THE PROBABILITY OF DYING IN A PLANE CRASH OR HAVING A SAFE FLIGHT

Ireneusz JOZWIAK, Aleksander MARIANSKI, Michal CWIK

Wrocław University of Technology
27 Wybrzeże Wyspiańskiego St. 50-370 Wrocław, Poland
E-mail: aleksander.marianski@pwr.edu.pl (corresponding author)

Ireneusz JOZWIAK, *Prof. Dr Habil*
Education: MSc in computer science, Department of Electronics, Wrocław University of Technology (WUT), 1975; doctor's degree in computer science, Institute of Engineering Cybernetics, Wrocław University of Technology, 1979; habilitation degree in computer science, automatic and robotics, Institute of System Research of the Polish Academy of Sciences, 1994; professor of Wrocław University of Technology, 1998.
Affiliations and functions: Wrocław University of Technology, Institute of Engineering Cybernetics (1979–1990), Department of Computer Science and Management (since 1990) Society of Polish Informatics (president since 1995); Society of Education (president since 2005); Committee on Automatic Control and Robotics of the Polish Academy of Sciences (member since 1994), Polish Society for Measurement, Automatic Control and Robotics (POLSPAR) (member since 2008); International Federation of Automatic Control (IFAC) (member of TC 3.3 i TC 5.3 since 2008).
Research interests: system safety and reliability.
Publications: author of 2 books, co-author of 3 books. Contributor to 220 articles in professional journals.

Aleksander MARIANSKI, *PhD student*
Education: 2006–2011 MSc degree in computer science, Wrocław University of Technology, Faculty of Computer Science and Management. 2012 – to present, PhD studies in computer science, Wrocław University of Technology, Faculty of Computer Science and Management. Specialization: system safety and reliability.
Affiliations and functions: 2009–2011 assistant in SOA project at Wrocław University of Technology.
Research interests: applied probability in computer science.
Publications: author of 3 conference papers.
Present position: PhD student.

Michal CWIK, *PhD student*
Education: 2006–2011, MSc degree in computer science, Wrocław University of Technology, Faculty of Computer Science and Management. 2012 – to present, PhD studies in computer science, Wrocław University of Technology, Faculty of Computer Science and Management. Specialization: operational research.
Research interests: operational research applied in computer science.
Publications: author of 1 conference paper.
Present position: PhD student.

Abstract. In the article, an overall study involving a plane crash and the likelihood of having a safe flight is presented. The probability of dying in a plane crash and the probability of getting to the destination safely were calculated within the period 1970–2008. The formula for calculating the average length of a sequence of consecutive safe flights was determined. The probability of dying in a car accident and the probability of dying as a result of a plane crash were compared as a conclusion.

Keywords: aviation, security, plane crashes, probability, estimation of probability.

1. Introduction

Ensuring a high level of safety and reliability is one of the key issues in aviation. Of all accidents, aircraft crashes are one of the major topics of research. Flights are and will be attractive targets for terrorists (Hougham 2009). Also, many articles (Colebunders 2011; Oakes, Bor 2010; Van Gerwen *et al.* 1997) show that fear of flying is a common phenomenon. In the authors' opinion fear of flying is not equivalent to the probability of being harmed as a result of a plane crash this hypothesis was proved using statistical analysis.

In previous publications the issues of security procedures in an event of an aviation disaster (Hougham 2009), extracting knowledge from data in the field of aviation security (Nazeri *et al.* 2001), or fear of flying (Van Gerwen *et al.* 1997) were discussed. In recent years, there have been many publications on improving safety in aviation engineering (Krause 2003; Shyur 2008; Sweet 2009). NASA has launched the "Aviation Safety & Security Program", in which they publish a number of reports on aviation security (NASA 2013). An extensive study of aviation accidents can be found in (Krause 2003). However, the authors did not come across a study on the likelihood of dying in a plane crash. In this publication the probability and its variation over 38 years are studied.

