16

Mathematical modeling of the Internet survey

Getka-Wilczyńska Elżbieta
Warsaw School of Economics
Poland

1. Introduction

Recently statistical research as for as sampling selection can be divided into representative surveys based on the probability sample and surveys based on the non-probability sample, e.g. the Internet surveys. After choosing the kind of the sample selection next stage of the survey is data gathering. In all surveys, the data is collected using an immediate interview, a telephone interview or a post and in recent years an interview over the Internet (Internet, 2001, Dillman, 2000).
Surveys over the Internet (Internet mediated research: online surveys, Internet surveys, Web surveys (Vehovar, 2007)) are in a process of intensive development and key of characteristic of it is their diversity. Collecting data through the Internet surveys is useful either for marketing and other private research societies either statistical agencies. The first graphic browser (NCSA Mosaic) was released in 1992, with Netscape Navigator following in 1994 and Internet explore in 1995. The first publishes papers on Web surveys appeared in 1996. Since then, there has been a virtual increasing tendency of interest in the Internet generally, and World Wide Web specifically, as a tool of data collection (www.WebSM.org). A special portal WebSM – Web survey methodology web site is a website dedicated to the methodology of Web surveys. It supported by EU since 2002 and it includes bibliography lists and software database.
Generally, difference between these researches rely on following aspects. In representative surveys based on the probability sample the frame of sampling is known, respondents are drawn to the sample by a statistician in accordance with sampling design (sampling scheme) and the methods of theory sampling are applied to data analysis (Bracha, 1996; Kordos, 1982; Tille, 2006; Särndal et al., 1992; Steczkowski, 1988; Wywiał, 1992). If in these surveys an electronic questionnaire is used, it is only one of modes of data collection and then the correct use of this data collection tool requires the suitable survey methodology (Biffignandi & Pratesi, 2000).
The Internet survey has several advantages, such as low costs of collecting information, the speed of the data transmission and a possibility to monitor it. Moreover, the computerized nature of Web surveys facilitates conducting experiments. The usage of the electronic questionnaire in the Internet survey makes the interview more efficient, lowers the workload of the respondents and controls the responds' quality.
But the basic problem in the surveys over the Internet is concerned with collecting data sets according to classical methods of the sampling theory. In the Internet surveys drawing the

sample is not possible and respondents are not randomly selected to the sample, but they participate in the survey with a subjective decision and they form an uncontrolled sample (the Internet sample). The methods of the sampling theory can not be used immediately for the data from such the sample because the probability inclusions are not known and statistics are calculated usually on the basis Internet data refer usually to the population surveyed (Getka-Wilczyńska, 2003). In theory and practice of Internet survey two approaches to dealing with this problem are identified (Couper & Miller, 2008). The first - the design based approach, attempts to build probability – based Internet panels by using other methods for sampling and recruitment and – if it is necessary- providing Internet access to those without. This approach is applied e.g. by Knowledge Networks in the USA and CentERdata's MESS panel in the Netherlands. The second – the model based approach begins with a volunteer or opt - in panel of Internet users, and attempts to correct for representation biases using e.g. propensity score adjustment (Lee, 2006) or some other weighting method for assessing Web panel quality (Callegaro & Disogra, 2008, Toeppoel et.al., 2008). In both approaches usually are used methodology of sampling theory to data analysis. Other an interesting proposition is an application of a dynamic theory of decision making and the decision field theory to theoretical explanation of survey behavior (Galesic, 2006).
In this study are proposed certain conceptions of modelling of the Internet survey as a random experiment or a life testing experiment by using notions and methods of the stochastic processes and the reliability theory (Kingman, 2002; Kopociński, 1973; Barlow & Proschan, 1965 (1996), 1975; Sołowiew, 1983). Generally this approach is presented in following way.
At the first, the process of the Internet data collection is considered as a process of registering questionnaires on the server at fixed interval of the time, the time of the survey conducted. An events appear in the Internet survey are interpreted as the moment of an arrival, a birth, a death of the population elements or a waiting time for these events. In this case the random size of the uncontrolled sample (Internet sample, a random set of the moments in which questionnaires - from respondents who participate in the Internet survey - are recorded on the server) is defined as a counting process by using Poisson processes (Getka-Wilczyńska, 2004 (in Polish), 2005, 2008).
At the second, the Internet survey is considered as a life test of the population surveyed by using the notions and methods of the reliability theory, (Getka-Wilczyńska, 2007). In this case the events which appear in the Internet survey are interpreted as the moment of failure, renew or functioning time of the population elements, when the population is treated as a coherent system of finite number of elements or as the lifetime of the population elements, when the length of the population lifetime is considered. The length of the population lifetime is defined by using the structure function of the population and the reliability function of the length of the population lifetime is defined for the series, parallel and partial structure of the population. Then the basic characteristics of the reliability function are described, calculated and estimated by using path and cut method (Barlow & Proschan, 1965).
More interesting is the life testing experiment, if different models of elements' dependence are considered, e.g. dependence on initial parameters of the population (the system), external conditions, as well as on the states of other elements. In this study is described a general model of functioning of the population (the system), when exist the dependence of the reliability of one element on the states of the other elements, (Sołowiew, 1983). Then the changes in time of the population states (the system states) are determined by multidimensional stochastic process which is chosen in such a way that the state of the population can be

explicitly defined in each moment (whether the population is in the state of life or in the state of death). As an example are considered the particularly cases of stochastic processes - a general death process (when the rate of the state change of the element depend only on the states of the elements) and a pure death process (when in the general death process the elements of the population are symmetrical), (Sołowiew, 1983). In an aspect of the Internet survey three characteristics are important: the time until the first change of state of the population, the length of the lifetime of the population per one change of the state and the residual time of the population lifetime.

