Error patterns in determining combined probability functions from continuous random variables

F Mulyatna¹ and W Nofiansyah²
¹Universitas Indraprasta PGRI, Jl. Nangka No. 58C, Jakarta Selatan, DKI Jakarta 12530, Indonesia
²STKIP Kumala Lampung, Jl. Yossudarso Gang Karet No. 15 Polos, Metro Pusat, Kota Metro, Lampung 34111, Indonesia

E-mail: fauzi.mulyatna@unindra.ac.id

Abstract. The purpose of this research is to analyze the patterns of error in determining the combined probability functions of continuous random variables. The type of this research is qualitative research. The data, which is obtained in this research, is then analyzed descriptively. Subjects selection are based on purposive sampling technique by taking 9 subjects with consideration/criteria for high, medium and low academic abilities. The method used to collect data in this research is a written test in the form of a description. The triangulation used in this research is time triangulation that collects data from 6th semester students in the academic year of 2015/2016, 2016/2017 and 2017/2018 who took mathematics statistics course. Conclusion indicates that the pattern of errors experienced by the subjects is concept errors, errors related to the concept of pre-requisite material and calculation errors.

1. Introduction
Mathematics learned in school is mathematics whose materials are chosen in such a way in order to make it easier to use for students’ daily life [1-3]. Due to its important role, mathematics must be studied and understood properly and correctly [4] so that it can be the key for the success of students and will give them a brilliant career. Success in learning mathematics is the estuary of all activities carried out by students in the learning process of mathematics. The implementation of its learning is the key in determining learning outcomes. Studies have shown that the learning process in the classroom contributes greatly to the success of delivery and the concept-formation process of university students [8].

Learning process in university class is not different from secondary schools stage because it also gives large contribution towards the success of the delivery and the concept establishment process of university students [8]. Different from school students, university students are demanded to be more independent in study so that Students-Centred Learning (SCL) is a must for them. Based on Yager research (1991) [9], learning is an active process and requires students to construct their own personal schema to assimilate new concepts. Learning conditions in school, in which most of its material is a contextual (problems of students’ daily life), became a differentiator when students has already been accepted in university. Mathematics education course, especially the exact courses, contains more concepts that are completely abstract and they cannot be associated with problems in everyday life. Even in science lessons, it still needs a model and it can make students confuse. According to Treagust, et, al [10], “teachers use models as aids to help explain scientific phenomena and students often make their own models of scientific phenomena to display their understanding.”
One of the subjects included in the curriculum of mathematics education is mathematical statistics. The researchers focus on taking the concept to determine the combined probability functions of continuous random variables. The reason is that because the concept is the basic that must be understood completely as foundation for further concepts [11].

Actually the concept to determine the combined probability functions of continuous random variables, which is the focus of this study, is still simple. The given function already exists and it just completes its coefficient. In fact, there are many students who are still experiencing difficulties.

In the era of the industrial revolution 4.0, people must follow the changes that are very quick in order not to miss anything. The integration of the science concept in digital systems, even further into the robotic system, is highly emphasized. Based on that fact, it can be said that there are problems of students’ knowledge gaps when problems are simple and conventional.

The discovery of analyzing the errors pattern in determining the combined probability functions of continuous random variables can be an easier way in next courses strategy. In fact, the pattern of same mistakes often appears in each academic year and it is very clear that it is done by different class of students.

The process to discover the errors pattern is a step to map out and to categorize the errors probability made by the students. This is useful for learning in the next courses. The emphasis of the concept in the materials tends to emerge errors that need to be fixed in further courses.

In the first year, which was the academic year of 2015/2016, researchers found the fact that there is no one who got perfect score of mathematical statistic course in the midterm. Therefore, the focus of this study is to determine the combined probability functions of continuous random variables. Different student made error in determining the combined probability functions of continuous random variables in this problem because there were no students who got maximal point (perfect score). Based on that fact, the researchers want to analyze the patterns of error in determining the combined probability functions of continuous random variables.