Reliability is the ability of an object to fulfill the required function in a specific environment and operating conditions in a given period of time. The reliability of an object is determined by a reliability function $R(t)$. This is the probability that the object does not fail in an interval $(0; t)$. It is equivalent to the probability that a random variable T – a time that elapses from the start of the operation of the object until its damage – does not gain a value from the interval $[0; t)$; therefore, $R(t) = P(T \geq t)$.

Publications in the field of aviation generally examine the reliability of plane components and materials (Al-Garni *et al.* 1997; Huang *et al.* 2012; Jun, Huibin 2012). These works are important for aviation engineers, pilots, air traffic controllers and other people directly connected with aviation. In order to minimize disasters, related to the unreliability of aircraft components, special training centers are created such as the East Midlands Training Center. Pilots undergo theoretical training regarding the main elements of an airplane and spend a few days on practical training in a simulator. The simulator allows simulating failure of any combination of items and flights in any conditions. Possible phenomena, such as turbulence and smoke in the plane, can also be simulated. Specialists of such training centers are trying to minimize the impact of human factor in a crash. Procedures are devised which the pilots have to know and have to repeat many times. The most important element of the training and later practice is the safety procedures. The pilots must repeat the safety procedures before starting up every plane. It is an assurance that the pilot will behave automatically in an emergency situation, which increases the likelihood of a safe flight.

From the passengers' point of view it is not relevant whether or not plane components work properly. The passengers are interested in a safe, timely and comfortable journey to their destination. On this basis the authors have specified two key safety indicators. The criterion was the safety of the passengers:

– The probability of completing at least n flights without any incident for a passenger.
– The probability of completing at least n flights without a death of a passenger.

In this paper an analysis was carried out in order to determine the probability of completing at least n flights without incurring a death of a passenger. This indicator was chosen for the analysis because passengers mostly have a fear of death in a plane crash (Van Gerwen *et al.* 1997). In the following part of the paper the concept "safe flight" is defined as a flight without a death of a given passenger.

The main goal of many publications is to raise the level of safety in aviation. The hypothesis is that the safety in aviation increases each year. The motivation was a widespread opinion that the thesis is correct but has hardly any statistical proof in publications. The authors used frequency analysis of fatal accidents in aviation to prove the thesis.

2. Passenger-plane system

In the article, an examination of the reliability of the passenger-plane system is performed using a similar pattern as for the analysis of the human-machine system. The goal, which is to met by the passenger-plane system, is undamaged passenger transport by aircraft. The system is defined as working properly, if within a given flight a given passenger flies safely. The system is defined as faulty, if within a given flight a given passenger dies.

In the passenger-plane system time points are defined. Consecutive time points represent subsequent flights taken by a passenger. Time is measured in successive natural numbers.

The system uses the concept of an elementary event. It is defined as the possibility of a given passenger to take at least n flights without the passenger's death. The elementary event was denoted as A_n, where n is the number of held flights. The space of the elementary events is, therefore, the sum of all elementary events. The designation $P(A_n)$ is defined as the probability of holding at least n flights with a given passenger without the passenger's death.

On the basis of the data on the number of passengers using airlines and the number of deceased people due to aircraft accidents, we studied how probability function $P(A_n)$ changed over time. For this purpose, a sample set was defined as a set of the number of passengers and the number of deaths as a result of air accidents from 1970 to 2008 for each year. The time range has been established according to the availability of complete source materials for the period (Bureau… 2012; International… 2008; Kilroy 2013; PlaneCrashInfo 2013).

3. Obtaining the sample data

The data of air crashes and flights was collected from two sources:

– The archives of the ACRO organization (data about air disasters, with the exception of the events of 11.09.2001) (Bureau… 2012).
– The data from the International Civil Aviation Organization (International… 2008).

These sources were selected because of the completeness of the data. ACRO (Aircraft Crashes Record Office) is an organization dedicated to archiving aviation accidents since 1918. ICAO (International Civil Aviation Organization) is an organization dedicated to archiving the statistical data on civil aviation since 1970. Other sources (Kilroy 2013; PlaneCrashInfo 2013) contain data from a shorter period of time or are incomplete.

