확률 및 통계학

- 3장 이산형 확률분포(1) -

송 영 준(youngjun@pel.smuc.ac.kr)

상명대학교 프로토콜공학연구실

목 차

- 이산형 균일분포
- 이항분포
- 다항분포

이산형 균일분포

- 균일분포(Uniform Distribution)
  - 과거의 경험이 미래를 예측하는데 어떠한 영향도 미치지 않으며, 나타날 가능성이 모두 동일한 분포
    - 이산형 균일분포, 연속형 균일분포

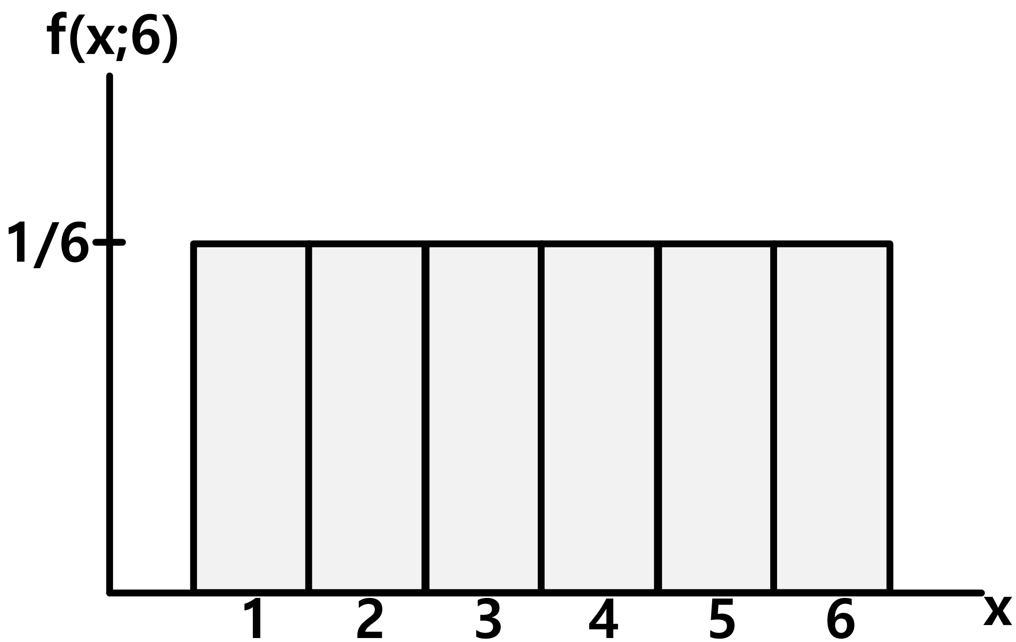
- 이산형 균일분포(Discrete Uniform Distribution)
  - 확률 변수가 취할 수 있는 각 값들이 모두 같은 확률을 가지는 경우
  - 확률변수 X가 $x_1, x_2, \dots, x_k$의 각 값을 취할 확률이 동일한 경우, $f(x;k) = \frac{1}{k}, x = x_1, x_2, x_3, \dots, x_k$로 주어짐
  - 모수 k에 종속되어 있음을 나타내기 위하여 $f(x)$대신에 $f(x;k)$라는 기호를 사용

이산형 균일분포

- 예제 5.1
  - 40와트, 60와트, 75와트, 100와트의 전구가 들어 있는 상자에서 임의로 하나의 전구를 꺼낼 때, 표본공간 $S = \{40,60,75,100\}$의 각 원소는 의 발생확률을 가짐
    - X의 확률분포 : $f(x;4) = \frac{1}{4}, x = 40,60,75,100$인 균일분포

이산형 균일분포

- 예제 5.2
  - 하나의 주사위를 던지는 실험에서 표본공간 $S = \{1,2,3,4,5,6\}$의 각 원소의 발생확률은 $\frac{1}{6}$로 같음
    - X의 확률분포는 $f(x;6) = \frac{1}{6}$, $x = 1,2,3,4,5,6$인 균일분포



