■ 초기하분포 (Hypergeometric Dist.)

- 크기가 N인 모집단이 크기가 M과 $N-M$인 두 개의 부모집단 (A, B)로 나누어진 경우 ⇨ **유한모집단**

- n개의 표본을 **비복원으로 추출**할 때, 부모집단(A)에서 추출될 표본 수의 분포 ⇨ 각 표본의 추출과정은 독립적이지 않음

SOOKMYUNG WOMEN'S UNIVERSITY K-MOOC

◉ 6개가 정상품과 4개의 불량품이 있는 상자에서 임의로 3개의 제품을 비복원 추출한 경우에 3개 중 1개가 불량품일 확률?

○ 3개 중 1개가 불량품일 사건:

{(불,정,정), (정,불,정), (정,정,불) }

$$P(\text{불},\text{정},\text{정})+P(\text{정},\text{불},\text{정})+P(\text{정},\text{정},\text{불})$$

$$=\frac{4}{10}\times\frac{6}{9}\times\frac{5}{8}+\frac{6}{10}\times\frac{4}{9}\times\frac{5}{8}+\frac{6}{10}\times\frac{5}{9}\times\frac{4}{8}$$

$$=3\times\frac{4\times6\times5}{10\times9\times8}=\frac{1}{2}$$

↳ 위치 중 하나를 선택해 “불” 대입하는 방법 수

4개에서 1개, 6개에서 2개를 비복원 추출 나열하는 방법

↑

ㅇ 확률 : $\dfrac{4\times6\times5}{10\times9\times8}$

↳ 10개에서 3개를 비복원 추출 나열하는 방법

ㅇ X: 불량품 개수

$$\Rightarrow P(X=1)=\binom{3}{1}\frac{\dfrac{4!}{(4-1)!}\,\dfrac{6!}{(6-2)!}}{\dfrac{10!}{(10-3)!}}=\frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}}=\frac{1}{2}$$

○ 확률질량함수

$$f(x) = \binom{3}{x} \frac{\dfrac{4!}{(4-x)!} \dfrac{6!}{(6-(3-x))!}}{\dfrac{10!}{(10-3)!}}$$

$$= \frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}}, \qquad x = 0, 1, 2, 3$$

- 확률질량함수 일반식:

$$f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

  - $x = \max(0, n-N+M), \ldots, \min(n, M)$

- N이 크고 N에 비해 n이 상대적으로 작은 경우
  - 비복원의 효과가 적기 때문에 베르누이 실험으로 근사
  - 초기화 분포는 $p = M/N$인 이분항분포로 근사

◉ 10000개의 제품 중 7000개가 정상, 3000개가 불량이라면 3개를 비복원 추출에서 불량품이 한 개일 확률

$$P(X=1)=\binom{3}{1}\frac{3000\times 7000\times 6999}{10000\times 9999\times 9998}\approx \binom{3}{1}\frac{3}{10}\left(\frac{7}{10}\right)^2$$

● 기댓값

○ 초기하분포도 각 시행에서 A 집단에서 추출되면 1 다른 집단에서 추출되면 0으로 표시한 확률변수의 합

$$\begin{array}{cccccc} & X_1 + & X_2 + & \cdots & + X_n = & X \\ A & 1 & 1 & \cdots & 1 & \downarrow \\ B & 0 & 0 & \cdots & 0 & A\text{에서 추출된 표본의 수} \end{array}$$

○ $P(X_i = 1) = P(A) = \frac{M}{N}$, $P(X_i = 0) = 1 - \frac{M}{N}$

○ $E(X_i) = \frac{M}{N} = p \Rightarrow E(X) = n\frac{M}{N} = np$

○ $E(X_i^2) = \frac{M}{N} = p \Rightarrow Var(X_i) = p - p^2 = p(1-p) = \frac{M}{N}\frac{N-M}{N}$

