05 CHAPTER

정규분포와 관련 분포

1

제5-2장
정규분포와 관련 분포

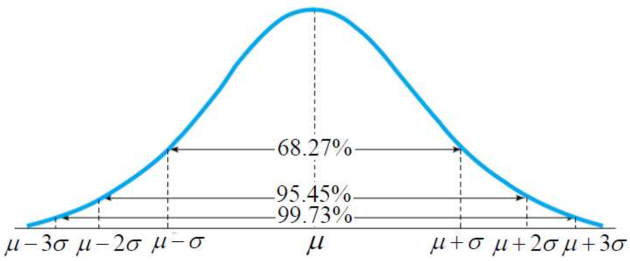
2

1. 정규분포

[정의 8-1] 정규분포(normal distribution) $X \sim N(\mu, \sigma^2)$

: 기대값을 중심으로 대칭이며, 중심위치는 기대값, 산포는 표준편차에 의해 결정되는 엎어 놓은 종 모양의 분포

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty$$



정규분포의 R 함수

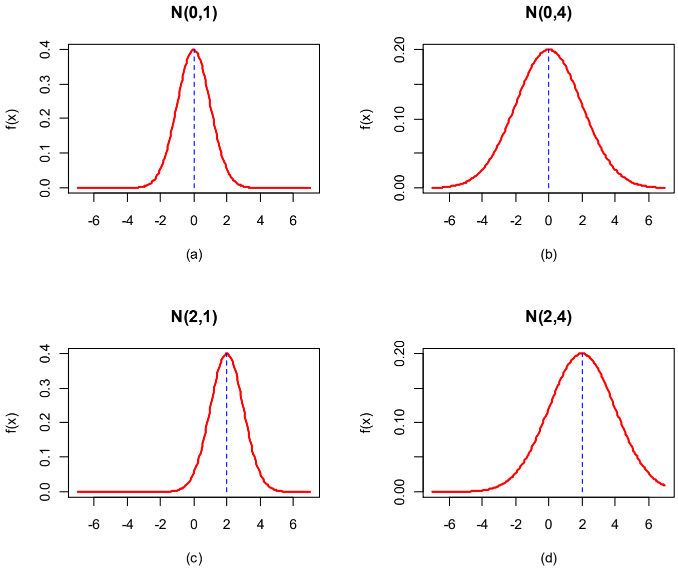
# 확률밀도함수 (mean=기대값, sd=표준편차)
# mean, sd를 생략하면 표준정규분포로 계산함 (mean = 0, sd = 1)
dnorm(x, mean = 0, sd = 1)
# Excel 함수 = NORM.DIST(x, mean, sd, FALSE)

# 누적분포함수 (q=분위수)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE)
# Excel 함수 = NORM.DIST(x, mean, sd, TRUE)

# 분위수 (p=누적확률)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE)
# Excel 함수 = NORM.INV(p, mean, sd)

# 정규 확률변수 (n=난수의 개수)
rnorm(n, mean = 0, sd = 1)
# Excel 함수 = NORM.INV(RAND( ), mean, sd) ⇒ 난수 1개 생성

[예 8-1] N(0,1), N(0,2^2), N(2,1), N(2,2^2)에 대한 확률밀도함수



5/33

5

1. 정규분포

[정의 8-2] 표준정규분포(standard normal distribution)

: 기댓값은 0, 표준편차는 1인 정규분포 $Z \sim N(0,1)$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \ -\infty < z < \infty$$

- CDF $\Phi(y) \equiv P(Z < y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$
- 분위수 $z_p \equiv \Phi^{-1}(p) = \{ y \mid \Phi(y) = p \}$
- MGF $m_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} \frac{e^{tz}}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(z^2 - 2tz + t^2) + \frac{t^2}{2}\right] dz$$

$$= \exp(\frac{t^2}{2}) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(z-t)^2\right] dz = \exp(\frac{t^2}{2})$$