2. Assumptions

We generally assume, that Internet survey begins at the moment $t = 0$, when the electronic questionnaire is put on the website and the survey is conducted for the time $T > 0$.
A set $\{u_1, u_2, ...,\}$ denotes the population of potential respondents. For each $n \geq 1$, the population of n units is surveyed and the respondent sent the questionnaires independently.
By X_j, $j = 1,2,..., n$, $n \geq 1$, we denote a moment of questionnaire record on the server after an initial moment $t = 0$, from each respondent $u_k, t = 1,2,..., n$, $n \geq 1$, belonging to the population of the size $n \geq 1$, who took part in the survey as j- th.
The moment of the questionnaire record is an event that can be interpreted

- as the moment of arrival or the respondent, who took part in the survey as j-th, when the size of the uncontrolled sample is defined,
- as the moment of failure or renew of j-th element of the population, when the population is treated as a coherent system,
- as the waiting time for the j-th questionnaire record after the initial moment $t = 0$ equal the length of lifetime of j-th element of the population until the moment $t \geq 0$
- the moment of death or birth of the j-th element of the population, when the length of the population lifetime is considered.

Theoretically, four cases which describe the relation between the time of the survey conducting , $T > 0$, and the size of the uncontrolled sample can be considered.
In the first case, the registering the questionnaires ends at the moment $T > 0$ specified in advance, independently of the questionnaires' number recorded. The size of the uncontrolled sample is then a random value in the interval $[0, T]$ and depends on the length of time of the survey and on a selection procedure applied in the survey (if it is used in the survey). An extreme situation occurs when no data was collected (an arrival set is empty or the questionnaire, which arrived were rejected by the selection procedure used in the survey).
In the second case the sample size is specified in advance and the survey ends when the assumed number of responses has arrived, independently of the length of time of the survey (a random value in this case). An extreme situation occurs when the length of time of the survey is infinite.

In the third case, both the length of time of the survey and the sample size are specified in advance and the survey ends in earlier of the assumed moments.
In the fourth case, the final moment of the survey is not specified in advance. The process of registering questionnaires lasts at the moment when the collected data set meets the demands of the survey organizers.

3. The random size of an uncontrolled sample

If the process of Internet data collection is considered as a process of registering questionnaires on the server in a fixed interval of the time $T > 0$ (the time of the survey conducting) then the size of the uncontrolled sample at the moment $t \geq 0$ equals the total number arrivals until the moment $t \geq 0$ and is defined as a counting process - Bernoulli, Poisson or compound Poisson process (Kingman, 2001).
In this part of the study paper we describe the two cases, the first and the second of the dependence between the time of the survey conducting and the size of the Internet sample (Getka-Wilczyńska, 2008).

3.1 The size of the uncontrolled sample as Bernoulli process

Definition 2.1. For fixed $n \geq 1$ the size of uncontrolled sample until the moment $t \geq 0$ is given by

$$N(t) = \text{card}\,\{1 \leq k \leq n : X_k \in [0, t]\}$$

and is equal to a sum

$$N(t) = N_1(t) + \ldots + N_n(t),$$

where

$$N_k(t) = \begin{cases} 0 & \text{if} \quad \chi_k > t \\ 1 & \text{if} \quad \chi_k \leq t \end{cases}, \; \chi_k,\, k = 1, 2, \ldots, n,\; n \geq 1,$$

are independent random variables with uniform distribution over $[0, T]$, $T > 0$, $\chi_{k-1} \leq \chi_k$ for $j \geq 1$, $\chi_0 = 0$ and at the initial moment $t = 0$ no arrivals occur.
The value of the random variable $N(t), t \geq 0$ equals the total number of arrivals until the moment $t \geq 0$ and the process $\{N(t), t \geq 0\}$ can be described in a following way. Each of n respondents, independently of others send only one questionnaire with the probability 1 in the interval $[0, T]$ for $T > 0$ (the time of the survey conducted). The probability of sending the questionnaire by the certain respondent in the interval of the length $\Delta \subset [0, T]$ is equal to ratio $\frac{|\Delta|}{T}$.
In this way, each respondent generates a stream consisted of only one arrival. A summary stream obtained by summing these streams is called a bound Bernoulli stream, that is, it consists of finite number of events.
To complete the definition of the counting process it remains to compute the distribution of $N(t)$ and the joint distribution of $N(t_1), N(t_2), \ldots, N(t_n)$ for any non-negative $t_0, t_1, \ldots, t_n$.

Let

$$P_k(t) = P\{N(t) = k\},\ k = 1,\dots,n$$

be the probability of event, that at the moment $t \geq 0$ the total number of arrivals $N(t)$ equals k.

