04 강

통계적 추론

이산형 확률분포

한국방송통신대학교 통계·데이터과학과
이긍희 교수

학습내용

1. 이산형 균등분포와 베르누이분포를 이해한다.
2. 이항분포와 초기하분포를 이해한다.
3. 기하분포와 음이항분포를 이해한다.
4. 포아송분포와 다항분포를 이해한다.

01

이산형 확률분포의 개요

1 이산형 확률분포

이산형 확률변수와 확률분포

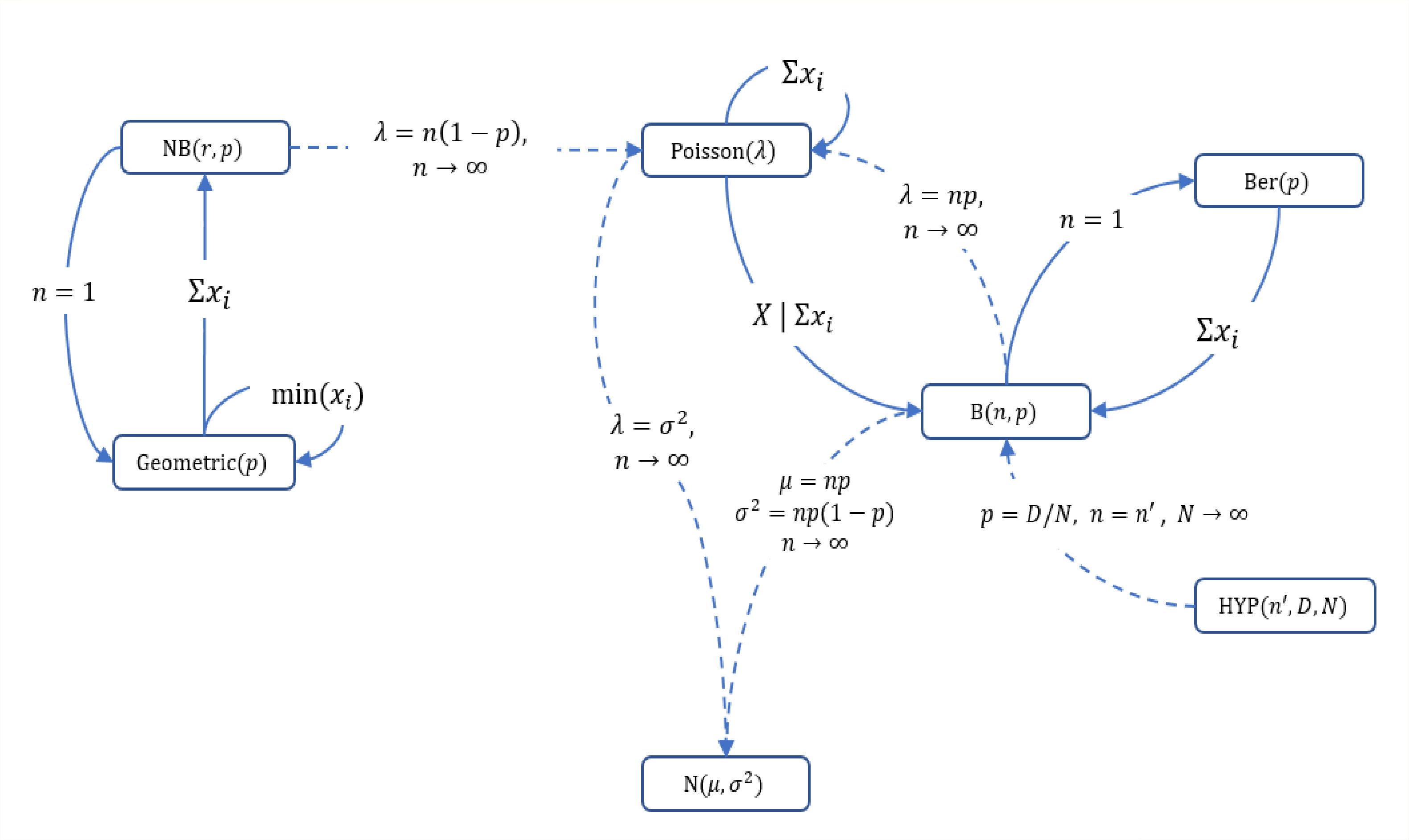
- **이산형 확률변수**
- **확률질량함수**

$$f(x|\theta) = P_\theta(X = x), \qquad x = 0,1,2,\cdots$$

- **이산형 균등분포, 초기하분포, 베르누이분포, 이항분포, 기하분포, 음이항분포, 포아송분포, 다항분포**

1 이산형 확률분포

이산형 확률분포의 관계



2 경우의 수

표본공간에서 원소 추출할 때 경우의 수

- **주머니(a, b, c)에서 공 2개 뽑을 때**

$$\{a,\ a\},\ \{a,\ b\},\ \{a,\ c\},$$
$$\{b,\ a\},\ \{b,\ b\},\ \{b,\ c\},$$
$$\{c,\ a\},\ \{c,\ b\},\ \{c,\ c\}$$

2 경우의 수

표본공간에서 원소 추출할 때 경우의 수

- **주머니(a, b, c)에서 공 2개 뽑을 때**

$$\{a,\ a\},\ \{a,\ b\},\ \{a,\ c\},$$
$$\{b,\ a\},\ \{b,\ b\},\ \{b,\ c\},$$
$$\{c,\ a\},\ \{c,\ b\},\ \{c,\ c\}$$

$$\{a,\ b\},\ \{a,\ c\},$$
$$\{b,\ a\},\ \{b,\ c\},$$
$$\{c,\ a\},\ \{c,\ b\}$$

2 경우의 수

표본공간에서 원소 추출할 때 경우의 수

- **주머니(a, b, c)에서 공 2개 뽑을 때**

$$\{a,\ a\},\ \{a,\ b\},\ \{a,\ c\},$$
$$\{b,\ a\},\ \{b,\ b\},\ \{b,\ c\},$$
$$\{c,\ a\},\ \{c,\ b\},\ \{c,\ c\}$$