6/33

6

1. 정규분포

- 정규분포의 MGF $X \sim N(\mu,\sigma^2)$

$$m_X(t)=E(e^{tX})=\int_{-\infty}^{\infty}\frac{e^{tx}}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx$$

$$z=\frac{x-\mu}{\sigma} \Rightarrow x=\mu+\sigma z \Rightarrow dx=\sigma dz$$

$$m_X(t)=e^{\mu t}\int_{-\infty}^{\infty}\frac{e^{\sigma tz}}{\sqrt{2\pi}\sigma}\exp\left(-\frac{z^2}{2}\right)\sigma dz$$

$$=e^{\mu t}\int_{-\infty}^{\infty}e^{\sigma tz}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz$$

$$=e^{\mu t}E(e^{\sigma tZ})=e^{\mu t}m_Z(\sigma t)=\exp(\mu t+\frac{\sigma^2t^2}{2})$$

$$X=\mu+\sigma Z$$

$$m_X(t)=E(e^{tX})=E(e^{t(\mu+\sigma Z)})=e^{\mu t}E(e^{\sigma tZ})$$

$$=e^{\mu t}m_Z(\sigma t)=\exp(\mu t+\frac{\sigma^2t^2}{2})$$

7/33

7

1. 정규분포

[정리 8-1] 정규분포의 선형변환

$$X\sim N(\mu,\sigma^2),\ Y=a+bX \Rightarrow Y\sim N(a+b\mu,b^2\sigma^2)$$

[증명] $m_X(t)=\exp(\mu t+\frac{\sigma^2t^2}{2})$

$$m_Y(t)=E(e^{tY})=E(e^{t(a+bX)})=e^{at}E(e^{btX})$$

$$=e^{at}m_X(bt)=e^{at}\exp\left[\mu bt+\frac{\sigma^2(bt)^2}{2}\right]=\exp\left[(a+\mu b)t+\frac{(b\sigma)^2t^2}{2}\right]$$

[따름정리 8-1] 정규분포의 표준화(standardization)

$$X\sim N(\mu,\sigma^2),\ Z=\frac{X-\mu}{\sigma} \Rightarrow Z\sim N(0,1)$$

[증명] $E(Z)=E\left(\frac{X-\mu}{\sigma}\right)=\frac{1}{\sigma}E(X-\mu)=\frac{1}{\sigma}(\mu-\mu)=0$

$$Var(Z)=Var\left(\frac{X-\mu}{\sigma}\right)=\frac{1}{\sigma^2}Var(X-\mu)=\frac{\sigma^2}{\sigma^2}=1$$

8/33

8

1. 정규분포

- 정규분포의 확률계산 → 표준정규분포 사용

$$P(a<X<b)=P(\frac{a-\mu}{\sigma}<Z<\frac{b-\mu}{\sigma})=\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})$$

[예 8-2] X~N(2,4)일 때, P(-1<X<4) 계산

$$P(-1<X<4)=P(\frac{-1-2}{2}<Z<\frac{4-2}{2})=P(-1.5<Z<1)=\Phi(1)-\Phi(-1.5)$$

[표 A-1] → $\Phi(1)\equiv P(Z\le 1)\doteq 0.8413,\ \Phi(1.5)\equiv P(Z<1.5)\doteq 0.9332$

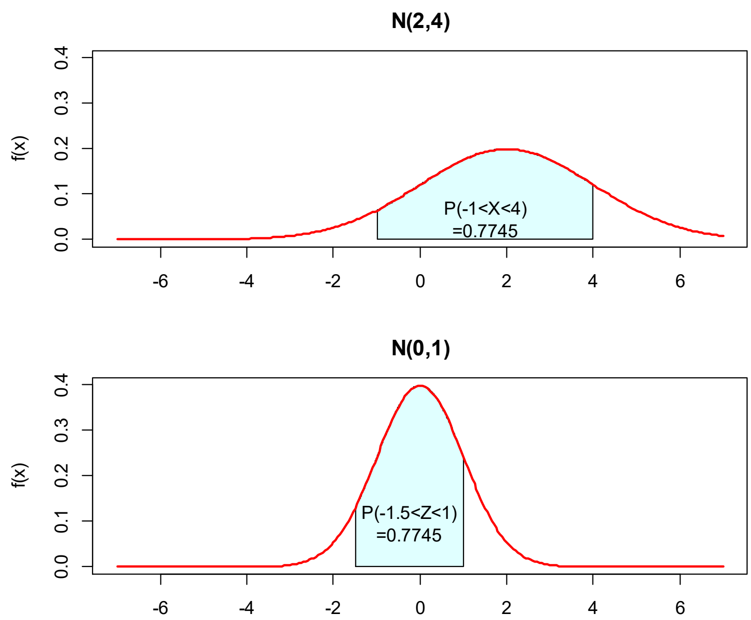
$$\Phi(-1.5)\equiv P(Z\le -1.5)=1-\Phi(1.5)\doteq 1-0.9332=0.0668$$

