Probability

Sung Won Kang

*2016*년 *11*월 *12*일

Section 1: 확률

1. 확률의 정의

(뭔가 매우 헷갈리는 말이 많이 쓰여 있으므로 무시)

- 시행 (trial)/실험 (experiment): 주사위 던지기
  - 결과는 사전에는 모르고 사후에는 확인
  - '⋯ 중 하나가 결과가 된다.'는 정보는 사전에 알려져 있음
- 결과 (outcome) : 주사위를 던져서 나온 값
- 표본공간 (Sample Space): 주사위 던져서 나올 수 있는 모든 값

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- 사건 (Event): 주사위 눈으로 만들 수 있는 부분집합

$$A \in \{\phi, \{1\}, \{1, 2\}, \{3, 4, 5\}, \{1, 2, 3, 4, 5, 6\}\}$$

  - 근원사건 (A_i): 사건 중 원소가 하나인 사건

$$A_i \in \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$$

- 확률 (probability): 모든 사건에 할당하는 척도 (얼마나 자주 일어날까?)
  - 0과 1 사이

$$0 \leq P(A) \leq 1$$

  - 표본 공간에 할당하는 확률은 1

$$P(\Omega) = 1$$

  - 겹치지 않는 두 사건이 함께 일어날 사건의 확률 = 각 사건의 확률의 합

$$P(A \cup B) = P(A) + P(B) \qquad (\forall \{A, B\} \quad A \cap B = \phi)$$

- 확률공간 : 확률이 정의된 실험
  - 표본공간, '사건의 집합', 확률

```r
library(prob)

## 표본공간

# 동전던지기
tosscoin(1)
```

```
##   toss1
## 1     H
## 2     T
```

```r
# 주사위 굴리기
rolldie(1)
```

```
##   X1
## 1  1
## 2  2
## 3  3
## 4  4
## 5  5
## 6  6
```

```r
# 1,2,3 중 서로 다른 2개를 고르기
urnsamples(1:3,size=2)
```

```
##   X1 X2
## 1  1  2
## 2  1  3
## 3  2  3
```

```r
# 1,2,3 중 2개를 고르기 (똑같은 것을 골라도 됨)
urnsamples(1:3, size=2, replace=T)
```

```
##   X1 X2
## 1  1  1
## 2  1  2
## 3  1  3
## 4  2  2
## 5  2  3
## 6  3  3
```

```r
# 빨간공 3개, 푸른공 2개 있는 주머니에서 2개 고르기
urnsamples(c(rep("R",3),rep("B",2)),size=2)
```

```
##    X1 X2
## 1   R  R
## 2   R  R
## 3   R  B
## 4   R  B
## 5   R  R
## 6   R  B
## 7   R  B
## 8   R  B
## 9   R  B
## 10  B  B
```

```r
urnsamples(c(paste("R",(1:3),sep=""),paste("B",(1:2),sep="")),size=2)
```

```
##    X1 X2
## 1  R1 R2
## 2  R1 R3
## 3  R1 B1
## 4  R1 B2
## 5  R2 R3
## 6  R2 B1
## 7  R2 B2
## 8  R3 B1
## 9  R3 B2
## 10 B1 B2
```

```r
# 통전 2개 던지는 실험의 확률공간 (정상적인 동전)
tosscoin(2,makespace=T)
```

```
##   toss1 toss2 probs
## 1     H     H  0.25
## 2     T     H  0.25
## 3     H     T  0.25
## 4     T     T  0.25
```

```r
probspace(tosscoin(2))
```

```
##   toss1 toss2 probs
## 1     H     H  0.25
## 2     T     H  0.25
## 3     H     T  0.25
## 4     T     T  0.25
```

```r
iidspace(c("H","T"),2)
```

```
##   X1 X2 probs
## 1  H  H  0.25
## 2  T  H  0.25
## 3  H  T  0.25
## 4  T  T  0.25
```

```r
# 동전 2개 던지는 실험의 확률공간 (야바위꾼의 동전)
prob.1=c(1,2)
prob.1=prob.1/sum(prob.1)
iidspace(c("H","T"),2,probs=prob.1)
```

```
##   X1 X2     probs
## 1  H  H 0.1111111
## 2  T  H 0.2222222
## 3  H  T 0.2222222
## 4  T  T 0.4444444
```

2. 확률법칙

1. 조건부 확률

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$A, B \text{ 독립} \rightarrow P(A|B) = P(A), P(B|A) = P(B)$$

