Article

Monte Carlo Analysis: An Application to Aircraft Design and Crash

Emre Soydemir [1] and Panagiotis Petratos [2,*]

1   International Baccalaureate Program, Modesto High School, Modesto, CA 95351, USA;
    emre.soydemir04@gmail.com
2   Department of Management Information Systems, California State University, Stanislaus, 1 University Circle,
    Turlock, CA 95382, USA
*   Correspondence: ppetratos@csustan.edu

Abstract: The current study investigates the application of statistical methods to flight, which have been used in science over time to understand complex physical and mathematical systems by using randomly generated numbers as input into those systems to generate a range of solutions and, specifically, how mathematics is used to examine airplane design and crash frequency. In order to make very accurate predictions, one also requires an appropriate mathematical model. Using randomly selected numbers, the Monte Carlo statistical method is able to make very accurate predictions. With the Monte Carlo statistical method, by using significantly larger numbers of trials, the likelihood of the solutions can be determined very accurately. Currently, Monte Carlo methods are widely used and play a key part in various fields of science. Monte Carlo methods have vast uses in trials with limited observations that cannot be replicated many times. This paper adds new findings to the knowledge base on causes of crashes by airplane design. First, mathematical methods are used in this paper to investigate what the most likely casualty number and range are in the five years after the first flight based on 5000 simulations. Second, an investigation is performed to determine if certain casualty numbers are outliers of certain airplane designs based on the number of casualties reported using Monte Carlo analysis.

Keywords: Monte Carlo methods; flight; casualties

1. Introduction

Monte Carlo methods were invented in the 1930s by Enrico Fermi [1,2] and were used to solve crucial problems in developing the atomic bomb in the 1940s. It was not possible to make many experiments of explosion. Therefore, scientists had to rely on simulations. Enrico Fermi invented the Monte Carlo method for studies in neutron diffusion in Rome. Fermi did not publish the Monte Carlo method as a stand-alone article but used it to solve many problems in his other publications. Fermi took great delight in impressing greatly his Roman colleagues with his remarkably accurate, "too-good-to-believe" predictions of experimental results. After indulging himself, he revealed that his "guesses" were really derived from Monte Carlo statistical sampling techniques. Fermi, during his hiatus from the ENIAC operation at Los Alamos National Laboratory, invented a simple but ingenious analog device for studies in neutron transport collision, and he persuaded his friend and collaborator Percy King to build such an instrument, later called the FERMIAC. Stanislaw Ulam then introduced the Markov Chain Monte Carlo method for the ENIAC operation at Los Alamos National Laboratory. John von Neumann understood its importance and programmed the ENIAC computer to perform Monte Carlo simulations [3,4]. Scientists working on the Manhattan Project had to model what would happen in a chain reaction in highly enriched uranium. Projections had to be accurate and could not deviate from actual results. Monte Carlo simulations were the answer [5,6]. Unlike a normal forecasting model, Monte Carlo simulation predicts a set of outcomes based on an estimated range

of values versus a set of fixed input values [7,8]. Scientists used the first "computers", which were calculators, and early IBM punched-card machines in which people entered numbers by hand in each simulation. However, the problem had so many dimensions that systematically plugging in and trying numbers in all these dimensions took far too long. Modern computer architecture provides a solution for this problem with the linear increase of computing performance as computing cores in the silicon microchip increase.

