

컴퓨터 응용통계

5 분포이론

최경미

5.1 베르누이 분포 (Bernouille Distribution)

정의 5.1 X 는 성공 또는 실패 두 경우를 표현한다.

$$X(\text{성공}) = 1, X(\text{실패}) = 0.$$

$$P(X = 1) = p, \quad P(X = 0) = 1 - p = q, \quad p + q = 1$$

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

$$X \sim \text{Bernoulli}(p)$$

예제 5.1 베르누이 시행의 대표적인 예를 들어보자.

- (1) 동전 던지기의 앞면 (H) 또는 뒷면 (T)
- (2) 게임의 승리 또는 패배
- (3) 질병의 유무
- (4) 제품품질의 불량 여부

체크

$$f(0) = P(X = 0) = p^0 q^{1-0} = q$$

$$f(1) = P(X = 1) = p^1 q^{1-1} = p$$

정리 5.1 베르누이 확률변수의 평균 μ 와 분산 σ^2

$$X \sim \text{Bernoulli}(p)$$

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

$$\mu = E[X] = \sum_x x f(x) = (0)(q) + (1)(p) = p$$

$$E[X^2] = \sum_x x^2 f(x) = (0^2)(q) + (1^2)(p) = p$$

$$\sigma^2 = \text{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p) = pq$$

예제 5.2 어떤 유명 농구선수(Stephen Curry)의 3점 슛 성공률이 44%로 알려져 있다. 이 선수가 3점 슛을 한 번 던질 때, 슛의 성공 또는 실패를 확률변수 X 로 나타내고 확률분포, 평균과 분산을 구해보자.

$$X(\text{성공}) = 1, X(\text{실패}) = 0.$$

$$\mu = E[X] = p = 0.44$$

$$P(X = 1) = 0.44, \quad P(X = 0) = 0.56.$$

$$\sigma^2 = \text{Var}(X) = pq$$

$$= (0.44)(0.56) = 0.2464$$

$$X \sim \text{Bernoulli}(0.44)$$

$$f(x) = (0.44)^x (0.56)^{1-x}, x = 0, 1$$

5.2 이항분포 (Binomial distribution)

예제 3점 슛 성공률이 0.44인 농구선수가 3점 슛을 10번 던질 때, 총 성공 횟수를 확률변수 X 로 표현해보자.

각 슛은 독립이다.

각 슛의 결과는 성공 또는 실패이다.

따라서 각 슛은 $Bernoulli(0.44)$ 를 따른다.

만약, 10번의 3점 슛 중에서 2, 4, 5, 7에서 성공하고, 나머지에서 실패했다고 가정하자.

각 슛 $X_1 = X_3 = X_6 = X_8 = X_9 = X_{10} = 0$

$$X_2 = X_4 = X_5 = X_7 = 1$$

총 성공횟수

$$X = X_1 + X_2 + \cdots + X_{10}$$

$$= 0 + 1 + 0 + 1 + 1 + 0 + 1 + 0 + 0 + 0 = 4$$

$$X_1, X_2, \dots, X_{10} \sim iid Bernoulli(0.44)$$

$$X = X_1 + X_2 + \cdots + X_{10} \sim B(10, 0.44)$$

성공확률이 p 이고 서로 독립인 n 번의 베르누이 시행에서 성공횟수는 이항분포를 따른다.

각 시행은 성공확률 p 이고 서로 독립인 베르누이 확률변수 X_1, X_2, \dots, X_n 이므로, 다음이 성립한다.

$$X_1, X_2, \dots, X_n \text{ iid } \sim Bernoulli(p)$$

iid=independent (독립)

identically distributed (동일분포)

$$X = X_1 + X_2 + \cdots + X_n$$

= 성공확률이 p 이고 서로 독립인 n 개의 베르누이 확률변수의 합

= 성공확률이 p 이고 서로 독립인 n 번의 베르누이 시행에서 성공횟수

$$X \sim B(n, p)$$

$$X \sim B(n, p)$$

$$X_1, X_2, \dots, X_n \text{ iid } \sim \text{Bernoulli}(p)$$

$$X = X_1 + X_2 + \dots + X_n$$

정의 5.2 확률변수 X 가 성공확률이 p 이고 서로 독립인 n 번의 베르누이 시행에서 성공횟수일 때, X 는 이항분포를 따르고,

$$X \sim B(n, p)$$

라고 나타낸다. B=binomial, n = 시행횟수, p = 성공확률.

X 의 확률함수는 다음과 같다.

$$f(x) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, \dots, n$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = n \text{ 번의 시행 중, } x \text{ 번의 성공이 나타나는 경우의 수}$$

$$p = P(X_1 = 1), q = P(X_1 = 0), p + q = 1$$

이항분포 $B\left(3, \frac{1}{6}\right)$ 의 확률함수

예제 5.3 주사위 한 개를 3번 던질 때, 1이 나타나는 횟수를 X 라고 두자.

각 시행이 $\{1\}$ 이면 성공(S), $\{2,3,4,5,6\}$ 이면 실패(F)라고 두자.

성공 $S = \{1\}$

실패 $F = \{2,3,4,5,6\}$

$$P(S) = p = \frac{1}{6}, P(F) = q = \frac{5}{6}$$

각 주사위 던지기는 서로 독립이고, 성공확률이 동일한 베르누이 시행이다.

표본공간 $S = \{FFF, SFF, FSF, FFS, SSF, SFS, FSS, SSS\}$

1이 0번 나옴
 $X=0$

1이 1번 나옴
 $X=1$

1이 2번 나옴
 $X=2$

1이 3번 나옴
 $X=3$

$$X_1, X_2, X_3 \sim iid \text{ Bernoulli } \left(\frac{1}{6}\right)$$

$$\text{첫번째 주사위 } X_1(F) = 0, X_1(S) = 1$$

$$\text{두번째 주사위 } X_2(F) = 0, X_2(S) = 1$$

$$\text{세번째 주사위 } X_3(F) = 0, X_3(S) = 1$$

$$\text{총 } S \text{의 개수 } X = X_1 + X_2 + X_3 \sim B\left(3, \frac{1}{6}\right)$$

$$X(FFF) = X_1(F) + X_2(F) + X_3(F) = 0 + 0 + 0 = 0$$

$$X(SFF) = X(FSF) = X(FFS) = 1$$

$$X(SSF) = X(SFS) = X(FSS) = 2$$

$$X(SSS) = 3$$

$$X=1$$

$$P(SFF)$$

$$= P(\text{시행1에서 1 나타남} \cap \text{시행2에서 1 안나타남} \cap \text{시행3에서 1 안나타남})$$

$$= P(S \cap F \cap F) = P(S)P(F)P(F) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$$

$$P(FSF) = P(F \cap S \cap F) = P(F)P(S)P(F) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$$

$$P(FFS) = P(F \cap F \cap S) = P(F)P(F)P(S) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$$

$$P(X=1) = \binom{3}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{3-1}$$

$$\binom{3}{1} = \frac{3!}{1!2!} = 3 \text{ (성공이 1번 나타나는 경우의 수)}$$

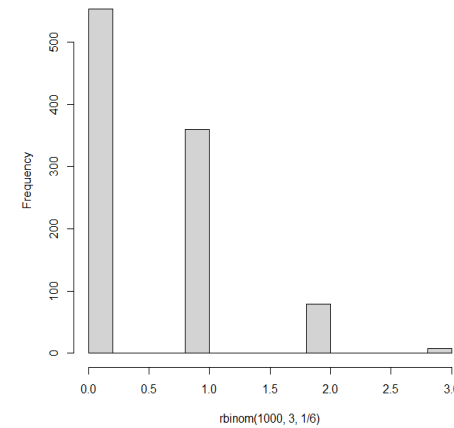
이항분포 $B\left(3, \frac{1}{6}\right)$ 의 확률함수

$X = x$	경우	확률	경우의 수	$P(X = x)$
0	FFF	$\left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^3$	$\binom{3}{0} = 1$ 셋 중 성공 없음	$\binom{3}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^3$
1	SFF	$\left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$	$\binom{3}{1} = 3$ 셋 중 성공 1	$\binom{3}{1} \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$
	FSF	$\left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$		
	FFS	$\left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2$		
2	SSF	$\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$	$\binom{3}{2} = 3$ 셋 중 성공 2	$\binom{3}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$
	SFS	$\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$		
	FSS	$\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$		
3	SSS	$\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^0$	$\binom{3}{3} = 1$ 모두 성공	$\binom{3}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^0$

X = 주사위 세 번 던지기에서 1이 나오는 횟수

$$X \sim B(3, 1/6)$$

$$f(x) = \binom{3}{x} \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{3-x}, x = 0, 1, 2, 3$$



$$X \sim B(n, p)$$

$$f(x) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, \dots, n$$

$$\text{Bernoulli}(p) = B(1, p)$$

정리 5.2 $X \sim B(n, p)$ 의 평균, 분산

$$X_1, X_2, \dots, X_n \text{ iid } \sim \text{Bernoulli}(p)$$

$$X = X_1 + X_2 + \dots + X_n \sim B(n, p)$$

$$\mu = E[X]$$

$$= E[X_1 + X_2 + \dots + X_n] \quad (\text{이항분포의 정의})$$

$$= E[X_1] + E[X_2] + \dots + E[X_n] \quad (\text{기대값의 선형성})$$

$$= np \quad (\text{동일분포의 가정})$$

$$\sigma^2 = \text{Var}(X)$$

$$= \text{Var}(X_1 + X_2 + \dots + X_n) \quad (\text{이항분포의 정의})$$

$$= \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) \quad (\text{독립성})$$

$$= pq + pq + \dots + pq \quad (\text{동일 분포})$$

$$= npq \quad (\text{동일분포의 가정})$$

두 확률변수가 독립이면, 합의 분산은 분산의 합이 된다.