2. Research Methods
This Research is implemented in Campus B University Indraprasta PGRI Jakarta which is located at 80 Raya Tengah Street, Gedong Village, Pasar Rebo District, East Jakarta 13760. The type of this research is qualitative descriptive. Data of this research is analyzed descriptively. This refers to the Groth [12] "The rich descriptions provided by qualitative research are often instrumental in helping readers understand the relationship between the context of a given study and their own circumstances."

Qualitative research is applied in this research because the researchers want to collect data which is focused on one study and it will be studied more profound. Sugiyono [13] states that qualitative method is applied in order to get more comprehensive data which is meaningful. The data in this research is error patterns done by subject in determining the combined probability functions of continuous distribution.

The subject of this research is students of mathematics education. The selection of the subject is based on purposive sampling technique. Purposive sampling is a technique collect data by using considerations/certain criteria of subjects. The subjects are 6th grade students of mathematics education program who take Mathematical Statistics in each academic year. This is because the data is taken from the result analysis of the Midterm test done by the subject in even semester of 2015/2016, 2016/2017 and 2017/2018. In addition, the criteria for selecting subject are based on high, medium and low academic ranking. The subjects are 3 students for each academic year so that there are 9 subjects in this research.

The main data of this research is result analysis of subjects’ written test. Instrument in this research consists of researchers as main instrument and descriptive test as auxiliary instruments. The tests in this research is written test in the form of essay. Instrument test has been validated by the coordinator lecturer. Coordinator lecturer is a lecturer who has advantages in scientific capability. According to Buchori and Setyawati [14], the presence of researchers in the field as the main instrument
are as data collectors, planners, implementers, action giver, data analysts, results reporters and caretakers for the entire process and results.

The technique to examine data validity (validity) is done after collecting data from the subject of the research. Data in this research is results analysis of subjects written test subject which is done in different times; they were in the academic year of 2015/2016, 2016/2017 and 2017/2018 which took 3 subjects in each academic year. Therefore, time triangulation is used in this research.

Data coding given to a subject is L1, M1 and H1 for students who are in 6th grade of the academic year of 2015/2016 with low, medium and high academic ability. Moreover, L2, H2 and M2 are used for students in 6th grade of the academic year of 2016/2017 with low, medium and high academic ability. The last is L3, M3 and H3 for students who are in 6th grade of the academic year of 2017/2018 with low, medium and high academic ability.

Data validity will show that data presented by researchers is appropriate with actual condition which is error patterns done by subjects in determining the combined probability functions of continuous distribution. Technique of data analysis is done by using data reduction, data display and verification and also conclusion taker.

3. Results and Discussions
3.1. The Results Answer of Subjects’ Written Test
The issue raised in this study is to determine the combined probability functions of continuous random variables by analyzing probability of error patterns conducted by students. Determiner of the combined probability functions is only limited to its determination of coefficient value. The function model is as follows.

\[ f(x, y) = cx^2y, a < y < x < b \]

First year, the data is collected in the academic year of 2015/2016. After collecting data from each subject, the results obtained are represented in Figure 1, Figure 2 and Figure 3 with the given question: Given combined density function is as follows: \( f(x, y) = cx^2y, 0 < y < x < 2 \). Determine the value of \( c \)!

In the Figure 1 below, it shows that the error, which occurs, is concept error. The use of operations \( \Sigma \) (sigma) is a settlement in combined probability functions of discrete random variables.

![Figure 1. The Answer Result of Subject L1](image)

Figure 2 indicates that error, which is experienced, is an error related to integral boundary determination so that this error is correspond with the concept of material pre requisites on the previous semester courses which is integral calculus course. Moreover, extracting the limit of 0 until 2 for \( x \) and 0 until 2 for \( y \) in the integral \( \int_0^2 \int_0^2 cx^2 y \, dx \, dy \) is not accurate.
Figure 2. The Answer Result of Subject M1

The correct answer is represented in Figure 3.

Figure 3. The Answer Result of Subject H1

In the academic year of 2016/2017 academic year, the same question pattern and quality were given. The density function is given as follows:

\[ f(x, y) = \begin{cases} kx^2y & ; 0 < y < x < 5 \\ 0 & ; \text{otherwise} \end{cases} \]

Determine the value of \( k \)!
The answer of each subject is represented in Figure 4, 5 and 6. In Figure 4., the error, which is experienced, is concept error. The use of operations $\Sigma$ (sigma) is a settlement in combined probability functions of discrete random variables.