Monte Carlo simulation [1] is a mathematical model or a multiple probability simulation that is used to compute the possible outcomes of an uncertain event. From a set of fixed input values, (e.g., a five-year data set for Boeing 737-Max), it predicts a set of outcomes based on an estimated range of values. It leverages a probability distribution, such as a uniform or normal distribution, to build a model of possible results for any variable that has inherent uncertainty. It, then, recalculates the results continually, each time using a different set of random numbers between the minimum and maximum values. In a typical Monte Carlo experiment, this procedure can reoccur thousands of times to produce a large number of likely outcomes. Monte Carlo simulations are also utilized for long-term predictions due to their accuracy. As the number of inputs increases, the number of forecasts also grows, allowing one to project outcomes farther out in time with more accuracy. When a Monte Carlo simulation is complete, it yields a range of possible outcomes with the probability of each result occurring. One simple example of a Monte Carlo simulation is to consider calculating the probability of rolling two standard dice. There are 36 combinations of dice rolls. Based on this, one can manually compute the probability of a particular outcome. Using a Monte Carlo simulation, one can simulate rolling the dice 10,000 times (or more) to achieve more accurate predictions [1].

This paper looks at some applications in flight that have been used over time and how mathematics is used to examine airplane design and crash frequency [9]. Using randomly selected numbers, the Monte Carlo statistical method is able to make very accurate predictions. With the Monte Carlo statistical method, by using significantly larger numbers of trials, the likelihood of the solutions can be determined extremely accurately [10]. Currently, it is widely used and plays a key part in various fields of science. Monte Carlo methods have vast uses in trials with limited observations that cannot be replicated many times [11]. This paper adds new findings to the knowledge base on causes of crashes by airplane design. First, mathematical methods are used in this paper to investigate what the most likely casualty number and range are in the five years after the first flight based on 5000 simulations. Second, an investigation is performed to see if certain casualty numbers are outliers of certain airplane designs based on the number of casualties reported using Monte Carlo analysis.

2. Methodology

The Monte Carlo method is a mathematical technique also known as statistical sampling [2]. Monte Carlo simulation can be developed to model the probability of different outcomes that present uncertainty and then play them out on a computer thousands of times. Monte Carlo simulation is a mathematical numerical method that uses random draws to perform calculations and solve complex problems.

One of the most common used generators is the following:

$$X_{n+1} = (aX_n + b) \bmod m \tag{1}$$

to generate numbers, where a, b and m (integer modulus m > 1) are large integers, and X_{n+1} is the next in X as a series of pseudo-random numbers. The maximum number the formula can produce is one less than the modulus, m − 1. To avoid certain non-random properties of a single linear congruential generator, several such random number generators with slightly different values of the multiplier coefficient, a, can be used in parallel, with a "master" random number generator that selects data sources from among the several different generators [12].

In reality, however, no random draw is truly random, as it depends on the root. Each time the root is different, a distinct random process occurs, similar to a Polaris cleaner in a pool. When the Polaris cleaner is tied to a different wall of the pool, resulting random movements differ.

Monte Carlo simulations can be utilized to replicate, say, 1000 trials of a limited occurrence. For example, the mean and dispersion of the damage done by less than a handful of atomic bomb explosions can be simulated by Monte Carlo trials. These can be used to project the actual radius of the damage in real-life explosions.

One can use the cumulative distribution function (CDF) to calculate the probability that the variable takes a value less than or equal to x [13]. The plot of the normal cumulative distribution function is S-shaped starting from zero on the *y*-axis [13]. Because the vertical axis is a probability, it must fall between zero and one. This is particularly suited to Excel command RAND( ), which generates random numbers between 0 and 1 (it is worth noting here that each time a key is pressed, a whole different set of Monte Carlo trials is generated). The probability increases from zero to one as we go from left to right on the horizontal axis. The CDF can be calculated using the VLOOKUP command in Excel to assign a real-life occurrence based on each randomly generated number.

The mean from the simulations assigned casualty numbers of 5000 iterations (N) can be found by using the basic mean calculation formula:

$$X_{mean} = \sum X_i/N \text{ where X is the number of casualties and } i = 1, \ldots, N \quad (2)$$

A measure of dispersion around the mean can be found by utilizing the following formula (one can also calculate the weighted average using the COUNTIF command in Excel). Thus, any value above or below the dispersion interval would warrant closer scrutiny:

$$\text{Dispersion measure} = \sum \sqrt{(X_i - X_{mean})^2/(N-1)} \quad (3)$$

It is possible to estimate if, say, one particular occurrence is an anomaly by calculating a range around the mean in the following manner:

$$U_b = \text{upper bound} = \text{mean} + \text{dispersion} \quad (4)$$

$$L_b = \text{lower bound} = \text{mean} - \text{dispersion} \quad (5)$$

If a particular occurrence falls outside of the upper or lower bound, it may be treated as an anomaly. The cause must then be looked at carefully to understand if this occurrence must be interpreted differently than the rest of the sample.

