확률 및 통계

Quiz 2,

Assignment 3

2024.05.24.

퀴즈 답안

- **퀴즈 #2 (5/24)**

1. A test in which the outcome is uncertain. → random experiment (확률 실험)

2. A complete collection of outcomes. All of the possible outcomes of an experiment.
   → sample space (표본공간)

3. A subset of the sample space of a random experiment. A single outcome or a collection of outcomes from a sample space. → event (사건)

4. A real number on a scale of 0 to 1 that represent the likelihood or chance of a certain event occurring. → probability (확률)

5. The ratio of the number of outcomes in which a specified event occurs to the total number of trials. The occurrence of an event in a large number of repetitions of the experiment. → relative frequency (상대빈도)

6. A tabular representation of data that shows the frequency of each category or class. It is commonly used in statistics to organize, summarize, and visualize data.
   → frequency distribution table (도수분포표)

퀴즈 답안

- **퀴즈 #2 (5/24)**

7. A statement or principle that is accepted to be true, but need not be so. Fundamental assumption without proof. → axiom (공리)

8. Events that cannot occur simultaneously. Only one event can occur at a time, i.e., the occurrence of one event precludes the occurrence of other events.
→ Mutually exclusive (상호배타)

9. Two experiments or events that have nothing to do with each other. The occurrence of the first event does not affect the probability of the occurrence of the second event.
→ statistically independent (통계적독립)

10. The probability of an event occurring given that another event has already occurred. When two events are independent, conditioning one upon the other has no effect.
→ conditional probability (조건부확률)

11. It describes the probability of an event, based on prior knowledge of conditions that might be related to the event, i.e., one conditional probability to be computed from the other. This rule can be used to compute the probability of an event based on new evidence or information. → Bayes' rule (베이즈규칙)

퀴즈 답안

- **퀴즈 #2 (5/24)**

12. A branch of mathematics that deal with the collection, organization, analysis, interpretation, and presentation of data. → statistics (통계)

13. A branch of statistics that uses various analytical tools to draw inferences about the population data from sample data. Based on assumptions about the underlying probability distribution, we want to estimate its parameters, make prediction, and test hypothesis.
→ inferential statistics (추론통계)

14. A method that focuses on collecting/selecting data from/about an entire population. After that, statistical inference about the entire population can be obtained from this subset. → sample survey (표본조사)

15. A function of sample values. Any quantity computed from values in a sample which is considered for a statistical purpose. → statistic (통계량)

16. This mathematical operation/function is to weight a random variable/function with its PMF or PDF. We can compute the average value of a random variable/function weighted by the probabilities. → expectation (기댓값)

퀴즈 답안

- **퀴즈 #2 (5/24)**

  17. This is an average value (or mean value) of observations taken from a larger population, the entirety of relevant data. It is a measure of central tendency that provides an estimate of the population mean. → sample mean (표본평균)

  18. Given observations taken from a population, it represents the dispersion or spread of a set of data points around their mean. It is calculated by taking the average of the squared differences between each data point and the sample mean.→ sample variance (표본분산)

  19. It means the number of independent pieces of information available to estimate something. In a statistical analysis (e.g. normalizing data), it is important to know how many independent variables are. Within a data set, some initial numbers can be chosen at random/freely, but sometimes one data can be constrained/fixed if the data set must add up to a specific sum or mean. → degree of freedom (자유도)

  20. It states that if you repeat an experiment independently a large number of times and average the result, what you obtain should be close to the expected value. It links the mean of a random variable to the average value of a number of realizations of the random variable, and thus justifies computing the mean experimentally as an average.
  → law of large numbers (큰 수의 법칙)

퀴즈 답안

- **퀴즈 #2 (5/24)**

21. A mapping from the sample space to the real line. A variable whose numerical value is determined by the outcome of a random experiment. → random variable (확률변수)

22. A function that describes the probabilities of values of a random variable.
→ probability distribution (확률분포)

23. A random variable all of whose values are equally likely.
→ uniform random variable (균일 확률변수)

24. A discrete random variable that takes on only one of two discrete values. e.g., 0 or 1.
→ Bernoulli random variable (베르누이 확률변수)

25. A discrete random variable that represents the number of independent and identical Bernoulli trials required to obtain the first successes.
→ geometric random variable (기하 확률변수)

퀴즈 답안

- **퀴즈 #2 (5/24)**

26. A continuous random variable that represents the waiting time between two successive events that occur independently.
→ exponential random variable (지수 확률변수)

27. It is a mathematical operation that combines two functions to produce a third function by taking a sum/integral of the product of two functions after one is reflected about the y-axis and shifted. It is commonly used in signal processing, image processing, and mathematics to describe the interaction or combination of two functions or signals.
→ convolution (합성곱)

28. A discrete random variable that represents the number of successes in a fixed number of independent and identical Bernoulli trials.
→ binomial random variable (이항 확률변수)

29. A discrete random variable that represents the number of independent and identical Bernoulli trials required to obtain a certain number of successes.
→ Pascal random variable (파스칼/음이항 확률변수)

퀴즈 답안

- **퀴즈 #2 (5/24)**

  30. A continuous random variable that represents the waiting time for a certain number of/multiple events that occur independently. Its probability distribution is also known as the gamma distribution with an integer shape parameter.
  → Erlang random variable (얼랑 확률변수)

  31. A continuous random variable that has a symmetry probability distribution with a single peak at its mean value and a dispersion parameter. Its probability distribution is expressed in terms of the squared difference from the mean.
  → Gaussian/Normal random variable (가우시안 확률변수)

  32. A symmetric, bell-shaped probability distribution that is defined as a Gaussian distribution with zero mean and unit variance.
  → z-distribution (z-분포/표준정규분포)

  33. A theorem that the distribution of the sum of a large number of independent and identically distributed (i.i.d.) random variables with a finite mean and variance approaches a Gaussian. → central limit theorem (중심극한정리)

퀴즈 답안

- **퀴즈 #2 (5/24)**

34. From the probability that random values will generate between lower and upper limits, it is commonly used to quantify the uncertainty associated with estimating population parameters. For example, the true mean of a random variable falls within this interval with certain probability. → confidence interval (신뢰구간)

35. We collect data to show that the null hypothesis is not true beyond a reasonable doubt. So we need to set the criterion to reject H0 and accept H1. According to this level, we can also make mistakes to reject previous notions of truth that are in fact true.
→ significance level (유의수준)

36. A method of statistical inference used to test claims or ideas about a true parameter based on given observations/a subset of a population. → hypothesis testing (가설검정)

37. A statistical measure that represents the strength of evidence for/against a null hypothesis. Given a sample data and the statistic, we can compute the probability of obtaining the value (or more extreme values), and compare it with a criteria to reject H_0.
→ p-value (유의확률)

퀴즈 답안

- **퀴즈 #2 (5/24)**

38. A probability distribution that is used when working with small sample sizes or when the population standard deviation is unknown. In a conservative approach, critical values of this distribution are used to determine the rejection or acceptance of a null hypothesis.

→ t-distribution (t-분포)

39. A statistical test used to determine if there is a significant association between two categorical variables. It involves comparing observed frequencies with expected frequencies to assess whether the differences between the observed and expected values are statistically significant. → chi-square test (카이제곱검정)

40. A method of estimating the parameters of an assumed/designed/learned probability distribution, given some observed data. We should choose the parameter value that makes the observed data the most likely result/most probable. It is achieved by maximizing a likelihood function. → maximum likelihood estimation (최대우도추정법)

41. A standard method in regression analysis to approximate the solution of overdetermined systems by minimizing the sum of the squares of the residuals. Here, a residual is the squared difference between an observed value and the fitted value provided by a model. The parameters of many models can be estimated by this method.

→ least square method (최소자승법)

퀴즈 답안

- **퀴즈 #2 (5/24)**

42. Discrete probability distribution that random variables take on any given pair of values. It is a function that describes the probability distribution of two or more discrete random variables simultaneously. → joint probability mass function (결합확률질량함수)

43. It is a function that provides the cumulative probability for multiple random variables taking on values less than or equal to certain values. It describes the probability that a set of random variables assumes specific values or falls within specific ranges.
→ joint cumulative distribution function (결합누적분포함수)

44. It is a function that describes the probability distribution of a single random variable in the presence of multiple random variables. It provides the probability density for that variable while integrating or summing out the other variables.
→ marginal probability density function (주변확률밀도함수)

45. A type of second order moment that measures the similarity between random variables on the average. It quantifies the extent to which the variables are related or associated with each other → correlation (상관)

퀴즈 답안

- 퀴즈 #2 (5/24)