$$\{a,\ b\},\ \{b,\ c\},\ \{a,\ c\}$$

2 경우의 수

표본공간에서 원소 추출할 때 경우의 수

- **복원추출여부, 순서 고려 여부**

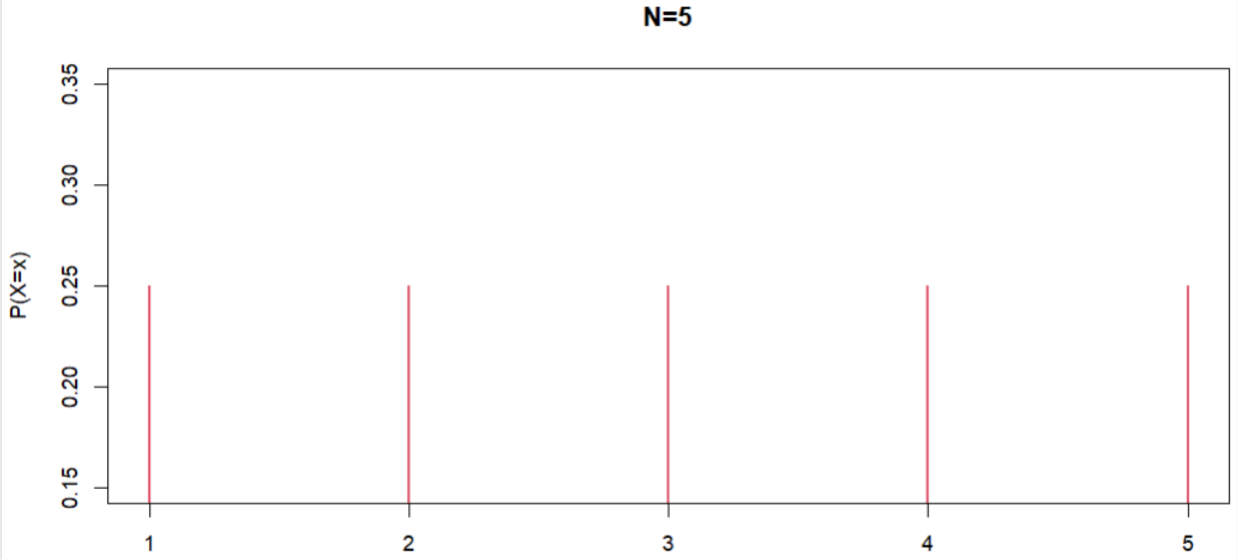
구분	순서	
	비고려	고려
비복원	${}_nC_r = \dfrac{n!}{r!(n-r)!}$	${}_nP_r = \dfrac{n!}{(n-r)!}$
복원	${}_nH_r = \dfrac{(n+r-1)!}{(n-1)!r!}$	${}_n\Pi_r = n^r$

02

이산형 확률분포

1 이산형 균등분포

- 모든 값이 동일한 확률을 가지는 확률분포
  - **주사위 던지기**
  - $f(x) = \frac{1}{N}, x = 1,2,\cdots,N$



1 이산형 균등분포

기댓값, 분산, 적률생성함수

- $E(X) = \frac{N+1}{2}, Var(X) = \frac{(N+1)(N-1)}{12}$
- $M(t) = \frac{1}{N}\sum_{i=1}^{N} e^{it}$

1 이산형 균등분포

예 주사위를 한 번 던져서 나오는 숫자를 확률변수 X라고 할 때 확률변수 X의 확률질량함수와 기댓값은?