Since the probability of arrival of the given respondent in the interval $[0,t] \subset [0,T]$ is equal to ratio $\frac{t}{T}$ and the arrivals came independently, hence the total number of arrivals $N(t)$ at the moment $t \geq 0$ is random variable with Bernoulli distribution

$$P_k(t) = \binom{n}{k}\left(\frac{t}{k}\right)^k \left(1 - \frac{t}{k}\right)^{n-k}$$

If the intervals $\Delta_1,\dots,\Delta_n$ are disjoint pairs and the interval $[0,T] = \Delta_1 \cup \dots \cup \Delta_n$ is a sum of $\Delta_1,\dots,\Delta_n$, then for any non-negative integers $k_1,\dots,k_n$ such that $k_1 + \dots + k_n = n$ holds

$$P\{N(\Delta_1) = k_1,\dots,N(\Delta_n) = k_n\} = \frac{n!}{k_1!\dots k_n!} p_1^{k_1}\dots p_n^{k_n},$$

where $N(\Delta_i)$ is the number of arrivals which occur in the interval Δ_i, $p_i = \frac{|\Delta_i|}{T}$ for $i = 1,\dots,n$, and $|\Delta_i|$ is the length of the interval $\Delta_i = t_i - t_{i-1}$, $i = 1,\dots,n$.

3.2 The size of the uncontrolled sample as Poisson process

Definition 2.2. For fixed $n \geq 1$ the size of uncontrolled sample until the moment $t \geq 0$ is given by

$$N'(t) = \text{card}\,\{n : X_n \in [0,t]\} = \max\{n \geq 0 : S_n \leq t\},$$

where $X_1, X_2, \dots$, denote as before the successive moments of questionnaires record, $X_{k-1} \leq X_k$ for $k \geq 1$ and $X_0 = 0$,

$(Y_k)_{k=1}^{\infty}$ is a sequence of independent and identically distributed random variables $Y_k = X_k - X_{k-1}$ with exponential distribution

$$G(t) = 1 - e^{-\lambda t},\ t \geq 0,\ \lambda > 0 \text{ and } Y_k = X_k - X_{k-1} \text{ for } k \geq 1$$

denotes k – th spacing between k – th and $(k-1)$ – th arrivals, $S_n = \sum_{k=1}^{n} Y_k$ is a random variable with Erlang distribution given by

$$P(S_n \leq t) = 1 - \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t} \text{ for } t \geq 0 \text{ and } \lambda > 0,$$

Then

$$P\{N'(t)=k\}=P\{N'(t)\le n+1\}-P\{N'(t)\le n\}=P\{S_{n+1}>t\}-P\{S_n>t\}=\frac{(\lambda t)^n}{n!}e^{-\lambda t}\ ,\ t\ge 0\ ,$$

the total number of arrivals until the moment $t \ge 0$, is a random variable with Poisson distribution with the parameter $\lambda > 0$ and $\{N'(t): t \ge 0\}$ is Poisson process.

Moreover, if in the Poisson process (stream) in the interval $[0,T]$, $T>0$, n arrivals occur, then process (stream) of arrivals in this interval is the Bernoulli process (stream), (Kingman, 2001). This fact is shown below.

If $0 \le t \le T$ and $0 \le k \le n$, then

$$P\{N'(t)=k|N'(T)=n\}=\frac{P\{N'(t)=k, N'(T)-N'(t)=n-k\}}{P\{N'(t)=n\}}=\frac{P\{N'(t)=k\}P\{N'(T-t)=n-k\}}{P\{N'(T)=n\}}=$$

$$=\frac{\frac{(\lambda t)^k}{k!}e^{-\lambda t}\frac{(\lambda(T-t))^{n-k}}{(n-k)!}e^{-\lambda(T-t)}}{\frac{(\lambda T)^n}{n!}e^{-\lambda(T-t)}}.$$

Hence

$$P\{N'(t)=k|N'(T)=n\}=\binom{n}{k}\left(\frac{t}{T}\right)^k\left(1-\frac{t}{T}\right)^{n-k}.$$

If the intervals $\Delta_1,...,\Delta_n$ are disjoint pairs and $[0,T]=\Delta_1\cup...\cup\Delta_n$, then for any nonnegative integers $k_1,...,k_n$ such that $k_1+...+k_n=n$ holds

$$P\{N'(\Delta_1)=k_1,...,N'(\Delta_n)=k_n|N'(T)=n\}=\frac{P_k\{N'(\Delta_1)=k_1,...,N'(\Delta_n)=k_n\}}{P\{N'(T)=n\}}=\frac{\prod_{i=1}^{n}P\{N'(\Delta_i)=k_i\}}{P\{N'(T)=n\}}=$$

$$=\frac{\prod_{i=1}^{n}(\lambda|\Delta_i|)^{k_i}e^{-\lambda|\Delta_i|}}{\frac{(\lambda T)^n}{n!}e^{-\lambda T}}.$$

Therefore

$$P\{N'(\Delta_1)=k_1,...,N'(\Delta_n)=k_n|N'(T)=n\}=\frac{n!}{k_1!...k_n!}p_1^{k_1}...p_n^{k_n}.$$

3.3 The size of the uncontrolled sample with a selection procedure as compound Poisson process

In the Internet surveys the electronic questionnaire is available to all Internet users and a part of the registered arrivals came from respondents who do not necessarily belong to the surveyed population.

In this case only the arrivals of these respondents whose questionnaires qualified for the data set based on the selection procedure are included in the sample. By this assumption and the assumptions made in case 2 the size of the uncontrolled sample is defined as a compound Poisson process (Kingman 2001).

Definition 3.3. For fixed $n \ge 1$ the size of uncontrolled sample until the moment $t \ge 0$ is given by $Y(t) = S_{N'(t)}$,

where

$$N'(t) = \text{card}\{n: X_n \in [0,t]\} = \max\{n \ge 0 : S_n \le t\},\ S_{N'(t)} = \sum_{j=1}^{N'(t)} U_j,$$

a sequence $(U_n)_{n=1}^{\infty}$ of independent and identically distributed random variables and the Poisson process $\{N'(t): t \ge 0\}$ are independent.