Data collected from these sources us not saved in a format that is easy to reuse. It has to be read from the web pages. Due to the large amount of data we decided to automate the process of collecting and processing using a dedicated author's application.

The data from the sources (International… 2008; Bureau… 2012) includes information from the period: 1918–2013 – ACRO and from 1970 to 2008 – ICAO. In order to analyze the probability of holding at least n flights without a fatal accident for a passenger, the data on accidents and the number of passengers in a given year are necessary. The period for analysis was chosen to be 1970–2008, as the data from this period was complete.

Sources (International… 2008; Bureau… 2012) are the websites where the organizations summarize the results of their work. From the ACRO archives we obtained the total number of deaths in a given year. The archives of ICAO contained information about the number of passengers carried. This data was divided into groups according to countries. Data downloaded for each country has been counted to calculate the sum of all passengers carried during the year.

To download the data PHP programming language was used, along with a library Simple html DOM (Chen 2012). The websites should be written in standard-compliant HTML so that it would be possible to parse the data using an XML parser. In contrast to other document formats, web pages are often written incorrectly. For example, the ACRO website (Bureau… 2012), according to the W3C Validator (W3C 2013) contains 1210 errors, while the ICAO website (International… 2008) contains 56 errors. Therefore, it is necessary to use tools that enable processing of the source data, despite errors. The Simple Html DOM library supports invalid HTML, and in known cases carries out correct parsing of invalid documents.

4. The probability function of carrying out *n* safe flights

We defined A_i' as an event opposite to event A_i (defined in section 2). A_i' is the event of having i flights, of which at least one was fatal to a given passenger. A_1 is an event of having one flight that was fatal to a given passenger. In order to determine $P(A_i)$ we first obtained the estimation of $P(A_1')$ in successive years from 1970. For this purpose, we calculated what is the part of passengers that had a fatal accident in relation to all passengers carried during the year. The data is illustrated in Figure 1.

The x-axis represents individual years, while the y-axis denotes the ratio of the number of people who died as a result of an air accident to the total number of passengers, expressed in percent and marked by *R*. Variable *R* in each period is indicated by squares. In Figure 1 there is also a B-spline curve of the 10th degree (Chen 2010).

The graph in Figure 1 shows that function *R(t)* is decreasing. This provides support to the hypothesis that the estimated probability of safe flights is increasing. According to graph, we can estimate the value of likelihood of having one flight without the death of a given passenger each year, denoted by *P(A₁)*:

$$P(A_1) = 1 - P(A_1') \ . \tag{1}$$

Using equation (1) and Figure 1, we estimated, that in 1972 the *P(A₁)* was equal to approximately 99.9991% while in 2000 it was equal to nearly 99.9999%.
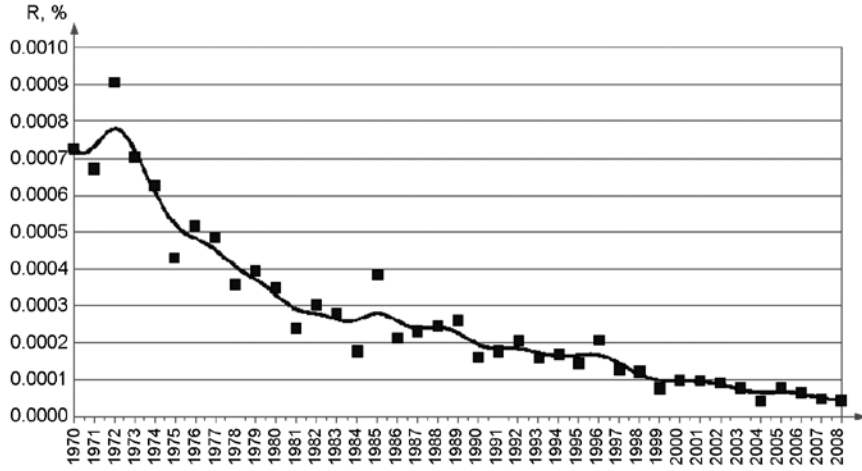
Fig. 1. The percentage of people who died in air accidents relative to all carried passengers in the period between 1970 and 2008

