■ 확률표본(random sample)

- 모집단에서 무작위로 선택되어진 관측값
- 서로 독립이고 동일한 분포를 따른다고 가정

⇨ independent and identically distributed (iid)

○ (독립, 동일한 분포) ⇒ 복원추출

○ 예】정규분포에서 추출한 경우

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$$

- 독립이기 때문에 결합분포는 각각의 주변분포 곱으로 표시

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

- 동일한 분포를 따르기 때문에 동일한 확률질량(밀도)함수를 가짐

$$\prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f(x_i)$$

- $X_1, X_2, \ldots, X_n$이 확률표본이고 $X_i \sim f(x)$이면

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f(x_i)$$

◉ 윷놀이

- X_i : i 번째 윷이 젖혀지면 1 엎어지면 0 ⇒ $X_i \sim B(1, p)$

$$f(x) = p^x (1-p)^{1-x}, \quad x = 0, 1$$

- $X_1, X_2, X_3, X_4 \overset{\text{iid}}{\sim} B(1, p)$

- 확률표본의 결합확률질량함수

$$f_{X_1, \cdots, X_4}(x_1, \cdots, x_4) = \prod_{i=1}^{4} p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{4-\sum x_i}$$

- 통계학적 관점에서 표본을 뽑는 이유
  ⇒ 모집단에 대한 추론

◉ $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$

  ○ μ, σ^2이 무엇인가? ⇨ 모수적추론

  ○ μ에 대해 알아보기 위해서는 $\overline{X}$

  ○ σ^2에 대해 알아보기 위해서는 S^2

- 통계량(statistic): 관측가능한 표본의 함수
  ○ 관측가능하다는 것은 미지의 모수를 포함하지 않음을 의미

- 추정량(estimator): 모수의 추정에서 사용되는 통계량
  - 확률변수
  - 추정값(estimate, 추정치): 추정량의 관측값

◉ $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, σ^2를 알고 있다고 가정

- 추정량: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ⇒ 확률변수(추출 전)
- 추정값: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ ⇒ 상수 (추출 후)

◉ 확률분포가 다음과 같을 때

x	0	1	2
$P(X=x)$	2/5	2/5	1/5

○ 두 개의 확률표본 추출: X_1, X_2

○ $Y=\max(X_1, X_2)$일 때 Y의 분포는?

y	0	1	2
$P(Y=y)$	4/25	12/25	9/25

● 통계량의 확률분포 ⇒ 표집분포(sampling distribution)