2 베르누이분포

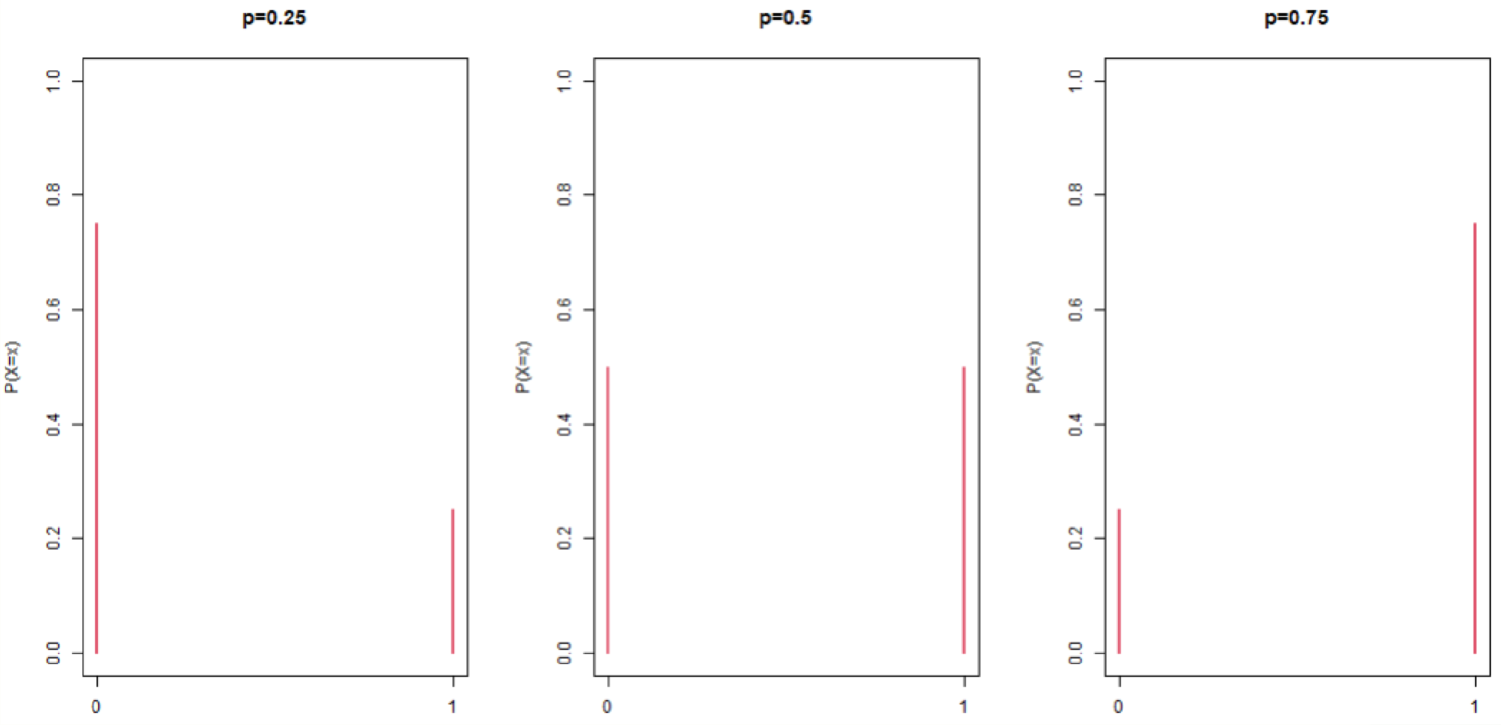
베르누이 시행

- 시행 결과 두 가지 중 하나가 나타나는 실험
  - 예 : 불량품 여부, 찬성 여부, 성공과 실패

2 베르누이분포

베르누이 분포

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



2 베르누이분포

베르누이 분포의 기댓값

- $E(X) = p$
- $Var(X) = p(1-p)$

2 베르누이분포

베르누이 분포의 적률생성함수

- $M(t) = (1-p) + pe^t$

3 이항분포

이항분포

- **성공률이 0.05인 베르누이 시행을 3회 독립 시행했을 때 성공횟수 X의 분포**
- $\binom{n}{x} = {}_nC_x$

X	0	1	2	3	합
$P(X)$	${}_3C_0\left(\frac{5}{100}\right)^0\left(\frac{95}{100}\right)^3$	${}_3C_1\frac{5}{100}\left(\frac{95}{100}\right)^2$	${}_3C_2\left(\frac{5}{100}\right)^2\frac{95}{100}$	${}_3C_3\left(\frac{5}{100}\right)^3\left(\frac{95}{100}\right)^0$	1

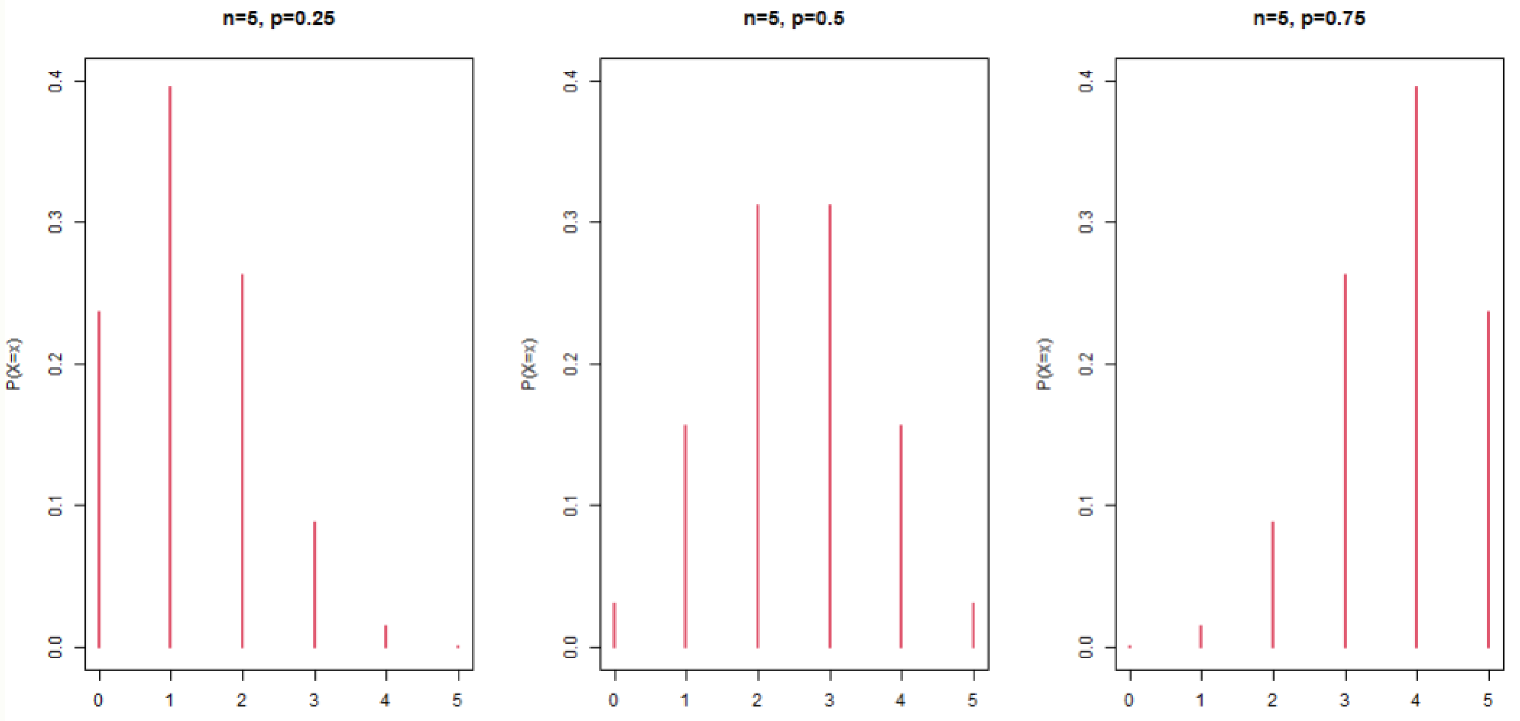
3 이항분포

이항분포

- **성공률이 p인 베르누이 시행을 n회 독립 시행했을 때 성공횟수 X의 분포 :** $X \sim B(n,p)$
- $f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0,1,2,\cdots,n$

