Mathematical Statistics I
Dec 2022 - Jan 2023
Mentoring

Mentor
Joonhyuk JUNG (정준혁)
Department of Statistics
25동 408호

학생(제2022-18호)

활동확인서

소 속	통계학과
생년월일	1999.04.20
성 명	정 준 혁

위 학생은 서울대학교 자연과학대학 학생회 사업 2022학년 겨울 계절학기 전공학교에서 **『수리 통계 1』** 과목 멘토로 활동하였음을 인증합니다.

2023년 1월 31일

서울대학교 자연과학대학장

수리통계학 교과서 1.2절까지 읽고 다음 물음에 답하여라.

1 확률의 공리

- (확률측도의 연속성) $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ 에 대하여 $\mathbb{P}\left(\cup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$
- (확률의 유한가법성) $\{A_k\}_{k=1}^{n}$ 에 대하여 $A_i \cap A_j = \emptyset \;\; (i \neq j)$ 이면 $\mathbb{P}\left(\cup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathbb{P}(A_k)$
- (확률의 가산가법성) $\{A_n\}_{n=1}^{\infty}$ 에 대하여 $A_i \cap A_j = \emptyset \;\; (i \neq j)$ 이면 $\mathbb{P}\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$

본문에서는 확률의 가산가법성을 공리로 받아들이고 확률측도의 연속성을 증명하였다. 반대로 확률측도의 연속성과 확률의 유한가법성을 가정하고 확률의 가산가법성을 유도하시오.

2 조건부확률과 확률의 가산가법성

(a) 앞면이 나올 확률이 $\frac{1}{2}$ 인 공정한 동전 한 닢이 있다. 성공률이 $\frac{1}{3}$ 인 실험을 정의하시오.
(b) 앞면이 나올 확률이 $\frac{1}{2}$ 인 공정한 동전 한 닢이 있다. 성공률이 $\frac{1}{\pi}$ 인 실험을 정의하시오.
Hint. 임의의 $0 < p < 1$ 에 대하여

$$\sum_{n=1}^{\infty} \frac{b_n}{2^n} = p \qquad\qquad \forall n, \;\; b_n \in \{0, 1\}$$

를 만족하는 수열 $\{b_n\}_{n=1}^{\infty}$ 이 존재한다. *유일하진 않다.*

3 확률측도의 연속성

$\{A_n\}_{n\geq 1}$ 이 집합의 열일 때, 다음의 질문에 답하시오.
(a)

$$A = \cup_{n=1}^{\infty} \cap_{k\geq n} A_k = \cap_{n=1}^{\infty} \cup_{k\geq n} A_k$$

를 만족한다고 할 때, $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$ 임을 보여라.
(b) A_n 이 증가하거나 감소할 때 $\cup_{n=1}^{\infty} \cap_{k\geq n} A_k = \cap_{n=1}^{\infty} \cup_{k\geq n} A_k$ 가 성립함을 보이시오.
(c) $\cup_{n=1}^{\infty} \cap_{k\geq n} A_k = \cap_{n=1}^{\infty} \cup_{k\geq n} A_k$ 를 만족하는 증가하거나 감소하지 않는 열 $\{A_n\}$ 의 예를 드시오.

4 조건부독립성

- 사건 A, B, E 에 대하여 $\mathbb{P}(A \cap B|E) = \mathbb{P}(A|E)\mathbb{P}(B|E)$ 가 성립하면 A 와 B 는 E 에 대해서 조건부독립 (A and B are conditionally independent given E.)이라고 하고 $A \perp\!\!\!\perp B|E$ 라고 나타낸다.

(a) $A \perp\!\!\!\perp B$ 이면 $A \perp\!\!\!\perp B|E$ 인가?
(b) $A \perp\!\!\!\perp B|E$ 이면 $A \perp\!\!\!\perp B$ 인가?
(c) $A \perp\!\!\!\perp B|E$ 이면 $A \perp\!\!\!\perp B^c|E$ 인가?
(d) $A \perp\!\!\!\perp B|E$ 이면 $A \perp\!\!\!\perp B|E^c$ 인가?

1

$A_i \cap A_j = \emptyset (i \neq j)$ 이면 $B_n = \cup_{k=1}^{n} A_k$ 에 대하여 $B_1 \subseteq B_2 \subseteq \cdots$ 가 성립하므로

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) &= \lim_{n\to\infty} \mathbb{P}(B_n) && \text{(확률측도의 연속성)} \\
&= \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(A_k) && \text{(확률의 유한가법성)} \\
&= \sum_{n=1}^{\infty} \mathbb{P}(A_n)
\end{aligned}$$

을 얻는다. 따라서 $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n$ 을 증명하면 충분하다. 임의의 n 에 대하여 $B_n \supseteq A_n$ 이므로 $\cup_{n=1}^{\infty} B_n \supseteq \cup_{n=1}^{\infty} A_n$ 이 성립한다. 반대로 $\cup_{n=1}^{\infty} B_n \subseteq \cup_{n=1}^{\infty} A_n$ 을 보이면 충분하다. 임의의 $x \in \cup_{n=1}^{\infty} B_n$ 에 대하여

$$\exists m, x \in B_m \Longrightarrow \exists k \leq m, x \in A_k \Longrightarrow x \in \cup_{n=1}^{\infty} A_n$$

이 성립하므로 증명 끝.

2

(a) omitted.
(b) *Hint*와 같이 $\{b_n\}$ 을 잡자. $(b_n \in \{0,1\})$ $n = 1, 2, \cdots$ 에 대하여 한 번씩 동전을 던지는 다음과 같은 실험을 생각하자.

$$a_n = \begin{cases} 1 & \text{앞면} \\ 0 & \text{뒷면} \end{cases}$$

에 대하여

$$\begin{cases} \text{성공으로 실험을 종료} & a_n < b_n \\ \text{실패로 실험을 종료} & a_n > b_n \\ n \leftarrow n+1 & a_n = b_n \end{cases}$$

n 번째 step에서 실험이 성공하는 사건을 A_n 이라고 나타내면

$$\mathbb{P}(A_n) = \underbrace{\left(\frac{1}{2}\right)^{n-1}}_{\text{처음 } n-1 \text{ 번 실험을 종료하지 않음}} \cdot \underbrace{\left(\frac{b_n}{2}\right)}_{\text{성공으로 실험을 종료}} = \frac{b_n}{2^n}$$

이므로 실험을 성공하는 사건 $A = \cup_{n=1}^{\infty} A_n$ 의 확률은 가산가법성에 의해 $\mathbb{P}(A) = \sum_{n=1}^{\infty} \frac{b_n}{2^n} = p$ 가 성립한다. ($\{A_n\}$ 이 disjoint한 사건열인 이유는 무엇인가?)

3 (a)

다음과 같이 사건열 $\{B_n\}_{n=1}^{\infty}, \{C_n\}_{n=1}^{\infty}$ 을 정의하자.

$$B_n = \bigcap_{k \geq n} A_k \qquad\qquad C_n = \bigcup_{k \geq n} A_k$$

- (a)번 문제의 조건은 $A = \cup_{n=1}^{\infty} B_n = \cap_{n=1}^{\infty} C_n$ 으로 요약된다.
- $\{B_n\}_{n=1}^{\infty}$ 은 증가하는 사건열이다.
- $\{C_n\}_{n=1}^{\infty}$ 은 감소하는 사건열이다.
- 임의의 자연수 n 에 대하여 $B_n \subseteq A_n \subseteq C_n$ 이 성립한다.

위의 1, 2, 3번째 고찰과 정리 1.1.3 (확률측도의 연속성)에 의하여

$$\lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \mathbb{P}(A) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} C_n\right) = \lim_{n\to\infty} \mathbb{P}(C_n)$$

을 얻는다. 또한, 4번째 고찰과 정리 1.1.1 (확률의 단조성)

$$\mathbb{P}(B_n) \leq \mathbb{P}(A_n) \leq \mathbb{P}(C_n)$$

이 성립한다. 마지막으로 샌드위치 정리에 의하여

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$$

가 성립한다. 증명 끝.

3 (b)

먼저 A_n 이 증가인 경우만 생각하자. 위의 (a)번과 같은 방식으로 B_n, C_n 을 정의한다. 그러면

$$B_n = A_n \qquad\qquad C_n \equiv C := \bigcup_{k=1}^{\infty} A_k$$

가 성립함을 먼저 보이겠다. B_n 의 경우는 쉽고, C_n 의 경우를 살펴보자. (수업시간에 $C_n = C$ 를 증명하였는데, 설명이 조금 부족한 것 같아 여기서 보충합니다.) C_n 에 정의에 의하여 A_n 의 증가 여부에 무관하게 $C_n \subseteq C$ 임은 쉽게 확인할 수 있으므로 $C_n \supseteq C$ 만 밝히면 된다. 임의의 $x \in C$ 에 대하여

$$\begin{aligned}
\exists m, x \in A_m &\Longrightarrow \exists m, \forall \ell \geq m, x \in A_\ell && (A_n \text{ 은 증가}) \\
&\Longrightarrow \exists k \geq n, x \in A_k && (\text{choose } k = \max\{m, n\}) \\
&\Longrightarrow x \in \cup_{k\geq n} A_k = C_n && (C_n \text{ 의 정의})
\end{aligned}$$

이 성립한다. 이제 $\bigcup_{n=1}^{\infty} B_n = C = \bigcap_{n=1}^{\infty} C_n$ 임을 확인하는 것은 어렵지 않다.
(A_n 이 감소인 경우에는 어떻게 되는가?)

3 (c)

자연수 m 에 대하여

$$[m] = \{1, 2, \cdots, m\} \subseteq \mathbb{N}$$

라고 정의하자. 이제 자연수 p, q 에 대하여

$$[p] \subseteq [q] \Longleftrightarrow p \leq q$$

임을 발견할 수 있다.

$$A_n = \begin{cases} [m] & n = 2m \\ \\ [2m-1] & n = 2m-1 \end{cases}$$

라고 정의하자. A_n 이 증가하거나 감소하지 않음을 쉽게 확인할 수 있다.

$$B_n = \begin{cases} [m] & n = 2m \\ \\ [m] & n = 2m-1 \end{cases} \qquad\qquad C_n \equiv \mathbb{N}$$

(Why?) 이제 $\bigcup_{n=1}^{\infty} B_n = \mathbb{N} = \bigcap_{n=1}^{\infty} C_n$ 임을 확인하는 것은 어렵지 않다.

4

네 개의 주장 중 유일하게 (c)만 참이다. (c)의 증명은 교과서 예 1.2.5의 증명에서 모든 $\mathbb{P}(\cdot)$ 를 $\mathbb{P}(\cdot|E)$ 로 대체하면 된다. 나머지 보기들의 반례를 다음과 같이 만들자. 공정한 정육면체 주사위를 두 번 굴린다. 형식적으로는 표본공간 $S = \{1, 2, 3, 4, 5, 6\}^2$ 과 사건들의 집합 $\mathcal{F} = 2^S$ 에 대하여 확률 $\mathbb{P}$ 를

$$\mathbb{P}(E) = \frac{|E|}{|S|} \qquad\qquad E \in \mathcal{F}$$

로 정의한다.
(a) 반례: $A = \{(i, j) \in S : i = 1\}, B = \{(i, j) \in S : j = 1\}, E = \{(i, j) \in S : ij$는 짝수$\}$
Note. $\mathbb{P}(A \cap B|E) = 0$.
(b) 반례: $A = \{(i, j) \in S : i = 1\}, B = \{(i, j) \in S : i + j = 6\}, E = \{(i, j) \in S : ij$는 홀수$\}$
(d) 반례: $A = \{(i, j) \in S : i = 1\}, B = \{(i, j) \in S : j = 1\}, E = \{(i, j) \in S : ij$는 홀수$\}$

1 누적분포함수의 성질

확률변수 X 의 누적분포함수(cdf)를 $F(x) = \mathbb{P}(X \leq x)$ 라고 쓴다면

$$x_1 < x_2 \Longrightarrow F(x_1) \leq F(x_2) \tag{증가}$$

$$\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1 \tag{전체 변동}$$

$$\lim_{h \downarrow 0} F(x+h) = F(x) \tag{오른쪽 연속}$$

이 성립하는 것을 교재에서 증명하였다. 여기서는 역으로 위의 성질을 만족하는 함수 F 가 존재한다면, F 는 반드시 어떤 확률변수 X 의 cdf가 될 수밖에 없음을 설명하겠다.
다음과 같은 pdf f_U 를 가지는 확률변수 U 에 대하여 $X = F^{-1}(U)$ 로 정의한다.

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

단, F^{-1} 는

$$F^{-1}(u) = \inf \underbrace{\{y \in \mathbb{R} : F(y) \geq u\}}_{A(u)} \qquad 0 < u < 1$$

로 정의한다. 임의의 $0 < u < 1$ 에 대하여 $\lim_{x \to \infty} F(x) = 1$ 이므로 $A(u) \neq \emptyset$ 이고 $\lim_{x \to -\infty} F(x) = 0$ 이므로 $A(u)$ 는 아래로 유계이다. (Why?) 따라서 $\inf A(u) \in \mathbb{R}$ 이 잘 정의된다. $\inf A(u) \leq x \Longleftrightarrow u \leq F(x)$ 임을 증명하게 되면

$$\begin{aligned}
\mathbb{P}(X \leq x) &= \mathbb{P}(F^{-1}(U) \leq x | U \sim f_U) \\
&= \mathbb{P}(U \leq F(x) | U \sim f_U) \\
&= F(x) && (0 \leq F(x) \leq 1)
\end{aligned}$$

가 되어 F 는 X 의 cdf임을 확인할 수 있다. 먼저 $\inf A(u) \leq x \Longleftarrow u \leq F(x)$ 는 F 의 오른쪽 연속성에 무관하게 성립하는 사실이다.
반대 방향을 보이기 위해 $u \leq F(x)$ 를 부정하자. F 가 증가함수이므로 $A(u)$ 는 적어도 x 보다 큰 원소들만 가질 수 있다. 따라서 $\inf A(u) \geq x$ 가 성립한다. 만약 $\inf A(u) = x$ 라고 가정하면, F 의 오른쪽 연속성에 의해 임의의 $\epsilon > 0$ 에 대하여 $\delta > 0$ 이 존재하여 $0 < h < \delta \Longrightarrow F(x+h) < F(x) + \epsilon$ 이 성립하고 그 δ 에 대하여 하한의 성질에 의해 $z \in (x, x+\delta)$ 가 존재하여 $z \in A(u)$ 가 성립한다. 즉,

$$u \leq F(z) < F(x) + \epsilon$$

정리하면 임의의 $\epsilon > 0$ 에 대하여 $u < F(x) + \epsilon$ 이 성립하므로 $u \leq F(x)$. 이는 처음에 $u \leq F(x)$ 를 부정한 것에 모순이다. 따라서 $u > F(x) \Longrightarrow \inf A(u) > x$.
Note. $X = F^{-1}(U)$ 일 때 X 의 cdf가 F 인 것은 실제로 컴퓨터에서 난수 생성을 통해 cdf를 알고 있는 어느 분포로부터 sampling을 하는 과정에서 사용하는 사실이다.