Figure 4. The Answer Result of Subject L2

Figure 5. shows that error, which is experienced, is an error related to integral boundary determination so that this error is correspond with the concept of material pre requisites on the previous semester courses which is integral calculus course. Moreover, extracting the limit of 0 until 5 for $x$ and 0 until 5 for $y$ in the integral $\int_0^5 \int_0^5 kx^2y \, dx \, dy$ is not accurate.

Figure 5. The Answer Result of Subject M2

Figures 6. Shows that the error is an error of calculating operation, $\frac{1}{3} \cdot 5^3$ is $\frac{15}{3}$

Figure 6. The Answer Result of Subject H2

In the academic year of 2017/2018 academic year, the problem given used different symbols as its variables, namely $x_1$ and $x_2$.

The problem given is as follows:

Given combined probability function is:

$$f(x_1, x_2) = \begin{cases} cx_1^2x_2 & ; 0 < x_1 < x_2 < 1 \\ 0 & ; \text{otherwise} \end{cases}$$

Determine the value of $c$ !
The first answer is shown in Figure 7 which indicates that error experienced is error related to integral boundary determination so that this error is correspond with the concept of material pre requisites on the previous semester courses which is integral calculus course. Moreover, extracting the limit of 0 until 1 for $x$ and 0 until 1 for $y$ in the integral $\int_0^1 \int_0^1 c x^2 \ y \ dx \ dy$ is not accurate. The use of variables $x$ and $y$ as random variables is also inappropriate, since the problems that prompted is $x_1$ and $x_2$. This is also a finding that the symbols is only sensed as calculus procedure, is not included as a concept. Therefore, the use of symbol only depends on the previous naming habits [15].

Likewise in Figure 8, the error experienced is error related to integral boundary determination so that this error is correspond with the concept of material pre requisites on the previous semester courses which is integral calculus course. Moreover, extracting the limit of 0 until 1 for $x$ and 0 until 1 for $y$ in the integral $\int_0^1 \int_0^1 c x_1^2 \ x_2 \ dx_1 \ dx_2$ is not accurate.
3.2. Analysis Based on Subjects’ Written Answers

The results of subjects’ descriptive test are then analyzed. The analysis is conducted to know the error patterns committed by subjects. Seen from the academic ability of each subject, is there any pattern of the same error?

In the academic year of 2015/2016, error found on the subject L1 is the fault concept. It is clear that the combined probability function of continuous random variables (the concept attributes to probability distribution) definitely uses integral. The integral boundary interval form of random variables is different and it is believed to be the reason why subject considers the given function as the combined probability function of discrete random variables. The general form of the combined probability function of continuous random variables is actually \( f(x, y) \) with \( a < y < b \) and \( a < x < b \). Therefore, a different interval form \( a < y < x < b \) or \( a < x < y < b \) causes the interpretation changing into combined probability function of continuous random variables. It is same as in the given problem, with its interval shape is \( 0 < y < x < 2 \).

Subject M1 made a mistake in taking integral limit. This shows that the concept of pre requisites material, which is integral calculus, is not yet completely mastered. The integral limit with interval forms of random variables \( a < y < b \) and \( a < x < b \) can easily be determined by the form of its integral, which is \( \int_a^b \int_a^b f(x, y) \, dx \, dy \). Different from integral limit with interval \( 0 < y < x < 2 \), this interval requires an analysis within the limits of its integral. The interpretation form of its integral limit becomes \( 0 < y < 2 \) and \( y < x < 2 \) or it can be with integral limit \( 0 < x < 2 \) and \( 0 < y < x \). In this case, subject of M1 hasn’t been able to do the analysis correctly because subject gave the integral limit \( 0 < x < 2 \) and \( 0 < y < 2 \).

The correct answer was obtained by subject H1. Subject H1 managed to find combined probability function of continuous random variables with the exact concept.

