A WAY TO ACCELERATE THE PROCESS OF GATHERING INFORMATION FOR DECISION-MAKING

Deyan Mihaylov

University of Economics-Varna
e-mail: dgmihaylov@ue-varna.bg
Bulgaria

Abstract: The decision-making is a delicate balance between competing subjects. The managers spend large part of time on gathering necessary information. If the process of data-collecting can be accelerated, the duration of decision-making will be reduced. A way to achieve that is to use not the whole possible information, but only sufficient part of it. This article explains how the duration of data gathering can be reduced if the decision-maker accepts this idea. The duration of actions of collecting information is modeled by time-measured random variables.

Key words: Decision-making, Gathering data, Distribution function.

1. INTRODUCTION

Decision-making is an important component of good management. There is a wide range of decision-makers varying from individual entrepreneur [1], teacher [2] or clinician [3], to military staff, assisted by systems for command, control, communications, computers, intelligence and surveillance [4, 5]. Often the decision process is a competition between rival organizations. So, the decision-makers work under time-pressure. It is necessary to construct good decisions but these decisions must be constructed rapidly. Usually decision process is decomposed into a sequence of particular steps. Every step includes some activities. The steps are analyzed and (if it is possible) activities are executed in parallel. In this way the duration of decision process can be reduced.

One of the first models is described by Herbert A. Simon. The decision-making is explained as a three-phase process. The first phase is surveying the economic, technical, political, social environment. The decision-makers spend a large fraction of time on this phase. Simon calls this phase “intelligence activity (borrowing the military meaning of intelligence)”. The second phase is “design activity”. The decision-makers “spend even larger fraction of their time,…, seeking to invent, design, and develop possible courses of action…”. The third phase is “choice activity”. The decision-makers “spend small part of their time in choosing among alternative actions already developed” [6].

Let us discuss the first phase. A object or a system is examined. The survey includes activities of collecting and analyzing data, testing the object, etc. If we receive data from many and different sources and test object in some ways, we hope that the object will be fully analyzed. Let us suppose that all possible data about this object are collected and all possible tests are made. Hence, the decision maker has all the evidence concerning the structure and the behavior of the object. Recently this is called “weight of evidence” [7].

Suppose that the “intelligence activities” can be performed in parallel. This is possible if the decision-maker has several assistants. (Here “assistant” means a person or a team. The assistant may be a part of the organization of the decision-maker or may be temporarily hired.) In this way the process can accelerate. But usually the number of assistants has some limitations. Then few data will be collected or a few tests will be performed. Hence, the decision-maker has a subset of evidence. The term “strength of evidence” is used in this case [7].

The purpose of this article is to explain a way to accelerate the first phase of decision-making. The constraints of this study are:

1. A single object is examined.
2. The full investigation of the object can be separated into several actions which can be performed in parallel. The duration of every action is time-measured random variable. The distribution functions of all the random variables are known.
3. Decision-maker does not have enough assistants to ensure all actions are executed in parallel. So, the principle of decision-making is “strength of evidence”. The decision-maker considers that it is sufficient to complete n investigating actions to ensure decision-making. Moment of ensuring of decision-making is the time, when all the n actions are completed with probability not less than a predetermined positive number $\alpha \le 1$.

The aim is to make the decision as fast as possible. The durations of the actions are random variables, so the time of completion of all n actions is a random variable, too. If the distribution functions of durations of actions are known, the distribution function of time, required to be completed all the n actions can be obtained.

2. PROBLEM AND SOLUTION

We investigate n parallel actions that run simultaneously. The process durations are random variables ξ_1, ξ_2, …, ξ_n, with cumulative distribution functions (CDF) F_1, F_2, …, F_n. F_j is the probability that the random variable ξ_j is less than or equal to t [8].

$$F_j(t) = P(\xi_j \le t). \tag{1}$$

The domain of each one CDF, given in (1) is $t \in [0, \infty)$. This follows from the nature of time as a physical quantity.