2 평균이 존재하지 않는 분포

다음과 같은 pdf를 가지는 분포를 (표준)Cauchy분포라고 한다.

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Cauchy분포는 pdf가 even function($f(-x) = f(x)$) 임에도 불구하고 평균이 존재하지 않는다. $\int_{-\infty}^{\infty} xf(x)dx$ 의 이상적분이 정의되지 않기 때문이다. $x \geq 1$ 일 때 $\frac{x}{1+x^2} \geq \frac{1}{2x}$ 가 성립하므로

$$\int_1^\infty xf(x)dx \geq \frac{1}{\pi}\int_1^\infty \frac{1}{2x} = \infty$$

인 까닭이다.

3 모든 차수의 적률은 존재하지만 적률생성함수는 존재하지 않는 분포

다음과 같은 pdf를 가지는 분포를 (표준)로그정규분포라고 한다.

$$f(x) = \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\log x)^2\right) \mathrm{I}_{(0,\infty)}(x)$$

(X 가 표준로그정규분포를 따르는 것의 다른 정의는 $\log X$ 가 표준정규분포를 따르는 것이다.)
표준로그정규분포는 임의의 $m = 1, 2, \cdots$ 에 대하여 m차 적률 $\mathbb{E}(X^m)$ 이 잘 정의되지만 적률생성함수를 가지지는 않는다. 여기서 적률생성함수를 가진다는 말은

$$\mathbb{E}(e^{tX}) < +\infty \qquad\qquad -\epsilon < t < \epsilon$$

를 만족하는 $\epsilon > 0$ 이 존재한다는 것으로 해석하면 된다. 임의의 $m = 1, 2, \cdots$ 에 대하여

$$\begin{aligned}
\mathbb{E}(X^m) = \int_0^\infty x^m f(x)dx &= \int_0^\infty \frac{x^m}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\log x)^2\right) dx \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(my - \frac{1}{2}y^2\right) dy \qquad (y = \log x) \\
&= \exp\left(\frac{m^2}{2}\right) \int_{-\infty}^\infty \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y-m)^2\right)}_{\text{평균이 } m \text{ 이고 분산이 1 인 정규분포의 pdf}} dy \\
&= \exp\left(\frac{m^2}{2}\right)
\end{aligned}$$

임을 확인할 수 있다. 이 논의는 $m = 0$ 일 때도 성립하므로 f 가 적절한 pdf임을 확인할 수 있다. ($\int_0^\infty f(x)dx = 1$)
한편 $t > 0$ 에 대하여

$$\mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} f(x)dx = \int_0^\infty \frac{e^{tx}}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\log x)^2\right) dx = \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left(tx - \log x - \frac{1}{2}(\log x)^2\right) dx$$

이고 $x > M \Longrightarrow tx - \log x - \frac{1}{2}(\log x)^2 > \frac{t}{2}x$ 을 만족하는 $M > 0$ 이 존재하는데

$$\int_M^\infty \exp\left(tx - \log x - \frac{1}{2}(\log x)^2\right) dx \geq \int_M^\infty \exp\left(\frac{t}{2}x\right) dx = \infty$$

이므로 $\mathbb{E}(e^{tX}) = \infty$ 이다.

4 모든 차수의 적률을 공유하지만 서로 일치하지는 않는 분포

앞서 살펴본 로그정규분포에 λ 만큼의 perturbation을 가한 분포를 생각하자.

$$f(x;\lambda) = \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\log x)^2\right)(1 + \lambda \sin(4\pi \log x))\, \mathrm{I}_{(0,\infty)}(x) \qquad -1 < \lambda < 1$$

먼저 $1 + \lambda \sin(4\pi \log x)$ 는 반드시 양의 값을 가지므로 $f(x;\lambda)$ 는 x, λ 에 관계 없이 항상 양의 값을 가진다. (분포의 support인 $(0,\infty)$ 내에서) $m = 0, 1, 2, \cdots$ 에 대하여

$$\begin{aligned}\int_0^\infty \frac{x^m}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\log x)^2\right) \sin(4\pi \log x) dx &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(my - \frac{1}{2}y^2\right) \sin(4\pi y) dy \qquad (y = \log x) \\ &= \exp\left(\frac{m^2}{2}\right) \int_{-\infty}^\infty \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \sin(4\pi z)}_{h(z)}\, dz \quad (z = y - m)\end{aligned}$$

가 성립하는데 $h(z)$ 는 odd function($h(-z) = -h(z)$) 이고

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

이므로 $\int_{-\infty}^\infty h(z) = 0$ 이 성립한다. 따라서 λ 의 값에 무관하게 $\int_0^\infty f(x;\lambda) dx = 1$ 이며, $\mathbb{E}(X^m) = \exp\left(\frac{m^2}{2}\right)$ 임을 알 수 있다. 따라서 모든 차수의 적률이 같은 두 분포는 서로 다른 분포일 수가 있다.

1 공분산의 이해

$\mathrm{Corr}(X,Y) = \mathrm{Corr}(Y,Z) = 0.8$ 일 때 $\mathrm{Corr}(X,Z)$ 의 값으로 가능한 범위는?

2 고차원 확률벡터

n 차원 확률벡터 $\mathbf{X}$ 의 기대값과 분산행렬이 각각 $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}, \mathrm{Var}(\mathbf{X}) = \Sigma$ 로 주어진다. $n \times n$ 대칭행렬 $\mathbf{A}$ 에 대하여 다음을 보이시오.

$$\mathbb{E}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) = \mathrm{tr}(\mathbf{A}\Sigma) + \boldsymbol{\mu}^\top \mathbf{A}\boldsymbol{\mu}$$

3 조건부확률밀도함수

양의 상수 a, b, c 에 대하여 확률변수 X_1, X_2, X_3 의 결합확률밀도함수는 다음과 같이 주어진다.

$$f_{1,2,3}(x,y,z) = Ke^{-(ax+by+cz)}\mathrm{I}_{(0<x<y<z<\infty)}$$

양의 상수 K 의 값을 a, b, c 로 표현하고, 조건부확률밀도함수 $f_{2,3|1}(y,z|x)$ 를 구하시오.

4 베이즈정리

조건부확률밀도함수에 대한 베이즈정리는 다음과 같이 기술된다.

$$\begin{aligned} f_{1|2}(x|y) &= \frac{f_1(x)f_{2|1}(y|x)}{\sum_x f_1(x)f_{2|1}(y|x)} && (X_1 \text{ 이산}) \\ &= \frac{f_1(x)f_{2|1}(y|x)}{\int f_1(x)f_{2|1}(y|x)dx} && (X_1 \text{ 연속}) \end{aligned}$$

(a) 양의 상수 α, β 에 대하여

$$\begin{aligned} f_1(x) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\mathrm{I}_{(0,1)}(x) \\ f_{2|1}(y|x) &= \binom{100}{y}x^y(1-x)^{100-y}\mathrm{I}_{\{0,1,\cdots,100\}}(y) \end{aligned}$$

일 때 $f_{1|2}(x|y)$ 를 구하시오. *Note*. 감마함수 Γ 에 친숙하지 않다면, $\alpha = \beta = 1$ 로 놓고 풀어도 좋습니다.
(b) 양의 상수 α, β 에 대하여

$$\begin{aligned} f_1(x) &= \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}\mathrm{I}_{(0,\infty)}(x) \\ f_{2|1}(y|x) &= \frac{x^y e^{-x}}{y!}\mathrm{I}_{\{0,1,\cdots\}}(y) \end{aligned}$$

일 때 $f_{1|2}(x|y)$ 를 구하시오. *Hint*. (a), (b) 모두 $\int f_1(x)dx = 1$ 을 받아들인다면, 암산으로 풀 수 있습니다.

1

$\mathrm{Corr}(X,Z)=\rho$라고 하자. 임의의 실수 a,b,c에 대하여

$$\mathrm{Var}(aX+bY+cZ)=\begin{bmatrix}a & b & c\end{bmatrix}\begin{bmatrix}\mathrm{Sd}(X) & & \\ & \mathrm{Sd}(Y) & \\ & & \mathrm{Sd}(Z)\end{bmatrix}\begin{bmatrix}1 & 0.8 & \rho \\ 0.8 & 1 & 0.8 \\ \rho & 0.8 & 1\end{bmatrix}\begin{bmatrix}\mathrm{Sd}(X) & & \\ & \mathrm{Sd}(Y) & \\ & & \mathrm{Sd}(Z)\end{bmatrix}\begin{bmatrix}a \\ b \\ c\end{bmatrix}\geq 0$$

이 성립하고 $\mathrm{Sd}(X),\mathrm{Sd}(Y),\mathrm{Sd}(Z)$는 0이 아니므로 상관계수행렬은 p.s.d.여야 한다. 이는 Sylvester's criterion에 의하여

$$\det\begin{bmatrix}1 & 0.8 \\ 0.8 & 1\end{bmatrix}\geq 0 \qquad \det\begin{bmatrix}1 & \rho \\ \rho & 1\end{bmatrix}\geq 0 \qquad \det\begin{bmatrix}1 & 0.8 & \rho \\ 0.8 & 1 & 0.8 \\ \rho & 0.8 & 1\end{bmatrix}\geq 0$$

임과 동치이고, 계산하면 $0.28\leq\rho\leq 1$을 얻는다.
Sylvester's criterion을 우회한 풀이방법도 있다. *($\rho=0.28$의 bound를 얻기 위해서는 상관계수행렬의 행렬식 값이 0 이어야 하고, 다시 말해 0 을 고유값으로 가져야 한다. $\rho=0.28$ 인 경우 고유벡터 $(1,-1.6,1)^\top$ 을 얻으니까...)*

$$\begin{bmatrix}1 & -1.6 & 1\end{bmatrix}\begin{bmatrix}1 & 0.8 & \rho \\ 0.8 & 1 & 0.8 \\ \rho & 0.8 & 1\end{bmatrix}\begin{bmatrix}1 \\ -1.6 \\ 1\end{bmatrix}=2(\rho-0.28)\geq 0$$

으로부터 $\rho\geq 0.28$을 얻고,

$$\begin{bmatrix}1 & 0 & -1\end{bmatrix}\begin{bmatrix}1 & 0.8 & \rho \\ 0.8 & 1 & 0.8 \\ \rho & 0.8 & 1\end{bmatrix}\begin{bmatrix}1 \\ 0 \\ -1\end{bmatrix}=2(1-\rho)\geq 0$$

으로부터 $\rho\leq 1$을 얻는다. Sylvester's criterion 없이 $0.28\leq\rho\leq 1$을 유도하였다.
실례를 만드는 방법은 다음과 같다. $0.28\leq\rho\leq 1$에 대하여

$$\overline{\rho}=\frac{\rho-0.64}{0.36}$$

으로 정의하면 $-1\leq\overline{\rho}\leq 1$이 성립하고 확률변수 $Y,\varepsilon,\delta\overset{\text{iid}}{\sim}N(0,1)$에 대하여

$$\begin{aligned}X&=0.8Y+0.6\varepsilon\\ Z&=0.8Y+0.6\left(\overline{\rho}\varepsilon+\sqrt{1-\overline{\rho}^2}\delta\right)\end{aligned}$$

로 정의하면 문제의 ρ에 대한 실례가 됨을 확인하여라.
질문: $\overline{\rho}=\pm 1$ 인 경우는 어떻게 되는가?
질문 하나 더: $\mathrm{Corr}(X,Z|Y)$ 를 정의할 수 있겠는가?