In the academic year of 2016/2017, the error discovered was from subject L2 which was concept error. The concept of the use of sigma (\( \Sigma \)) is clear and that is a combined probability function of discrete random variables. This case was also found on the academic year of 2015/2016, the concept attributed to the distribution of continuous probability was certain in using integral. Subject L2 hasn’t been able to characterize or distinguish between the distribution of combined probability of discrete random variables as well as combined probability of continuous random variables. It was same as subject L1, the interval form of integral limit of different random variables was believed to be the cause why subject considered given function as combined probability function of discrete random variables. The general form of combined probability function of continuous random variables is actually \( f(x, y) \) with \( a < y < b \) and \( a < x < b \) so that, with a different interval form \( a < y < x < b \) or \( a < x < y < b \), it changes...
the interpretation into combined probability function of discrete random variables. It is like as the given problem, the form of its interval is \(0 < y < x < 5\). Therefore, the result is absolutely wrong in determining the value of \(k\).

Error in taking integral limit was also experienced by Subject M2. In this case, the subject hasn’t been able to perform analysis correctly, by placing integral limit \(0 < x < 5\) and \(0 < y < 5\). This shows that he concept of pre requisites material, which is integral calculus, is not yet completely mastered by Subject M2 because Subject M2 takes integral limit \(\int_0^5 \int_0^5 kx^2 \,dy \,dx\) which is incorrect. Based on interval of continuous random variables \(0 < y < x < 5\), the correct integration is \(\int_0^5 \int_y^5 kx^2 \, dy \, dx\) or \(\int_0^5 \int_0^x kx^2 \, dy \, dx\) so that, in order to obtain value of \(k\), it should be associated with the distribution nature of combined probability function of continuous random variables which is \(\int_0^5 \int_0^y kx^2 \, dy \, dx = 1\) or \(\int_0^5 \int_0^x kx^2 \, dy \, dx = 1\).

Subject H2 has had the right concepts in solving the problem which were distribution concept of combined probability function of continuous random variables as well as concept of integral calculus. However, Subject H2 made calculation error in which Subject H2 calculated the calculating operation of \(\frac{15}{3}\) to be \(\frac{15}{3}\). This was due to Subject H2’s inadvertence.

In the academic year of 2017/2018, it was found on Subject L3, M3 and H3 papers that those subject experienced errors related to integral limit determination and its integrating process. Those errors were associated to pre requisite material concept in integral calculus course of previous semester. The case experienced by Subject L3 was that the taking of integral limit \(\int_0^1 \int_0^1 \) was not right. The given interval was \(0 < x_1 < x_2 < 1\) so that the correct interpretation related to integral limit should be \(0 < x_1 < 1\) and \(x_1 < x_2 < 1\) or it was permitted with integral limit on interval \(0 < x_2 < 1\) and \(0 < x_1 < x_2\). The use of variables \(x\) and \(y\) as random variables was also inappropriate since the prompted problems was for random variables \(x_1\) and \(x_2\).

The mistake was also experienced by Subject M3 which takes incorrect integral limit \(\int_0^1 \int_0^1 cx_1^2 x_2 \, dx_1 \, dx_2\). As described earlier, the given interval was \(0 < x_1 < x_2 < 1\) so that the correct interpretation related to integral limit should \(0 < x_1 < 1\) and \(x_1 < x_2 < 1\) or it was permitted with integral limit on interval \(0 < x_2 < 1\) and \(0 < x_1 < x_2\). Therefore, the integral form after being associated with the distribution nature of combined probability becomes \(\int_0^1 \int_0^x cx_1^2 x_2 \, dx_1 \, dx_2 = 1\) or \(\int_0^1 \int_0^x x_1^2 x_2 \, dx_2 \, dx_1 = 1\).

Subject H3 successfully interpreted the integral limit. The integral limit \(\int_0^1 \int_0^1 \) is the interpretation result of interval \(0 < x_1 < 1\) and \(x_1 < x_2 < 1\). However, Subject H3 incorrectly took \(dx \, dy\) because the requested variables were \(x_1\) and \(x_2\) as their continuous random variables. If meant variable \(dx \, dy\) was \(dx_1 \, dx_2\), then it will be an error related to the interpretation of limit integral towards the order of its integrating process. Therefore, the correct one should be \(\int_0^1 \int_0^1 f(x_1, x_2) \, dx_2 \, dx_1\).