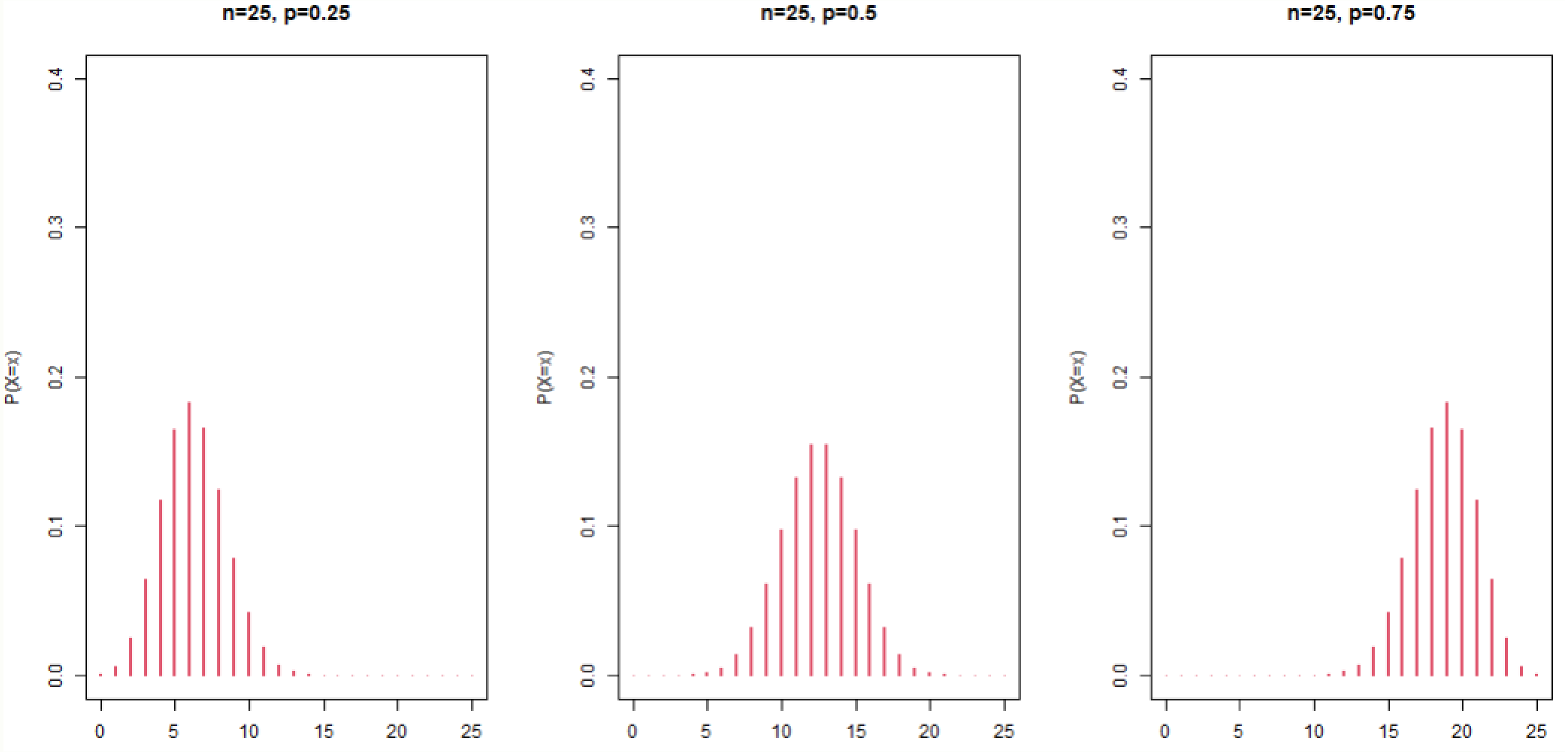
3 이항분포

- 이항분포 : $X \sim B(5, p), p = 0.25, 0.5, 0.75$



3 이항분포

- 이항분포 : $X \sim B(25, p), p = 0.25, 0.5, 0.75$



3 이항분포

이항정리

- $\sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} = (a+b)^n$

3 이항분포

- 이항분포 $B(n, p)$의 기댓값
  - $E(X) = np$
  - $\mathrm{Var}(X) = np(1-p)$

3 이항분포

- 이항분포 $B(n, p)$의 기댓값
  - $E(X) = np$
  - $\mathrm{Var}(X) = np(1-p)$

3 이항분포

예 동전을 3번 던져서 앞면이 나온 횟수를 확률변수 X 라고 할 때 가 X 가 2 이상일 확률과 기댓값은?

3 이항분포

- 이항분포 $B(n, p)$의 적률생성함수
  - $M(t) = [(1-p) + pe^t]^n$

3 이항분포

약대수의 법칙(weak law of large numbers)

- $\lim_{n\to\infty} P(|\frac{X}{n} - p| \geq \varepsilon) = 0, \quad \varepsilon > 0$

3 이항분포

이항분포의 성질

- $X_1 \sim B(n_1, p), X_2 \sim B(n_2, p)$

  $\Rightarrow \; X = X_1 + X_2 \sim B(n_1 + n_2, p)$

4 초기하분포

초기하분포

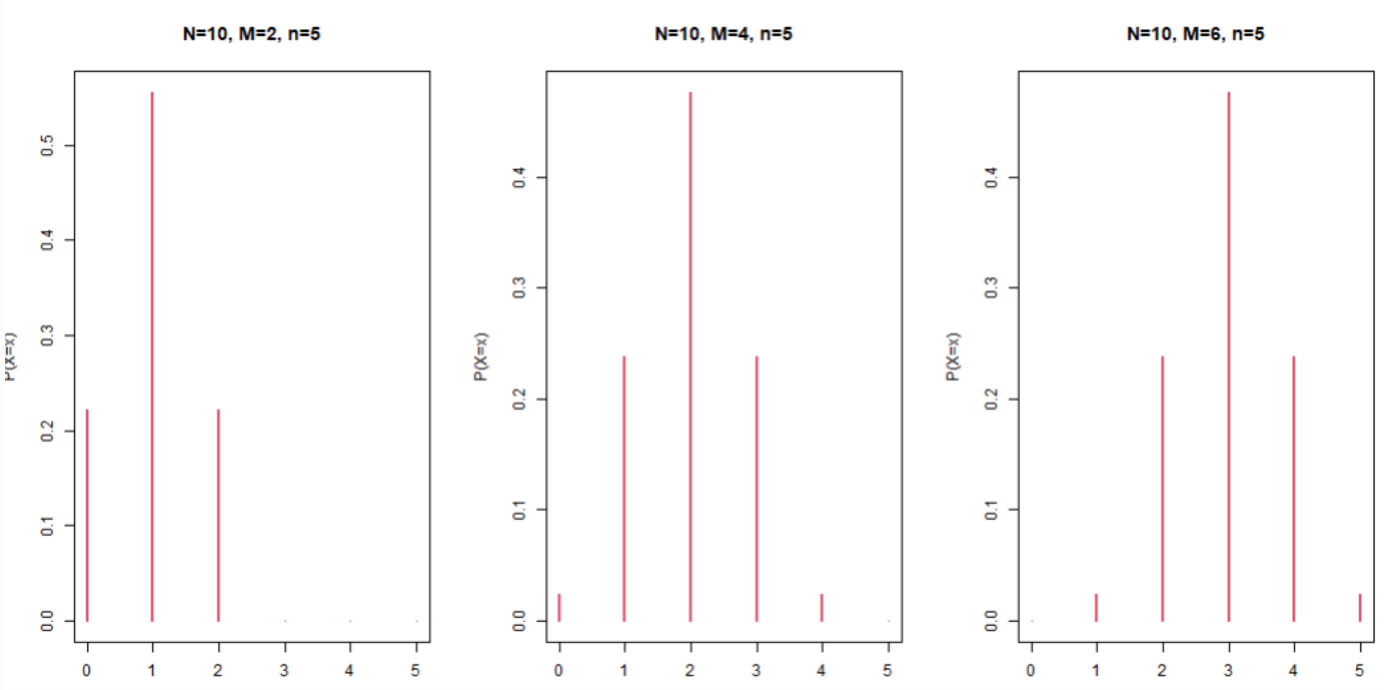
- **항아리 속 공 N개 = 흰색 공 M개 + 검정색 공 $(N-M)$개로 구성 → n 개의 공 비복원 추출. 흰색공 개수 X의 분포**