Suppose that the random variables ξ_1, ξ_2, $\ldots \xi_n$ are mutually independent, and let by the random variable η be denoted the time required "the n specific processes to be completed". Then

$$F_\eta(t) = P(\eta \le t). \quad (2)$$

According to the definition given above,

$$\eta = \max(\xi_1, \xi_2, \ldots \xi_n).$$

The CDF of the maximum for mutually independent random variables is known to be given by:

$$F_\eta(t) = F_1(t) \cdot F_2(t) \cdot \ldots \cdot F_n(t). \quad (3)$$

The range of every CDF is between 0 and 1. Hence

$$F_\eta(t) = F_1(t) \cdot F_2(t) \cdot \ldots \cdot F_n(t) \le F_j(t), \; j \in \{1, 2, \ldots n\}, 0 \le t < \infty. \quad (4)$$

Let α be the desired probability to ensure decision-making. The CDFs are monotonically increasing, hence exist $t_1, t_2, \ldots, t_n, t_\eta$, such that $F_1(t_1) = \alpha$, $F_2(t_2) = \alpha$, …, $F_n(t_n) = \alpha$, $F_\eta(t_\eta) = \alpha$. Moreover,

$$t_\eta \ge \max(t_1, t_2, \ldots t_n), \quad (5)$$

i.e. by increasing the number of investigating actions the necessary time is increased.

A way to decrease the required time is to add resource. Let us include $n+1$ actions. The ensuring of decision-making occurs when n of all the actions are completed. So, one of the actions is "hot reserve".

The CDF is equal to the probability that the process has been completed before t. Let t be fixed. We denote the event "the j-th action is completed" by A_j. Then

$$P_j = F_j(t) = P(\xi_j \le t) \quad (6)$$

is the probability of occurrence of A_j at the moment t.

Let A (without index) is the event "at least n actions are completed at the time t". For example, combination of events $A_1 A_2 \ldots A_n \overline{A_{n+1}}$ implies A. $\overline{A_{n+1}}$ means the complement of A_{n+1}. All the combinations of events, which imply A are

$A_1 A_2 A_3 \ldots A_n \overline{A_{n+1}}$;

$A_1 A_2 \ldots \overline{A_n} A_{n+1}$;

…

$A_1\overline{A_2}...A_n A_{n+1}$;

$\overline{A_1}A_2...A_n A_{n+1}$;

$A_1 A_2...A_3 A_4$.

The last one takes into consideration possibility that all the events occur simultaneously.

Then the probability of A is

$$P(A) = P\left(A_1 A_2...A_n \overline{A_{n+1}}\right) + P\left(A_1 A_2...\overline{A_n} A_{n+1}\right) + ... + \\ + P\left(A_1 \overline{A_2}...A_n A_{n+1}\right) + P\left(\overline{A_1} A_2...A_n A_{n+1}\right) + P\left(A_1 A_2...A_n A_{n+1}\right).$$

The events are mutually independent, so

$$P\left(A_1 A_2...A_n \overline{A_{n+1}}\right) = P_1 P_2...P_n (1 - P_{n+1}) = P_1 P_2...P_n - \prod_{j=1}^{n+1} P_j \ ;$$

$$P\left(A_1 A_2...\overline{A_n} A_{n+1}\right) = P_1 P_2...(1 - P_n) P_{n+1} = \prod_{j=\{1,...,n+1\}\backslash n} P_j - \prod_{j=1}^{n+1} P_j \ ;$$

…

$$P\left(A_1 \overline{A_2}...A_n A_{n+1}\right) = P_1 (1 - P_2)...P_n P_{n+1} = \prod_{j=\{1,...,n+1\}\backslash 2} P_j - \prod_{j=1}^{n+1} P_j \ ;$$

$$P\left(\overline{A_1} A_2,,, A_n A_{n+1}\right) = (1 - P_1) P_2...P_n P_{n+1} = \prod_{j=\{1,...,n+1\}\backslash 1} P_j - \prod_{j=1}^{n+1} P_j \ ;$$

$$P(A_1 A_2...A_3 A_4) = \prod_{j=1}^{n+1} P_j \ .$$

$\prod_{j=\{1,...,n+1\}\backslash k} P_j$ means the production of all the P_j, excluding P_k.