2. A 와 B 가 동시에 일어날 확률

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

$$A, B \text{ 독립} \rightarrow P(A \cap B) = P(A) \times P(B)$$

3. A 혹은 B가 일어날 확률

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

4. A 가 일어나지 않을 확률

$$P(A^c) = 1 - P(A)$$

3. 확률변수

- 확률변수(X): '나올수 있는 모든 결과'를 숫자(실수)에 대응시키는 함수
  - 이산확률변수(descrete Random Variable): 값이 '끊어져 있는' 확률변수(동전 2개를 던져서 나오는 앞면의 갯수)
  - 연속확률변수(Continuous Random Variable): 값이 '이어져 있는' 확률변수(100m 달리기 기록)
- 확률분포 : 확률변수 X 가 특정한 값 x 를 갖는 사건의 확률
  - 이산확률분포(descrete Random Variable)

$$P(X = x)$$

  - 연속확률분포(Continuous Random Variable)

$$F(x) = P(X \leq x) \qquad \text{누적분포함수, 분포함수(distribution function)}$$

$$f(x) = \frac{dF(x)}{dx} \qquad \text{확률밀도함수(density function)}$$

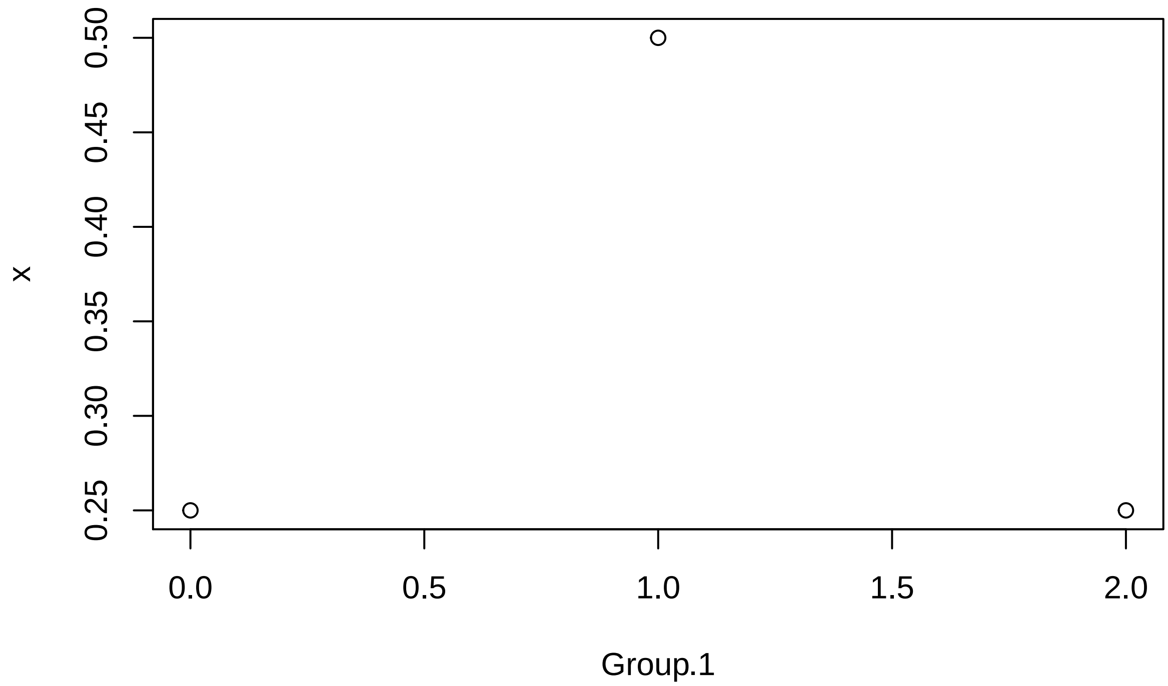
```
# 동전 2개 던저셔 나오는 앞면의 갯수
xii=iidspace(c("H","T"),2)
nhead=function(x){
  as.numeric(x[,1]=="H")+as.numeric(x[,2]=="H")
}
NH=nhead(xii[,1:2])
D.nhead=data.frame(xii[,1:2],NH,xii[,3])
RV.nhead=aggregate(D.nhead[,4],by=list(D.nhead[,3]),FUN=sum)
## 확률변수
D.nhead
```

```
##   X1 X2 NH xii...3.
## 1  H  H  2     0.25
## 2  T  H  1     0.25
## 3  H  T  1     0.25
## 4  T  T  0     0.25
```