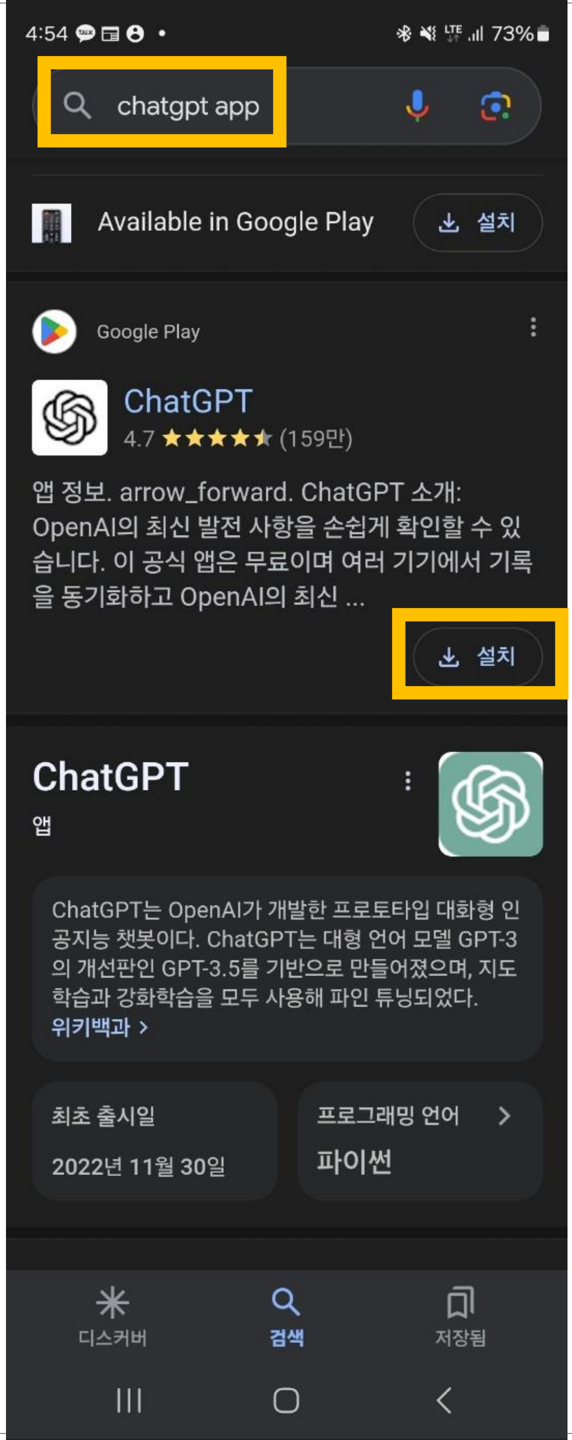
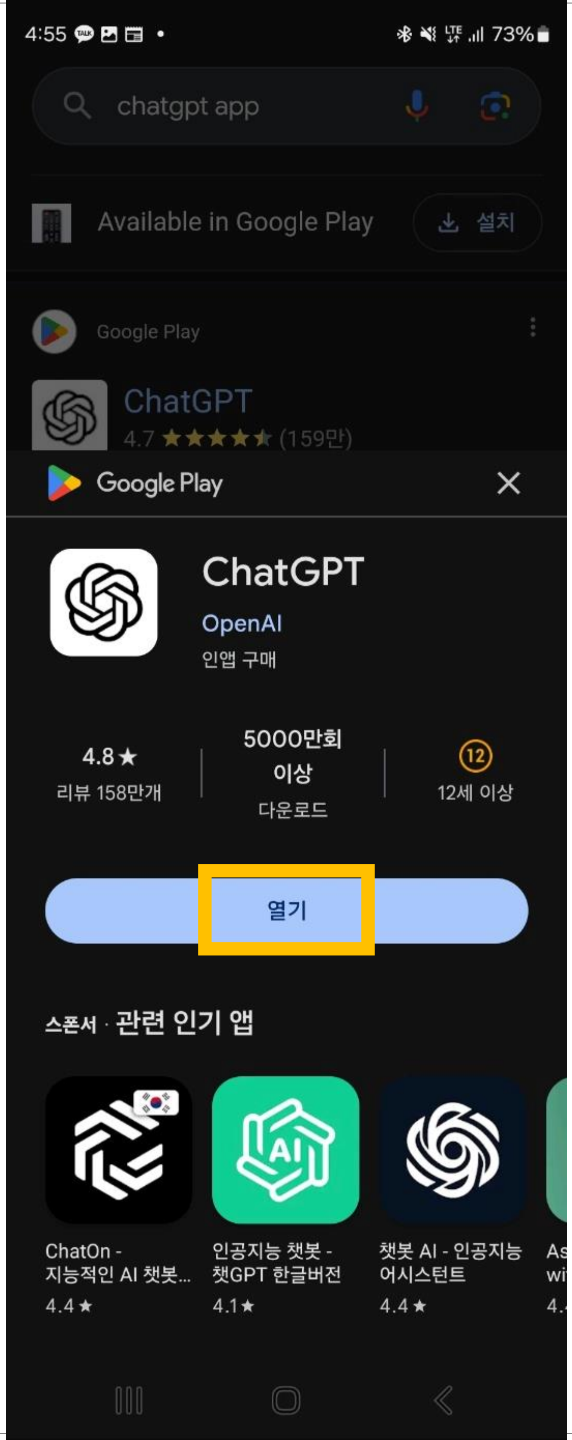
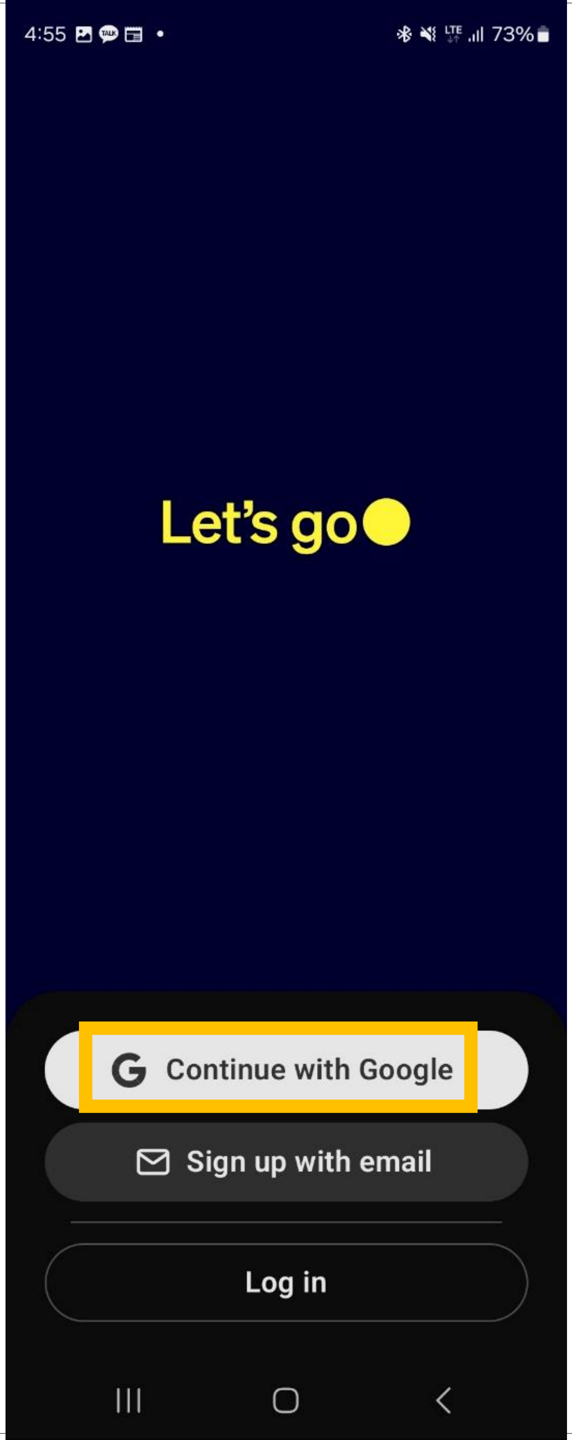
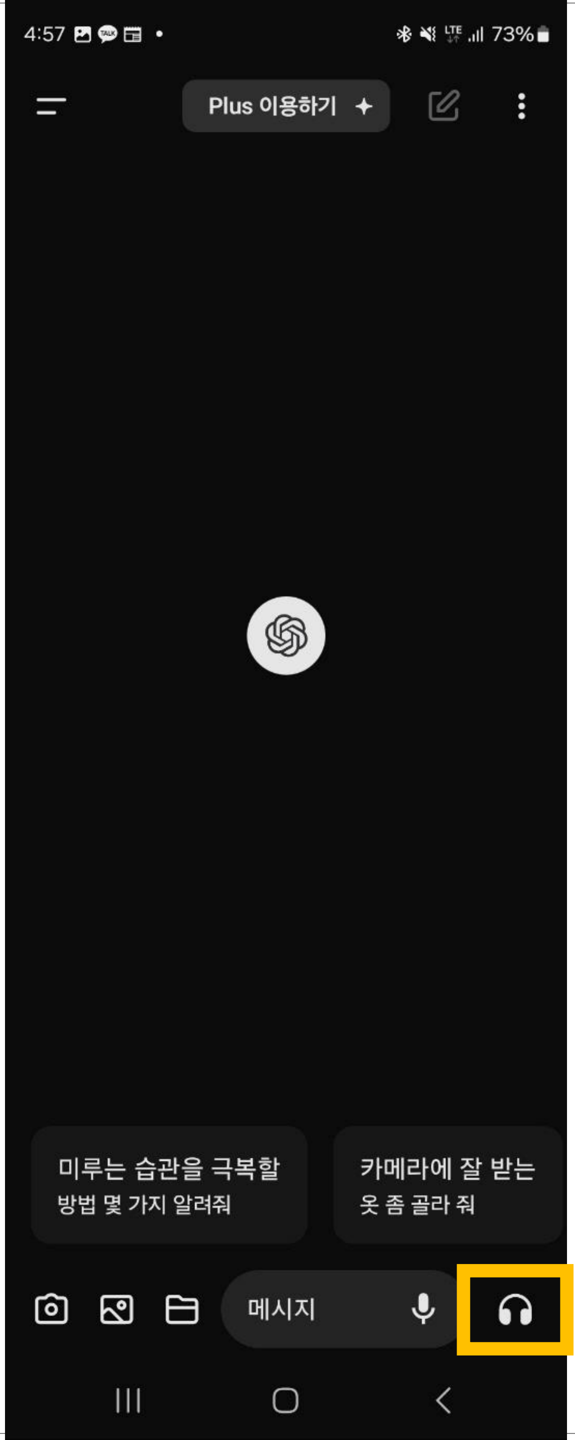
46. A type of second order central moment related to the means of the both variables. It measures how changes in one variable are associated with changes in another variable.
→ covariance (공분산)

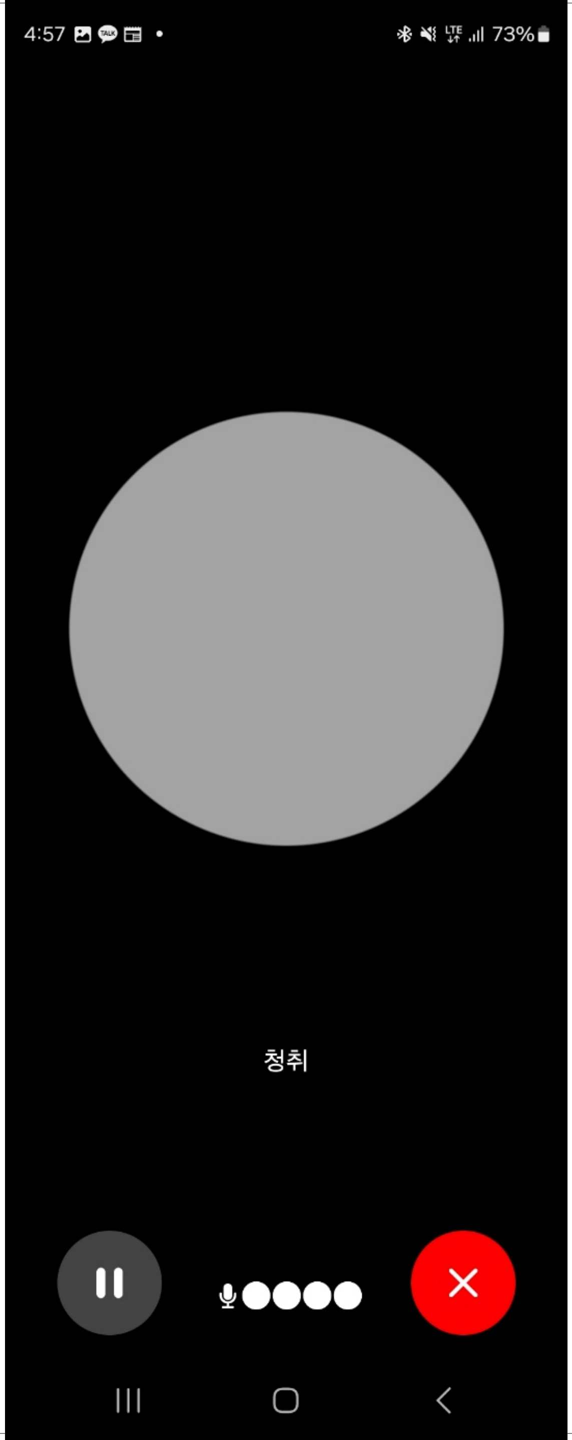
47. It scales the covariance by the standard deviations of the variables, resulting in a value between -1 and 1 that represents the strength and direction of the linear relationship
→ correlation coefficient (상관계수)

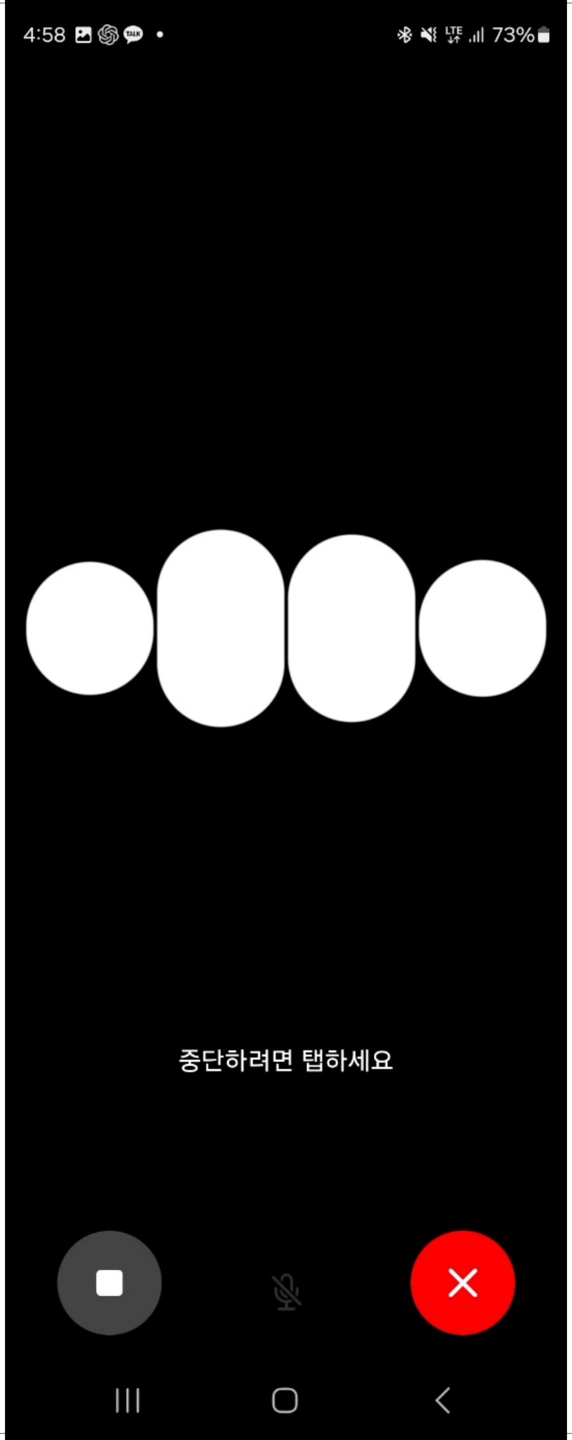
48. Multi-dimensional array of numbers. It is a mathematical object that represents a relationship between vectors, scalars, and other tensors. It is a generalization of vectors and matrices to higher-dimensional spaces. → tensor (텐서)

