

■ 표집분포(sampling distribution)

- 통계량의 확률분포
 - 통계량 : 측정 가능한 확률표본의 함수
 - 관심 통계량
 - 표본평균 : \bar{X} (표본비율 포함)
 - 표본분산: S^2 (표본표준편차)
 - 극한값: $X_{(n)} - X_{(1)} \Rightarrow$ 범위($X_{(1)}, X_{(n)}$)
 - 순위(rank): X_i 의 크기 순서
 - 이들 통계량의 통계적 성질?

◎ 확률분포가 다음과 같을 때

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

- $\mu = E(X) = \frac{4}{5}$
- $\sigma^2 = Var(X) = \frac{14}{25}$
- 두 개의 확률표본을 추출한 경우, 두 표본의 평균 \bar{X} 의 분포는?
 - $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$

		X_2		
		0	1	2
X_1	0	4/25	4/25	2/25
	1	4/25	4/25	2/25
	2	2/25	2/25	1/25

○ 표본평균 \bar{X} 의 분포는?

\bar{x}	0	1/2	1	3/2	2
$P(\bar{X} = \bar{x})$	4/25	8/25	8/25	4/25	1/25

$$E(\bar{X}) = \frac{4}{5}, \quad Var(\bar{X}) = \frac{14}{50} = \frac{1}{2} \times \frac{14}{25}$$

- 평균 μ , 분산 σ^2 인 분포에서 n 개의 확률표본을 추출했을 경우, 표본평균 \bar{X} 의 분포는?
 - $E(\bar{X}) = ? \mu$
 - $Var(\bar{X}) = ? \sigma^2/n$
 - $SD(\bar{X}) = ? \sigma/\sqrt{n}$
 - 표준오차(standard error, SE): 통계량의 표준편차
 - 분포의 형태는?

- 정규분포인 경우

- $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ 이고, X_1 과 X_2 가 독립이면

$$\Rightarrow X_1 \pm X_2 \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$$

- $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$

- $X_1 + \dots + X_n \sim N(n\mu, n\sigma^2)$

- $\bar{X} = \frac{X_1 + \dots + X_n}{n} \sim N(\mu, \sigma^2/n) \Rightarrow \bar{X} \text{의 표준오차는 } \frac{\sigma}{\sqrt{n}}$

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

- 지수족(exponential family): 정규분포포함
 - 이항분포: $X_1 + X_2 \sim B(2n, p)$
 - 포아송분포: $X_1 + X_2 \sim Poi(2\lambda)$
 - 음이항분포: $X_1 + X_2 \sim G(2\alpha, \beta)$
 - 감마분포: $X_1 + X_2 \sim G(2\alpha, \beta)$
 - 지수분포, 카이제곱분포 포함
- 다른 분포는?
 - 직접유도
 - 근사분포 유도
 - Monte Carlo 모의실험을 통해 표집분포 추정

● $X_i \sim Poi(1),$

○ $X = X_1 + \dots + X_{15} \Rightarrow X \sim Poi(15)$

○ $P(X \leq x)$

• Monte Carlo 모의실험: 10만 번 반복

x	3	5	10	15	20
정확확률	0.0002	0.0028	0.1185	0.5681	0.9170
모의실험결과	0.0002	0.0028	0.1172	0.5696	0.9170
허용오차	0.0001	0.0003	0.0020	0.0031	0.0017

• 확률이 0 근거나 1 근거의 경우 모의실험결과가 상당히 정확

- 정리
 - 표집분포: 통계량의 (모집단) 분포
 - 정규확률표본의 표본평균의 분포는 정규분포
 - 독립인 지수족 표본들 합의 분포는 해당 지수족의 분포