$$X \sim Hyper(n; N, M)$$

4 초기하분포

확률질량함수

- $f(x) = \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, x = 0,1,2,\cdots,n$



4 초기하분포

- 초기하분포 $Hyper(n;N,M)$의 기댓값
  - $E(X) = \frac{nM}{N}$

4 초기하분포

- 초기하분포 $Hyper(n; N, M)$의 기댓값
  - $E(X) = \frac{nM}{N}$

4 초기하분포

- 초기하분포 $Hyper(n; N, M)$의 기댓값
  - $E(X) = \frac{nM}{N}$
  - $\mathrm{Var(X)} = \frac{nM}{N}\frac{(N-M)}{N}\frac{(N-n)}{(N-1)}$

4 초기하분포

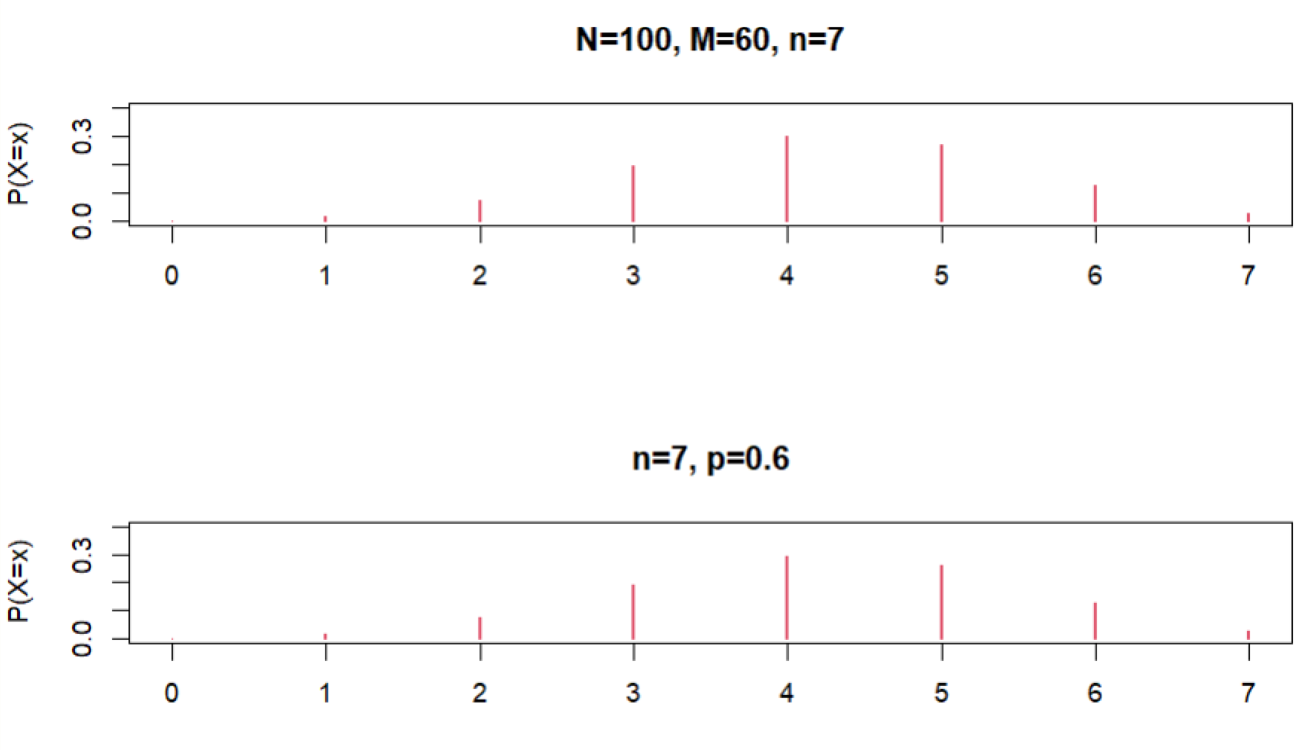
예 흰 공 2개와 검정색 공 3개가 있다. 2개의 공을 비복원 추출했을 때 흰색공의 개수의 분포는?

