4장. 표본 분포

4.1 통계량의 분포

용이성과 근사성 때문에 통계적 추측 방법의 성질을 연구할 때 **흔히 유한모집단에서의 복원추출을 개념화 하여 동일한 모집단을 독립적으로 관측하는 것을 전제로 한다.**

랜덤표본(random sample): 서로 독립이고 동일한 분포를 따르는 확률변수들
모수(母數 population parameter): 모집단 분포들을 모형으로 설정할 때 사용되는 매개변수 **모수공간**(母數空間 parameter space):가능한 모수 전체의 집합

모수는 흔히 θ로, 모수공간은 Ω로, 모집단 분포는 확률밀도함수 $f(x;\theta)$로 나타낸다.

예 4.1.1
(a) 공정의 불량률에 대한 통계 조사의 경우에 모집단분포는 베르누이분포 Bernoulli(p)로나타낼 수 있고, 불량률 p가 모수이고 미지인 불량률의 집합 $\Omega = \{p : 0 \leq p \leq 1\}$가 모수공간이다.
(b) 하천의 오염도를 조사할 때 하천 물의 단위 부피당 부유생물의 수에 대한 모형으로 포아송분포 Poisson(λ)를 설정한다면, 평균 부유생물의 수 λ가 모수이고 미지인 λ의 집합
$\Omega = \{\lambda : \lambda \geq 0\}$가 모수공간이다.
(c) 많은 과학 실험에서 오차의 분포에 대한 모형으로 정규분포 $\mathrm{N}(\mu, \sigma^2)$을 모형으로 설정한다. 이 경우에는 평균과 분산을 나타내는 $\theta = (\mu, \sigma^2)^t$ 가 모수이고, 미지의 평균과 분산의 집합 $\Omega = \{(\mu, \sigma^2)^t : -\infty < \mu < +\infty, \sigma^2 > 0\}$가 모수공간이다.

통계량(統計量 statistic):모집단 분포 또는 모수에 대한 추측을 목적으로 사용하는 랜덤표본의 관측 가능한 함수

랜덤표본과 통계량:
모집단 분포가 확률밀도함수 $f(x;\theta), \theta \in \Omega$ 인 모집단에서의 랜덤표본 $\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_n$이란

$$\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_n \overset{iid}{\sim} f(x;\theta), \theta \in \Omega$$

을 뜻하고, 통계량 $u(\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_n)$이란 랜덤표본의 함수로서 랜덤표본의 값이 주어지면 그 값이 정해지는 함수를 뜻한다.

예 4.1.2

(a) 불량률 p를 모수로 하는 베르누이분포 Bernoulli(p)에서의 랜덤표본을 $X_1, X_2, \cdots, X_n$이라고 할 때, 표본 중 불량 개수의 합계 $X_1 + X_2 + \cdots + X_n$와 표본에서의 불량률을 나타내는 표본비율(標本比率 sample proportion)

$$\hat{p} = (X_1 + X_2 \cdots + X_n)/n$$

은 모두 통계량이다.

(b) 랜덤표본 $X_1, X_2, \cdots, X_n$의 함수로서

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n), \ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

을 각각 표본평균(標本平均 sample mean), 표본분산(標本分散 sample variance)이라고 하며, 모평균과 모분산의 추측에 사용되는 통계량이다.

(c) 모집단 분포가 연속형인 경우에 이 모집단에서의 랜덤표본 $X_1, X_2, \cdots, X_n$을 크기 순서로 늘어놓은 것을 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$이라고 할 때, 이들을 순서통계량(順序統計量 order statistics)이라고 한다. 또한 이들 크기의 가운데를 나타내는 통계량

$$\underset{1 \le i \le n}{med}\ X_i = \begin{cases} X_{(m+1)}, & n = 2m+1\text{일 때} \\ (X_{(m)} + X_{(m+1)})/2, & n = 2m\text{일 때} \end{cases}$$

을 표본중앙값(標本中央값 sample median)이라고 하며, 모집단 분포의 중심부분에 대한 추측에 사용되는 통계량이다.

랜덤표본 $X_1, X_2, \cdots, X_n$의 함수인 통계량 $u(X_1, X_2, \cdots, X_n)$의 분포를 일반적으로 표본분포(標本分布 sampling distribution)라고 한다.

예 4.1.3

(a) 불량률 p를 모수로 하는 베르누이분포 Bernoulli(p)에서의 랜덤표본을 $X_1, X_2, \cdots, X_n$이라고 할 때, 표본 중 불량품의 개수인 $X_1 + \cdots + X_n$은 이항분포 $\mathrm{Bin}(n, p)$를 갖는다. 따라서 표본비율 $\hat{p} = (X_1 + \cdots X_n)/n$은

$$P(\hat{p} = k/n) = \binom{n}{k} p^k (1-p)^{n-k}, \ k = 0, 1, \cdots, n$$

으로 주어지는 분포를 가지며 그 평균과 분산은 각각

$$E(\hat{p}) = p, \ \mathrm{Va}(\hat{p}) = p(1-p)/n$$

이다. 즉 표본비율 $\hat{p}$은 표본 크기가 커짐에 따라 모비율 p 주위에 집중되는 분포를 갖는다.

(b) 단위 부피당 평균 부유생물의 수가 λ인 포아송분포 Poisson(λ)에서의 랜덤표본을 $X_1, X_2, \cdots, X_n$이라고 할 때, 정리 3.4.1로부터 표본 중 부유 생물의 합계 $X_1 + X_2 + \cdots + X_n$은 포아송분포 Poisson$(n\lambda)$를 갖는다. 따라서 표본평균 $\overline{X}$는

$$P(\overline{X} = k/n) = e^{-\lambda}\lambda^k/k!, \quad k = 0, 1, \cdots$$

으로 주어지는 분포를 가지며 그 평균과 분산은 각각

$$E(\overline{X}) = \lambda, \ \mathrm{Var}(\overline{X}) = \lambda/n$$

이다. 즉 표본평균 $\overline{X}$는 표본 크기가 커짐에 따라 모평균 λ주위에 집중되는 분포를 갖는다.

랜덤표본의 함수인 통계량의 분포를 일반적인 경우에 구하는 방법:
확률변수 $X=(X_1,\cdots,X_n)^t$의 분포로부터 $Y=u(X)=u(X_1,\cdots,X_n)$의 분포를 구하려면 Y에 관한 확률을 대응하는 X에 관한 확률로 바꾸어 계산해야 한다.

변수 변환과 확률밀도함수의 기본 관계:
(a) 이산형 경우

$$P(Y=y)=P(u(X)=y)=\sum_{x:u(x)=y} P(X=x)$$

$$\therefore\ pdf_Y(y)=\sum_{x:u(x)=y} pdf_X(x)$$

(b) 연속형 경우

$$P(Y\in y\pm|\Delta y|)=P(u(X)\in y\pm|\Delta y|)\approx\sum_{x:u(x)=y} P(X\in x\pm|\Delta x|)$$

$$\therefore\ pdf_Y(y)|\Delta y|\approx\sum_{x:u(x)=y} pdf_X(x)|\Delta x|,\ \ \text{즉}\ \ pdf_Y(y)\approx\sum_{x:u(x)=y} pdf_X(x)|\Delta y/\Delta x|^{-1}$$

예 4.1.4
(b) 서로 독립이고 각각 포아송분포 $\text{Poisson}(\lambda_1),\text{Poisson}(\lambda_2)$를 따르는 확률변수 X_1,X_2에 대하여 $Y=X_1+X_2$의 확률밀도함수는 다음과 같이 구할 수 있다.

$$\begin{aligned} pdf_Y(y)&=\sum_{x:x_1+x_2=y}\frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!}\frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}\\ &=\left\{\sum_{x_1=0}^{y}\binom{y}{x_1}\lambda_1^{x_1}\lambda_2^{y-x_1}\right\}\frac{e^{-\lambda_1-\lambda_2}}{y!}\\ &=\frac{e^{-\lambda_1-\lambda_2}(\lambda_1+\lambda_2)^y}{y!},\ y=0,1,\cdots \end{aligned}$$

즉 $Y=X_1+X_2$의 분포는 포아송분포 $\text{Poisson}(\lambda_1+\lambda_2)$로서 정리 3.4.1의 결과와 일치한다.

(c) 표준정규분포를 따르는 확률변수 X에 대하여 $Y=X^2$의 확률밀도함수는 누적분포함수를 미분하여 다음과 같이 구할 수 있다.

$$\frac{d}{dy}cdf_Y(y)=\frac{d}{dy}\{P(X\le\sqrt{y})-P(X\le-\sqrt{y})\}=\frac{1}{2}y^{-1/2}\{pdf_X(\sqrt{y})+pdf_X(-\sqrt{y})\}$$

$$\therefore pdf_Y(y)=\frac{d}{dy}cdf_Y(y)=\sum_{x:x^2=y} pdf_X(x)|\frac{dy}{dx}|^{-1}=\frac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2},\ y>0$$

그런데 $\Gamma(1/2)=\sqrt{\pi}$ 이므로, $Y=X^2$의 분포는 $\text{Gamma}(1/2,2)$분포임을 알 수 있다.

--------------------부록 I (p.500):정리 I.6.1 치환적분----------------------

정의역 $\mathcal{Y}$ 가 n차원 공간 R^n에서의 열린집합인 일대일 함수 $\omega:\mathcal{Y} \to \mathcal{X}$ 가 미분가능하고, 1차 편도함수가 연속함수로서 야코비안(Jacobian)행렬식

$$J_\omega(y) = \det(\frac{\partial}{\partial y_i}\omega_j(y))$$

이 정의역에서 0이 아닐 때, 함수 ω의 치역에서 정의된 적분가능한 함수 f에 대하여 다음이 성립한다.

$$\int_{\mathrm{A}} f(x)dx = \int_{\omega^{-1}(\mathrm{A})} f(w(y))|J_\omega(y)|dy$$

정리 4.1.1: 연속형 변수의 일대일 변환과 확률밀도함수

연속형의 k차원 확률변수 $\mathrm{X} = (\mathrm{X}_1,\cdots,\mathrm{X}_k)^t$와 함수 $u = (u_1,\cdots,u_k)^t : \mathcal{X} \to \mathcal{Y}$ 에 대하여 다음이 성립한다고 하자.

