Honam Mathematical J. **28** (2006), No. 1, pp. 165–183

TEACHING PROBABILISTIC CONCEPTS AND PRINCIPLES USING THE MONTE CARLO METHODS

Sang-Gone Lee

Abstract. In this article, we try to show that concepts and principles in probability can be taught vividly through the use of the Monte Carlo method to students who have difficulty with probability in the classrooms. We include some topics to demonstrate the application of a wide variety of real world problems that can be addressed.

1. Preliminary

The Curriculum and Evaluation Standards for School Mathematics (NCTM,[6]) recommend that high school students extend their experiences with simulations to continue to improve their intuition and build more formal concepts of theoretical probability based on these experiences. Particularly, the notions of chance, of variation, and of statistical inference are increasingly a part of the experience of living in today's technological society. The mathematical methods dealing with such concepts is surely important knowledge for students to acquire. In the classroom, teachers and students are finding that these methods can infuse vitality into the teaching and learning situation. In keeping with this idea, our approach bring into focus on developing the real world

Received January 18, 2006. Revised March 7, 2006.

2000 Mathematics Subject Classification : 97U70.

Key words and phrases : Monte Carlo Methods, Problem Solving, Pseudocode.

†This paper was supported by research funds of Chonbuk National University (BS-2005-05).

mathematics application of the Monte Carlo methods in studying and investigating probabilistic situations with both high school students and future secondary mathematics teachers.

We believe that one method for studying probability is through the simulation of mathematical problems by a procedure that involves the use of sampling techniques based on probability to approximate the solution. This method is sometimes referred to as the Monte Carlo methods. The Monte Carlo methods are a very powerful tool for simulating probabilistic situations, not only for simple problems but also for rather complex problem situations. Although the method had been used before, the Monte Carlo method was named and developed during the Second World War by a group of mathematicians including J. V. Neumann to solve problems that arose in the design of atomic reactors. Also, the availability of computers has enabled to use the Monte Carlo method to solve a wide range of real world problems including the development of statistical tests.

Probability is usually a difficult subject for students. Having been introduced to probability relatively late, students have little intuition for it or experience with it. We believe that simulation gives students a feeling of power over probability. Recently, teachers have found simulation useful as a way of verifying results that students are reluctant to believe from a purely analytic explanation. The Monte Carlo simulation is easy to do in the classroom. Most needed materials can be quickly made. Students will find this kind of mathematics fun and enjoy trying to devise new and clever variations. If a computer is available in the classroom, we think that the Monte Carlo method is a convincing way of demonstrating its power. Participation in the Monte Carlo method will give students a better understanding of probabilistic concepts and principles involved, computers can also programmed to perform the simulation.

In this article, we illustrate some topics using the Monte Carlo methods for solving probability problems. Since the programming syntax is different for languages such as Pascal and Logo, we adopt the Pseudocodes that can be applied for Microsoft Excel, TI-graphing calculator, and the others with the same logic and degree of effectiveness through simple modifications of the programs used to simulate the problems. But we should consider some kind of struggle against the somewhat negative attitude toward the use of the new technology in the classroom, held by many of our colleagues.

2. The Monte Carlo Simulation Technique

The Monte Carlo simulation technique is a method of approximately solving mathematical problems by the simulation of random quantities. The name 'Monte Carlo' comes from the city of Monte Carlo in Monaco, famous for its gambling house. The systematic development of the Monte Carlo methods dates from about 1944. In the nineteenth century, statistical problems were sometimes solved with the help of random selections by the Monte Carlo method. In the last century, the Monte Carlo method has come back into favor. The Monte Carlo method becomes necessary when mathematical analysis is difficult or impossible and experimentation is expensive or otherwise impractical. This is physically different and easier to operate but has the same mathematical characteristics as the use of simple random devices such as dice, coins, random numbers from tables, or the random number generator of a computer. The theoretical basis for the Monte Carlo method is called the Law of Large Numbers, which states that for a larger number of times, the simulated estimate becomes approximately equal to the theoretical parameter. Prior to the appearance of electronic computers, this method was not widely applicable since the simulation of random quantities by hand is a very laborious.

Almost any probability or expected value problem can be solved by a suitable Monte Carlo method. It may help students believe the results by carrying out a simple classroom experiment that requires only one die for each student to simulate the die many times. The Monte Carlo simulation can be carried out at a very reasonable cost in the classroom. Since such experiments are instructive, students seem to get the added benefit of introducing them to the real world mathematics application of the Monte Carlo methods. But the point is gradually made that the laws of probability really do work if a sufficient number of trials are permitted to occur.