4 초기하분포

초기하분포의 성질

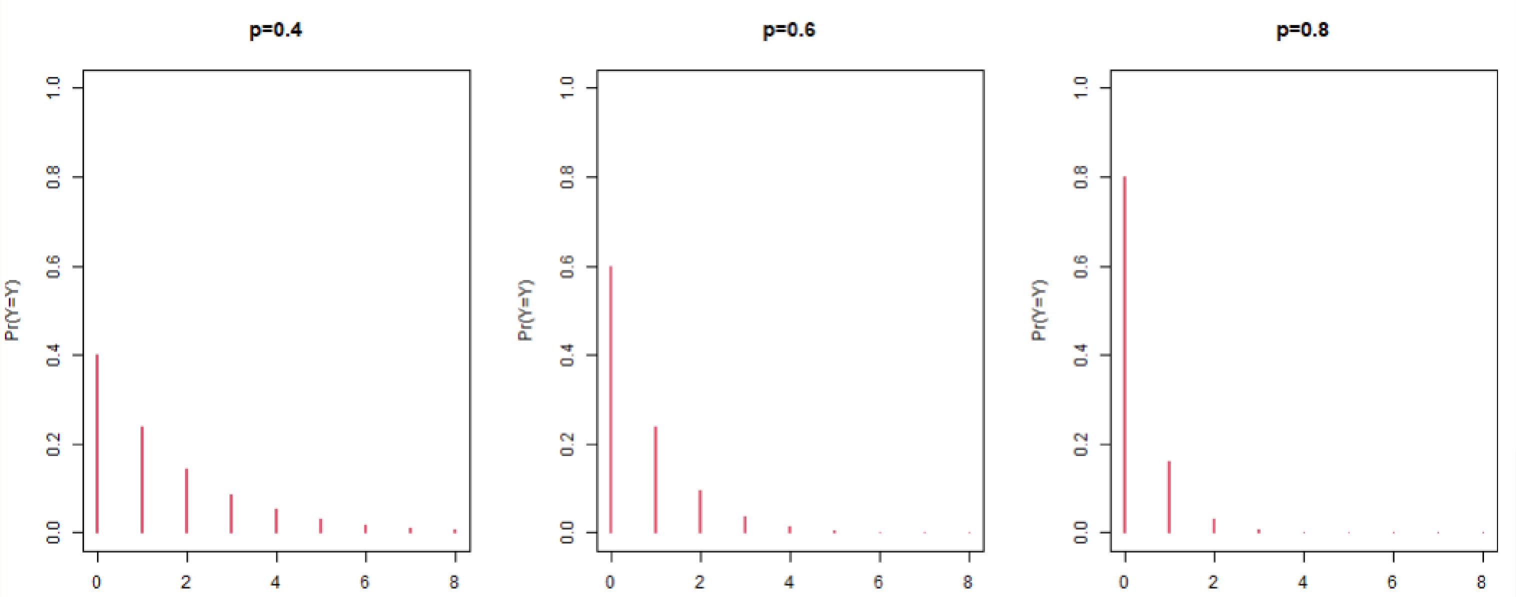
- $N \longrightarrow \infty, p = \frac{M}{N}, X \sim HYP(n; N, M)$

  $\Longrightarrow \; X \sim B(n, p)$



5 기하분포

- 베르누이 시행을 지속할 때 첫번째 성공에 이르기까지의 시행 횟수 X의 분포 : $X \sim Geometric(p)$
  - $f(x) = p(1-p)^{x-1}, x = 1,2,\cdots$



5 기하분포

- 기하분포 $Geometric(p)$의 적률생성함수
  - $M(t) = (1-(1-p)e^t)^{-1}e^t p, \quad t < -\log(1-p)$

5 기하분포

기하분포 $Geometric(p)$의 기댓값

- $\mathrm{E}(X) = \frac{1}{p}, \; Var(X) = \frac{1-p}{p^2}$

5 기하분포

- 기하분포 $Geometric(p)$ 의 특징 : 무기억성
  - $P(X > m + n|X > n) = P(X > m)$

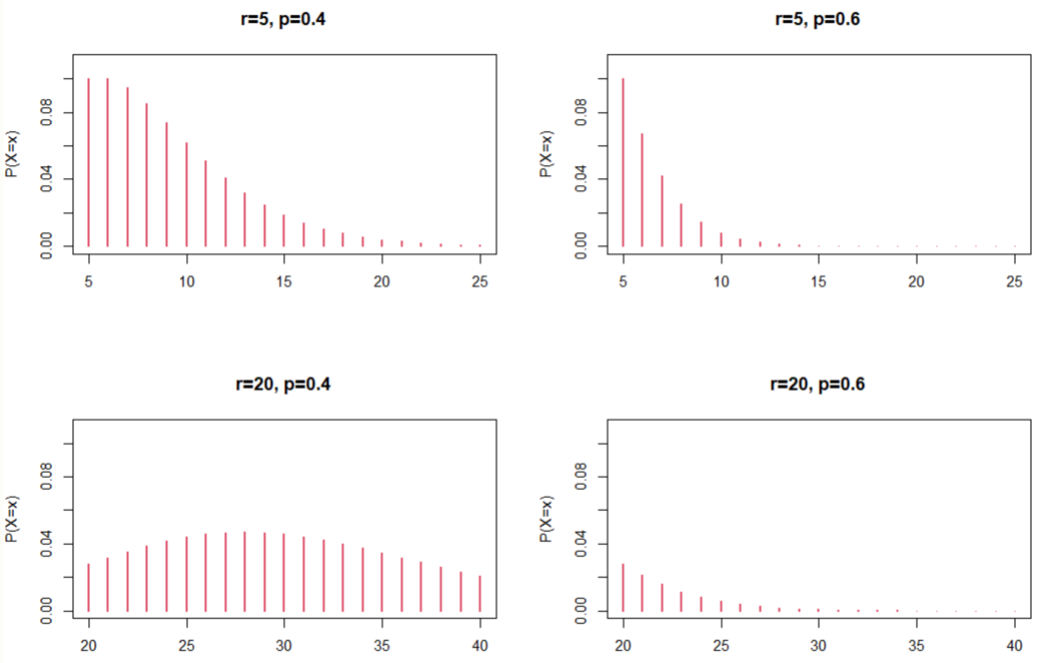
5 기하분포

- 기하분포 $Geometric(p)$ 의 특징 : 무기억성
  - $P(X > m + n | X > n) = P(X > m)$

6 음이항분포

- 베르누이 시행을 지속할 때 r번째 성공에 이르기까지의 시행 횟수 X의 분포 : $X \sim NB(r, p)$

  - $f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, x = r, r+1, r+2, \cdots$



6 음이항분포

음이항분포 $NB(r,p)$의 기댓값과 적률생성함수

- $E(X) = \frac{r}{p}$, $Var(X) = \frac{r(1-p)}{p^2}$
- $M(t) = [(1-(1-p)e^t)^{-1}e^t p]^r$, $t < -\log(1-p)$

