통계학 기초(확률분포)

이산, 연속 확률분포

- 이산확률분포: 셀 수 있는 경우

균일분포, 베르누이분포, 이항분포, 초기하분포, 포아송 분포 등이 있다.

- 연속확률분포

균일분포, 정규분포, 표준정규분포, t 분포, 카이제곱 분포, F 분포 등이 있다.

(이산) 균일분포

- (이산)균일분포(uniform) :

$$X \sim Unif\,(a,\, b)$$

➢ 모수: a =최소값, b =최대값

➢ 분포식(확률질량함수): $P(x) = \dfrac{1}{b-a+1}$

➢ 값의 범위: $X = a,\; a+1,\; \ldots, b$

➢ 특징: equally likely

예) X = 주사위를 던져서 나오는 점의 개수

$$E(X) = ? \qquad V(X) = ?$$

베르누이분포

- 베르누이분포(Bernoulli) :

  X~ $Bernoulli(\pi)$

- 모수: π =P(success)
- 분포식(확률질량함수): P(x)= $\pi^{x}(1-\pi)^{1-x}$
- 값의 범위: X= 0, 1
- 특징: 1) 결과물이 성공과 실패 두 개 밖에 없고, X=1 을 성공(success)으로 X=0을 실패(fail)로 규정지을 수 있는 경우

  2) E(X)=π , V(X)=π(1-π)

예) X = 동전던지기를 하여 앞면이 나오는 개수

베르누이분포

- 베르누이분포

```
pi=0.5; # 모수: 성공확률
size=1; # 고정값
x=0; # 확률변수의 특정 값. 베르누이분포의 확률변수 X는 0 또는 1
f=dbinom(x,size,pi); # 베르누이분포의 확률질량함수
F=pbinom(x,size,pi); # 베르누이분포의 누적확률분포함수
```

이항분포

- 이항분포(binomial) :

  X~ Binomial(n, π)

- 모수: n= 베르누이 시행횟수, π= P(success)
- 분포식(확률질량함수): P(x)= $\binom{n}{x}\pi^{x}(1-\pi)^{n-x}$ 참고) $\binom{n}{x}$ = $\frac{x!}{n!(n-x)!}$
- 값의 범위: X = 0, 1, …, n
- 특징: 1) 독립의 베르누이 시행을 n번 시행

  2) 베르누이 시행으로 얻은 값들의 합

  X=Y1+Y2+…+Yn 여기서 Yi~Bernoulli(π)

  3) E(X)=nπ, V(X)=nπ(1-π)

예) X = 동전던지기 3회 시행할 때 앞면의 개수

이항분포

- 이항분포 :

```
pi=0.5; # 모수: 성공확률
size=3; # 모수: 베르누이 시행횟수
x=2; # 확률변수의 특정 값. 이항분포의 확률변수 X는 0, 1, ..., n
f=dbinom(x,size,pi); # 이항분포의 확률질량함수
F=pbinom(x,size,pi); # 이항분포의 누적확률분포함수
```

- 연습1) 자동차 정비소에 차를 맡겼을 때 약속된 시간 안에 정비를 마칠 확률이 70%라고 할 때, 10대의 자동차를 맡겨서 이 중 2대 이상이 약속된 시간에 정비를 완료할 확률은?

- 연습2) 주식시장에 상장된 10개의 전자제품관련 주식 가운데 1년 안에 부도날 확률이 3%라고 할 때, 적어도 9개의 주식이 부도 없이 정상상태를 유지할 확률은?

