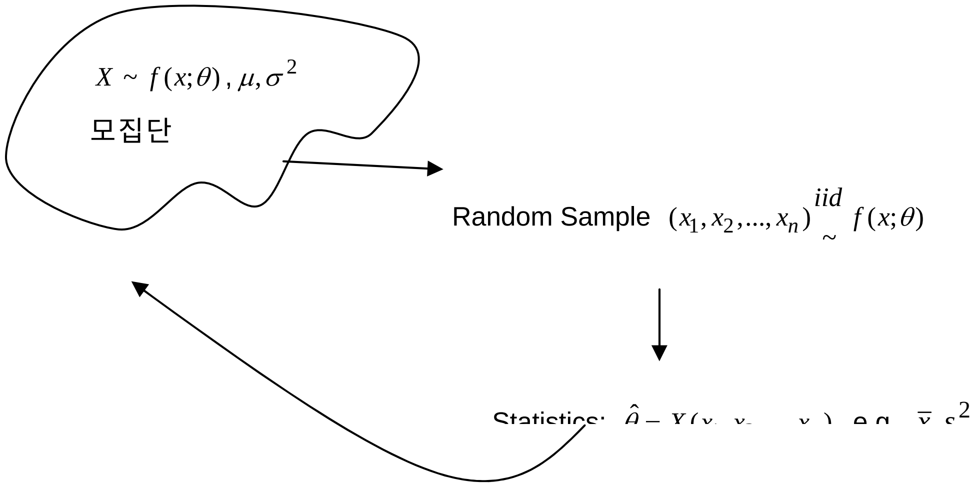

<u>기대치(Expected value)</u> $E(X) = \sum_x xp(x)$ (평균), $V(X) = E[(X - E(X))^2]$ (분산)

•평균(mean): 관측치 크기의 중심

•분산(variance): 관측치들이 중심으로부터 얼마나 흩어져 있는지에 대한 측정

3.4 이항분포 (Binomial)

베르누이 시행(Bernoulli experiment)

•확률실험의 결과가 두 가지이다. (성공 success=1, 실패 fail=0)

•각 실험은 서로 독립(independent)이다.

•각 실험의 성공 확률은 p 으로 동일하다.

확률변수 X 을 베르누이 실행의 결과라 정의하자. 즉 성공이면 $X = 1$, 실패면 $X = 0$ 이다. 이 경우 확률변수 X 의 확률밀도함수(probability density function)은 다음과 같다. 이를 베르누이 분포(Bernoulli distribution)라 한다.

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

Notation $X \sim Bernoulli(p)$

베르누이 분포에서 p 을 모수(parameter)라 한다. 여기서 모수의 의미는 그 값을 알면 확률밀도함수를 그릴 수 있다는 것이다. 즉 확률을 계산할 수 있다는 의미이다.

EXAMPLE 3-6 베르누이 분포 평균과 분산

확률변수 $X \sim Bernoulli(p)$ 의 평균과 분산이 p, pq 임을 보이시오.

$$E(X) = \sum xp(x) = \sum_{i=0}^{1} xp^x q^{1-x} = p\,,\;\; E(X^2) = \sum x^2 p(x) = \sum_{i=0}^{1} x^2 p^x q^{1-x} = p$$

$$V(X) = E(X) - E(X^2) = pq$$

EXAMPLE 3-7 베르누이 분포와 이항 분포

학생들 중 감기 환자가 10%이다. 학생 3명을 임의로 선택하였을 때 확률변수 X 을 감기 걸린 학생 수라 정의할 때 X 의 확률밀도함수를 구하시오.

X	$P(X = x)$
0	
1	
2	
3	

정의(DEFINITION)

확률변수 X 을 n번의 베르누이 시행에서 성공 회수 정의하자. 확률변수 X 의 확률밀도함수는 다음과 같고 이를 이항분포(binomial distribution)라 한다.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0,1,2,...,n \quad \text{---} \; ②$$

Notation $X \sim Binomial(n, p)$

EXAMPLE 3-8 이항분포 PDF?

식 ②의 이항분포 PDF가 확률밀도함수임을 보이시오. 그리고 확률밀도함수의 의미는?

EXAMPLE 3-9 이항분포 이용

EXAMPLE 3-7에서 학생을 15명 임의 선택하였을 때 그 중 적어도 한 명이 환자일 확률을 구하시오.