2

다음과 같은 trace trick은 아주, 아주 유용하다. 가령, 고차원 회귀분석에서 σ^2 의 unbiased estimator를 구할 때 다시 등장할 것이다.

$$\begin{aligned}
\mathbb{E}(\mathbf{X}^\top\mathbf{A}\mathbf{X}) &= \mathbb{E}(\mathrm{tr}(\mathbf{X}^\top\mathbf{A}\mathbf{X})) \\
&= \mathbb{E}(\mathrm{tr}(\mathbf{A}\mathbf{X}\mathbf{X}^\top)) && (\mathrm{tr}(PQ) = \mathrm{tr}(QP)) \\
&= \mathrm{tr}(\mathbb{E}(\mathbf{A}\mathbf{X}\mathbf{X}^\top)) \\
&= \mathrm{tr}(\mathbf{A}\mathbb{E}(\mathbf{X}\mathbf{X}^\top)) \\
&= \mathrm{tr}(\mathbf{A}(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^\top)) && (\Sigma = \mathbb{E}(\mathbf{X}\mathbf{X}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top) \\
&= \mathrm{tr}(\mathbf{A}\Sigma) + \mathrm{tr}(\mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}^\top) \\
&= \mathrm{tr}(\mathbf{A}\Sigma) + \mathrm{tr}(\boldsymbol{\mu}^\top\mathbf{A}\boldsymbol{\mu}) \\
&= \mathrm{tr}(\mathbf{A}\Sigma) + \boldsymbol{\mu}^\top\mathbf{A}\boldsymbol{\mu}
\end{aligned}$$

3

$$\begin{aligned}
1 &= \int_0^\infty \int_x^\infty \int_y^\infty f_{1,2,3}(x,y,z)dzdydx \\
&= \int_0^\infty \int_x^\infty \int_y^\infty Ke^{-ax-by-cz}dzdydx \\
&= \int_0^\infty \int_x^\infty \frac{K}{c}e^{-ax-by-cy}dzdydx \\
&= \int_0^\infty \underbrace{\frac{K}{c(b+c)}e^{-ax-bx-cx}}_{(\bigstar)}dx \\
&= \frac{K}{c(b+c)(a+b+c)}
\end{aligned}$$

로부터

$$K = c(b+c)(a+b+c)$$

이고 $(\bigstar) = \int_x^\infty \int_y^\infty f_{1,2,3}(x,y,z)dzdy$ 와 주변확률밀도함수 f_1 의 정의로부터

$$f_1(x) = \frac{K}{c(b+c)}e^{-(a+b+c)x}$$

이므로 조건부확률밀도함수 $f_{2,3|1}$ 의 정의로부터

$$\begin{aligned} f_{2,3|1}(y,z|x) &= \frac{f_{1,2,3}(x,y,z)}{f_1(x)} \\ &= \frac{Ke^{-(ax+by+cz)}}{\frac{K}{c(b+c)}e^{-(a+b+c)x}}\mathrm{I}_{(x<y<z<\infty)} \\ &= c(b+c)\exp\left(-b(y-x)-c(z-x)\right)\mathrm{I}_{(x<y<z<\infty)} \qquad \text{(defined assuming } x>0) \end{aligned}$$

을 얻는다.

4

(a) 베이즈정리를 적용하면

$$\begin{aligned} f_{1|2}(x|y) &= \frac{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\binom{100}{y}x^y(1-x)^{100-y}}{\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\binom{100}{y}x^y(1-x)^{100-y}dx}\mathrm{I}_{(0,1)}(x) \\ &= \frac{x^{\alpha+y-1}(1-x)^{\beta+100-y-1}}{\int_0^1 x^{\alpha+y-1}(1-x)^{\beta+100-y-1}dx}\mathrm{I}_{(0,1)}(x) \end{aligned}$$

여기서 중요한 것은 x가 포함되어 있지 않은 항은 분자, 분모에서 cancel된다는 것이다. (즉, y는 상수취급한다.) 더 나아가서 분모를 고려할 필요도 없다. 분모는 x에 의존하지 않는 상수항이며, x의 support 내에서 적분한 결과가 1이 되도록 맞춰주는 normalization의 역할만 하기 때문이다. 그리고 베이즈정리로 얻은 조건부확률밀도함수 $f_{1|2}$의 식을 보면 원래 f_1의 식에서 모수 α, β 자리에 $\alpha+y, \beta+100-y$를 넣은 꼴임을 알 수 있다. 따라서 normalization constant를 직접 계산할 필요가 없으며,

$$f_{1|2}(x|y) = \frac{\Gamma(\alpha+\beta+100)}{\Gamma(\alpha+y)\Gamma(\beta+100-y)}x^{\alpha+y-1}(1-x)^{\beta+100-y-1}\mathrm{I}_{(0,1)}(x)$$

으로 주어짐을 바로 알 수 있다. 참고로 이 분포의 이름은 $\mathrm{Beta}(\alpha+y, \beta+100-y)$이다.
(b) 베이즈정리를 적용하되 x가 포함된 항만 남기고 나머지는 cancel하면

$$\begin{aligned} f_{1|2}(x|y) &= \frac{x^{\alpha-1}e^{-x/\beta}x^y e^{-x}}{\int_0^\infty x^{\alpha-1}e^{-x/\beta}x^y e^{-x}dx}\mathrm{I}_{(0,\infty)}(x) \\ &= \frac{x^{\alpha+y-1}e^{-\left(\frac{1}{\beta}+1\right)x}}{\int_0^\infty x^{\alpha+y-1}e^{-\left(\frac{1}{\beta}+1\right)x}dx}\mathrm{I}_{(0,\infty)}(x) \end{aligned}$$

으로 주어지는데, 이는 원래 f_1의 식에서 모수 α, β 자리에 $\alpha+y, \frac{\beta}{\beta+1}$을 넣은 꼴이다. 따라서

$$f_{1|2}(x|y) = \frac{1}{\Gamma(\alpha+y)}\left(\frac{\beta+1}{\beta}\right)^{\alpha+y}x^{\alpha+y-1}e^{-\left(\frac{1}{\beta}+1\right)x}\mathrm{I}_{(0,\infty)}(x)$$

참고로 이 분포의 이름은 $\mathrm{Gamma}(\alpha+y, \frac{\beta}{\beta+1})$이다.

꼭 증명할 줄 알아야 하는 정리들

정리 2.3.4

(a) $\mathbb{E}(|Y|) < +\infty$ 이면 $\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}(Y)$
(b) $\mathbb{E}(Y^2) < +\infty, \mathbb{E}(v^2(X)) < +\infty$ 이면 함수 $v : \mathbb{R} \to \mathbb{R}$ 의 형태에 관계없이 $\mathrm{Cov}(Y - \mathbb{E}(Y|X), v(X)) = 0$
Note1. (a)를 Law of iterated expectations 혹은 Law of total expectation이라고 부른다.
Note2. (b)에서 $Y - \mathbb{E}(Y|X)$ 의 expectation은 항상 0 이므로 공분산이 0 이라는 것은 사실 곱의 expectation이 0 이라는 것과 같은 의미이다.

정리 2.3.5

($\mathbb{E}(Y^2) < +\infty$ 일 때) 모든 함수 $u : \mathbb{R} \to \mathbb{R}$ 중에서 $\mathbb{E}[(Y - u(X))^2]$ 을 최소로 하는 것은 $u(X) = \mathbb{E}(Y|X)$ 이다. 보다 구체적으로, 다음이 성립한다.

$$\mathbb{E}[(Y - \mathbb{E}(Y|X))^2] \le \mathbb{E}[(Y - u(X))^2] \qquad \text{(for all } u\text{)}$$
$$\underset{①}{\mathbb{E}}[(Y - \mathbb{E}(Y|X))^2] = \underset{②}{\mathbb{E}}[\mathrm{Var}(Y|X)]$$

Note. ①은 X, Y 의 joint distribution에 걸린 기대값이고, ②는 X 의 marginal distribution에 걸린 기대값이라고 생각하면 해석이 쉽다.

정리 2.3.6

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}[\mathbb{E}(Y|X)]$$

Note. 이를 Law of total variance라고 부른다.

확률변수를 벡터로 바라보는 시각

확률공간 $(S, \mathcal{F}, \mathbb{P})$ 가 주어져 있을 때 확률변수는 $X : S \to \mathbb{R}$ 꼴의 함수로 정의된다. 즉 모든 확률변수를 모은 집합을 $S \to \mathbb{R}$ 함수들의 집합인 $\mathbb{R}^S$ 로 쓸 수 있다. 다행히도, 수리통계 1에서 다루는 대부분의 확률변수들은 "예쁜" 함수들이다. 양수 m 에 대하여

$$L^m(\mathbb{P}) = \left\{X \in \mathbb{R}^S : \mathbb{E}(|X|^m) < +\infty\right\}$$

을 정의하면 정리 1.6.2(리아푸노프 부등식)에 의하여 $0 < r < s$ 이면 $L^r(\mathbb{P}) \supseteq L^s(\mathbb{P})$ 임을 알 수 있다. 따라서 다음과 같은 집합의 서열을 생각할 수 있다.

$$L^1(\mathbb{P}) \supseteq L^2(\mathbb{P}) \supseteq \cdots \supseteq L^\infty(\mathbb{P}) \supseteq \left\{X \in \mathbb{R}^S : \exists \epsilon > 0, \forall t \in (-\epsilon, \epsilon), \mathbb{E}(e^{tX}) < +\infty\right\}$$

여기서 $L^\infty(\mathbb{P}) = \cap_{m=1}^\infty L^m(\mathbb{P})$ 는 모든 차수의 moment(적률)가 존재하는 확률변수들의 집합으로 생각할 수 있고, 마지막 집합은 적률생성함수가 존재하는 확률변수들의 집합이라고 생각하면 된다. 지난 주에 살펴보았듯이, 적률생성함수가 존재할 경우 모든 차수의 moment가 존재하지만, 그 역은 성립하지 않는다. (cf. 로그정규분포)
이제 $m = 2$ 인 경우만 생각하자. 즉 $V = L^2(\mathbb{P})$ 라고 놓자. V 가 vector space임은 쉽게 확인할 수 있다.

Inner Product Space

$V = L^2(\mathbb{P})$는 vector space일 뿐만 아니라 inner product space가 된다. 다음과 같이 내적을 정의할 수 있다.

$$\langle X, Y \rangle_V = \mathbb{E}(XY)$$

이 내적이 잘 정의되는 이유는 $X, Y \in V$에 대하여 $\mathbb{E}(X^2), \mathbb{E}(Y^2) < +\infty$인데 $|XY| \le \frac{1}{2}(X^2 + Y^2)$이 성립하므로 $\mathbb{E}(|XY|) < +\infty$이기 때문이다. $V-$norm은 다음과 같이 정의된다.

$$\|X\|_V^2 = \langle X, X \rangle_V = \mathbb{E}(X^2)$$

함수 $\mathbf{1} : S \to \mathbb{R}$을 $\forall s \in S, \mathbf{1}(s) = 1$로 정의하면 $\mathbf{1}$은 확률변수일 뿐만 아니라 V의 원소가 된다. 즉 벡터이다. 지난 주에는 정리 2.3.5의 쉬운 버전인 연습문제 1.13을 살펴보았다:
($Y \in V$ 일 때) 모든 $u \in \mathbb{R}$ 중에서 $\mathbb{E}[(Y-u)^2]$을 최소로 하는 것은 $u = \mathbb{E}(Y)$이다. 보다 구체적으로,

$$\mathbb{E}[(Y - \mathbb{E}(Y))^2] \le \mathbb{E}[(Y-u)^2] \qquad \text{(for all } u \in \mathbb{R})$$
$$\mathbb{E}[(Y - \mathbb{E}(Y))^2] = \mathrm{Var}(Y)$$

이 성립한다. 이것은 inner product space V의 언어로 해석하면

$$\|Y - \mathbb{E}(Y)\mathbf{1}\|_V^2 \le \|Y - u\mathbf{1}\|_V^2 \qquad \text{(for all } u \in \mathbb{R})$$
$$\|Y - \mathbb{E}(Y)\mathbf{1}\|_V^2 = \mathrm{Var}(Y)$$

추가적으로 $\langle Y - \mathbb{E}(Y)\mathbf{1}, \mathbf{1} \rangle_V = \mathbb{E}(Y) - \mathbb{E}(Y) = 0$인 것도 금방 확인할 수 있다. 따라서 V 공간 상의 벡터 Y에서 $\mathbf{1}$에 내린 수선의 발이 $\mathbb{E}(Y)\mathbf{1}$이며, 그 수선의 길이의 제곱이 $\mathrm{Var}(Y)$라고 말할 수 있다. 이제 $X \in V$를 고정하고 다음과 같은 집합을 생각하자.

$$\widetilde{X} = \{Z \in V : \exists v : \mathbb{R} \to \mathbb{R}, Z = v \circ X\}$$

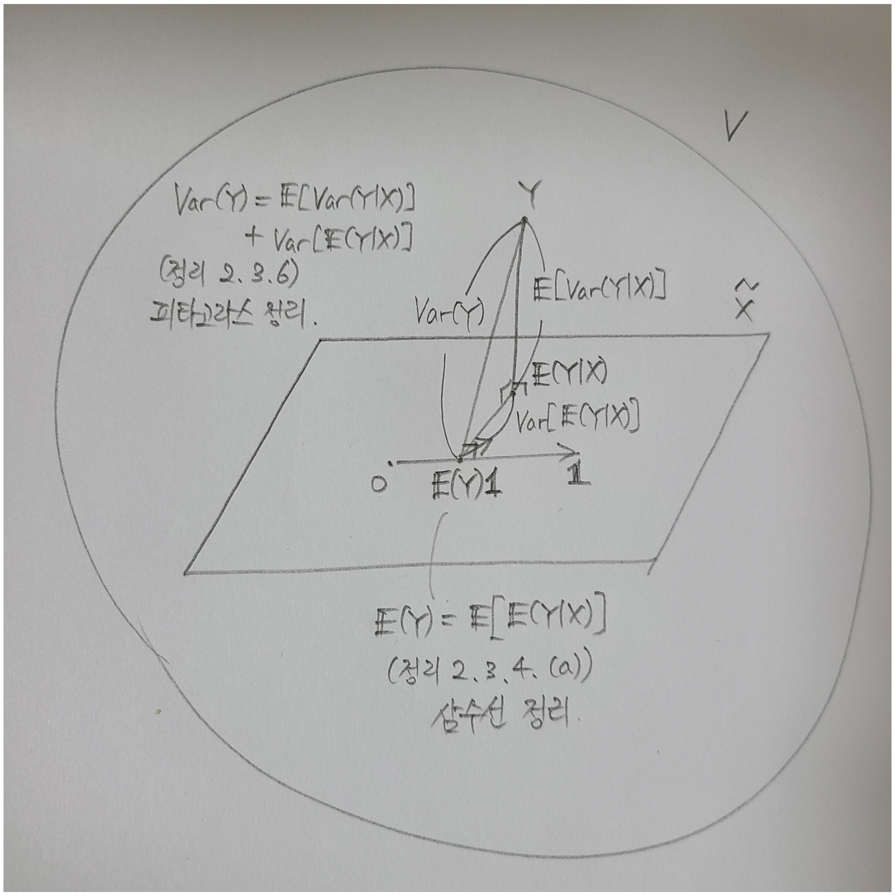
여기서 $Z = v \circ X$는 $s \in S$에 대하여 $Z(s) = v(X(s))$가 성립한다는 뜻으로 해석할 수 있다. 이것을 수리통계학 교과서에서는 $Z = v(X)$로 나타내고는 한다. v는 $\mathbb{R} \to \mathbb{R}$의 함수이므로 혼동하지 않도록 하자. 정의 상 $\widetilde{X}$는 V의 부분집합인데, 사실 부분공간이기도 하다. V의 언어로 해석하면 정리 2.3.4의 (b)는

$$Y \in V, Z \in \widetilde{X} \Longrightarrow \langle Y - \mathbb{E}(Y|X), Z \rangle_V = 0$$