○ 다른 점은 추출이 비복원으로 각각의 시행이 독립이 아님

⇨ $Var(X) = Var(X_1 + \cdots + X_n)$

$$= \sum_i Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

↳ $\binom{n}{2}$

- $Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$

- $E(X_i X_j) = P(X_i = 1, X_j = 1)$

$$= P(X_i = 1)P(X_i = 1 \mid X_i = 1) = \frac{M}{N}\frac{M-1}{N-1}$$

$$\Rightarrow Cov(X_i, X_j) = \frac{M}{N}\frac{M-1}{N-1} - \left(\frac{M}{N}\right)^2$$

$$= -\frac{M}{N}\frac{N-M}{N(N-1)} = -\frac{p(1-p)}{N-1} \le 0$$

○ $$Var(X) = np(1-p) - n(n-1)\frac{p(1-p)}{N-1}$$

$$= np(1-p)\frac{N-n}{N-1} \le np(1-p)$$

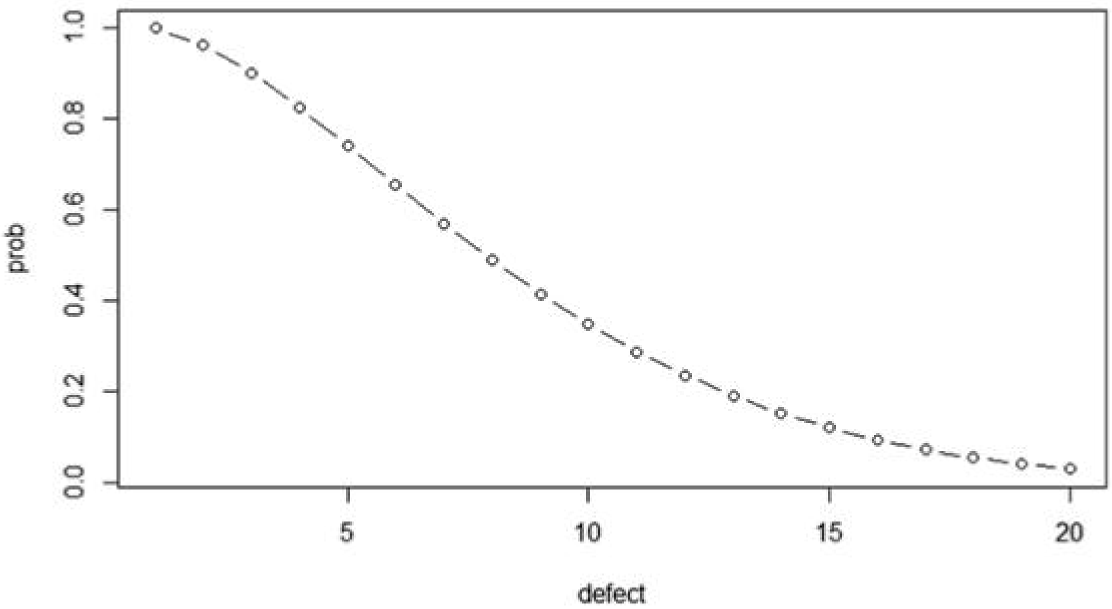
↳ 유한모집단 수정계수 ≤ 1

◉ 품질관리 – Operating Characteristic(OC) curve

- 50개의 전구들이 들어 있는 상자에서 10개의 전구를 무작위로 선택하여 검사
- 불량전구의 개수가 1개 이하이면 이 회사의 전구를 구매
- 만약 이 상자에 5개의 불량품이 있을 때, 구매할 확률은?
  - X= 10개 중 불량품의 수

$$P(X \le 1) = \frac{\binom{5}{0}\binom{45}{10}}{\binom{50}{10}} + \frac{\binom{5}{1}\binom{45}{9}}{\binom{50}{10}} = 0.311 + 0.431 = 0.742$$

○ 만약 k개 불량품이 있을 때, 구매할 확률은?



ㅇ OC curve 계산

· 몇 개의 표본을 추출할 것인가?

· 불량품이 몇 개일 때 까지 구매할 것인가?

◉ 연못에 사는 물고기는 몇 마리?

ㅇ 꼬리표를 붙인 20마리의 물고기를 연못에 넣고 어느 정도 지난 후 물고기 15마리를 잡았을 때 꼬리표가 있는 물고기의 분포는?

· $N-20$: 꼬리표가 없는 물고기

○ 15마리 중 4마리가 꼬리표가 있는 물고기라면?

· 비례식: N = 75

- 요약
  - 크기가 N인 모집단이 두 그룹 (크기가 M과 $N-M$)으로 나뉘고 n개의 표본을 **비복원으로 추출**할 때, 특정그룹에서 추출될 표본 수의 분포

$$f(x) = \binom{M}{x}\binom{N-M}{n-x}/\binom{N}{n},$$

. $x = \max(0, n-N+M), \ldots, \min(n, M)$

· $n \ll N$인 경우 $p = M/N$**인 이항분포로 근사**할 수 있음

  - $E(X) = \frac{nM}{N}$, $Var(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$