Our government has launched the 7th educational curriculum in 2000 for a movement to reform teaching at all levels. Almost all the texts that we use are written from a traditional point of view. Most of our math teachers teach probabilistic concepts and principles in the same way that are presented in the text they are using. Teachers want to teach students in an exploratory and constructive fashion with explicit guidance to build their knowledge with active learning on an efficient but passive transfer of information. We want to encourage the Monte Carlo methods to foster motive action and to improve problem solving on the basis of theoretical probability. When doing a judicious choice of the simulations, the Monte Carlo methods make students act, think, and evolve by their own motivation. Since the simulation may be consider as a fair play, student get to know its working rules which respond to anything they are doing. Understanding the simulation involves determining what the problem is seeking and what information and conditions are provided. Devising a Pseudocode involves finding a strategy for logical thinking that may help solve the problem. Carrying out the simulation involves following the solution procedure and checking it for flaws. Teacher can seek to be involved in a simulation with the system of interaction of the student with the problems. According to the nature of the class, teachers can decide if all the teams will work a problem at a time or if the different problems will be solved at the same time. The teamwork allows standardization of knowledge among peers, fosters discussion of different strategies of solution, and develops in the ability of communicate mathematical ideas. The simulation permits the justification that present a validation process that can be established between student and teacher. After summarizing the simulated results, teachers may try to introduce theoretical background for the simulation discussed previously and compare the result with the exact value. Thus, we believe that building the concepts and principles of probability based on the experiment may not only improve students' problem solving performance, but it may also help students more clearly understand the problem solving process.

In this article, we define three steps with slight modification in using the Monte Carlo methods:

• *Choose a problem and a random device.*

Begin by identifying an approximate model to employ, and find a random device applying the problem by the Monte Carlo method.

• *Identify the possible outcomes of each trial, and determine the probability of each outcome.*

A random device embody the mathematical characteristics of the problem. Show that the possible outcomes of the problem are matched to outcomes of the random device that have the same probability. Determine the probability of each outcome. State whether or not the trials in the original problem are independent.

• *Compute the simulated answer, and do a large number of trials.*

Show that by running additional trials with a new number, matches occur with the simulation of different sizes. Consider that many runs of the simulations are needed to give a sufficiently accurate estimate. In the section 4, we will show that to cut the error in half, the number of trials must be quadrupled.

Although students and future teachers can profit a great deal from studying and investigating the Monte Carlo methods, the Monte Carlo methods do not answer all the problems in teaching and learning probability. The Monte Carlo methods require careful planning on the part of the teacher and, careful attention to students' responses. When using the Monte Carlo methods, we believe that future mathematics teachers should notice that there is always the danger that students will be led through the problems too quickly, not being allowed the opportunity to think carefully about what they have done and to move with confidence from one stage of thinking to another.

3. Topics Suitable for the Monte Carlo Methods

Multiple Choice Test

When Students take a multiple choice quiz in our basic probability theory, some students are willing to take his chances. Our multiple choice quiz ask a student to recognize a correct answer among 4 answers that include 3 wrong answers. Suppose that we take the multiple choice quiz many times and provide students with an answer key after the quiz has been taken. We can keep a record of the number of correct answers obtained each time the quiz is taken.

What are his chances of getting k or more correct answers of those n questions by guessing?

The three steps for the Monte Carlo method are discussed below.

• *Choose a problem and a random device.* A way to think of answering a question by guessing is to draw a ball. Let an urn contain 4 colored balls. Suppose that n balls are drawn with independently and with replacement from the urn. At this time, we can use the following Pseudocode 1 because we have n Bernoulli trials, with success corresponding to a ball in a correct answer, failure to a ball in 3 wrong answers.

Pseudocode 1;

[Algorithm to count the number of correct answers]

[Use a conditional statement to tell whether the answer to each item is success or failure]

begin
$score \leftarrow 0$
$totscore \leftarrow 0$
for $j \leftarrow 1 \quad to \quad N$
  begin
  $genum \leftarrow RND(1)$
  if $genum \leq 0.5$ then
  $score \leftarrow 1$
  else $score \leftarrow 0$
  output $totscore \leftarrow totscore + score$
  end
end

• *Identify the possible outcomes of each trial, and determine the probability of each outcome.* Our experiment consists of drawing a ball n times, one for each question on the quiz, and then using the answer key to determine the number of correct answers. At each drawing the probability of success is $\frac{1}{4}$, and the probability of failure is $\frac{3}{4}$.