(a) $\mathrm{P}(\mathrm{X} \in \mathcal{X}) = 1$

(b) 벡터값 함수 $u = (u_1,\cdots,u_k)^t : \mathcal{X} \to \mathcal{Y}$ 는 정의역이 $\mathcal{X}$ 이고 치역이 $\mathcal{Y}$ 인 일대일 함수이다.

(c) 함수 $u = (u_1,\cdots,u_k)^t$의 정의역 $\mathcal{X}$ 는 k차원 공간 R^k에서 열린집합이고, 함수 u가 미분가능하며 1차 편도함수가 연속함수로서 0이 아닌 야코비안(Jacobian)행렬식을 갖는다. 즉

$$J_u(x) = \det(\frac{\partial}{\partial x_i}u_j(x)) \neq 0 \quad \forall x \in \mathcal{X}$$

이 때, 확률변수 $\mathrm{X} = (\mathrm{X}_1,\cdots,\mathrm{X}_k)^t$와 함수 u를 이용하여 정의된 k차원 확률변수

$$\mathrm{Y} = u(\mathrm{X}),\ \text{즉} \quad \mathrm{Y} = (\mathrm{Y}_1,\cdots,\mathrm{Y}_k)^t = (u_1(\mathrm{X}),\cdots,u_k(\mathrm{X}))^t$$

의 확률밀도함수는 다음과 같이 주어진다.

$$pdf_{\mathrm{Y}}(y) = pdf_{\mathrm{X}}(u^{-1}(y))|J_{u^{-1}}(y)|,\ y \in \mathcal{Y}$$

즉

$$pdf_{\mathrm{Y}}(y) = pdf_{\mathrm{X}}(x)|\det(\frac{\partial y}{\partial x})|^{-1},\ y = u(x) \in \mathcal{Y}$$

[증명] 부록 I의 치환적분법으로부터 $\mathcal{Y}$ 의 임의의 부분 집합 B에 대하여 다음이 성립함을 알 수 있다. 따라서 $\mathrm{Y} = u(\mathrm{X})$의 확률밀도함수가 위에서와 같이 주어진다.

$$\begin{aligned}\mathrm{P}(\mathrm{Y} \in \mathrm{B}) &= \mathrm{P}(u(\mathrm{X}) \in \mathrm{B}) \\ &= \int_{x \in u^{-1}(B)} pdf_{\mathrm{X}}(x)dx \\ &= \int_{y \in \mathrm{B}} pdf_X(u^{-1}(y))|J_{u^{-1}}(y)|dy \quad (\because x = u^{-1}(y)\text{로 치환})\end{aligned}$$

예 4.1.5 위치모수와 척도모수를 이용한 확률분포:
연속형 확률변수 Z의 확률밀도 함수가 $f(z)$일 때, 양수 σ와 실수 μ에 대하여

$$\mathrm{X} = \sigma \mathrm{Z} + \mu$$

로 정의된 확률변수 X의 확률밀도함수는

$$pdf_{\mathrm{X}}(x) = f(z)\left|\frac{dx}{dz}\right|^{-1}, x = \sigma z + \mu$$
$$= \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

로 주어진다. 이러한 꼴의 확률밀도함수에서 μ와 σ를 각각 위치모수(位置母數 location parameter)와 척도모수(尺度母數 scale parameter)라고 한다.

(a) 정규분포 $\mathrm{N}(\mu,\sigma^2)$

$$pdf_{\mathrm{X}}(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right),\ \ \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \mathrm{I}_{(-\infty < z < +\infty)}$$

$$\mathrm{X} \sim \mathrm{N}(\mu,\sigma^2) \Leftrightarrow \mathrm{X} \overset{d}{\equiv} \sigma \mathrm{Z} + \mu,\ \mathrm{Z} \sim \mathrm{N}(0,1)$$

(b) 로지스틱(logistic)분포 $\mathrm{L}(\mu,\sigma)$

$$pdf_{\mathrm{X}}(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right),\ \ f(z) = \frac{e^z}{(1+e^z)^2} \mathrm{I}_{(-\infty < z < +\infty)}$$

$$\mathrm{X} \sim \mathrm{L}(\mu,\sigma) \Leftrightarrow \mathrm{X} \overset{d}{\equiv} \sigma \mathrm{Z} + \mu,\ \mathrm{Z} \sim \mathrm{L}(0,1)$$

(c) 이중지수(double exponential)분포 $\mathrm{DE}(\mu,\sigma)$

$$pdf_{\mathrm{X}}(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right),\ \ f(z) = \frac{1}{2} e^{-|z|} \mathrm{I}_{(-\infty < z < +\infty)}$$

$$\mathrm{X} \sim \mathrm{DE}(\mu,\sigma) \Leftrightarrow \mathrm{X} \overset{d}{\equiv} \sigma \mathrm{Z} + \mu,\ \mathrm{Z} \sim \mathrm{DE}(0,1)$$

(d) 코쉬(Cauchy)분포 $\mathrm{C}(\mu,\sigma)$

$$pdf_{\mathrm{X}}(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right),\ \ f(z) = \frac{1}{\pi}\frac{1}{1+z^2} \mathrm{I}_{(-\infty < z < +\infty)}$$

$$\mathrm{X} \sim \mathrm{C}(\mu,\sigma) \Leftrightarrow \mathrm{X} \overset{d}{\equiv} \sigma \mathrm{Z} + \mu,\ \mathrm{Z} \sim \mathrm{C}(0,1)$$

(e) 지수분포 $\mathrm{Exp}(\sigma)$

$$pdf_{\mathrm{X}}(x) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right),\ \ f(z) = e^{-z} \mathrm{I}_{(0 \le z < +\infty)}$$

$$\mathrm{X} \sim \mathrm{Exp}(\sigma) \Leftrightarrow \mathrm{X} \overset{d}{\equiv} \sigma \mathrm{Z},\ \mathrm{Z} \sim \mathrm{Exp}(1)$$

(f) 감마분포 $\mathrm{Gamma}(\alpha,\beta)$

$$pdf_{\mathrm{X}}(x) = \frac{1}{\beta} f\left(\frac{x}{\beta}\right),\ \ f(z) = \frac{1}{\Gamma(\alpha)} z^{\alpha-1} e^{-z} \mathrm{I}_{(0 \le z < +\infty)}$$

$$\mathrm{X} \sim \mathrm{Gamma}(\alpha,\beta) \Leftrightarrow \mathrm{X} \overset{d}{\equiv} \beta \mathrm{Z},\ \mathrm{Z} \sim \mathrm{Gamma}(\alpha,1)$$

예 4.1.6 :서로 독립이고 각각 감마분포 $\mathrm{Gamma}(\alpha_1,\beta)$, $\mathrm{Gamma}(\alpha_2,\beta)$를 따르는 확률변수 $\mathrm{X}_1,\mathrm{X}_2$에 대하여

$$\mathrm{Y}_1 = \frac{\mathrm{X}_1}{\mathrm{X}_1+\mathrm{X}_2},\ \ \mathrm{Y}_2 = \mathrm{X}_1+\mathrm{X}_2$$

라고 할 때, $\mathrm{Y}_1,\mathrm{Y}_2$의 결합확률밀도함수와 각각의 주변확률밀도함수를 구하여라.

[풀이] $\mathrm{X}_1,\mathrm{X}_2$의 결합확률밀도함수는

$$pdf_{\mathrm{X}_1,\mathrm{X}_2}(x_1,x_2) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} x_1^{\alpha_1-1}x_2^{\alpha_2-1}e^{-(x_1+x_2)/\beta}\mathrm{I}_{(x_1>0,\,x_2>0)}$$

----정의역이 $\mathcal{X} = \left\{(x_1,x_2)^t : pdf_{\mathrm{X}_1,\mathrm{X}_2}(x_1,x_2)>0\right\} = \left\{(x_1,x_2)^t : x_1>0, x_2>0\right\}$

인 함수 $u(x_1,x_2) = (\frac{x_1}{x_1+x_2}, x_1+x_2)^t$에 대하여 정리 4.1.1의 조건을 확인-----

$$u : \begin{cases} y_1 = \dfrac{x_1}{x_1+x_2} \\ y_2 = x_1+x_2 \end{cases} \qquad u^{-1} : \begin{cases} x_1 = y_1y_2 \\ x_2 = y_2(1-y_1) \end{cases}$$

$$\mathcal{Y} = \left\{(y_1,y_2)^t : y_1y_2>0,\ y_2(1-y_1)>0\right\} = \left\{(y_1,y_2)^t : 0<y_1<1, y_2>0\right\}$$

함수 u는 $\mathcal{X}$에서 $\mathcal{Y}$로의 일대일 함수로서 정리 4.1.1의 조건을 만족시키고

$$J_{u^{-1}} = \det\begin{pmatrix} y_2 & -y_2 \\ y_1 & 1-y_1 \end{pmatrix} = y_2$$

따라서 정리 4.1.1로부터 $\mathrm{Y}_1,\mathrm{Y}_2$의 결합확률밀도함수는

$$\begin{aligned} pdf_{\mathrm{Y}_1,\mathrm{Y}_2}(y_1,y_2) &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}}(y_1y_2)^{\alpha_1-1}(y_2(1-y_1))^{\alpha_2-1}e^{-y_2/\beta}\mathrm{I}_{(0,1)}(y_1)\mathrm{I}_{(0,\infty)}(y_2)|y_2| \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1}\mathrm{I}_{(0,1)}(y_1)y_2^{\alpha_1+\alpha_2-1}e^{-y_2/\beta}\mathrm{I}_{(0,\infty)}(y_2) \end{aligned}$$

로 주어진다. 이는 y_1만의 함수와 y_2만의 함수의 곱이므로 $\mathrm{Y}_1,\mathrm{Y}_2$는 서로 독립이고 각각의 주변확률밀도함수는 다음과 같이 주어진다.

$$pdf_{\mathrm{Y}_1}(y_1) = \int_{-\infty}^{+\infty} pdf_{\mathrm{Y}_1,\mathrm{Y}_2}(y_1,y_2)dy_2 = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1}\mathrm{I}_{(0,1)}(y_1)$$

$$pdf_{\mathrm{Y}_2}(y_2) = \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}} y_2^{\alpha_1+\alpha_2-1}e^{-y_2/\beta}\mathrm{I}_{(0,\infty)}(y_2)$$

베타분포[1]의 정의:

$$\mathrm{X} \sim \mathrm{Beta}(\alpha_1,\alpha_2) \Leftrightarrow \mathrm{X} \overset{d}{\equiv} \mathrm{Z}_1/(\mathrm{Z}_1+\mathrm{Z}_2),\ \mathrm{Z}_i \sim \mathrm{Gamma}(\alpha_i,\beta)(i=1,2)\text{이고 서로 독립}$$

$$\Leftrightarrow pdf_{\mathrm{X}}(x) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1}(1-x)^{\alpha_2-1}\mathrm{I}_{(0,1)}(x)$$

----

1) 예 4.1.6에서 Y_1의 확률밀도함수의 적분 값은 1이므로, 양수 α_1,α_2에 대하여

$$\int_0^1 x^{\alpha_1-1}(1-x)^{\alpha_2-1}dx = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$$

임을 알 수 있다. 이러한 정적분으로 정의된 함수를 베타함수라고 하며 $B(\alpha_1,\alpha_2)$로 나타내기도 한다.