0.5409

IN SAS

```
%macro loop1(n);
data one;
    do x=0 to &n;
    %do p=1 %to 9 %by 1;
        pr=&p/10;
        p&p=pdf('binomial',x,pr,&n);
    %end;
    output;
    end;
run;
proc print data=one;
    var x p1-p9;
run;
%mend loop1;
%loop1(15);
quit;
```

함수 PDF대신 CDF 사용하면 누적확률밀도함수를 얻는다.

$$F(x)=P(X\le x)=\sum_{t=0}^{x}\binom{n}{t}p^{t}(1-p)^{n-t}$$

x	p1	p2	p3	p4	p5	p6	p7	p8	p9
0	0.20589	0.03518	0.00475	0.00047	0.00003	0.00000	0.00000	0.00000	0.00000
1	0.34315	0.13194	0.03052	0.00470	0.00046	0.00002	0.00000	0.00000	0.00000
2	0.26690	0.23090	0.09156	0.02194	0.00320	0.00025	0.00001	0.00000	0.00000
3	0.12851	0.25014	0.17004	0.06339	0.01389	0.00165	0.00008	0.00000	0.00000
4	0.04284	0.18760	0.21862	0.12678	0.04166	0.00742	0.00058	0.00001	0.00000
5	0.01047	0.10318	0.20613	0.18594	0.09164	0.02449	0.00298	0.00010	0.00000
6	0.00194	0.04299	0.14724	0.20660	0.15274	0.06121	0.01159	0.00067	0.00000
7	0.00028	0.01382	0.08113	0.17708	0.19638	0.11806	0.03477	0.00345	0.00003
8	0.00003	0.00345	0.03477	0.11806	0.19638	0.17708	0.08113	0.01382	0.00028
9	0.00000	0.00067	0.01159	0.06121	0.15274	0.20660	0.14724	0.04299	0.00194
10	0.00000	0.00010	0.00298	0.02449	0.09164	0.18594	0.20613	0.10318	0.01047
11	0.00000	0.00001	0.00058	0.00742	0.04166	0.12678	0.21862	0.18760	0.04284
12	0.00000	0.00000	0.00008	0.00165	0.01389	0.06339	0.17004	0.25014	0.12851
13	0.00000	0.00000	0.00001	0.00025	0.00320	0.02194	0.09156	0.23090	0.26690
14	0.00000	0.00000	0.00000	0.00002	0.00046	0.00470	0.03052	0.13194	0.34315
15	0.00000	0.00000	0.00000	0.00000	0.00003	0.00047	0.00475	0.03518	0.20589

정리(THEOREM)

확률변수 $X \sim Binomial(n, p)$ 일 때 $\mu = E(X) = np,\ \sigma^2 = V(X) = npq$ 을 밝히시오.

PROOF

$$E(X) = \sum_x xp(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$$

$$E(X) = \sum_{x=1}^{n} \frac{n!}{(n-x)!(x-1)!} p^x q^{n-x} (\because x = 0 \rightarrow xp(x) = 0)$$
$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} q^{n-x} (z = x-1)$$
$$= np$$

$E(X(X-1)) = E(X^2) - E(X)$ 을 이용하여 $V(X)$ 을 계산하자. ($\because$ 이항분포의 경우 $E(X^2)$ 을 계산하는 것은 쉽지 않다)

$E(X(X-1)) = n(n-1)p^2$

그러므로 $E(X^2) = n(n-1)p^2 + np \Rightarrow V(X) = E(X^2) - E(X)^2 = npq$. Q.E.D.

HOMEWORK #6-1 DUE 3월 29일(화)

확률변수 $X \sim Binomial(n, p)$ 이다. 확률변수 $(n-X)$ ("실패 개수")의 PDF를 구하시오. 그리고 어떤 분포를 따르는지 밝히시오.

HOMEWORK #6-2 DUE 3월 29일(화)

헌혈 지원자의 80%는 헌혈 가능자라 하자.

(1)5명을 임의 추출했을 때 적어도 한 명이 헌혈 가능한 사람일 확률을 계산하시오.

(2)많아야 4명이 한혈 가능한 사람일 확률을 구하시오.

```
data one;
    do n=5 to 20;
    p=1-cdf('binomial',4,0.8,n);
    output;
    end;
run;
```

3.5 기하분포 (Geometric)

정의(DEFINITION)

베르누이 시행에서 <u>첫 성공</u>이 있기까지 실험(시행) 회수를 확률변수 X 라 정의하면 확률 밀도함수는 다음과 같고 이를 기하분포라 한다.

$$f(x) = p(1-p)^{x-1}, x = 1, 2, \ldots$$

Notation $X \sim Geometric(p)$

IN SAS, PDF('GEOMETRIC',m,p); m=실패 회수, $m+1=x$

EXAMPLE 3-10 기하분포 이용

기계 A가 1시간 동안 오작동 할 확률은 0.02이다. 이 기계가 두 시간 동안 오작동 없이 작동될 확률을 구하시오.