As a safe flight for a passenger we understand a flight during which the passenger survived. Let's denote by B_i the event of *i-th* flight being safe for a given passenger. $\{B_i\}$ is a sequence of independent repetitions of the same trial. The probability of carrying out n safe flights for a given passenger, denoted as $P(A_n)$ is obtained the following equation:

$$P(A_n) = \prod_{i=1}^{n} P(B_i). \tag{2}$$

It must be noted, that $P(B_1) = P(B_2) = \ldots = P(B_n)$ and that $A_1 = B_1$. Taking this into consideration for equation (2), we get:

$$P(A_n) = P(A_1)^n. \tag{3}$$

Based on equations (1) and (3) and the data illustrated in Figure 1 the estimation of the function $P(A_n)$ for a given period was calculated. Graphs of functions $P(A_n)$ are illustrated in Figure 2. The x axis represents the number of flights, denoted by n, while the y axis represents the probability of carrying out *n* safe flights for a given passenger, which is denoted by $P(A_n)$. The graph was drawn on the basis of the data from 1972, 1992, 2008 years and from the calculations of $P(A_n)$ for these years. The range of *n* was fixed at *a* [0, 210.000] interval.

From the results shown in Figure 2 it can be concluded that the probability of death in a consequence of air accident has been decreasing in consecutive years. It is worth mentioning that in year 2008 one should take approximately 200000 flights in order to reach a 10% probability of death in air accident.

Let C_n be an event of carrying out exactly $n – 1$ safe flights for a given passenger. Then C_n is an intersection of two events: A_{n-1} of carrying out at least $n – 1$ safe flights and B_n', that the *n-th* flight will end with a passenger's
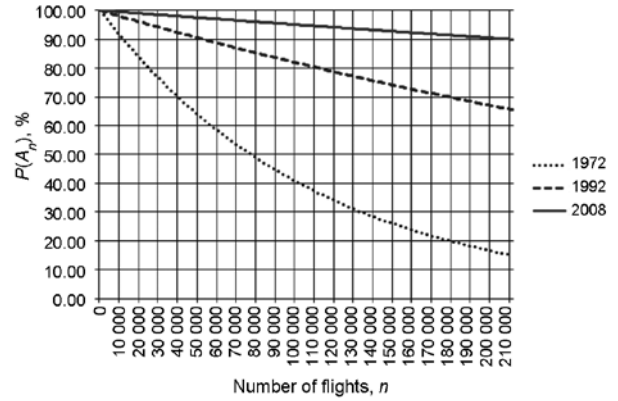

Fig. 2. The probability of carrying out *n* safe flights for a given passenger in 1972, 1992 and 2008

death. Using equations $P(B_1') = P(B_2') = \ldots = P(B_n')$ and $A_1' = B_1'$ we obtain that:

$$P(C_n) = P(A_{n-1})P(A_1'). \tag{4}$$

Let us denote $X(C_i)$ as:

$$X(C_i) = i. \tag{5}$$

X is a random variable assigned to an event of carrying out exactly $i – 1$ safe flights for a given passenger number *i*. Using definition (5) and equation (4) a formula for the average length of safe flights for a given passenger is obtained. It is the expected value of random variable *X*. It is assumed that the probability of event A_1 is obtained before. The expected value of random variable X is equal to:

$$EX = \sum_{n=1}^{\infty} nP(A_{n-1})P(A_1'). \tag{6}$$

Substituting into equation (6) the formula obtained in equations (3) and (1) we get:

$$EX = \sum_{n=1}^{\infty} nP(A_1)^{n-1}(1 - P(A_1)). \tag{7}$$

Next the following transformations in order to calculate the sum of the above series are made:

$$EX = (1 - P(A_1)) \sum_{n=1}^{\infty} nP(A_1)^{n-1} =$$

$$(1 - P(A_1)) \left(1 + \sum_{n=2}^{\infty} nP(A_1)^{n-1} \right) =$$

$$(1 - P(A_1)) \left(1 + \sum_{n=1}^{\infty} (n+1)P(A_1)^n \right) = \qquad (8)$$

$$(1 - P(A_1)) \left(1 + \sum_{n=1}^{\infty} nP(A_1)^n + \sum_{n=1}^{\infty} P(A_1)^n \right).$$