(연속)균일분포

- (연속)균일분포(Uniform) :

  X ~Unif(a, b)

- 모수: a = 최솟값, b = 최댓값
- 분포식(확률밀도함수): f(x) = $\frac{1}{b-a}$
- 값의 범위: a ≤ X ≤ b
- 특징: Equally Likely

(연속)균일분포

- (연속)균일분포 :

```
a=0; # 모수: 최솟값
b=1; # 모수: 최댓값
x=0.2; # 확률변수의 특정 값.
f=dunif(x, a, b); # 균일분포의 확률밀도함수
F=punif(x, a, b); # 균일분포의 누적확률분포함수
```

정규분포

- 정규분포(Normal) :

  $X \sim N(\mu, \sigma^2)$

- ➢ 모수: $\mu = E(X)$, $\sigma^2 = V(X)$
- ➢ 분포식(확률밀도함수): $f(x) = \ldots$
- ➢ 값의 범위: $-\infty < X < \infty$
- ➢ 특징: 1)경험률(empirical rule)

  2) 종모양의 좌우 대칭 확률밀도함수

정규분포

- 정규분포 :

```
mu=0.5; # 모수: 평균
sigma=2; # 모수: 표준편차
x=1.2; # 확률변수의 특정 값.
f=dnorm(x, mu, sigma); # 정규분포의 확률밀도함수
F=pnorm(x, mu, sigma); # 정규분포의 누적확률분포함수
```

정규분포

- 연습) A 은행 대출 고객의 신용점수가 정규분포를 따르고 평균이 70점이고
  분산이 9라고 하자.

(1) 무작위로 추출한 고객의 신용점수가 80에서 85일 확률은?

(2) 신용점수가 상위 15%에 들려면 몇 점을 얻어야 하는가? (퍼센타일 개념)

(3) 어느 고객의 신용점수가 90점이라면, 이 고객보다 신용점수가 낮은 고객은 전체의 몇 퍼센트인가?

표준정규분포

- 표준정규분포() :

  $Z \sim N(0, 1)$

- 모수: $0 = E(Z)$, $1 = V(Z)$
- 분포식(확률밀도함수): $f(z) = \ldots$
- 값의 범위: $-\infty < Z < \infty$
- 특징: 1) 경험룰(empirical rule)

  2) 종모양의 좌우 대칭 확률밀도함수

표준정규분포

- 표준정규분포 :

```
mu=0; # 모수: 평균 (고정값)
sigma=1; # 모수: 표준편차 (고정값)
x=1.2; # 확률변수의 특정 값.
f=dnorm(x, mu, sigma); # 표준정규분포의 확률밀도함수
F=pnorm(x, mu, sigma); # 표준정규분포의 누적확률분포함수
```

모집단 백분위수

- 모집단 백분위수:
  - 데이터 수치화에서 학습하였던 백분위수는 표본의 백분위수였음 (즉, 통계량임)
  - 모집단의 백분위수는 모집단의 분포를 알면 R 프로그램을 이용해서 쉽게 구할 수 있음 (모집단의 백분위수는 모수(parameter)임)

- 연속확률변수의 경우 백분위수

  0≤p≤1을 만족하는 p에 대해서 100p 번째 퍼센타일(percentile)은 확률변수 X에 대해서 x_p 로 표현하고 아래의 해에 해당된다.

$$p = F(x_p) \quad \text{또는} \quad x_p = F^{-1}(p)$$

- 이산확률변수의 백분위수는 연속확률변수와 단순히 역함수를 계산하는 방식으로 정의할 수 없다. 다만, R 프로그램을 이용하여 쉽게 구할 수 있다.

모집단 백분위수

- 모집단 백분위수:

```
p=seq(0,1,1,0,1)
qbinom(p, size, pi) # 이항분포
qhyper(p, s, N-s, n) # 초기하분포
qpois(p, lambda) # 포아송분포
qunif(p, min, max) # 균일분포
qnorm(p, mu, sigma) # 정규분포
qt(p, df) # t분포
qf(p, df1, df2) # F분포
qchisq(p, df) # 카이제곱분포
```

- 연습1) A 은행 대출 고객의 신용점수가 정규분포를 따르고 평균이 70점이고 분산이 9라고 하자. 신용점수의 70번째 퍼센타일을 구하시오.

- 연습2) B 은행 대출 고객의 신용점수가 정규분포를 따르고 평균이 78점이고 분산이 7라고 하자. 신용점수의 70번째 퍼센타일을 구하시오.