



Using Card Games for Conditional Probability, Explaining Gamma vs. Poisson Distributions, and Weighing Central Limit Theory

Dr. Roes Arief Budiman P.Eng., University of Calgary

Received PhD in Materials Science and Engineering at University of Toronto in 2001. Currently a Senior Instructor at University of Calgary and have been teaching Probability & Statistics for Engineers course in the past three years. Maintain a small research group (1 PhD, 1 MEng) on pipeline failure and reliability.

USING CARD GAMES FOR CONDITIONAL PROBABILITY, EXPLAINING GAMMA VS. POISSON DISTRIBUTIONS, AND WEIGHING CENTRAL LIMIT THEORY

Abstract

Students taking probability course for the first time are often struggling with conditional probability. To help explain the concept better, card games are used to explain especially differences of conditional probabilities of sequential and simultaneous card draws. Several card game experiments are discussed and typical probability results are shown and compared with predictions. These simple experiments can be demonstrated in classroom and students can use them to test the predictions. Another concept students often struggle is distinguishing gamma from Poisson distribution. An identity connecting them, generalizing the connection between exponential and Poisson distribution, will be used to discuss their differences and to point out nuances in the wording of some probability problems that yield different answers when both distributions are used. Lastly, a teaching tool for explaining central limit theorem is discussed based on guessing weights of books. This guessing game proves useful to explain sampling distribution.

1. Introduction

Although it has been argued since 1960s [1] that probability and statistics is as important as calculus as a mathematical foundation for engineering students who have to cope with uncertainty and variability in their professional careers, majority of engineering programs in North America have one course for both probability and statistics. In this one course, engineering students must learn combinatorics, conditional probability, discrete and continuous probability density functions and their parameter estimators, statistical inferences from confidence intervals to hypothesis tests, linear regression method, and possibly even joint probability density functions [2]. Hogg [3] proposed a list of basic statistical concepts, which emphasizes experimental designs and quality control, for a second-year engineering course. One main problem for implementing such statistics course is most second-year students have not taken their program-specific courses, such as machine component design course in mechanical engineering, to appreciate the interplay between statistics (e.g., safety factor for mechanical engineers), performance (e.g., fatigue strength and tensile strength), and cost (e.g., machining and material choice).

With these stringent constraints, a most practical way for an instructor of the course is to teach a topic effectively and efficiently using a real-world analog such as card game. This paper aims to discuss three effective teaching tools that can be adopted in the course to teach concepts of conditional probability, (continuous) gamma vs. (discrete) Poisson distribution, and central limit theorem that forms a basis for statistical inferences.

ENGG 319 Probability and Statistics for Engineers course in Schulich School of Engineering is offered to second-year engineering students who have taken two calculus courses, covering differential calculus to introductory vector calculus, and linear algebra course in their first year. After learning complicated calculus and linear transformations,

students are initially pleased with the simplicity of probability definition, i.e., the ratio between the number of ways for the event we're interested in and the number of ways for all possible events. But they quickly realize that counting the number of ways for either numerator or denominator of the probability definition is quite often more daunting than solving calculus problems because the counting problems do not have clear prescribed methods to follow. Counting them is similar to solving linear transformation problems; they require visualization, but one can argue that the former can be more challenging because different counting problems require different visualization strategies.

The basic counting formula is given by

$$\frac{n!}{n_1! n_2! \cdots n_k!},$$

where $n = n_1 + \cdots + n_k$, which counts the number of ways for dividing n labelled objects into k groups, each with n_1, n_2, \cdots, n_k members. If n objects are divided into n groups, each with one member, then there are

$$\frac{n!}{1! 1! \cdots 1!} = n!$$

ways. The connection to the $n!$ ways confirms the need to label each n objects. If there is only one group with n members, then there is

$$\frac{n!}{n! 0!} = 1$$

way to do it, which shows that the group label is different from the object label. Students use the basic formula in different counting problems by adapting its application to different problem types.

One problem type requires multiplying same basic formula several times with some constraint. One example is "4 cards are drawn from a deck of 52 cards. How many ways are there to get 2 kings and 1 queen?" The answer is

$$\binom{4}{2} \binom{4}{1} \binom{44}{1} = 1056,$$

where $\binom{4}{2}$ is a shorthand notation for $4!/(2! 2!)$, which counts the number of ways of getting 2 kings from 4 available. The constraint is given by $52 = 4 + 4 + 44$. In this example students also learn to incorporate *statistical independence* concept which is reflected by multiplying the three formulas together to arrive at the answer. The concept allows them to understand that picking 2 kings out of 4 is independent of picking 1 queen from the available 4.