이산형 균일분포

- 이산형 균일분포(Discrete Uniform Distribution)
  - 이산형 균일분포의 평균, 분산
    - 평균 : $\mu = \frac{1}{k}\sum_{i=1}^{k} x_i$
    - 분산 : $\sigma^2 = \frac{1}{k}\sum_{i=i}^{k}(x_i - \mu)^2$
  - 증명
    - $\mu = E(x) = \sum_{i=i}^{k} x_i f(x_i; k) = \sum_{i=1}^{k} \frac{x_i}{k} = \frac{1}{k}\sum_{i=1}^{k} x_i$
    - $\sigma^2 = \mathrm{E}[(\mathrm{X} - \mu^2)] = \sum_{i=1}^{k}(x_i - \mu)^2 f(x_i; k) = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2$

이산형 균일분포

- 예제5.3
  - 예제5.2에서 X의 평균과 분산
    - X의 확률분포는 $f(x;6)=\frac{1}{6}$, $x=1,2,3,4,5,6$인 균일분포
      - $\mu=\frac{1}{k}\sum_{i=1}^{k}x_i$, $\frac{1}{6}\sum_{i=1}^{6}x_i=\frac{1+2+3+4+5+6}{6}=\frac{21}{6}=3.5$
      - $\sigma^2=\frac{1}{k}\sum_{i=1}^{k}(x_i-\mu)^2$, $\frac{1}{6}\sum_{i=1}^{6}(x_i-\mu)^2=\frac{1}{6}\sum_{i=1}^{6}(x_i-3.5)^2$
        $=\frac{1}{6}[(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2=2.92$

이항분포

- 어떤 실험을 하거나 또는 표본을 뽑을 때, 그 실험의 결과 또는 표본을 뽑는 결과가 상호 배타적인 두 가지 사건으로만 나타나는 경우
  - 동전 앞/뒤, 어떤 사건의 성공/실패, 환자가 병에 걸리거나 정상인 경우, 제품 확인 시 정상/비정상
  - 베르누이 시행(Bernoulli trial)
    - 실험은 n번의 반복시행으로 구성
    - 각 시행의 결과는 성공이나 실패의 두 가지 중 하나가 됨
    - p로 표시되는 성공확률은 매 시행마다 일정
    - 각 시행은 서로 독립

이항분포

- 이항분포(Binomial Distribution)
  - n번의 베르누이 시행에서의 성공횟수 X를 이항확률변수(Binomial Random Variable)이라 하고, 이 확률분포를 이항분포(Binomial Distribution)이라 함
    - 확률은 시행횟수와 각 시행에서의 성공확률에 종속되어 있으므로 $b(x;n,p)$로 표현
      - x: 성공 횟수
      - n: n회 시행
      - p: 성공 확률
  - 성공확률 : p, 실패확률 : $q\ (1-p)$인 베르누이 시행의 n회 독립시행에서 성공의 횟수를 나타내는 이항확률변수 X의 확률분포
    - $b(x;n,p) = \binom{n}{x} p^x q^{n-x}, x = 0,1,2,\ldots,n$

이항분포

- 예제5.4
  - 어떤 종류의 부품이 충격실험에서 충격을 견딜 확률이 $\frac{3}{4}$일 때, 4개의 부품에 대한 실험에서 정확하게 2개가 충격을 견딜 확률
    - $\mathrm{b}\left(2; 4, \frac{3}{4}\right) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \left(\frac{4!}{2!2!}\right)\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = \frac{27}{128}$

이항분포

- 이항분포(Binomial Distribution)
  - $(q+p)^n$의 이항전개에서 각 항이 $x = 0,1,2,\ldots,n$일 때, $b(x;n,p)$의 각 값들과 일치한다는 사실에서 연유
    - 이항정리
      - 두 개의 항을 곱셈에 대하여 전개한 식
      - $(a+b)^n$의 원하는 항의 계수를 찾을 때 유용
      - 이항정리를 사용해 차수가 큰 식의 전개를 빠르게 할 수 있음
    - 정의
      - n이 양의 정수일 때, $(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{r}a^{n-r}b^r + \cdots + \binom{n}{n}b^n$, 여기서 $\binom{n}{r}a^{n-r}b^r$를 일반항, $\binom{n}{r}$를 이항계수라고 함
      - 일반항을 써서 간단히 정리하면, $(a+b)^n = \sum_{r=0}^{n}\binom{n}{r}a^{n-r}b^r$