$$P(-1<X<4)\doteq 0.8413-0.0668=0.7745$$

```
# pnorm(x, mean, sd) 사용
pnorm(4, 2, 2) - pnorm(-1, 2, 2); pnorm(1)-pnorm(-1.5)
[1] 0.7745375      [1] 0.7745375
```

9/33

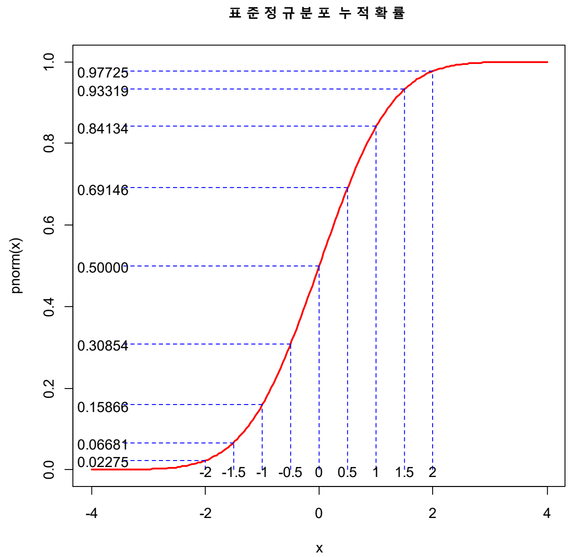
9



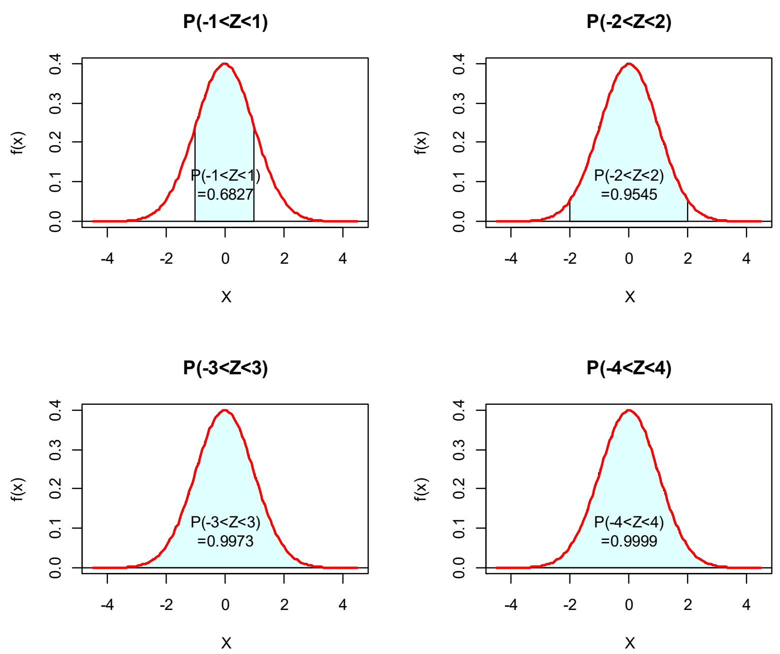
10/33

10

[예 8-3] <부록 B-1>의 표준정규 누적분포표 작성

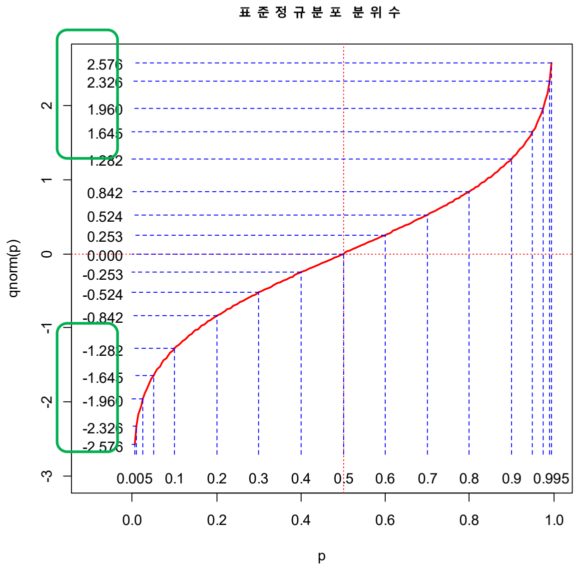


11/33



12/33

[예 8-4] 누적확률 0.5%, 1.0%, 2.5%, 5.0%, 10%~90%, 95%, 97.5%, 99%, 99.5%에 대한 표준정규분포의 분위수 표 작성