으로, 정리 2.3.5는

$$Y \in V, Z \in \widetilde{X} \Longrightarrow \begin{cases} \|Y - \mathbb{E}(Y|X)\|_V^2 \le & \|Y - Z\|_V^2 \\ \|Y - \mathbb{E}(Y|X)\|_V^2 = & \mathbb{E}[\mathrm{Var}(Y|X)] \end{cases}$$

으로 이해할 수 있다. 따라서 V 공간 상의 벡터 Y에서 $\widetilde{X}$에 내린 수선이 발이 $\mathbb{E}(Y|X)$이며, 그 수선의 길이의 제곱이 $\mathbb{E}[\mathrm{Var}(Y|X)]$라고 말할 수 있다.
지금까지의 story를 따르면 inner product space V에서 정리 2.3.4는 삼수선 정리, 정리 2.3.6은 피타고라스 정리에 불과하다. (다음 쪽의 그림 참조) 이러한 story를 알고 있으면 책에 나와 있는 정리들의 증명 과정에 어떤 직관이 담겨 있는지 알 수 있을 것이다. *책의 정리들을 보지 않고 서술하고 증명해보자.*



예제

Law of iterated expectations는 앞으로 만나게 될 일이 매우 많을 것이다. 가능한 하나의 예제는 다음과 같다: 확률변수 $Y(1), Y(0), Z, X$ 를 생각하자. 여기서 Z 는 binary variable로 $0, 1$ 의 값만을 가진다.

$$Y = ZY(1) + (1 - Z)Y(0)$$
$$\pi(X) = \mathbb{P}(Z = 1|X)$$

을 정의하자. $\pi(X) > 0$ for all X 및 Conditional independence

$$(Y(1), Y(0)) \perp\!\!\!\perp Z|X$$

를 가정하고

$$\mathbb{E}\left[\frac{ZY}{\pi(X)}\right] = \mathbb{E}(Y(1))$$

임을 증명하여라.
주의: $Y \perp\!\!\!\perp Z|X$ 는 참이 아니다. 풀이는 다음 쪽에

$$\begin{aligned}
\mathbb{E}\left[\frac{ZY}{\pi(X)}\right] &= \mathbb{E}\left[\mathbb{E}\left(\frac{ZY}{\pi(X)}\middle| X\right)\right] && \text{(Law of iterated expectations)}\\
&= \mathbb{E}\left[\frac{1}{\pi(X)}\mathbb{E}\left(ZY|X\right)\right] && (\mathbb{E}(\cdot|X) \text{ 내부에서 } X \text{는 상수취급})\\
&= \mathbb{E}\left[\frac{1}{\pi(X)}\mathbb{E}\left(ZY(1)|X\right)\right] && (ZY = ZY(1))\\
&= \mathbb{E}\left[\frac{1}{\pi(X)}\mathbb{E}\left(Z|X\right)\mathbb{E}\left(Y(1)|X\right)\right] && (Z \perp\!\!\!\perp Y(1)|X.\ \text{정리 2.4.3의 응용})\\
&= \mathbb{E}\left[\mathbb{E}\left(Y(1)|X\right)\right] && (Z\text{는 binary이므로 } \mathbb{E}(Z|X) = \mathbb{P}(Z=1|X))\\
&= \mathbb{E}(Y(1)) && \text{(Law of iterated expectations)}
\end{aligned}$$

보충 문제: joint mgf (결합적률생성함수)

시간의 부족으로 수업 시간에 joint mgf는 언급하지 못했습니다. 정리 2.4.1을 읽고 다음을 보여라:
양의 상수 a, b, c에 대하여 확률변수 X_1, X_2, X_3의 결합확률밀도함수는 다음과 같이 주어진다.

$$f_{1,2,3}(x,y,z) = Ke^{-(ax+by+cz)}\mathrm{I}_{(0<x<y<z<\infty)}$$
$$K = c(b+c)(a+b+c)$$

이때 $X_1, X_2 - X_1, X_3 - X_2$는 모두 독립임을 보여라.
Comment. $X_1, X_2 - X_1, X_3 - X_2$의 joint mgf를 계산하여라.

pdf를 작성할 때 support를 반드시 명시해야 한다. 가령 이항분포의 pdf로 올바른 표현은 $\binom{n}{x}p^x(1-p)^{n-x}\mathrm{I}_{\{0,1,\cdots,n\}}(x)$ 이다.

name	notation	support	probability density function	moment generating function
이항분포 Binomial	$\mathrm{B}(n,p)$ $0 \le p \le 1$	$\{0,1,\cdots,n\}$	$\binom{n}{x}p^x q^{n-x}$ where $q=1-p$	$(pe^t+q)^n$ for $t \in \mathbb{R}$
음이항분포 Neg. Bin.	$\mathrm{Negbin}(r,p)$ $0 \le p \le 1$	$\{r, r+1, \cdots\}$	$\binom{x-1}{r-1}p^r q^{x-r}$ where $q=1-p$	$\left(\frac{pe^t}{1-qe^t}\right)^r$ for $t < -\log q$
포아송분포 Poisson	$\mathrm{Poisson}(\lambda)$ $\lambda \ge 0$	$\{0,1,\cdots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	$e^{\lambda(e^t-1)}$ for $t \in \mathbb{R}$
다항분포 Multinomial	$\mathrm{Multi}(n,(p_1,\cdots,p_k)^\top)$ $\sum_{j=1}^k p_j = 1, p_j \ge 0$	①	$\frac{n!}{x_1!\cdots x_k!}p_1^{x_1}\cdots p_k^{x_k}$	$\left(\sum_{j=1}^k p_j e^{t_j}\right)^n$ for $t_j \in \mathbb{R}$
감마분포 Gamma	$\mathrm{Gamma}(\alpha,\beta)$ $\alpha,\beta>0$	$(0,\infty)$	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}$	$(1-\beta t)^{-\alpha}$ for $t < \frac{1}{\beta}$
베타분포 Beta	$\mathrm{Beta}(\alpha,\beta)$ $\alpha,\beta>0$	$(0,1)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\mathbb{E}(X^k)=\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$ mgf exists for $t \in \mathbb{R}$
베타이항분포 Beta Bin.	$\mathrm{Betabin}(n,\alpha,\beta)$ $\alpha,\beta>0$	$\{0,1,\cdots,n\}$	$\binom{n}{x}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+x)\Gamma(\beta+n-x)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)}$	mgf exists for $t \in \mathbb{R}$
역감마분포 Inv. Gamma	$\mathrm{invGamma}(\alpha,\beta)$ $\alpha,\beta>0$	$(0,\infty)$	$\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{-\alpha-1}e^{-\frac{1}{\beta x}}$	mgf does not exist
로지스틱분포 Logistic	$\mathrm{L}(\mu,\sigma)$ $\mu \in \mathbb{R}, \sigma>0$	$(-\infty,\infty)$	$\frac{e^{-z}}{\sigma(1+e^{-z})^2}$ where $z=\frac{x-\mu}{\sigma}$	$e^{\mu t}\Gamma(1-\sigma t)\Gamma(1+\sigma t)$ for $-\frac{1}{\sigma}<t<\frac{1}{\sigma}$
정규분포 Normal	$\mathrm{N}(\mu,\sigma^2)$ $\mu \in \mathbb{R}, \sigma>0$	$(-\infty,\infty)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$e^{\mu t+\frac{1}{2}\sigma^2 t^2}$ for $t \in \mathbb{R}$
로그정규분포 Log Norm.	$\mathrm{Lognormal}(\mu,\sigma^2)$ $\mu \in \mathbb{R}, \sigma>0$	$(0,\infty)$	$\frac{1}{x\sqrt{2\pi}\sigma}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)$	$\mathbb{E}(X^k)=e^{k\mu+\frac{1}{2}k^2\sigma^2}$ HOWEVER mgf does not exist
$t-$분포 Student's t	t_ν $\nu>0$	$(-\infty,\infty)$	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	mgf does not exist
웨이불분포 Weibull	$\mathrm{Weibull}(\alpha,\beta)$ $\alpha,\beta>0$	$(0,\infty)$	$\frac{\alpha}{\beta^\alpha}x^{\alpha-1}e^{-(x/\beta)^\alpha}$	$\mathbb{E}(X^k)=\beta^k\Gamma(1+\frac{k}{\alpha})$ mgf exists if $\alpha \ge 1$

질문. 베르누이(Bernoulli), 기하(Geometric), 지수(Exponential), 카이제곱(χ^2), 코시(Cauchy)분포를 이 표에서 찾을 수 있겠는가?

Notes

Parameters

$$f(x;\mu,\sigma)=\frac{1}{\sigma}f_0\left(\frac{x-\mu}{\sigma}\right)$$

꼴이면, μ,σ를 각각 location, scale parameter라고 부른다. 꼭 평균, 표준편차일 필요는 없다. θ^{-1}가 scale parameter인 경우 보통 θ를 rate parameter라고 부른다. 나머지 경우 일반적으로 shape parameter라고 부른다.

Gamma Integral

$\alpha>0$에 대하여 감마함수는 다음과 같이 정의된다.

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx$$

먼저 부분적분을 통해 $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$를 보일 수 있다:

$$\begin{aligned}\Gamma(\alpha+1)&=\int_0^\infty x^{\alpha}e^{-x}dx\\&=\left[x^\alpha e^{-x}\right]_\infty^0+\int_0^\infty \alpha x^{\alpha-1}e^{-x}dx=\alpha\Gamma(\alpha)\end{aligned}$$

$\Gamma(1)=1$이므로 $n=0,1,\cdots$ 일 때 $\Gamma(n+1)=n!$ 임을 알 수 있다.

Beta Integral

$\alpha,\beta>0$에 대하여

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta)&=\int_0^\infty x^{\alpha-1}e^{-x}dx\int_0^\infty y^{\beta-1}e^{-y}dy\\&=\int_0^\infty\int_0^\infty x^{\alpha-1}y^{\beta-1}e^{-x-y}dxdy\end{aligned}$$

여기서 $x=zw, y=z(1-w)$ 치환하면 $\dfrac{\partial(x,y)}{\partial(z,w)}=\begin{vmatrix}w & z\\ 1-w & -z\end{vmatrix}=-z$이므로

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta)&=\int_0^1\int_0^\infty (zw)^{\alpha-1}(z(1-w))^{\beta-1}e^{-z}zdzdw\\&=\int_0^1 w^{\alpha-1}(1-w)^{\beta-1}dw\underbrace{\int_0^\infty z^{\alpha+\beta-1}e^{-z}dz}_{=\Gamma(\alpha+\beta)}\end{aligned}$$

따라서 $\alpha,\beta>0$에 대하여 베타함수를 $\mathrm{B}(\alpha,\beta)=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$라고 정의하면

$$\mathrm{B}(\alpha,\beta)=\int_0^1 w^{\alpha-1}(1-w)^{\beta-1}dw$$

가 성립한다. *질문.* $\Gamma(\frac{1}{2})=\sqrt{\pi}$임을 보일 수 있겠는가? *Hint.* $\theta\in[0,\frac{\pi}{2}]$에 대하여 $w=\sin^2\theta$.

