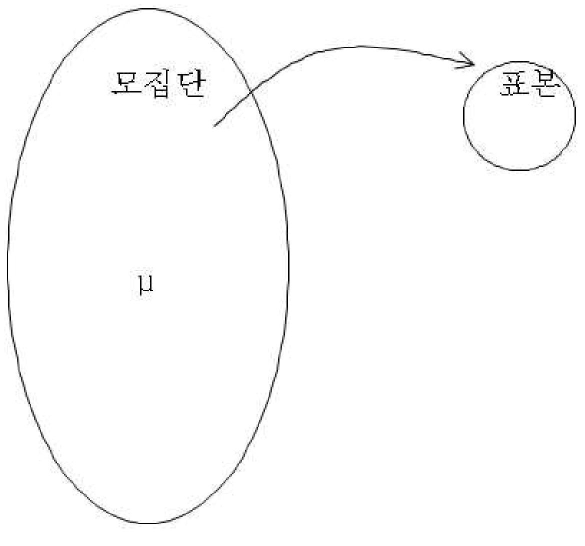
6장. 중심극한정리

1. 표본분포

통계 이론은 이미 알고 있는 약간의 정보(흔히 표본을 통해서)를 이용하여 미지의 상황을 파악하는데 도움을 주는 과학적인 방법이다. 미지의 상황이란 항상 확정적 사건이라기보다는 어떤 사건이 더 발생할 가능성이 많은가로 평가되어야 함은 당연하다. 그래서 확률의 개념은 매우 빈번하게 사용하고 있다.

그러나 정확한 확률을 계산한다는 것은 매우 지리하고 힘든 과정을 요한다. 그래서 어떤 사건이 일어날 수 있는 가능성을 표로 요약한 것을 확률분포라 한다. 이러한 확률분포를 만들기 위해서는 각 사건이 일어날 가능성을 찾아야 하는데 때론 이 과정이 매우 어렵다. 다행히 어떤 사건이 규칙성이 있어 수학적 함수로 표현 가능하다면 확률의 계산은 매우 편리할 것이며 더욱이 이 수학적 함수가 우리 실생활에서 흔히 일어날 수 있는 사건과 유사하다면 매우 유용할 것이다. 이러한 유용한 함수를 확률분포함수라 하며 정규분포, 이항분포, χ^2분포, t-분포, F-분포 등 이미 유용하게 사용하고 있는 분포는 다양하다.

[6, 7주차 강의]



1) 표본분포

표본분포란 표본을 이용하기 때문에 발생하는 것으로 표본통계량의 분포를 지칭한다. 표본을 뽑게되면 뽑힌 자료를 이용하여 통계량을 계산할 수 있으나 이 통계량의 값은 표본이 어떻게 뽑히느냐에 따라 달라질 것이다. 그래서 통계량은 확률변수가 되며 통계이론에 의해 일정한 분포함수를 갖게 된다.

· **모집단 분포의 추론** : 모집단에서 임의로 뽑은 표본(random sample)들의 분포가 모집단의 분포와 항상 비슷하게 나온다는 보장이 없다. 특히 표본이 개수가 적은 경우 표본의 분포로 모집단의 형태를 추론하기는 어렵다. 그래서 실제로는 표본들을 통해 모집단의 분포를 결정하기보다는 모집단의 분포로 미리 가정한 분포와 다르다고 할 수 있는지를 표본을 통해 판단하는 정도이다.

· **모수의 추론** : 모집단에서 실제로 관심을 갖는 것은 분포의 형태보다는 모수의 값이다. 모수를 추론하기 위하여 표본들을 적당한 형태로 변환한 표본들의 함수인 통계량(statistic, 예를 들어 표본평균, 표본분산 등)을 구하게 된다. 그런데 이러한 통계량 값은 우리가 얻은 표본에 따라 매번 다른 값이 나올 것이다. 즉 통계량은 확률변수이며 어떤 확률분포를 따르게 된다.