예 4.1.7

확률변수 X_1, X_2이 서로 독립이고 동일한 확률밀도함수

$$f(x) = I_{(0,1)}(x)$$

를 가질 때, $Y = (X_1 + X_2)/2$의 확률밀도함수를 구하여라.

[풀이] 이차원 확률변수 사이의 일대일 변환을 생각하기 위하여

$$Y = (X_1 + X_2)/2, \ \underline{Z = (X_1 - X_2)/2}$$

라고 하면, 다음과 같이 함수 u와 그 정의역 $\mathcal{X}$, 치역 $\mathcal{Y}$, 역함수 u^{-1}를 생각할 수 있다.

$$u : \begin{cases} y = (x_1 + x_2)/2 \\ z = (x_1 - x_2)/2 \end{cases} \qquad u^{-1} : \begin{cases} x_1 = y + z \\ x_2 = y - z \end{cases}$$

$$\mathcal{X} = \left\{ (x_1, x_2)^t : pdf_{X_1, X_2}(x_1, x_2) > 0 \right\} = \left\{ (x_1, x_2)^t : 0 < x_1 < 1, 0 < x_2 < 1 \right\}$$

$$\mathcal{Y} = \left\{ (y, z)^t : 0 < y + z < 1, 0 < y - z < 1 \right\}$$

이로부터 함수 u는 $\mathcal{X}$ 에서 $\mathcal{Y}$ 로의 일대일 함수로서 정리 4.1.1의 조건을 만족시키고 그 역함수 u^{-1}의 야코비안이 다음과 같이 주어지는 것을 알 수 있다.

$$J_{u^{-1}} = \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = -2$$

따라서 정리 4.1.1로부터 Y,Z의 결합확률밀도함수는

$$pdf_{Y,Z}(y,z) = I_{(0 < y+z < 1, 0 < y-z < 1)} |-2|$$

이로부터 Y의 주변확률밀도함수를 구하기 위하여, 지표함수의 연립부등식을 $\underline{z\text{에 대한 부등식으로}}$ 나타내면

$$-y < z < 1 - y, \ y - 1 < z < y$$

즉 $a \vee b = \max(a,b), a \wedge b = \min(a,b)$이라고 하면

$$(-y) \vee (y-1) < z < (1-y) \wedge y$$

그러므로 $y \le 1/2$인 경우와 $y > 1/2$인 경우로 나누어 Y의 주변확률밀도함수를 구하면 다음과 같다.

(i) $y \le 1/2$인 경우에, $(-y) \vee (y-1) = -y$, $(1-y) \wedge y = y$ 이고 $-y < y \Leftrightarrow y > 0$ 이므로

$$pdf_Y(y) = \int_{-\infty}^{+\infty} pdf_{Y,Z}(y,z) dz = \int_{-y}^{y} 2dz \, I_{(0,1/2]}(y) = 4y \, I_{(0,1/2]}(y)$$

(ii) $y > 1/2$인 경우에, $(-y) \vee (y-1) = y-1, (1-y) \wedge y = 1-y$이고 $y - 1 < 1 - y \Leftrightarrow y < 1$이므로

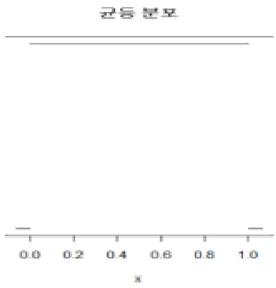
$$pdf_Y(y) = \int_{-\infty}^{+\infty} pdf_{Y,Z}(y,z) dz = \int_{y-1}^{1-y} 2dz \, I_{(1/2,1)}(y) = 4(1-y) I_{(1/2,1)}(y)$$

따라서 Y의 주변확률밀도함수를 하나의 식으로 나타내면 예 1.4.3에서와 같이

$$pdf_Y(y) = (2 - 4|y - 1/2|) I_{(0 \ \ 1)}(y)$$

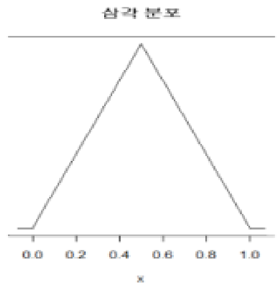
임을 알 수 있다.

예 4.1.7에서 서로 독립인 X_1, X_2의 공통 분포를 구간 (0,1)에서의 균등분포(均等分布 uniform distribution)라고 하며, 이들의 평균인 $Y = (X_1 + X_2)/2$의 분포를 구간 (0,1)에서의 삼각분포(三角分布 triangular distribution)라고 한다.



$f(x) = \mathrm{I}_{(0,\,1)}(x)$

그림 4.1.1 균등분포



$f(y) = (2 - 4|y - 1/2|)\mathrm{I}_{(0,\,1)}(y)$

그림 4.1.2 삼각분포

한편, 일반적인 구간 (a, b)에서의 균등분포는 다음과 같이 정의한다.

균등분포의 정의:

$$\begin{aligned} \mathrm{X} \sim \mathrm{U}(a,b) &\Leftrightarrow pdf_{\mathrm{X}}(x) = \frac{1}{b-a}\mathrm{I}_{(\mathrm{a,b})}(\mathrm{x}) \\ &\Leftrightarrow \mathrm{X} \overset{d}{\equiv} (b-a)\mathrm{Z} + a,\ \mathrm{Z} \sim \mathrm{U}(0,1) \end{aligned}$$

예 4.1.8
서로 독립이고 표준정규분포 $\mathrm{N}(0,1)$를 따르는 확률변수 X,Y에 대하여, 극좌표 변환인

$$\begin{cases} \mathrm{X} = \mathrm{R}\cos\Theta \\ \mathrm{Y} = \mathrm{R}\sin\Theta, \quad 0 \le \mathrm{R} < +\infty,\ 0 \le \Theta < 2\pi \end{cases}$$

를 만족시키는 R,Θ의 결합확률밀도함수를 구하여라.

[풀이] X,Y 결합확률밀도함수는

$$pdf_{\mathrm{X,Y}}(x,y) = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}\mathrm{I}_{(-\infty<x<+\infty,\ -\infty<y<+\infty)}$$

이고, 주어진 극좌표 변환

$$w : \begin{cases} x = r\cos\theta \\ y = r\sin\theta, \quad 0 \le r < +\infty,\ 0 \le \theta < 2\pi \end{cases}$$

는 정의역이

$$\mathcal{Y} = \{(r,\theta) : 0 \le r < +\infty,\ 0 \le \theta < 2\pi\}$$

이고 치역이

$$\mathcal{X} = \{(x,y) : -\infty < x < +\infty, -\infty < y < +\infty\}$$

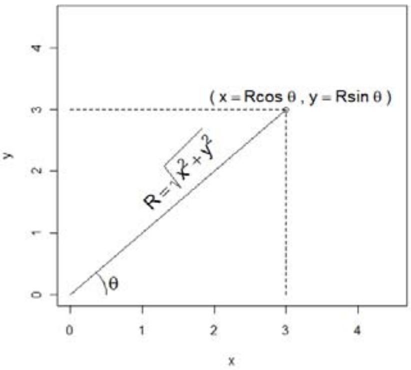


그림 4.1.3 직교좌표와 극좌표

인 일대일 함수로서, 미분가능하고 그 야코비안이

$$J_w(y) = \det\begin{pmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial y}{\partial r} \\ \dfrac{\partial x}{\partial \theta} & \dfrac{\partial y}{\partial \theta} \end{pmatrix} = \det\begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix} = r$$

임을 알 수 있다. 따라서 극좌표 변환 w의 역함수를 $u = w^{-1}$라고 하면 역함수 정리로부터 함수 $u(x,y)$가 정리 4.1.1의 조건을 만족시키는 것을 알 수 있다.

그러므로 정리 4.1.1로부터 R,Θ의 결합확률밀도함수는

$$\begin{aligned} pdf_{\mathrm{R},\Theta}(r,\theta) &= \frac{1}{2\pi}e^{-\frac{1}{2}r^2(\cos^2\theta+\sin^2\theta)}\mathrm{I}_{[0,+\infty)}(r)\mathrm{I}_{[0,2\pi)}(\theta)|r| \\ &= re^{-\frac{1}{2}r^2}\mathrm{I}_{[0,+\infty)}(r)\frac{1}{2\pi}\mathrm{I}_{[0,2\pi)}(\theta) \end{aligned}$$

로 주어진다. 즉 R,Θ는 서로 독립이고

$$\mathrm{R}^2/2 \sim \mathrm{Exp}(1),\ \ \Theta \sim \mathrm{U}(0,2\pi)$$

예 4.1.9

서로 독립이고 각각 감마분포 Gamma(α_i,β)를 따르는 확률변수 $X_i(i=1,\cdots,k+1)$들에 대하여

$$Y_1=\frac{X_1}{X_1+\cdots+X_{k+1}},\ \cdots\cdots, Y_k=\frac{X_k}{X_1+\cdots+X_{k+1}}$$

라고 할 때, $Y_1,\cdots,Y_k$의 결합확률밀도함수를 구하여라.