Derivation

$X \sim \mathrm{B}(n,p)$ 이면 $t \in \mathbb{R}$ 에 대하여

$$\mathbb{E}(e^{tX}) = \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x q^{n-x} = \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x q^{n-x} = (pe^t+q)^n$$

$X \sim \mathrm{Negbin}(r,p)$ 이면 $qe^t < 1$ 일 때

$$\mathbb{E}(e^{tX}) = \sum_{x=r}^{\infty} e^{tx}\binom{x-1}{r-1}p^r q^{x-r} = \sum_{x=r}^{\infty}\binom{x-1}{r-1}(pe^t)^r(qe^t)^{x-r} = \left(\frac{pe^t}{1-qe^t}\right)^r \underbrace{\sum_{x=r}^{\infty}\binom{x-1}{r-1}(1-qe^t)^r(qe^t)^{x-r}}_{=1}$$

$X \sim \mathrm{Poisson}(\lambda)$ 이면 $t \in \mathbb{R}$ 에 대하여

$$\mathbb{E}(e^{tX}) = \sum_{x=0}^{\infty} e^{tx}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = e^{-\lambda+\lambda e^t}$$

$X \sim \mathrm{Multi}(n,(p_1,\cdots,p_k)^\top)$ 이면 $t_1,\cdots,t_k \in \mathbb{R}$ 에 대하여

$$\begin{aligned}\mathbb{E}(e^{t_1X_1+\cdots+t_kX_k}) &= \sum_{x_1+\cdots+x_k=n} e^{t_1x_1+\cdots+t_kx_k}\frac{n!}{x_1!\cdots x_k!}p_1^{x_1}\cdots p_k^{x_k} \\ &= \sum_{x_1+\cdots+x_k=n}\frac{n!}{x_1!\cdots x_k!}(p_1e^{t_1})^{x_1}\cdots(p_ke^{t_k})^{x_k} \\ &= \left(p_1e^{t_1}+\cdots+p_ke^{t_k}\right)^n\end{aligned}$$

$X \sim \mathrm{Gamma}(\alpha,\beta)$ 이면 $\beta t < 1$ 일 때

$$\begin{aligned}\mathbb{E}(e^{tX}) &= \int_0^\infty e^{tx}\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{1}{\beta}x}dx \\ &= \frac{1}{(1-\beta t)^\alpha}\underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha)}\left(\frac{1}{\beta}-t\right)^\alpha x^{\alpha-1}e^{-\left(\frac{1}{\beta}-t\right)x}dx}_{=1}\end{aligned}$$

$X \sim \mathrm{Beta}(\alpha,\beta)$ 이면 $k=1,2,\cdots$ 에 대하여

$$\begin{aligned}\mathbb{E}(X^k) &= \int_0^1 x^k\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{\Gamma(\alpha+\beta+k)}{\Gamma(\alpha+k)\Gamma(\beta)}\right)^{-1}\underbrace{\int_0^1\frac{\Gamma(\alpha+\beta+k)}{\Gamma(\alpha+k)\Gamma(\beta)}x^{\alpha+k-1}(1-x)^{\beta-1}dx}_{=1}\end{aligned}$$

와 같이 $k-$th moment를 얻을 수 있을 뿐만 아니라, $t \in \mathbb{R}$ 에 대하여

$$\mathbb{E}(e^{tX}) = \int_0^1 e^{tx}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx$$

인데 $x \in [0,1]$ 에서 $e^{tx} \le e^{|t|}$ 이므로 $\mathbb{E}(e^{tX}) \le e^{|t|}$ 이다. 따라서 mgf가 모든 $t \in \mathbb{R}$ 에 대하여 존재하고, 정리 1.5.2에 의거하여

$$\mathbb{E}(e^{tX}) = \sum_{k=0}^{\infty}\frac{\mathbb{E}(X^k)}{k!}t^k = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\sum_{k=0}^{\infty}\frac{t^k}{k!}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha+\beta+k)}$$

$p \sim \mathrm{Beta}(\alpha, \beta)$ 이고 $X|p \sim \mathrm{B}(n, p)$ 이면 $x \in \{0, \cdots, n\}$ 에 대하여

$$\begin{aligned}\mathbb{P}(X = x) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\binom{n}{x}p^x(1-p)^{n-x}dp \\ &= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}\right)^{-1}\underbrace{\int_0^1 \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}p^{\alpha+x-1}(1-p)^{\beta+n-x-1}dp}_{=1}\end{aligned}$$

이고

$$\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|p)] = \mathbb{E}[np] = n \cdot \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)} = \frac{n\alpha}{\alpha+\beta}$$

일반적으로 $t \in \mathbb{R}$ 에 대하여

$$\begin{aligned}\mathbb{E}(e^{tX}) = \mathbb{E}[\mathbb{E}(e^{tX}|p)] &= \mathbb{E}\left[\left(1 + p(e^t - 1)\right)^n\right] \\ &= \mathbb{E}\left[\sum_{k=0}^n \binom{n}{k} p^k (e^t-1)^k\right] \\ &= \sum_{k=0}^n \binom{n}{k}\mathbb{E}\left[p^k\right](e^t-1)^k \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\sum_{k=0}^n \binom{n}{k}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha+\beta+k)}(e^t-1)^k\end{aligned}$$

$X \sim \mathrm{invGamma}(\alpha, \beta)$ 이면 $t > 0$ 에 대하여

$$\mathbb{E}(e^{tX}) = \int_0^\infty e^{tx}\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{-\alpha-1}e^{-\frac{1}{\beta x}}dx$$

인데 $x \geq M \implies tx - (\alpha+1)\log x - \frac{1}{\beta x} > \frac{t}{2}x$ 를 만족하는 $M > 0$ 이 존재하고

$$\int_M^\infty e^{\frac{t}{2}x}dx = \infty$$

이므로 mgf는 존재하지 않는다. *Note.* 역감마분포에서 $k-$th moment의 존재성은 α 와 k 의 대소와 관련되어 있다. $\alpha > 0$ 이기만 하면 분포가 잘 정의되지만, $k-$th moment를 가지려면 $\alpha > k$ 여야 한다. 이 경우,

$$\begin{aligned}\mathbb{E}(X^k) &= \int_0^\infty x^k \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{-\alpha-1}e^{-\frac{1}{\beta x}}dx \\ &= \frac{\Gamma(\alpha-k)}{\Gamma(\alpha)\beta^k}\underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha-k)\beta^{\alpha-k}}x^{-(\alpha-k)-1}e^{-\frac{1}{\beta x}}dx}_{=1}\end{aligned}$$

$X \sim \mathrm{L}(\mu, \sigma)$ 이면 $|\sigma t| < 1$ 일 때 $z = \frac{x-\mu}{\sigma}$ 에 대하여 $dz = \frac{1}{\sigma}dx$ 이고 $w = \frac{1}{1+e^{-z}}$ 에 대하여 $dw = \frac{e^{-z}}{(1+e^{-z})^2}dz$ 이므로

$$\begin{aligned}\mathbb{E}(e^{tX}) = \int_{-\infty}^\infty e^{tx}\frac{e^{-z}}{\sigma(1+e^{-z})^2}dx &= \int_{-\infty}^\infty e^{t(\mu+\sigma z)}\frac{e^{-z}}{(1+e^{-z})^2}dz \\ &= e^{\mu t}\int_0^1 \left(\frac{w}{1-w}\right)^{\sigma t}dw \\ &= e^{\mu t}\Gamma(1-\sigma t)\Gamma(1+\sigma t)\underbrace{\int_0^1 \frac{\Gamma(2)}{\Gamma(1+\sigma t)\Gamma(1-\sigma t)}w^{1+\sigma t-1}(1-w)^{1-\sigma t-1}dw}_{=1}\end{aligned}$$

$X \sim \mathrm{N}(\mu, \sigma^2)$ 이면 $t \in \mathbb{R}$ 일 때 $z = \frac{x-\mu}{\sigma}$ 에 대하여 $dz = \frac{1}{\sigma}dx$ 이므로

$$\begin{aligned}\mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \int_{-\infty}^{\infty} e^{t(\mu+\sigma z)}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \\ = e^{\mu t+\frac{1}{2}\sigma^2t^2}\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(z-\sigma t)^2/2}dz}_{=1}\end{aligned}$$

$X \sim \mathrm{Lognormal}(\mu, \sigma^2)$ 이면 $k = 0, 1, \cdots$ 에 대하여

$$\begin{aligned}\mathbb{E}(X^k) &= \int_0^{\infty} x^k\frac{1}{x\sqrt{2\pi}\sigma}\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)dx \\ &= \int_{-\infty}^{\infty} e^{ky}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)dy && (y = \log x) \\ &= e^{k\mu+\frac{1}{2}k^2\sigma^2} && \text{(정규분포의 mgf 유도과정을 다시 살펴보자)}\end{aligned}$$

그럼에도 불구하고 임의의 $t > 0$ 에 대하여 $\mathbb{E}(e^{tX}) = \infty$ 임을 보인 적이 있다.
$X \sim t_\nu$ 면 $t > 0$ 에 대하여

$$\mathbb{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}dx$$

인데 역감마함수와 같은 논증으로 mgf는 존재하지 않는다. ($x \to \infty$ 일 때 피적분함수가 발산한다.) *Note*. 역감마분포와 마찬가지로 $k-$th moment의 존재성은 ν 와 k 의 대소와 관련되어 있다. $\nu > 0$ 이기만 하면 분포가 잘 정의되지만, $k-$th moment를 가지려면 $\nu > k$ 여야 한다. 이 경우, pdf가 even function이므로 k 가 odd일 때 $\mathbb{E}(X^k) = 0$ 이고 k 가 even일 때 $x \geq 0$ 에 대하여

$$z = \frac{x^2}{\nu+x^2} \qquad x = \sqrt{\frac{\nu z}{1-z}} \qquad dx = \frac{1}{2}\sqrt{\frac{\nu}{z(1-z)^3}}dz$$

로 치환하면

$$\begin{aligned}\mathbb{E}(X^k) &= \int_{-\infty}^{\infty} x^k\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}dx \\ &= \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\int_0^{\infty} x^k\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}dx && \text{(pdf and } k \text{ are even)} \\ &= \frac{2\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\int_0^1\left(\frac{\nu z}{1-z}\right)^{\frac{k}{2}}(1-z)^{\frac{\nu+1}{2}}\frac{1}{2}\sqrt{\frac{\nu}{z(1-z)^3}}dz && \text{(substitute } x \text{ by } z\text{)} \\ &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\nu^{\frac{k+1}{2}}\int_0^1 z^{\frac{k+1}{2}-1}(1-z)^{\frac{\nu-k}{2}-1}dz \\ &= \frac{1}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\nu^{\frac{k+1}{2}}\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{\nu-k}{2}\right) && \text{(Beta Integral)} \\ &= \frac{\Gamma\left(\frac{1}{2}+m\right)\Gamma\left(\frac{\nu}{2}-m\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}\nu^m && \text{(Let } m = \tfrac{k}{2} \in \mathbb{Z}\text{)}\end{aligned}$$

다시 한 번 강조하지만, 이 모든 논의는 $\nu > k$ 일 때 가능한 것이다.
$X \sim \mathrm{Weibull}(\alpha, \beta)$ 이면 $(\alpha, \beta > 0)$

$$z = \left(\frac{x}{\beta}\right)^\alpha \qquad dz = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}dx$$

로 치환하여

$$\mathbb{E}(X^k) = \int_0^\infty x^k \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} dx = \beta^k \int_0^\infty z^{k/\alpha} e^{-z} dz = \beta^k \Gamma\left(1 + \frac{k}{\alpha}\right)$$

를 얻는다.
이상의 논의는 모든 $\alpha, \beta > 0$에 대하여 성립했으나, 웨이불분포의 mgf가 존재하기 위해서는 $\alpha \geq 1$이어야 함이 알려져 있다. $\alpha > 1$인 경우에는 $t \in \mathbb{R}$에 대하여

$$\mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} dx = \int_0^\infty \exp\left(\beta t z^{1/\alpha} - z\right) dz$$

인데 $z \geq M \implies \beta t z^{1/\alpha} \leq \frac{z}{2}$가 성립하는 $M > 0$이 존재하고

$$\int_M^\infty \exp\left(-\frac{z}{2}\right) < \infty$$

이므로 mgf가 모든 $t \in \mathbb{R}$에 대하여 존재한다. 이제는 정리 1.5.2에 의거하여

$$\mathbb{E}(e^{tX}) = \sum_{k=0}^\infty \frac{\mathbb{E}(X^k)}{k!} t^k = \sum_{k=0}^\infty \frac{(\beta t)^k}{k!} \Gamma\left(1 + \frac{k}{\alpha}\right)$$

라고 적을 수 있다. 정확하게 $\alpha = 1$인 경우에는 $\beta t < 1$인 t에 대하여 mgf가 존재할 것이다. 그리고 이 경우는 지수분포에 해당한다. (다음 절을 참조하라.)

Related Distributions

$\text{Bernoulli}(p) \overset{d}{=} \text{B}(1,p)$	$p^x(1-p)^{1-x} \text{I}_{\{0,1\}}(x)$	(베르누이 Bernoulli 분포)
$\text{Geo}(p) \overset{d}{=} \text{Negbin}(1,p)$	$p(1-p)^{x-1} \text{I}_{\{1,2,\cdots\}}(x)$	(기하 Geometric 분포)
$\text{Exp}(\beta) \overset{d}{=} \text{Gamma}(1,\beta)$	$\frac{1}{\beta} e^{-x/\beta} \text{I}_{(0,\infty)}(x)$	(지수 Exponential 분포)
$\chi^2_\nu \overset{d}{=} \text{Gamma}\left(\frac{\nu}{2}, 2\right)$	$\frac{1}{\Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2}} x^{\frac{\nu}{2}-1} e^{-x/2} \text{I}_{(0,\infty)}(x)$	(카이제곱 χ^2 분포)
$\text{Cauchy}(0,1) \overset{d}{=} t_1$	$\frac{1}{\pi(1+x^2)} \text{I}_{(-\infty,\infty)}(x)$	(코시 Cauchy 분포)
$\text{Unif}(0,1) \overset{d}{=} \text{Beta}(1,1)$	$\text{I}_{(0,1)}(x)$	(균등 Uniform 분포)

Generalized Gamma Distribution

name	notation	support	probability density function	moment generating function
일반화된 감마분포 Generalized Gamma	$\text{GG}(d,p,\beta)$ $d,p,\beta > 0$	$(0,\infty)$	$\frac{p}{\Gamma(d/p)\beta^d} x^{d-1} e^{-(x/\beta)^p}$	$\mathbb{E}(X^k)$ mgf exists if $p \geq 1$

Note. $p = 1$이면 감마분포가 되고, $d = p$이면 웨이불분포가 되고, $d = 1, p = 2$이면 반정규 Half Normal 분포(정규분포를 따르는 확률변수의 절대값을 생각)가 된다.
$X \sim \text{GG}(d,p,\beta)$이면 $(d,p,\beta > 0)$

$$z = \left(\frac{x}{\beta}\right)^p \qquad dz = \frac{p}{\beta}\left(\frac{x}{\beta}\right)^{p-1} dx$$

로 치환하여 $k = 0, 1, 2, \cdots$ 에 대하여

$$\begin{aligned}
\mathbb{E}(X^k) &= \int_0^\infty x^k \frac{p}{\Gamma(d/p)\beta^d} x^{d-1} e^{-(x/\beta)^p} dx \\
&= \frac{\beta^k}{\Gamma(d/p)} \int_0^\infty \left(\frac{x}{\beta}\right)^{d+k-p} \cdot e^{-(x/\beta)^p} \cdot \frac{p}{\beta} \left(\frac{x}{\beta}\right)^{p-1} dx \\
&= \frac{\beta^k}{\Gamma(d/p)} \int_0^\infty z^{\frac{d+k-p}{p}} e^{-z} dz \\
&= \beta^k \frac{\Gamma\left(\frac{d+k}{p}\right)}{\Gamma\left(\frac{d}{p}\right)}
\end{aligned}$$

더 나아가 $p \geq 1$ 이라면 mgf가 존재한다. 정확히 $p = 1$ 이라면 앞서 다룬 감마분포에 해당하게 된다. $p > 1$ 이라면 $t \in \mathbb{R}$ 에 대하여

$$\begin{aligned}
\mathbb{E}(e^{tX}) &= \int_0^\infty e^{tx} \frac{p}{\Gamma(d/p)\beta^d} x^{d-1} e^{-(x/\beta)^p} dx \\
&= \frac{1}{\Gamma(d/p)} \int_0^\infty e^{tx} \cdot \left(\frac{x}{\beta}\right)^{d-p} \cdot e^{-(x/\beta)^p} \cdot \frac{p}{\beta} \left(\frac{x}{\beta}\right)^{p-1} dx \\
&= \frac{1}{\Gamma(d/p)} \int_0^\infty \exp\left(\beta t z^{1/p} + \frac{d-p}{p} \log z - z\right) dz
\end{aligned}$$

$M > 0$ 이 존재하여 $z > M$ 이면 $\exp(\cdot)$ 안의 항이 $-z/2$ 보다 작게 된다. 그리고

$$\int_M^\infty \exp\left(-\frac{z}{2}\right) dz < \infty$$

이므로 mgf가 모든 $t \in \mathbb{R}$ 에 대하여 존재한다. 이제 정리 1.5.2를 적용할 수 있다.