After substitution we find

$$P(A) = P_1 P_2...P_n + \sum_{k=1}^{n} \prod_{j=\{1,...,n+1\}\backslash k} P_j - n \cdot \prod_{j=1}^{n+1} P_j \ . \qquad (7)$$

Let by the random variable ν be denoted the time required "at least n of the $n+1$ processes to be completed". According to formula (6) we can substitute $P_j = F_j(t)$ and $P(A) = F_\nu(t) = P(\nu < t)$ in (7):

$$F_\nu = F_1 F_2...F_n + \sum_{k=1}^{n} \prod_{j=\{1,...,n+1\}\backslash k} F_j - n \cdot \prod_{j=1}^{n+1} F_j \ . \qquad (8)$$

Let us compare $F_\nu(t)$ and $F_\eta(t)$. We find

$$F_\nu - F_\eta = F_1F_2...F_n + \sum_{k=1}^{n} \prod_{j=\{1,...,n+1\}\backslash k} F_j - n \cdot \prod_{j=1}^{n+1} F_j - F_1F_2...F_n =$$

$$= \sum_{k=1}^{n} \left(\prod_{j=\{1,...,n+1\}\backslash k} F_j - \prod_{j=1}^{n+1} F_j \right) = \sum_{k=1}^{n} \left[(1-F_k) \cdot \prod_{j=\{1,...,n+1\}\backslash k} F_j \right].$$

All the addends in the right-hand-side of the equilibrium are non-negative, hence

$$F_\nu \geq F_\eta .$$

Both functions are monotonically increasing. Analogically to the aforementioned by formulas (4) and (5) there are t_η and t_ν for every positive $\alpha \leq 1$, such that $t_\eta \geq t_\nu$ and $F_\nu(t_\nu) = F_\eta(t_\eta) = \alpha$.

Hence, the inclusion of an additional action accelerates the decision-making.

3. EXAMPLES

Let the durations of four actions are modeled by absolutely continuous random variables ξ_1, ξ_2, ξ_3 and ξ_4, which are identically distributed, i.e. $F_1 = F_2 = F_3 = F_4 = F$. It is sufficient to complete 3 actions to ensure decision-making. According to (3) and (8), the CDF of η and ν are

$$F_\eta = F^3;$$

$$F_\nu = F^4 + 4F^3 - 4F^4 = 4F^3 - 3F^4.$$

Let the desired probability, ensuring the decision-making, is α. The values of t_η and t_ν are solutions of equations

$$F_\eta(t_\eta) = [F(t_\eta)]^3 = \alpha; \qquad (9)$$

$$F_\nu(t_\nu) = 4[F(t_\nu)]^3 - 3[F(t_\nu)]^4 = \alpha. \qquad (10)$$

The values of functions $F_\eta(F)$ and $F_\nu(F)$ are 0 when $F = 0$ and 1 when $F = 1$. Hence, they take any value between 0 and 1 at some point within the interval [0, 1], i.e. the solutions of $F^3 = \alpha$ and $4F^3 - 3F^4 = \alpha$ exist. The derivatives $F'_\eta(F) = 3F^2$ and $F'_\nu(F) = 12F^2 - 12F^3$ are positive when $0 \leq F \leq 1$, $F_\eta(F)$ and $F_\nu(F)$ are strictly increasing on the interval [0, 1] and the solutions are unique. Hence, the inverse functions of $F_\eta(F)$ and $F_\nu(F)$ exist. The domain of the inverse functions is the interval (0, 1) and they can be obtained from (9) and (10):