In order to simplify the notation this denotation was introduced:

$$y = \sum_{n=1}^{\infty} nP(A_1)^n. \qquad (9)$$

Also, the sum of the series below was calculated by:

$$\sum_{n=1}^{\infty} P(A_1)^n = \frac{P(A_1)}{1 - P(A_1)}. \qquad (10)$$

When we equate the expressions obtained in the equations (7) and (8) and make the following transformations we see that:

$$\sum_{n=1}^{\infty} nP(A_1)^{n-1}(1 - P(A_1)) =$$

$$(1 - P(A_1)) \left(1 + \sum_{n=1}^{\infty} nP(A_1)^n + \sum_{n=1}^{\infty} P(A_1)^n \right) \times$$

$$\frac{1 - P(A_1)}{P(A_1)} \sum_{n=1}^{\infty} nP(A_1)^n =$$

$$(1 - P(A_1)) \left(1 + \sum_{n=1}^{\infty} nP(A_1)^n + \sum_{n=1}^{\infty} P(A_1)^n \right). \qquad (11)$$

When we substitute variable y from (9) into equation (11) and use equation (10) we get:

$$\frac{1 - P(A_1)}{P(A_1)} y = (1 - P(A_1)) \left(1 + y + \frac{P(A_1)}{1 - P(A_1)} \right) \times$$

$$\frac{1}{P(A_1)} y = 1 + y + \frac{P(A_1)}{1 - P(A_1)} \times$$

$$\left(\frac{1}{P(A_1)} - 1 \right) y = \frac{1}{1 - P(A_1)} \times y = \frac{p(A_1)}{(1 - P(A_1))^2}. \qquad (12)$$

Using equation (7) and the definitions of y from (9) and from (12) we get that:

$$EX = \sum_{n=1}^{\infty} nP(A_1)^{n-1}(1 - P(A_1)) =$$

$$\frac{y}{P(A_1)} (1 - P(A_1)) = \frac{1}{1 - P(A_1)}. \qquad (13)$$

From equation (13) the average length of the sequence of safe flights for a given passenger can be calculated. For the data from 2008, where $P(A_1') = 0.0000005$ the average length of the sequence of safe flights was equal to 2000000, while in 1982, where $P(A_1') = 0.000003$ the average length of the sequence of safe flights was equal to 333333.

5. Conclusions

In the article the analysis of the works on reliability in the field of aviation is presented. To the knowledge of the author's there are no papers describing the probability of carrying out n safe flights. We introduced three groups of elementary events. By A_i we denoted the event of carrying out at least i safe flights for a given passenger. By B_i we denoted the event of carrying out the *i-th* flight safely for a given passenger. By C_i we denoted the event of carrying out exactly $i - 1$ safe flight by a given passenger.

On the basis of the source materials from ACRO and ICAO we calculated the percentage of people who died in the air accidents relatively to all carried passengers during the year. This quantity was denoted by R. The data was illustrated in the Figure 1. On the basis of the results presented in Figure 1 it was evident that R is a decreasing function from which it was concluded that the safety of air travels has been increasing in the period 1970–2008.

It has also been found that in year 2008 the probability of taking a flight during which a given passenger suffers death, is equal to 0.0000005. The authors compared this data with the data from NHTSA the National Highway Traffic Safety Agency for the year 2008 (National… 2008). On the basis of the NHTSA data and the result of this work, the authors noticed that driving about 40 miles with a car results in the same probability of a death accident as one flight does.

In the paper the authors also devised formula (13) to calculate the average length of the sequence of safe flights for a given passenger. From the formula the authors calculated the average length of the sequence of safe flights on the basis of data from the year 2008.