3. Data

Table 1 reports data from the Aviation Safety Net Database [14]. Casualty numbers for each airplane design for the five years after the first flight are reported. The numbers are cumulative. For example, the third-year number is the sum of the past years, the fourth-year number reported is the sum of the past four years, and so on.

Table 1. Cumulative probability with Boeing 737-Max in the sample.

Design	Probability *	Casualties
Boeing 737-200	0	45
Airbus A320-200	0.06	92
DC9-32	0.19	262
Boeing 737-max	0.54	346

* 0.06, 0.13, 0.35, 0.46 expressed cumulatively.

It is possible to infer from Table 1 that all four designs of passenger airplanes reported casualties in their first five years. Boeing 737-Max had the maximum number of casualties,

while Boeing 737-200 had the least number of casualties reported in their first five years. Boeing 737-Max makers can claim that the casualty numbers are normal.

However, it is possible to apply mathematical tools such as Monte Carlo analysis to investigate whether they are normal or constitute a significant outlier for regulators agencies such as the Federal Aviation Association to halt the flying of Boeing, Chicago, IL, USA, 737-Max airplanes for safety reasons until further tests are performed on airplane safety.

Significant studies quantify the risk of extreme aviation accidents [15] and provide a survey of aviation risk and safety modeling [16].

4. Monte Carlo Analysis

Table 1 reports the cumulative distribution function based on the casualty numbers, and as the way the CDF is reported, it starts with zero, and probabilities are added. At the end of the five-year interval, there were a total of 745 casualties. Out of 745, about 6% belong to Airbus A320-200, 13% belong to Airbus A30-200, 35% belong to DC9-32 and 46% belong to Boeing 737-Max, totaling 1.0, as it is supposed to.

Casualty numbers are a very limited sample, and the experiments cannot be controlled. There are only a few observations over the years. However, one can resort to simulations using the Monte Carlo approach to generate, say, 5000 numbers from a standard normal distribution to arrive at a number that is more representative of the population mean and dispersion. To illustrate the mathematical approach, only 50 simulations will be reported; however, the Supplementary Materials report all of the 5000 simulations and the corresponding numbers in Table 2. The Corresponding Casualty Value column in Table 2 lists the values from the Monte Carlo method: 5000 randomly selected casualty numbers from the five-year Boeing 737-Max data set. The mean from the simulations assigned casualty numbers of 5000 iterations (N) can be found by using the basic mean calculation formula:

$$X_{mean} = \sum X_i / N \quad \text{where X is the number of casualties, and } i = 1, \dots, N$$

Table 2. Monte Carlo simulations *.

Simulation	Random Number	Corresponding Casualty Value
1	0.029511	45
2	0.877784	346
3	0.049684	45
4	0.375842	262
5	0.97594	346
6	0.909059	346
7	0.206874	262
8	0.326738	262
9	0.2458	262
10	0.976921	346
11	0.49824	262
12	0.708151	346
13	0.752177	346
14	0.055357	45
15	0.587237	346
16	0.44498	262
17	0.485908	262
18	0.481106	262
19	0.045698	45
20	0.108847	92
21	0.700739	346

Table 2. *Cont.*

Simulation	Random Number	Corresponding Casualty Value
22	0.951193	346
23	0.279288	262
24	0.072672	92
25	0.377888	262
26	0.156489	92
27	0.521516	262
28	0.239952	262
29	0.856471	346
30	0.611255	346
31	0.089922	92
32	0.067326	92
33	0.923868	346
34	0.143525	92
35	0.444293	262
36	0.319392	262
37	0.743319	346
38	0.636609	346
39	0.65753	346
40	0.205481	262
41	0.447495	262
42	0.095009	92
43	0.794265	346
44	0.481688	262
45	0.946968	346
46	0.39313	262
47	0.444996	262
48	0.654459	346
49	0.191648	262
50	0.833734	346

* = rand( ) = vlookup(lookup_value,table_array,col_index_num,[range_lookup]).