$$F(t_\eta) = F_\eta^{-1}(\alpha) = \alpha^{\frac{1}{3}}, \tag{11}$$

$$F(t_\nu) = F_\nu^{-1}(\alpha) = \frac{1}{6}\left(\sqrt{2}\sqrt{\frac{3\alpha}{\sqrt[3]{\sqrt{\alpha^2-\alpha^3}+\alpha}} + 3\sqrt[3]{\sqrt{\alpha^2-\alpha^3}+\alpha} + 2} + 2\right) -$$

$$-\frac{1}{2}\left(-\frac{2\alpha}{3\sqrt[3]{\sqrt{\alpha^2-\alpha^3}+\alpha}} - \frac{2}{3}\sqrt[3]{\sqrt{\alpha^2-\alpha^3}+\alpha} + \right.$$

$$\left. + \frac{8\sqrt{2}}{9\sqrt{\frac{3\alpha}{\sqrt[3]{\sqrt{\alpha^2-\alpha^3}+\alpha}} + 3\sqrt[3]{\sqrt{\alpha^2-\alpha^3}+\alpha} + 2}} + \frac{8}{9}\right)^{\frac{1}{2}}. \tag{12}$$

The graphs of $F_\eta^{-1}(\alpha)$, $F_\nu^{-1}(\alpha)$ and $F = \alpha$ are shown at Figure 1.

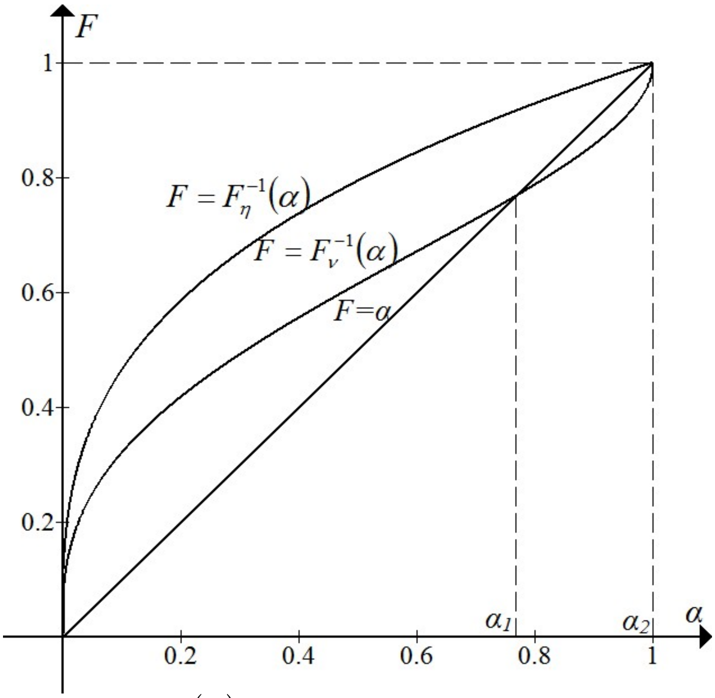

Fig. 1. Graphs of $F(t_\eta) = F_\eta^{-1}(\alpha)$, $F(t_\nu) = F_\nu^{-1}(\alpha)$ *and* $F(t) = \alpha$.

If $\alpha = F(t)$ is strictly increasing, the inverse function $t = F^{-1}(\alpha)$ exists and it is possible to obtain

$$t_\eta(\alpha) = F^{-1}\left[F_\eta^{-1}(\alpha)\right] \;; \tag{13}$$

$$t_\nu(\alpha) = F^{-1}\left[F_\nu^{-1}(\alpha)\right] \text{ and} \tag{14}$$

$$t(\alpha) = F^{-1}(\alpha). \tag{15}$$

The analysis of the functions of Figure 1 shows that in their domain there is an interval, such, that $F_\nu^{-1}(\alpha) < F^{-1}(\alpha)$. The limits of this interval are points of intersection of the lines $F = F_\nu^{-1}(\alpha)$ and $F = F^{-1}(\alpha) = \alpha$ (or of the inverses $\alpha = F_\nu(F)$ and $\alpha = F(F) = F$). The lower limit is $\alpha_1 = \frac{1+\sqrt{13}}{6}$ and the upper one is $\alpha_2 = 1$ (see Figure 1). Using (14) and (15) it is obtained that $t_\nu(\alpha) < t(\alpha)$, i.e. it is possible that the time to collect data from n sources to be shorter than time to collect data from the only one, if an additional resource is included.