• *Compute the simulated answer, and do a large number of trials.* Looking at the results of trials with the Monte Carlo method, teachers should carry out the simulation as a class activity in which each student draws a ball or generates random number by a computer. Encourage students to conduct their own table keeping a record obtained from their trials. Introduce them to estimate the probability of getting k or more correct answers by the ratio:

$$\frac{The\ number\ of\ getting\ k\ or\ more\ correct\ answers}{The\ total\ number\ of\ trials} \text{ for } n \geq k.$$

Finally, on the mathematical background of students, teacher must allow for students to compare the ratio with the theoretical probability that get k or more correct answers of n questions by guessing, can be shown to be $\sum_{i=k}^{n} C(n,i)\frac{1}{4}^{i}\frac{3}{4}^{n-i}$ for $k = 0, \cdots, n$.

Distribution of Correct Answers

k	n_1	n_2	n_3
0	4	5	11
1	6	20	37
2	12	26	42
3	15	22	60
4	5	22	33
5	6	4	14
6	2	0	2
7	0	1	1
8	0	0	0
9	0	0	0
10	0	0	0
Trials	50	100	200

Table 1

Table 1 shows a summery of the number of correct answers obtained from 50, 100, and 200 trials respectively. The summery shows that the probability of getting 5 or more correct answers of 10 questions (based on 200 trials) is estimated as $\frac{17}{200} = 0.085$, which compares with the theoretical probability $\sum_{k=5}^{10} C(10,k)\frac{1}{4}^{k}\frac{3}{4}^{10-k} = 0.0781$. Figure 1 shows that as the number of trials increase, the simulated chances of getting k correct answers of 10 questions by guessing approaches the theoretical

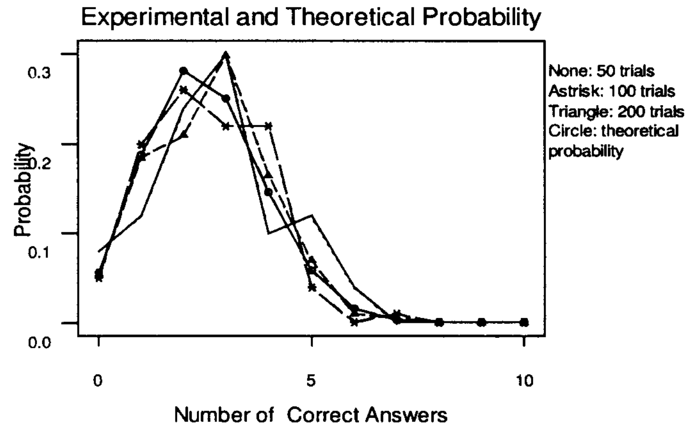

FIGURE 1

probability.

Definite Integration

The problem of determining the probability of some event, or its mathematical expectation, can be reduced to the computation of some integral. Assume that we want to evaluate a definite integral $\int_a^b f(x)dx$. If the function $f(x)$ fails to have a continuous derivative of moderate order, then the traditional numerical analytic techniques, such as Newton-Raphson and Simpson's Rule becomes less popular. In the classroom, the Monte Carlo methods becomes more impressive.

• *Choose a problem and a random device.* Develop the use of random numbers to evaluate a definite integral $\int_a^b f(x)dx$ for $a \leq x \leq b$. Perform independently the Monte Carlo simulation by selecting arbitrary points lying within the square randomly in the plane. We can use the Pseudocode 2 with the aid of a computer.

Pseudocode 2;

[Algorithm to count the points which happen to fall exactly on the actual area]

[M is a maximum value of a given function FN]

begin

input A,B, and M

for $j \leftarrow 1 \quad to \quad N$

begin

$X = RND(1) * (B - A) + A$

$Y = RND(2) * M$

If $Y \leq FN(X)$ then

$H = H + 1$

end

$I = \frac{H}{N} * M * (B - A)$

output I

end

• *Identify the possible outcomes of each trial, and determine the probability of each outcome.* Figure 2 shows the curve and the shaded area. The actual area S is bounded by the line $x = a, x = b, and\ curve\ f(x)$. Consider the desired area A to line within the square bounded by $x = a, x = b, and\ M$(maximum value of $f(x)$). Choose at random points in A and designate the number of points lying inside S. If the point falls within S, it is considered as a 'success' n_1. If it does not fall in S, it is considered a 'failure' n_2 and simply count in the total count n.