The arrivals $\chi_n, n = 1,2...$, are selected for the uncontrolled sample in the following way (the sequence of arrivals $\chi_n, n = 1,2...$ is thinned): the arrival $\tau_n, n = 1,2...$., is omitted with the probability p, $p \in [0,1]$ (independently of the process taking place), if the respondent does not belong to the population and the arrival $\chi_n, n = 1,2...$., is left with the probability $1 - p$, otherwise.

The random variable U_i is equal to 1, if the arrival χ_i remains, and 0, if the arrival χ_i is omitted.

The probability p, $p \in [0,1]$ is defined by the procedure of selection used in the survey and consequently, process $\{Y(t), t \ge 0\}$ is compound Poisson process with expected number of the arrivals $\lambda(1-p)$.

4. Length of the population lifetime

In the remaining part of this study we assume that the population of n units for $n \ge 1$ is treated as a finite coherent system of n components (Barlow & Proschan, 1965) and the Internet survey begins at time $t = 0$ and it is conducted for the time $T, T > 0$. In this case the process of Internet data collection can be considered as a random experiment or a life testing experiment in which the basic characteristics of length of the population lifetime are analysed by using the methods of reliability theory (Barlow & Proschan, 1965 (1996), 1975; Kopociński, 1973, Sołowiew, 1983).

We assume that a non-negative independent random variables $\chi_k, k = 1,2..., n, n \ge 1$ with distribution function

$$F_k(t) = P(\chi_k \le t),\ \text{for}\ t \ge 0,\ k = 1,2,...,n,\ n \ge 1,$$

and probability density function

$$f_k(t) = F_k'(t) \quad \text{and} \quad F_k(t) = \int_0^t f_k(x)dx$$

are interpreted as the length of lifetime of k-th element of the population until the moment $t \ge 0$ or the moment of death of k-th element of the population until the moment $t \ge 0$ or the waiting time of arrival of k-th element of the population until the moment $t \ge 0$.

The probability

$$\overline{F}_k(t) = 1 - F_k(t)\ \text{for}\ t \ge 0,\ k = 1,2,...,n,\ n \ge 1$$

is the reliability function of the length of lifetime of k-th element at the moment $t \geq 0$, (the reliability of the k-th element in short) and equal to the probability of the length of lifetime of k-th element at least $t \geq 0$ or the probability of event that k-th respondent is in the state of life at least $t \geq 0$.

The conditional probability density function

$$\lambda_k(t) = \frac{f_k(t)}{\bar{F}_k(t)} \text{ for } t \geq 0, \ k = 1,2,\ldots,n, \ n \geq 1$$

is called arrival or failure rate of k-th element of the population, (Kopociński, 1973).

The elements of the population are not renewed - each record of questionnaires decreases the size of the population in one and the element which arrived is not replaced by a new one. This way of selection is called random sampling without replacement.

The length of the population lifetime and is defined by using the structure function of the population as follows (Sołowiew, 1983).

4.1 States of elements of the population

The state of i-th elements of the population (as the system) is defined by the values of the binary function

$$Y_i(t) = \begin{cases} 0 & \text{if } i\text{-th element is in the state of life or } i\text{-th element did not arrive until the moment } t \\ 1 & \text{if } i\text{-th element is in the state of death or } i\text{-th element arrive until the moment } t \end{cases}$$

where $Y : [0,\infty) \times \{1,2,\ldots,n\} \rightarrow \{0,1\}$ and $Y(t,i) = Y_i(t)$ for $i \in \{1,\ldots,n\}$, $t \in [0,\infty)$.

Then the state of all elements of the population of size n, for $n \geq 1$, is determined by n-dimension vector $\underline{Y}(t) = (Y_1(t), Y_2(t),\ldots, Y_n(t))^T$ and we assume that at the initial moment $t = 0$ all elements of the population of size n, for $n \geq 1$, are in the states of life. This assumption means that at the moment $t = 0$ no arrivals occurred.

4.2 States of the population

The state of the population of the size n, for $n \geq 1$, at the moment $t \geq 0$ is defined by the values of the binary function

$$\varphi_0(t) = \begin{cases} 0 & \text{if the population is in state of life (the survey (test) is conducted) at the moment } t \\ 1 & \text{if the popualtion is in the state of death (the survey (test) ended) until the moment } t \end{cases}$$

and at each moment $t \geq 0$ it depends on the states of the elements through the values of the function $\varphi_0(t) = \varphi(Y_1(t),\ldots, Y_n(t)) = \varphi(\underline{Y}(t))$.

In the process of Internet data collection treated as a life test of the population of size n, for $n \geq 1$, the population can be found at the moment $t \geq 0$ in the state of life during the conducting of the survey in following cases.

In the first case, at the moment $0 \le t \le T$ where T is the time of the survey conducting specified in advance and the number of death (arrivals) in interval of the length $t \ge 0$ is a random value but it is less than the size of the population. Otherwise, until the moment T specified in advance.
In the second case, until the moment $t \ge 0$, in which the number of death (arrivals) is equal to the size of the sample specified in advance (it is equals or less than the size of the population) and the time of the survey conducting T is not specified in advance.
In the third case, until the earlier of the time of the survey conducting T and the moment $t \ge 0$, when the number of death (arrivals) is equal to the sample size, where both the length of time of the survey T and the sample size are specified in advance.
In the fourth case, until the moment $t \ge 0$, when the collected data set (it can be a subset of the population or the population surveyed) meets the demands of the survey organizers and the final moment of the survey is not specified in advance.