Another problem type requires students to remove several objects before counting is performed. One example is the following: "There are 4 red balls and 3 blue balls in a bin. You select 7 balls randomly from the bin without replacement. How many ways are there so that the first and the fifth are of the same color?" The first and the fifth can be either blue or red, which implies that these two cases are *mutually exclusive*. Assigning the first and the fifth to red balls will give

$$\frac{(2 + 3)!}{2! 3!} = 10$$

for the remaining 5 sites, while assigning the first and the fifth to blue balls will give

$$\frac{(1 + 4)!}{1! 4!} = 5$$

so that there are total 15 ways.

Probability and statistics for engineers textbooks [2] cover the basic counting formula and give many exercise problems. The presentation of combinatorics in such textbooks is not yet systematic to allow students to learn it methodically. Probability and statistics for engineers course typically spends maximum 1 week on combinatorics at the beginning of the course. It is not an easy topic to teach and is at an inconveniently early time in a course, when most students are grappling with the concept of probability. Combinatorics teaches students the ability to spot "hidden patterns" in numbers and graphs when working on other problems, such as those they later find in thermodynamics and mechanics. One way to alleviate this minimal exposure and workout is to lengthen the combinatorics coverage within the conditional probability and to show to students the close connections between combinatorics and other more applied topics in statistics.

2. Conditional Probability Using Card Games

Conditional probability formula

$$P(A|B) = \frac{P(AB)}{P(B)},$$

where event B becomes a condition for event A , is a difficult formula for students to understand. The first hurdle for students is understanding how the condition B affects the counting of events. This hurdle can be illustrated by two counting problems. The first problem is as follows: "5 cards are dealt randomly from a deck of 52. Given that the first 4 are heart, how many ways are there for the fifth is also heart?" If event A corresponds to the fifth heart, then event B corresponds to the first 4 hearts, so that there are

$$\binom{13}{4 + 1} = 1287$$

ways to have 5 hearts. The probability that the fifth card is a heart given that the first four are heart, under the condition that 5 cards are drawn simultaneously, is therefore

$$P(A|B) = \frac{\binom{13}{5} \binom{39}{0} / \binom{52}{5}}{\left[\binom{13}{5} \binom{39}{0} + \binom{13}{4} \binom{39}{1} \right] / \binom{52}{5}} = \frac{\binom{13}{5}}{\binom{13}{5} + 39 \binom{13}{4}} = \frac{3}{68}.$$

The numerator $\binom{13}{5}$ is the number of ways for the event we want (i.e., five hearts), while the denominator $\binom{13}{5} \binom{39}{0} + 3 \binom{13}{4} \binom{39}{1}$ captures the total number of ways that can happen when 5 cards are taken simultaneously given that 4 are hearts.

The second problem is related to the first and is as follows: "4 cards are drawn from a deck of 52. They are opened and turn out to be hearts. How many ways are there for a fifth to be heart if one more card is drawn?" The number of ways for the fifth to be a heart is

$$\binom{9}{1} \binom{39}{0} = 9,$$

so that under the condition that 4 cards drawn first are heart, we obtain different conditional probability

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\binom{13}{5} \binom{39}{0} / \binom{52}{5}}{\binom{13}{4} \binom{39}{0} / \binom{52}{4}} = \frac{\binom{4! 9!}{5! 8!} \binom{5! 47!}{4! 48!}}{\binom{4! 9!}{5! 8!} \binom{5! 47!}{4! 48!}} = \frac{9}{48}.$$

An astute student might recognize that there are 9 heart cards left when 4 hearts are drawn, which represent the numerator, and that there are 48 cards left when 4 are drawn, representing the denominator. The connection between 9/48 and the basic counting formula will be lost if the answer 9/48 is given directly.