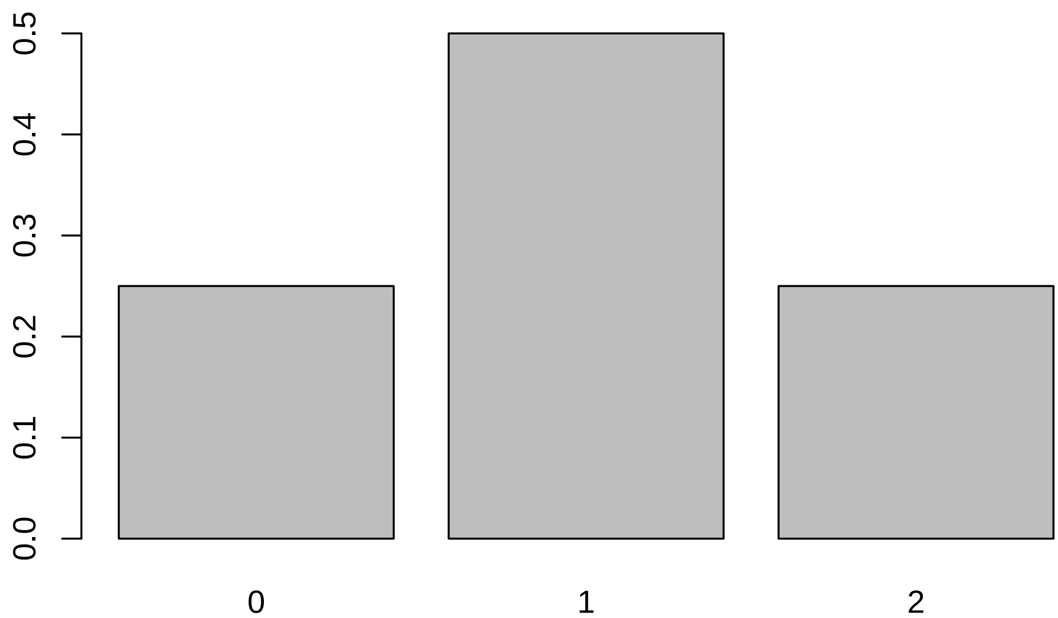
```
## 확률분포
RV.nhead
```

```
##   Group.1    x
## 1       0 0.25
## 2       1 0.50
## 3       2 0.25
```

```
plot(RV.nhead)
```




```
barplot(RV.nhead[,2],names.arg=RV.nhead[,1])
```



4. 평균과 분산 (parameter)

- 이산확률변수
  - 평균

$$E(X) = \sum_x x \cdot P(X = x)$$

  - 분산

$$E(X - E(X))^2 = \sum_x (x - E(X))^2 \cdot P(X = x) = E(X^2) - (E(X))^2$$

- 연속확률변수
  - 평균

$$E(X) = \int_x x \cdot f(x) dx$$

  - 분산

$$E(X - E(X))^2 = \int_x (x - E(X))^2 f(x) dx = E(X^2) - (E(X))^2$$

- parameter? : 모집단의 특성

```
colnames(RV.nhead)=c("x","px")
Ex=sum(RV.nhead$x*RV.nhead$px)
Ex
```

```
## [1] 1
```

```r
Vx1=sum(((RV.nhead$x)^2)*RV.nhead$px)-Ex^2
Vx2=sum(((RV.nhead$x-Ex)^2)*RV.nhead$px)
Vx1
```

```
## [1] 0.5
```

```r
Vx2
```

```
## [1] 0.5
```

Section 2: 분포함수

베르누이 분포 (Bernolli distribution)

- p의 확률로 성공하면 1, 실패하면 0이 되는 확률변수의 확률분포
- 확률

$$P(X) = p^x \cdot (1-p)^{1-x} \qquad x \in \{0, 1\}$$

- 모수 $= p$
- 평균, 분산

$$E(X) = 0 \times p^0 \cdot (1-p)^1 + 1 \times p^1 \cdot (1-p)^0 = p$$
$$V(X) = E(X^2) - (E(X))^2 = 1^2 \times p^1 \cdot (1-p)^0 - p^2 = p(1-p)$$

이항분포 (binomial distribution)

-성공확률이 p인 베르누이 시행을 n 번 반복하여 발생하는 성공 횟수의 확률분포 + n번의 독립적이고 동일한 베르누이 분포를 하는 확률변수의 합

- 확률

$$P(X) = \binom{n}{x} p^x \cdot (1-p)^{n-x} \qquad x \in \{0, 1, 2, , , , n\}$$

- 평균,분산

$$E(X) = \sum_{n}^{x=0} x \binom{n}{x} p^x \cdot (1-p)^{n-x} = np$$
$$V(X) = \sum_{n}^{x=0} x^2 \binom{n}{x} p^x \cdot (1-p)^{n-x} - (np)^2 = np(1-p)$$

$$E(X) = E(\sum_{i=1}^{n} Z_i)) = \sum_{i=1}^{n} E(Z_i) = np \qquad Z_i \sim B(1,p)$$

$$V(X) = E(\sum_{i=1}^{n} Z_i - E(\sum_{i=1}^{n} Z_i))^2 = E(\sum_{i=1}^{n} Z_i - \sum_{i=1}^{n} E(Z_i))^2$$