Table 3 reports the cumulative probabilities excluding Boeing 737-Max based on the casualty numbers from the data of the Aviation Safety Net Database for all other aircraft [14]. At the end of the five-year interval after the first flight, there were a total of 399 casualties for all designs. Out of 399, about 12% belong to Boeing 737-200, 23% belong to Airbus A30-200, and 65% belong to DC9-32, totaling, again, 1.0, as it is supposed to.

Table 3. Cumulative probability without Boeing 737-Max in the sample.

Design	Probability *	Casualties
Boeing 737-200	0	45
Airbus A320-200	0.12	92
DC9-32	0.35	262

* 0.12, 0.23, 0.65, expressed cumulatively.

Table 4 reports Monte Carlo simulation results excluding Boeing 737-Max only for the first 50 simulations. The results of the 5000 simulations are provided in the Supplementary Materials of this paper. Excel commands to use to generate both simulations are the random number generator = rand( ) and to assign the value on the cumulative probability table the Excel command to use is:

=vlookup(lookup_value,table_array,col_index_num,[range_lookup]).

Table 4. Monte Carlo simulations with Boeing 737-Max excluded *.

Simulation	Random Number	Corresponding Casualty Value
1	0.938014	262
2	0.877194	262
3	0.506338	262
4	0.164717	92
5	0.369072	262
6	0.698868	262
7	0.661901	262
8	0.807135	262
9	0.757239	262
10	0.06584	45
11	0.429569	262
12	0.237415	92
13	0.898775	262
14	0.927234	262
15	0.374385	262
16	0.692643	262
17	0.223326	92
18	0.50979	262
19	0.306565	92
20	0.843186	262
21	0.62789	262
22	0.033259	45
23	0.512249	262
24	0.452089	262
25	0.432025	262
26	0.811374	262
27	0.923514	262
28	0.113079	45
29	0.473194	262
30	0.434762	262
31	0.9362	262
32	0.195844	92
33	0.632611	262
34	0.375676	262
35	0.434593	262
36	0.755377	262
37	0.569038	262
38	0.462824	262
39	0.053393	45
40	0.312179	92
41	0.158278	92
42	0.195863	92
43	0.860629	262
44	0.979724	262
45	0.210679	92
46	0.380913	262
47	0.084189	45
48	0.065844	45
49	0.032934	45
50	0.652622	262

* = rand( ) = vlookup(lookup_value,table_array,col_index_num,[range_lookup]).

Table 5 reports Monte Carlo analysis of 5000 simulations with Boeing 737-Max included in the top part of Table 5 and excluded in the bottom part of Table 5 from the sample. The top part of Table 5 produces a mean value of 263 casualties and dispersion around a mean of 101 casualties. The upper and lower bounds are attained by adding and subtracting the dispersion measure from the mean, respectively. The upper bound of casualties is 364, while the lower bound of casualties is 163 (provided at the end of the 5000 simulations).

Table 5. Mean and dispersion of the 5000 Monte Carlo simulations.