13/33

13

1. 정규분포

[예 8-5] 20대 남성의 신장이 평균이 175, 분산이 64인 정규분포를 따른다고 할 때, 20대 남성 중 180과 185 사이의 신장을 갖는 비율

$$X \sim N(175, 8^2)$$

$$P(180 < X < 185) = P\left(\frac{180-175}{8} < Z < \frac{185-175}{8}\right)$$
$$= P(0.625 < Z < 1.25) = \Phi(1.25) - \Phi(0.625)$$
$$\Phi(1.25) \doteq 0.8944$$
$$\Phi(0.62) \doteq 0.7324,\ \Phi(0.63) \doteq 0.7357 \quad \Rightarrow \ \Phi(0.625) \doteq 0.7341$$
$$\Rightarrow \ P(180 < X < 185) \doteq 0.8944 - 0.7341 = 0.1603$$

```
# pnorm(x, mean, sd) 사용
pnorm(185, 175, 8) - pnorm(180, 175, 8)
[1] 0.1603358
```

14/33

14

1. 정규분포

[예 8-6] A사 Q모델의 엔진 수명은 평균 10년, 표준편차가 1.5년인 정규분포를 따름. Q모델 엔진의 무상보증비율을 5% 이내로 유지 → 무상보증기간은 최대 몇 년?

$$X \sim N(10, 1.5^2) \qquad P(X < m) \le 0.05$$

$$P(X < m) = P(Z < \frac{m-10}{1.5}) \le 0.05$$

$$\Rightarrow \frac{m-10}{1.5} \le \Phi^{-1}(0.05) \doteq -1.645$$

$$\Rightarrow m \le -1.645 \times 1.5 + 10 = 7.5325$$

```
# qnorm(p, mean, sd) 사용
qnorm(0.05, 10, 1.5)
[1] 7.53272

10+qnorm(0.05)*1.5
[1] 7.53272
```

15/33

15

1. 정규분포

[정리 8-2] 정규분포의 가법성 1

$$X \sim N(\mu_1, \sigma_1^2),\ Y \sim N(\mu_2, \sigma_2^2)\ \& X \perp Y$$
$$\Rightarrow X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

[증명] $m_X(t) = \exp(\mu_1 t + \frac{\sigma_1^2 t^2}{2}),\ m_Y(t) = \exp(\mu_2 t + \frac{\sigma_2^2 t^2}{2})$

$$\Rightarrow m_{X+Y}(t) = m_X(t) m_Y(t) = \exp\left[(\mu_1 + \mu_2)t + \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}\right]$$

$$\Rightarrow X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

- X와 Y가 독립이 아닌 경우

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\sigma_{XY})$$

16/33

16

1. 정규분포

[예 8-7] 어떤 볼트의 직경은 평균 20mm, 표준편차 0.3mm인 정규분포를 따르고, 너트의 직경은 평균 21.5mm, 표준편차 0.4mm인 정규분포를 따른다. 이 볼트가 너트에 안 들어갈 확률?

$$X \sim N(20,\ 0.3^2),\ \ Y \sim N(21.5,\ 0.4^2)$$

$$\Rightarrow Y - X \sim N(1.5,\ 0.4^2 + 0.3^2)$$

$$P(X > Y) = P(Y - X < 0)$$

$$= P(Z < \frac{0-1.5}{\sqrt{0.25}}) = \Phi(-3) \doteq 1 - 0.9987 = 0.0013$$

```
# pnorm(p, mean, sd) 사용
pnorm(0, 21.5-20, sqrt(0.4^2 +0.3^2))
[1] 0.001349898
```