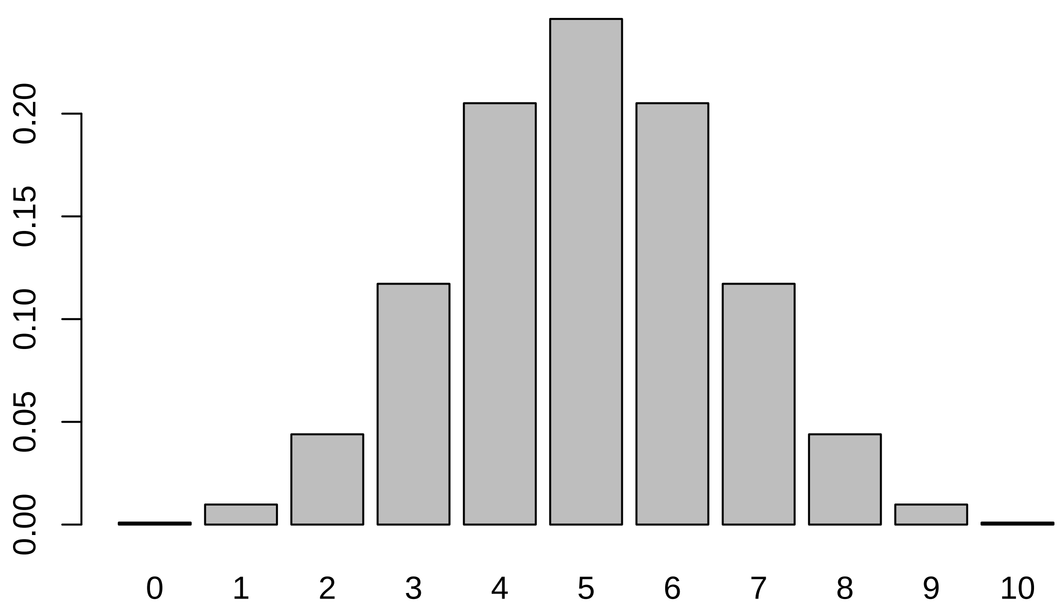
$$= E(\sum_{i=1}^{n} Z_i - n \times E(Z_i))^2 = E(\sum_{i=1}^{n}(Z_i - E(Z_i)))^2$$

$$= E(\sum_{i=1}^{n}(Z_i - E(Z_i))^2 + \sum_{i \neq j}(Z_i - E(Z_i))(Z_j - E(Z_j)))$$

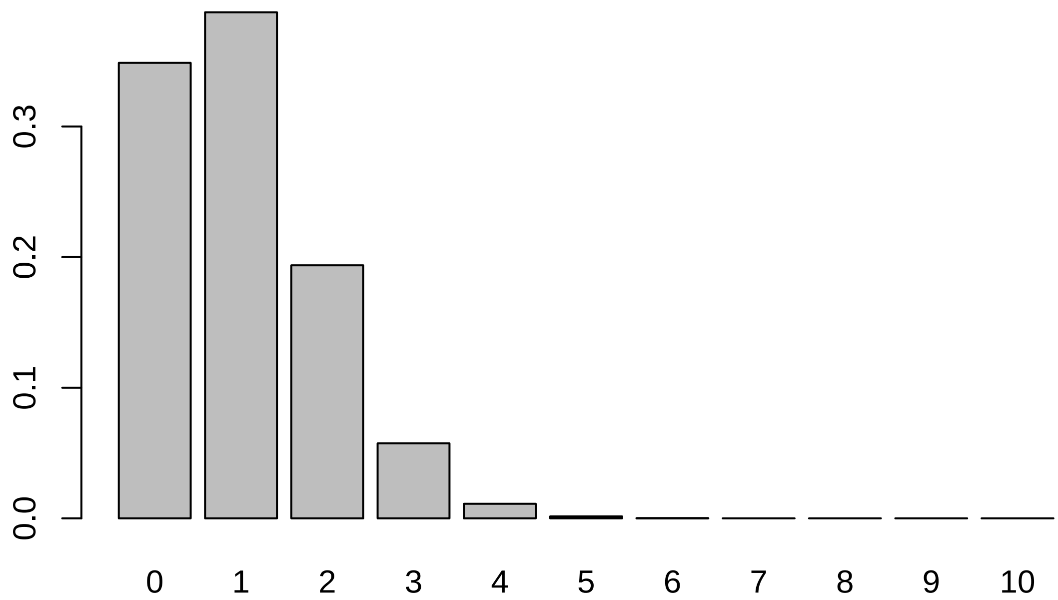
$$= E(\sum_{i=1}^{n}(Z_i - E(Z_i))^2) + \sum_{i \neq j} E(Z_i - E(Z_i))E(Z_j - E(Z_j)))$$

$$= E(\sum_{i=1}^{n}(Z_i - E(Z_i))^2) = \sum_{i=1}^{n} E(Z_i - E(Z_i))^2 = np(1-p)$$