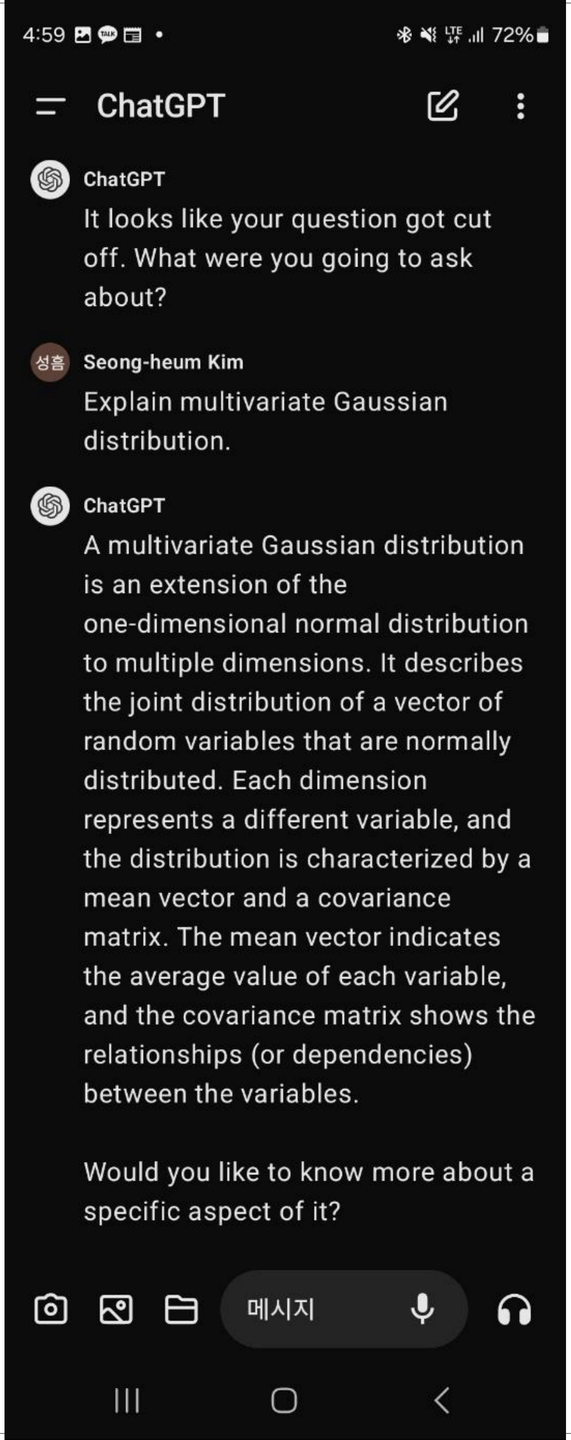
49. An array or list of random variables. It is a mathematical concept that represents a collection of random variables. It can be thought of as a vector where each component corresponds to a different random variable. → random vector (확률벡터)

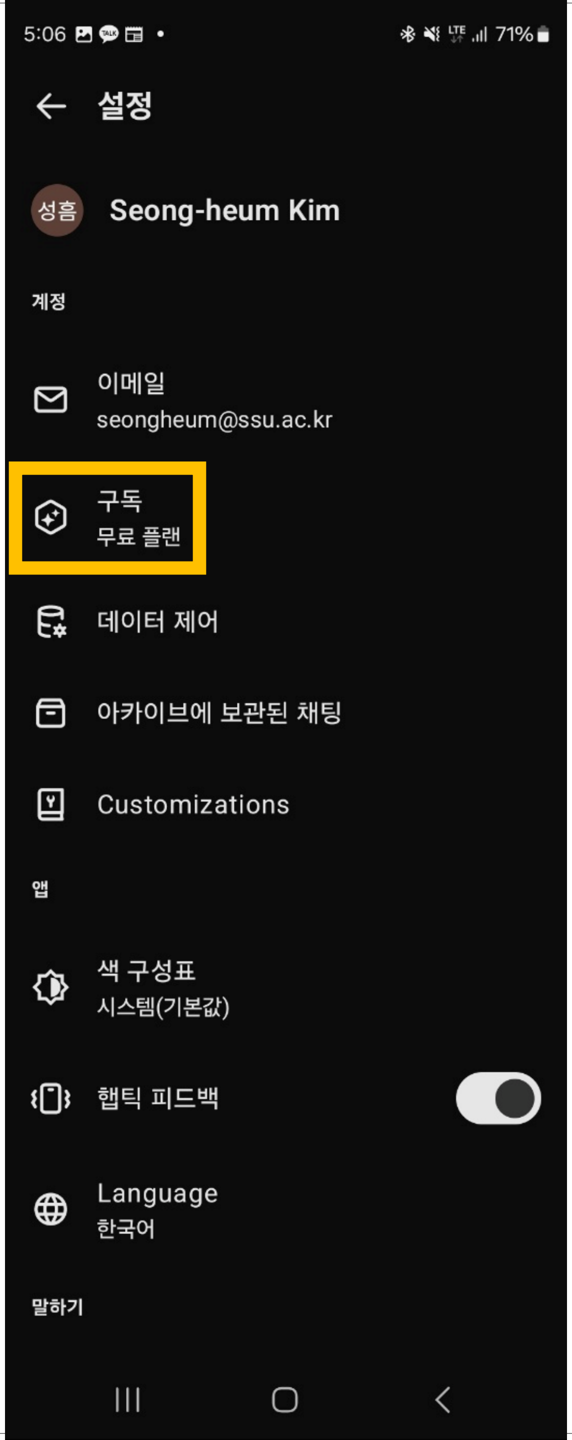
50. It is a probability distribution that describes the joint behavior of multiple Gaussian random variables. It is an extension of the univariate Gaussian distribution to higher dimensions. It is characterized by two parameters: a mean vector and a covariance matrix.
→ multivariate Gaussian distribution (다변량 가우시안 분포)

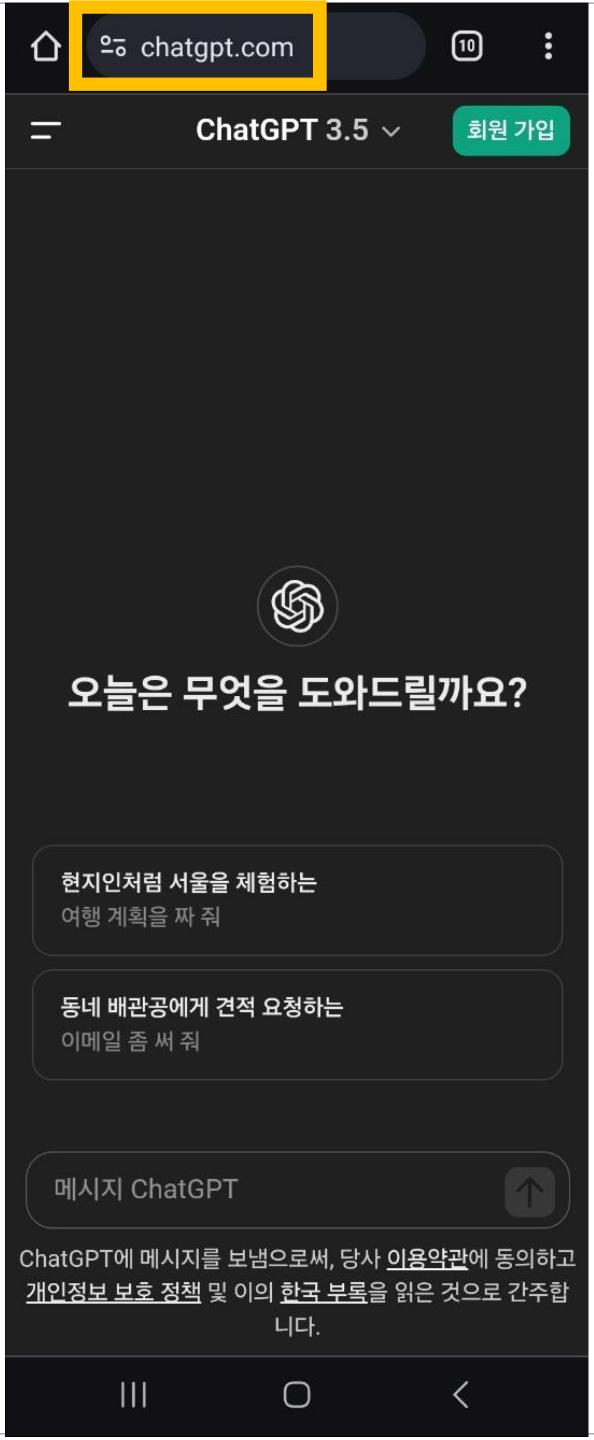

4:54
chatgpt app
Available in Google Play
설치
Google Play
ChatGPT
4.7 ★★★★★ (159만)
앱 정보. arrow_forward. ChatGPT 소개: OpenAI의 최신 발전 사항을 손쉽게 확인할 수 있습니다. 이 공식 앱은 무료이며 여러 기기에서 기록을 동기화하고 OpenAI의 최신 ...
설치
ChatGPT
앱
ChatGPT는 OpenAI가 개발한 프로토타입 대화형 인공지능 챗봇이다. ChatGPT는 대형 언어 모델 GPT-3의 개선판인 GPT-3.5를 기반으로 만들어졌으며, 지도 학습과 강화학습을 모두 사용해 파인 튜닝되었다.
위키백과 >
최초 출시일
2022년 11월 30일
프로그래밍 언어
파이썬
디스커버
검색
저장됨

4:55
chatgpt app
Available in Google Play
설치
Google Play
ChatGPT
4.7 ★★★★★ (159만)
Google Play
ChatGPT
OpenAI
인앱 구매
4.8★
리뷰 158만개
5000만회 이상
다운로드
12
12세 이상
열기
스폰서 · 관련 인기 앱
ChatOn - 지능적인 AI 챗봇...
4.4★
인공지능 챗봇 - 챗GPT 한글버전
4.1★
챗봇 AI - 인공지능 어시스턴트
4.4★