References

Al-Garni, A.; Sahin, A.; Al-Farayedhi, A. 1997. A reliability study of Fokker f-27 airplane brakes, *Reliability Engineering & System Safety* 56(2): 143–150. http://dx.doi.org/10.1016/S0951-8320(97)00004-5

Bureau of Aircraft Accidents Archives 2012. *Aircraft crashes record office* [online], [cited 20 December 2013]. Available from Internet: http://www.baaa-acro.com

Chen, S. 2012. *Php simple html dom parser* [online], [cited 20 December 2013]. Available from Internet: http://simple-htmldom.sourceforge.net

Chen, W.-K. 2010. *Feedback, Nonlinear and Distributed Circuits*. CRC Press.

Colebunders, J. 2011. Cured of fear of flying, *Travel Medicine and Infectious Disease* 9(2): 82.
http://dx.doi.org/10.1016/j.tmaid.2011.01.001

Hougham, C. 2009. Aviation security in the face of tragedy, chapter 8 in M. R. Haberfeld, A. von Hassell (Eds.). *A New Understanding of Terrorism.* Springer, 143–158.

Huang, W.; Wang, T.-J.; Garbatov, Y., *et al.* 2012. Fatigue reliability assessment of riveted lap joint of aircraft structures, *International Journal of Fatigue* 43: 54–61.
http://dx.doi.org/10.1016/j.ijfatigue.2012.02.005

International Civil Aviation Organization. 2008. *Air Transport, Passengers Carried by Country. Civil Aviation Statistics of the World and Icao Staff Estimates* [online], [cited 20 December 2013]. Available from Internet: http://www.nationmaster.com/country-info/stats/Transport/Air-transport/Passengers-carried

Jun, L.; Huibin, X. 2012. Reliability analysis of aircraft equipment based on fmeca method, *Physics Procedia* 25: 1816–1822. http://dx.doi.org/10.1016/j.phpro.2012.03.316

Kilroy, C. 2013. *Airdisaster.com solutions for safer skies* [online], [cited 20 December 2013]. Available from Internet: http://www.airdisaster.com

Krause, S. S. 2003. *Aircraft Safety: Accident Investigations, Analyses, and Applications.* McGraw-Hill.

NASA. 2013. *NASA Aviation Safety Program* [online], [cited 20 December 2013]. Available from Internet: http://www.aeronautics.nasa.gov/programs-aosp.htm

National Highway Traffic Safety Administration. 2008. *Traffic Safety Facts* [online], [cited 20 December 2013]. Available from Internet: http://www-nrd.nhtsa.dot.gov/pubs/811170.pdf

Nazeri, Z.; Bloedorn, E.; Ostwald, P. 2001. Experiences in mining aviation safety data, *ACM SIGMOD Record* 30: 562–566. http://dx.doi.org/10.1145/376284.375743

Oakes, M.; Bor, R. 2010. The psychology of fear of flying: a critical evaluation of current perspectives on the nature, prevalence and etiology of fear of flying, *Travel Medicine and Infectious Disease* 8(6): 327–338.
http://dx.doi.org/10.1016/j.tmaid.2010.10.001

PlaneCrashInfo. 2013. *Plane crash info Incidents* [online], [cited 20 December 2013]. Available from Internet: http://www.planecrashinfo.com

Shyur, H.-J. 2008. A quantitative model for aviation safety risk assessment, *Computers & Industrial Engineering* 54(1): 34–44. http://dx.doi.org/10.1016/j.cie.2007.06.032

Sweet, K. 2009. *Aviation and Airport Security. Terrorism and Safety Concerns.* CRC Press.

Van Gerwen, L. J.; Spinhoven, P.; Diekstra, R. F., *et al.* 1997. People who seek help for fear of flying: typology of flying phobics, *Behavior Therapy* 28(2): 237–251.
http://dx.doi.org/10.1016/S0005-7894(97)80045-7

W3C. 2013. *Markup validation service* [online], [cited 20 December 2013]. Available from Internet: http://validator.w3.org