17/33

17

1. 정규분포

[따름정리 8-2] 정규분포의 가법성 2

$$X_i \overset{ind}{\sim} N(\mu_i, \sigma_i^2),\ \ i = 1, 2, \ldots, n\ \ \Rightarrow Y \equiv \sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$$

[증명] $m_{X_i}(t) = \exp(\mu_i t + \frac{\sigma_i^2 t^2}{2})$

$$\Rightarrow m_Y(t) = \prod_{i=1}^{n} m_{X_i}(t) = \prod_{i=1}^{n} \exp(\mu_i t + \frac{\sigma_i^2 t^2}{2}) = \exp\left(\sum_{i=1}^{n} \mu_i t + \frac{\sum_{i=1}^{n} \sigma_i^2 t^2}{2}\right)$$

[예 8-8] 각각의 길이가 N(12.5, 4)를 따르는 4개의 막대를 일렬로 연결했을 때, 전체 막대의 길이가 40과 60 사이일 확률?

$$X \sim N(50,\ 4 \times 4)$$

$$P(40 < X < 60) = P(\frac{40-50}{\sqrt{16}} < Z < \frac{60-50}{\sqrt{16}}) = \Phi(2.5) - \Phi(-2.5)$$

$$\doteq 0.9938 - (1 - 0.9938) = 0.9876$$

```
# pnorm(x, mean, sd) 사용
pnorm(60, 50, 4)-pnorm(40, 50, 4)
[1] 0.9875807
```

18/33

18

2. 이항분포의 정규근사

[정리 8-3] 이항분포의 정규근사

$$X \sim B(n,p) \Rightarrow Z \equiv \frac{X-np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

[예 8-9] 불량률 p=0.2인 공정에서 n=25개의 제품을 검사했을 때, 4개 이하의 불량품이 발견될 확률을 구하시오.

$$X \sim B(25,\, 0.2) \quad \Rightarrow np(1-p) = 25 \times 0.2 \times 0.8 = 4$$

$$P(X \le 4) = \sum_{x=0}^{4} \binom{25}{x} (0.2)^x (0.8)^{25-x} \doteq 0.4207$$

$$P(X \le 4) = P(\frac{X-5}{2} \le \frac{4-5}{2}) = P(Z \le -0.5) \doteq 0.3085$$

- 연속성 보정(continuity correction)

$$P(X \le 4) \simeq P(X < 4.5) = P(Z \le \frac{4.5-5}{2}) = P(Z \le -0.25) \doteq 0.4013$$

19/33

19

2. 이항분포의 정규근사

[예 8-9] 불량률 p=0.2인 공정에서 n=25개의 제품을 검사했을 때, 4개 이하의 불량품이 발견될 확률을 구하시오. (계속)

```
# 정확한 계산 → pbinom(x, n, p) 사용
pbinom(4, 25, 0.2)
[1] 0.4206743

# 정규근사 → pnorm(x, mean, sd) 사용
pnorm(4, 5, 2)
[1] 0.3085375

# 연속성 보정
pnorm(4.5, 5, 2)
[1] 0.4012937
```

20/33

20

2. 이항분포의 정규근사

[예 8-10] 주사위를 100번 굴렸을 때 4 이상 나온 회수가 40개 이상 45개 이하일 확률을 구하시오.

$X \sim B(100, 0.5)$

$$P(40 \le X \le 45) = \sum_{x=40}^{45} \binom{100}{x} (\frac{1}{2})^x (\frac{1}{2})^{100-x} \doteq 0.1665$$

$$P(40 \le X \le 45) \simeq P(39.5 < X < 45.5) = P(\frac{39.5-50}{5} \le Z \le \frac{45.5-50}{5})$$
$$= P(-2.1 \le Z \le -0.9) = [1-\Phi(0.9)] - [1-\Phi(2.1)]$$
$$\doteq (1-0.8159) - (1-0.9821) = 0.1662$$

```
# 정확한 계산
pbinom(45, 100, 0.5)-pbinom(39, 100, 0.5)
[1] 0.1665007
# 정규근사 (연속성 보정)
pnorm(-0.9)-pnorm(-2.1)
[1] 0.1661957
```


3. 카이제곱분포

[정의 8-3] 카이제곱분포(chi-square distributiion)

$$Z_1, Z_2, \dots, Z_n \overset{iid}{\sim} N(0,1) \Rightarrow \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$$