이항분포

- 이항분포(Binomial Distribution)
  - $(q+p)^n$의 이항전개에서 각 항이 $x = 0,1,2,\ldots,n$일 때, $b(x;n,p)$의 각 값들과 일치한다는 사실에서 연유
    - 이항정리
      - $(a+b)^3$의 전개식에서 ab^2의 계수를 찾고 싶을 경우
        - $(a+b)\ (a+b)\ (a+b)$에서 각 항에서 순서에 상관 없이 a를 한번, b를 두 번 뽑는 경우의 수
        - 계수는 $\binom{3}{2} = 3$
      - $(a+b)^5 = \binom{5}{0}a^5 + \binom{5}{1}a^4b^1 + \binom{5}{2}a^3b^2 + \binom{5}{3}a^2b^3 + \binom{5}{4}a^1b^4 + \binom{5}{5}b^5$

이항분포

- 누적이항분포
  - 이항분포의 누적 합 $B(r;n,p) = \sum_{x=0}^{r} b(x;n,p)$
    - 누적이항분포 표
    - $p + q = \sum_{x=0}^{n} b(x;n,p) = 1$

n	x	P 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
1	0	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
	1	0.0500	0.1000	0.1500	0.2000	0.2500	0.3000	0.3500	0.4000	0.4500	0.5000
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
	2	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
	2	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
	3	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
	2	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
	3	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
	4	0.0000	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1563
	2	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
	3	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1811	0.2304	0.2757	0.3125
	4	0.0000	0.0005	0.0022	0.0064	0.0146	0.0284	0.0488	0.0768	0.1128	0.1563
	5	0.0000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0313

이항분포

- 예제5.5
  - 빈혈환자가 회복될 확률이 0.4일 때, 15명이 빈혈에 걸렸을 때 확률
    - (a) 적어도 10명이 회복될 확률
      - $P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15,0.4) = 1 - 0.966$
        $= 0.0338$
    - (b) 3명에서 8명 사이의 사람이 회복될 확률
      - $P(3 \leq X \leq 8) =$
        $\sum_{x=3}^{8} b(x; 15,0.4) = \sum_{x=0}^{8} b(x; 15,0.4) - \sum_{x=0}^{2} b(x; 15,0.4)$
        $= 0.9050 - 0.0271 = 0.8779$
    - (c) 정확히 5명이 회복될 확률
      - $P(X = 5) = b(5; 15,0.4) = \sum_{x=0}^{5} b(x; 15,0.4) - \sum_{x=0}^{4} b(x; 15,0.4)$
        $= 0.4032 - 0.2173 = 0.1859$

이항분포

- 예제5.6
  - 어느 대형 할인점 상인이 전자장비를 제조업자로부터 납품 받을 때, 장비의 불량률이 3%인 경우
    - (a) 상인이 납품된 제품 중 20개를 임의로 선택하여 검사할 때, 이 20개 중 적어도 한 대 의 불량품이 있을 확률
      - $P(X \geq 1) = 1 - P(X = 0) = 1 - b(0; 20{,}0.03)$
        $1 - 0.03^{0}(1 - 0.03)^{20-0} = 0.4562$
    - (b) 한 달에 10번의 납품을 받고, 납품 시 마다 20개의 장비를 검사한다고 할 때, 적어도 한 대의 불량품이 포함된 납품이 3번 있을 확률
      - 각 납품 시 검사 확률 $p{=}0.4562$
      - $b(y; 10{,}0.4562) = \mathrm{P}(\mathrm{Y} = 3) = \binom{10}{3}0.4562^{3}(1 - 0.4562)^{7} = 0.1602$