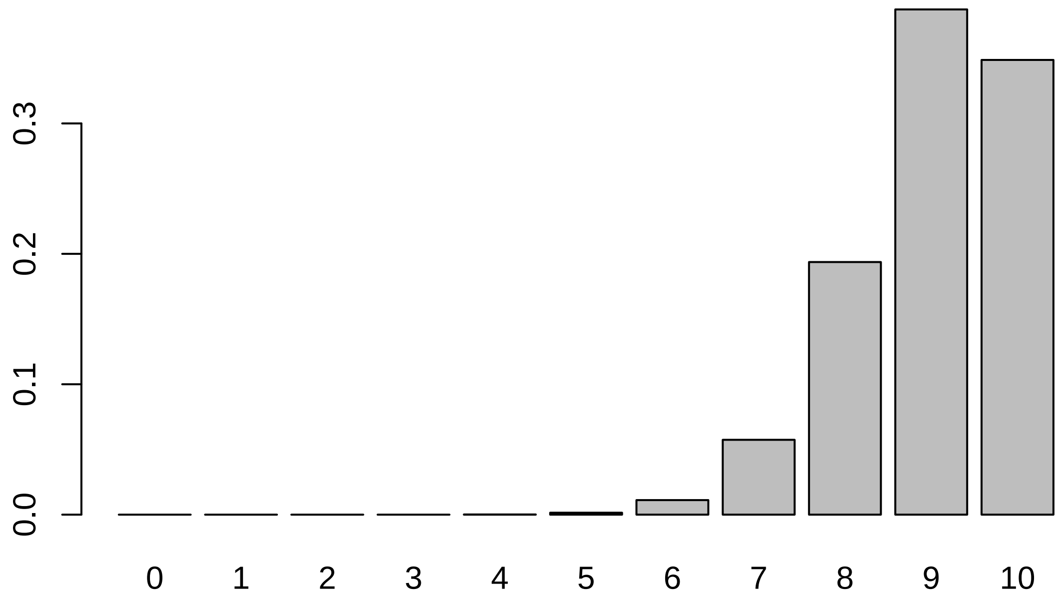
```
n=10
p=0.5
x=(0:n)
px=dbinom(x,size=n,prob=p)
barplot(px,names.arg=x)
```



```r
p1=0.1
px1=dbinom(x,size=n,prob=p1)
barplot(px1,names.arg=x)
```




```r
p2=0.9
px2=dbinom(x,size=n,prob=p2)
barplot(px2,names.arg=x)
```



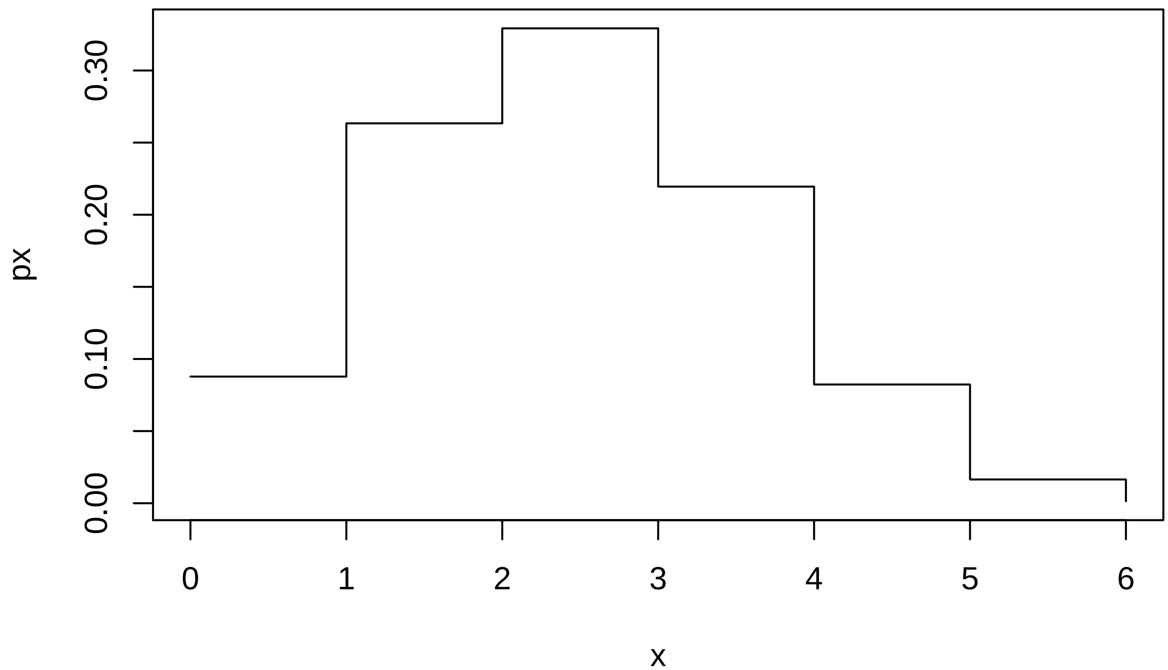

```r
n=6
p=1/3
x=0:n
# Probability
dbinom(2,size=n,prob=p)
```

```
## [1] 0.3292181
```

