7-2 이론적 확률분포(이산형) 2

제주권역 대학 e-러닝 지원센터

제주권역 대학 e-러닝 지원센터

1. 베르누이 시행(Bernoulli trial)

- 실험에서 결과가 둘 중의 하나로 나타나는 실험을 말한다.
- 그 중 하나를 성공(success : s)이라 하고 다른 하나를 실패(failure : f)라고 정의한다.
- 따라서 베르누이 시행은 실험의 결과가 s 또는 f인 확률실험이라고 할 수 있으며, 표본공간은 Ω={s, f}이 된다.
- 이 실험에서 결과가 s일 확률이 p라면 확률의 기본원리에 의하여 실험결과가 f일 확률은 1-p이다.

예.베르누이 시행의 예

(a) 동전 하나를 던지는 실험	결과	앞면(Head)/뒷면(Tail)
(b) 대학에 지원한 한 학생의 시험결과	결과	합격/불합격
(c) 한 공장의 생산제품의 불량여부	결과	합격품/불량품
(d) 활을 쏘아서 과녁 맞추기	결과	성공/실패

2. 베르누이 확률변수

- 베르누이 시행에서 결과가 s 이면 '1'이고, 결과가 f 이면 '0'이라고 정의된 확률변수를 말한다.

3. 베르누이 확률분포

- 확률변수 X의 분포가 다음과 같을 때, X를 베르누이 확률변수라 하고, **X~Bernoulli(p)**로 정의한다.

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

- 베르누이 분포의 모양은 확률 p 의 값에 의하여 결정되므로 **이 분포의 모수는 p**이다.

x	p(x)
0	1-p
1	p

4. 베르누이 확률분포의 평균과 분산

① 평균

$$E(X) = \sum x_i p_i = 0 \times (1-p) + 1 \times p = p$$

② 분산

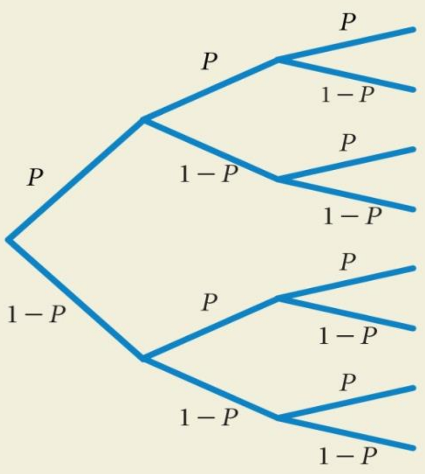
$$E(X^2) = \sum x_i^2 p_i = 0 \times (1-p) + 1 \times p = p$$

$$\sigma_X^2 = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$$

8-2-2.이항분포

제주권역 대학 e-러닝 지원센터

베르누이 시행을 세 번 행할 때의 성공횟수와 발생확률의 예



첫째 시행	둘째 시행	셋째 시행	가능한 결과	발생확률	계약 체결의 수
		S	SSS	ppp	3
		F	SSF	$pp(1-p)$	2
		S	SFS	$p(1-p)p$	2
		F	SFF	$p(1-p)(1-p)$	1
		S	FSS	$(1-p)pp$	2
		F	FSF	$(1-p)p(1-p)$	1
		S	FFS	$(1-p)(1-p)p$	1
		F	FFF	$(1-p)(1-p)(1-p)$	0

방문자가 3명일 때의 이항분포

계약 체결의 수(x)	$P(x)$
0	$(1-p)^3$
1	$3p(1-p)^2$
2	$3p^2(1-p)$
3	p^3

1. 이항분포

- 이항분포 (binomial distribution) 는 베르누이 시행을 독립적으로 n번 반복했을 때 나타나는 결과에 있어서 성공(s)의 횟수에 대한 분포를 구하는 것이다
- 성공의 확률이 p 이고 실패의 확률이 q (q = 1 – p) 인 베르누이 시행을 독립적으로 n 번 반복하였을 때 나타나는 성공의 횟수를 확률변수 X 라고 할 때, X 를 이항확률변수 (binomial random variable) 라 하고, **X~B (n, p)**로 정의하며 확률분포는 다음과 같다.

$$P_r(X = k) = \binom{n}{k} p^k q^{n-k}, k = 0,1,2,\dots,k$$

$$\frac{n!}{k!\,(n-k)!}$$

- 이항분포에서 **모수는 각 시행에서 성공이 나타날 확률 p와 시행횟수 n**이다.

2. n개 독립인 확률변수의 합의 평균과 분산

· X_1 , X_2 , … , X_n 이 서로 독립이며 각각의 평균이 μ 이고, 분산이 σ^2 이라면 이 확률변수들의 합 $X = \sum_{i=1}^{n} X_i$ 의 평균과 분산은 다음과 같다.

$$E(X) = n\mu$$
$$Var(X) = n\sigma^2$$

3. 이항분포의 평균과 분산

· 위의 설명 2번을 이용하여 이항분포의 평균과 분산을 다음과 같이 유도할 수 있다.

$X_1, X_2, \dots, X_n$ 이 각각 서로 독립인 베르누이 확률변수이고 각각의 확률은

$P_r(x_i = 1) = p,\ P_r(x_i = 0) = q, i = 1,2,\dots,n$ 이라고 할 때, 이항확률변수 X는

$X = \sum_{i=1}^{n} X_i$ 라고 정의할 수 있다.