4.3 Properties of the structure function

The structure function $\varphi(\underline{Y})$ is increasing, if for any two vectors $\underline{Y}^{(1)}$ and $\underline{Y}^{(2)}$ is satisfied the condition:

$$\text{if } \underline{Y}^{(1)} \le \underline{Y}^{(2)}, \text{ then } \varphi\left(\underline{Y}^{(1)}\right) \le \varphi\left(\underline{Y}^{(2)}\right),$$

where $\underline{Y}^{(1)} \le \underline{Y}^{(2)}$, if for all $i = 1,...,n$, $Y_i^{(1)} \le Y_i^{(2)}$.
This property of the structure function introduce a partial order in a set of the binary vectors and means that additional death of the element can not change the state of the population from the state of death to the state of life.
The function $\varphi(Y(t))$ define a division of a set $E = \left\{\underline{Y}: \ \underline{Y}: Y^n \to \{0,1\}^n\right\}$ of all n - dimension and binary vectors which describe the state of the population to two sets:

$$E_+ = \{\underline{Y} : \varphi(\underline{Y}(t)) = 0\}, \text{ a set of states of life of the population and}$$
$$E_- = \{\underline{Y} : \varphi(\underline{Y}(t)) = 1\}, \text{ a set of states of death of the population.}$$

If the structure function is increasing, then the division of the set E to two sets E_+ and E_- is called a monotonic structure (Barlow & Proschan, 1965).

4.4 Length of the population lifetime

Let us denote by X the length of the population lifetime and

$$X = \inf\{t : \varphi(\underline{Y}(t)) = 1\}.$$

Then

$$F(t) = P(X \le t)$$

is the probability of ending of the survey (test) until the moment $t \ge 0$ or the probability of the event that the population is in the state of death until the moment $t \ge 0$
and

$$\overline{F}(t) = P(X > t)$$

is the probability of the conducting survey (test) at least $t \geq 0$, the probability of the event that the population is in the state of life at least $t \geq 0$ or the reliability function of the X, the length of the population lifetime at the moment $t \geq 0$, (the reliability of the population in short).

4.5 Calculation of the reliability of the length of the population lifetime

The formula which expresses the relation between X, the reliability of the length of the population lifetime and X_k, $k = 1,2...,n$, $n \geq 1$ reliabilities of elements at the moment $t \geq 0$ is given by

$$\overline{F}(t) = \sum_{\underline{Y}(t) \in E_+} p(\underline{Y}(t)),$$

where

$$p(\underline{Y}(t)) = \prod_{k=1}^{n} \overline{F}_k^{1-Y_k}(t) F_k^{Y_k}(t),$$

(there is adopted the convention $0^0 = 1$) is a probability of event that the population is in the state $\underline{Y}$.

If $X_k, k = 1,2...,n, n \geq 1$ are non-negative independent random variables, the elements are not renewed and the function $\varphi(e)$ is increasing, then the reliability function of the length of the population lifetime $\overline{F}(t)$ is increasing respectively to each coordinate of the reliability function of the length of the element lifetime $\overline{F}_k(t)$.
Thus an upper or a lower bound on the reliability of X, the length of the population lifetime may be obtained from the upper or lower bounds on the reliabilities of the elements.
When the number of the states is large (the number of all states is equal to 2^n) and the function $\varphi(\underline{Y})$ is very complicated, then a formulae given above is not efficient and the other methods of calculation are applied e.g. the method of path and cut (Barlow & Proschan, 1975; Koutras et al.,2003) or the recurrence method of Markov chain or generally, the Markov methods (Sołowiew, 1983).

5. Basic structures of the population

5.1 Length of the population lifetime for the series structure

The population (as the system) of n - elements for $n \geq 1$ is called a series structure, when the population is in the state of life if and only if each element is in the state of life.
In this case, the change of the state of any element causes the change of the population state. The length of the population lifetime is equal to the waiting time of the first death and the size of the uncontrolled sample equals zero for the first death. Then the basic characteristics of the reliability function of the series structure are given as follows.

From definition of the series structure follows that the length of the population lifetime is equal

$$X_{1,n} = \min(X_1, X_2, ..., X_n)$$

and the probability of it (duration of the survey) at least $t \geq 0$ is equal to

$$\overline{F}(t) = \prod_{i=1}^{n} \overline{F}_i(t).$$

From inequality (Hardy, et. al, 1934)

$$1 - \sum_{i=1}^{n} F_i \leq \prod_{i=1}^{n}(1 - F_i) \leq 1 - \sum_{i=1}^{n} F_i + \sum_{i<j} F_i F_j \leq 1 - \sum_{i=1}^{n} F_i + \frac{1}{2}\left(\sum_{i=1}^{n} F_i\right)^2$$

it follows that

$$\left|F(t) - \sum_{i=1}^{n} \overline{F}_i\right| \leq \frac{1}{2}\left(\sum_{i=1}^{n} F_i\right)^2.$$

The change rate of the population equals the sum of the change rate of the elements

$$\lambda(t) = -\frac{\overline{F}'(t)}{\overline{F}(t)} = -\left[\ln \overline{F}(t)\right]' = -\sum_{i=1}^{n}\left[\ln \overline{F}_i(t)\right] = -\sum_{i=1}^{n} \frac{\overline{F}_i{}'(t)}{\overline{F}_i(t)} = \sum_{i=1}^{n} \lambda_i(t).$$