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

$$X_1, X_2, \dots, X_n \text{ iid } \sim \text{Bernoulli}(p)$$

$$X = X_1 + X_2 + \dots + X_n \sim B(n, p)$$

$$\mu = E[X]$$

$$= \sum x f(x) \quad (\text{기대값 정의})$$

$$= \sum x \binom{n}{x} p^x q^{n-x} \quad (\text{pdf})$$

$$= np \quad (\text{계산복잡})$$

$$\sigma^2 = \text{Var}(X)$$

$$= E[X^2] - (E[X])^2$$

$$= \sum x^2 \binom{n}{x} p^x q^{n-x} - \mu^2 \quad (\text{pdf})$$

$$= npq \quad (\text{계산복잡})$$

예제 5.4 $X \sim B(3, 1/6)$ 의 평균과 분산, 확률

$$f(x) = \binom{3}{x} \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{3-x}, x = 0, 1, 2, 3$$

$$\mu = E[X] = np = (3) \left(\frac{1}{6}\right) = \frac{1}{2}$$

$$\sigma^2 = \text{Var}(X) = npq = (3) \left(\frac{1}{6}\right) \left(\frac{5}{6}\right) = \frac{5}{12}$$

$$P(X < 2)$$

$$= P(X \leq 1)$$

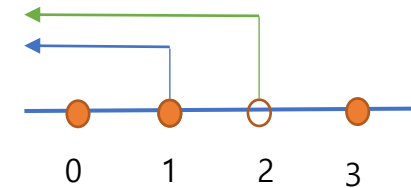
$$= P(X = 0) + P(X = 1)$$

$$= \binom{3}{0} \left(\frac{5}{6}\right)^3 + \binom{3}{1} \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2 = \frac{5^3 + 3 \cdot 5^2}{6^3}$$

$$= 0.9259$$

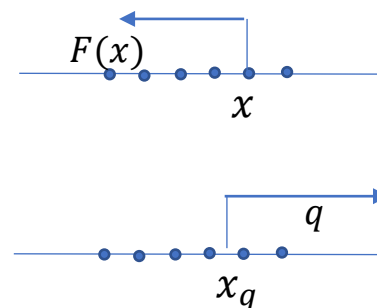
$$= \text{pbinom}(1, 3, 1/6)$$

- X 가 이산형 확률변수이므로, $P(X < 2) = P(X \leq 1)$ 임에 주의하자



R 실습 $X \sim B(n, p)$

- 확률함수 $\text{dbinom}(x, n, p) = f(x) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, \dots, n$
- 분포함수 $\text{pbinom}(x, n, p) = F(x) = P(X \leq x)$
- 분위수 $\text{qbinom}(1 - q, n, p) = x_q, P(X > x_q) = q$
- 난수생성 $\text{rbinom}(x, n, p), B(n, p)$ 를 따르는 확률변수의 값을 x 개 생성함.



예제 5.5 R을 이용하여, 주사위를 3번 던질 때 1이 나오는 횟수 X 를 생각해보자. 확률변수 X 의 분포는 $X \sim B\left(3, \frac{1}{6}\right)$ 이다. 세 번 중 한 번만 1이 나올 확률을 구해보자.

$$P(X = 1) = f(1) = \binom{3}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{3-1} = \text{dbinom}\left(1, 3, \frac{1}{6}\right) = 0.3472$$

세 번 중 두 번 이상 1이 나올 확률을 두 가지 방법으로 구해보자.

$$P(X \geq 2) = P(X = 2) + P(X = 3) = f(2) + f(3) = \text{dbinom}\left(2, 3, \frac{1}{6}\right) + \text{dbinom}\left(3, 3, \frac{1}{6}\right) = 0.0741$$

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - \text{pbinom}\left(1, 3, \frac{1}{6}\right) = 0.0741 \quad \square$$

연습문제

1. 자연 상태에서 신생아가 남자일 확률은 0.51이며, 여자일 확률은 0.49이고, 각 자녀의 성별은 독립이라고 가정하자. 어떤 부부가 3명의 자녀를 낳을 때 자녀 수에 대한 설명 중 옳은 것은 무엇인가?

a. 각 자녀가 아들인지 아닌지는 베르누이분포 $\text{Bernoulli}(0.49)$ 를 따른다.

b. 세 자녀 중 딸의 수는 이항분포 $B(3,0.49)$ 를 따른다.

c. 세 자녀 중 아들이 1명일 확률은 0.3673이다.

d. 세 자녀 중 딸이 없을 확률은 0.1327이다.

딸 수 $X \sim B(3,0.49)$

아들 수 $Y \sim B(3,0.51)$

(풀이) a. 각 자녀가 아들인지 아닌지는 베르누이분포 $\text{Bernoulli}(0.51)$ 를 따른다.

b. 세 자녀 중 딸의 수는 이항분포 $B(3,0.49)$ 를 따른다.

c. 세 자녀 중 아들의 수를 Y 라고 두자.

$$P(Y = 1) = \binom{3}{1} (0.51)^1 (0.49)^2 = \text{dbinom}(1,3,0.51) = 0.3673$$

d. 세 자녀 중 딸의 수를 X 라고 두자.

$$P(X = 0) = \binom{3}{0} (0.49)^0 (0.51)^3 = \text{dbinom}(0,3,0.49) = 0.1327$$

혼자 풀기

1. 공정한 주사위 한 개를 4번 던질 때 2가 나타나는 횟수를 X 라고 두자.

(1) X 의 확률분포를 구하시오.

(2) X 의 평균과 분산을 구하시오.

(3) 2가 2번 이상 나타날 확률을 구하시오.

pbinom사용

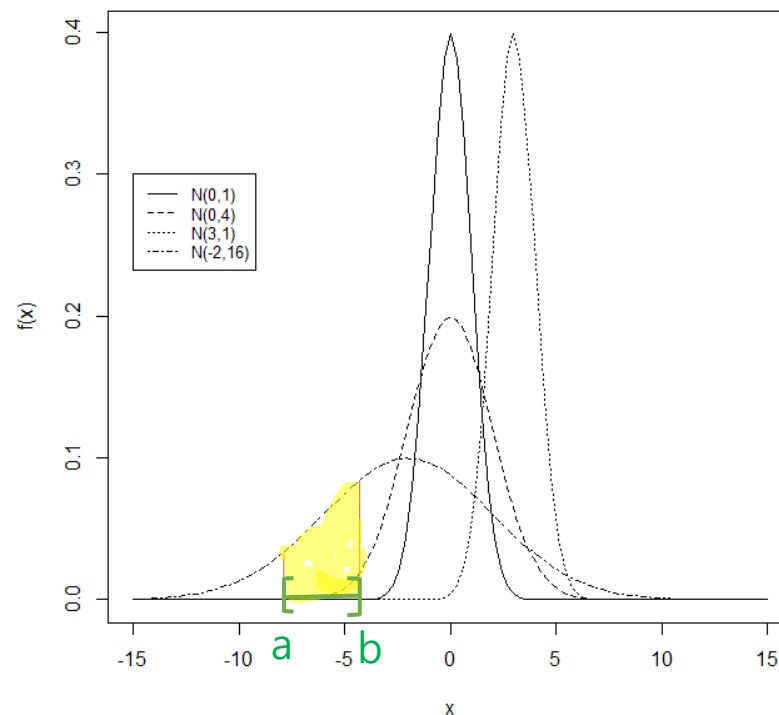
5.3 정규분포

- $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$
- $X \sim N(\mu, \sigma^2)$
- 평균은 $-\infty < \mu < \infty$ 이며, 분산은 σ^2 이고, σ 은 표준편차이다.
- 평균을 중심으로 대칭이다.
- 종모양이다.
- 연속형 확률분포이다.
- 한 점에서의 확률은 0이다.
- 확률변수가 어떤 구간에 속할 확률은 확률함수 아래의 면적이다.