Example 1. Let all the actions have minimal execution time a and maximal execution time b. Suppose that the time of execution is evenly distributed i.e. all ξ-s have uniform distribution $U(a,b)$. The CDF is given by

$$F(t) = \begin{cases} 0, & t \le a \\ \dfrac{t-a}{b-a}, & a < t \le b \\ 1, & b < t\,. \end{cases}$$

According to (3) and (8), the CDF-s of η and ν are given respectively by

$$F_\eta(t) = F(t)^3 = \begin{cases} 0, & t \le a \\ \left(\dfrac{t-a}{b-a}\right)^3, & a < t \le b \\ 1, & b < t \end{cases}$$

and

$$F_\nu(t) = 4F(t)^3 - 3F(t)^4 = \begin{cases} 0, & t \le a \\ 4\left(\dfrac{t-a}{b-a}\right)^3 - 3\left(\dfrac{t-a}{b-a}\right)^4, & a < t \le b \\ 1, & b < t\,. \end{cases}$$

If $F(t) = u$, then

$$F^{-1}(u) = (b-a)\cdot u + a\,.$$

Using (13) and (14) we find

$$t_\eta(\alpha) = F^{-1}\left[F_\eta^{-1}(\alpha)\right] = (b-a)\cdot F_\eta^{-1}(\alpha) + a;$$

$$t_\nu(\alpha) = F^{-1}\left[F_\nu^{-1}(\alpha)\right] = (b-a)\cdot F_\nu^{-1}(\alpha) + a.$$

The acceleration time $\Delta t(\alpha)$ is

$$\Delta t(\alpha) = t_\eta(\alpha) - t_\nu(\alpha) = (b-a)\cdot F_\eta^{-1}(\alpha) + a - \left[(b-a)\cdot F_\nu^{-1}(\alpha) + a\right] =$$
$$= \left[F_\eta^{-1}(\alpha) - F_\nu^{-1}(\alpha)\right]\cdot(b-a).$$

The acceleration time is directly proportional to the width of the interval of the distribution.

The graphs of F, F_η and F_ν are shown at Figure 2.

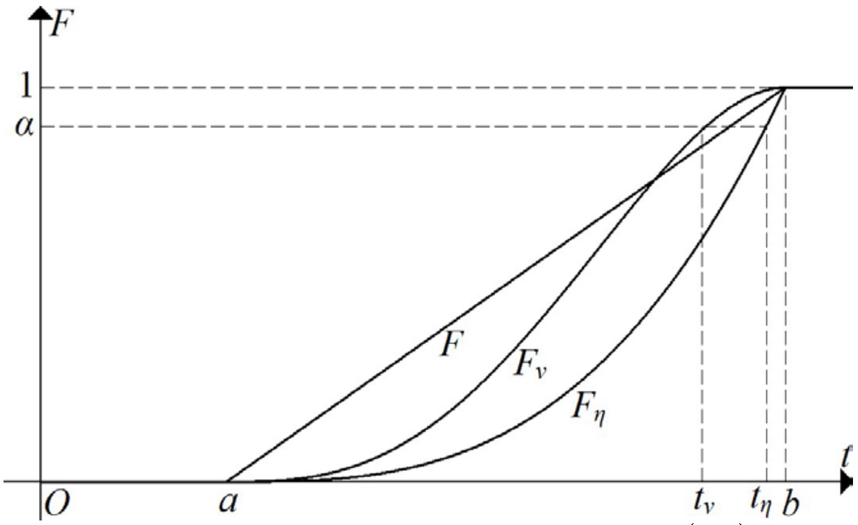

Fig. 2. Cumulative distribution functions: $F \sim U(a,b)$.