6 음이항분포

음이항분포 $NB(r,p)$의 성질

- $X_1 \sim NB(r_1, p), X_2 \sim NB(r_2, p)$

  $\Rightarrow X = X_1 + X_2 \sim NB(r_1 + r_2, p)$

- $X_i \sim Geometric(p), 1 \le p \le r$

  $\Leftrightarrow X = X_1 + X_2 + \cdots + X_r \sim NB(r, p)$

7 포아송분포

- 특정 기간 또는 영역에서 일어나는 사건 수 X :

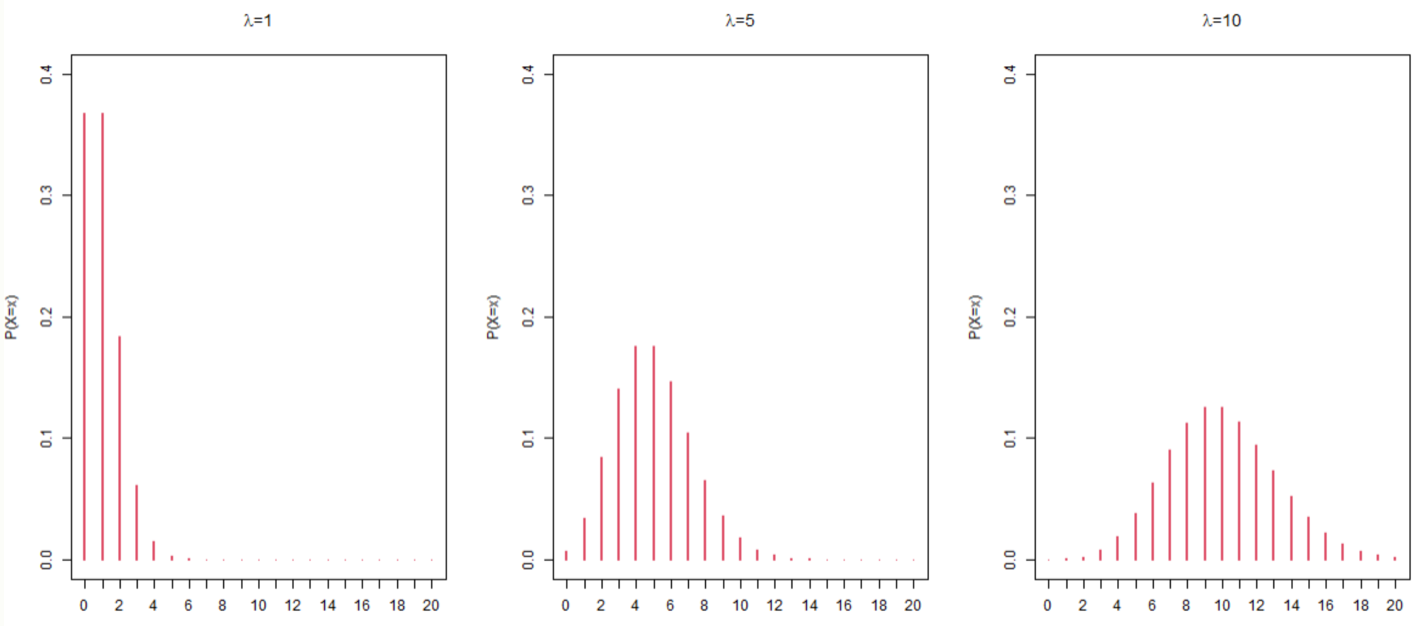
$$X \sim Poisson(\lambda)$$

  - S. D. Poisson(1781-1840)가 1837년 소개
  - 첫 소개는 1711년 A. de Movire가 제시

7 포아송분포

- $Poisson(\lambda)$의 확률질량함수

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0,1,2,\cdots, \lambda > 0$$



7 포아송분포

- $Poisson(\lambda)$의 기댓값

$$E(X) = Var(X) = \lambda$$

7 포아송분포

- $Poisson(\lambda)$의 적률생성함수

$$M(t) = e^{\lambda(e^t - 1)}$$

7 포아송분포

예 어떤 지역에서 하루 평균 1건 교통사고 발생.
일주일 동안 사고가 3건 이하로 일어날 확률은?

7 포아송분포

포아송분포의 성질

- n은 크고 p는 작으면서 $\lambda = np$는 일정

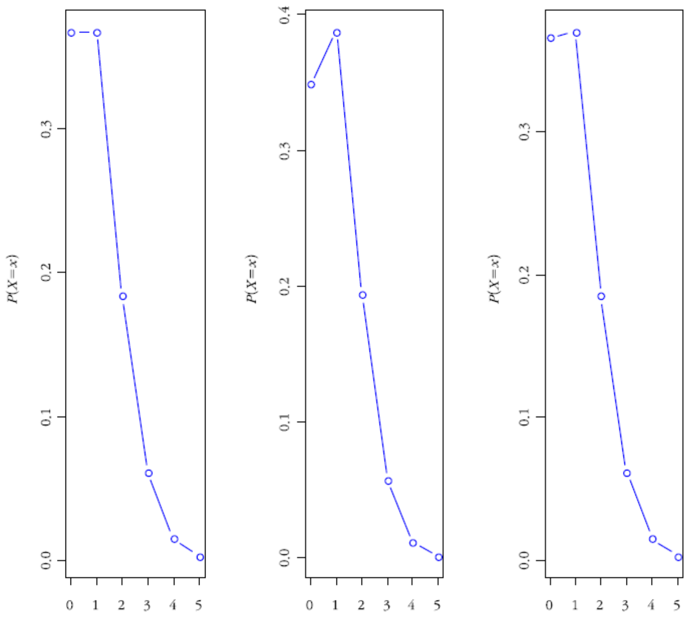
$$\Rightarrow \lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!}$$

7 포아송분포

포아송분포의 성질

- n은 크고 p는 작으면서 $\lambda = np$는 일정

$$\Rightarrow \lim_{n\to\infty} \binom{n}{x} p^{x} (1-p)^{n-x} = \frac{\lambda^{x} e^{-\lambda}}{x!}$$