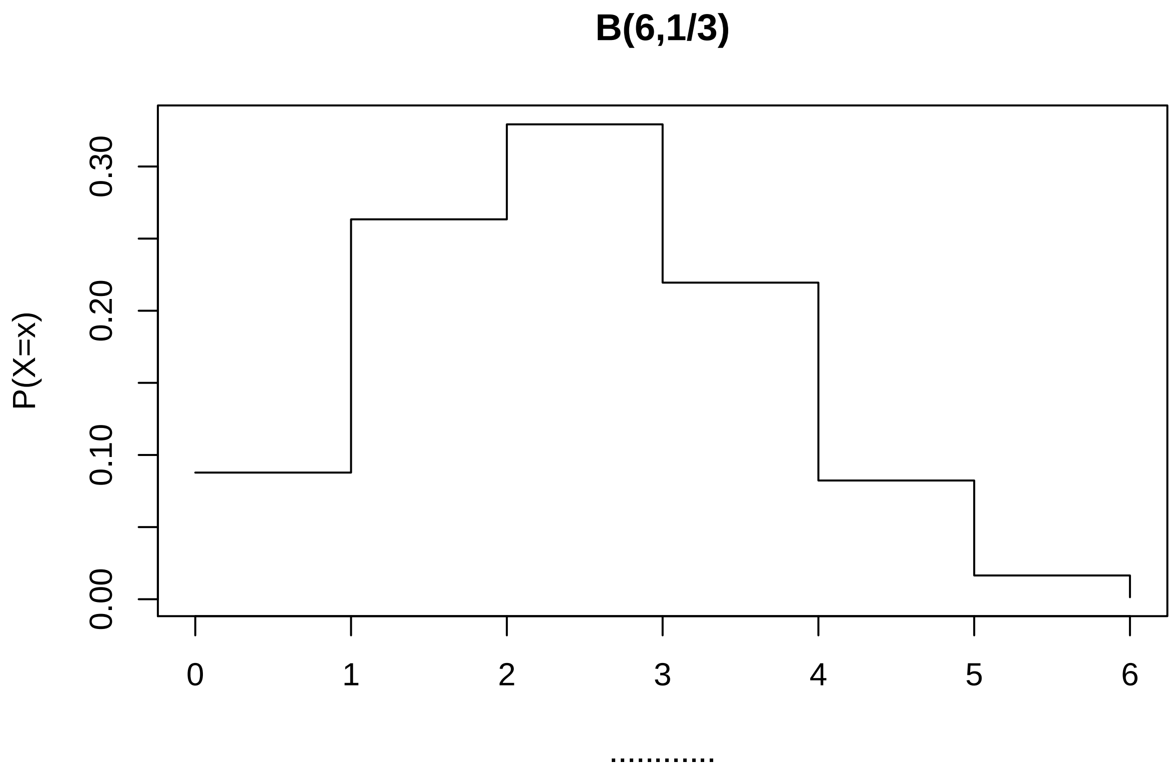
```r
dbinom(4,size=n,prob=p)
```

```
## [1] 0.08230453
```

```r
px=dbinom(x,size=n,prob=p)
plot(x,px,type="s")
```




```r
plot(x,px,type="s",xlab="성공횟수",ylab="P(X=x)",main="B(6,1/3)")
```




```r
# distribution P(X\le x)
pbinom(2,size=n,prob=p) #P(x\le 2)
```

```
## [1] 0.6803841
```

```r
pbinom(4,size=n,prob=p) #P(x\le 4)
```

```
## [1] 0.9821674
```

```r
pbinom(4,size=n,prob=p)-pbinom(2,size=n,prob=p) # P((2,4]))
```

```
## [1] 0.3017833
```

```r
#Quantile from Random variable
qbinom(0.1,size=n,prob=p) #z . Pr(x \le z ) =0.1
```

```
## [1] 1
```

```r
qbinom(0.5,size=n,prob=p) #z . Pr(x \le z ) =0.5
```

```
## [1] 2
```

```r
#random number generation
```

```r
rbinom(10,size=n,prob=p)
```

```
##  [1] 2 2 3 3 3 1 2 1 2 0
```

```r
# mean


ex=sum(x*px)
ex2=sum(x^2*px)
vx=ex2-ex^2
ex
```

```
## [1] 2
```

```r
vx
```

```
## [1] 1.333333
```

정규분포 (normal distribution)