The expected time of the length of the population lifetime is equal to

$$E(X) = \int_0^{\infty} \overline{F}(t) dt.$$

5.2 Length of the population lifetime for the parallel structure

The population (as the system) of n - elements for $n \geq 1$ is called a parallel structure, when the population is in the state of death if and only if all elements are in the state of death.
In this case, the change of the state of the population (death of the population) takes places only if changes of all population elements occur - all elements of the population died and the size of the uncontrolled sample is equal to the size of the population (all elements of the populations arrived).
From definition of the parallel structure follows that the length of the population lifetime is equal to

$$X_{n,n} = \max(X_1, X_2, ..., X_n)$$

and the probability of the length of the population lifetime (duration of the survey) at least $t \geq 0$ is equal to

$$F(t) = \prod_{i=1}^{n} F_i(t) \text{ or } F(t) = F_0^n(t)$$

and the expected time of the length of the population lifetime is equal to

$$E(X) = \int_0^{\infty} [1 - F(t)] dt \text{ or } E(X) = \int_0^{\infty} [1 - F_0^n(t)] dt \text{, when } F_i(t) = F_0(t), i = 1,2,...,n.$$

5.3 Length of the population lifetime for the partial structure

The population (as the system) of n - elements for $n \geq 1$ is called a partial structure if all elements of the population are identical and the population is in the state of life if at least m elements of the population are in the state of life (that is, at most $n-m$ death occur) and the size of the uncontrolled sample equals $n-m$.
The reliability function of the length of the population lifetime is equal to

$$\overline{F}(t) = 1 - P(X_{n-m+1,n} \leq t) = \sum_{k=m}^{n} \binom{n}{k} \overline{F}_0^k(t) F_0^{n-k}(t).$$

6. Estimate of the reliability of the length of the population lifetime - the method of path and cut

In this method are defined notions of minimal path minimal cut that used to estimate the reliability function of the length of the population lifetime, (Barlow & Proschan, 1965).

Definition 6.1. The set of elements $A = \{u_1, ..., u_k\}$ of the population of the size n, $n \geq 1$, is called a minimal path if all the elements of this set are in the state of life (the population is in the state of life, the survey is being conducted) and no subset of the set A has this property.

From the monotonic property of the structure function, the set $A = \{u_1, ..., u_k\}$ is a minimal path if and only if $\underline{Y} \in E_+$, where e is a vector in which coordinates $i_1, ..., i_k$ take on the value zero and the remaining coordinates take on the value one, with any state greater than $\underline{Y}$ belonging to E_-.
Therefore, every minimal path determines a bordering state of the life of the population in which the occur of death of any element causes a change of the state of the population into the state of death one (ending of the survey (test)).
In term of the size of the uncontrolled sample, it means, that the number of arrivals is equal to the number of elements of the minimal path is smaller per one than the size of the sample assumed in the survey.
Let $\{A_1, A_2, ..., A_m\}$ be a sets of all minimal paths with the corresponding bordering states $\underline{Y}^{(1)}, \underline{Y}^{(2)}, ..., \underline{Y}^{(m)}$.
A_s is an event in which all elements of the minimal path A_s are in the state of life. Since $E_+ = \bigcup_{s=1}^{m} A_s$, (Sołowiew, 1983), the reliability function of the length of the population lifetime is calculated from the formula

$$\overline{F}(t) = P\left(\bigcup_{s=1}^{m} A_s\right) = \sum_{i=1}^{m} P(A_i) + \sum_{i,j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) + ... + (-1)^{m+1} P(A_1 A_2 ... A_m)$$

The number of the elements of the sum on the right is equal to $2^m - 1$ and the probability of any event which is given by $A_{i_1} A_{i_2} ... A_{i_k}$, where $i_1 < i_2, ... < i_k$ is equal to

$$P\left(A_{i_1}A_{i_2}\dots A_{i_k}\right)=\overline{F}_{s_1}(t)\overline{F}_{s_2}(t)\dots\overline{F}_{s_l}(t),$$

where s_1, s_2,. s_l are different indices of elements of the minimal paths (that is, in the case of the elements belonging to overlapping parts of different paths, each element is calculated only once).

In the order to lower the number of calculation are introduced the following notions.

The two minimal paths are called crossing, when they have at least one common element.

The two minimal paths are called relevant, if there exists a chain of crossing paths which connects them.

The relevant relation is the equivalent relation which divides a set of all minimal paths into classes of relevant minimal paths.

Let $\{A_1\dots A_{k_1}\},\{A_{k_1+1}\dots A_{k_2}\}\dots$ be the successive classes of the relevant minimal paths.

Because

$$F(t)=1-\overline{F}(t)=P\left(\prod_{i=1}^{m}\overline{A}_i\right),$$

(the symbol $\prod_{i=1}^{m}\overline{A}_i$ means an intersection of sets $\overline{A}_i$ for $i=1,\dots,m$) and the events belonging to different classes are independent, then

$$F(t)=P\left(\prod_{i=1}^{k_1}\overline{A}_i\right)P\left(\prod_{i=1}^{k_2}\overline{A}_{k_1+1}\right)\dots\dots .$$

A dual notion of the minimal path is a minimal cut (a critical set).