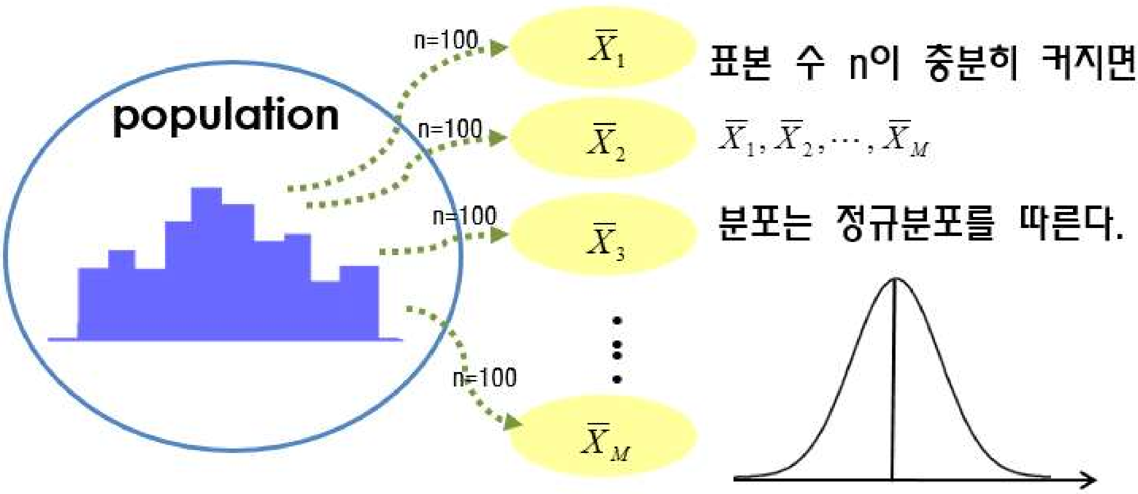
· **통계량의 확률분포** : 표본을 이용하여 구한 통계량이 모수에 얼마나 근접한 값일까? 하는 의문은 당연히 발생한다. 왜냐하면 모수와 멀어져 있는 통계량은 가치를 없으며 그래서 대부분의 사람들은 통계량을 모수와 가까운 값으로 찾기를 원한다. 그래서 구한 통계량이 모수와 얼마나 다를까 혹은 같을까를 나타낼 수 있는 방법이 필요하다. 이를 위하여 확률적으로 설명하고자 한다. 앞서 모수의 추론에서 설명하였듯이 통계량은 확률변수이므로 확률분포를 갖게 되며 특정 이론에 따라 규칙성을 찾을 수 있다. 이런 이론적 바탕으로 확률계산을 위해서 통계량의 분포를 얻게되며 이 분포로부터 확률을 계산하여 모수와의 관계를 확률적으로 설명하고 한다. 통계량의 분포는 통계량에 따라 t, F, χ^2 분포 등을 따른다.

2) 표본통계량의 분포

- 표본평균의 분포

$X_1,\ X_2,\ \cdots, X_n$이 평균이 μ이고 분산이 σ^2인 임의의 분포에서 뽑은 확률표본일 때, $n \to \infty$이면 $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$를 따른다. 그래서 표준화한 통계량은

$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$을 따르게 된다.



[중심극한 정리]

n이 클 때 => $E(\overline{X}) = \mu,\ Var(\overline{X}) = \frac{\sigma^2}{n},\ \overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

6.1. 이산확률분포에서의 중심극한정리

1) 이항분포 ($\hat{p}$의 표본분포)