4:55
Let's go
Continue with Google
Sign up with email
Log in

4:57
Plus 이용하기
미루는 습관을 극복할 방법 몇 가지 알려줘
카메라에 잘 받는 옷 좀 골라 줘
메시지

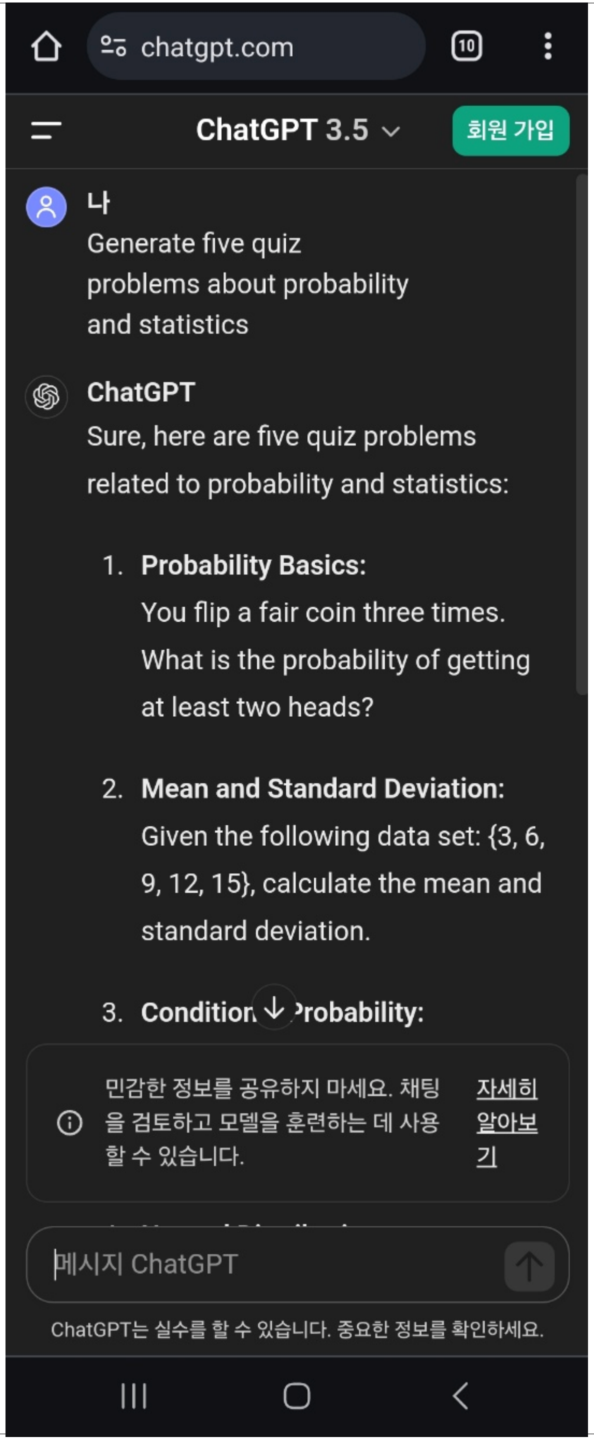

4:57
73%
청취


4:58
73%
중단하려면 탭하세요

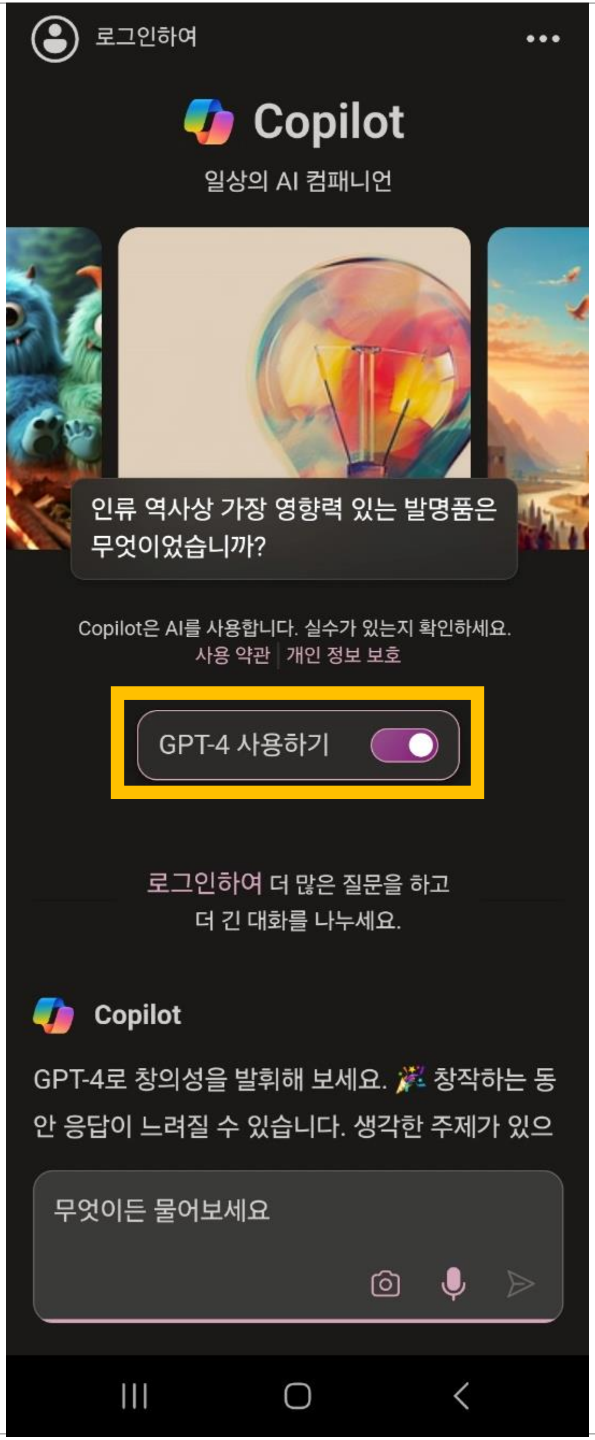

4:59
72%
ChatGPT
ChatGPT
It looks like your question got cut off. What were you going to ask about?
성흠
Seong-heum Kim
Explain multivariate Gaussian distribution.
ChatGPT
A multivariate Gaussian distribution is an extension of the one-dimensional normal distribution to multiple dimensions. It describes the joint distribution of a vector of random variables that are normally distributed. Each dimension represents a different variable, and the distribution is characterized by a mean vector and a covariance matrix. The mean vector indicates the average value of each variable, and the covariance matrix shows the relationships (or dependencies) between the variables.
Would you like to know more about a specific aspect of it?
메시지


5:06
71%
설정
성흠
Seong-heum Kim
계정
이메일
seongheum@ssu.ac.kr
구독
무료 플랜
데이터 제어
아카이브에 보관된 채팅
Customizations
앱
색 구성표
시스템(기본값)
햅틱 피드백
Language
한국어
말하기


chatgpt.com
ChatGPT 3.5
회원 가입
오늘은 무엇을 도와드릴까요?
현지인처럼 서울을 체험하는
여행 계획을 짜 줘
동네 배관공에게 견적 요청하는
이메일 좀 써 줘
메시지 ChatGPT
ChatGPT에 메시지를 보냄으로써, 당사 이용약관에 동의하고 개인정보 보호 정책 및 이의 한국 부록을 읽은 것으로 간주합니다.


chatgpt.com
ChatGPT 3.5
회원 가입
나
Generate five quiz problems about probability and statistics
ChatGPT
Sure, here are five quiz problems related to probability and statistics:
1. Probability Basics:
You flip a fair coin three times. What is the probability of getting at least two heads?
2. Mean and Standard Deviation:
Given the following data set: {3, 6, 9, 12, 15}, calculate the mean and standard deviation.
3. Condition Probability:
민감한 정보를 공유하지 마세요. 채팅을 검토하고 모델을 훈련하는 데 사용할 수 있습니다.
자세히 알아보기
메시지 ChatGPT
ChatGPT는 실수를 할 수 있습니다. 중요한 정보를 확인하세요.


ms copilot app
Google Play
Microsoft Copilot
4.7 ★★★★★ (23만)
Copilot is a pioneering personal AI chat assistant, powered by the latest OpenAI models and DALL·E 3. Copilot uses AI technologies that provide fast, complex, ...
설치
마이크로소프트 코파일럿
앱
마이크로소프트 코파일럿은 마이크로소프트에서 개발한 대형 언어 모델 기반 챗봇이다. 2023년 2월 7일에 마이크로소프트 엣지의 내장 확장 기능으로 빙 챗으로 출시되었으며 코타나의 간접적인 후속 제품이다.
위키백과 >
최초 출시일
2023년 2월 7일
개발
마이크로소프트
관련 질문
Is MS Copilot available now?
디스커버
검색
저장됨


로그인하여
Copilot
일상의 AI 컴패니언
인류 역사상 가장 영향력 있는 발명품은 무엇이었습니까?
Copilot은 AI를 사용합니다. 실수가 있는지 확인하세요.
사용 약관 | 개인 정보 보호
GPT-4 사용하기
로그인하여 더 많은 질문을 하고
더 긴 대화를 나누세요.
Copilot
GPT-4로 창의성을 발휘해 보세요. 🎉 창작하는 동안 응답이 느려질 수 있습니다. 생각한 주제가 있으
무엇이든 물어보세요

과제 범위

- **Assignment #3 (~5/24)**
  - 연습 16. Problems 5.3
  - 연습 17. Problems 5.9
  - 연습 18. Problems 5.41
  - 연습 19. Problems 5.45
  - 연습 20. Problems 6.24
  - 연습 21. Problems 7.5
  - 연습 22. Problems 7.7
  - 실습 7. 분류 에러 분석 실습
  - 실습 8. 피어슨의 적합도 검정

연습 16. Problems 5.3

5.3 Consider the joint PMF specified below:

$i_1 \rightarrow$ (columns), $i_2 \downarrow$ (rows)

$f_{I_1 I_2}[i_1, i_2]$	1	2	3	4
1	$6/32$	$2/32$	$4/32$	$3/32$
2	$2/32$	$3/32$	$2/32$	$1/32$
3	$1/32$	$1/32$	$2/32$	0
4	$1/32$	$1/32$	$1/32$	$2/32$

(a) Determine the marginal PMFs $f_{I_1}(i_1)$ and $f_{I_2}(i_2)$.

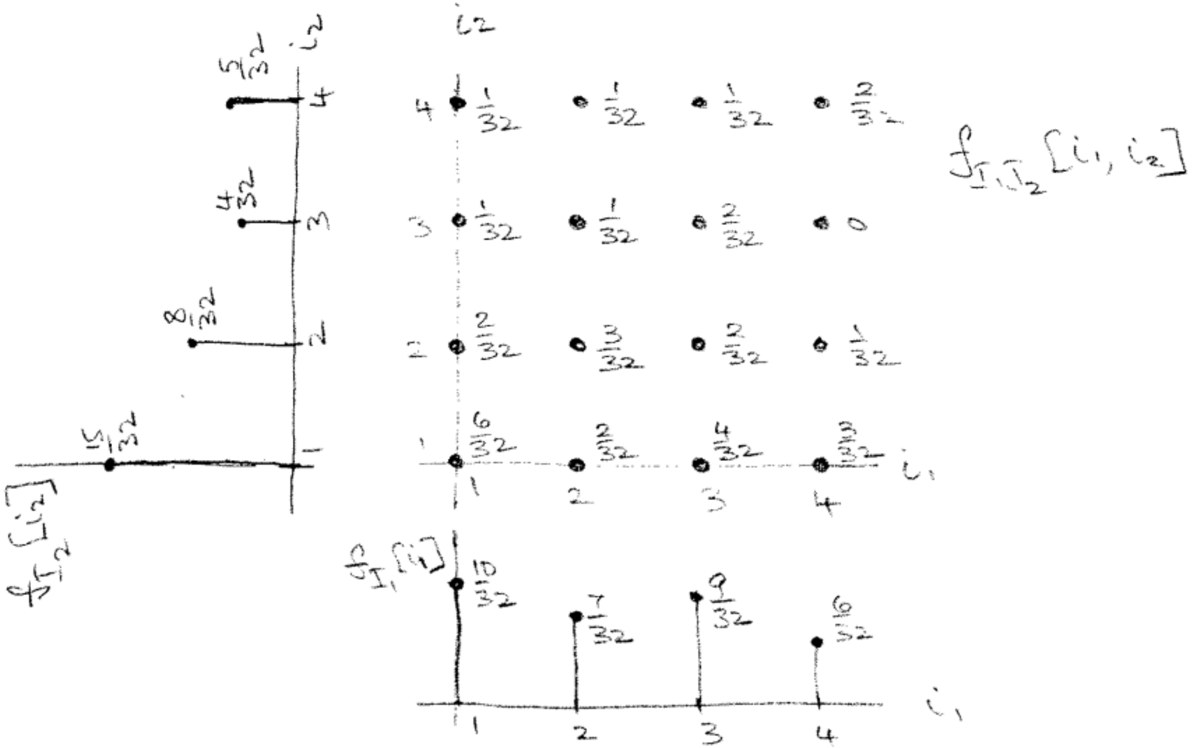
(b) Determine the conditional PMF $f_{I_1|I_2}(i_1|i_2)$.

(c) Find the CDF $F_{I_1,I_2}[i_1, i_2]$.

연습 16. Problems 5.3

5.3 Consider the joint PMF specified below:

$f_{I_1 I_2}[i_1,i_2]$ ($i_2 \downarrow$, $i_1 \rightarrow$)	1	2	3	4
1	$6/32$	$2/32$	$4/32$	$3/32$
2	$2/32$	$3/32$	$2/32$	$1/32$
3	$1/32$	$1/32$	$2/32$	0
4	$1/32$	$1/32$	$1/32$	$2/32$



(a) Determine the marginal PMFs $f_{I_1}(i_1)$ and $f_{I_2}(i_2)$.

$$f_{I_1}[i_1] = \sum_{i_2=1}^{4} f_{I_1,I_2}[i_1,i_2], \quad f_{I_2}[i_2] = \sum_{i_1=1}^{4} f_{I_1,I_2}[i_1,i_2]$$

Problem 5.3

5.3 Consider the joint PMF specified below:

$f_{I_1I_2}[i_1,i_2]$	1	2	3	4
1	$^6\!/_{32}$	$^2\!/_{32}$	$^4\!/_{32}$	$^3\!/_{32}$
2	$^2\!/_{32}$	$^3\!/_{32}$	$^2\!/_{32}$	$^1\!/_{32}$
3	$^1\!/_{32}$	$^1\!/_{32}$	$^2\!/_{32}$	0
4	$^1\!/_{32}$	$^1\!/_{32}$	$^1\!/_{32}$	$^2\!/_{32}$

($i_1 \rightarrow$ across columns; $i_2 \downarrow$ down rows)

(b) Determine the conditional PMF $f_{I_1|I_2}(i_1|i_2)$.

(c) Find the CDF $F_{I_1,I_2}[i_1,i_2]$.