With Boeing 737-Max in the Sample	
Mean	263
Median	262
MAD	076
Dispersion	101
Upper bound	364
Lower bound	163
Without Boeing 737-Max in the Sample	
Mean	197
Median	262
MAD	083
Dispersion	089
Upper bound	287
Lower bound	107

The bottom part of Table 5 reports sample results when Boeing 737-Max is excluded to avoid bias, as would otherwise be expected. Only three comparator aircraft were used because of the data provided by the Aviation Safety Database [14], which uses the same set of comparator aircraft. However, some aircraft that were reported had zero casualties in the first five years and higher casualties in the following five years. These were excluded to prevent bias against Boeing 737-Max. This was done to increase the reliability and robustness of the study. Simulations produce a mean value of 197 casualties and dispersion around a mean of 89 casualties. The upper and lower bounds are attained by adding and subtracting the dispersion measure from the mean. The upper bound of casualties is 287, while the lower bound of casualties is 107. The Standard Deviation and the Mean Absolute Deviation MAD are robust tools to flag outliers in the data set. The top part of Table 5 reports MAD with a value of 76, and the bottom part reports MAD with a value of 83. The median is the same for both samples because 262 is the most frequently occurring value.

5. Discussion

New findings of the current study and statistical analysis results demonstrate that the number of casualties reported by the Aviation Safety Net Database, as well as the number of casualties predicted by the statistical analysis methods for the Boeing 737-Max aircraft, is significantly different than the number of casualties caused by the other types of aircraft (Boeing, Chicago, IL, USA, 737-200, Airbus, Leiden, The Netherlands, A320-200, McDonnell Douglas, St. Louis, MO, USA, DC9-32) included in the current study. These new findings warrant further investigation into the paradox of the unusually high number of casualties for the type of aircraft Boeing 737-Max. Limitations of the current study include the limited Aviation Safety Net Database five-year data available after the first flight of Boeing 737-Max. This is the reason the Monte Carlo method was selected to add 5000 data points for the statistical analysis. Robust statistical analysis measures, including median, standard deviation, and mean absolute deviation, are included to verify the Monte Carlo analysis results and to detect the outliers in the data set. There is a need to explain better the intended impact of this work for the readers to understand the novel application of the Monte Carlo statistical method to aviation. It is noteworthy to explain the potential advantages to be

gained from this method in future studies if more data are available and how the confidence in the approach would increase when applied to more data.

To clarify and clearly explain, the same simple example of the atomic bomb explosions can be used herein. For example, instead of detonating 5000 atomic bombs to study the resulting nuclear explosions and their impact, the scientists used the Monte Carlo method for 5000 instances to re-create the exact same conditions of the nuclear explosions.

In the same modus operandi, instead of 5000 test flights to study the resulting crashes and potential casualties, the scientists can use the Monte Carlo method for 5000 instances to re-create the exact same conditions of the test flights and of the crashes.

Furthermore, the intent, due to the five-year available data, is to analyze the probability of accidents in new designs, over their first five years, to reveal information about the likelihood of new designs crashing. This suggests that the data used for the study should be only from accidents where the cause was design related, ideally. For example, if an aircraft crashed due to weather or pilot error, rather than due to aspects of its design, this should not be included, ideally.

Unfortunately, this is not the case with the cumulative data available from the Aviation Safety Net Database. We know, after the grounding of Boeing 737-Max, there are design issues with Boeing 737-Max; however, this is not the case for all the other aircraft data from the Aviation Safety Net Database.

In other words, the casualties reported from the database are cumulative due to all causes and not strictly design related only.

Hence, as the ideal data set is not available, we must use the available data from the Aviation Safety Net Database, which reports cumulative casualties due to all causes and not design related only.

It is necessary for the data used in this study to be explained in the context of these issues so that the interpretation of the results can be meaningful.

The scientific value of a statistical process to future aircraft safety is irreplaceable compared to the usual detailed assessment of each case in respect to its specific circumstances. The statistical evaluation of casualties in airliner accidents can provide an objective framework by which to confirm perceptions of a particular aircraft being an outlier and to relate this to the specific circumstances of the crashes such as, for example, to indicate whether management had taken appropriate decisions after a first accident.