[풀이] $X_1,\cdots,X_{k+1}$의 결합확률밀도함수는

$$pdf_{1,\cdots,k+1}(x_1,\cdots,x_{k+1})=\prod_{i=1}^{k+1}\left\{\frac{1}{\Gamma(\alpha_i)\beta^{\alpha_i}}x_i^{\alpha_i-1}e^{-x_i/\beta}I_{(0,+\infty)}(x_i)\right\}$$

이고, $(k+1)$개 확률변수 사이의 일대일 변환을 생각하기 위하여 다음 변수를 생각해보자.

$$Y_{k+1}=X_1+\cdots+X_{k+1}$$

$$u:\begin{cases}y_1=x_1/(x_1+\cdots+x_{k+1})\\ \quad\vdots\\ y_k=x_k/(x_1+\cdots+x_{k+1})\\ y_{k+1}=x_1+\cdots+x_{k+1}\end{cases}\qquad u^{-1}:\begin{cases}x_1=y_1y_{k+1}\\ \quad\vdots\\ x_k=y_ky_{k+1}\\ x_{k+1}=y_{k+1}(1-y_1-\cdots-y_k)\end{cases}$$

$$\boldsymbol{X}=\left\{(x_1,\cdots,x_{k+1})^t : x_i>0\,(i=1,\cdots,k+1)\right\}$$

$$\boldsymbol{Y}=\left\{(y_1,\cdots,y_k,y_{k+1})^t : y_i>0\,(i=1,\cdots,k), y_1+\cdots+y_k<1,\ y_{k+1}>0\right\}$$

함수 u는 $\boldsymbol{X}$ 에서 $\boldsymbol{Y}$ 로의 일대일 함수로서 정리 4.1.1의 조건을 만족시키고

$$J_{u^{-1}}=\det\begin{pmatrix}y_{k+1} & \cdots & 0 & -y_{k+1}\\ \vdots & \vdots & \vdots & -y_{k+1}\\ 0 & \cdots & y_{k+1} & -y_{k+1}\\ y_1 & \cdots & y_k & (1-y_1-\cdots-y_k)\end{pmatrix}=y_{k+1}^k$$

따라서 $Y_1,\cdots,Y_k,Y_{k+1}$의 결합확률밀도함수는

$$pdf_{Y_1,\cdots,Y_k,Y_{k+1}}(y_1,\cdots,y_k,y_{k+1})=\left\{\prod_{i=1}^{k+1}\frac{1}{\Gamma(\alpha_i)\beta^{\alpha_i}}\right\}\times\left\{\prod_{i=1}^{k}(y_iy_{k+1})^{\alpha_i-1}\right\}\times\left\{y_{k+1}(1-y_1-\cdots-y_k)\right\}^{\alpha_{k+1}-1}\times\Big\}$$

$$\times e^{-y_{k+1}/\beta}|y_{k+1}^k|,\ (y_1,\cdots,y_{k+1})^t\in\boldsymbol{Y}$$

$$\therefore\ pdf_{Y_1,\cdots,Y_k,Y_{k+1}}=\left\{\prod_{i=1}^{k}\frac{1}{\Gamma(\alpha_i)}y_i^{\alpha_i-1}\right\}(1-y_1-\cdots y_k)^{\alpha_{k+1}-1}I_{(y_1>0,\cdots,y_k>0,y_1+\cdots+y_k<1)}\times$$
$$\times\left\{\frac{1}{\Gamma(\alpha_{k+1})\beta^{\alpha_1+\cdots+\alpha_{k+1}}}y_{k+1}^{\alpha_1+\cdots+\alpha_{k+1}-1}e^{-y_{k+1}/\beta}I_{(y_{k+1}>0)}\right\}$$

그런데

$$\int_0^{+\infty}\frac{1}{\beta^{\alpha_1+\cdots+\alpha_{k+1}}}y_{k+1}^{\alpha_1+\cdots\,\alpha_{k+1}-1}e^{-y_{k+1}/\beta}dy_{k+1}=\Gamma(\alpha_1+\cdots+\alpha_{k+1})$$

이므로 $Y_1,\cdots,Y_k,Y_{k+1}$의 결합확률밀도함수를 y_{k+1}에 대하여 적분하면

$$\begin{aligned}pdf_{Y_1,\cdots,Y_k}(y_1,\cdots,y_k)&=\int_{-\infty}^{+\infty}pdf_{Y_1,\cdots,Y_k,Y_{k+1}}(y_1,\cdots,y_k,y_{k+1})\,dy_{k+1}\\&=\frac{\Gamma(\alpha_1+\cdots+\alpha_{k+1})}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{k+1})}\left\{\prod_{i=1}^{k}y_i^{\alpha_i-1}\right\}(1-y_1-\cdots y_k)^{\alpha_{k+1}-1}I_{(y_1>0,\cdots,y_k>0,y_1+\cdots+y_k<1)}\end{aligned}$$

디리클레분포의 정의:

$Y=(Y_1,\cdots,Y_k)^t\sim$ Dirichlet$(\alpha_1,\cdots,\alpha_k,\alpha_{k+1})$

$\Leftrightarrow Y\overset{d}{\equiv}\left(\frac{X_1}{X_1+\cdots+X_{k+1}},\cdots,\frac{X_k}{X_1+\cdots+X_{k+1}}\right)$, $X_i\sim$ Gamma(α_i,β)이고 서로독립$(i=1,\cdots,k+1)\Leftrightarrow$

$$pdf_{Y_1,\cdots,Y_k}(y_1,\cdots,y_k)=\frac{\Gamma(\alpha_1+\cdots+\alpha_{k+1})}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{k+1})}\left\{\prod_{i=1}^{k}y_i^{\alpha_i-1}\right\}(1-y_1-\cdots y_k)^{\alpha_{k+1}-1}I_{(y_1>0,\cdots,y_k>0,y_1+\cdots+y_k<1)}$$

정리 4.1.2: 연속형 변수의 다대일 변환과 확률밀도함수

연속형의 k차원 확률변수 $\mathrm{X}=(\mathrm{X}_1,\cdots,\mathrm{X}_k)^t$와 함수 $u=(u_1,\cdots,u_k)^t : \mathcal{X} \to \mathcal{Y}$ 에 대하여 다음의 조건이 성립한다고 하자.

(a) $\mathrm{P}(\mathrm{X}\in\mathcal{X})=1$

(b) 벡터값 함수 $u=(u_1,\cdots,u_k)^t : \mathcal{X} \to \mathcal{Y}$ 는 정의역이 $\mathcal{X}$ 이고 치역이 $\mathcal{Y}$ 인 m대일 함수이다.

(c) 함수 $u=(u_1,\cdots,u_k)^t$의 정의역 $\mathcal{X}$ 는 서로 공통 부분이 없는 열린집합 $\mathcal{X}_1,\cdots,\mathcal{X}_m$의 합집합으로 나타낼 수 있고, 함수 u에서 정의역을 $\mathcal{X}_r\ (r=1,\cdots,m)$로 제한한 함수

$$u^r(x)=u(x),\ x\in\mathcal{X}_r \quad (r=1,\cdots,m)$$

는 각각 $\mathcal{X}_r$에서 $\mathcal{Y}$ 로의 일대일 함수로서 미분가능하며 1차 편도함수가 연속함수로서 0이 아닌 야코비안(Jacobian) 행렬식을 갖는다. 즉 $u^r(x)=(u_1^r(x),\cdots,u_k^r(x))^t$라고 할 때

$$J_{u^r}(x)=\det\left(\frac{\partial}{\partial x_i}u_j^r(x)\right)\neq 0 \quad \forall x\in\mathcal{X}_r$$

이러한 조건하에서 확률변수 $\mathrm{X}=(\mathrm{X}_1,\cdots,\mathrm{X}_k)^t$와 함수 u를 이용하여 정의된 k차원 확률변수

$$\mathrm{Y}=u(\mathrm{X}),\ \text{즉}\quad \mathrm{Y}=(\mathrm{Y}_1,\cdots,\mathrm{Y}_k)^t=(u_1(\mathrm{X}),\cdots,u_k(\mathrm{X}))^t$$

의 확률밀도함수는 다음과 같이 주어진다.

$$pdf_{\mathrm{Y}}(y)=\sum_{x:u(x)=y} pdf_{\mathrm{X}}(x)\left|\det\left(\frac{\partial u(x)}{\partial x}\right)\right|^{-1},\ y\in\mathcal{Y}$$

[증명] 정리 4.1.1의 증명에서와 같이 분할된 각각의 정의역 $\mathcal{X}_r\ (r=1,\cdots,m)$에서의 일대일 함수 $y=u^r(x)\ (r=1,\cdots,m)$의 역함수 $x=(u^r)^{-1}(y)$를 이용한 치환적분법을 적용하면 $\mathcal{Y}$ 의 임의의 부분 집합 B에 대하여 다음이 성립함을 알 수 있다.

$$\begin{aligned}\mathrm{P}(\mathrm{Y}\in\mathrm{B})&=\mathrm{P}(u(\mathrm{X})\in\mathrm{B})\\&=\sum_{r=1}^m \mathrm{P}(u^r(\mathrm{X})\in\mathrm{B},\mathrm{X}\in\mathcal{X}_r)\\&=\sum_{r=1}^m\int_{x\in(u^r)^{-1}(\mathrm{B})} pdf_{\mathrm{X}}(x)dx\\&=\int_{y\in\mathrm{B}}\sum_{r=1}^m pdf_X((u^r)^{-1}(y))|J_{(u^r)^{-1}}(y)|\,dy \quad (\because x=(u^r)^{-1}(y))\end{aligned}$$

$$\therefore\ pdf_{\mathrm{Y}}(y)=\sum_{r=1}^m pdf_{\mathrm{X}}((u^r)^{-1}(y))|J_{(u^r)^{-1}}(y)|,\ y\in\mathcal{Y}$$

따라서
$$pdf_{\mathrm{Y}}(y)=\sum_{x:u(x)=y} pdf_{\mathrm{X}}(x)\left|\det\left(\frac{\partial u(x)}{\partial x}\right)\right|^{-1},\ y\in\mathcal{Y}$$

예 4.1.10

(-1, 1)에서 균등분포를 따르는 확률변수 X에 대하여 $\mathrm{Y}=\mathrm{X}^2$의 확률밀도함수를 구하여라.