CDF and Sampling Theory

$\mathbb{P}(X \leq x) = F(x)$ 이고 $U \sim \text{unif}(0,1)$ 이라고 할 때 X 와 $F^{-1}(U)$ 는 정확히 같은 분포가 됨을 지난 번 `2.pdf`에서 밝혔었다. F 의 inverse가 존재하면 그대로 사용하면 되고, 존재하지 않는다면 다음의 generalized version을 사용하는 것이다.

$$F^{-1}(u) = \inf\{y \in \mathbb{R} : F(y) \geq u\} \qquad 0 < u < 1$$

이것은 항상 잘 정의되는 것을 역시 밝혔었다.
$X \sim \text{Exp}(\beta)$ 면 pdf와 cdf가 각각 $f(x) = \frac{1}{\beta} e^{-x/\beta} \mathrm{I}_{(0,\infty)}(x)$ 와

$$F(x) = \begin{cases} 1 - e^{-x/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

으로 주어지므로 $X \overset{d}{=} -\beta \log(1-U)$ 가 성립한다.
응용. CDF를 통한 sampling은 로지스틱분포에서도 유효하다.
$X \sim \text{Gamma}(\alpha, \theta), Y \sim \text{Gamma}(\beta, \theta)$ 이고 $X \perp\!\!\!\perp Y$ 이면 X, Y 의 joint pdf는

$$f(x,y) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}} x^{\alpha-1} y^{\beta-1} e^{-(x+y)/\theta} \mathrm{I}_{(0,\infty)}(x) \mathrm{I}_{(0,\infty)}(y)$$

로 주어지므로 $x = zw, y = z(1-w)$ 로 치환하여 $W = \frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$ 임을 확인할 수 있다.

흔히 표준정규분포의 pdf를 ϕ로 나타낸다.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{I}_{(-\infty,\infty)}(x)$$

Student's t 분포 pdf의 점별수렴

자유도가 $\nu > 0$인 t분포의 pdf는

$$f_{t(\nu)}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

로 주어지며, $k = 1, 2, \cdots$ 에 대하여

$$\mathbb{E}(X^k) = \begin{cases} \text{undefined} & \nu \le k \\ 0 & \nu > k = 2m - 1 \\ \dfrac{\Gamma\left(\frac{1}{2} + m\right)\Gamma\left(\frac{\nu}{2} - m\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} \nu^m & \nu > k = 2m \end{cases}$$

가 성립함을 보였었다.
Step 1. $X \sim t_\nu$ 에 대하여 $\lim_{\nu\to\infty} \mathrm{Var}(X) = 1$임을 보여라.
Step 2. 감마함수의 정의로부터 다음을 보여라.

$$\left(\Gamma\left(\frac{\alpha+\beta}{2}\right)\right)^2 \le \Gamma(\alpha)\Gamma(\beta) \qquad \alpha, \beta > 0$$

Step 3. 감마함수가 양의 실수에서 정의된 연속함수임을 이용하여 $\log\Gamma(\alpha)$가 볼록함수임을 보여라.
Step 4. $\log\Gamma(\alpha)$가 볼록함수임을 이용하여 $n = 2, 3, \cdots$ 및 $\epsilon \in (0, 2]$에 대하여 다음을 보여라.

$$(n-1)^\epsilon \le \frac{\Gamma(n+\epsilon)}{\Gamma(n)} \le (n+1)^\epsilon$$

Step 5. 위의 사실을 이용하여 다음을 보여라.

$$\lim_{\alpha\to\infty} \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\sqrt{\alpha}\,\Gamma(\alpha)} = 1$$

Step 6. 고정된 $x \in \mathbb{R}$에 대하여 다음을 보여라.

$$\lim_{\nu\to\infty} f_{t(\nu)}(x) = \phi(x)$$

Solution

먼저 $\nu > 2$일 때 $\mathbb{E}(X) = 0$이고

$$\mathbb{E}(X^2) = \frac{\Gamma\left(\frac{1}{2}+1\right)\Gamma\left(\frac{\nu}{2}-1\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}\nu = \frac{\frac{1}{2}\nu}{\frac{\nu}{2}-1} = \frac{\nu}{\nu-2} \to 1 \qquad \text{as } \nu \to \infty$$

이 성립한다. 한편 임의의 양수 $\alpha, \beta > 0$에 대하여 정적분의 코시 슈바르츠 부등식에 의하여

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta) &= \int_0^\infty x^{\alpha-1}e^{-x}dx \int_0^\infty x^{\beta-1}e^{-x}dx \\ &\geq \left(\int_0^\infty x^{\frac{\alpha+\beta}{2}-1}e^{-x}dx\right)^2 \\ &= \left(\Gamma\left(\frac{\alpha+\beta}{2}\right)\right)^2\end{aligned}$$

이므로 $\log\Gamma\left(\frac{\alpha+\beta}{2}\right) \leq \frac{1}{2}\left(\log\Gamma(\alpha) + \log\Gamma(\beta)\right)$가 성립한다. 감마함수는 양의 실수에서 연속이므로 임의의 $0 < t < 1$에 대하여

$$\log\Gamma\left((1-t)\alpha + t\beta\right) \leq (1-t)\log\Gamma(\alpha) + t\log\Gamma(\beta)$$

가 성립하는 것을 확인할 수 있다. *Comment*. 임의의 $\epsilon > 0$에 대하여 근방 $(t-\epsilon, t+\epsilon)$에 속하는 $t' = n/2^m \in (0,1)$을 찾을 수 있다. 그리고 $t = n/2^m$ 꼴인 경우에는 위의 부등식이 성립하는 것을 m에 대한 수학적 귀납법으로 보일 수 있다.

이러한 논의로 $\log\Gamma$가 볼록함수임을 알 수 있다. *Step 4*는 감마함수의 근사적인 범위를 제공하는 유용한 식이기는 하나, 여기서는 무시하고 *Step 5*로 넘어갈 수 있다. (아이디어는 같다.) 구체적으로, 서로 다른 양수 α, β에 대하여

$$S(\alpha,\beta) = \frac{\log\Gamma(\alpha) - \log\Gamma(\beta)}{\alpha - \beta}$$

라고 정의하면 $\log\Gamma$의 볼록성에 의하여 임의의 $\alpha > 1$에 대하여

$$\log(\alpha-1) = \log\frac{\Gamma(\alpha)}{\Gamma(\alpha-1)} = S(\alpha-1,\alpha) \leq \underbrace{S\left(\alpha, \alpha+\frac{1}{2}\right)}_{=2\log\frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)}} \leq S(\alpha,\alpha+1) = \log\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \log\alpha$$

이므로

$$\sqrt{\alpha-1} \leq \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)} \leq \sqrt{\alpha}$$

각 변을 $\sqrt{\alpha}$로 나누고 샌드위치 정리를 적용하면 *Step 5*가 성립함을 확인할 수 있다. 이제 고정된 x에 대하여

$$\lim_{\nu\to\infty} f_{t(\nu)}(x) = \lim_{\nu\to\infty} \underbrace{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}}_{\to 1} \cdot \frac{1}{\sqrt{2\pi}} \cdot \underbrace{\left(\left[1+\frac{x^2}{\nu}\right]^{\nu/x^2}\right)}_{\to e}{}^{-\frac{\nu+1}{2\nu}x^2} = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} = \phi(x)$$

4.1 통계량의 분포 및 4.2 대표적인 표본분포

수업시간에 정리 4.1.1 (일대일 변환을 통한 확률변수의 치환법) 및 정리 4.1.2 (다대일 변환을 통한 확률변수의 치환법)를 다루고 다음의 예시를 확인하였다.

코시분포와 베타분포의 관계

$X \sim \mathrm{Cauchy}(0,1)$ 일 때 $Y = \frac{X^2}{1+X^2}$ 의 분포는?
X 는 $(-\infty, \infty)$ 를 support로 가지는 확률변수다. 그런데 $\mathbb{P}(X \in (-\infty, 0) \cup (0, \infty)) = 1$ 이고 각각의 open set $(-\infty, 0), (0, \infty)$ 에서 $x \mapsto \frac{x^2}{1+x^2}$ 는 (미분가능하고 도함수가 연속이며 영이 아니고 치역이 $(0,1)$ 인) 일대일 변환이 된다. 역변환은 각각 $y \mapsto -\sqrt{\frac{y}{1-y}}$ 및 $y \mapsto \sqrt{\frac{y}{1-y}}$ 로 주어지는데, 절대값 $|dx/dy|$ 는 동일하므로 정리 4.1.2에 의하여 $y = x^2/(1+x^2)$ 일 때

$$\begin{aligned}
pdf_Y(y) &= pdf_X(x) \left| \frac{dx}{dy} \right| \cdot 2 \\
&= \frac{1}{\pi(1+x^2)} \frac{\frac{1}{(1-y)^2}}{\sqrt{\frac{y}{1-y}}} \\
&= \frac{1}{\pi\sqrt{y(1-y)}} \mathrm{I}_{(0,1)}(y) \\
&= \frac{\Gamma(1)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)} y^{\frac{1}{2}-1}(1-y)^{\frac{1}{2}-1} \mathrm{I}_{(0,1)}(y)
\end{aligned}$$

마지막에 support $(0,1)$ 을 명시해주는 것을 잊지 말자. Y 의 분포는 $\mathrm{Beta}(\frac{1}{2}, \frac{1}{2})$ 이 된다.

정규분포와 감마분포의 관계

$X \sim \mathrm{N}(0,1)$ 일 때 $Y = X^2$ 의 분포는? 위와 동일한 논증을 통해 정리 4.1.2를 적용하면 $y = x^2$ 일 때

$$\begin{aligned}
pdf_Y(y) &= pdf_X(x) \left| \frac{dx}{dy} \right| \cdot 2 \\
&= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{y}} \\
&= \frac{1}{\sqrt{2\pi y}} e^{-y/2} \mathrm{I}_{(0,\infty)}(y) \\
&= \frac{1}{\Gamma\left(\frac{1}{2}\right) 2^{\frac{1}{2}}} y^{\frac{1}{2}-1} e^{-y/2} \mathrm{I}_{(0,\infty)}(y)
\end{aligned}$$

따라서 Y 의 분포는 $\mathrm{Gamma}\left(\frac{1}{2}, 2\right)$ 가 된다. 우리는 이것을 자유도가 1인 카이제곱분포라고 부르며 $\chi^2(1)$ 로 나타낸다. 감마분포의 가산성(감마분포 mgf 식의 꼴을 참조하라)에 의하여 $X_1, \cdots, X_\nu \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ 일 때 $V = X_1^2 + \cdots X_\nu^2$ 이 따르는 분포는 $\mathrm{Gamma}\left(\frac{\nu}{2}, 2\right)$ 가 된다. 이것은 자유도가 ν 인 카이제곱분포이며 $\chi^2(\nu)$ 로 나타낸다.

t 분포와 F 분포의 pdf

이제 예 4.2.3과 예 4.2.4를 읽고 t 분포 및 F 분포의 pdf 유도과정을 직접 써보자.

정리 4.2.2부터 4.2.6까지

이 부분의 내용은 t 분포와 F 분포의 대의적 정의와 표본분포 사이를 연결하는 것이다. 수업시간에 정리 4.2.2와 4.2.3을 유도하였는데, 직접 손으로 써보기를 바란다. 4단원에서 가장 중요한 내용을 두 가지 뽑으라면 하나는 정리 4.2.2이고 다른 하나는 순서통계량의 분포이다.