0.9604

정리(THEOREM)

$X \sim Geometric(p)$ 이면 $\mu = E(X) = 1/p,\ \sigma^2 = V(X) = (1-p)/p^2 = q/p^2$ where $q = (1-p)$ 이다.

PROOF

$$E(X) = \sum_x xp(x) = \sum_{x=1}^{\infty} xpq^{x-1} = p\sum_{x=1}^{\infty} xq^{x-1}$$

$\frac{d}{dq}(q^x) = xq^{x-1}$ 을 이용하면 $\frac{d}{dq}(\sum_{x=1}^{\infty} q^x) = \sum_{x=1}^{\infty} xq^{x-1}$ 임을 알 수 있다.

그러므로 $$E(X) = p\sum_{x=1}^{\infty} xq^{x-1} = p\frac{d}{dq}(\sum_{x=1}^{\infty} q^x) = p\frac{d}{dq}(\frac{q}{1-q}) = \frac{1}{p}$$. 분산은 숙제로 남겨둔다. Q.E.D.

EXAMPLE 3-11 기하분포 이용

은행에 고객이 임의의 1초에 창구를 찾을 확률은 0.1이고 고객이 창구를 찾는 사건은 서로 독립이라고 가정하자.

①3초에 첫 손님이 올 확률을 계산하시오.

②적어도 3초 안에는 첫 손님이 오지 않을 확률을 계산하시오.

③첫 손님이 오려면 몇 초나 기다려야 할까?

0.081 / 0.81 / 10

HOMEWORK #6-3 DUE 3월 29일(화)

확률변수 $X \sim Geometric(p)$ 일 때 $P(X > a+b \mid X > a) = P(X > b)$ 임을 보이시오.

이를 기하분포의 **무기억성(memoryless property)**이라 한다.

HOMEWORK #6-4 (optional) DUE 3월 29일(화)

$X \sim Geometric(p)$ 일 때 $\sigma^2 = V(X) = (1-p)/p^2 = q/p^2$ where $q = (1-p)$ 임을 보이시오.

TIP $\frac{d^2}{dq^2}(\sum_{x=2}^{\infty} q^x)$을 이용하여 $E(X(X-1))$을 구한 후 $E(X^2)$을 얻으시오.

3.6 음이항 분포(Negative Binomial)

정의(DEFINITION)

베르누이 시행에서 r 번 성공할 때까지 시행하는 실험 회수를 확률변수 X 라 하면 확률밀도함수는 다음과 같고 이를 음이항분포라 한다.

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

Notation $X \sim NB(r, p)$

IN SAS, PDF('NEGBINOMIAL',x-r,p,r);

“Negative” Binomial 은 의미는 “이항분포의 역”을 의미한다. 이항분포는 n번 베르누이 시행에서 성공의 회수라면 음이항분포는 r번 성공할 때까지 시행의 회수에 대한 분포이다.

정리(THEOREM)

확률변수 $X \sim NB(r,p)$ 이면 평균과 분산은 다음과 같다.

$\mu = E(X) = r/p,\ \sigma^2 = V(X) = r(1-p)/p^2$

PROOF

확률변수 $X_1, X_1, ..., X_r$ 을 기하분포 $Geomeric(p)$ 라 하자.($X_i \underset{iid}{\sim} Geo(p)$) $\sum_{i=1}^{r} X_i$ 는 $NB(r,p)$ 을 따른다(이 부분은 5장에서 다루기로 한다).

평균 $E(Y) = E(\sum X_i) = \sum E(X_i) = \frac{r}{p}$ 이고 분산은 $V(Y) = V(\sum X_i) = \sum V(X_i) = r\frac{q}{p^2}$ 이다.

EXAMPLE 3-12 음이항 분포 이용

지질학 연구 결과 석유 탐사 중 석유 발견 확률은 0.2이다.
①3번째 탐사에서 처음 석유가 발견될 확률을 계산하시오.
②7번째 탐사에서 3번째 석유가 발견될 확률을 계산하시오.
③석유가 3개 발견될 때까지 시추를 계속하기로 하였다. 몇 번 정도 시추해야 하나?

0.128 / 0.049 /15

HOMEWORK #7-1 DUE 4월 7일(목)

제품의 10%는 불량이다. 제품을 하나씩 검사하는 과정에서
(1) 첫 불량품이 2번째 검사에서 발견될 확률을 계산하시오.
(2) 5번째 시행에서 세 번째 불량 제품이 발견될 확률을 계산하시오.
(3) 5번째 시행 혹은 그 전 시행에서 세 번째 불량 제품이 발견될 확률을 계산하시오.