정규분포

정의

- 응용과 통계적 추론에서 특히 중요한 분포 및 분포족

1. I = $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$를 정의하면

1) 적분 계산을 위해, 이 적분을 제곱하면

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

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

$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2} - \frac{z^2}{2}\right) dzdz$

(2) 하나의 z를 w로 치환하면 $\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{z^2+w^2}{2}\right) dzdw$

정의

- 응용과 통계적 추론에서 특히 중요한 분포 및 분포족

2) 이 이중적분을 극좌표 치환하면

(1) $z=r\sin(\theta)$,$w=r\cos(\theta)$ 일 때

- $|J| = \begin{bmatrix} \frac{\partial r\sin(\theta)}{\partial r} & \frac{\partial r\sin(\theta)}{\partial \theta} \\ \frac{\partial r\cos(\theta)}{\partial r} & \frac{\partial r\cos(\theta)}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \sin(\theta) & r\cos(\theta) \\ \cos(\theta) & -r\sin(\theta) \end{bmatrix} = r(\sin^2(\theta)\text{-}\cos^2(\theta)) = r$

- $\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(-\frac{z^2+w^2}{2}\right)dzdw = \frac{1}{2\pi}\int_{0}^{2\pi}\int_{-\infty}^{\infty} r\cdot\exp\left(-\frac{r^2}{2}\right)drd\theta$

(2) $\frac{r^2}{2}$=u로 치환하면 $r\cdot\exp\left(-\frac{r^2}{2}\right) \rightarrow [\exp(-u)]_0^{\infty}=1$

- $\frac{1}{2\pi}\int_0^{2\pi} 1\, d\theta = 1$

정의

- 정규 분포의 mgf

1. 정규분포의 mgf는 다음과 같이 구한다.

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

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

3) U = $(z-t)$ 로 치환하면

(1) $e^{\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = e^{\frac{1}{2}t^2}$ =M(t)

(2) M′(t) = $te^{\frac{1}{2}t^2}$, M′(0) = 0

(3) M″(t) = $e^{\frac{1}{2}t^2} + t^2 e^{\frac{1}{2}t^2}$, M″(0) = 1

정의

- 정규분포의 가법성

1. $x_1, \cdots, x_n$을 i=1, ⋯,n에 대해 모두 N(μ_i, σ_i^2) 분포를 따르는 독립인 확률변수라고 할 때,

1) Y= $\sum a_i x_i$ 일 때, Y의 분포는 N($\sum a_i \mu_i$, $a_i^2 \sigma_i^2$)을 따른다.

2) 따라서, E(x) = $\overline{x}$ = E($\sum \frac{x_i}{n_i}$) = $\frac{1}{n_i}$ $\sum E(\ x_i)$ 인 결합분포는 N(μ, $\frac{\sigma}{n}$)을 따른다.

정의

- 표준정규분포

1. 정규분포의 변환을 $\mathrm{x} = bz + a$ 로 정의하면 $\mathrm{z} = (x - a) \cdot b^{-1}$

1) 야코비안 |J| = $\mathrm{z} = \frac{\partial(x-a)\cdot b^{-1}}{\partial \mathrm{x}} = b^{-1}$

2) 변환을 다시 정리하면

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}b} \exp\left(-\frac{1}{2}(\frac{x-a}{b})^2\right) dz$$

2. 위 변환의 mgf를 구하면

1) $E(e^{(bz+a)t})$ = $e^{at}E(e^{tbz}) = e^{at} \cdot e^{\frac{1}{2}t^2b^2} = e^{at+\frac{1}{2}t^2b^2}$ (MGF의 변환)

2) M′(0) = a$e^{a0+\frac{1}{2}0^2b^2}$ + $b^2 0 e^{a0+\frac{1}{2}0^2b^2}$ = a

3) M″(0) = $a^2 e^{a0+\frac{1}{2}0^2b^2}$ + ab$^2 0 e^{a0+\frac{1}{2}0^2b^2}$ + $b^2 e^{a0+\frac{1}{2}0^2b^2}$ + ab$^2 0 e^{a0+\frac{1}{2}0^2b^2}$ + $b^4 0^2 e^{a0+\frac{1}{2}0^2b^2}$ $= a^2 + b^2$

4) Var(x) = $a^2 + b^2$ - a^2 = b^2

정의

- 표준정규분포와 정규분포는 서로 밀접한 관계에 놓여있다.

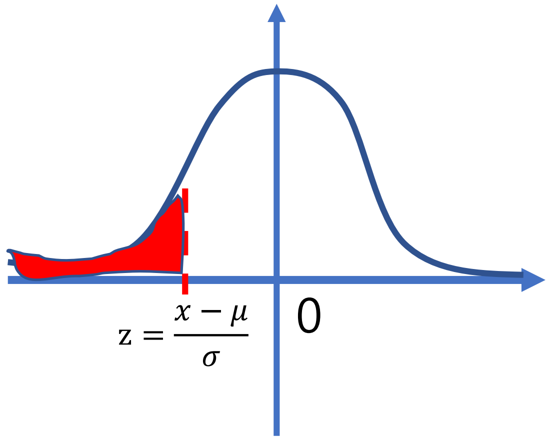
1. $\mathrm{x} = \sigma z + \mu$ 이므로, $\mathrm{z} = \frac{x-\mu}{\sigma}$ 일 때

   1) X가 N(a,b)를 따를경우, 그 역함수 $\frac{x-a}{b}$는 N(0,1)을 따른다.