The conditions, whether 5 or 4 cards are drawn, yield different probabilities. Most students will readily accept the 9/48 probability but not the 3/68 probability. To show their differences, one might use an analogy based on the answers above. If 5 cards are drawn simultaneously, the number of ways to have them is 1287, while for the simultaneously 4 hearts drawn *followed by* a fifth heart

$$\binom{13}{4} \binom{9}{1} = 6435.$$

The increased number of ways when the cards are drawn *sequentially* thus results in a higher probability. "How could the number increase?", students might ask. It has to do with the *ordering* of the cards drawn. When 5 cards are drawn simultaneously, and they are then opened to reveal their ranks, the information on the rank order of appearance is

erased. Sequential draw, however, is more preserved for the 4 simultaneous draw. For instance, the sequential draw distinguishes between having a queen heart in the first 4 and having a queen heart in the fifth. So, whichever the heart rank to appear in the first 4 and the fifth, there is a factor of

$$\frac{5!}{1!4!} = 5,$$

which is equal to 6435/1287, between the two conditions.

Ultimately, their probability difference must be settled by card experiments. Randomly obtaining 4 hearts in a simultaneous draw of 5 cards would be exceedingly rare, so this card experiment is ideal for a class assignment. A group of 3 students can run one card experiment and having 30 such groups for a 90-student class will increase the number of times the event will occur. Such card experiments, when done first before theoretical prediction is revealed, will remove instructor's reluctance to explore more subtle counting problems out of embarrassment of incorrectly predicting an answer. There are several good combinatorics books [4], which can be used as references, but a course instructor of probability and statistics for engineers course has to be willing to go beyond textbook problems with ready solution manuals in order to make the course practical.

In one card experiment I picked 2 cards simultaneously and opened the 2 cards. This process is repeated 100 times and I looked for having 2 clubs given that the first is a club. During this experiment, I was able to see a possibility that the second card is not a club and the first one is; there were 27 times such event occurred, and within the 27 there were 4 that resulted in two clubs. The 27/4 ratio is approaching the theoretical prediction

$$\frac{\binom{13}{2} + \binom{13}{1}\binom{39}{1}}{\binom{13}{2}} = \frac{15}{2}.$$

This ratio could be compared with sequential draws of 2 cards, increasing the number of card experiments to be performed, which is ideal if done as a class experiment.

This card game can be expanded to include n simultaneous draws and to open m cards, where m doesn't have to be equal to $n - 1$. Theoretical predictions and card experiments from the simultaneous draws can then be compared against n sequential draws' predictions and card experiments.

Such card game and other variants involving sports and other games [5] provide real-world applications of conditional probability concept. Students can also easily compare theoretical predictions using combinatorics against experiments using a deck of cards. The game approach to explaining conditional probability also provides practice for students to improve their combinatoric (counting) skills. This game approach is more practical than another approach using Bayes' box [6] to explain conditional probability.

3. Poisson vs. Gamma Distribution

Poisson probability mass function

$$f(r) = \frac{\lambda^r e^{-\lambda}}{r!}$$

describes the probability of having r number of events within a *fixed* time interval whose average number of events is λ . Poisson distribution can be used to model naturally occurring, periodic processes. Gamma probability density function

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)},$$

where $\Gamma(r) = (r - 1)!$ is gamma function with an integer argument, treats the integral number r of events as a parameter unlike in Poisson. The random variable for gamma distribution is x , which is either distance or time duration. The two functions look similar and students often confuse their respective underlying assumptions and thus their applicability.

One problem that illustrates their overlap is the following. "Calls to a call center have a mean of 6 calls per hour. What's the probability that 4 calls occur within one hour?" According to Poisson distribution, the answer is

$$\frac{6^4 e^{-6}}{4!} = 0.134,$$

while gamma distribution gives

$$\int_0^1 \frac{6^4 x^3 e^{-6x}}{3!} dx = 0.849.$$

The two numbers occupy two probability extremes and thus don't represent consistent approximations. Course textbooks often casually use Poisson distribution for problems that require gamma distribution. Resolution to this problem is often stated as an exercise problem that requires rather involved mathematical treatment for second year students. I'd like to bring this connection to our attention to prevent such confusion.