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

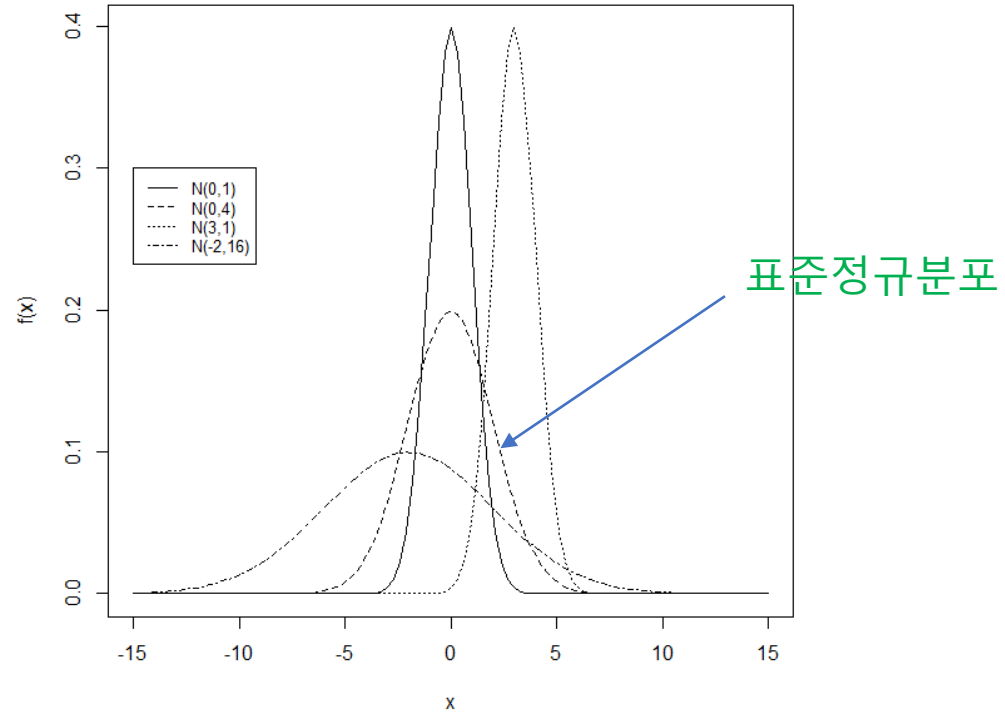
- 확률함수 아래의 전체 면적은 1이다.

$$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$



표준정규분포 (standard normal distribution) $N(0,1)$

- $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad -\infty < z < \infty$
- 평균 $\mu = 0$, 표준편차 $\sigma = 1$



예제 5.7 확률변수 Z 가 표준정규분포 $N(0,1)$

$$P(-1.5 \leq Z \leq 2.0) = \int_{-1.5}^{2.0} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

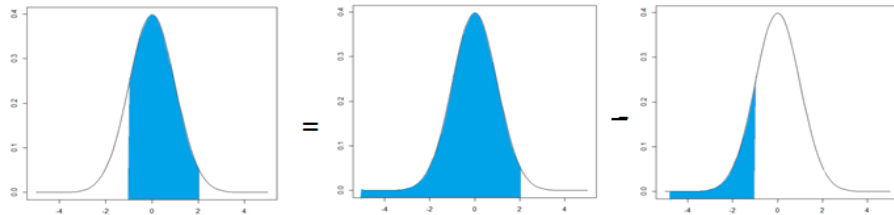
$$= P(Z \leq 2.0) - P(Z \leq -1.5) \text{ (그림 5.2)}$$

$$= P(Z \leq 2.0) - P(Z \geq 1.5) \text{ (대칭성)}$$

$$= P(Z \leq 2.0) - (1 - P(Z \leq 1.5)) \text{ (여사건의 확률)}$$

$$= 0.9772 - 0.0668 \text{ (표 찾기)}$$

$$= 0.9104.$$



$$\begin{aligned} P(Z \leq -1.5) &= P(Z \geq 1.5) \\ &= 1 - P(Z \leq 1.5) \\ &= 1 - 0.9332 \\ &= 0.0668 \end{aligned}$$

z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628							
0.5	0.6915	0.695	0.6985							
0.6	0.7257	0.7291	0.7324							
0.7	0.758	0.7611	0.7642							
0.8	0.7881	0.791	0.7939							
0.9	0.8159	0.8186	0.8212							
1.0	0.8413	0.8438	0.8461							
1.1	0.8643	0.8665	0.8686							
1.2	0.8849	0.8869	0.8888							
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

$$P(Z \leq 1.50) = 0.9332$$

$$P(Z \leq 2.00) = 0.9772$$

정리 5.3 정규분포의 성질

① **선형성.** X 가 정규분포 $N(\mu, \sigma^2)$ 를 따르면, $aX + b$ 는 정규분포 $N(a\mu + b, a^2\sigma^2)$ 을 따른다.

증명

$$E[aX + b] = aE[X] + b = a\mu + b$$

$$\text{Var}(aX + b) = a^2\text{Var}(X) = a^2\sigma^2$$

정규분포를 선형변환시켜도 정규분포이다. 증명 생략

② **표준화.** X 가 정규분포 $N(\mu, \sigma^2)$ 를 따르면, $Z = \frac{X-\mu}{\sigma}$ 는 표준정규분포 $N(0,1)$ 을 따른다.

증명

$$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = 0$$

$$\text{Var}(Z) = \text{Var}\left(\frac{X-\mu}{\sigma}\right) = 1$$

성질 ①에 따라서 $Z \sim N(0,1)$ 가 된다. \square

예제 5.8 $X \sim N(3, 2^2)$

$$Z = \frac{X-3}{2} \sim N(0,1)$$

$$P(1 < X \leq 5)$$

$$= P\left(\frac{1-3}{2} < \frac{X-3}{2} \leq \frac{5-3}{2}\right) \quad (\text{표준화})$$

$$= P(-1 < Z \leq 1)$$

$$= P(Z \leq 1) - P(Z \leq -1)$$

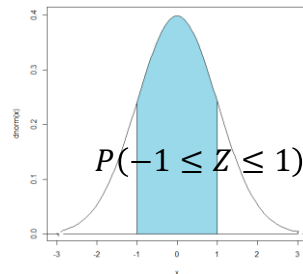
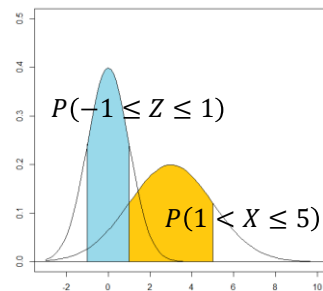
$$= P(Z \leq 1) - P(Z \geq 1) \quad (\text{대칭성})$$

$$= P(Z \leq 1) - (1 - P(Z \leq 1)) \quad (\text{여사건의 확률})$$

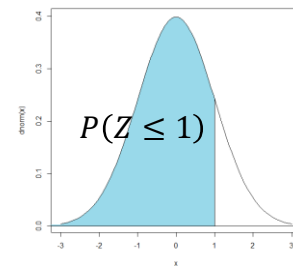
$$= 2P(Z \leq 1) - 1 \quad (\text{표 찾기})$$

$$= 2(0.8413) - 1$$

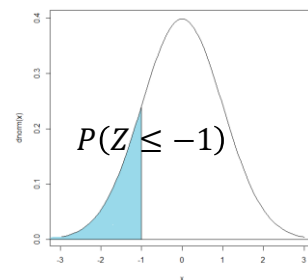
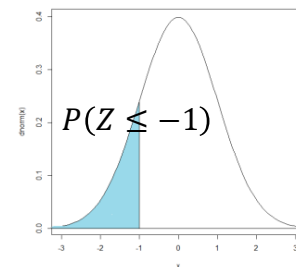
$$= 0.6826$$



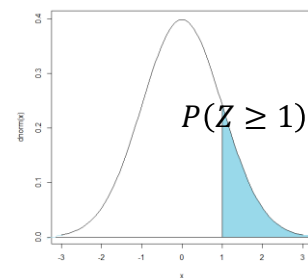
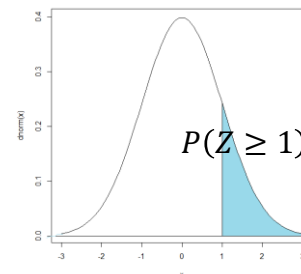
=



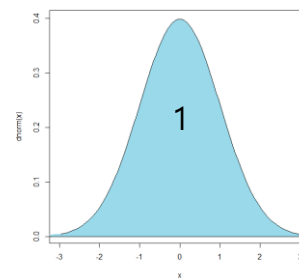
-



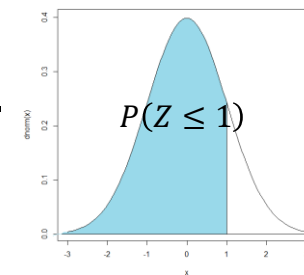
=



=

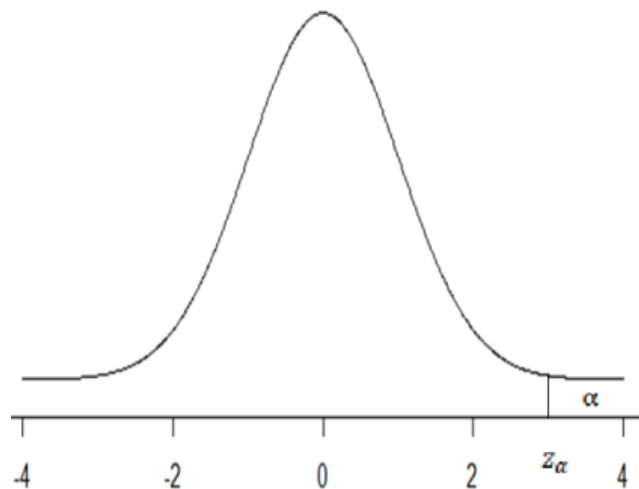


-



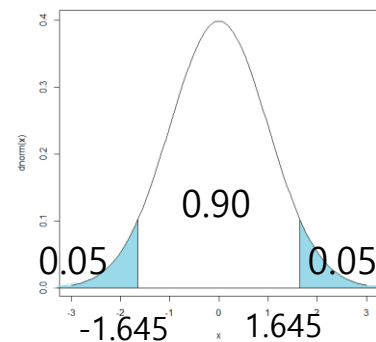
100(1- α) 백분위수 또는 α 분위수 (quantile) z_α

정의 5.4 $Z \sim N(0,1)$ 일 때, $P(Z > z_\alpha) = \alpha$ 를 만족하는 z_α 를 100(1- α) 백분위수 또는 α 분위수 (quantile)라고 정의한다.