[풀이] 함수 $y=x^2$은 $\mathcal{X}=(-1,0)\cup(0,1)$에서 $\mathcal{Y}=(0,1)$로의 이대일 함수이고, 정리 4.1.2의 조건을 만족시킨다. 따라서

$$\begin{aligned} pdf_{\mathrm{Y}}(y) &= \sum_{x:x^2=y} pdf_{\mathrm{X}}(x)|dx^2/dx|^{-1}\mathrm{I}_{(0,1)}(y) \\ &= \sum_{x:x^2=y}\frac{1}{2}|2x|^{-1}\mathrm{I}_{(0,1)}(y) \\ &= \frac{1}{2\sqrt{y}}\mathrm{I}_{(0,1)} \end{aligned}$$

예 4.1.11

확률변수 $\mathrm{X}=(\mathrm{X}_1,\mathrm{X}_2)^t$의 결합확률밀도함수가

$$pdf_{\mathrm{X}}(x_1,x_2)=\frac{1}{\pi}\mathrm{I}_{(0<x_1^2+x_2^2<1)}$$

일 때, 다음과 같이 정의된 $\mathrm{Y}=(\mathrm{Y}_1,\mathrm{Y}_2)^t$의 결합확률밀도함수를 구하여라.

$$\mathrm{Y}_1=\mathrm{X}_1^2+\mathrm{X}_2^2,\quad \mathrm{Y}_2=\frac{\mathrm{X}_1^2}{\mathrm{X}_1^2+\mathrm{X}_2^2}$$

[풀이] 함수

$$\begin{cases} y_1=x_1^2+x_2^2 \\ y_2=\dfrac{x_1^2}{x_1^2+x_2^2} \end{cases}$$

의 표현에서 x_1,x_2를 y_1,y_2로 나타내면

$$\begin{cases} x_1^2=y_1y_2 \\ x_2^2=y_1-y_1y_2 \end{cases} \qquad \therefore \begin{cases} x_1=\pm\sqrt{y_1y_2} \\ x_2=\pm\sqrt{y_1(1-y_2)} \end{cases}$$

따라서 위에서 정의된 함수는 다음과 같이 정의된 $\mathcal{X}$ 에서 $\mathcal{Y}$ 로의 사대일 함수이다.

$$\mathcal{X}=\left\{(x_1,x_2)^t: 0<x_1^2+x_2^2<1, x_1\neq 0, x_2\neq 0\right\},\ \mathcal{Y}=\left\{(y_1,y_2)^t: 0<y_1<1, 0<y_2<1\right\}$$

또한, 이 함수의 야코비안 행렬식의 절대값은

$$|\det(\frac{\partial y}{\partial x})|=|\det\begin{pmatrix} 2x_1 & 2x_2 \\ \dfrac{2x_1x_2^2}{(x_1^2+x_2^2)^2} & \dfrac{-2x_1^2x_2}{(x_1^2+x_2^2)^2}\end{pmatrix}^t|=|\frac{-4x_1x_2}{x_1^2+x_2^2}|=4\sqrt{y_2(1-y_2)}$$

$$\begin{aligned} \therefore pdf_{\mathrm{Y}}(y_1,y_2) &= \sum_{\substack{x:x_1^2+x_2^2=y_1 \\ x_1^2/(x_1^2+x_2^2)=y_2}} \frac{1}{\pi}|\det(\frac{\partial y}{\partial x})|^{-1}\mathrm{I}_{(0<y_1<1,0<y_2<1)} \\ &= \frac{4}{\pi}(4\sqrt{y_2(1-y_2)})^{-1}\mathrm{I}_{(0,1)}(y_1)\mathrm{I}_{(0,1)}(y_2) \\ &= \mathrm{I}_{(0,1)}(y_1)\frac{1}{\pi}y_2^{-1/2}(1-y_2)^{-1/2}\mathrm{I}_{(0,1)}(y_2) \end{aligned}$$

그런데 $\Gamma(1/2)=\sqrt{\pi}$. 이므로 $\mathrm{Y}_1,\mathrm{Y}_2$는 서로 독립이고 $\mathrm{Y}_1\sim\mathrm{U}(0,1), \mathrm{Y}_2\sim\mathrm{Beta}(1/2,1/2)$ 임을 알 수 있다.

4.2 대표적인 표본분포

예 4.2.1

서로 독립이고 표준정규분포 $\mathrm{N}(0,1)$를 따르는 확률변수들 $X_1,\cdots,X_r$에 대하여

$$Y = X_1^2 + \cdots + X_r^2$$

의 확률밀도함수를 구하여라.

[풀이] 예 4.1.4의 (c)로부터 $X_i^2 (i=1,\cdots,r)$은 각각 Gamma(1/2,2) 분포를 따르고 이들이 서로 독립이므로, 정리 3.5.2에 주어진 감마분포의 성질로부터

$$Y = X_1^2 + \cdots + X_r^2 \sim \mathrm{Gamma}(r/2, 2)$$

$$\therefore\ pdf_Y(y) = \frac{1}{\Gamma(r/2)2^{r/2}} y^{r/2-1} e^{-y/2} I_{(0,\infty)}(y)$$

자유도(自由度 degrees of freedom)가 r인 카이제곱분포(chi-squared distribution):

$$\begin{aligned} Y \sim \chi^2(r)(r>0) &\Leftrightarrow Y \sim \mathrm{Gamma}(r/2,2) \\ &\Leftrightarrow pdf_Y(y) = \frac{1}{\Gamma(r/2)2^{r/2}} y^{r/2-1} e^{-y/2} I_{(0,\infty)}(y) \\ &\Leftrightarrow Y \overset{d}{\equiv} X_1^2 + \cdots + X_r^2,\ X_i \overset{iid}{\sim} \mathrm{N}(0,1)(i=1,\cdots,r)\ (r\text{이 자연수인 경우}) \end{aligned}$$

---카이제곱분포의 확률밀도함수나 누적확률은 패키지 R을 이용하여 계산

--- $Y \sim \chi^2(r)$일 때

$$P(Y > \chi^2_{\alpha}(r)) = \alpha \ \ (0 < \alpha < 1)$$

를 만족시키는 값 $\chi^2_\alpha(r)$를 자유도 r인 카이제곱분포 $\chi^2(r)$의 상방 α 분위수라고 한다.

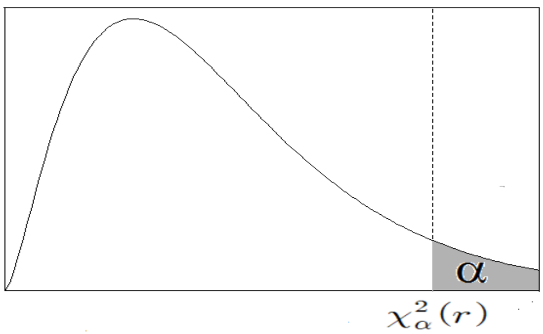


그림 4.2.1 카이제곱분포 $\chi^2(r)$의 형태와 상방 α 분위수

---카이제곱분포의 상방 분위수는 부록 Ⅳ 또는 패키지 R로부터

$$\chi^2_{0.990}(5) = 0.554,\ \ \chi^2_{0.975}(5) = 0.831,\ \ \chi^2_{0.025}(5) = 12.833,\ \ \chi^2_{0.010}(5) = 15.086$$

--- $X \sim \mathrm{Gamma}(\nu,\beta) \Leftrightarrow X/\beta \sim \mathrm{Gamma}(\nu,1) \Leftrightarrow 2X/\beta \sim \mathrm{Gamma}(\nu,2) = \chi^2(2\nu)$이므로

$$P(X > \beta\chi^2_\alpha(2\nu)/2) = P(2X/\beta > \chi^2_\alpha(2\nu)) = \alpha$$

이므로 감마분포 $\mathrm{Gamma}(\nu,\beta)$ 의 상방 α 분위수는 $\beta\chi^2_\alpha(2\nu)/2$ 로 주어진다.

정리 4.2.1:카이제곱분포의 성질

(a) $\mathrm{Y} \sim \chi^2(r)$이면

$$\mathrm{E}(\mathrm{X}) = r,\ \ \mathrm{Var}(\mathrm{X}) = 2r$$

(b) $\mathrm{Y} \sim \chi^2(r)$이면 그 적률생성함수는

$$mgf_{\mathrm{Y}}(t) = (1-2t)^{-r/2},\ t < 1/2$$

(c) $\mathrm{Y}_1 \sim \chi^2(r_1), \mathrm{Y}_2 \sim \chi^2(r_2)$이고 $\mathrm{Y}_1, \mathrm{Y}_2$가 서로 독립이면

$$\mathrm{Y}_1 + \mathrm{Y}_2 \sim \chi^2(r_1 + r_2)$$

[증명] 카이제곱분포 $\chi^2(r)$는 감마분포 $\mathrm{Gamma}(r/2, 2)$이므로 이 정리는 정리 3.5.2에 주어진 감마분포의 성질을 특별한 경우에 정리해 놓은 것이다.

예 4.2.3

두 확률변수 Z와 V가 서로 독립이고, $\mathrm{Z} \sim \mathrm{N}(0,1), \mathrm{V} \sim \chi^2(r)$일 때

$$\mathrm{X} = \frac{\mathrm{Z}}{\sqrt{\mathrm{V}/r}}$$

의 확률밀도함수를 구하여라.