Example 2. Let all the actions have minimal distributed time a, but maximal execution time does not have a limit. Suppose that the probability that actions complete nearly to a is high. However, the probability that actions complete after a long period, also exists. So, it is reasonably to model all ξ-s by shifted exponential distribution $Exp(\lambda, a)$. The CDF is given by

$$F(t) = \begin{cases} 0 & t \le a \\ 1 - e^{-\lambda(t-a)} & a < t. \end{cases}$$

According to formulas (3) and (8), the CDF-s of η and ν are given by

$$F_\eta(t) = F(t)^3 = \begin{cases} 0 & t \le a \\ \left(1 - e^{-\lambda(t-a)}\right)^3 & a < t \end{cases}$$

and

$$F_\nu(t) = 4F(t)^3 - 3F(t)^4 = \begin{cases} 0 & t \le a \\ 4\left(1-e^{-\lambda(t-a)}\right)^3 - 3\left(1-e^{-\lambda(t-a)}\right)^4 & a < t. \end{cases}$$

If $F(t) = u$, then

$$F^{-1}(u) = -\frac{\ln(1-u)}{\lambda} + a.$$

Using (13) and (14) we find

$$t_\eta(\alpha) = F^{-1}\left[F_\eta^{-1}(\alpha)\right] = -\frac{\ln\left[1-F_\eta^{-1}(\alpha)\right]}{\lambda} + a;$$

$$t_\nu(\alpha) = F^{-1}\left[F_\nu^{-1}(\alpha)\right] = -\frac{\ln\left[1-F_\nu^{-1}(\alpha)\right]}{\lambda} + a.$$

The acceleration time $\Delta t(\alpha)$ is

$$\Delta t(\alpha) = t_\eta(\alpha) - t_\nu(\alpha) = -\frac{\ln\left[1-F_\eta^{-1}(\alpha)\right]}{\lambda} + a - \left[-\frac{\ln\left[1-F_\nu^{-1}(\alpha)\right]}{\lambda} + a\right] =$$

$$= \ln\left[\frac{1-F_\nu^{-1}(\alpha)}{1-F_\eta^{-1}(\alpha)}\right] \cdot \frac{1}{\lambda}.$$

The acceleration time is inversely proportional to the parameter λ of the distribution.

The graphs of F, F_η and F_ν are shown at Figure 3.

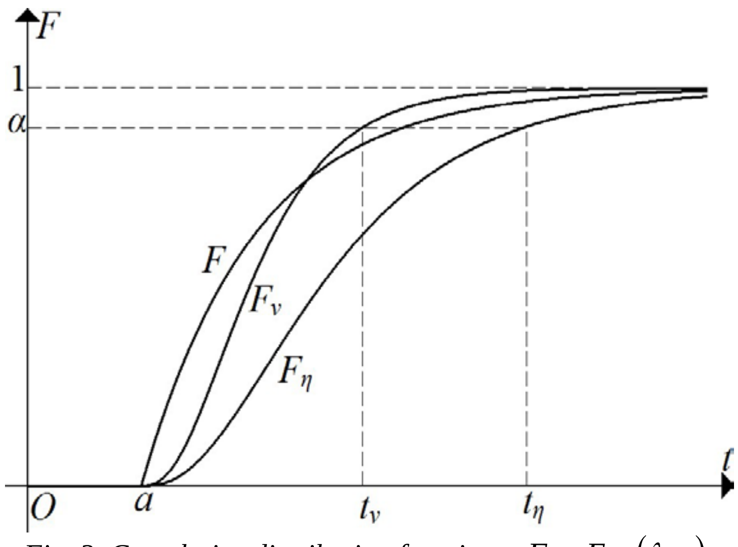

Fig. 3. Cumulative distribution functions: $F \sim Exp(\lambda, a)$.

The acceleration time Δt as a function of α is shown at Figure 4 for two cases. One of them is the case of uniformly distributed durations of activities and the other is the case of exponentially distributed durations of activities. Every case is explained by two examples with different parameters of distributions. The behavior of Δt depends on the distribution of random variables ξ_j.