4.3 순서통계량의 분포 및 4.4 다변량 정규분포

예 4.3.3 (응용)

모수가 μ, σ 인 지수분포 $\mathrm{Exp}(\mu, \sigma)$ 는 다음의 pdf를 가진다. $(\sigma > 0)$

$$f(x) = \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} \mathrm{I}_{(\mu,\infty)}(x)$$

이 지수분포 랜덤표본 n 개에 기초한 순서통계량 $X_{(1)} < \cdots < X_{(n)}$ 에 대하여

$$Z_r = \frac{n-r+1}{\sigma}(X_{(r)} - X_{(r-1)}) \qquad\qquad r = 1, \cdots, n$$

라고 정의하면 (단, $X_{(0)} = \mu$ 로 정의한다.) $Z_1, \cdots, Z_n$ 은 서로 독립이고 표준지수분포를 따른다.
연습문제 7.13을 살펴보자. 최대가능도비 검정은 수리통계학 I 범위가 아니기 때문에 아직 몰라도 괜찮다. 그런데 최대가능도비 검정의 결과 기각역이

$$Y = \frac{X_{(1)} - \mu_0}{\overline{X} - X_{(1)}} \geq c \qquad\qquad \text{또는} \qquad\qquad X_{(1)} < \mu_0$$

의 꼴로 주어진다는 것을 알 때 (단, $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ 이고 $c > 0$) 통계량 Y 의 귀무(null) 분포에 대해서는 논할 수 있다. 그래야만 $c > 0$ 의 값을 결정할 수 있다. 앞서 예 4.3.3에서 정의한 Z_r 을 사용하면, under H_0,

$$\begin{aligned}
X_{(1)} - \mu_0 &= \frac{\sigma}{n} Z_1 \\
\overline{X} - X_{(1)} &= \frac{1}{n}\sum_{r=1}^n (X_{(r)} - X_{(1)}) \\
&= \frac{1}{n}\sum_{r=2}^n (n-r+1)(X_{(r)} - X_{(r-1)}) \\
&= \frac{\sigma}{n}\sum_{r=2}^n Z_r
\end{aligned}$$

이므로 $X_{(1)} - \mu_0 \perp \overline{X} - X_{(1)}$ 이고

$$(n-1)Y = \frac{2Z_1/2}{\sum_{r=2}^n 2Z_r/(2n-2)} \sim F(2, 2n-2)$$

따라서 적절한 상수 c 의 값은 $\frac{1}{n-1}F_\alpha(2, 2n-2)$ 로 주어진다.

정리 4.4.3 Note

정리 4.4.3의 (b) 및 예 4.4.2를 보고 "절대" 주의해야 할 것이 있다. (b)의 내용은 X_1, X_2 이 각각 normal이고 covariance가 0일 때 X_1, X_2 가 서로 독립이라는 뜻이 아니다. covariance가 0일 때 서로 독립이라는 이 성질은 X_1, X_2 가 "jointly" normal일 때 성립하는 것이다. 예를 들어,

$$X \sim \mathrm{N}(0,1)$$
$$Z \sim \mathrm{Ber}\left(\frac{1}{2}\right)$$
$$Y = (-1)^Z X$$

이면 $Y \sim \mathrm{N}(0,1)$ 이고 $\mathrm{Cov}(X,Y) = 0$ 이지만 X, Y 는 서로 독립이 아니다.

Lecture 4. Multivariate normal distribution

Sungkyu Jung
Seoul National University

August 24, 2021

4.1 Multivariate normal distribution - nonsingular case

Recall that the univariate normal distribution with mean μ and variance σ^2 has density

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp[-\frac{1}{2}(x-\mu)\sigma^{-2}(x-\mu)].$$

Similarly, the multivariate normal distribution for the special case of nonsingular covariance matrix Σ is defined as follows.

Definition 1. Let $\mu \in \mathbb{R}^p$ and $\Sigma_{(p\times p)} > 0$. A random vector $\mathbf{X} \in \mathbb{R}^p$ has p-variate normal distribution with mean μ and covariance matrix Σ if it has probability density function

$$f(\mathbf{x}) = |2\pi\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right], \tag{1}$$

for $\mathbf{x} \in \mathbb{R}^p$. We use the notation $\mathbf{X} \sim N_p(\mu, \Sigma)$.

One may check that $f(\mathbf{x})$ on $\mathbb{R}^p$ is a joint p.d.f.; i.e. for any $\mathbf{x} \in \mathbb{R}^p$, $f(\mathbf{x}) \geq 0$ (in fact, $f(\mathbf{x}) > 0$) and $\int f(\mathbf{x})d\mathbf{x} = 1$.

Theorem 1. *If* $\mathbf{X} \sim N_p(\mu, \Sigma)$ *for* $\Sigma > 0$, *then*

1. $\mathbf{Y} = \Sigma^{-\frac{1}{2}}(\mathbf{X}-\mu) \sim N_p(\mathbf{0}, \mathbb{I}_p)$,
2. $\mathbf{X} \overset{\mathcal{L}}{=} \Sigma^{\frac{1}{2}}\mathbf{Y} + \mu$ *where* $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbb{I}_p)$,
3. $E(\mathbf{X}) = \mu$ *and* $Cov(\mathbf{X}) = \Sigma$,
4. *for any fixed* $\mathbf{v} \in \mathbb{R}^p$, $\mathbf{v}'\mathbf{X}$ *is univariate normal.*
5. $U = (\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu) \sim \chi^2(p)$.

Example 1 (Bivariate normal). Suppose that two random variables X and Y are jointly normally distributed, with $E(X) = \mu_x$, $E(Y) = \mu_Y$, $\mathrm{Var}(X) = \sigma_X^2$, $\mathrm{Var}(Y) = \sigma_Y^2$, and $\mathrm{corr}(X,Y) = \rho$. Write down the joint p.d.f of (X, Y).

Example 2 (Standard multivariate normal distribution). If $\mathbf{X} \sim N_p(\mathbf{0}, \mathbb{I}_p)$, then $\mathbf{X}$ is said to have the standard multivariate normal distribution.

4.2 Geometry of multivariate normal

The multivariate normal distribution has location parameter μ and the shape parameter $\Sigma > 0$. In particular, let's look into the contour of equal density

$$\begin{aligned} E_c &= \{\mathbf{x} \in \mathbb{R}^p : f(\mathbf{x}) = c_0\} \\ &= \{\mathbf{x} \in \mathbb{R}^p : (\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu) = c^2\}. \end{aligned} \tag{2}$$

Moreover, consider the spectral decomposition of $\Sigma = \mathbf{U}\Lambda\mathbf{U}'$ where $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ and $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$ with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$. The E_c, for any $c > 0$, is an ellipsoid centered around μ with principal axes $\mathbf{u}_i$ of length proportional to $\sqrt{\lambda_i}$. If $\Sigma = \mathbb{I}_p$, the ellipsoid is the surface of a sphere of radius c centered at μ.

As an example, consider the bivariate normal distribution $N_2(\mathbf{0}, \Sigma)$ with

$$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix}'.$$

The location of the distribution is the origin ($\mu = \mathbf{0}$), and the shape (Σ) of the distribution is determined by the ellipse given by the two principal axes (one at 45 degree line, the other at -45 degree line). Figure 1 shows the density function and the corresponding E_c for $c = 0.5, 1, 1.5, 2, \dots$.

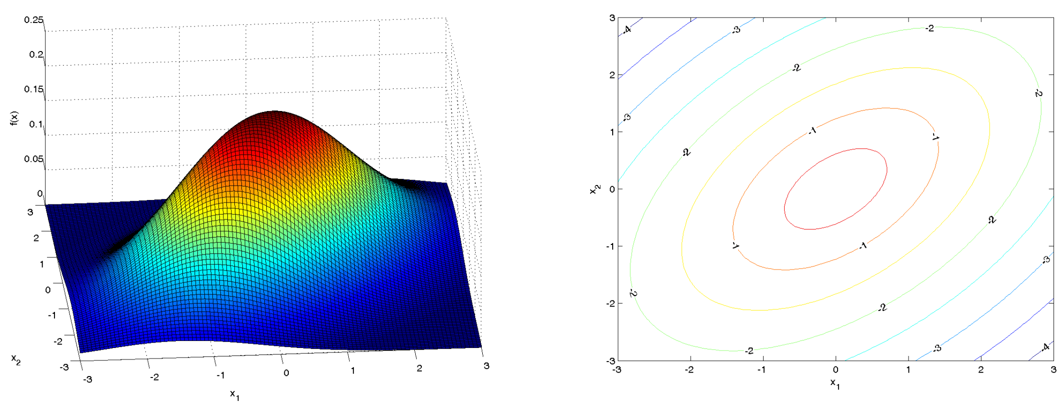


Figure 1: Bivariate normal density and its contours. Notice that an ellipses in the plane can represent a bivariate normal distribution. In higher dimensions $d > 2$, ellipsoids play the similar role.

Note that the constant $c = \{(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\}^{1/2} := d_M(\mu, \mathbf{x})$ in (2) is the Mahalanobis distance between $\mathbf{x}$ and μ.

4.3 General multivariate normal distribution

The characteristic function of a random vector $\mathbf{X}$ is defined as

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E(e^{i\mathbf{t}'\mathbf{X}}), \quad \text{for} \quad \mathbf{t} \in \mathbb{R}^p.$$

Note that the characteristic function is $\mathbb{C}$-valued, and always exists. We collect some important facts.

1. $\varphi_{\mathbf{X}}(t) = \varphi_{\mathbf{Y}}(t)$ if and only if $\mathbf{X} \stackrel{\mathcal{L}}{=} \mathbf{Y}$.
2. If $\mathbf{X}$ and $\mathbf{Y}$ are independent, then $\varphi_{\mathbf{X}+\mathbf{Y}} = \varphi_{\mathbf{X}}(t)\varphi_{\mathbf{Y}}(t)$.
3. $\mathbf{X}_n \Rightarrow \mathbf{X}$ if and only if $\varphi_{\mathbf{X}_n}(t) \to \varphi_{\mathbf{X}}(t)$ for all t.

An important corollary follows from the uniqueness of the characteristic function.

Corollary 2 (Cramer–Wold device)**.** *If $\mathbf{X}$ is a $p \times 1$ random vector then its distribution is uniquely determined by the distributions of linear functions of $\mathbf{t}'\mathbf{X}$, for every $\mathbf{t} \in \mathbb{R}^p$.*

Corollary 2 paves the way to the definition of (general) multivariate normal distribution.

Definition 2. A random vector $\mathbf{X} \in \mathbb{R}^p$ has a multivariate normal distribution if $\mathbf{t}'\mathbf{X}$ is an univariate normal for all $\mathbf{t} \in \mathbb{R}^p$.

The definition says that $\mathbf{X}$ is MVN if every projection of $\mathbf{X}$ onto a 1-dimensional subspace is normal, with a convention that a degenerate distribution δ_c has a normal distribution with variance 0, i.e., $c \sim N(c, 0)$. The definition does not require that $\mathrm{Cov}(\mathbf{X})$ is nonsingular.

Theorem 3. *The characteristic function of a multivariate normal distribution with mean μ and covariance matrix $\Sigma \geq 0$ is, for $t \in \mathbb{R}^p$,*

$$\varphi(t) = \exp[it'\mu - \frac{1}{2}t'\Sigma t].$$

If $\Sigma > 0$, then the pdf exists and is the same as (1).

In the following, the notation $\mathbf{X} \sim N(\mu, \Sigma)$ is valid for a non-negative definite Σ. However, whenever Σ^{-1} appears in the statement, Σ is assumed to be positive definite.

Proposition 4. *If $\mathbf{X} \sim N_p(\mu, \Sigma)$ and $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ for $\mathbf{A}_{(q\times p)}$ and $\mathbf{b}_{(q\times 1)}$, then $\mathbf{Y} \sim N_q(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}')$.*

Next two results are concerning independence and conditional distributions of normal random vectors. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be the partition of $\mathbf{X}$ whose dimensions are r and s, $r + s = p$, and suppose μ and Σ are partitioned accordingly. That is,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_p\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right).$$

Proposition 5. *The normal random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $Cov(\mathbf{X}_1, \mathbf{X}_2) = \Sigma_{12} = \mathbf{0}$.*

Proposition 6. *The conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is*

$$N_r(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Proof. Consider new random vectors $\mathbf{X}_1^* = \mathbf{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2$ and $\mathbf{X}_2^* = \mathbf{X}_2$,

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1^* \\ \mathbf{X}_2^* \end{bmatrix} = \mathbf{A}\mathbf{X}, \quad \mathbf{A} = \begin{bmatrix} \mathbb{I}_r & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0}_{(s\times r)} & \mathbb{I}_s \end{bmatrix}.$$

By Proposition 4, $\mathbf{X}^*$ is multivariate normal. An inspection of the covariance matrix of $\mathbf{X}^*$ leads that $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$ are independent. The result follows by writing

$$\mathbf{X}_1 = \mathbf{X}_1^* + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2,$$

and that the distribution (law) of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is $\mathcal{L}(\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2) = \mathcal{L}(\mathbf{X}_1^* + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2 \mid \mathbf{X}_2 = \mathbf{x}_2) = \mathcal{L}(\mathbf{X}_1^* + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2 \mid \mathbf{X}_2 = \mathbf{x}_2)$, which is a MVN of dimension r. $\square$

4.4 Multivariate Central Limit Theorem

If $\mathbf{X}_1, \mathbf{X}_2, \ldots \in \mathbb{R}^p$ are i.i.d. with $E(\mathbf{X}_i) = \mu$ and $\mathrm{Cov}(\mathbf{X}) = \Sigma$, then

$$n^{-\frac{1}{2}} \sum_{j=1}^{n} (\mathbf{X}_j - \mu) \Rightarrow N_p(\mathbf{0}, \Sigma) \quad \text{as} \;\; n \to \infty,$$

or equivalently,

$$n^{\frac{1}{2}}(\bar{\mathbf{X}}_n - \mu) \Rightarrow N_p(\mathbf{0}, \Sigma) \quad \text{as} \;\; n \to \infty,$$

where $\bar{\mathbf{X}}_n = \frac{1}{2}\sum_{j=1}^n \mathbf{X}_j$.