- 확률밀도함수 $f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \; x > 0$

  $\chi^2(\nu) \equiv \Gamma(\alpha = \frac{\nu}{2}, \theta = 2)$
- 모멘트생성함수 $m(t) = (1-\theta t)^{-\alpha} = (1-2t)^{-\nu/2}$
- 기대값 $E(X) = \alpha\theta = (\nu/2) \times 2 = \nu$
- 분산 $Var(X) = \alpha\theta^2 = (\nu/2) \times 2^2 = 2\nu$

3. 카이제곱분포

카이제곱분포의 R 함수

# 카이제곱분포의 확률밀도함수 (df=자유도, ncp=비중심모수(미사용))
dchisq(x, df, ncp = 0)
# Excel 함수 = CHISQ.DIST(x, df, FALSE)

# 카이제곱분포의 누적분포함수 F(x)
pchisq(x, df, ncp = 0, lower.tail = TRUE)
# Excel 함수 = CHISQ.DIST(x, df, TRUE)

# 카이제곱분포의 분위수 (p=누적확률)
qchisq(p, df, ncp = 0, lower.tail = TRUE)
# Excel 함수 = CHISQ.INV(p, df)

# 카이제곱분포의 확률변수 (n=난수의 개수)
rchisq(n, df, ncp = 0)
# Excel 함수 = CHISQ.INV(RAND( ), df) ⇒ 한 개의 난수 생성

23/33

23

3. 카이제곱분포

[정리 8-4] 카이제곱분포의 가법성

$$X_1, X_2, \ldots, X_k \text{ are indep. } \& \ X_i \sim \chi^2(\nu_i)$$
$$\Rightarrow \sum_{i=1}^{k} X_i \sim \chi^2(\nu_1 + \nu_2 + \cdots + \nu_k)$$

[증명] $Y = X_1 + X_2 + \cdots + X_k$ $\qquad E(e^{tX_i}) = (1-\theta t)^{-\nu_i}$

$$m_Y(t) = E(e^{tX_1})E(e^{tX_2})\cdots E(e^{tX_k}) = (1-\theta t)^{-(\nu_1+\nu_2+\cdots+\nu_k)}$$

- 자유도(degree of freedom)
  제곱합에 포함되는 '독립적인 항'의 개수

- 표본분산의 자유도는? $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

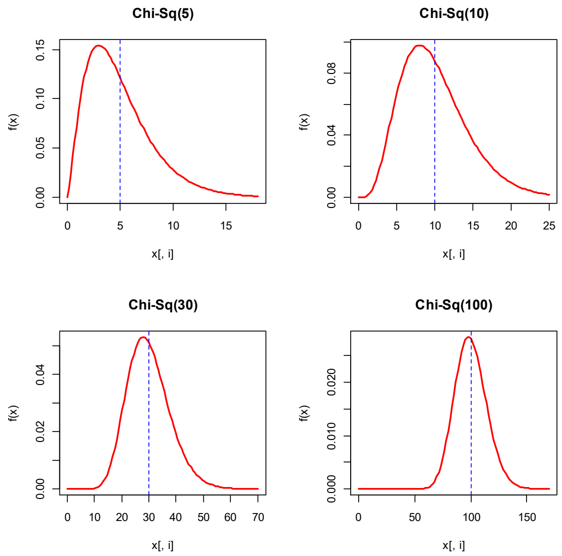
$$\sum_{i=1}^{n}(X_i - \bar{X}) = \sum_{i=1}^{n} X_i - n\bar{X} = 0 \quad \Rightarrow df = n-1$$

24/33

24

[예 8-11] 카이제곱 확률분포의 분위수 [표] 작성 $P(\chi^2 > \chi^2_{1-\alpha;\nu}) = \alpha$

[예 8-12] 자유도 5, 10, 30, 100인 카이제곱분포의 확률밀도함수



25/33

25

3. 카이제곱분포

[예 8-13] 표준정규분포를 따르는 모집단으로부터 추출한 n개의 확률표본

$Z_1, Z_2, \ldots, Z_n \Rightarrow P\left(\sum_{i=1}^{n} Z_i^2 \le a\right) = 0.95$ → 상수 a의 값은?