(b)

$$f_{I_1|I_2}[i_1|i_2] = \frac{f_{I_1,I_2}[i_1,i_2]}{f_{I_2}[i_2]}$$

$$f_{I_1|I_2}[i_1|i_2=1] = \frac{f_{I_1,I_2}[i_1,1]}{f_{I_2}[1]} =$$

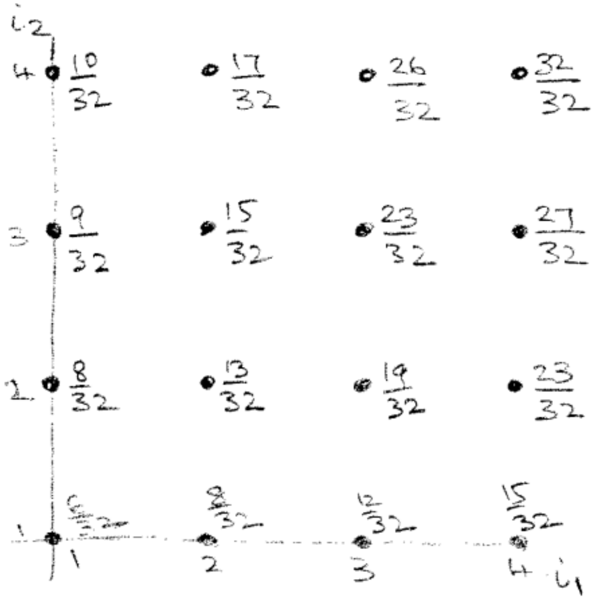
$$f_{I_2|I_1}[i_1|i_2=2] = \frac{f_{I_1,I_2}[i_1,2]}{f_{I_2}[2]} =$$

$\vdots$

$i_2 \downarrow$ \ $i_1 \rightarrow$	1	2	3	4
1	$\frac{6}{15}$	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{3}{15}$
2	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{1}{8}$
3	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$	0
4	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

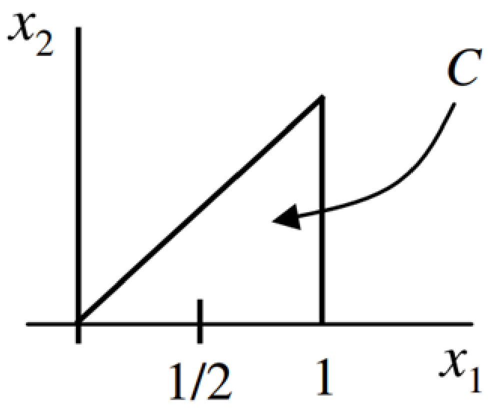
(c)

$$F_{I_1I_2}[i_1,i_2] = \sum_{l_1=1}^{i_1}\sum_{l_2=1}^{i_2} f_{I_1,I_2}[i_1,i_2]$$



Example 5.3

Example 5.3: Two random variables are uniformly distributed over the region of the x_1, x_2 plane shown below. Find the constant C and the probability of the event "$X_1 > \frac{1}{2}$." Also find the two marginal PDFs.



$$f_{X_1X_2}(x_1, x_2) = C$$
$$0 \le x_2 \le x_1 \le 1$$

The constant C is found by integrating the density over the region where it is nonzero and setting that equal to 1.

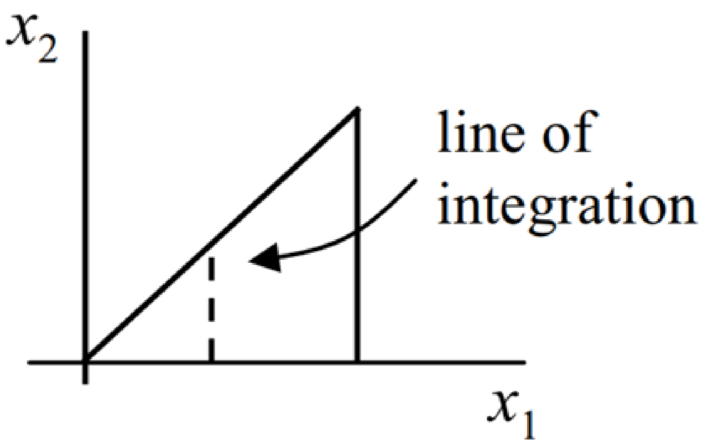
$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1X_2}(x_1, x_2)dx_2dx_1 &= \int_0^1\int_0^{x_1} C\,dx_2dx_1 \\ &= C\int_0^1 x_1\,dx_1 = C\,\frac{x_1^2}{2}\bigg|_0^1 = \frac{C}{2} = 1\end{aligned}$$

Example 5.3

The probability that $X_1 > 1/2$ is obtained by integrating the joint density over the appropriate region.

$$\Pr[X_1 > 1/2] = \int_{1/2}^{1}\int_{0}^{x_1} 2\,dx_2 dx_1 = \int_{1/2}^{1} 2x_1\,dx_1 = x_1^2\Big|_{1/2}^{1} = 3/4$$

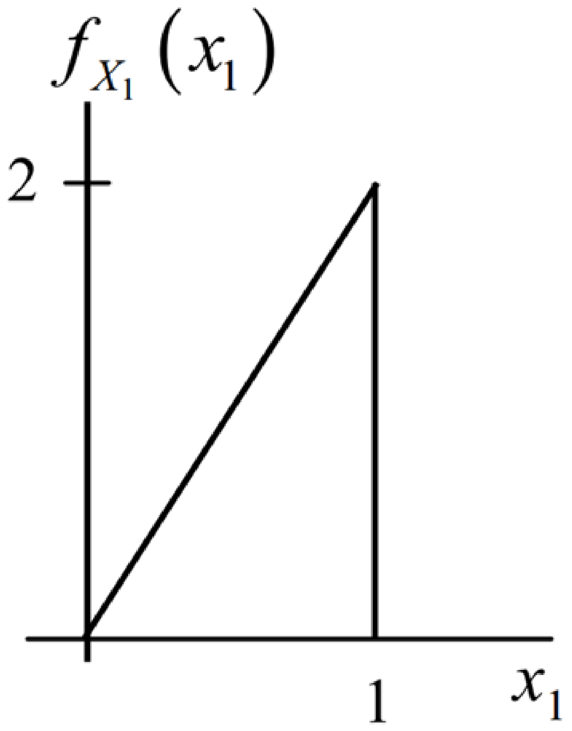
Finally, the marginals are obtained by integrating the joint PDF over each of the other variables. To obtain $f_{X_1}(x_1)$ the joint density function is integrated as follows:



$$f_{X_1}(x_1) = \int_0^{x_1} 2\,dx_2 = 2x_1$$

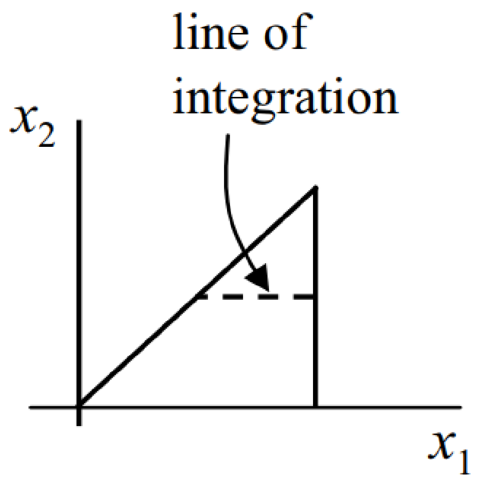
Example 5.3

The density function is sketched below.



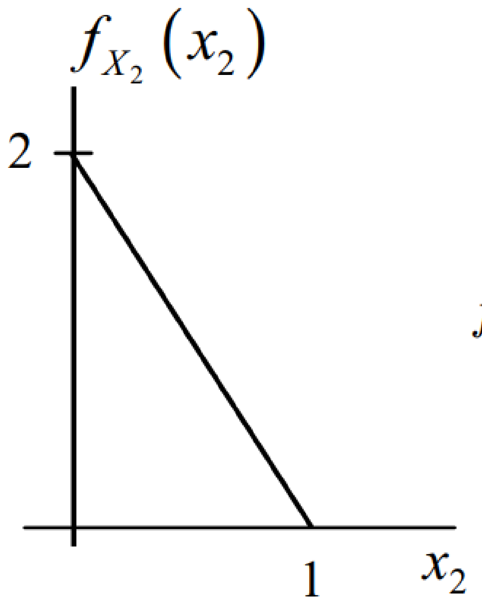
$$f_{X_1}(x_1) = 2x_1 \qquad 0 \leq x_1 \leq 1$$

To obtain $f_{X_2}(x_2)$ the integration is as shown below:



$$f_{X_2}(x_2) = \int_{x_2}^{1} 2\,dx_1 = 2(1-x_2)$$

The density function is sketched below.



$$f_{X_2}(x_2) = 2(1-x_2) \qquad 0 \leq x_2 \leq 1$$

Notice that for both of these marginal densities the limits of definition $0 \leq x_1 \leq 1$ or $0 \leq x_2 \leq 1$ are very important.

Example 5.4

Example 5.4: Let us continue with the random variables described in Example 5.3. In particular, let us check for independence of the random variables and determine the two conditional PDFs.

The product of the marginal densities computed in the previous example is

$$f_{X_1}(x_1) \cdot f_{X_2}(x_2) = 2x_1 \cdot 2(1 - x_2)$$

This is clearly not equal to the joint density $f_{X_1 X_2}(x_1, x_2)$, which is a constant $(C = 2)$ over the region of interest. Therefore the random variables are *not* independent.

연습 17. Problems 5.9

5.9 The joint PDF $f_{X_1,X_2}(x_1, x_2)$ of two random variables X_1 and X_2 is given by

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} C(4 - x_1 x_2), & 0 \le x_1 \le 4, 0 \le x_2 \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find C to make this a valid PDF.

(b) Find the marginal density functions of X_1 and X_2. Clearly define the ranges of values they take.

(c) Are the random variables independent?

연습 17. Problems 5.9

5.9 The joint PDF $f_{X_1,X_2}(x_1,x_2)$ of two random variables X_1 and X_2 is given by

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} C(4-x_1x_2), & 0 \le x_1 \le 4, 0 \le x_2 \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find C to make this a valid PDF.

$$C\int_0^4\int_0^1 (4-x_1x_2)\,dx_2\,dx_1 = 1$$

$$C\int_0^4 \left(4-\frac{x_1}{2}\right) dx_1 = C(16-4) = 1$$

$$C = 1/12$$

연습 17. Problems 5.9

(b) Find the marginal density functions of X_1 and X_2. Clearly define the ranges of values they take.

(c) Are the random variables independent?

$$f_{X_1}(x_1) = \frac{1}{12}\int_0^1 (4 - x_1 x_2)\, dx_2$$

$$= \frac{1}{12}\left(4 - \frac{x_1}{2}\right) \qquad 0 \le x_1 \le 4$$

$$f_{X_2}(x_2) = \frac{1}{12}\int_0^4 (4 - x_1 x_2)\, dx_1$$

$$= \frac{1}{12}(16 - 8x_2)$$

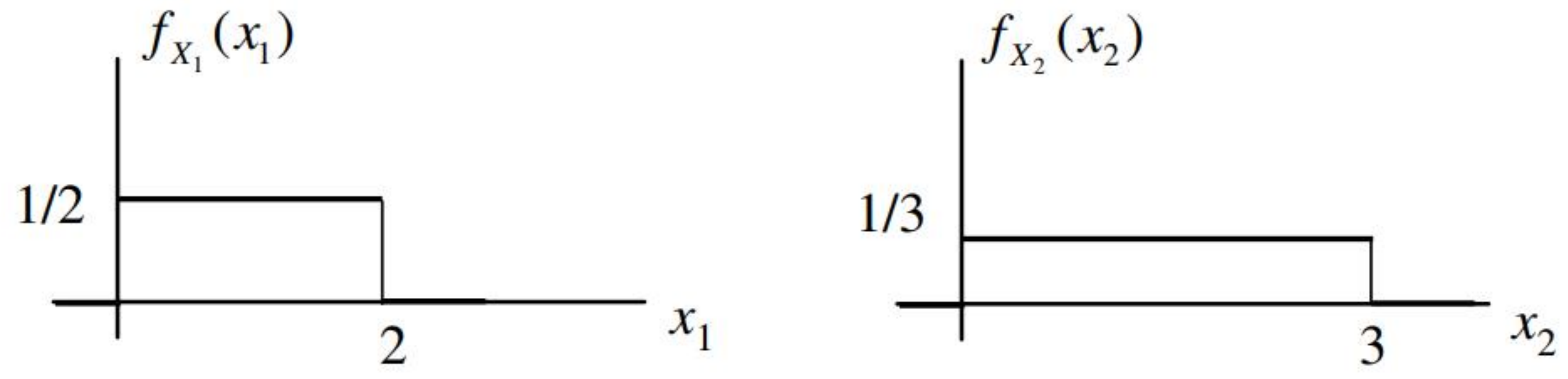
$$= \frac{2}{3}(2 - x_2), \qquad 0 \le x_2 \le 1$$

$$f_{X_1 X_2}(x_1, x_2) \overset{?}{=} f_{X_1}(x_1)\, f_{X_2}(x_2)$$

Not independent

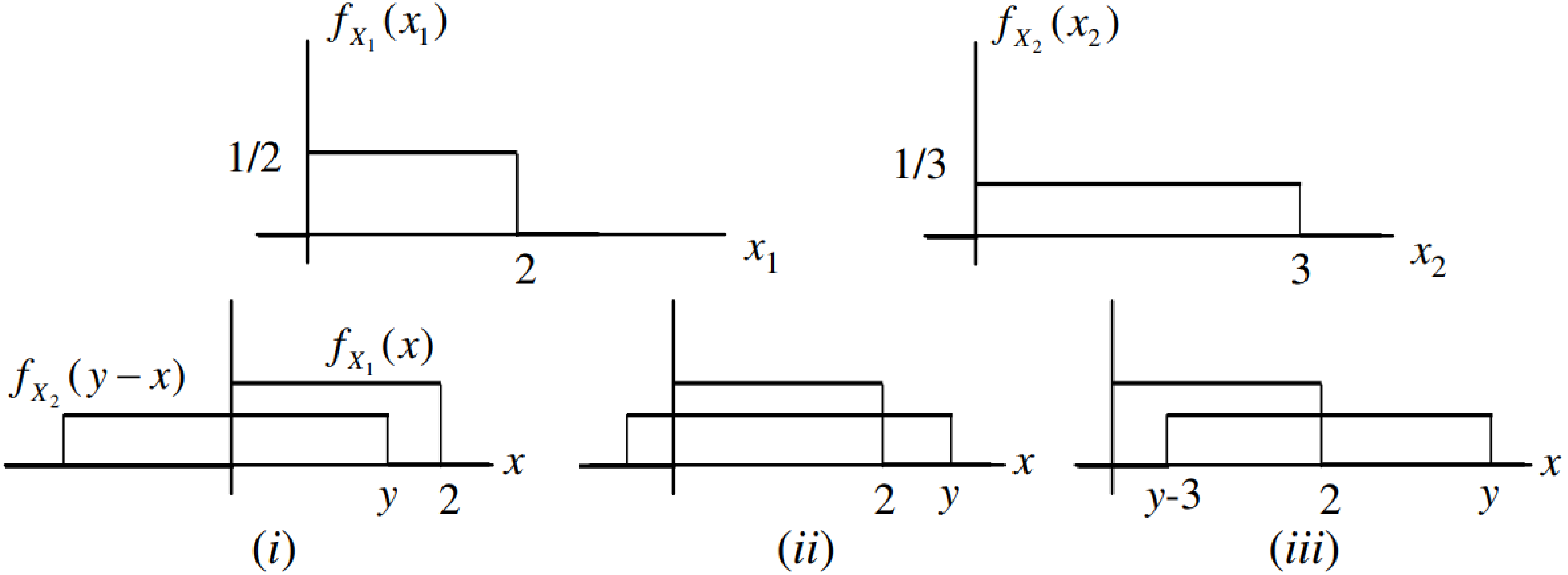
Example 5.8

Example 5.8: Suppose that X_1 and X_2 are two independent uniform random variables with the density functions shown below.