• *Compute the simulated answer, and do a large number of trials.* When the number of trials become great, the estimation S of the exact area $\int_a^b f(x)dx$ is $A * \frac{n_1}{n_1+n_2} = \frac{n_1}{n} * A$. If a function $f(x)$ is $x^3 - \frac{9}{2}x^2 + 6x + 1$, the exact area of S is

$\int_0^{2.5}(x^3 - \frac{9}{2}x^2 + 6x + 1)dx \approx 7.578$ over the interval $0 \leq x \leq 2.5$.

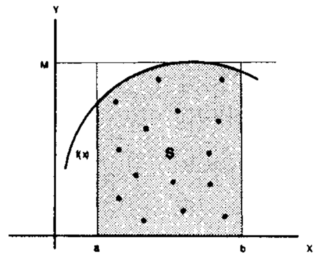

FIGURE 2

For attempting to evaluate the area by the Monte Carlo simulation, we counted respectively points lying within the square A and the actual area S randomly by generating random numbers. We obtained a typical approximate 7.56 of the desired trial 2592 for a total of 3000 samples, compared with the exact area 7.578. Figure 3 shows that the simulated area begin to converge as the number of trials increase.

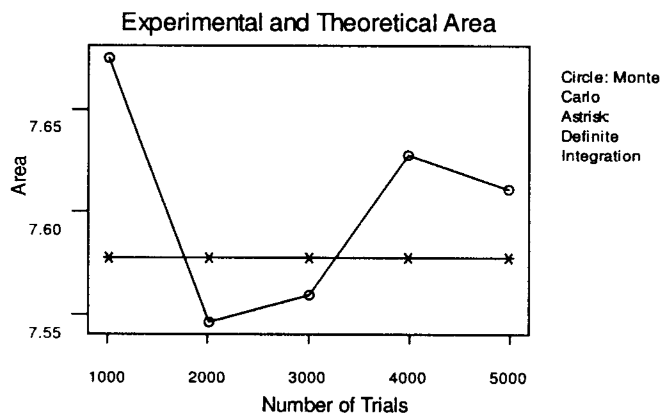

FIGURE 3

Birthday Problems

Our birthday problem finds the probability that at least two people in a group of some people will have the same birthday. We believe that the problem creates interest in the classroom. A random number table or a computer random digit generator lends itself well to this task. Other sources of random digits include spinners with ten equal divisions and the last digits in a series of telephone number. We can find that with the aid of the Monte Carlo method, the birthday problem can be simulated with little difficulty.[11]

Suppose that given a set of n numbers, each person in a group of k people choose a number at random from the set.($n \geq k$) Let us define A_k to be the event that a group of k people have distinct numbers. If we write the probability of the event A_k,

$$\begin{aligned} P(A_k) &= \frac{P(n,k)}{n^k} \quad for \quad k \leq n \\ &= \prod_{i=1}^{k-1}(1-\frac{i}{n}). \end{aligned}$$

Since the geometric mean is strictly less than the arithmetic mean, taking the respective means yields the inequality.

$$\begin{aligned} \prod_{i=1}^{k-1}(1-\frac{i}{n})^{\frac{1}{k-1}} &\leq \frac{\sum_{i=1}^{k-1}(1-\frac{i}{n})}{k-1} \\ &= 1-\frac{k}{2n} \end{aligned}$$

Therefore, the desired upper bound of $P(A_k)$ is given by

$$P(A_k) = \prod_{i=1}^{k-1}(1-\frac{i}{n}) \leq (1-\frac{k}{2n})^{k-1}.$$

Consider the probability that at least two people in a group of k people will have the same numbers.

$$
\begin{aligned}
P(There\ will\ be\ at\ least\ two\ repetitions) &= 1 - P(A_k) \\
&= 1 - \prod_{i=1}^{k-1}(1 - \frac{i}{n}) \\
&\geq 1 - (1 - \frac{k}{2n})^{k-1} \quad for \quad k \leq n
\end{aligned}
$$

Since the birthday problem can be handled in much the same way as the birthday of the year, we made three tables using the permutation that can find exact values of the theoretical probability to have a greater than 60 percent chance that at least two people in the group share the same birthday, birth week , or birth month.