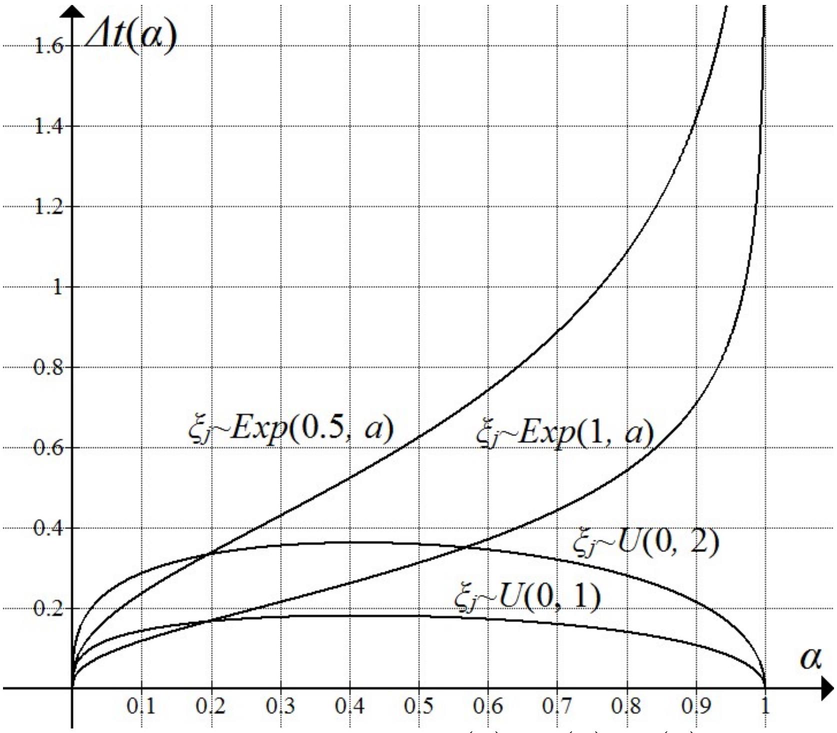

Fig. 4. Acceleration time $\Delta t(\alpha) = t_\eta(\alpha) - t_\nu(\alpha)$.

The acceleration time in both examples can be presented as a product of two multipliers. The first one is a function of inverse distribution functions $f\left(F_\eta^{-1}, F_\nu^{-1}\right)$. The second one is a function of parameters of the distribution of random variable ξ_j. This is the width of the interval of distribution when $\xi_j \sim U(a,b)$ or λ when $\xi_j \sim Exp(\lambda, a)$.

The functions F_η^{-1} and F_ν^{-1} are functions of α. Hence, $f\left(F_\eta^{-1}, F_\nu^{-1}\right)$ is a function of α only. Let it be noted that different types of functions can exist.

Some values of f and the way to evaluate the acceleration time are presented in Table 1.