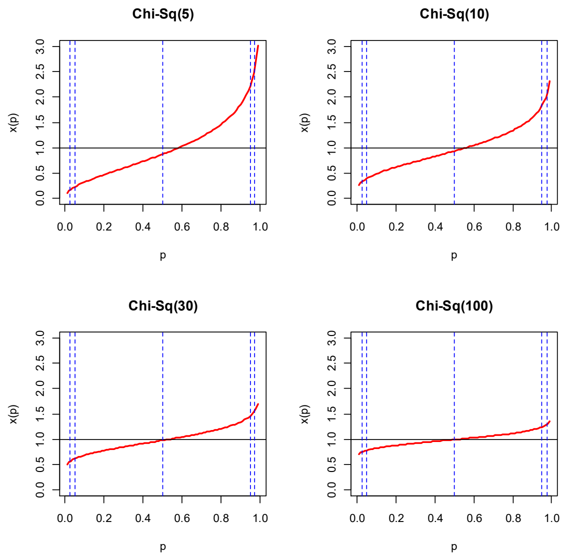
```
# 자유도, 누적확률 입력
nu <- c(5, 10, 30, 100)
p <- 0.95

# 분위수  (상수 a)
qchisq(p, nu)
[1]  11.07050  18.30704  43.77297 124.34211

# 기대값 대비 95% 분위수 비율
qchisq(p, nu) / nu
[1] 2.214100 1.830704 1.459099 1.243421
```

☞ 자유도가 커질수록 기대값 대비 95% 분위수 비율이 1에 가까워지므로, 중심집중 경향이 증가함을 알 수 있음.

26/33

26

[예 8-14] 자유도 5, 10, 30, 100인 카이제곱분포의 분위수/기대값 곡선을 작성하여 비교



27/33

27

4. t-분포

[정의 8-4] t-분포

: 표준정규분포를 따르는 확률변수를 Z라 하고, Z와는 독립적으로 자유도 ν인 카이제곱 분포를 따르는 확률변수를 Y라 하면,

$$T \equiv \frac{Z}{\sqrt{Y/\nu}} \sim t(\nu)$$

- 응용

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\frac{S}{\sigma}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{Z}{\sqrt{Y/(n-1)}} \sim t(n-1)$$

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

28/33

28

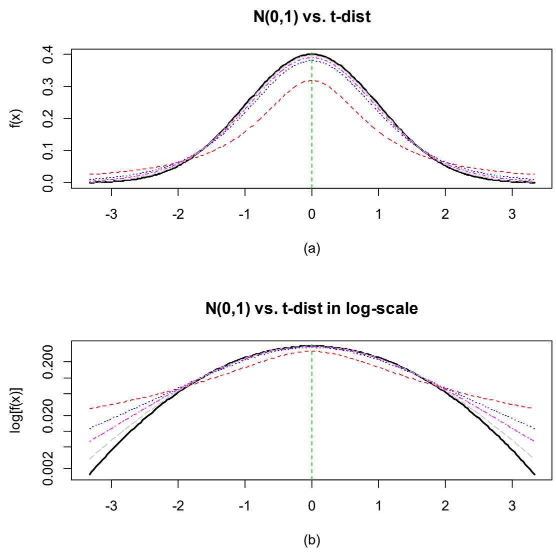
4. t-분포

```
# t-분포의 확률밀도함수 f(x) (df=자유도, ncp=비중심모수(미사용))
dt(x, df, ncp)
# Excel 함수 = T.DIST(x, df, FALSE)

# t-분포의 누적분포함수 F(x)
pt(x, df, ncp, lower.tail = TRUE)
# Excel 함수 = T.DIST(x, df, TRUE)

# t-분포의 분위수 (p=누적확률)
qt(p, df, ncp, lower.tail = TRUE)
# Excel 함수 = T.INV(p, df)

# t-분포의 확률변수 (n=난수의 개수)
rt(n, df, ncp)
# Excel 함수 = T.INV(RAND( ), df) ⇒ 한 개의 난수 생성
```

29/33

29

[예 8-15] 표준정규분포와 자유도 1, 5, 10, 30인 t-분포의 확률밀도함수를 작성하고, 구간 (-1, 1), (-2, 2), (-3, 3)의 확률 비교