At least two people share birthday

n	k	Probability
365	30	0.7063
365	29	0.6810
365	28	0.6545
365	27	0.6269
365	26	0.5982
365	25	0.5687
365	24	0.5383

Table 2

At least two people share birth month

n	k	Probability
12	7	0.8886
12	6	0.7772
12	5	0.6181
12	4	0.4271
12	3	0.2361

Table 3

At least two people share birth week

n	k	Probability
7	5	0.8501
7	4	0.6501
7	3	0.3878

Table 4

The tables above show that 4 people is sufficient to give a better than 60 percent chance that at least two people have birthdays on the same week, 5 people is sufficient on the same month, and 27 people is sufficient on the same birthday of the year respectively. Since the birth month problem can be easily applied to solve the other birth problems with the Monte Carlo methods, we only discuss the birth month problem in the following.

What is the minimum number of people needed for the probability that two of them were born on the same month of the year to be 60 percent ?

In terms of the Monte Carlo method, we will try to have the three steps.

• *Choose a problem and a random device.* Use a twelve-sided die, one side for each month of the year. For the birth month problem, students can use a graphic calculator or a simple computer software providing the necessary random number.

Pseudocode 3;

[Algorithm to generate the birth month]
begin
input N
for $i \leftarrow 1$ to N
output INT $(12 * RND(1)) + 1$
end

• *Identify the possible outcomes of each trial, and determine the probability of each outcome.* Assume that a trial consists of rolling the twelve-sided die independently five times, once for each of the five people in the group. Count the same side of the die obtained at least twice in five rolls.

• *Compute the simulated answer, and do a large number of trials.* The probability that at least two people share the same birth month to be 60 percent is estimated by the ratio:

$$\frac{\textit{Number of the same side of trials obtained at least twice}}{\textit{Number of trials}}$$

In each trial, we count the number of the same side obtained at least twice in five rolls. When the die was tossed 100 times, we got 14 successes of same side. The probability that at least two people share birth month to be 60 percent is 0.7. After tossing the die 500 times, we obtained 68 successes of same side die. The probability that in a group of five people at least two people were born in the same month to be 60 percent is estimated as 0.68, which compares with the lower bound of 0.6072 and the theoretical probability of 0.6181.

4. Accuracy of the Monte Carlo Method

The Monte Carlo method is also subject to some error as all measurements are subject to some error. It is difficult to determine how many trials should be run for a sufficiently accurate estimate using the Monte Carlo method. Approaching a problem with the Monte Carlo method using the same data is different, we might consider that the mean value is often equal to the solution being sought. In order to illustrate the general nature of the Monte Carlo methods, we begin the discussion with a very simple example.

We denote X_i $(i = 1, \cdots, n)$ a random variable as a result which is obtained for each of n independent trials. Let A be an event with a probability p of occurring. Assume that each random variable X_i has a finite expectation $E(X_i) = p$ and a finite variance $\text{Var}(X_i) = \sigma^2$. If we let $S_n = \sum_{i=1}^{n} X_i$ as the total number of trials in which the event A occurs, the relative frequency of occurrence of the event A is $\frac{S_n}{n}$, which is a random variable with $E(\frac{S_n}{n}) = p$ and $Var(\frac{S_n}{n}) = \frac{\sigma^2}{n}$.

The Law of Large numbers says that for $\epsilon > 0$ and $\delta > 0$, there exists a number n of trials such that with probability less than ϵ, the relative frequency of occurrence of the event A will differ from the probability p of the occurrence of this event by not less than δ:

$$P(|\frac{S_n}{n} - p| \geq \delta) < \epsilon \text{ for large enough n.}$$

Since $\frac{S_n}{n}$ means the approximate value obtained for p by the Monte Carlo method, then the difference $\frac{S_n}{n} - p$ is the error of the Monte Carlo method. This means that, with a probability greater than or equal to $1 - \epsilon$, the following inequality holds:

$$|\frac{S_n}{n} - p| < \frac{\sigma}{\sqrt{n\epsilon}} \text{ for large enough n,}$$

which follows from Chebyshev's Inequality. The relation $\delta \sim \frac{1}{\sqrt{n}}$ holds. Letting $k = \frac{\sigma}{\sqrt{\epsilon}}$, the role of the variance of X_i is clearly evident as is that of the number n of trials. Thus, the accuracy of the Monte Carlo method is determined only by the number of independent trials and by the variance.