Starting with Poisson distribution

$$P(r < N) = \sum_{r=0}^{N-1} \frac{\lambda^r e^{-\lambda}}{r!},$$

we have

$$P(r \geq N) = 1 - \sum_{r=0}^{N-1} \frac{(\lambda x)^r e^{-\lambda x}}{r!}$$

by rescaling the parameter $\lambda \rightarrow \lambda x$ so that we can insert a continuous variable x based on a constant λ . Because $P(r \geq N)$ is a (cumulative) distribution function, taking a derivative with respect to x gives a gamma probability density function

$$f(x) = \frac{d}{dx} P(r \geq N) = \sum_{r=0}^{N-1} \frac{(\lambda x - r)(\lambda x)^r e^{-\lambda x}}{r! \tau} = \frac{\lambda^N x^{N-1} e^{-\lambda x}}{\Gamma(N)}.$$

This implies by integrating the derivative that

$$\int_0^{\tau} \frac{\lambda^N x^{N-1} e^{-\lambda x}}{\Gamma(N)} dx = \sum_{r=N}^{\infty} \frac{(\lambda \tau)^r e^{-\lambda \tau}}{r!}.$$

This identity clarifies the relationship between Poisson and gamma distribution. Probability from gamma distribution for a specified number N of events is equal to having at least N when the probability is assigned to a Poisson process. Gamma distribution assumes that the N th event occurs exactly at the end of the time duration $0 \leq x \leq \tau$; however, Poisson process allows for these N events to occur anywhere within the interval. It is therefore possible that in Poisson process there are more than N events within the duration $0 \leq x \leq \tau$.

The 0.134 probability for the 4 calls in one hour obtained by Poisson distribution means that it is very unlikely that there are only 4 phone calls. The 0.849 probability from gamma distribution confirms that it is more likely that there are more than 4 phone calls when the phone calls are thought of as a Poisson process. However, the question does specify that there are *only* 4 phone calls, instead of *at least* phone calls. This implies that the fourth phone call should occur exactly at the end of one-hour period. The 0.849 probability answer should prevail as a result.

4. Central Limit Theorem

One class exercise that helps students understand central limit theorem is to ask them one by one to weigh with their hands an unknown book by comparing it with another book of known weight. Students are asked not to tell or discuss their weight guess with their class mates. Their decisions are then polled by a show of hands. This active learning activity might help students remember the tenet of central limit theorem longer than showing them its proof [7].

The weight comparison task mimics a random process with a well-defined average and a finite variance required by central limit theorem. The presence of a well-defined average is guaranteed when each student is asked to estimate the weight since most of them will

seriously compare their weights. There was only one student in a class of 70+ students, who arbitrarily picked a number for the weight. The weight comparison task also removes a possibility of an infinite variance which can occur if each student is simply asked to pick a random number. In addition, weighing an item still cannot be performed by a smartphone or a laptop so that their predictions will vary considerably.

A class exercise of weighing a 3.75-lb book gave an average guess weight of 4.28 lb and a variance of 1.20 lb^2 from 73 students. In another class of 71 students, the average guess weight was 4.38 lb and the variance was 1.14 lb^2 .

5. Acknowledgments

I would like to acknowledge students of ENGG 319 in lecture sections L01 and L02, who have given me a motivation to explain these three topics to them as clearly as possible.

References

1. G. Blom and L. Rade, Probability and Statistics Courses for Students of Engineering at the University Level, Review of the International Statistical Institute Vol. 34, No. 2, 165-173 (1966).
2. For example: R. E. Walpole et al., Probability & Statistics for Engineers and Scientists (Prentice Hall, 9th Edition, 2012). D. C. Montgomery and G. C. Runger, Applied Statistics and Probability for Engineers (Wiley, 6th Edition, 2014).
3. R. V. Hogg, A Core in Statistics for Engineering Students, The American Statistician Vol. 48, No. 4, 285-287 (1994).
4. For example: M. Bona, A Walk Through Combinatorics: An Introduction to Enumeration and Graph Theory (World Scientific, 2nd Edition, 2003). R. J. Wilson, Introduction to Graph Theory (Prentice-Hall, 5th Edition, 2010).
5. R. J. Gould, Mathematics in Games, Sports, and Gambling – The Games People Play (CRC Press, Boca Raton, 2010).
6. J. Albert, Teaching Bayes' Rule: A Data-Oriented Approach, The American Statistician Vol. 51 No. 3, 247-253 (1997).
7. P. H. Kvam, The Effect of Active Learning Methods on Student Retention in Engineering Statistics, The American Statistician Vol. 54 No. 2, 136-140 (2000).