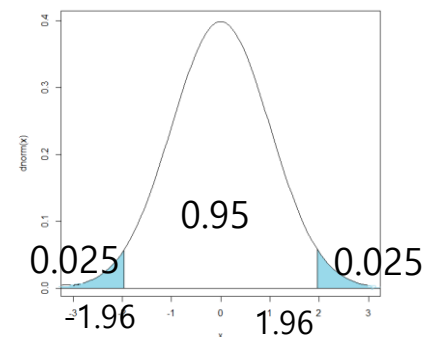
$$P(Z > z_{0.05}) = 0.05, z_{0.05} = 1.645$$

$$P(Z \leq z_{0.05}) = P(Z \leq 1.645) = 0.95$$



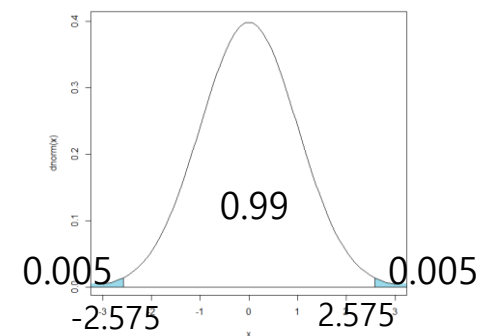
$$P(Z > z_{0.025}) = 0.025, z_{0.025} = 1.96$$

$$P(Z \leq z_{0.025}) = P(Z \leq 1.96) = 0.975$$



$$P(Z > z_{0.005}) = 0.005, z_{0.005} = 2.575$$

$$P(Z \leq z_{0.005}) = P(Z \leq 2.575) = 0.99$$

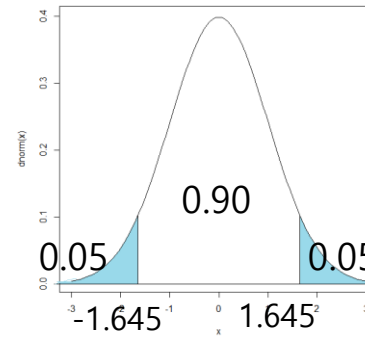
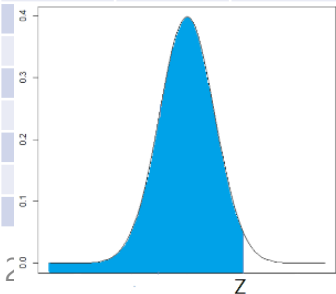


$$P(-z_{0.025} < Z < z_{0.025}) = 0.95$$

$$P(-1.96 < Z < 1.96) = 0.95$$

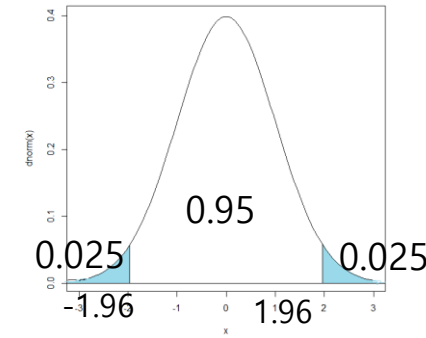
표준정규분포표 $P(Z \leq z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8				0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9				0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0				0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1				0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2				0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3				0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4				0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998



$$P(Z > z_{0.05}) = 0.05, z_{0.05} = 1.645$$

$$P(Z \leq z_{0.05}) = P(Z \leq 1.645) = 0.95$$

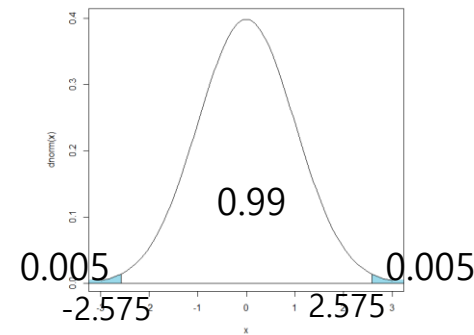


$$P(Z > z_{0.025}) = 0.025, z_{0.025} = 1.96$$

$$P(Z \leq z_{0.025}) = P(Z \leq 1.96) = 0.975$$

$$P(-z_{0.025} < Z < z_{0.025}) = 0.95$$

$$P(-1.96 < Z < 1.96) = 0.95$$



$$P(Z > z_{0.005}) = 0.005, z_{0.005} = 2.575$$

$$P(Z \leq z_{0.005}) = P(Z \leq 2.575) = 0.99$$

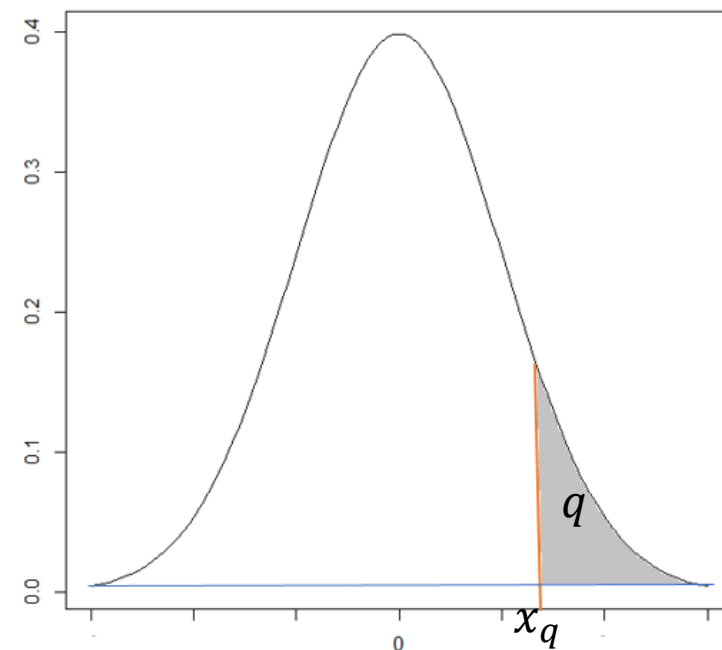
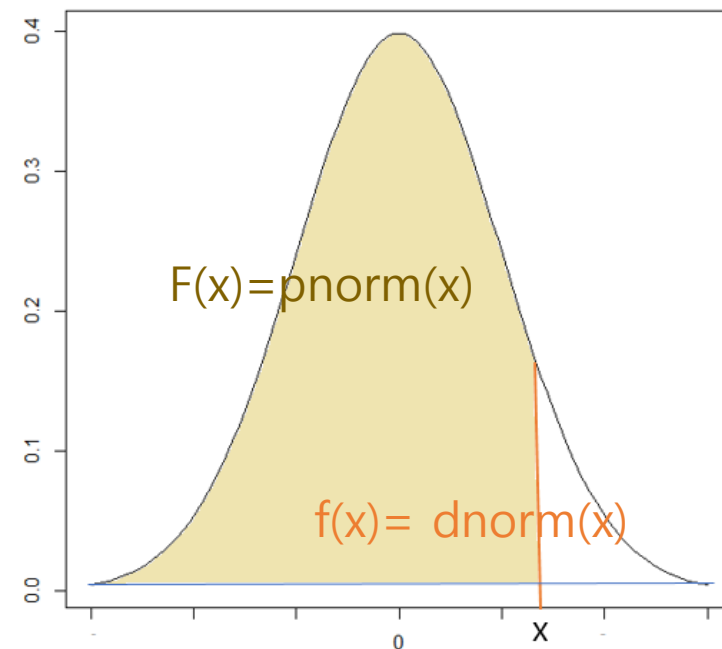
R 실습 $X \sim N(\mu, \sigma^2)$

확률함수 $dnorm(x, mu, sigma) = f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$

분포함수 $pnorm(x, mu, sigma) = F(x) = P(X \leq x)$

분위수 $qnorm(1 - q, mu, sigma) = x_q, P(X > x_q) = q$

난수생성 $rnorm(n, mu, sigma)$. $N(\mu, \sigma^2)$ 를 따르는 확률변수의 값을 n개 생성함.



예제 5.10 ① qnorm을 이용하여, 두 가지 방법으로 $Z_{0.05}, Z_{0.025}, Z_{0.005}$ 를 구해보고, ② pnorm을 이용하여, $X \sim N(3, 2^2)$ 일 때, $P(1 < X \leq 5)$ 를 계산해보자.

①

qnorm(0.995) # 2.58로 사용함

qnorm(0.975) # 1.96으로 사용함

qnorm(0.950) # 1.645로 사용함

qnorm(0.005, lower.tail=F) # 2.58로 사용함

qnorm(0.025, lower.tail=F) # 1.96으로 사용함

qnorm(0.050, lower.tail=F) # 1.645로 사용함

②

$$\begin{aligned} P(1 < X \leq 5) &= P(X \leq 5) - P(X \leq 1) \\ &= pnorm(5, 3, 2) - pnorm(1, 3, 2) \end{aligned}$$

$$\begin{aligned} P(1 < X \leq 5) &= P\left(\frac{1-3}{2} < \frac{X-3}{2} \leq \frac{5-3}{2}\right) \\ &= P(-1 < Z \leq 1) \\ &= P(Z \leq 1) - P(Z \leq -1) = pnorm(1) - pnorm(-1) \\ &= P(Z \leq 1) - P(Z > 1) \\ &= P(Z \leq 1) - (1 - P(Z \leq 1)) \\ &= 2P(Z \leq 1) - 1 = 2 * pnorm(1, 0, 1) - 1 \end{aligned}$$

이거 보기로 만들어서, 시험 출제함.

5.6 표본평균의 분포

증명. $aX + bY$ 가 정규분포임을 보이는 증명을 생략하고, 이들의 평균과 분산을 구해보자.