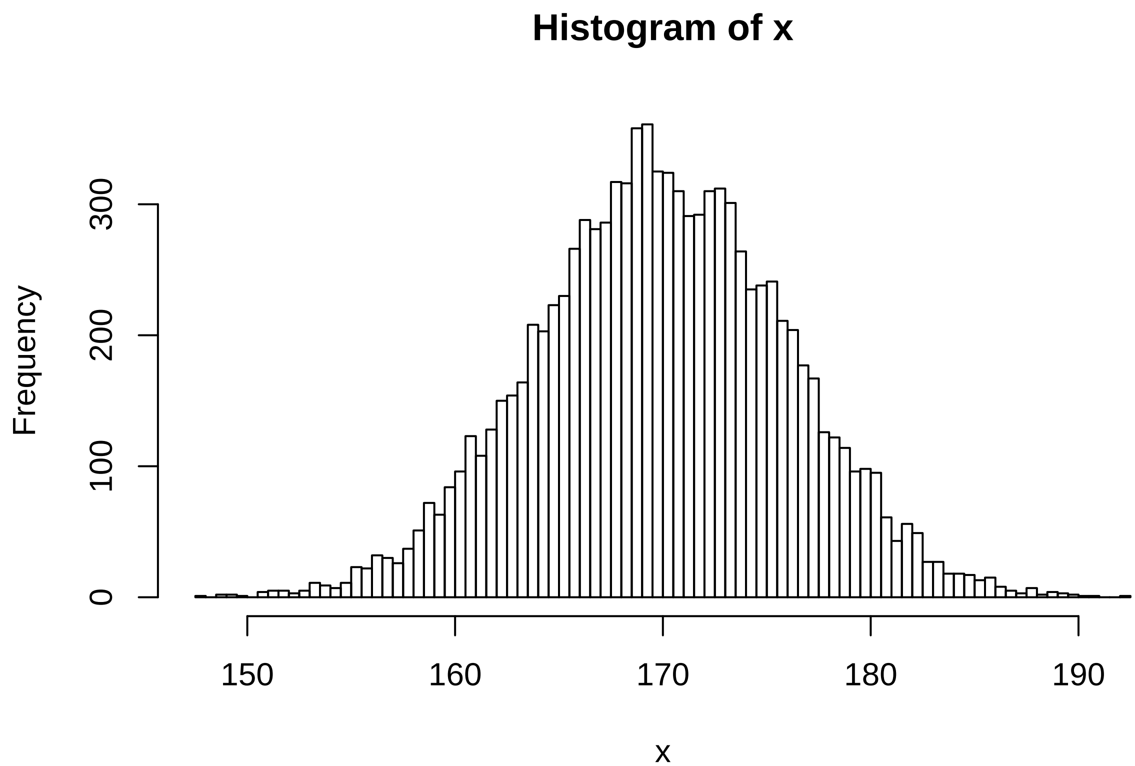
- 확률밀도함수, 분포함수

$$
f(X) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right)
$$
$$
F(X) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(t-\mu)^2}{\sigma^2}\right) dt
$$