In other words, the Monte Carlo statistical method is exceedingly valuable for future aircraft safety, to minimize casualties and study flight conditions, especially during the aircraft development and test flight phases.

6. Conclusions

In conclusion, Boeing 737-Max had 346 casualties in the five-year interval. It is important to exclude Boeing 737-Max data in the second phase of the statistical analysis to arrive at unbiased results. As can be seen, 346 casualties are above the upper bound of 287 casualties when Boeing 737-Max is excluded from the sample. Therefore, the casualty numbers for Boeing 737-Max are significantly different than the rest of the sample of passenger airplanes considered in this study and can be seen as a mathematical anomaly, constituting evidence that the casualties of Boeing 737-Max were exceptionally high, warranting closer scrutiny.

One weakness of the study is only less than a handful of airplane designs are investigated due to limitations of the data. The strength of the study is the simulation technique that replicates a normal distribution by way of repeated sampling. Monte Carlo trials, therefore, allow us to arrive at relatively robust results despite the data limitation on airplane crashes.

References

1. IBM Cloud Education. Monte Carlo Simulation. Available online: https://www.ibm.com/cloud/learn/monte-carlo-simulation (accessed on 25 March 2021).
2. Metropolis, N. The beginning of the Monte Carlo Method. *Los Alamos Sci.* **1897**, *15*, 125–130. [CrossRef]
3. Grabowski, J.G.; Curriero, F.C.; Baker, S.P.; Li, G. Exploratory spatial analysis of pilot fatality rates in general aviation crashes using geographic information systems. *Am. J. Epidemiol.* **2002**, *155*, 398–405. [CrossRef] [PubMed]
4. Gotoh, H.; Takezawa, M.; Maeno, Y. Risk analysis of airplane accidents due to bird strikes using monte carlo simulations. *Trans. Inf. Commun. Technol.* **2012**, *44*, 398–403. [CrossRef]
5. Fleisher, H.J.; Benaroya, H. Investigation of Monte Carlo simulation in FAA program KRASH. *J. Aircr.* **1994**, *31*, 367–375. [CrossRef]
6. Lyle, K.H.; Stockwell, A.E.; Hardy, R.C. Application of Probability Methods to Assess Airframe Crash Modeling Uncertainty. *J. Aircr.* **2007**, *44*, 1568–1573. [CrossRef]
7. Dunn, W.L.; Shultis, J.K. *Exploring Monte Carlo Methods*; Elsevier: Amsterdam, The Netherlands, 2012.
8. Winston, L.W. *Microsoft Office Excel 2007: Data Analysis and Business Modeling*; Microsoft Press: Redmond, WA, USA, 2007.
9. Kroese, P.D.; Taimre, T.; Botev, Z.I. *Handbook of Monte Carlo Methods*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 706.
10. Rubinstein, R.Y. *Simulation and the Monte Carlo Method*; John Wiley & Sons: New York, NY, USA, 1981.
11. Wonnacott, H.T.; Wonnacott, R.J. *Introductory Statistics*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 1990.
12. Shaukat, S.S.; Zafar, A.; Noor, N.; Khan, A.M. Random numbers and monte carlo simulation: Applications in genetic drift and random walk models of predator-prey interaction. *Int. J. Biol. Biotechnol.* **2020**, *17*, 195–218.
13. CDF vs. PDF: What's the Difference? Available online: https://www.statology.org/cdf-vs-pdf/ (accessed on 25 March 2021).
14. Aviation Safety Net Database. Available online: https://www.riskope.com/2019/04/03/boeing-737-max-8-set-in-risk-context/ (accessed on 25 March 2021).
15. Das, K.P.; Dey, A.K. Quantifying the risk of extreme aviation accidents. *Phys. A Stat. Mech. Appl.* **2016**, *463*, 345–355. [CrossRef]
16. Netjasov, F.; Janic, M. A review of research on risk and safety modelling in civil aviation. *J. Air Transp. Manag.* **2008**, *14*, 213–220. [CrossRef]