정리 5.6 정규분포를 따르는 두 확률변수의 선형결합은 정규분포를 따른다.

$$X \sim N(\mu_X, \sigma_X^2),$$

$$Y \sim N(\mu_Y, \sigma_Y^2),$$

X 와 Y 가 독립이라고 가정하자.

그러면, X 와 Y 의 선형결합의 분포는

$$aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$$

가 성립한다.

$$E[X] = \mu_X, \text{Var}(X) = \sigma_X^2$$

$$E[Y] = \mu_Y, \text{Var}(Y) = \sigma_Y^2$$

$$E[aX + bY] = aE[X] + bE[Y] = a\mu_X + b\mu_Y$$

X 와 Y 가 독립이면, 합의 분산을 다음과 같이 얻을 수 있다.

$$\text{Var}(aX + bY)$$

$$= \text{Var}(aX) + \text{Var}(bY) + 2\text{Cov}(aX, bY)$$

$$= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

$$= a^2\sigma_X^2 + b^2\sigma_Y^2 \text{ (독립이므로, Cov}(X, Y)=0)$$

예제 5.13 정규분포의 선형결합 (시험)

$X \sim N(-2, 1)$ 이고, $Y \sim N(0, 4)$ 이며, X 와 Y 가 독립이라고 가정하자.

$2X - 3Y$ 는 어떤 분포를 따르는가?

풀이.

$$E[2X - 3Y] = 2E[X] - 3E[Y] = 2(-2) - 3(0) = -4.$$

$$\text{Var}(2X - 3Y) = 2^2\text{Var}(X) + (-3)^2\text{Var}(Y) = 4(1) + 9(4) = 40$$

따라서 $2X - 3Y \sim N(-4, 40)$ 이다.

5.8 표본평균의 분포

정규분포의 성질

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

$$\frac{X - \mu}{\sigma} \sim N(0,1)$$

$$\mu = E[X], \quad \sigma^2 = \text{Var}(X), \quad \frac{X - E[X]}{\sqrt{\text{Var}(X)}} \sim N(0,1)$$

정리 5.7 표본평균의 분포

표본 X_1, X_2, \dots, X_n 이 독립이고, 동일분포 (iid) $N(\mu, \sigma^2)$ 를 따른다고 가정하자.

① 표본평균 \bar{X} 는 $N\left(\mu, \frac{\sigma^2}{n}\right)$ 를 따른다.

② 이를 표준화하여 표준정규분포를 따르는 확률변수 Z 를 정의할 수 있다.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

증명. 정리 5.6에 따라서, 표본평균 \bar{X} 는 정규분포를 따른다. 이때, 평균과 분산을 다음과 같이 구할 수 있다.

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \dots + \frac{1}{n}X_n$$

$$E[\bar{X}] = E\left[\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right]$$

$$= \frac{1}{n}(E[X_1] + E[X_2] + \dots + E[X_n])$$

$$= \mu$$

$$\text{Var}(\bar{X})$$

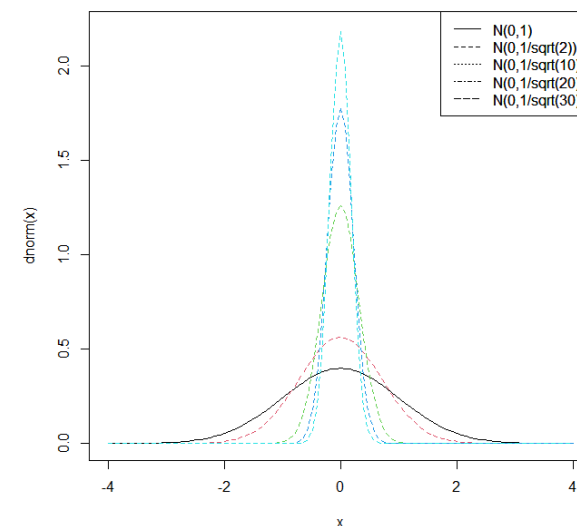
$$= \text{Var}\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right)$$

$$= \frac{1}{n^2}\text{Var}(X_1) + \frac{1}{n^2}\text{Var}(X_2) + \dots + \frac{1}{n^2}\text{Var}(X_n) \quad (\text{독립})$$

$$= \frac{\sigma^2}{n} \quad (\text{동일분포})$$

정규분포의 선형결합은 정규분포를 따른다.

$$\therefore \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$



예제 5.14 표본평균의 분포 (시험)

표본 X_1, X_2, \dots, X_{25} 가 독립이고 동일한 정규분포 $N(60, 36)$ 를 따르고, $P(\bar{X} > c) = 0.05$ 이면, c 는 얼마인가?

풀이.

$$X_1, X_2, \dots, X_{25} \sim \text{iid } N(60, 6^2)$$

$$\bar{X} \sim N\left(60, \left(\frac{6}{5}\right)^2\right)$$

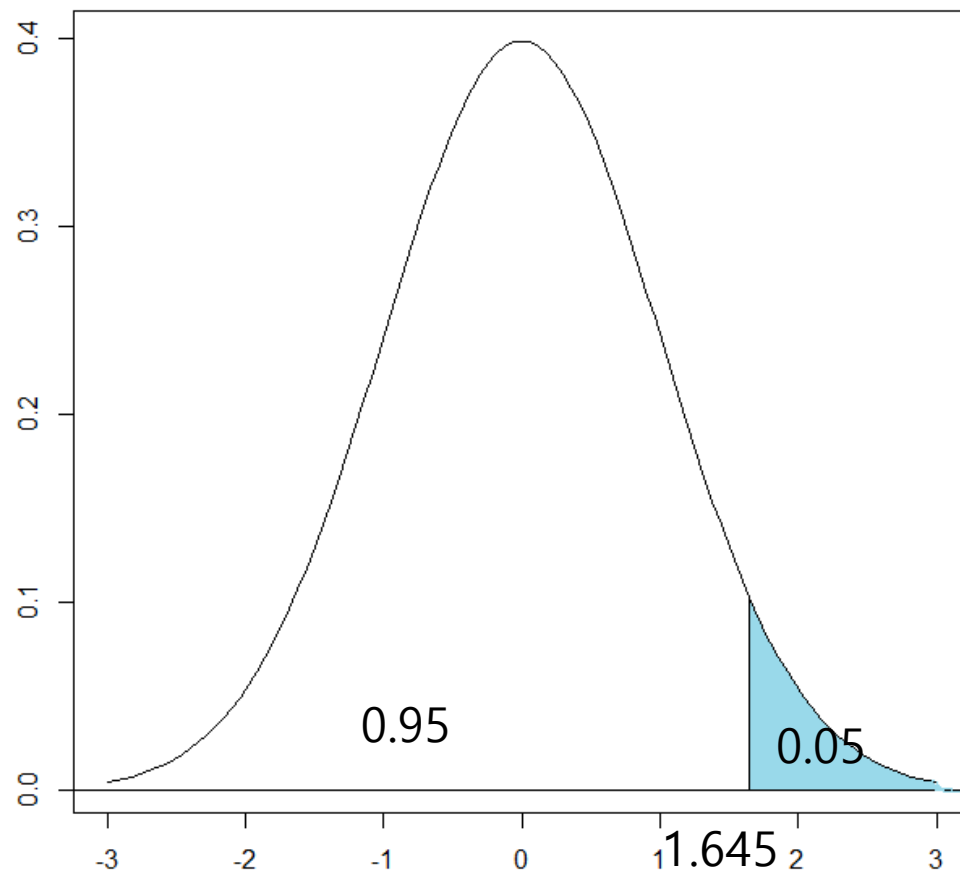
$$Z = \frac{\bar{X} - 60}{6/\sqrt{25}} \sim N(0, 1)$$

$$P\left(\frac{\bar{X} - 60}{\frac{6}{5}} > \frac{c - 60}{\frac{6}{5}}\right) = 0.05$$

$$P\left(Z > \frac{c - 60}{\frac{6}{5}}\right) = 0.05$$

$$\frac{c - 60}{1.2} = 1.645$$

$$c = 60 + (1.2)(1.645) = 61.974$$



연습문제

5. 어떤 회사의 공산품의 배터리 용량(W)이 평균이 36개월이고 표준편차가 2개월인 정규분포를 따른다고 가정하자. 이 회사에서 생산된 배터리 한 개의 용량이 30개월 미만일 확률은 얼마인가?

- ① 0.0013
- ② 0.0668
- ③ 0.9332
- ④ 0.9987
- ⑤ 위 보기 중 답 없음

(풀이) 1.

$$W \sim N(36, 2^2), P(W < 30) = P\left(Z < \frac{30-36}{2}\right) = P(Z < -3) = \text{pnorm}(-3, 0, 1) = 0.0013$$

연습문제

6. $X \sim N(5, 2^2)$ 일 때, $Y = \frac{5X-1}{2}$ 의 분포에 대한 설명 중 맞는 것은 어느 것인가?

- ① Y 는 t 분포를 따른다.
- ② Y 의 평균은 $22/3$ 이다.
- ③ Y 의 표준편차는 4이다.
- ④ $\frac{Y-12}{5}$ 는 표준정규분포를 따른다.
- ⑤ 위 보기 중 답 없음

(풀이) 4. $Y = \frac{5X-1}{2} \sim N(12, 5^2)$, $\frac{Y-12}{5} \sim N(0, 1)$

혼자 풀기

1. $X \sim N(-1, 2^2)$ 이고, $Y = \frac{-X+3}{2}$ 일 때, 다음 설명이 옳으면 참, 틀리면 거짓으로 나타내시오.

(1) X 의 분포는 평균이 -1 이고, 분산이 2 인 정규분포이다.