Table 1. Acceleration time

α	$F_{\eta}^{-1}(\alpha)$	$F_{\nu}^{-1}(\alpha)$	$\Delta t(\alpha)$	
			$\left[F_{\eta}^{-1}(\alpha)-F_{\nu}^{-1}(\alpha)\right](b-a)$	$\ln\left[\frac{1-F_{\nu}^{-1}(\alpha)}{1-F_{\eta}^{-1}(\alpha)}\right]\cdot\lambda^{-1}$
1	2	3	4	5
0	0	0	0	0
0,10	0,46416	0,32046	0,14370(*b*–*a*)	0,23758·λ^{-1}
0,20	0,58480	0,41755	0,16726(*b*–*a*)	0,33850·λ^{-1}
0,30	0,66943	0,49160	0,17784(*b*–*a*)	0,43047·λ^{-1}
0,40	0,73681	0,55550	0,18131(*b*–*a*)	0,52406·λ^{-1}
0,50	0,79370	0,61427	0,17943(*b*–*a*)	0,62580·λ^{-1}
0,60	0,84343	0,67083	0,17260(*b*–*a*)	0,74308·λ^{-1}
0,70	0,88790	0,72762	0,16029(*b*–*a*)	0,88786·λ^{-1}
0,75	0,90856	0,75698	0,15158(*b*–*a*)	0,97747·λ^{-1}
0,80	0,92832	0,78768	0,14063(*b*–*a*)	1,08584·λ^{-1}
0,85	0,94727	0,82062	0,12665(*b*–*a*)	1,22431·λ^{-1}
0,90	0,96549	0,85744	0,10805(*b*–*a*)	1,41849·λ^{-1}
0,95	0,98305	0,90239	0,08066(*b*–*a*)	1,75058·λ^{-1}
0,96	0,98648	0,91337	0,07312(*b*–*a*)	1,85788·λ^{-1}
0,97	0,98990	0,92561	0,06429(*b*–*a*)	1,99659·λ^{-1}
0,98	0,99329	0,93986	0,05343(*b*–*a*)	2,19284·λ^{-1}
0,99	0,99666	0,95800	0,03865(*b*–*a*)	2,53032·λ^{-1}
1,00	1,00000	1,00000	0	∞

4. CONCLUSION

The collection of information is an important part of decision-making. The good decision is based on many and different sources. The gathering of information takes time and resources. But "even the right decision is wrong if it's made too late" [9]. So, the decision-maker is forced to work with incomplete data. This article is an attempt to show a way of finding a reasonable compromise. If incomplete information is allowed, the process can be accelerated. In addition, a method to evaluate the effect is explained.

Generally, the equation (10) is

$$F_{\nu}(t_{\nu}) = (n+1)\left[F(t_{\nu})\right]^{n} - n\left[F(t_{\nu})\right]^{n+1} = \alpha .$$

If $n+1>4$, the equation does not have exact solution. Then the inverse function should be represented numerically.

However, it has to be noted that if the distributions of durations are different, the mathematical formulas will be very complicated. In this case, it is better to use computer simulations.

REFERENCES

[1] Narlev, Y. Entrepreneurial Management in the Concept of the Organization's Project Cycle *Proc.of Sat Conf. "Management and Engineering 2013"*, Technical University of Sofia, vol. 2, 2013, pp. 796-804. (in Bulgarian)

[2] Skeriene, S., Augustiniene, A. The Theoretical Framework of Factors Influencing the Pedagogical Decision-Making. *Pedagogika*, 131 (vol. 3), 2018, pp. 5-25.

[3] Witteman, Cilia L.M. Concluding Commentary: Clinical Decision Making. *Clinical Psychological Science*, 6 (vol. 2), 2018, pp. 266-270.

[4] Hristozov, I. S. Application of information technologies in integrated environment for jointly work, *Proc. of Int'l Conf."Military technologies and systems for defence" (MT&S 2013)*, Defence Institute, Sofia, Bulgaria, 2013, pp. II-101–108. (in Bulgarian)

[5] Lambeva, M.H, I. S. Hristozov, A comparative analysis of existing information systems for command and control in NATO armies, *Proc of Conf "Knowledge management in scientific and educational organizations"*, Rakovski National Defence College, Sofia, Bulgaria, 2009, pp. 155-162. (in Bulgarian)

[6] Simon, H. A. *The New Science of Management Decision.* Harper & Brothers, New York, USA, 1960.

[7] Weed D. L. Weight of Evidence: A Review of Concept and Methods. *Risk Analysis,* 25 (vol. 6), 2005, pp. 1545-1557.

[8] Von Mises, R. *Mathematical Theory of Probability and Statistics. Academic Press Inc.*, New York, 1964.

[9] Iacocca, L., W. Novak. *Iacocca: An Autobiography.* Bantam Books, New York, 1984.

Information about author:

Deyan Mihaylov – PhD in informatics, chief assistant-professor at University of Economics – Varna, Statistics and Applied Mathematics Department. His area of interest is operations research.

Manuscript received on 15 August 2019