■ 이산분포(Discrete distribution)

- 확률변수의 치역이 셀 수 있는 경우
  - 확률질량함수로 표시
- 베르누이 분포(Bernoulli distribution)
  - 베르누이 시행(두가지 결과(S, F), 독립, 확률 불변)
  - S이면 1, F이면 0
  - 확률질량함수

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

  - $E(X) = 0 \times (1-p) + 1 \times p = p$
  - $Var(X) = p - p^2 = p(1-p) \quad \Rightarrow SD(X) = \sqrt{p(1-p)}$

- 이항분포(Binomial distribution)
  - 성공할 확률이 p인 베르누이 실험을 n 번 반복했을 때, 성공 횟수(X)의 분포
  - 베르누이 확률변수의 합
  - $E(X) = E(X_1 + \cdots + X_n) = np$
  - $Var(X) = Var(X_1 + \cdots + X_n) = np(1-p)$
  - $SD(X) = \sqrt{np(1-p)}$

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

- 통계학 문제: 모수(주로 p)는 얼마인가?

- 기하분포(Geometric distribution)
  - 성공할 확률이 p 인 베르누이 시행을 성공할 때 까지 시행하는 경우
  - 실패(시행) 횟수의 분포
  - 확률질량함수: $f(x)=(1-p)^x p, \quad x=0,1,2, \ldots$
  - 무기억성(memoryless): x 번째까지 실패했다고 할 때, 다음($x+1$번째)시행에서의 성공확률

$$\frac{f_X(x)}{P(X \geq x)}=\frac{(1-p)^x p}{(1-p)^x}=p$$

  - $$E(X)=\sum_{x=0}^{\infty} xp(1-p)^x=p(1-p)/p^2=(1-p)/p$$
- 통계학 문제: 모수(주로 p)는 얼마인가?

- 음이항분포(Negative Binomial distribution)
  - 성공할 확률이 p 인 베르누이 시행을 r 번 성공할 때까지 시행하는 경우 실패(시행)횟수의 분포

$$f(x) = \binom{x+r-1}{r-1} p^r (1-p)^x, \quad x = 0, 1, 2, \ldots$$

  - $E(X) = E(X_1 + \cdots + X_r) = r(1-p)/p$
  - 계수자료 분석에서 포아송분포의 대안으로 사용가능

    포아송분포: $E(X) = \lambda = Var(X)$

    $\Rightarrow$ $\bar{x}$ 와 s^2의 차이가 심하면($\bar{x} \ll s^2$), 포아송분포?

- 다항분포(Multinomial distribution)
  - 이항분포: 결과가 두가지 $\Rightarrow$ k가지
  - $(X_1, X_2, \ldots, X_k)$: n번 시행했을 때, 각 결과의 횟수

$$f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \quad \sum_{i=1}^{k} x_i = n, \ \sum_{i=1}^{k} p_i = 1$$

  - $X_i \sim B(n, p_i)$
  - $Cov(X_i, X_j) = -n p_i p_j \Rightarrow Cor(X_i, X_j) = -\sqrt{\dfrac{p_i p_j}{(1-p_i)(1-p_j)}}$
  - 범주가 k 개인 도수분포표의 확률모형

- 포아송분포(Poisson distribution)
  - 계수(count)자료에 대한 통계모형
    - 구간을 나누었을 때 각 구간의 발생 빈도는 서로 독립 (independent increment)
    - 구간의 위치와 관계없이 동일 길이의 구간에서의 평균발생 빈도는 동일 (stationary increment)
  - n 이 크고 p 가 작은 이항분포의 근사모형
  - 확률질량함수: $f(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!},\ x = 0, 1, 2, \ \ldots$
  - $E(X) = \lambda = Var(X)$
    - $\bar{x} \ll s^2$ 이면 포아송분포?

- 초기하분포 (Hypergeometric Dist.)
    - 크기가 N인 모집단이 크기가 M과 $N-M$인 두 개의 부모집단 (A, B)로 나누어진 경우 ⇨ 유한모집단
    - n개의 표본을 비복원으로 추출할 때, 부모집단(A)에서 추출될 표본 수의 분포 ⇨ 각 표본의 추출과정은 독립적이지 않음
    - 확률질량함수: $f(x) = \binom{M}{x}\binom{N-M}{n-x}/\binom{N}{n}$,
        - · $n \ll N$인 경우 $p = M/N$인 이항분포로 근사할 수 있음
    - $E(X) = \frac{nM}{N}$, $Var(X) = n\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}$

■ 연속분포(Continuous distribution)

- 확률변수의 치역이 실수인 분포
  - 확률밀도함수로 표시
- 균일분포(Uniform distribution)
  - 구간 $[\boldsymbol{A}, \boldsymbol{B}]$ 에서 균일하게 발생: $\boldsymbol{X} \sim \boldsymbol{U}(\boldsymbol{A}, \boldsymbol{B})$
  - 확률밀도함수:

$$f(x) = \frac{1}{B-A},\ F(x) = \frac{x-A}{B-A}, \quad A \le x \le B$$

- 정규분포(normal distribution)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty$$

  - Gaussian distribution
  - 확률계산: $\mu = 0$ 이고 $\sigma^2 = 1$ 인 경우 ⇒ 0을 중심으로 대칭
  - 표준화: $X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
    - $Z \sim N(0,1) \Rightarrow X = \sigma Z + \mu \sim N(\mu, \sigma)$
  - 정규확률변수의 선형결합도 정규분포를 따름
  - 표본평균(합)의 근사분포

- 확률분포
  - 모집단의 형태
  - 표본을 어떻게 선택 했는가에 따라 달라짐
  - 불확실성을 내포한 현상을 수리적으로 모델링

    ⇒ 다양한 분포를 알고 있다는 것은 설명할 수 있는 현상이 많아지고 보다 정확한 모델링이 가능