Definition 6.2. A set of elements $B=\left(j_1,j_2,\dots,j_l\right)$ is called a minimal cut, if all elements of this set are in the state of death (the population is in the state of death, the survey (test) ended) and no subset of the set B has this property.

In this case we are interesting in those cut set in which the number of elements is equal to the size of the uncontrolled sample specified in advance (all elements belonging to the set B arrived).

If $\{B_1,B_2,\dots,B_s\}$ is a set of all the minimal cuts, then the probability of an event that the survey (test) ends until the moment $t\geq 0$ is equal to

$$F(t)=P\left(\bigcup_{i=1}^{s}B_i\right)=\sum_{i=1}^{s}P(B_i)+\sum_{i<j}P\left(B_iB_j\right)+\sum_{i<j<k}P\left(B_iB_jB_k\right)+\dots+\;+(-1)^{m+1}P(B_1B_2\dots B_m)=$$
$$=S_1-S_2+S_3+\dots+(-1)^{s+1}S_s$$

From this formula the estimation of the length of the population lifetime (during the survey) as the estimation of the reliability function of the population can be obtained.

In this case the number of elements in the successive minimal cuts is interpreted as the possible sample sizes which can be collected in the survey on condition that the survey ends after collection of the sample of the assumed size.

From proof of this formula for the non-crossing minimal cuts holds

$$S_2 \le \frac{S_1^2}{2}, \quad S_3 \le \frac{S_1^3}{6},$$

and so on, and in the case of the crossing minimal cuts the partial sums maintain an order $S_k = O\left(S_1^2\right)$.

Moreover, for any k,

$$S_1 - S_2 + \ldots - S_{2k} \le F(t) \le S_1 - S_2 + \ldots - S_{2k} + S_{2k+1}.$$

Thus there exists the possibility of the estimation of the probability of length of the population lifetime with assumed precision because partial sums on the right of the last formula are the interchangeable upper and the lower bounds of the reliability function.

7. Reliability of the length of the population lifetime – Markov methods

So far we have assumed that the length of the lifetime of the population elements (the waiting time for arrival) are independent distributed random variables. In literature devoted to research of reliability of the system of order n considered dependence on initial parameters of the system, external conditions, as well as on the states of other elements. Let us consider the last dependence.

7.1 Model of dependence of the element of the population

In the most general way, if no dependence of initial parameters exists and there is no dependence of the lifetimes of the elements on the common external conditions, then the only type of dependence is the dependence of the reliability of one element on the states of the other elements. This dependence is described as follows (Sołowiew, 1983).
Let the states of the population of n - elements for $n \ge 1$ are defined by the binary vector

$$\underline{Y}(t) = (Y_1(t), Y_2(t), \ldots, Y_n(t))^T$$

and by Γ_t is denoted a realisation of the process $\underline{Y}(x)$ in interval $[0, t]$.
If the realisation Γ_t of the process $\underline{Y}(x)$ is determined until the moment t, then the transition probability in the time Δt to the state $\underline{Y}$ equals

$$\lambda(\Gamma_t, \underline{Y})\Delta t + o(\Delta t) \tag{7.1.1}$$

There are can be considered the following cases:
1) at the same time the change of the states of a few elements occur,
2) in each moment the change of the state of only one element occurs (failure, arrival).
If case 2 is considered, then the process $\underline{Y}(t)$ is described by the arrival (failure) rates of the elements and $\lambda_i(\Gamma_t)$ is the conditional arrival (failure) rate of the i-th element, on condition that the realisation of the process is determined until the moment t.
If the arrival (failure) rate of the element given by formulae $\lambda_i(\Gamma_t) = \lambda_i[t, \underline{Y}(t)]$, depends only on the moment t and the state of the process in this moment then the process $\underline{Y}(t)$ is non-homogeneous Markov process with a finite number of states.

If the arrival (failure) rate of the element does not depends on time, but depends only on the states of the elements $\lambda_i(t,\underline{Y}) = \lambda_i(\underline{Y})$, then such a homogeneous Markov process is a generalised pure death process.
If in the generalised pure death process the elements of the population are symmetrical, then all arrival (failure) rates of elements are equal to $\lambda_i(\underline{Y}) = \lambda(\underline{Y})$ and the arrival (failure) rate of the population not depend on the vector $\underline{Y}$, but depends only on the number of states changes (arrivals, failures) of the elements $\lambda(\underline{Y}) = \lambda\|\underline{Y}\|$, where $\|\underline{Y}\| = \sum_{k=1}^{n} Y_k^2$. The process $\|\underline{Y}(t)\|$, in which the transition from the state $\|\underline{Y}\| = i$ is possible only to the state $i+1$ with the rate $(n-1)\lambda(i)$ is a pure death process.
The pure death process describes functioning of the non-renewable the system consisting of identical elements while the generalised one describes functioning of the system consisting of different elements. It seems to me that the application of the Markov methods for the research of the length of the population lifetime when the population is treated as a system of n, $n \geq 1$ elements is natural.
Namely – changes in time of the population states (as the system of n, $n \geq 1$ elements) are determined by multidimensional stochastic process. This process is chosen in such a way that when we investigate the reliability of the population and we know the state of the process, we can in each moment explicitly define the state of the population (whether the system is functioning or failed, whether the population is in the state of life or death).
Bellow is presented a general model of functioning of that population (as the system of n, $n \geq 1$ elements), (Sołowiew, 1983).