1 참 2 거짓

(2) Y 의 분포는 평균이 2 이고, 표준편차가 1 인 정규분포이다.

1 참 2 거짓

(3) $Y - 2$ 의 분포는 표준정규분포이다.

1 참 2 거짓

2 1 1

2. 어떤 화장품의 용량 X 가 평균 80ml , 표준편차가 2 ml 인 정규분포를 따른다고 가정하자. 이 화장품 용량이 78ml 이상이고, 83ml 이하일 확률에 대한 다음 설명이 옳으면 참, 틀리면 거짓으로 나타내시오.

(1) 이 확률은 $P(78 \leq X \leq 83)$ 이다.

1 참 2 거짓

(2) 표준화를 이용하면, 이 확률은 $P(Z \leq 1.5) - P(Z > -1)$ 와 동일하다.

1 참 2 거짓

(3) 이 확률은 0.7745375 이다.

1 참 2 거짓

1 2 1

5.9 중심극한정리 (Central Limit Theorem)

복습

표본

$$X_1, X_2, \dots, X_n \sim \text{iid } N(\mu, \sigma^2)$$

표본평균의 분포 $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ (정확한 분포)

표준화 $Z \sim \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

중심극한정리 5.8

표본 $X_1, X_2, \dots, X_n \sim \text{iid Any distribution}(\mu, \sigma^2)$

$n \gg 30, n \rightarrow \infty$

표본평균의 분포 $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$

표준화 $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1)$

중심극한정리 5.8 표본 X_1, X_2, \dots, X_n 가 독립이고, 평균이 μ 이고, 분산이 σ^2 인 임의의 분포를 동일하게 따른다고 가정하자 (**iid** any distribution). 만약 표본의 크기 n 이 충분히 크면 ($n \gg 30$), 표본평균 \bar{X} 는 **근사적으로** $N\left(\mu, \frac{\sigma^2}{n}\right)$ 를 따른다.

$n \rightarrow \infty$ 일 때, $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1)$

주의 정규분포 가정이 없음

정리 5.8 중심극한정리 (Central Limit Theorem) 표현

표준화 $\frac{X - E[X]}{\sqrt{Var(X)}}$

평균으로 표현

독립이고 동일한분포(μ, σ^2) X_1, X_2, \dots, X_n

$$E[\bar{X}] = \mu, Var(\bar{X}) = \sigma^2/n$$

$n \rightarrow \infty$ 일 때,

$$Z = \frac{\bar{X} - E[\bar{X}]}{\sqrt{Var(\bar{X})}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0,1) \text{ (CLT)}$$

합으로 표현

독립이고 동일한분포를 따르는 X_1, X_2, \dots, X_n 의 합 S_n 을 아래와 같이 정의하자.

$$S_n = X_1 + X_2 + \dots + X_n = n\bar{X}$$

$$E[S_n] = E[X_1 + X_2 + \dots + X_n]$$

$$= E[X_1] + \dots + E[X_n]$$

$$= \mu + \dots + \mu = n\mu$$

$$Var(S_n) = Var(X_1 + X_2 + \dots + X_n)$$

$$= Var(X_1) + \dots + Var(X_n)$$

$$= n\sigma^2$$

$$\sqrt{Var(S_n)} = \sqrt{n\sigma^2} = \sqrt{n}\sigma$$

$$Z = \frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{n(\bar{X} - \mu)}{n\sigma/\sqrt{n}} \rightarrow N(0,1)$$

예제 5.15 $X_1, X_2, \dots, X_{49} \sim \text{iid Bernoulli}(0.4)$ (시험)

$$X_1, X_2, \dots, X_{49} \sim \text{iid Bernoulli}(0.4)$$

$$E[X_1] = p = 0.4, \text{Var}(X_1) = pq = 0.24$$

$$Z = \frac{\bar{X} - p}{\sqrt{pq}/\sqrt{n}} = \frac{\bar{X} - 0.4}{\sqrt{(0.4)(0.6)}/\sqrt{49}} \rightarrow N(0,1)$$

$$X = S_{49} = X_1 + X_2 + \dots + X_{49} \sim B(49, 0.4)$$

$$E[X] = np = (49)(0.4) = 19.6$$

$$\text{Var}(X) = npq = (49)(0.4)(0.6) = 11.76$$

$$Z = \frac{X - np}{\sqrt{npq}} = \frac{X - 19.6}{\sqrt{11.76}} \rightarrow N(0,1)$$

$$\text{근사 } P(9 \leq X \leq 25) \text{ (이산형)}$$

$$= P(8.5 \leq X \leq 25.5) \text{ (원래 구간이 포함되도록)}$$

연속보정, 아직 이산형)

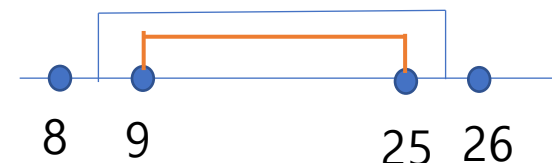
$$= P\left(\frac{8.5 - 19.6}{\sqrt{11.76}} \leq \frac{X - 19.6}{\sqrt{11.76}} \leq \frac{25.5 - 19.6}{\sqrt{11.76}}\right) \text{ (표준화)}$$

$$\cong P\left(\frac{8.5 - 19.6}{\sqrt{11.76}} \leq Z \leq \frac{25.5 - 19.6}{\sqrt{11.76}}\right) \text{ (중심극한정리, 연속형)}$$

$$= P(-3.24 \leq Z \leq 1.72)$$

$$= P(Z \leq 1.72) - P(Z \leq -3.24) \text{ (누적 확률 이용)}$$

$$= \text{pnorm}(1.72) - \text{pnorm}(-3.24) = 0.9567$$



$$P(9 \leq X \leq 25)$$

$$= P(X = 9) + \dots + P(X = 25)$$

$$= P(X \leq 25) - P(X \leq 8)$$

$$= \text{pbinom}(25, 49, 0.4) - \text{pbinom}(8, 49, 0.4)$$

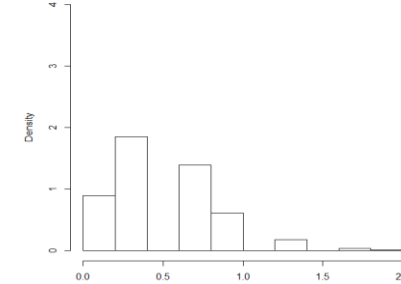
$$= 0.9558$$

$$\text{오차} \sim 0.0009$$

중심극한정리. $B\left(3, \frac{1}{6}\right)$ 의 표본평균의 분포 ① $n = 3$ ② $n = 10$ ③ $n = 30$

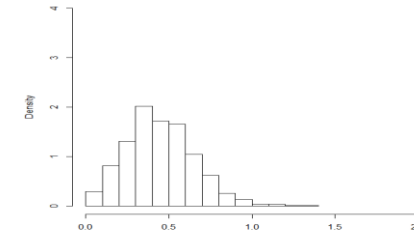
① $B\left(3, \frac{1}{6}\right)$ 에서 발생시킨 난수 $n=3$ 개의 평균을 계산하자. 이를 1000번 반복하여, 1000개 평균의 히스토그램을 그려보자.

$$\left. \begin{array}{l} X_1, X_2, X_3 \sim iid B\left(3, \frac{1}{6}\right) \Rightarrow \bar{X} \\ X_1, X_2, X_3 \sim iid B\left(3, \frac{1}{6}\right) \Rightarrow \bar{X} \\ \dots \dots \\ X_1, X_2, X_3 \sim iid B\left(3, \frac{1}{6}\right) \Rightarrow \bar{X} \end{array} \right\} 1000 \text{ 개의 평균}$$



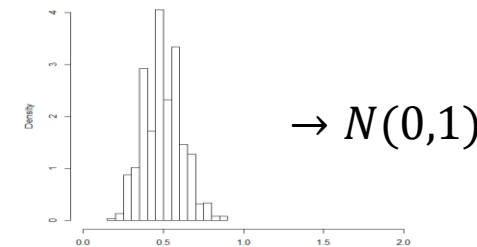
② $B\left(3, \frac{1}{6}\right)$ 에서 발생시킨 난수 $n=10$ 개의 평균을 계산하자. 이를 1000번 반복하여, 1000개 평균의 히스토그램을 그려보자.

$$\left. \begin{array}{l} X_1, X_2, X_3, \dots, X_{10} \sim iid B\left(3, \frac{1}{6}\right) \Rightarrow \bar{X} \\ X_1, X_2, X_3, \dots, X_{10} \sim iid B\left(3, \frac{1}{6}\right) \Rightarrow \bar{X} \\ \dots \dots \\ X_1, X_2, X_3, \dots, X_{10} \sim iid B\left(3, \frac{1}{6}\right) \Rightarrow \bar{X} \end{array} \right\} 1000 \text{ 개의 평균}$$



③ $B\left(3, \frac{1}{6}\right)$ 에서 발생시킨 난수 $n=30$ 개의 평균을 계산하자. 이를 1000번 반복하여, 1000개 평균의 히스토그램을 그려보자.

$$\left. \begin{array}{l} X_1, X_2, X_3, \dots, X_{30} \sim iid B\left(3, \frac{1}{6}\right) \Rightarrow \bar{X} \\ X_1, X_2, X_3, \dots, X_{30} \sim iid B\left(3, \frac{1}{6}\right) \Rightarrow \bar{X} \\ \dots \dots \\ X_1, X_2, X_3, \dots, X_{30} \sim iid B\left(3, \frac{1}{6}\right) \Rightarrow \bar{X} \end{array} \right\} 1000 \text{ 개의 평균}$$