### 3.3 Discussion

Based on the analysis of subjects’ written answers, error patterns, which were generally found, were concept errors, errors related to the concept of pre-requisite material and calculation errors. The errors related to pre requisite material are the concepts in integral calculus course which are not yet fully understood. The errors related to integrating process are the errors in determining double integral limit as well as the order of variable calculation which should be done first in integrating process itself.

This study only takes subjects with low academic ability that experience errors related to the differentiation of combined probability function of continuous random variables as well as discrete random variables. These concept errors are experienced by Subject L1 and L2. Meanwhile, Subject L3
can identify that the given problem is combined probability function of continuous random variables. However, Subject L3 has not been able to adjust the use of variable with different letter. This is a habit in learning by using random variables \(x\) and \(y\). Although the given problem is \(x_1\) and \(x_2\), Subject L3 still uses random variables \(x\) and \(y\) in his/her work[15].

All of the subjects with medium academic ability, M1, M2 and M3, make mistake related to pre-requisite material which is included in integral calculus course. The concept of double integrals limit determination by interpreting given interval is not yet mastered. This concept becomes important, because it can be determined the prompted coefficient value based on the distribution nature of combined probability through a process of double integrals with probability value of 1.

Subjects with high academic ability provide various results. Subject H1 managed to understand the concept well. It is shown by his/her correct answer. Subject H2 subject is also understand the concept as good as Subject H1. Subject H2 only makes calculating mistake which causes its last result becomes incorrect.

Based on the results of this research, there are still much efforts, which should be improved, in order to pursuit industrial revolution 4.0. University students are still having trouble in doing question of LOTS (Low Order Thinking Skills) and are having much trouble to integrate mathematics concepts with technology.

In fact, in other studies, they shows that students as prospective teachers still have difficulties in very basic calculating operations. The result of Guler and Celik research [16] shows that “teacher candidates had difficulties in conceptualizing even basic mathematical statements such as \(2^0\) and generating explanations about teaching it”.

In the era of industrial revolution 4.0., the role of human beings will be replaced by robotic systems. Therefore, the learning process should be oriented to programming systems. Before taling further about that, students’ independent in shaping new knowledge is a must. Ntemngwa and Oliver [17] states that “In constructing the robot, the construction kit and the instructional strategy the teachers used provided the student a favorable learning environment where they could construct their own knowledge”.

In today's education system, it is emphasized that teachers should be able to integrate technology to teaching. This development and spread of technology in daily life has necessitated diversity and innovations in learning and teaching methods. For this reason, in today’s education, teachers have to improve themselves well in the use of technological tools in order to obtain the necessary competences needed in their profession. The studies of recent years generally emphasize teacher's technological pedagogical content knowledge. It is expected that today’s teachers should be well informed about curriculum, how the curriculum will be taught and relation of the field with other fields, latest developments in the field, basic concepts, means and structures of the field and integration of the content that will be taught with technology. This paragraph refers to Özdemir [18].

After finding error patterns from each subject, then learning strategies need to be considered because each subject with a different academic levels requires different treatment. According to Yerison, et. al. research [19] states that “A constraint faced by teachers in teaching mathematics is the difficulty in delivering learning materials. Because the ability of learners here varies widely, students with high abilities often feel bored because certain material is often repeated. We recommend that teachers provide additional material or provide challenging questions. They will be greatly helped to develop their competence. For almost all time, teachers focus on the completion of the course. For students with low ability, teachers have to repeat several times so that learners better understand”.

This study illustrates that, in order to achieve optimal learning target, it must be accompanied with a strong concept. It will be very difficult to integrate in programming systems as well as in other technologies, if students’ basic concepts of mathematics is weak. Perhaps in the next research, it can be considered a strategy for using of robotic systems in mathematical statistics teaching as it is stated in Barker and Ansorge [20], “research support the use of 4-h robotics to teach SET concepts in an after school program and that the evaluation instrument developed to test the SET concepts is reliable and valid”. 
Review the analysis of written answers of each subject, error patterns generally found are errors related to pre-requisite material. In fact, students have basic knowledge and it must be confirmed that the basic knowledge is the right concept. When concepts are introduced in school, they are not transmitted to students, but students will attempt to fit them to the models or concepts they currently have [21].