Let us consider the fundamental case in which the distribution of $\frac{S_n}{n}$ is approximately Normal from the Central Limit Theorem. Choosing the confidence level for the estimate of the error with a finite degree of certainty $\delta = \pm 10$ percent of the population mean, we would do $\frac{1}{0.1^2} = 100$ trials. It is clear that to decrease the error by a factor of 10, it is necessary to increase n by a factor of 100. Therefore, the Monte Carlo method cannot give any solution of very high accuracy. Exceptionally when events occur with small probability p, the Poisson distribution frequently holds for integer values of S_n. The relation $p \sim \frac{1}{n}$ become a condition for the Poisson distribution to appear as the limiting law for S_n.

Ordinarily it is rather difficult to estimate the variance σ^2 prior to solving the problem. In most practical applications of the Monte Carlo methods, it is always advisable to reduce this variance as much as possible, if necessary by employing a special procedure called variance reducing techniques. H. Kahn[3] says that the Monte Carlo methods to reduce variance can be sharply dependent upon the probability model and the techniques used to generate the values of the random variables. If the problem mandates little variance from trial to trial, then you should use a large sample size. If the size of the variance is of little consequence, then a smaller number of trials is appropriate.

5. Conclusion

The Monte Carlo method is a technique for solving problems using random outcomes of experiment, which in their simplest form, involve such activities as drawing balls, rolling dice, or generating the random number. The Monte Carlo methods can improve directly the solution of real world problems with a minimal amount of mathematical background and a reasonable cost in the classrooms. By creating the data processing, students can challenge a variety of problems encountered in daily life rather than contrived, 'textbook' problems. We present some real world topics as students can tackle in a conventional probability class until they provide a concrete interpretation of a solution as well as ways of organizing and analyzing data. The Monte Carlo methods lead the class into such mathematical activities as exploring, conjecturing, examining, and proving. By employing the Monte Carlo method in teaching, it is possible to make important probabilistic concepts and principles accessible to both high school students and future secondary mathematics teachers. We think that instructional applications of the Monte Carlo methods promise important new directions in the teaching and learning of concepts and principles in probability.

Particularly the Pseudocodes can be considered to get a better picture of problem solving applying the mathematical logic for which given topics can be modified easily when the data is transferred to a special programming language. During simulating the topics using the Monte Carlo methods, students can build more concepts and principles of theoretical probability based on random experiences. At this time, we hope that students take a considerable opportunity to explore and move with confidence from one stage of thinking to another.

References

[1] R. George. The Birthday Problem: A Short Lesson in Probability, 2005. http://www.mste.uiuc.edu/reese/birthday/intro.html.

[2] J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods.* Spottiswoode, Barllantyne and Co, London, 1964.

[3] H. Kahn. *Use of Different Monte Carlo Sampling Techniques.* John Wiley and Sons, 1956. Symposium on Monte Carlo Methods.

[4] E. S. Keeping. *Introduction of Statistical Inference.* Van Nostrand, 1962.

[5] L. M. Lesser. Exploring the Birthday Problem with Spreadsheets. *Mathematics Teacher*, 92:407–411, 1999.

[6] National Council of Teachers of Mathematics. *Curriculum and Evaluation Standards for School Mathematics.* NCTM, 1989.

[7] I. Peterson. Mathtrek, Birthday Surprises, 2005. http://www.sciencenews.org.

[8] R. Y. Rubenstein. *Simulation and Monte Carlo Method.* John Wiley and Sons, 1981.

[9] J. M. Shaughnessy. Research in Probability and Statistics: Reflections and Directions. *In Handbook of Research on Mathematics Teaching and Learning*, pages 465–494, 1992.

[10] Y. A. Shreider. *The Monte Carlo Method.* Pergamon Press, 1966.

[11] H. Susan. Welcome to Probability by Surprise, 2005. http://www-stat./stanford.edu/ susan/surprise/Birthday.html.

[12] M. C. Whitney. Exploring the Birthday Paradox Using a Monte Carlo Simulation and Graphing Calculators. *Mathematics Teacher*, 94:258–262, 2001.

Sang-Gone Lee
Department of Mathematics Education
Chonbuk National University
ChonJu, Korea
Email : sgl@chonbuk.ac.kr