R

①

```
x<-rbinom(3000, 3, 1/6)
```

B(3,1/6)에서 난수를 3000 개 생성함

```
x <- matrix(x, 1000, 3)
```

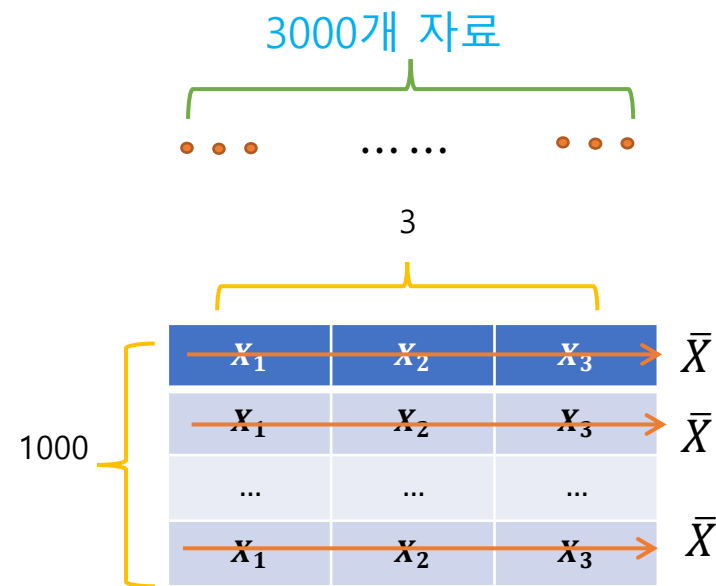
1000x3 행렬로 변환

```
x.mean <- apply(x,1,mean)
```

각 행에서 3개 난수의 평균 계산

```
hist(x.mean, main=" ", freq=F, xlab=" ", xlim=c(0,2.0), ylim=c(0,4.0))
```

1000개 표본평균의 히스토그램



혼자풀기 검사(5점)

20. 어느 해 추신수 선수의 미국 MLB 정규시즌 타율이 0.265이었다고 한다. 만약 추신수 선수가 바로 다음 해에도 동일한 타율을 유지한다고 가정한다면, 추신수 선수가 100 번 중 25번 이상 안타를 칠 확률은 무엇인가? 중심극한정리를 사용하고, p norm을 사용하자.

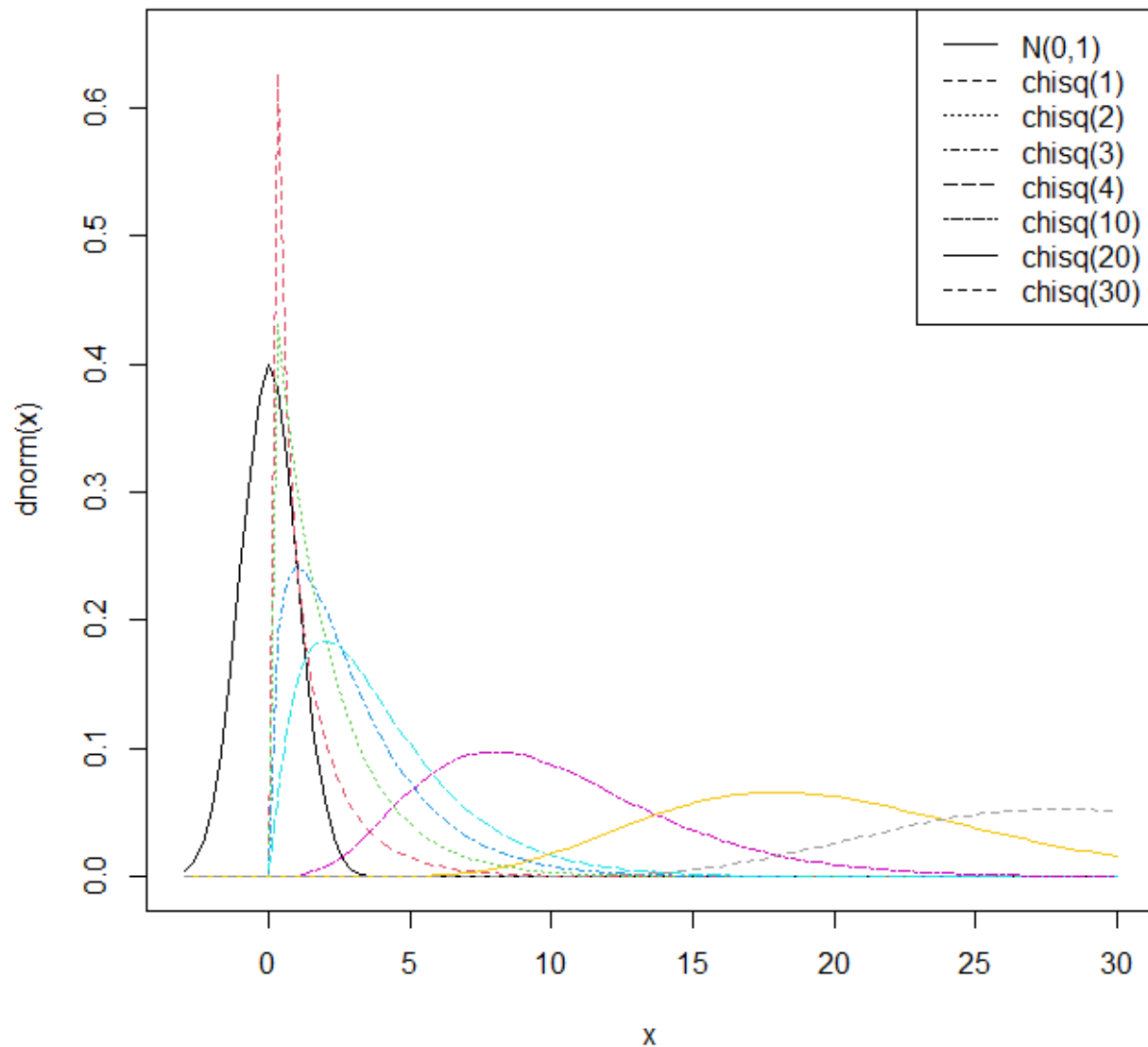
(1) 0.05635795 (2) 0.3252123 (3) 0.5130086 (4) 0.6747877 5 위 보기 중 답 없음

5.6 카이제곱 분포

정의 5.5 $Z \sim N(0,1)$ 이면 $Z^2 \sim \chi^2(1)$ 이다.

이때 1은 자유도(degrees of freedom)로 정의되며,

한 개의 표준정규분포를 사용했다는 의미이다.



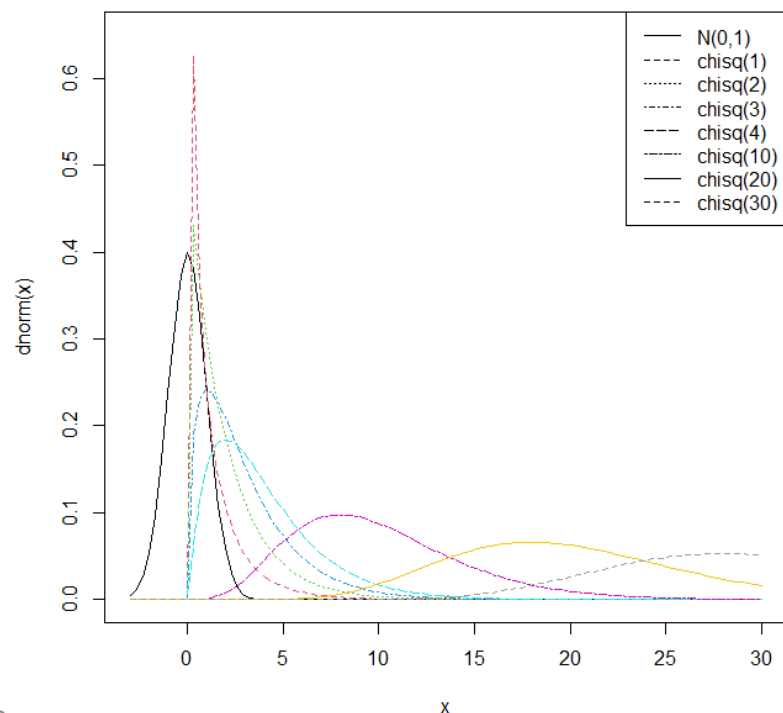
정리 5.4 카이제곱의 가법성

① $Z_1, Z_2, \dots, Z_r \sim iid N(0,1)$ 이라 두자.

$Z_1^2 \sim \chi^2(1), Z_2^2 \sim \chi^2(1), \dots, Z_r^2 \sim \chi^2(1)$, 그리고 독립

$$V = Z_1^2 + Z_2^2 + \dots + Z_r^2 \sim \chi^2(r)$$

자유도 r 은 독립인 표준정규분포 확률변수 r 개를 제공해서 더했다는 의미이다.