Since X_1 and X_2 are independent, (5.47) applies; the convolution is given by the integral (5.46) where the density f_{X_2} is reversed and slid over f_{X_1}. The arguments of the integral are shown below for a fixed value of y in three different cases.

Example 5.8



In each case we need to integrate only over the "overlap" region, i.e., the region where both functions are nonzero. The specific integrations are:

(i) for $0 \le y \le 2$

$$f_Y(y) = \int_0^y \tfrac{1}{2} \cdot \tfrac{1}{3}\, dx = \tfrac{1}{6} y$$

Example 5.8

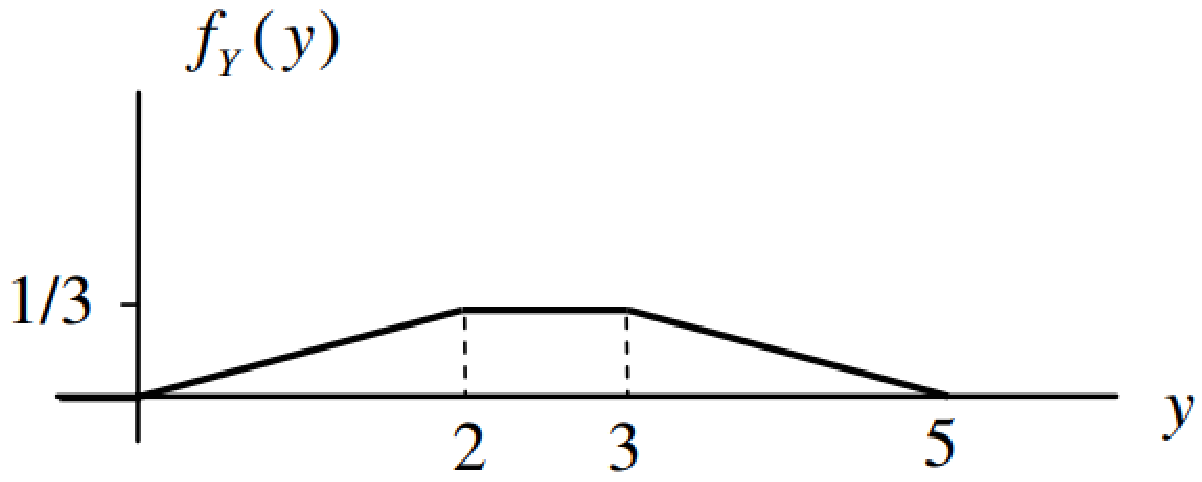
(ii) for $2 < y \le 3$

$$f_Y(y) = \int_0^2 \tfrac{1}{2} \cdot \tfrac{1}{3}\, dx = \tfrac{1}{3}$$

(iii) for $3 < y \le 5$

$$f_Y(y) = \int_{y-3}^2 \tfrac{1}{2} \cdot \tfrac{1}{3}\, dx = \tfrac{1}{6}(5 - y)$$

For all other values of y the integrands have no overlap, so f_Y is 0. The complete density function for Y is sketched below.



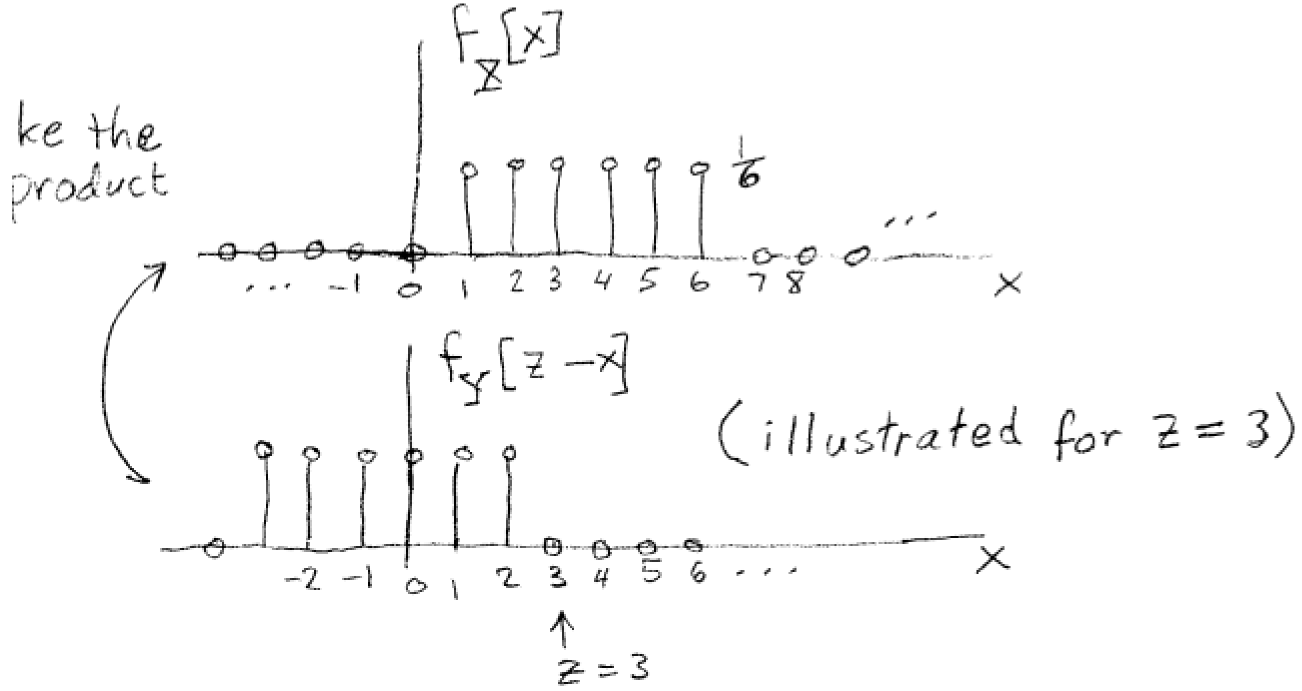
연습 18. Problems 5.41

Sums of common random variables

5.41 Let X and Y be the number shown on each of two dice and let Z be the sum ($Z = X + Y$). Assume that X and Y are independent and each is uniformly distributed over the integers 1 through 6. Using discrete convolution, show that the PMF for Z has a triangular shape. Compare it to Fig. 3.2 of Chapter 3.

Since X and Y are independent, we can write

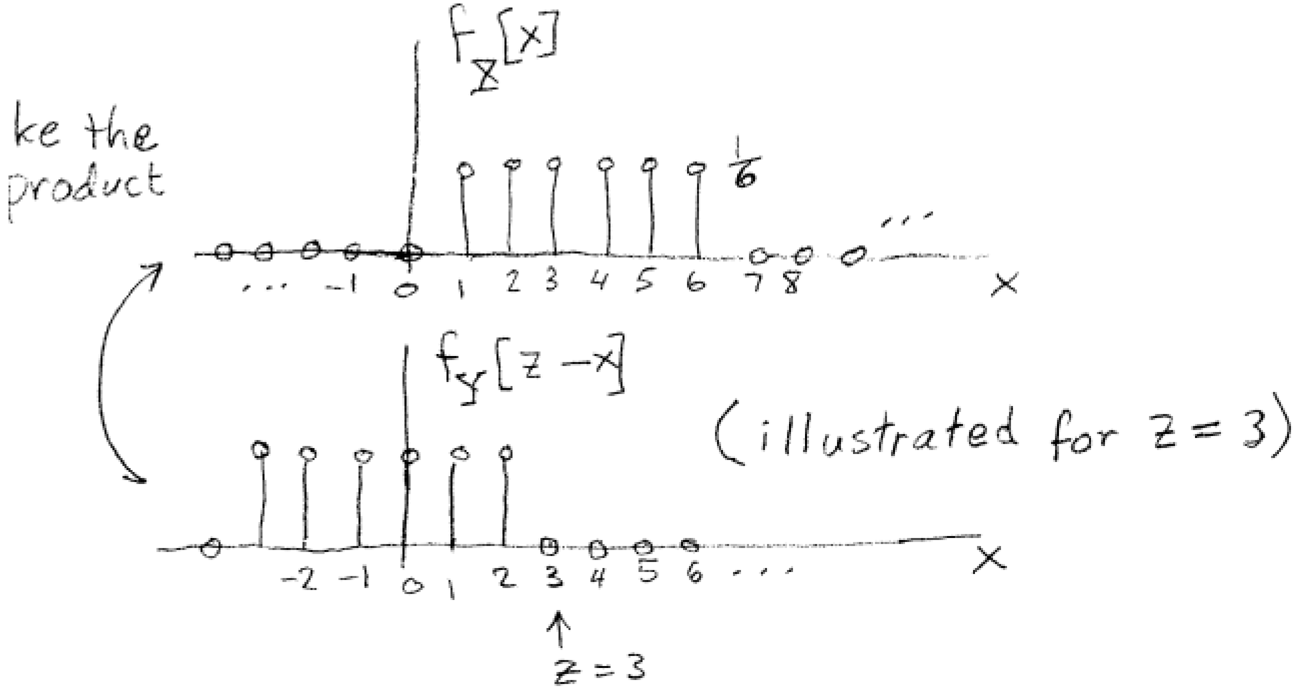
$$f_Z[z] = \sum_x f_X[x]\, f_Y[z-x]$$

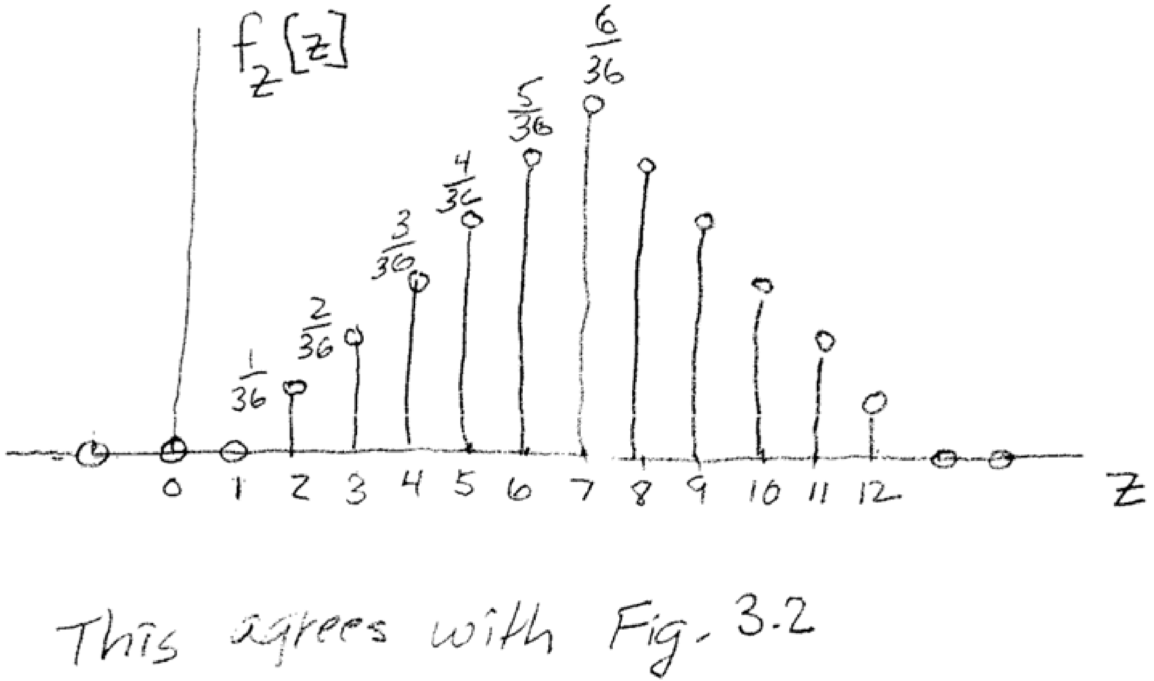


연습 18. Problems 5.41

Sums of common random variables

5.41 Let X and Y be the number shown on each of two dice and let Z be the sum ($Z = X + Y$). Assume that X and Y are independent and each is uniformly distributed over the integers 1 through 6. Using discrete convolution, show that the PMF for Z has a triangular shape. Compare it to Fig. 3.2 of Chapter 3.