[풀이] $\mathrm{Y} = \mathrm{V}$라고 하면, (Z, V)에서 (X, Y)로의 변환은 그 역변환이

$$\begin{cases} \mathrm{Z} = \mathrm{X}\sqrt{\mathrm{Y}/r} \\ \mathrm{V} = \mathrm{Y} \end{cases}$$

로 주어지는 일대일 변환이므로 X, Y의 결합확률밀도함수는 다음과 같이 주어진다.

$$\begin{aligned} pdf_{\mathrm{X,Y}}(x,y) &= pdf_{\mathrm{Z,V}}(z,v)\left|\det\left(\frac{\partial(z,v)}{\partial(x,y)}\right)\right| \\ &= \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\frac{1}{\Gamma(r/2)2^{r/2}}v^{r/2-1}e^{-v/2}\mathrm{I}_{(v>0)}\left|\det\begin{pmatrix} \sqrt{y/r} & 0 \\ xy^{-1/2}/2\sqrt{r} & 1 \end{pmatrix}\right| \\ &= \frac{1}{\Gamma(1/2)\Gamma(r/2)2^{(r+1)/2}\sqrt{r}}y^{(r+1)/2-1}e^{-(1+x^2/r)y/2}\mathrm{I}_{(\mathrm{y}>0)} \end{aligned}$$

이로부터 X의 주변확률밀도함수를 구하면

$$\begin{aligned} pdf_{\mathrm{X}}(x) &= \int_{-\infty}^{+\infty} pdf_{\mathrm{X,Y}}(x,y)dy \\ &= \frac{1}{\Gamma(1/2)\Gamma(r/2)2^{(r+1)/2}\sqrt{r}}\int_0^{+\infty} y^{(r+1)/2-1}e^{-(1+x^2/r)y/2}dy \end{aligned}$$

여기에서 $(1+x^2/r)y/2 = t$로 치환하여 적분하면

$$\begin{aligned} pdf_{\mathrm{X}}(x) &= \frac{1}{\Gamma(1/2)\Gamma(r/2)2^{(r+1)/2}\sqrt{r}}\int_0^{+\infty} t^{(r+1)/2-1}e^{-t}dt\,((1+x^2/r)/2)^{-(r+1)/2} \\ &= \frac{\Gamma((r+1)/2)}{\Gamma(1/2)\Gamma(r/2)\sqrt{r}}(1+x^2/r)^{-(r+1)/2} \end{aligned}$$

자유도가 r인 t분포(t distribution):

$$\begin{aligned} \mathrm{X} \sim t(r)(r>0) &\Leftrightarrow \mathrm{X} \overset{\mathrm{d}}{\equiv} \frac{\mathrm{Z}}{\sqrt{\mathrm{V}/r}}, \mathrm{Z} \sim \mathrm{N}(0,1), \mathrm{V} \sim \chi^2(r), \mathrm{Z}\text{와 V는 서로 독립} \\ &\Leftrightarrow pdf_{\mathrm{X}}(x) = \frac{\Gamma((r+1)/2)}{\Gamma(1/2)\Gamma(r/2)\sqrt{r}}(1+x^2/r)^{-(r+1)/2} \end{aligned}$$

--- t 분포의 확률밀도함수나 누적확률은 패키지 R을 이용하여 계산

--- t 분포는 0을 기준으로 좌우 대칭인 형태로서 그림 4.2.2와 같다.

--- $\mathrm{X} \sim t(r)$일 때

$$\mathrm{P}(\mathrm{X} > t_\alpha(r)) = \alpha\ \ (0 < \alpha < 1)$$

를 만족시키는 값 $t_\alpha(r)$을 자유도 r인 t분포의 상방 α 분위수라고 한다.

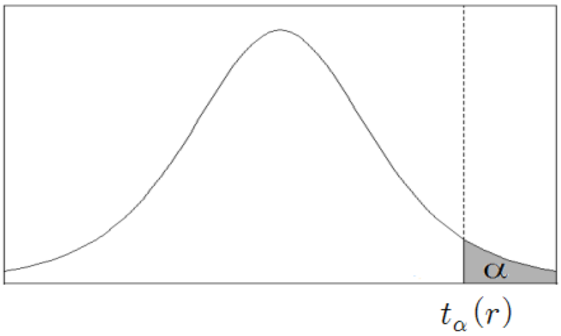


그림 4.2.2 t 분포 $t(r)$의 형태와 상방 α 분위수

--- t 분포의 상방 분위수는 부록 Ⅳ 또는 패키지 R로부터

$$t_{0.10}(5) = 1.476,\ \ t_{0.05}(5) = 2.015,\ \ t_{0.025}(5) = 2.571,\ \ t_{0.005}(5) = 4.032$$

--- 또한 t 분포의 대칭성으로부터

$$t_{0.90}(5) = -\,t_{0.10}(5) = -\,1.476,\ \ t_{0.975}(5) = -\,t_{0.025}(5) = -\,2.571$$

정리 4.2.2: 정규모집단 경우의 표본분포에 관한 기본 정리

정규분포 $\mathrm{N}(\mu,\sigma^2)$에서의 랜덤표본을 $\mathrm{X}_1,\mathrm{X}_2,\cdots,\mathrm{X}_n$이라고 할 때, 다음이 성립한다.

(a) 표본평균 $\overline{\mathrm{X}}=(\mathrm{X}_1+\cdots+\mathrm{X}_n)/n$ 의 분포는 $\mathrm{N}(\mu,\sigma^2/n)$ 이고,

(b) 표본분산 $\mathrm{S}^2=\sum_{i=1}^{n}(\mathrm{X}_i-\overline{\mathrm{X}})^2/(n-1)$와 표본평균 $\overline{\mathrm{X}}$는 서로 독립이며,

(c) $(n-1)\mathrm{S}^2/\sigma^2$ 은 자유도 $(n-1)$인 카이제곱분포를 따른다. 즉

$$(n-1)\mathrm{S}^2/\sigma^2=\sum_{i=1}^{n}(\mathrm{X_i}-\overline{\mathrm{X}})^2/\sigma^2\sim\chi^2(\mathrm{n}-1)$$

[증명] (a) 정리 2.5.11에 주어진 합의 적률생성함수의 공식을 이용.

(b) 표본분산 $\mathrm{S}^2=\sum_{i=1}^{n}(\mathrm{X}_i-\overline{\mathrm{X}})^2/(n-1)$은 $(\mathrm{X}_1-\overline{\mathrm{X}},\cdots,\mathrm{X}_n-\overline{\mathrm{X}})^t$의 함수이므로, 정리 2.5.10에 따라 $\overline{\mathrm{X}}$ 와 $(\mathrm{X}_1-\overline{\mathrm{X}},\cdots,\mathrm{X}_n-\overline{\mathrm{X}})^t$가 서로 독립임을 밝히면 $\overline{\mathrm{X}}$와 S^2도 서로 독립임을 알 수 있다. 한편 $\overline{\mathrm{X}}$ 와 $\mathrm{Y}=(\mathrm{X}_1-\overline{\mathrm{X}},\cdots,\mathrm{X}_n-\overline{\mathrm{X}})^t$의 결합적률생성함수를 구하면 다음과 같다.

$$\begin{aligned}
mgf_{\overline{\mathrm{X}},\mathrm{Y}}(s,(t_1,\cdots,t_n)) &= \mathrm{E}[\exp\{s\overline{\mathrm{X}}+t_1(\mathrm{X}_1-\overline{\mathrm{X}})+\cdots+t_n(\mathrm{X_n}-\overline{\mathrm{X}})\}] \\
&= \mathrm{E}[\exp\{(s/n+(t_1-\bar{t}))\mathrm{X}_1+\cdots+(s/n+(t_n-\bar{t}))\mathrm{X}_n\}] \\
&= \prod_{i=1}^{n} mgf_{\mathrm{X}_i}(s/n+(t_i-\bar{t})) \\
&= \exp[\sum_{i=1}^{n}\left\{\mu(s/n+(t_i-\bar{t}))+\frac{1}{2}\sigma^2(s/n+(t_i-\bar{t}))^2\right\}] \\
&= \exp(\mu s+\frac{1}{2}\frac{\sigma^2}{n}s^2)\exp(\frac{1}{2}\sigma^2\sum_{i=1}^{n}(t_i-\bar{t})^2)
\end{aligned}$$

이로부터 $\overline{\mathrm{X}}$ 와 $\mathrm{Y}=(\mathrm{X}_1-\overline{\mathrm{X}},\cdots,\mathrm{X}_n-\overline{\mathrm{X}})^t$의 결합적률생성함수는 각각의 주변적률생성함수의 곱인 s의 함수와 $t=(t_1,\cdots,t_n)^t$의 함수의 곱으로 나타내어진다. 따라서 정리 2.5.9로부터 $\overline{\mathrm{X}}$ 와 $\mathrm{Y}=(\mathrm{X}_1-\overline{\mathrm{X}},\cdots,\mathrm{X}_n-\overline{\mathrm{X}})^t$는 서로 독립이다.