이항분포

- 이항분포(Binomial Distribution)
  - 평균과 분산
    - $\mu = np, \sigma^2 = npq$
  - $\mu = np$ 증명
    - $E(X) = \sum_{r=0}^{n} r * nCr p^r q^{n-r}$
    - $r * nCr = r\frac{n!}{(n-r)!r!} = \frac{n*(n-1)!}{(n-r)!(r-1)!} = n * \binom{n-1}{r-1}$
    - $\sum_{r=1}^{n} n * \binom{n-1}{r-1} pp^{r-1} q^{n-r} = np \sum_{r=1}^{n} \binom{n-1}{r-1} p^{r-1} q^{n-r}$
    - $np * (p+q)^{n-1}, (p+q) = 1$
    - $\therefore E(X) = np$

이항분포

- 이항분포(Binomial Distribution)
  - 평균과 분산
    - $\mu = np,\ \sigma^2 = npq$

  - $\sigma^2 = npq$ 증명
    - $\sigma^2 = E[X^2] - \mu^2$
    - $\sigma^2 = E[X^2] - (np)^2$
    - $E[X^2] = \sum_{r=0}^{n} r^2 \binom{n}{r} p^r q^{n-r},\ (q = 1 - p)$
    - $E[X^2] = \sum_{r=0}^{n} (r^2 - r + r) \binom{n}{r} p^r q^{n-r} = \sum_{r=0}^{n} r(r-1) \binom{n}{r} p^r q^{n-r} + \sum_{r=0}^{n} r \binom{n}{r} p^r q^{n-r}$,
    $\sum_{r=0}^{n} r \binom{n}{r} p^r q^{n-r} = np$
    - $\sum_{r=0}^{n} r(r-1) \binom{n}{r} p^r q^{n-r} + np$

이항분포

- 이항분포(Binomial Distribution)
  - $\sigma^2 = npq$ 증명
    - $E[X^2] = \sum_{r=0}^{n} r(r-1)\binom{n}{r} p^r q^{n-r} + np$
    - $\sum_{r=2}^{n} r(r-1)\frac{n!}{(n-r)!r!} p^r q^{n-r} + np$
    - $\sum_{r=2}^{n} r(r-1)\frac{n\cdot(n-1)\cdot(n-2)!}{(n-r)!\cdot r\cdot(r-1)\cdot(r-2)!} p^2 p^{r-2} q^{n-r} + np$
    - $n(n-1)p^2 \sum_{r=2}^{n} \frac{(n-2)!}{(n-r)!(r-2)!} p^{r-2} q^{(n-2)-(r-2)} + np$, $(r-2) = k$로 치환
    - $\frac{(n-2)!}{(n-r)!(r-2)!} = \binom{(n-2)}{k}$
    - $n(n-1)p^2 \sum_{k=0}^{n-2}\binom{(n-2)}{k} p^r q^{n-2-k} + n$
    - $\sum_{k=0}^{n-2}\binom{(n-2)}{k} p^r q^{n-2-k} = (p+q)^{n-2} = 1^{n-2} = 1, (p+q=1)$

이항분포

- 이항분포(Binomial Distribution)
  - $\sigma^2 = npq$ 증명
    - $E[X^2] = n(n-1)p^2 + np$
    - $\sigma^2 = n(n-1)p^2 + np - (np)^2$
    - $n^2p^2 - np^2 + np - (np)^2$
    - $np(1-p) = npq$
    - $\therefore \sigma^2 = npq$

이항분포

- 예제5.7
  - 어느 시골에 있는 우물 중 30%가 불순물이 들어있을 때, 10개의 우물을 임의로 선정하여 검사
    - (a) 정확히 3개 우물에 불순물이 있을 확률의 이항분포
      - $b(3; 10{,}0.3) = P(X = 3) = \sum_{x=0}^{3} b(x; 10{,}0.3) - \sum_{x=0}^{2} b(x; 10{,}0.3)$
        $= 0.6496 - 0.3828 = 0.2668$
    - (b) 3개를 초과한 우물에 불순물이 있을 확률
      - $P(X > 3) = 1 - P(X \le 3) = 1 - \sum_{x=0}^{3} b(x; 10{,}0.3) = 1 - 0.6496 = 0.3504$