Sum of Two Random Variables

- **PDF for a Sum of Two Random Variables**
  - Example) Exponential random variables
    - X_1 and X_2 are lifetimes of two light bulbs that are used sequentially.
    - The lifetimes X_i are independent and identically distributed:

      → $f_{X_i}(x_i) = \lambda e^{-\lambda x_i} \quad (x_i \geq 0)$

    - Combined lifetime: $Y = X_1 + X_2$

      → PDF: $f_{X_1} * f_{X_2}$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(x_1)\, f_{X_2}(y - x_1) dx_1$$

$$= \int_0^y \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda(y - x_1)}\, dx_1 = \lambda^2 e^{-\lambda y} \int_0^y dx_1$$

$$= \lambda^2 y e^{-\lambda y} \quad (y \geq 0) \quad [\text{ Erlang distribution }]$$

연습 19. Problems 5.45

5.45 The sum of two independent random variables X_1 and X_2 is given by

$$X = X_1 + X_2$$

where X_1 is an exponential random variable with parameter $\lambda = 2$, and X_2 is another exponential random variable with parameter $\lambda = 3$.

(a) Find the mean and the variance of X.

(b) Determine the PDF of X.

(a) $E\{X\} = E\{X_1\} + E\{X_2\} = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$

$Var[X] = Var[X_1] + Var[X_2] = \frac{1}{(2)^2} + \frac{1}{(3)^2} = \frac{13}{36}$

연습 19. Problems 5.45

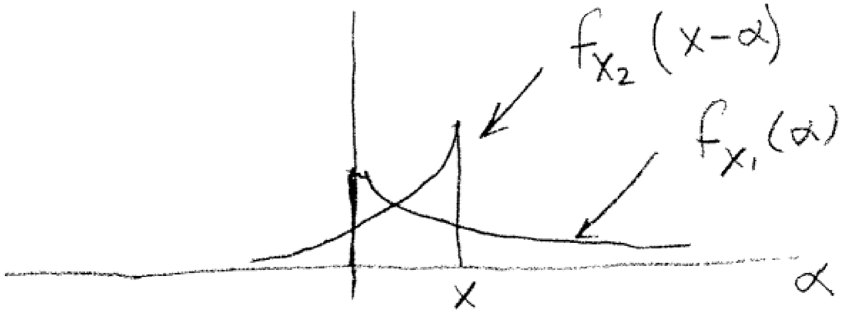
5.45 The sum of two independent random variables X_1 and X_2 is given by

$$X = X_1 + X_2$$

where X_1 is an exponential random variable with parameter $\lambda = 2$, and X_2 is another exponential random variable with parameter $\lambda = 3$.

(b) Determine the PDF of X.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X_1}(\alpha)\, f_{X_2}(x-\alpha)\, d\alpha$$



$$f_X(x) = \int_0^x 2e^{-2\alpha} \cdot 3e^{-3(x-\alpha)}\, d\alpha$$

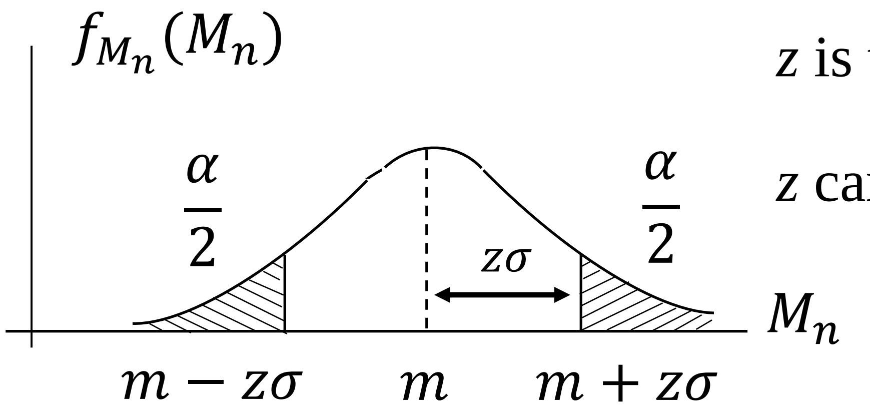
$$= 6e^{-3x} \int_0^x e^{\alpha}\, d\alpha$$

$$= 6e^{-3x}(e^{x} - 1) = 6(e^{-2x} - e^{-3x}) \qquad x \geq 0$$

$$f_X(x) = 0 \qquad x < 0$$

Confidence Intervals

- **Example 6.9**
  - CLT for sample mean
    - A set of IID random variables have unknown mean and variance 24.
    - The sample mean for 100 of these random variables is found to be 2.75
    - Find the corresponding 95% confidence interval (significant level: 0.05).
  - Calculation of the confidence interval (CI)
    - The sample mean can be assumed Gaussian with $\sigma = \sqrt{24/100} \cong 0.490$



z is the value needed to achieve probability $\alpha/2$

z can be found from Q tables:

$$\alpha/2 = 0.025 \rightleftarrows z = 1.96$$

The 95% CI is (2.75–1.96×0.490, 2.75+1.96×0.490) = (1.79, 3.71)

연습 20. Problems 6.24

Confidence Intervals

6.24 Five hundred observations of a random variable X with variance $\sigma_X^2 = 25$ are taken. The sample mean based on 500 samples is computed to be $M_{500} = 3.25$. Find 95% and 98% confidence intervals for this estimate.

The variance of the sample mean is

$$\sigma_{500}^2 = \frac{25}{500} = 0.05 \longrightarrow \sigma_{500} = 0.2236$$

For the 95% CI we can use the value

$z = 1.96$ from the table in Fig. 6.6

The upper and lower limits are $3.25 \pm (1.96)(0.2236)$

The CI is therefore $(2.81, 3.69)$

연습 20. Problems 6.24

Confidence Intervals

6.24 Five hundred observations of a random variable X with variance $\sigma_X^2 = 25$ are taken. The sample mean based on 500 samples is computed to be $M_{500} = 3.25$. Find 95% and 98% confidence intervals for this estimate.

For the 98% CI, we use the table of Q functions to find $Q(2.325) = 0.01$, which is half the probability of the critical region. With $z = 2.325$ the computation is $3.25 \pm (2.325)(0.2236)$

or $(2.73, 3.77)$

Example 7.3

Example 7.3: Given a set of four two-dimensional sample vectors

$$\boldsymbol{X}_1 = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \quad \boldsymbol{X}_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \quad \boldsymbol{X}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{X}_4 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

the sample mean is computed as

$$\boldsymbol{M}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i = \frac{1}{4}\left\{\begin{bmatrix} 2 \\ -2 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \end{bmatrix}\right\} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}$$

The sample correlation matrix is

$$\begin{aligned} \boldsymbol{R}_n &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T \\ &= \frac{1}{4}\left\{\begin{bmatrix} 2 \\ -2 \end{bmatrix}\begin{bmatrix} 2 & -2 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \end{bmatrix}\begin{bmatrix} -2 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \end{bmatrix}\begin{bmatrix} 0 & 2 \end{bmatrix}\right\} \\ &= \begin{bmatrix} \frac{9}{4} & -\frac{3}{2} \\ -\frac{3}{2} & \frac{9}{4} \end{bmatrix} \end{aligned}$$

Example 7.3

Finally, the (unbiased) sample covariance matrix is

$$
\begin{aligned}
\boldsymbol{C}_n &= \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i-\boldsymbol{M}_n)(\boldsymbol{X}_i-\boldsymbol{M}_n)^T \\
&= \frac{1}{3}\left\{\begin{bmatrix} \frac{7}{4} \\ -\frac{9}{4}\end{bmatrix}\begin{bmatrix} \frac{7}{4} & -\frac{9}{4}\end{bmatrix} + \begin{bmatrix} -\frac{9}{4} \\ \frac{3}{4}\end{bmatrix}\begin{bmatrix} -\frac{9}{4} & \frac{3}{4}\end{bmatrix} + \begin{bmatrix} \frac{3}{4} \\ -\frac{1}{4}\end{bmatrix}\begin{bmatrix} \frac{3}{4} & -\frac{1}{4}\end{bmatrix} + \begin{bmatrix} -\frac{1}{4} \\ \frac{7}{4}\end{bmatrix}\begin{bmatrix} -\frac{1}{4} & \frac{7}{4}\end{bmatrix}\right\} \\
&= \begin{bmatrix} \frac{35}{12} & -\frac{25}{12} \\ -\frac{25}{12} & \frac{35}{12}\end{bmatrix}
\end{aligned}
$$

or, using (7.21)

$$
\begin{aligned}
\boldsymbol{C}_n &= \frac{n}{n-1}(\boldsymbol{R}_n - \boldsymbol{M}_n\boldsymbol{M}_n^T) \\
&= \frac{4}{3}\left(\begin{bmatrix} \frac{9}{4} & -\frac{3}{2} \\ -\frac{3}{2} & \frac{9}{4}\end{bmatrix} - \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4}\end{bmatrix}\begin{bmatrix} \frac{1}{4} & \frac{1}{4}\end{bmatrix}\right) = \begin{bmatrix} \frac{35}{12} & -\frac{25}{12} \\ -\frac{25}{12} & \frac{35}{12}\end{bmatrix}
\end{aligned}
$$

연습 21. Problems 7.5

7.5 Find the sample mean, sample correlation matrix, and sample covariance matrix (7.21) for the following data:

$$X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad X_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad X_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad X_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Check your results using (7.22).

$$M_4 = \frac{1}{4}\left[\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right] = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$R_4 = \frac{1}{4}\left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}\begin{bmatrix} 2 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \end{bmatrix}\right] = \begin{bmatrix} 3/2 & 3/4 \\ 3/4 & 3/2 \end{bmatrix}$$

연습 21. Problems 7.5

7.5 Find the sample mean, sample correlation matrix, and sample covariance matrix (7.21) for the following data:

$$X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad X_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad X_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad X_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$C_4 = \frac{1}{3}\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix}\begin{bmatrix} 1 & -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \end{bmatrix}\begin{bmatrix} -1 & 0 \end{bmatrix}\right]$$

$$= \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$$

연습 21. Problems 7.5

7.5 Find the sample mean, sample correlation matrix, and sample covariance matrix (7.21) for the following data:

$$X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad X_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad X_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad X_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Check your results using (7.22).

$$C_4 = \frac{4}{3}\left(R_4 - M_4 M_4^T\right)$$

$$= \frac{4}{3}\left[\begin{bmatrix} 3/2 & 3/4 \\ 3/4 & 3/2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix}\right] = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}$$

Examples of a Random Vector

- **Multivariate Gaussian Density Function**
  - For a K-dimensional Gaussian random vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_K]^T$

$$f_{X_1, X_2, X_2, \cdots, X_K}(\boldsymbol{x}) = f_X(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{K}{2}} |\mathbf{C}_X|^{\frac{1}{2}}} \exp -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{m}_X)^T \mathbf{C}_X^{-1} (\boldsymbol{x} - \boldsymbol{m}_X)$$