· 따라서 이항확률변수 X의 평균과 분산은 다음 과 같다.

① 평균

$$E(X) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = p + p + \cdots + p = np$$

② 분산

$$Var(X) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = pq + pq + \cdots + pq = npq$$

4. 이항분포의 확률계산

- 이항확률분포에서 확률의 계산은 부록III의 [표1]에 의하여 구할 수 있다. 표본의 수 n=5, 10, 15, 20, 25인 경우에 있어서 성공의 확률이 p인 이항확률 변수 X의 값 a까지의 누적확률을 제시하여 준다.

$$P_r(X \le a) = \sum_{x=0}^{a} P(x)$$

[표 1] 누적이항확률분포표

$$P(X \leq a) = \sum_{x=0}^{a} P(x)$$

(a) $n=5$

P

a	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	a
0	.951	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000	.000	0
1	.999	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000	.000	1
2	1.000	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001	.000	2
3	1.000	1.000	1.000	.993	.969	.913	.812	.663	.472	.263	.081	.023	.001	3
4	1.000	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226	.049	4

```r
binom11<-rep(NA,5)
binom12<-rep(NA,5)
binom13<-rep(NA,5)

binom11[1]<-pbinom(0, 5, 0.1)
binom12[1]<-pbinom(0, 5, 0.2)
binom13[1]<-pbinom(0, 5, 0.3)

for(i in 2:5) {
  binom11[i]<-pbinom(i-1, 5, 0.1)
}

for(i in 2:5) {
  binom12[i]<-pbinom(i-1, 5, 0.2)
}

for(i in 2:5) {
  binom13[i]<-pbinom(i-1, 5, 0.3)
}

(binom<-cbind(binom11,binom12, binom13))
```

```
     binom11 binom12 binom13
[1,] 0.59049 0.32768 0.16807
[2,] 0.91854 0.73728 0.52822
[3,] 0.99144 0.94208 0.83692
[4,] 0.99954 0.99328 0.96922
[5,] 0.99999 0.99968 0.99757
```

예

재벌 구조조정에 대한 한 신문사의 여론조사 결과, 우리 국민들 중 80%는 강도 있는 재벌 구조조정에 찬성하고 20%는 반대하는 것으로 나타났다고 하자. 임의로 5명이 선택되었을 때 3명이 반대할 확률은 얼마인가? 또한 임의로 선택된 5명 중 3명 이상이 반대할 확률은 얼마인가?

답

$$P_X(3) = {}_5C_3 0.2^3 \times 0.8^2 = 0.0512$$

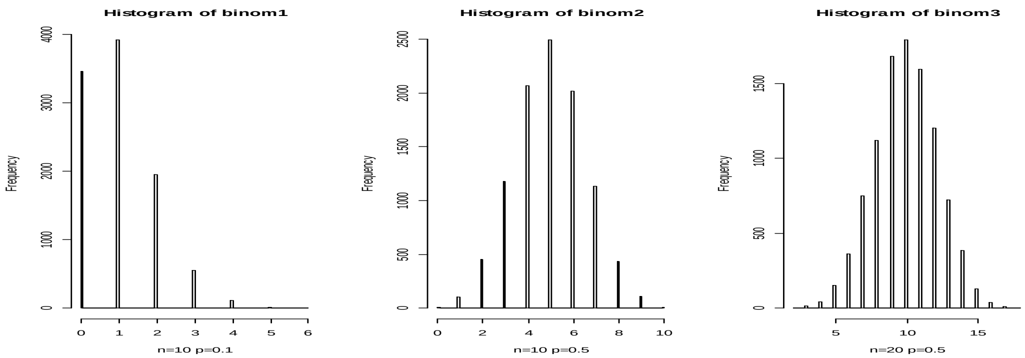
따라서 임의로 선택된 5명 중 3명이 반대할 확률은 0.0512이다. 그리고 3명 이상이 반대할 확률은 다음과 같다.

$$\begin{aligned} P(X \geq 3) &= P_X(3) + P_X(4) + P_X(5) \\ &= 0.0512 + 0.0064 + 0.0003 = 0.0579 \end{aligned}$$

5. 이항분포의 그래프

- 이항분포는 성공확률 p가 0.5이면 평균 $\mu = np = \frac{n}{2}$을 중심으로 좌우대칭인 분포를 가지며, p 와 n 의 크기에 따라 모양이 결정된다. 세 가지 경우에 대한 이항분포의 그림은 다음과 같다.

```r
set.seed(12345)

r<-10000

binom1<-rbinom(r, 10, 0.1)
binom2<-rbinom(r, 10, 0.5)
binom3<-rbinom(r, 20, 0.5)

par(mfrow=c(1,3))

hist(binom1, breaks=100, xlab="n=10 p=0.1")
hist(binom2, breaks=100, xlab="n=10 p=0.5")
hist(binom3, breaks=100, xlab="n=20 p=0.5")

```



```r
par(mfrow=c(2,2))

n<-25 # 시행횟수

p_list<-c(0.1, 0.2, 0.5, 0.7) # 발생확률

for (i in 1:length(p_list)) {
    p_x<-dbinom(x=1:n, n, p_list[i])
    plot(x=1:n, p_x, xlab="x", ylab="P(X=x)",
         ylim=c(0, 0.3), xlim=c(1,n), main=paste("p=", p_list[i]))
    x_seq<-seq(1,n,1)
    lines(x_seq, p_x, type="h", col="blue")
}
```