(c)

$$\sum_{i=1}^{n}(\mathrm{X}_i-\mu)^2=\sum_{i=1}^{n}\left\{(\mathrm{X_i}-\overline{\mathrm{X}})+(\overline{\mathrm{X}}-\mu)\right\}^2=\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{X}_i-\overline{\mathrm{X}})^2+n(\overline{\mathrm{X}}-\mu)^2$$

그러므로

$$\mathrm{U}=\sum_{i=1}^{n}(\mathrm{X}_i-\mu)^2/\sigma^2,\ \ \mathrm{V}=\sum_{i=1}^{n}(\mathrm{X}_i-\overline{\mathrm{X}})^2/\sigma^2=(n-1)\mathrm{S}^2/\sigma^2,\ \ \mathrm{W}=n(\overline{\mathrm{X}}-\mu)^2/\sigma^2$$

라고 하면 $\mathrm{U}=\mathrm{V}+\mathrm{W}$ 이고, (b)로부터 V와 W는 서로 독립이다. 따라서

$$mgf_{\mathrm{U}}(t)=mgf_{\mathrm{V}}(t)\ mgf_{\mathrm{W}}(t)$$

한편, $(\mathrm{X}_i-\mu)/\sigma\,(i=1,\cdots,n)$는 서로독립이고 표준정규분포 $\mathrm{N}(0,1)$을 따르므로

$$\mathrm{U}=\sum_{i=1}^{n}(\frac{\mathrm{X}_i-\mu}{\sigma})^2\sim\chi^2(\mathrm{n}),\ \ 즉\ \ mgf_{\mathrm{U}}(t)=(1-2t)^{-n/2}\mathrm{I}_{(t<1/2)}$$

또한 (a)로부터 $(\overline{\mathrm{X}}-\mu)/(\sigma/\sqrt{n})$도 표준정규분포 $\mathrm{N}(0,1)$을 따르므로

$$\mathrm{W}=(\frac{\overline{\mathrm{X}}-\mu}{\sigma/\sqrt{n}})^2\sim\chi^2(1),\ \ 즉\ \ mgf_{\mathrm{W}}(t)=(1-2t)^{-1/2}\mathrm{I}_{(t<1/2)}$$

$$\therefore\ \ mgf_{\mathrm{V}}(t)=(1-2t)^{-(n-1)/2}\mathrm{I}_{(t<1/2)},\ \ 즉\ \ V=(n-1)S^2/\sigma^2\sim\chi^2(n-1)$$

정리 4.2.3: 정규모집단에서 모평균의 추론

정규분포 $N(\mu,\sigma^2)$에서의 랜덤표본을 $X_1,X_2,\cdots,X_n$이라고 할 때, 다음이 성립한다.

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

$$P\left\{\overline{X}-t_{\alpha/2}(n-1)S/\sqrt{n} \le \mu \le \overline{X}+t_{\alpha/2}(n-1)S/\sqrt{n}\right\}=1-\alpha$$

[증명] $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$, $V=(n-1)S^2/\sigma^2$이라고 하면, 정리 4.2.2로부터

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

이고 Z 와 V는 서로 독립이다. 따라서

$$\frac{\overline{X}-\mu}{S/\sqrt{n}}=\frac{Z}{\sqrt{V/(n-1)}} \sim t(n-1)$$

이고 분위수 정의에 따라

$$P\left\{\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right| \le t_{\alpha/2}(n-1)\right\}=1-\alpha$$

정리 4.2.3에서의 모집단의 평균 μ에 관한 구간

$$[\overline{X}-t_{\alpha/2}(n-1)S/\sqrt{n},\ \ \overline{X}+t_{\alpha/2}(n-1)S/\sqrt{n}]$$

을 **신뢰수준(信頼水準 confidence level)**$(1-\alpha)$**의 신뢰구간(信頼區間 confidence interval)**이라고 하며, 이 때 신뢰수준의 의미는 표본으로부터 계산되는 구간이 미지의 모수 μ를 포함하게 되는 경우가 전체의 $100(1-\alpha)\%$일 것이라는 적중률의 뜻이다.

모집단의 분포가 정규분포인 경우에 평균에 관한 신뢰구간과 마찬가지로 모집단의 분산에 관한 신뢰구간도 다음과 같이 주어진다.

정리 4.2.4: 정규모집단에서 모분산의 추론

정규분포 $N(\mu,\sigma^2)$에서의 랜덤표본을 $X_1,X_2,\cdots,X_n$이라고 할 때, 다음이 성립한다.

$$P\left\{\frac{n-1}{\chi^2_{\alpha/2}(n-1)}S^2 \le \sigma^2 \le \frac{n-1}{\chi^2_{1-\alpha/2}(n-1)}S^2\right\}=1-\alpha$$

[증명] 정리 4.2.2로부터 $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$이므로, 분위수의 정의에 따라

$$P\left\{\chi^2_{1-\alpha/2}(n-1) \le (n-1)S^2/\sigma^2 \le \chi^2_{\alpha/2}(n-1)\right\}=1-\alpha$$

이고 부등식을 σ^2에 관하여 정리하여 신뢰구간을 나타낸 것이다.

예 4.2.4

두 확률변수 V_1과 V_2가 서로 독립이고 $\mathrm{V}_1 \sim \chi^2(r_1)$, $\mathrm{V}_2 \sim \chi^2(r_2)$일 때

$$\mathrm{X} = \frac{\mathrm{V}_1/r_1}{\mathrm{V}_2/r_2}$$

의 확률밀도함수를 구하여라.

[풀이] $\mathrm{Y} = \mathrm{V}_2$라고 하면, $(\mathrm{V}_1, \mathrm{V}_2)$에서 (X, Y)로의 변환은 그 역변환이

$$\begin{cases} \mathrm{V}_1 = r_1 \mathrm{XY}/r_2 \\ \mathrm{V}_2 = \mathrm{Y} \end{cases}$$

로 주어지는 일대일 변환이므로 X, Y의 결합확률밀도함수는 다음과 같이 주어진다.

$$\begin{aligned} pdf_{\mathrm{X,Y}}(x,y) &= pdf_{\mathrm{V}_1,\mathrm{V}_2}(v_1,v_2)|\det(\frac{\partial(v_1,v_2)}{\partial(x,y)})| \\ &= \prod_{i=1}^{2}\left\{\frac{1}{\Gamma(r_i/2)2^{r_i/2}} v_i^{r_i/2-1} e^{-v_i/2} \mathrm{I}_{(v_i>0)}\right\} |\det\begin{pmatrix} r_1 y/r_2 & 0 \\ r_1 x/r_2 & 1 \end{pmatrix}| \\ &= \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} (\frac{r_1}{r_2}xy)^{r_1/2-1} y^{r_2/2-1} e^{-(1+r_1x/r_2)y/2} (\frac{r_1}{r_2}y) \mathrm{I}_{(\mathrm{x}>0,\, \mathrm{y}>0)} \end{aligned}$$

이로부터 X의 주변확률밀도함수를 구하면

$$\begin{aligned} pdf_{\mathrm{X}}(x) &= \int_{-\infty}^{+\infty} pdf_{\mathrm{X,Y}}(x,y)dy \\ &= \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_2)/2}} (\frac{r_1}{r_2})^{r_1/2} x^{r_1/2-1} \mathrm{I}_{(x>0)} \int_0^{+\infty} y^{(r_1+r_2)/2-1} e^{-(1+r_1x/r_2)y/2} dy \end{aligned}$$

여기에서 $(1+r_1x/r_2)y/2 = t$로 치환하여 적분하면

$$\begin{aligned} pdf_{\mathrm{X}}(x) &= \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2)2^{(r_1+r_1)/2}} (\frac{r_1}{r_2})^{r_1/2} x^{r_1/2-1} \mathrm{I}_{(x>0)} \int_0^{+\infty} t^{(r_1+r_2)/2-1} e^{-t} dt\, ((1+r_1x/r_2)/2)^{-(r_1+r_2)/2} \\ &= \frac{\Gamma((r_1+r_2)/2)}{\Gamma(r_1/2)\Gamma(r_2/2)} (\frac{r_1}{r_2})^{r_1} x^{r_1/2-1} (1+r_1x/r_2)^{-(r_1+r_2)/2} \mathrm{I}_{(x>0)} \end{aligned}$$

자유도가 (r_1, r_2)인 F분포(F distribution):

$$\begin{aligned} & \mathrm{X} \sim \mathrm{F}(r_1, r_2)(r_i > 0, i = 1,2) \\ \Leftrightarrow\ & \mathrm{X} \overset{\mathrm{d}}{\equiv} \frac{\mathrm{V}_1/r_1}{\mathrm{V}_2/r_2}, \mathrm{V}_i \sim \chi^2(r_i)(i=1,2), \mathrm{V}_1\text{과 } \mathrm{V}_2\text{는 서로 독립} \\ \Leftrightarrow\ & pdf_{\mathrm{X}}(x) = \frac{\Gamma((r_1+r_2)/2)}{\Gamma(r_1/2)\Gamma(r_2/2)} (\frac{r_1}{r_2})^{r_1/2} x^{r_1/2-1} (1+r_1x/r_2)^{-(r_1+r_2)/2} \mathrm{I}_{(x>0)} \end{aligned}$$

--- F 분포의 확률밀도함수나 누적확률은 패키지 R을 이용하여 계산

--- $\mathrm{X} \sim \mathrm{F}(r_1, r_2)$일 때

$$\mathrm{P}(\mathrm{X} > \mathrm{F}_\alpha(r_1, r_2)) = \alpha \ \ (0 < \alpha < 1)$$

를 만족시키는 값 $\mathrm{F}_\alpha(\mathrm{r}_1, \mathrm{r}_2)$을 자유도가 (r_1, r_2)인 F분포의 상방 α 분위수라고 한다.