(a) $Poisson(1)$ (b) $B(10,\ 0.1)$ (c) $B(100,\ 0.01)$

7 포아송분포

포아송분포의 성질

- $X_1 \sim Poisson(\lambda_1), X_2 \sim Poisson(\lambda_2)$

  $\Rightarrow X = X_1 + X_2 \sim Poisson(\lambda_1 + \lambda_2)$

8 다항분포

삼항분포

- 결과가 세 가지(찬성, 반대, 무의견)중 하나인 실험을 독립적 n번 시행, 찬성의 수 X와 반대의 수 Y의 분포
- $f(x,y) = \binom{n}{x,y,n-x-y} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}$

8 다항분포

삼항분포의 성질

- $X \sim B(n, p_1),\ Y \sim B(n, p_2)$
- $Y|X = x \sim B(n - x, \frac{p_2}{1-p_1})$
- $Cov(X, Y) = -np_1p_2$

8 다항분포

다항분포(multinomial distribution)

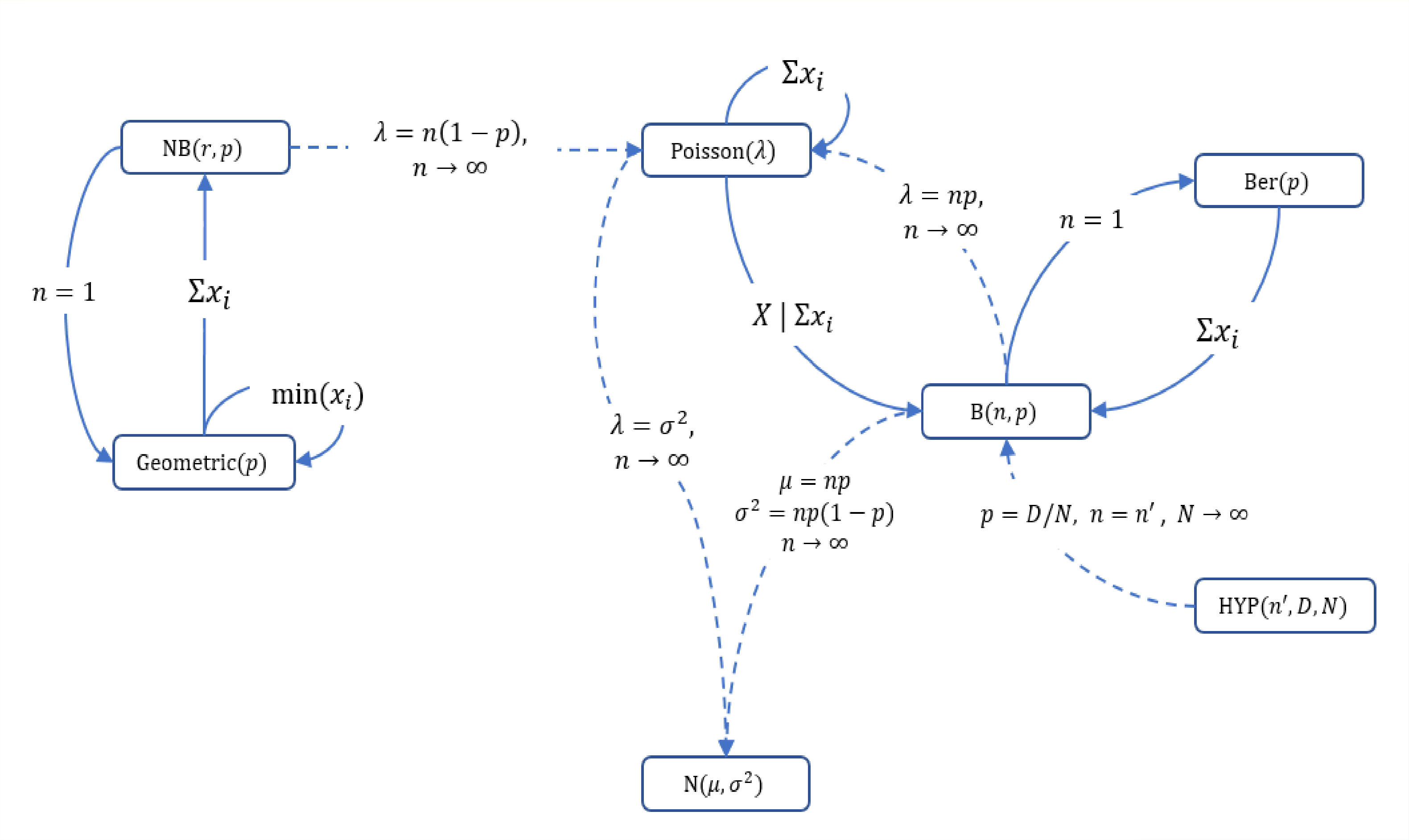
- 결과가 k 개 범주 중 하나인 실험을 독립적 n번 시행, k 범주로 나타난 수 $X_1, X_2, \cdots, X_k$ 의 분포

$$f(x_1, x_2, \cdots, x_k) = \binom{n}{x_1, x_2, \cdots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

$$\sum_{i=1}^{k} x_i = n, \sum_{i=1}^{k} p_i = 1, x_i = 0,1, \cdots, n$$

8 이산형 확률분포

이산형 확률분포의 관계



정리하기

- 이산형 균등분포는 이산형 표본공간의 확률이 모두 같을 때의 분포이다.
- 베르누이 분포는 확률변수가 서로 배반적인 두 값을 가질 때의 분포이다.
- 이항분포는 베르누이 독립시행에서 성공(실패) 횟수의 분포이다.
- 초기하분포는 두가지 속성으로 구성된 집단에서 비복원 추출했을 때 한 속성이 추출된 건수의 분포이다.

정리하기

- 기하분포는 첫번째 성공이 일어나기까지 베르누이 독립시행 횟수의 분포이다.
- 음이항분포는 일정 횟수의 성공이 일어나기까지 베르누이 독립시행 횟수의 분포이다.
- 포아송분포는 일정기간 동안 발생한 희귀 사건 건수의 분포이다.
- 다항분포는 여러 범주 중 하나만 나타날 수 있는 실험에서 각 범주가 나타나는 수들의 분포이다.

05 강

다음시간 안내

연속형 확률분포