- **Bivariate Gaussian Density Function (K=2)**
  - For a 2-dimensional Gaussian random vector $\boldsymbol{x} = [x_1, x_2]^T$

$$f_X(\boldsymbol{x}) = \frac{1}{2\pi |\mathbf{C}_X|^{\frac{1}{2}}} \exp -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{m}_X)^T \mathbf{C}_X^{-1} (\boldsymbol{x} - \boldsymbol{m}_X)$$

$$\boldsymbol{m}_X = [m_1 \quad m_2]^T$$

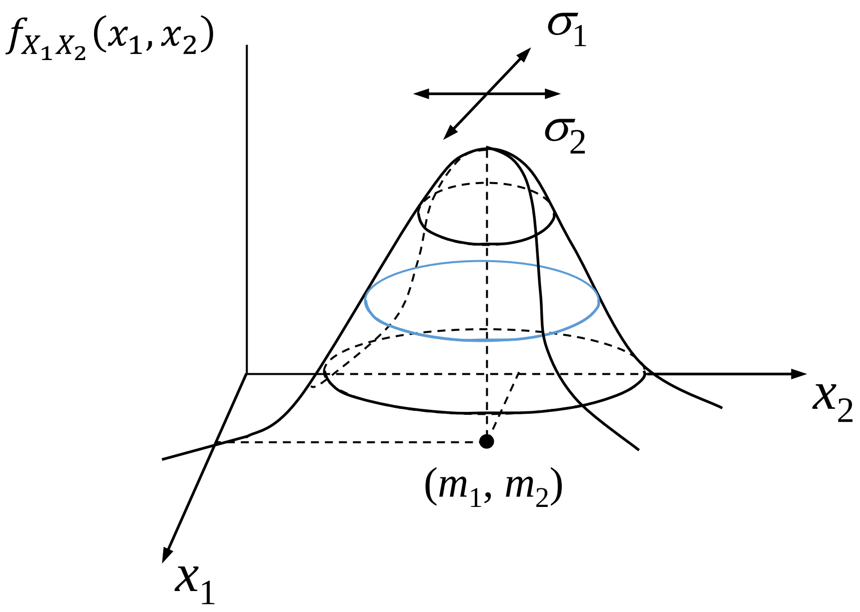
$$\rho = \frac{c_{12}}{\sigma_1 \sigma_2}$$

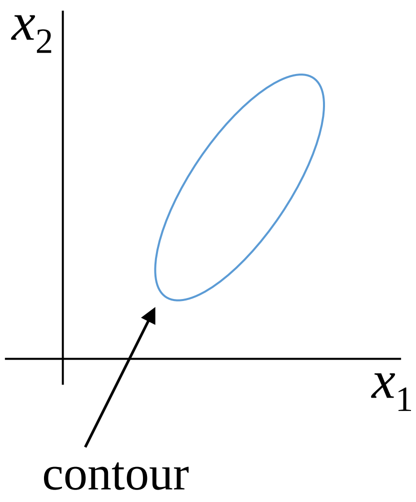
$$\mathbf{C}_X = \begin{bmatrix} {\sigma_1}^2 & c_{12} \\ c_{21} & {\sigma_2}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\mathbf{C}_{\boldsymbol{X}}^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}$$

Examples of a Random Vector

- **Bivariate Gaussian Density Function (K=2)**
  - Gaussian distribution with two variables
    - Two parameters (mean, standard deviation) of each dimension (x_1, x_2)
    - Covariance / correlation coefficient





$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right)$$

연습 22. Problems 7.7

7.7 The mean vector and covariance matrix for a Gaussian random vector $\boldsymbol{X}$ are given by

$$\mathbf{m}_{\boldsymbol{X}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{R}_{\boldsymbol{X}} = \begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix}$$

(a) Compute the covariance matrix $\mathbf{C}_{\boldsymbol{X}}$.

(b) What is the correlation coefficient $\rho_{X_1 X_2}$?

(c) Invert the covariance matrix and write an explicit expression for the Gaussian density function for $\boldsymbol{X}$.

연습 22. Problems 7.7

7.7 The mean vector and covariance matrix for a Gaussian random vector $\boldsymbol{X}$ are given by

$$\mathbf{m}_{\boldsymbol{X}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{R}_{\boldsymbol{X}} = \begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix}$$

(a) Compute the covariance matrix $\mathbf{C}_{\boldsymbol{X}}$.

$$C_X = R_X - \underline{m}_X \underline{m}_X^T$$

$$= \begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix}$$

연습 22. Problems 7.7

7.7 The mean vector and covariance matrix for a Gaussian random vector $\boldsymbol{X}$ are given by

$$\mathbf{m}_{\boldsymbol{X}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{R}_{\boldsymbol{X}} = \begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix}$$

(b) What is the correlation coefficient $\rho_{X_1 X_2}$?

$$\rho_{X_1 X_2} = \frac{E[(X_1 - m_{X_1})(X_2 - m_{X_2})]}{\sigma_{X_1}\ \sigma_{X_2}} = \frac{-2}{\sqrt{3}\sqrt{2}} = -\sqrt{\frac{2}{3}}$$

연습 22. Problems 7.7

7.7 The mean vector and covariance matrix for a Gaussian random vector $\boldsymbol{X}$ are given by

$$\mathbf{m}_{\boldsymbol{X}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{R}_{\boldsymbol{X}} = \begin{bmatrix} 4 & -1 \\ -1 & 3 \end{bmatrix}$$

(c) Invert the covariance matrix and write an explicit expression for the Gaussian density function for $\boldsymbol{X}$.

$$C_x^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & \frac{3}{2} \end{bmatrix}$$

$$f_{\underline{X}}(\underline{X}) = \frac{1}{2\pi |C_x|^{1/2}} e^{-\frac{1}{2}(\underline{X}-\underline{m}_x)^T C_x^{-1} (\underline{X}-\underline{m}_x)}$$

연습 22. Problems 7.7

(c) Invert the covariance matrix and write an explicit expression for the Gaussian density function for $\boldsymbol{X}$.

$$f_{\underline{X}}(\underline{X}) = \frac{1}{2\pi |C_X|^{1/2}} e^{-\frac{1}{2}(\underline{X}-\underline{m}_X)^T C_X^{-1} (\underline{X}-\underline{m}_X)}$$

N-D 가우시안 PDF 기본식과 함께
m_X 와 C_X 행렬식 그리고 역행렬 값까지 구하면 정답

$$= \frac{1}{2\pi\sqrt{2}}$$

$$\cdot \exp\left\{-\frac{1}{2}\left[(x_1-1)^2 + 2(x_1-1)(x_2-1) + \frac{3}{2}(x_2-1)^2\right]\right\}$$

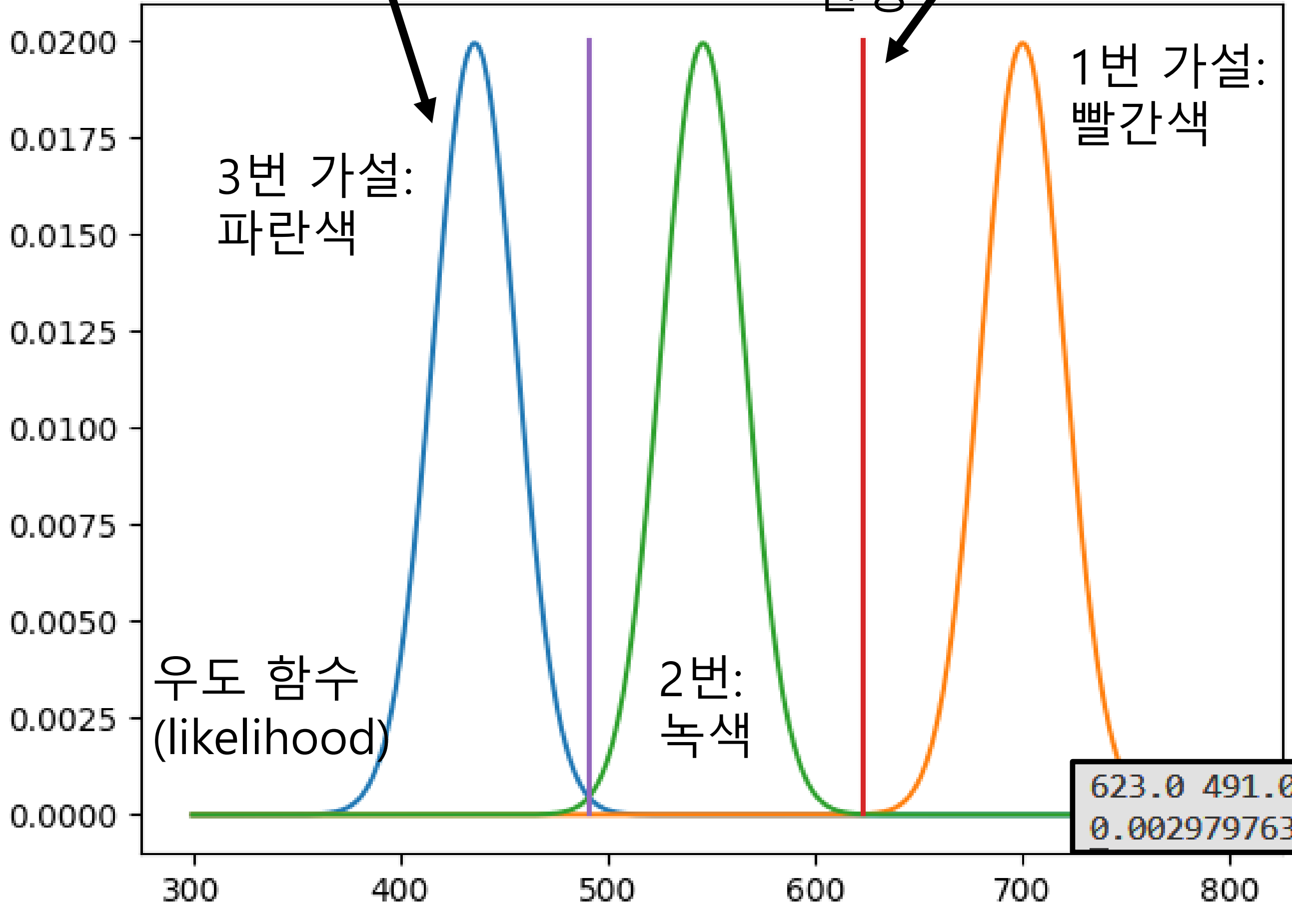
$$= \frac{1}{2\pi\sqrt{2}} \exp\left\{-\left[\frac{1}{2}(x_1-1)^2 + (x_1-1)(x_2-1) + \frac{3}{4}(x_2-1)^2\right]\right\}$$

실습 7. 공학적 응용 사례: 분류 에러 분석 → 코드/결과값 스크린샷

```
# example 3.22
H1_mu, H1_sig = 700, math.sqrt(400)
H2_mu, H2_sig = 546, math.sqrt(400)
H3_mu, H3_sig = 436, math.sqrt(400)
```

```
th12 = (H1_mu+H2_mu)/2
th23 = (H2_mu+H3_mu)/2
print(th12, th23)
```

판정




```
# simulation models
x = np.linspace(300, 800, 1000)
h1 = st.norm.pdf(x, H1_mu, H1_sig)
h2 = st.norm.pdf(x, H2_mu, H2_sig)
h3 = st.norm.pdf(x, H3_mu, H3_sig)
plt.plot(x, h3)
plt.plot(x, h1)
plt.plot(x, h2)
plt.plot([th12,th12], [0,0.02])
plt.plot([th23,th23], [0,0.02])
plt.show()
```

```
e3 = st.norm.sf((th23-H3_mu)/H3_sig, 0, 1)
e1 = st.norm.sf(-(th12-H1_mu)/H1_sig, 0, 1)
e2Left = st.norm.sf(-(th23-H2_mu)/H2_sig, 0, 1)
e2Right = st.norm.sf((th12-H2_mu)/H2_sig, 0, 1)
e2 = e2Left + e2Right
print(e3, e1, e2)
```

```
623.0 491.0
0.0029797632350545547 5.9058912418922374e-05 0.0030388221474 73477
```

실습 8. 카이제곱 검정 실습 (적합도) → 코드/결과값 스크린샷

- **피어슨의 적합도 검정**
  - 관측값(또는 표본값)이 가정한 확률분포의 값과 잘 일치하는지, 또는 가정이 적합한지 판단
    - 검정통계량 $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$
    - 예) 지금 내 주사위의 각 경우는 모두 1/6의 확률로 동일한 것인지?

H_0: 주사위의 값의 확률분포는 uniform 하다.

기대도수 = [100, 100, ..., 100]
→ 관찰값과 비교하여 적합한지 검정

```
import pandas as pd
import scipy.stats as stats
f_ob1 = [99, 101, 102, 97, 101, 100]
f_ob2 = [117, 119, 120, 115, 119, 10]
f_exp = [100, 100, 100, 100, 100, 100] # uniform
print("1:", stats.chisquare(f_ob1, f_exp=f_exp))
print("2:", stats.chisquare(f_ob2, f_exp=f_exp))
```

주사위 값: 1	주사위 값: 2	주사위 값: 3	주사위 값: 4	주사위 값: 5	주사위 값: 6
99번 발생	101번 발생	102번 발생	97번 발생	101번 발생	100번 발생
100번 예상	100번 예상	100번 예상	100번 예상	100번 예상	100번 예상

```
1: Power_divergenceResult(statistic=0.16, pvalue=0.9994854883416188)
```

H_0를 기각할 수 없음

주사위 값: 1	주사위 값: 2	주사위 값: 3	주사위 값: 4	주사위 값: 5	주사위 값: 6
117번 발생	119번 발생	120번 발생	115번 발생	119번 발생	10번 발생
100번 예상	100번 예상	100번 예상	100번 예상	100번 예상	100번 예상

```
2: Power_divergenceResult(statistic=97.36, pvalue=1.9021569776120922e-19)
```

H_0 기각 후 H_1 채택함.