--- F분포의 상방 분위수는 부록 Ⅳ 또는 패키지 R로부터

$$\mathrm{F}_{0.05}(5,10) = 3.330, \quad \mathrm{F}_{0.025}(5,10) = 4.240, \quad \mathrm{F}_{0.01}(5,10) = 5.640$$

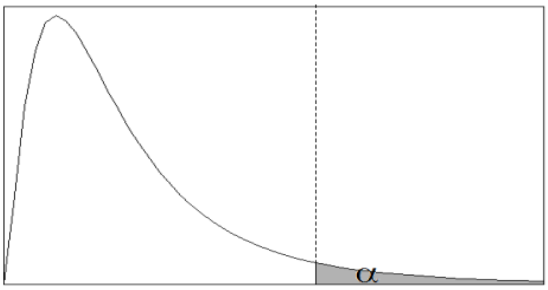


그림 4.2.3 F 분포 $F(r_1,r_2)$의 형태와 상방 α 분위수

정리 4.2.5: F 분포의 성질

(a) $X \sim F(r_1,r_2)$이면 $1/X \sim F(r_2,r_1)$이다. 따라서

$$F_{1-\alpha}(r_1,r_2) = 1/F_\alpha(r_2,r_1)$$

(b) $X \sim t(r)$이면 $X^2 \sim F(1,r)$이다. 따라서

$$t^2_{\alpha/2}(r) = F_\alpha(1,r)$$

[증명] (a) F 분포의 정의로부터

$$X \sim F(r_1,r_2) \Leftrightarrow X \overset{d}{\equiv} \frac{V_1/r_1}{V_2/r_2}, V_i \sim \chi^2(r_i)(i=1,2), V_1\text{과 } V_2\text{는 서로 독립}$$

$$\therefore\ 1/X \overset{d}{\equiv} \frac{V_2/r_2}{V_1/r_1}, V_i \sim \chi^2(r_i)(i=1,2), V_1\text{과 } V_2\text{는 서로 독립}$$

$$\therefore 1/X \sim F(r_2,r_1)$$

한편 분위수 사이의 관계는 다음 등식으로부터 알 수 있다.

$$1-\alpha = P\{X \ge F_{1-\alpha}(r_1,r_2)\},\ \ \alpha = P\{1/X \ge F_\alpha(r_2,r_1)\}$$

$$\therefore \alpha = P\{X \le F_{1-\alpha}(r_1,r_2)\} = P\{X \le 1/F_\alpha(r_2,r_1)\}$$

(b) t 분포의 정의로부터

$$X \sim t(r) \Leftrightarrow X \overset{d}{\equiv} \frac{Z}{\sqrt{V/r}}, Z \sim N(0,1), V \sim \chi^2(r), Z\text{와 } V\text{는 서로 독립}$$

$$\therefore X^2 \overset{d}{\equiv} \frac{Z^2}{V/r}, Z \sim N(0,1), V \sim \chi^2(r), Z\text{와 } V\text{는 서로 독립}$$

그런데 $Z^2 \sim \chi^2(1)$ 이고 Z^2과 V가 서로 독립이므로 $X^2 \sim F(1,r)$이다. 한편 분위수 사이의 관계는 다음 등식으로부터 알 수 있다.

$$\alpha = P\{X^2 \ge F_\alpha(1,r)\},\ \ \alpha = P\{|X| \ge t_{\alpha/2}(r)\} = P\{X^2 \ge t^2_{\alpha/2}(r)\}$$

예 4.2.5

부록 Ⅳ의 분포표로부터

$$F_{0.975}(5,10) = 1/F_{0.025}(10,5) = 1/6.620 = 0.151$$

$$t^2_{0.025}(5) = 2.571^2 = 6.610,\ F_{0.05}(1,5) = 6.610$$

정리 4.2.6:두 정규모집단에서 모분산의 비교

두 정규분포 $\mathrm{N}(\mu_1,\sigma_1^2),\mathrm{N}(\mu_2,\sigma_2^2)$에서의 랜덤표본이 각각 $\mathrm{X}_{11},\mathrm{X}_{12},\cdots,\mathrm{X}_{1n_1}$과 $\mathrm{X}_{21},\mathrm{X}_{22},\cdots,\mathrm{X}_{2n_2}$ 이고 두 랜덤표본이 서로 독립일 때,

$$\overline{\mathrm{X}_i}=\sum_{j=1}^{n_i}\mathrm{X}_{ij}/n_i,\ \ \mathrm{S}_i^2=\sum_{j=1}^{n_i}(\mathrm{X}_{ij}-\overline{\mathrm{X}_i})^2/(n_i-1)\ \ (i=1,2)$$

라고 하면

$$\frac{\mathrm{S}_1^2/\sigma_1^2}{\mathrm{S}_2^2/\sigma_2^2}\sim \mathrm{F}(n_1-1,n_2-1)$$

$$\mathrm{P}\left\{\frac{\mathrm{S}_1^2}{\mathrm{S}_2^2}\frac{1}{\mathrm{F}_{\alpha/2}(n_1-1,n_2-1)}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{\mathrm{S}_1^2}{\mathrm{S}_2^2}\mathrm{F}_{\alpha/2}(n_2-1,n_1-1)\right\}=1-\alpha$$

[증명] 표본분포에 관한 기본 정리인 정리 4.2.2로부터

$$(n_1-1)\mathrm{S}_1^2/\sigma_1^2\sim\chi^2(n_1-1),\quad (n_2-1)\mathrm{S}_2^2/\sigma_2^2\sim\chi^2(n_2-1)$$

이고 두 랜덤표본이 독립이므로 S_1^2, S_2^2은 서로 독립이다. 따라서 F 분포의 정의로부터

$$\frac{\mathrm{S}_1^2/\sigma_1^2}{\mathrm{S}_2^2/\sigma_2^2}\sim \mathrm{F}(n_1-1,n_2-1)$$

$$\therefore\ \ \mathrm{P}\left\{\mathrm{F}_{1-\alpha/2}(n_1-1,n_2-1)\le\frac{\mathrm{S}_1^2/\sigma_1^2}{\mathrm{S}_2^2/\sigma_2^2}\le\mathrm{F}_{\alpha/2}(n_1-1,n_2-1)\right\}=1-\alpha$$

이 확률 표현에서 부등식을 σ_1^2/σ_2^2에 관하여 정리하면

$$\frac{\mathrm{S}_1^2}{\mathrm{S}_2^2}\frac{1}{\mathrm{F}_{\alpha/2}(n_1-1,n_2-1)}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{\mathrm{S}_1^2}{\mathrm{S}_2^2}\frac{1}{\mathrm{F}_{1-\alpha/2}(n_1-1,n_2-1)}$$

이고, 정리 4.2.5에 주어진 F 분포의 분위수에 관한 성질을 이용하면 정리에 주어진 σ_1^2/σ_2^2에 관한 신뢰구간의 표현을 얻게 된다.

대표적 표본분포

분포의 명칭과 기호 (distribution)	대의적 정의 (representational definition)
	확률밀도함수 (pdf)
카이제곱분포 $\chi^2(r)$ $(r=1,2,\cdots)$	$\mathrm{X}\sim\chi^2(r)\Leftrightarrow\mathrm{X}\sim\mathrm{Gamma}(r/2,2)$ $\Leftrightarrow\mathrm{X}\overset{d}{\equiv}\mathrm{Z}_1^2+\cdots+\mathrm{Z}_r^2,\ \mathrm{Z}_i\overset{iid}{\sim}\mathrm{N}(0,1)(i=1,\cdots,r)$
	$pdf_{\mathrm{X}}(x)=\dfrac{1}{\Gamma(r/2)2^{r/2}}x^{r/2-1}e^{-x/2}\mathrm{I}_{(0,\infty)}(x)$
t 분포 $t(r)$ $(r=1,2,\cdots)$	$\mathrm{X}\sim t(r)$ $\Leftrightarrow\mathrm{X}\overset{\mathrm{d}}{\equiv}\dfrac{\mathrm{Z}}{\sqrt{\mathrm{V}/r}},\mathrm{Z}\sim\mathrm{N}(0,1),\mathrm{V}\sim\chi^2(r),\mathrm{Z},\mathrm{V}$는 서로 독립
	$pdf_{\mathrm{X}}(x)=\dfrac{\Gamma((r+1)/2)}{\Gamma(1/2)\Gamma(r/2)\sqrt{r}}(1+x^2/r)^{-(r+1)/2}$
F 분포 $\mathrm{F}(r_1,r_2)$ $(r_i=1,2,\cdots)(i=1,2)$	$\mathrm{X}\sim\mathrm{F}(r_1,r_2)$ $\Leftrightarrow\mathrm{X}\overset{\mathrm{d}}{\equiv}\dfrac{\mathrm{V}_1/r_1}{\mathrm{V}_2/r_2},\mathrm{V}_i\sim\chi^2(r_i)(i=1,2),\mathrm{V}_1,\ \mathrm{V}_2$는 서로 독립
	$pdf_{\mathrm{X}}(x)$ $=\dfrac{\Gamma((r_1+r_2)/2)}{\Gamma(r_1/2)\Gamma(r_2/2)}(\dfrac{r_1}{r_2})^{r_1/2}x^{r_1/2-1}(1+\dfrac{r_1}{r_2}x)^{-(r_1+r_2)/2}\mathrm{I}_{(x>0)}$
베타분포 $\mathrm{Beta}(\alpha_1,\alpha_2)$ $(\alpha_1>0,\alpha_2>0)$	$\mathrm{X}\sim\mathrm{Beta}(\alpha_1,\alpha_2)$ $\Leftrightarrow\mathrm{X}\overset{d}{\equiv}\mathrm{Z}_1/(\mathrm{Z}_1+\mathrm{Z}_2),\ \mathrm{Z}_i\sim\mathrm{Gamma}(\alpha_i,\beta)(i=1,2)$이고 서로 독립
	$pdf_{\mathrm{X}}(x)=\dfrac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}x^{\alpha_1-1}(1-x)^{\alpha_2-1}\mathrm{I}_{(0,1)}(x)$
디리클레분포 $\mathrm{Dirichlet}(\alpha_1,\cdots,\alpha_{k+1})$ $(\alpha_i>0,i=1,\cdots,k+1)$	$\mathrm{X}=(\mathrm{X}_1,\cdots,\mathrm{X}_k)^t\sim\mathrm{Dirichlet}(\alpha_1,\cdots,\alpha_k,\alpha_{k+1})$ $\Leftrightarrow\mathrm{X}\overset{d}{\equiv}(\dfrac{\mathrm{X}_1}{\mathrm{X}_1+\cdots+\mathrm{X}_{k+1}},\cdots,\dfrac{\mathrm{X}_k}{\mathrm{X}_1+\cdots+\mathrm{X}_{k+1}}),$ $\mathrm{X}_i\sim\mathrm{Gamma}(\alpha_i,\beta)$이고 서로독립$(i=1,\cdots,k+1)$
	$pdf_{\mathrm{X}_1,\cdots,\mathrm{X}_k}(x_1,\cdots,x_k)$ $=c\left\{\prod_{i=1}^{k}x_i^{\alpha_i-1}\right\}(1-x_1-\cdots x_k)^{\alpha_{k+1}-1}\mathrm{I}_{(x_1>0,\cdots,x_k>0,x_1+\cdots+x_k<1)}$ $c=\dfrac{\Gamma(\alpha_1+\cdots+\alpha_{k+1})}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{k+1})}$