cdots+X_n}{n}\sim N(\mu,\sigma^2/n)$$

⇨ $\overline{X}$ **의 표준오차는** $\frac{\sigma}{\sqrt{n}}$

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

- 지수족(exponential family): 정규분포포함
  - 이항분포, 포아송분포, 음이항분포, 감마분포 등
  - 확률표본의 합의 분포도 해당 분포
- 다른 분포는?
  - 직접유도
  - 근사분포 유도 ⇒ 중심극한정리
  - Monte Carlo 모의실험을 통해 표집분포 추정

- 중심극한정리(Central Limit Theorem, CLT)
  - $X_1, X_2, \ldots, X_n$ 평균 μ, 분산 σ^2인 모집단에서 추출된 확률표본
  - n 이 커질수록 모집단의 형태와 관계없이 $\overline{X}$ 의 분포(표집분포)는 정규분포에 근사

$$\overline{X} \simeq N(\mu, \sigma^2/n) \;\Rightarrow\; Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \simeq N(0,1)$$

  - $Y = X_1 + \cdots + X_n$라면

$$Z = \frac{Y - n\mu}{\sqrt{n}\,\sigma} \simeq N(0,1) \;\Rightarrow\; Y \simeq N(n\mu, n\sigma^2)$$

- 이항분포의 정규근사
  - ○ $X \sim B(n,p)$ ⇒ n 개의 베르누이 확률변수의 합
  - ○ 표본비율 : $\hat{p} = X/n = \overline{X}$ n이 큰 경우, 중심극한정리에 의해,

$$\hat{p} \simeq N\left(p, \frac{p(1-p)}{n}\right) \Rightarrow X \simeq N(np, np(1-p))$$

  - ○ 이항분포는이산형이고 정규분포는 연속형
    ⇒ 연속성수정
- 기타 통계량의 표집분포
  - ○ 표본분산, 최댓값, 최솟값 등

확률(probability) 통계적 추론(statistical inference)

⇓ ⇓ ⇓ ⇓

- **통계학의 이해 1: 기술통계 & 확률**
- **통계학의 이해 2: 통계학의 이해 1의 내용을 기반으로 한 모집단(모수)에 대해 추론하는 방법론**