The delta-method can be used for asymptotic normality of $h(\bar{\mathbf{X}}_n)$ for some function $h : \mathbb{R}^p \to \mathbb{R}$. In particular, denote $\nabla h(\mathbf{x})$ for the gradient of h at $\mathbf{x}$. Using the first two terms of Taylor series,

$$h(\bar{\mathbf{X}}_n) = h(\mu) + (\nabla h(\mu))'(\bar{\mathbf{X}}_n - \mu) + O_p(\|\bar{\mathbf{X}}_n - \mu\|_2^2),$$

Then Slutsky's theorem gives the result,

$$\begin{aligned} \sqrt{n}(h(\bar{\mathbf{X}}_n) - h(\mu)) &= (\nabla h(\mu))'\sqrt{n}(\bar{\mathbf{X}}_n - \mu) + O_p(\sqrt{n}(\bar{\mathbf{X}}_n - \mu)'(\bar{\mathbf{X}}_n - \mu)) \\ &\Rightarrow (\nabla h(\mu))' N_p(\mathbf{0}, \Sigma) \quad \text{as} \;\; n \to \infty, \\ &= N_p(\mathbf{0}, (\nabla h(\mu))'\Sigma(\nabla h(\mu))) \end{aligned}$$

4.5 Quadratic forms in normal random vectors

Let $\mathbf{X} \sim N_p(\mu, \Sigma)$. A quadratic form in $\mathbf{X}$ is a random variable of the form

$$Y = \mathbf{X}'\mathbf{A}\mathbf{X} = \sum_{i=1}^{p}\sum_{j=1}^{p} X_i a_{ij} X_j,$$

where $\mathbf{A}$ is a $p \times p$ symmetric matrix. We are interested in the distribution of quadratic forms and the conditions under which two quadratic forms are independent.

Example 3. A special case: If $\mathbf{X} \sim N_p(0, \mathbb{I}_p)$ and $\mathbf{A} = \mathbb{I}_p$,

$$Y = \mathbf{X}'\mathbf{A}\mathbf{X} = \mathbf{X}'\mathbf{X} = \sum_{i=1}^{p} X_i^2 \sim \chi^2(p).$$

Fact 1. Recall the following:

1. A $p \times p$ matrix $\mathbf{A}$ is idempotent if $\mathbf{A}^2 = \mathbf{A}$.
2. If $\mathbf{A}$ is symmetric, then $\mathbf{A} = \Gamma'\Lambda\Gamma$, where $\Lambda = \text{diag}(\lambda_i)$ and Γ is orthogonal.
3. If $\mathbf{A}$ is symmetric idempotent,
   (a) its eigenvalues are either 0 or 1,
   (b) $\text{rank}(\mathbf{A}) = \#\{\text{non zero eigenvalues}\} = \text{trace}(\mathbf{A})$.

Theorem 7. *Let $\mathbf{X} \sim N_p(0, \sigma^2\mathbb{I})$ and $\mathbf{A}$ be a $p \times p$ symmetric matrix. Then*

$$Y = \frac{\mathbf{X}'\mathbf{A}\mathbf{X}}{\sigma^2} \sim \chi^2(m)$$

if and only if $\mathbf{A}$ is idempotent of rank $m < p$.

Corollary 8. *Let $\mathbf{X} \sim N_p(0, \Sigma)$, where Σ is positive definite, and $\mathbf{A}$ be a $p \times p$ symmetric matrix. Then*

$$Y = \mathbf{X}'\mathbf{A}\mathbf{X} \sim \chi^2(m)$$

if and only if either i) $\mathbf{A}\Sigma$ is idempotent of rank m or ii) $\Sigma\mathbf{A}$ is idempotent of rank m.

Example 4. If $\mathbf{X} \sim N_p(\mu, \Sigma)$, where Σ is positive definite, then $(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) \sim \chi^2(p)$.

Theorem 9. *Let $\mathbf{X} \sim N_p(0, \mathbb{I})$ and $\mathbf{A}$ be a $p \times p$ symmetric matrix, and $\mathbf{B}$ be a $k \times p$ matrix. If $\mathbf{B}\mathbf{A} = 0$, then $\mathbf{B}\mathbf{X}$ and $\mathbf{X}'\mathbf{A}\mathbf{X}$ are independent.*

Example 5 (Normal sampling theory). Let $X_i \sim N(\mu, \sigma^2)$ i.i.d. Then

- The sample mean $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n}$.
- The sample mean $\bar{X}_n$ and the sample variance $S_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ are independent.
- Moreover, $(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$.

Example 6. The residual sum of squares in the standard linear regression has a scaled chi-squared distribution and is independent with the coefficient estimates.

다변량 정규분포 보충: 정성규 교수님 자료와 함께 보세요

$\Sigma > 0$(p.d.)인 경우의 확률밀도함수

p차원 다변량 정규분포 $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$는 pdf를 가지며 이는 다음과 같이 주어진다.

$$f(\mathbf{x}) = (\det(2\pi\boldsymbol{\Sigma}))^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

$\boldsymbol{\Sigma} > 0$이므로 $\boldsymbol{\Sigma} = \left(\boldsymbol{\Sigma}^{1/2}\right)\left(\boldsymbol{\Sigma}^{1/2}\right)^\top$의 Cholesky decomposition을 생각할 수 있고, 이때 $\mathbf{Y} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu})$의 변수 변환을 생각하면 일대일 변환이므로 $\mathbf{Y}$는 다음과 같은 pdf를 가진다.

$$\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\mathbf{x}) \left|\det\left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right)\right| \\
&= f_{\mathbf{X}}\left(\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{y}\right)\left|\det \boldsymbol{\Sigma}^{1/2}\right| \\
&= (\det(2\pi\boldsymbol{\Sigma}))^{-1/2} \exp\left[-\frac{1}{2}\mathbf{y}^\top\mathbf{y}\right](\det\boldsymbol{\Sigma})^{1/2} && (\det\boldsymbol{\Sigma} > 0) \\
&= (2\pi)^{-p/2}\exp\left[-\frac{1}{2}\mathbf{y}^\top\mathbf{y}\right] \\
&= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_i^2} && (\mathbf{y} = (y_1, \cdots, y_p)^\top)
\end{aligned}$$

따라서 $\mathbf{Y}$의 분포는 평균이 $\mathbf{0}$이고 분산이 $\mathbb{I}_p$인 정규분포, 즉 표준정규분포이다. 원래 분포 $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$의 mgf를 구해보자. 임의의 벡터 $\mathbf{t} \in \mathbb{R}^p$에 대하여

$$\begin{aligned}
\mathbb{E}(e^{\mathbf{t}^\top\mathbf{X}}) &= \mathbb{E}\left(e^{\mathbf{t}^\top(\boldsymbol{\mu}+\boldsymbol{\Sigma}^{1/2}\mathbf{Y})}\right) \\
&= e^{\mathbf{t}^\top\boldsymbol{\mu}}\mathbb{E}\left(e^{\mathbf{t}^\top\boldsymbol{\Sigma}^{1/2}\mathbf{Y}}\right) \\
&= e^{\mathbf{t}^\top\boldsymbol{\mu}}\int_{\mathbb{R}^p}(2\pi)^{-p/2}\exp\left[\mathbf{t}^\top\boldsymbol{\Sigma}^{1/2}\mathbf{y} - \frac{1}{2}\mathbf{y}^\top\mathbf{y}\right]d\mathbf{y} \\
&= e^{\mathbf{t}^\top\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t}}\underbrace{\int_{\mathbb{R}^p}(2\pi)^{-p/2}\exp\left[-\frac{1}{2}\mathbf{z}^\top\mathbf{z}\right]d\mathbf{z}}_{=1} && (\mathbf{z} = \mathbf{y} - (\boldsymbol{\Sigma}^{1/2})^\top\mathbf{t})
\end{aligned}$$

이 성립한다. 따라서 구하는 mgf는 $e^{\mathbf{t}^\top\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t}}$이고, $\mathbf{X}$의 평균과 분산이 각각 $\boldsymbol{\mu}, \boldsymbol{\Sigma}$임을 확인할 수 있다. 이 mgf를 사용하면 $\boldsymbol{\Sigma}$가 zero eigenvalue를 하나 이상 가지는 degenerate case에 대해서도 다변량정규분포를 정의할 수 있다. 다만 이 경우에는 pdf가 존재하지 않는다. 한편 임의의 $\mathbf{v} \in \mathbb{R}^p$와 $t \in \mathbb{R}$에 대하여

$$\begin{aligned}
\mathbb{E}(e^{t\mathbf{v}^\top\mathbf{X}}) = \mathbb{E}\left(e^{(t\mathbf{v})^\top\mathbf{X}}\right) &= \exp\left[(t\mathbf{v})^\top\boldsymbol{\mu} + \frac{1}{2}(t\mathbf{v})^\top\boldsymbol{\Sigma}(t\mathbf{v})\right] \\
&= \exp\left[t(\mathbf{v}^\top\boldsymbol{\mu}) + \frac{1}{2}t^2(\mathbf{v}^\top\boldsymbol{\Sigma}\mathbf{v})\right]
\end{aligned}$$

가 성립하는데 이는 일변량 정규분포 $\mathrm{N}(\mathbf{v}^\top\boldsymbol{\mu}, \mathbf{v}^\top\boldsymbol{\Sigma}\mathbf{v})$의 mgf와 같음을 알 수 있다. 따라서 적률생성함수의 분포 결정성에 의해 $\mathbf{v}^\top\mathbf{X} \sim \mathrm{N}(\mathbf{v}^\top\boldsymbol{\mu}, \mathbf{v}^\top\boldsymbol{\Sigma}\mathbf{v})$가 성립한다. `8.pdf`의 Proposition 4 역시 같은 논증으로 확인 가능하다.

다변량 정규분포의 조건부분포

각각 r, s차원인 확률벡터 $\mathbf{X}_1, \mathbf{X}_2$에 대하여 $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1^\top & \mathbf{X}_2^\top\end{bmatrix}^\top$가 $r+s=p$차원 정규분포를 따른다면 $\mathbf{X}_1, \mathbf{X}_2$는 jointly normal이라고 한다. 보다 구체적으로 다음과 같은 상황을 가정하자.

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 \\ \mathbf{X}_2\end{bmatrix} \sim \mathrm{N}_p\left(\begin{bmatrix}\boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$$

만약 $\boldsymbol{\Sigma}_{12} = \mathbf{0}$이라면 $\mathbf{X}_1, \mathbf{X}_2$의 joint mgf는 marginal mgf 둘의 곱으로 작성할 수 있으므로 $\mathbf{X}_1, \mathbf{X}_2$는 서로 독립이다. 그리고 $\boldsymbol{\Sigma}_{22} > 0$인 경우에 일대일변환

$$\begin{cases}\mathbf{X}_1^* = & \mathbf{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2 \\ \\ \mathbf{X}_2^* = & \mathbf{X}_2\end{cases}$$

을 생각하면 (이 변환은 암기하자) Proposition 4에 의하여 약간의 계산을 통해

$$\mathbf{X}^* = \begin{bmatrix}\mathbf{X}_1^* \\ \mathbf{X}_2^*\end{bmatrix} \sim \mathrm{N}_p\left(\begin{bmatrix}\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right)$$

임을 확인할 수 있다. 여기서 Schur complement matrix $\boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$로 정의된다. 따라서 $\mathbf{X}_1^*$는 $\mathbf{X}_2$와 독립이고, 이 사실을 통해 $\mathbf{X}_1|\mathbf{X}_2$의 conditional distribition을 구할 수 있다.

다변량 정규분포와 카이제곱분포

수업 시간에 `8.pdf`의 Theorem 7부터 9까지 (단방향으로) 증명하였다. 여기서는 Example 5를 살펴보도록 한다. Example 5의 경우 수리통계학 4장의 앞부분에서 다변량 정규분포의 지식 없이 일변량 정규분포의 mgf 공식만 가지고 이미 증명한 바 있다.

$X_i \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$ 인 것은 $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_n \end{bmatrix}^\top$ 에 대하여 $\mathbf{X} \sim \mathrm{N}_n(\mu \mathbf{1}_n, \sigma^2 \mathbb{I}_n)$ 과 같은 의미이다. 간단한 고찰과 Proposition 4에 의하여,

$$\begin{aligned} \overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n}\mathbf{1}_n^\top \mathbf{X} &\sim \mathrm{N}\left(\frac{1}{n}\mathbf{1}_n^\top \mu \mathbf{1}_n, \frac{1}{n}\mathbf{1}_n^\top \sigma^2 \mathbb{I}_n \frac{1}{n}\mathbf{1}_n\right) \\ &= \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) && (\mathbf{1}_n^\top \mathbf{1}_n = n) \end{aligned}$$

이고 $n \times n$ 행렬 $\mathbf{A} := \mathbb{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ 에 대하여

$$\begin{aligned} (n-1)\frac{S_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X_n})^2 &= \frac{1}{\sigma^2}\mathbf{X}^\top \mathbf{A}\mathbf{X} \\ &= \frac{1}{\sigma^2}(\mathbf{X} - \mu\mathbf{1}_n)^\top \mathbf{A}(\mathbf{X} - \mu\mathbf{1}_n) && (\mathbf{1}_n^\top \mathbf{A} = \mathbf{0}) \\ &\sim \chi^2(n-1) && (\mathbf{X} - \mu\mathbf{1}_n \sim \mathrm{N}_n(\mathbf{0}, \sigma^2\mathbb{I}_n) \text{ 이고, Theorem 7 적용}) \end{aligned}$$

이다. 여기서 real symmetric idempotent (i.e, projection) matrix인 $\mathbf{A} = \mathbb{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ 의 rank는 trace와 같으므로 $n-1$ 임을 쉽게 알 수 있다. 마지막으로, $\overline{X_n}$ 과 S_n^2 의 독립성은 Theorem 9로부터 확인할 수 있다. (Theorem 9는 Proposition 5로부터 얻은 것이다!)

Note. $\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ 는 내적공간 $\mathbb{R}^n$ 에서 $\mathbf{1}_n$ 벡터의 (one-dimensional) span 위로 정사영하는 선형사상이다. 뿐만 아니라 항등사상 (identity map)에서 이 사상을 뺀 $\mathbf{A}$ 는 $\mathbf{1}_n$ 벡터의 span의 직교여공간 (orthogonal complement)

$$\mathbf{1}_n^\perp = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{1}_n^\top \mathbf{x} = 0\right\}$$

위로 정사영하는 선형사상이다. 좌표공간에서 선형사상 $\mathbf{P}$ 가 정사영 (projection)일 필요충분조건은 real symmetric idempotent, 즉 $\mathbf{P}^\top = \mathbf{P} = \mathbf{P}^2$ 인 것이다.