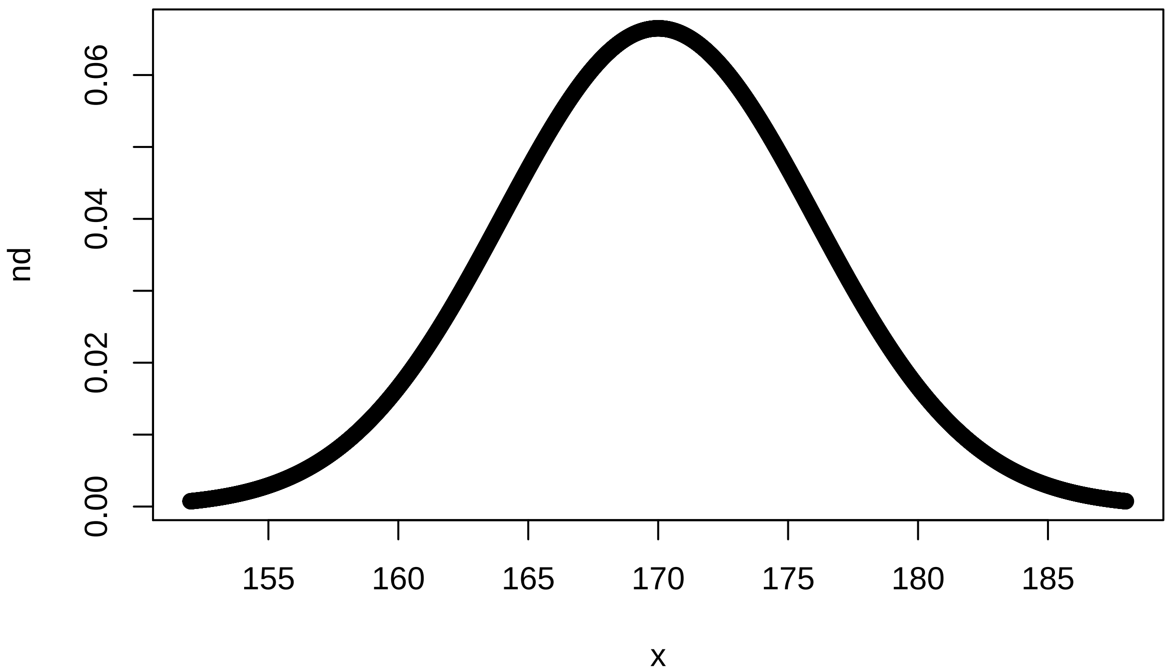
- 모수 $= \mu, \sigma$
- 평균, 분산

$$
E(x) = \int x \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right) dx = \mu
$$
$$
v(x) = \int (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{\sigma^2}\right) dx = \sigma^2
$$

```r
x=rnorm(10000,mean=170,sd=6)
hist(x,breaks=100)
```




```
options(digits=3)
mu=170
sigma=6
ll=170-3*sigma
ul=170+3*sigma
x=seq(ll,ul,by=0.01)
nd=dnorm(x,mean=mu,sd=sigma)
plot(x,nd)
```




```
# distribution funciton
pnorm(mu,mean=mu,sd=sigma)
```

```
## [1] 0.5
```

```
pnorm(159,mean=mu,sd=sigma) # P(x<159)
```

```
## [1] 0.0334
```

```
pnorm(180,mean=mu,sd=sigma)-pnorm(160,mean=mu,sd=sigma) # P(160<x<180)
```

```
## [1] 0.904
```

```
#quantile
qnorm(0.25,mean=mu,sd=sigma)
```

```
## [1] 166
```

```
qnorm(0.5,mean=mu,sd=sigma)
```

```
## [1] 170
```

```
qnorm(0.75,mean=mu,sd=sigma)
```

```
## [1] 174
```

```r
#random sample
options(digits=5)
set.seed(5)
smp=rnorm(400,mean=mu,sd=sigma)
mean(smp)
```

```
## [1] 170.02
```

```r
sd(smp)
```

```
## [1] 6.0054
```

```r
options(digits=4)
mu=0
sigma=1
p0.05=qnorm(0.05,mean=mu,sd=sigma)
p0.025=qnorm(0.025,mean=mu,sd=sigma)

pnorm(-1*p0.05)-pnorm(p0.05)
```

```
## [1] 0.9
```

```r
pnorm(-1.645)-pnorm(1.645)
```

```
## [1] -0.9
```

```r
pnorm(-1.96)-pnorm(1.96)
```

```
## [1] -0.95
```

3. 다음시간 준비: Loop

```r
v=c(1,4,5)
for (i in v){
  print(i)
}
```

```
## [1] 1
## [1] 4
## [1] 5
```

```r
r.n=rnorm(10)
SUM=0
for (i in 1:10){
```

```r
  SUM=SUM+r.n[i]
}
print(SUM)
```

```
## [1] -3.549
```

```r
sum(r.n)
```

```
## [1] -3.549
```

```r
dan=2
for (i in 2:9){
  times=dan*i
  print(paste(dan, "X" ,i, "=",times))
}
```

```
## [1] "2 X 2 = 4"
## [1] "2 X 3 = 6"
## [1] "2 X 4 = 8"
## [1] "2 X 5 = 10"
## [1] "2 X 6 = 12"
## [1] "2 X 7 = 14"
## [1] "2 X 8 = 16"
## [1] "2 X 9 = 18"
```

```r
m=matrix(1:12,ncol=3)
for (i in 1:nrow(m)){
  for (j in 1:ncol(m)){
    cat(i, "행", j, "열=", m[i,j],"\n")
}
}
```

```
## 1 행 1 열= 1
## 1 행 2 열= 5
## 1 행 3 열= 9
## 2 행 1 열= 2
## 2 행 2 열= 6
## 2 행 3 열= 10
## 3 행 1 열= 3
## 3 행 2 열= 7
## 3 행 3 열= 11
## 4 행 1 열= 4
## 4 행 2 열= 8
```

```
## 4 행 3 열= 12
```

```
times=1
i=0
while (times<(2000/2)){
  times=times*2
  i=i+1
    print(paste(2, "^" ,i, "=",times))
}
```

```
## [1] "2 ^ 1 = 2"
## [1] "2 ^ 2 = 4"
## [1] "2 ^ 3 = 8"
## [1] "2 ^ 4 = 16"
## [1] "2 ^ 5 = 32"
## [1] "2 ^ 6 = 64"
## [1] "2 ^ 7 = 128"
## [1] "2 ^ 8 = 256"
## [1] "2 ^ 9 = 512"
## [1] "2 ^ 10 = 1024"
```

```
S=1234
r=0.01
i=0
while(S>=(1234/2)){
  print("i= ")
  print(i)
  print("S= ")
  print(S)
  S=S/{(1+r)^i}
  i=i+1
}
```

```
## [1] "i= "
## [1] 0
## [1] "S= "
## [1] 1234
## [1] "i= "
## [1] 1
## [1] "S= "
## [1] 1234
```

```
## [1] "i= "
## [1] 2
## [1] "S= "
## [1] 1222
## [1] "i= "
## [1] 3
## [1] "S= "
## [1] 1198
## [1] "i= "
## [1] 4
## [1] "S= "
## [1] 1162
## [1] "i= "
## [1] 5
## [1] "S= "
## [1] 1117
## [1] "i= "
## [1] 6
## [1] "S= "
## [1] 1063
## [1] "i= "
## [1] 7
## [1] "S= "
## [1] 1001
## [1] "i= "
## [1] 8
## [1] "S= "
## [1] 933.9
## [1] "i= "
## [1] 9
## [1] "S= "
## [1] 862.5
## [1] "i= "
## [1] 10
## [1] "S= "
## [1] 788.6
## [1] "i= "
## [1] 11
## [1] "S= "
```

```
## [1] 713.9
## [1] "i= "
## [1] 12
## [1] "S= "
## [1] 639.9
```