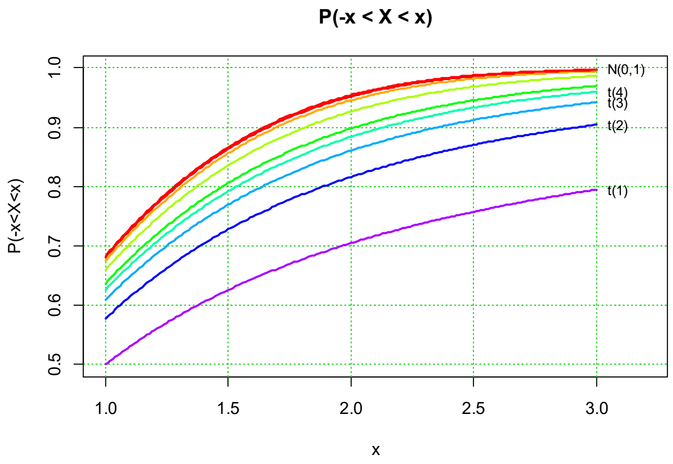


30/33

30

[예 8-16] t-분포 분위수 [표] 작성 $P(T > t_{1-\alpha;\nu}) = \alpha$

[예 8-17] 1에서 3 사이의 x에 대하여 표준정규분포와 t-분포의 중심 집중경향 확률 표시 (ν=1, 2, 3, 4, 5, 10, 30)



5. F-분포

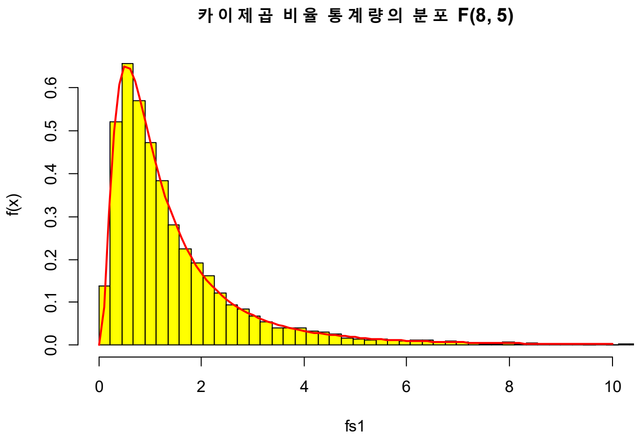
[정의 8-5] F-분포

$$U \sim \chi^2(\nu_1) \perp V \sim \chi^2(\nu_2) \Rightarrow F \equiv \frac{U/\nu_1}{V/\nu_2} \sim F(\nu_1, \nu_2)$$

$$\alpha \equiv P(F_{\nu_1,\nu_2} < F_{\alpha;(\nu_1,\nu_2)}) = P\left(F_{\nu_2,\nu_1} > \frac{1}{F_{\alpha;(\nu_1,\nu_2)}}\right) \equiv P\left(F_{\nu_2,\nu_1} > F_{1-\alpha;(\nu_2,\nu_1)}\right)$$

```
# F-분포의 R 함수
# F-분포의 확률밀도함수 (df1, df2 =자유도, ncp=비중심모수(사용하지 않음))
df(x, df1, df2, ncp)
# Excel 함수 = F.DIST(x, df1, df2, FALSE)
# F-분포의 누적분포함수 F(x)
pf(x, df1, df2, ncp, lower.tail = TRUE)
# Excel 함수 = F.DIST(x, df1, df2, TRUE)
# F-분포의 분위수 (p=누적확률)
qf(p, df1, df2, ncp, lower.tail = TRUE)
# Excel 함수 = F.INV(p, df1, df2)
# F-분포의 확률변수 (n=난수의 개수)
rf(n, df1, df2, ncp)
# Excel 함수 = F.INV(RAND( ), df1, df2)
```

[예 8-18] 각각 자유도 8과 5인 독립적인 카이제곱분포 모집단에서 10,000개씩 표본을 랜덤 추출하여 표준화 비율을 구하여 히스토그램 작성 및 [정의 8-5]에서 제시하는 F(8,5) 분포 확인



[예 8-19] F-분포의 분위수 [표] 작성 $\alpha \equiv P(F_{\nu_1,\nu_2} > F_{1-\alpha;(\nu_1,\nu_2)})$

33/33