$E(\hat{p}) = p,\ Var(\hat{p}) = \frac{p(1-p)}{n},\ \overline{X} \sim N(p, \frac{p(1-p)}{n})$, n이 클 때

par(mfrow=c(2,2))

n1 = 30 ; p = 0.1 ; mu =n1*p ; var = n1*p*(1-p) # 모수 : 모평균=15, 모분산= 7.5

① 표본평균의 평균과 분산 계산

```
m = 10                              # 자료 set의 개수 ( xbar의 개수)
n = 100                             # 난수 개수(표본의 개수)를 여러 가지로...
# 모수 : 모평균, 모분산 = 25

xbar = rep(0 , m)                   # 초기치를 0으로

for  (i in 1 : m) {
        x =  rbinom(n, n1, p)       # 이항분포에서 난수발생
        xbar[i] = mean(x)     }

xbar_mu = mean(xbar)                # xbar의 평균
xbar_var = var(xbar)                # xbar의의 분산

cat( "표본의 개수 = ", n,  ", 자료 set의 개수 = ", m, "\n",
 "모평균 = ", mu, ",  표본평균의 평균 = ", xbar_mu, "\n",
 "모분산 = ", var, ", var / n = ", var/n, ", 표본평균의 분산 = ", xbar_var, "\n")
```

③ 표본평균의 분포와 정규분포 비교

```
hist( xbar, breaks = "fd", prob = T )                      # 히스토그램
     curve( dnorm( x, mu, sqrt(var/n)), add = T)            # + 정규분포그림
```

[NOTE] 동영상강의와 다름(모든 실습을 아래와 같이 해주십시오...)

[실습1] 모수를 변화시키고, 자료 set의 개수(m=1000)으로 고정하고,
표본의 수(n=10, 100, 1000, 100000)를 증가시켜, 결과가 어떻게 변하는지 확인(해석)

par(mfrow=c(2,2)) => 4개의 변화를 해석 => 이후 실습도 계속

[실습2] 자료 set의 개수(m), 표본의 개수(n)를 커게해서, 결과를 확인(해석)

par(mfrow=c(1,1)) => 어떤 결과인지 해석 => 이후 실습도 계속

2) 포아송 분포

```
par(mfrow=c(2,2))
lambda = 5 ;  mu = lambda ; var = lambda      # 모수 : 모평균 = 5, 모분산= 5
```

① 표본평균의 평균과 분산 계산

```
m = 10                              # 자료 set의 개수 ( xbar의 개수)
n = 100                             # 난수 개수(표본의 개수)를 여러 가지로...
# 모수 : 모평균, 모분산 = 25

xbar = rep(0 , m)                   # 초기치를 0으로

for  (i in 1 : m) {
      x =  rpois(n, lambda)              # 포아송분포에서 난수발생
      xbar[i] = mean(x)     }

xbar_mu = mean(xbar)                # xbar의 평균
xbar_var = var(xbar)                # xbar의의 분산

cat( "표본의 개수 = ", n,  ", 자료 set의 개수 = ", m, "\n",
 "모평균 = ", mu, ",  표본평균의 평균 = ", xbar_mu, "\n",
 "모분산 = ", var, ", var / n = ", var/n, ", 표본평균의 분산 = ", xbar_var, "\n")
```

③ 표본평균의 분포와 정규분포 비교

```
hist( xbar, breaks = "fd", prob = T )                    # 히스토그램
      curve( dnorm( x, mu, sqrt(var/n)), add = T)        # + 정규분포그림
```

[실습3] 모수를 변화시키고, 자료 set의 개수(m=1000)으로 고정하고,
표본의 수(n=10, 100, 1000, 100000)를 증가시켜, 결과가 어떻게 변하는지 확인(해석)

[실습4] 자료 set의 개수(m), 표본의 개수(n)를 커게해서, 결과를 확인(해석)

[과제17] ([실습1] ~ [실습4] 실습내용)

동영상을 잘 들어시고, 과제를 하시고, 해석을 해주시길..

첨부파일 : 학번이름17.hwp (예 : 20192260홍길동17.hwp)

- m의 개수를 증가시켜 결과가 어떻게 변하는지 확인(해석)
- m, n의 개수를 증가시켜 결과가 어떻게 변하는지 확인(해석)
- 그래프 창의 그래프(해석)