② 동일한 정규분포 $N(\mu, \sigma^2)$ 를 따르는 독립인 확률변수 X_1, X_2, \dots, X_n 를 표준화한 후

제공하여 더하면 $\chi^2(n)$ 을 따른다.

$$X_1, X_2, \dots, X_n \sim iid N(\mu, \sigma^2)$$

$$\frac{X_1 - \mu}{\sigma}, \frac{X_2 - \mu}{\sigma}, \dots, \frac{X_n - \mu}{\sigma} \sim iid N(0,1)$$

$$\left(\frac{X_1 - \mu}{\sigma}\right)^2, \left(\frac{X_2 - \mu}{\sigma}\right)^2, \dots, \left(\frac{X_n - \mu}{\sigma}\right)^2 \sim iid \chi^2(1)$$

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$$

③ $V_1 \sim \chi^2(r_1), V_2 \sim \chi^2(r_2)$ 이고, V_1 과 V_2 가 독립이면, $V_1 + V_2 \sim \chi^2(r_1 + r_2)$ 이다

- R `dchisq`, `pchisq`, `qchisq`, `rchisq`

예제 5.12 (시험)

$Z_1, Z_2, \dots, Z_{10} \sim iid N(0,1)$ 이면, 다음과 같은 분포를 정의할 수 있다.

$$Z_1^2, Z_2^2, Z_5^2 \sim iid \chi^2(1)$$

$$Z_1^2 + Z_2^2 + Z_5^2 \sim \chi^2(3)$$

$X_1, X_2, \dots, X_{10} \sim iid N(-3, 2^2)$ 이면, 다음과 같은 분포를 정의할 수 있다.

$$\frac{X_1+3}{2}, \frac{X_2+3}{2}, \dots, \frac{X_5+3}{2} \sim iid N(0,1)$$

$$\left(\frac{X_1+3}{2}\right)^2, \left(\frac{X_2+3}{2}\right)^2, \left(\frac{X_5+3}{2}\right)^2 \sim iid \chi^2(1)$$

$$\left(\frac{X_1+3}{2}\right)^2 + \left(\frac{X_2+3}{2}\right)^2 + \left(\frac{X_5+3}{2}\right)^2 \sim \chi^2(3)$$

5.7 t 분포

정의 5.6

$$Z \sim N(0,1), V \sim \chi^2(r)$$

Z 와 V 가 독립일 때,

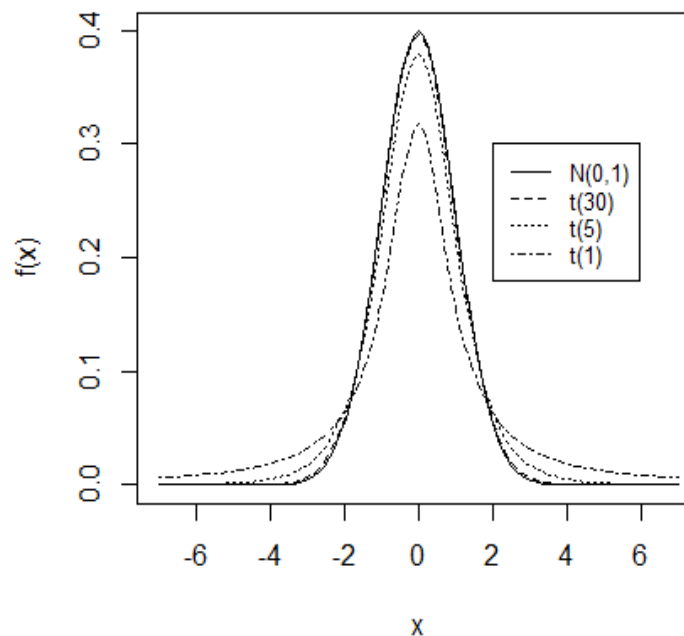
t 분포가 다음과 같이 정의된다.

$$t = \frac{Z}{\sqrt{V/r}} \sim t(r)$$

$$Z, Z_1, Z_2, \dots, Z_r \sim iid N(0,1)$$

$$V = Z_1^2 + Z_2^2 + \dots + Z_r^2 \sim \chi^2(r)$$

$$t = \frac{Z}{\sqrt{(Z_1^2 + Z_2^2 + \dots + Z_r^2)/r}} \sim t(r)$$



자유도가 커질 때 t 분포가 표준정규분포 $N(0,1)$ 에 가까워진다.

즉, $\lim_{r \rightarrow \infty} t(r) = N(0,1)$ 이 성립한다.

t 분포는 표준정규분포보다 꼬리가 두껍다.

R dt , pt , qt , rt

예제 5.12 (시험)

$Z_1, Z_2, \dots, Z_{10} \sim iid N(0,1)$ 라고 두자.

$$Z_2^2, Z_3^2 \sim iid \chi^2(1)$$

$$Z_2^2 + Z_3^2 \sim \chi^2(2)$$

$$\frac{Z_1}{\sqrt{(Z_2^2 + Z_3^2)/2}} \sim t(2)$$

$$\frac{Z_2}{\sqrt{(Z_2^2 + Z_3^2)/2}} \not\sim t(2)$$

$X_1, X_2, \dots, X_{10} \sim iid N(-3, 2^2)$ 이면, 다음과 같은 분포를 정의할 수 있다.

$$\frac{X_1+3}{2}, \frac{X_2+3}{2}, \dots, \frac{X_{10}+3}{2} \sim iid N(0,1)$$

$$\left(\frac{X_2+3}{2}\right)^2, \left(\frac{X_5+3}{2}\right)^2 \sim iid \chi^2(1)$$

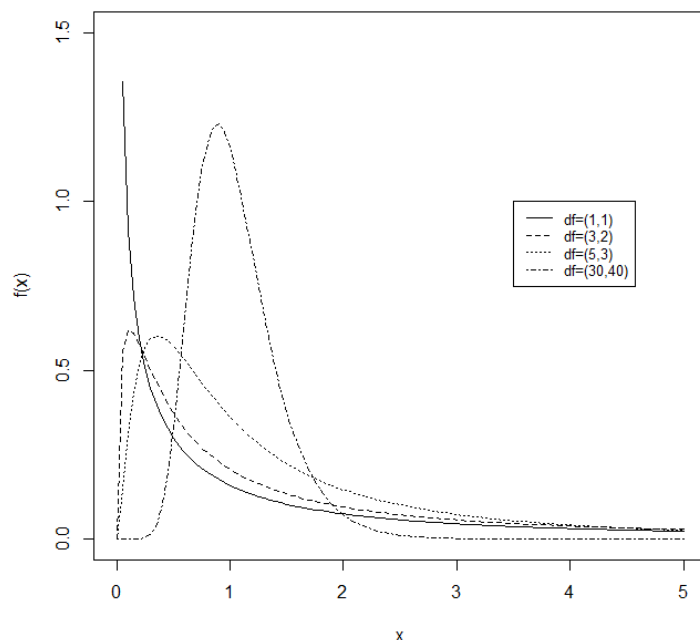
$$\left(\frac{X_2+3}{2}\right)^2 + \left(\frac{X_5+3}{2}\right)^2 \sim \chi^2(2)$$

$$\frac{\frac{X_1+3}{2}}{\sqrt{\left(\left(\frac{X_2+3}{2}\right)^2 + \left(\frac{X_5+3}{2}\right)^2\right)/2}} \sim t(2)$$

5.7 F 분포

정의 5.7 $V_1 \sim \chi^2(r_1)$, $V_2 \sim \chi^2(r_2)$ 이고, V_1 와 V_2 가 독립일 때, F 분포가 다음과 같이 정의된다.

$$F = \frac{V_1/r_1}{V_2/r_2} \sim F(r_1, r_2)$$



정리 5.5

① $F \sim F(r_1, r_2)$ 이면, $\frac{1}{F} \sim F(r_2, r_1)$ 이다.

$$F = \frac{V_1/r_1}{V_2/r_2}, \quad V_1 \sim \chi^2(r_1), \quad V_2 \sim \chi^2(r_2), \quad V_1, V_2 \text{ 독립}$$

$$\frac{1}{F} = \frac{V_2/r_2}{V_1/r_1} \sim F(r_2, r_1)$$

② $t \sim t(r)$ 이면, $t^2 = F(1, r)$ 이다.

$$Z \sim N(0,1), V \sim \chi^2(r), \quad Z, V \text{ 독립}$$

$$t = \frac{Z}{\sqrt{V/r}} \sim t(r)$$

$$t^2 = \left(\frac{Z}{\sqrt{V/r}} \right)^2 = \frac{Z^2/1}{V/r} \sim F(1, r)$$

R df, pf, qf, rf

예제 5.12 (시험)

$Z_1, Z_2, \dots, Z_{10} \sim iid N(0,1)$ 라고 두자.

$$Z_1^2, Z_2^2, \dots, Z_{10}^2 \sim iid \chi^2(1)$$

$$Z_1^2 + Z_2^2 \sim \chi^2(2)$$

$$Z_3^2 + Z_4^2 + \dots + Z_{10}^2 \sim \chi^2(8)$$

$$\frac{(Z_1^2 + Z_2^2)/2}{(Z_3^2 + Z_4^2 + \dots + Z_{10}^2)/8} \sim F(2,8)$$

} 독립

연습문제 (시험)

9. Z_1, Z_2, \dots, Z_{30} iid $\sim N(0,1)$ 이다. 다음 분포의 정의가 틀린 것은 어느 것인가? 3

1 $Z_{20}^2 \sim \chi^2(1)$ 2 $Z_3^2 + Z_5^2 + Z_9^2 \sim \chi^2(3)$ 3 $\frac{(Z_1^2 + Z_2^2 + Z_3^2)/3}{(Z_1^2 + Z_2^2 + \dots + Z_{10}^2)/10} \sim F(3,10)$

4 $\frac{Z_2}{\sqrt{(Z_3^2 + Z_4^2 + \dots + Z_8^2)/6}} \sim t(6)$ 5 위 보기 중 답 없음

과제

- 2, 3, 4, 7, 16, 17, 18 (각자 풀어보기, 제출하지 않음)