4. Conclusion
The error patterns in determining the combined probability function of continuous random variables are concept errors, errors related to the concept of pre-requisite material and calculation errors. In terms of academic ability, concept errors are experienced by a subject with low ability while a subject with medium ability experiences difficulties in understanding pre condition material. Last, a subject with high academic has succeeded to answer correctly. Errors experienced by a subject with high academic emerge due to calculation errors and incorrect integrating process.

References
[1] Guire P M and Kinzie M B 2013 Analysis of place value instruction and development in pre-kindergarten mathematics Early Childhood Education Journal 41 355
[2] Turnuklu E B and Yesildere S 2007 The pedagogical content knowledge in mathematics: Preservice primary mathematics teachers’ perspectives in Turkey IUMPTST: The Journal 1 1
[3] Tan D A, Orongan R C and Guayan D A 2015 Coherence of pre-service teachers’ conceptions of mathematics and its teaching CMU Journal of Science 19 16
[4] d’Ambrosio U 1985 Ethnomathematics and its place in the history and pedagogy of mathematics For the Learning of Mathematics 5 44
[5] Novak J D 2002 Meaningful learning: The essential factor for conceptual change in limited or appropriate propositional hierarchies (liphs) leading to empowerment of learners Science Education 86 548
[6] Chen Q and Leung F K S 2013 A comparative study of teachers’ mathematics beliefs in the context of curriculum reform in Hong Kong and Chongqing 15 46
[7] Ünlü M 2017 Investigating mathematics teachers candidates’ knowledge about problem solving strategies through problem posing Journal of Education and Practice 8 218
[8] Xenofontos C and Andrews P 2014 Defining mathematical problems and problem solving: Prospective primary teachers’ beliefs in Cyprus and England Mathematics Education Research Journal 26 279
[9] Yager R E 1991 The constructivist learning model: towards real reform in science education The Science Teacher 58 52
[10] Treagust D F, Chittleborough G and Mamiala T L 2002 Students’ understanding of the role of scientific models in learning science Int. J. Sci. Educ. 24 357
[11] Remillard J T 2005 Examining Key Concepts in Research on Teachers’ Use of Mathematics Curricula Review of Educational Research 75 211
[12] Groth R E 2010 Situating qualitative modes of inquiry within the discipline of statistics education research Statistics Education Research Journal 9 7
[13] Sugiyono 2010 Metode Penelitian Pendidikan: Pendekatan kuantitatif kualitatif dan R&D (Bandung: Alfabeta)
[14] Buchori A and Setyawati R D 2015 Development learning model of character education through e-comic in elementary school International Journal of Education and Research 3 369
[15] Mulyatna F and Kusumaningtyas W 2017 Simbolisasi dalam Metode Numerik sebagai Representasi Konsep dan Prosedur NUMERICAL: Jurnal Matematika dan Pendidikan Matematika 1 147
[16] Guler M and Celik D 2016 A research on future mathematics teachers’ instructional explanations: the case of algebra Educational Research and Reviews 11 1500
[17] Ntemngwa C and Oliver J S 2018 The implementation of integrated science technology, engineering and mathematics (STEM) instruction using robotics in the middle school science classroom *International Journal of Education in Mathematics, Science and Technology (IJEEMST)* 6 12

[18] Özdemir E 2018 Investigation of prospective Math teachers’ perceptions about the use of technology in mathematics teaching *Educational Research and Reviews* 13 674

[19] Yerizon , Putra A A and Subhan M 2018 Mathematics learning instructional development based on discovery learning for students with intrapersonal and interpersonal intelligence (preliminary research stage) *International Electronic Journal of Mathematics Education* 13 97

[20] Barker B S and Ansorge J 2007 Robotics as Means to Increase Achievement Scores in an Informal Learning Environment *Journal of Research on Technology in Education* 39 229

[21] Mumu J, Prahmana R C I and Tanujaya B 2017 Construction and reconstruction concept in mathematics instruction *Journal of Physics: Conf. Series* 943 4