2) 위 특성을 이용하여

(1) $F_{,x}(x) = p(X \leq x) = P(\sigma z + \mu \leq \mathrm{x}) = \mathrm{P}(\mathrm{Z} \leq \frac{x-\mu}{\sigma})$

(2) 해당 변수가 표준정규분포로 환산했을 때 어디에 위치하는지를 알 수 있다.



정의

- 오염된 정규분포

1. 개요

1) 혼합분포의 한 형태로서, 독립인 다른 분포와 결합하여 그 형태가 변한 정규분포

2) 결합된 형태를 정규분포로 환원하여 그 구조를 밝혀낼 수 있다.

2. Z가 N(0,1)을 따르는 확률변수이고, $I_{1-\varepsilon}$는 $I_{1-\varepsilon} = \begin{cases} 1, & \text{확률이 } 1-\varepsilon \\ 0, & \text{확률이 } \varepsilon \end{cases}$ 일 때

1) Z과 $I_{1-\varepsilon}$가 서로 독립이고, 결합 분포가 W = $ZI_{1-\varepsilon} + \sigma_\varepsilon Z(1-I_{1-\varepsilon})$ 라면, $I_{1-\varepsilon}$는 스위치 역할을 한다.

정의

- 오염된 정규분포

2) $P(\mathrm{W} \leq w) = P[\mathrm{W} \leq w \cap (I_{1-\varepsilon} = 1)] + P[\mathrm{W} \leq w \cap (I_{1-\varepsilon} = 0)]$

(1)조건부 확률의 성질을 사용하면

$P[\mathrm{W} \leq w \mid (I_{1-\varepsilon} = 1)] \cdot P(I_{1-\varepsilon} = 1) + P[\mathrm{W} \leq w \mid (I_{1-\varepsilon} = 0)] \cdot P(I_{1-\varepsilon} = 0)$

(2) 위 식은

$P[\mathrm{W} \leq w \mid (I_{1-\varepsilon} = 1)] \cdot (1 - \varepsilon) + P[\mathrm{W} \leq w \mid (I_{1-\varepsilon} = 0)] \cdot \varepsilon$

(3) $\mathrm{Z}I_{1-\varepsilon} + \sigma_\varepsilon \mathrm{Z}(1\text{-}I_{1-\varepsilon})$ 에서 $\boxed{I_{1-\varepsilon} = 1 \text{ 일 때 } \mathrm{Z}I_{1-\varepsilon} = \mathrm{Z}}$ $\boxed{I_{1-\varepsilon} = 0 \text{ 일 때 } \sigma_\varepsilon \mathrm{Z}}$

$P[\mathrm{W} \leq w \mid (I_{1-\varepsilon} = 1)] \cdot (1 - \varepsilon) + P[\mathrm{W} \leq w \mid (I_{1-\varepsilon} = 0)] \cdot \varepsilon$

$= P[\mathrm{Z} \leq w] \cdot (1 - \varepsilon) + P[\mathrm{Z} \leq \frac{w}{\sigma_\varepsilon}] \cdot \varepsilon$

정의

- 오염된 정규분포

3) 이 때, 각각의 기댓값을 구하면

(1) $E(w) = E[Z(w) \cdot (1-\varepsilon) + Z(\frac{w}{\sigma_\varepsilon}) \cdot \varepsilon] = 0 \cdot (1-\varepsilon) + 0 \cdot \varepsilon = 0$

(2) $E[(w-\mu)^2] = E[z^2 \cdot I^2 + 2\sigma_\varepsilon \cdot zI(1-I) + {\sigma_\varepsilon}^2 z^2 (1\text{-}I)^2]$

$= \underbrace{E[z^2]}_{1} \cdot E[I^2] + \cancel{2\sigma_\varepsilon \cdot E[z]E[I(1-I)]} + {\sigma_\varepsilon}^2 \underbrace{E[z^2]}_{1} E[(1-I)^2]$

$= E[I^2] + {\sigma_\varepsilon}^2\ E[1 - 2I - I^2]$, $I^2 = I$ 이므로

$= (1-\varepsilon) + {\sigma_\varepsilon}^2\ E[1-(1-\varepsilon)] = 1 + \varepsilon({\sigma_\varepsilon}^2 - 1)$

예제

- N(μ, σ) 인 정규분포에서 P(x ≤ 60)일 때 10%이고, P(x ≤ 90)일 때 5%이다. μ, σ를 구하시오

1. $\frac{60-\mu}{\sigma}$ = 1.282, $\frac{90-\mu}{\sigma}$ =1.645 이므로, 연립해서 풀면 μ=73.1, σ=10.2 이다.