이항분포

- 예제5.8
  - 예제5.5의 이항확률변수에 대한 평균과 분산을 구하고, 체비셰프 정리를 사용하여 구간 $\mu \pm 2\sigma$의 의미
    - $n = 15,\ p = 0.4, q = 0.6$
    - $\mu = np = 15 * 0.4 = 6$
    - $\sigma^2 = npq = 15 * 0.4 * 0.6 = 3.6$
    - $\sigma = 1.897$
    - $\mu \pm 2\sigma = 6 \pm 2(1.897) = 2.206 \sim 9.794$
    - 자료가 이산형이므로 3~9명까지가 될 확률이 적어도 $\frac{3}{4}$(75%) 이상

이항분포

- 체비셰프 정리
  - 자료의 분포가 정규분포가 아니거나 혹은 분포를 모르는 경우 체비셰프의 정리가 적용
    - 확률분포의 평균과 분산만을 알고 있어 이 값이 정확하게 어떤 분포를 하고 있는지 알 수 없는 상태일 때 체비셰프 정리를 이용
  - 어떤 확률 변수 X가 주어지고, 확률변수의 평균과 표준편차가 각 각 μ, σ일 때, 임의의 상수 $K > 0$에 대하여 항상 다음 식을 만족
    - $P(|X - \mu| < K\sigma) \geq 1 - \frac{1}{k^2}$

이항분포

- 예제5.9
  - 예제5.7에서 30%가 오염이라는 것은 단순 추측된 가설일 경우, 10개 우물을 검사한 결과 6개 우물에서 불순물이 발견되었을 때, 이것은 가설에 대한 의미, 확률적 개념을 이용한 설명
    - $P(X \geq 6) = \sum_{x=0}^{10} b(x;10,0.3) - \sum_{x=0}^{5} b(x;10,0.3) = 1 - 0.9527 = 0.0473$

다항분포

- 다항실험과 다항분포
  - 각 시행에 가능한 결과가 둘 이상이 되면 다항실험이 됨
  - 각 시행에서 $p_1, p_2, \ldots, p_k$의 확률로 가능한 결과 $E_1, E_2, \ldots E_k$ 중 어느 하나가 일어난다고 하면, 다항분포(Multinomial Distribution)는 n번의 독립시행에서 E_1이 x_1번, E_2가 x_2번,…, E_k이 x_k번 일어날 확률($x_1 + x_2 + \cdots + x_k = n$)
    - $f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n)$으로 표기
      - $(p_1 + p_2 + \cdots + p_k = 1)$

다항분포

- 다항실험과 다항분포
  - 각 시행에서 $p_1, p_2, \dots, p_k$의 확률로 k개의 결과 $E_1, E_2, \dots, E_k$ 중 어느 하나가 발생한다면, n번의 독립시행에서 각 각 $E_1, E_2, \dots, E_k$의 발생횟수를 나타내는 확률변수 $X_1, X_2, \dots, X_k$의 확률 분포
    - $f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n) = \binom{n}{x_1, x_2, \dots, x_k} {p_1}^{x_1} {p_2}^{x_2} \dots {p_k}^{x_k}$
    - $\sum_{i=0}^{k} x_i = n, \sum_{i=0}^{k} p_i = 1$

다항분포

- 예제 5.10
  - 어느 공항의 항공기 이착륙 상황에 대한 이상적인 조건을 알아보기 위해 컴퓨터 시뮬레이션이 수행되었을 때, 3개의 활주로가 있는 이 공항에서 각 활주로가 사용될 확률은 다음과 같고, 임의로 도착하는 6대의 비행기가 다음과 같이 활주로에 도착할 확률
    - 활주로 1 : $p_1 = \frac{2}{9}$, 활주로 2 : $p_2 = \frac{1}{6}$, 활주로 3 : $p_3 = \frac{11}{18}$
    - 활주로 1 : 2대, 활주로 2 : 1대, 활주로 3: 3대
      - $f\left(2,1,3;\frac{2}{9},\frac{1}{6},\frac{11}{18},6\right) = \binom{6}{2,1,3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3$

$$= \frac{6!}{2!\,1!\,3!}.\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3 = 0.1127$$

감사합니다!