7.2 General model functioning of the population

Let $\xi(t)$ denote the stochastic process chosen in the way described above.
A set E of this process states decomposed into two disjoined subsets $E = E_+ \cup E_-$.
If $\xi(t) \in E_+$, then the population is in state of life at the moment t and if $\xi(t) \in E_-$, then the population is in state of death at the moment t.
The transition of the process from the set E_+ to the set E_- is called death of the population, and a reverse transition from the set E_- to the set E_+ is called a renewal of the population.
The process observed in time changes the state E_+ into the state E_- and reverse.
Let

$$Z_0', Z_1', \ldots, Z_k', \ldots \text{ and } Z_0'', Z_1'', \ldots, Z_k'', \ldots$$

denote successive intervals respectively, in which the population is in the state of life or death.
If there exists a stationary distribution for the process $\xi(t)$, which describes the functioning of the renewable population, then there exist limits $Z_k' \xrightarrow{P} Z'$ $Z_k'' \xrightarrow{P} Z''$.
A random variable
Z_0' - is called the time until the first change of the state of the population
Z' - is called the length of the lifetime of the population per one change of the state

$Z(t)$ - is the lifetime of the population from the moment t until the first moment after the change of the state of the population, residual time of the population lifetime.
As before, if there exists a stationary distribution for the process $\xi(t)$, then there exists the limit $Z(t) \xrightarrow{P} Z$.
Distributions of random variables $Z_0^{'}, Z_k^{'}, Z_k^{''}, Z(t), Z', Z'', Z$ and their an expected values $E(Z_0^{'})$, $E(Z_k^{'})$, $E(Z_k^{''})$, $E(Z(t))$, $E(Z')$, $E(Z'')$, $E(Z)$) are basic characteristics of the reliability function.
One of the simpler type of the processes used in the reliability calculations is the class of Markov processes with a finite or uncountable number of states. In this sense the pure death or general death process could be a model describing the changes of the population states.

In aspect of Internet survey three characteristics are important:
$Z_0^{'}$ - the time until the first change of the state of the population,
Z' - the length of the lifetime of the population per one change of the state,
$Z(t)$ - the lifetime of the population from the moment t until the first moment after the change of the state of the population, residual time of lifetime of the population.

Below the probability distribution of these random values and their the expected values are derived.

7.3 Formulae of characteristics of Markov process

Assume that the process $\xi(t)$ is a homogeneous Markov process with a finite number of the states which is assigned $1,2,...,N$.
Let the set $E_+ = \{0,1,2,...,n\}$ denote a set of the functioning states of the system (life states of the population) and $E_- = \{k+1, k+2,...,n\}$ a set of the failed states of the system (a set of death states of the population which is interpreted as break between repeated surveys).
The Markov process is analysed in following steps.

1. The Markov process has two properties which are equivalent to the definition the Markov process:

- the interval in which the process finds itself in the state i, does not depend on the process taking place outside the interval and it has exponential distribution

$$P\{t_i > t\} = e^{-\lambda_{i,i} t} \text{ , where } \lambda_{i,i} = -\sum_{j \neq i} \lambda_{i,j} ,$$

- a sequence of the states through which the process passes is the homogenous Markov chain with transition probabilities $\pi_{i,j} = -\frac{\lambda_{i,j}}{\lambda_{i,i}}$. At the same time, if $\lambda_{i,i} = 0$, then the state i is called an absorbing state, because when the process has entered the state, it will remain in it forever. In this case we assume that $\pi_{i,j} = 0$.

2. Solution of the Kolmogorov equations

Let

$$p_i(t) = P\{\xi(t) = i\}.$$

The probabilities of the states $p_i(t)$ satisfy the Kolmogorov equations

$$p_j'(t) = \sum_{i=0}^{N} \lambda_{i,j} p_i(t), \quad j = 0, 1, \ldots, N \tag{7.3.1}$$

These equations are rewritten in a matrix form

$$p'(t) = p(t)\Lambda,$$

where $p(t) = \lfloor p_0(t), \ldots, p_n(t) \rfloor$ is the vector of the states probabilities, $\Lambda = \lfloor \lambda_{i,j} \rfloor$ is the matrix of the transition rate with the properties:

a) all elements of the matrix $\Lambda = \lfloor \lambda_{i,j} \rfloor$ satisfy the condition $\lambda_{i,j} > 0$,

b) $\sum_{j=0}^{N} \lambda_{i,j} = 0$.

After assuming the initial distribution of the process $p_i(0) = p_{i0}$ the system of the Kolmogorov equations $(7.3.1)$ has the explicit solution

$$p_i(0) = p_{i0}.$$

The system of the Kolmogorov equations $(7.3.1)$ is solved by means of Laplace transform.

Let

$$a_i(z) = \int_0^\infty e^{-zt} p_i(t) dt.$$

Taking the properties of Laplace transform is obtained

$$-p_{j0} + z a_j(z) = \sum_{i=0}^{N} a_i(z) \lambda_{i,j}, \quad j = 0, 1, 2 \ldots, N,$$

and from the Cramer's formulae the solution of the system is given by

$$a_i(z) = \frac{\Delta_i(z)}{\Delta(z)} \tag{7.3.2}$$

where $\Delta(z) = \left\| z\delta_{i,j} - \lambda_{i,j} \right\|$, the determinant in the numerator is calculated from the determinant in the denominator by the replacement of the i- th line with the line of the initial probabilities p_{j0}, $\delta_{i,j}$ is a Kronecker's symbol. After inversion of the Laplace transform the formulae of the probability of the process states is obtained.

3. Conditions of existing of the stationary distribution of the process.

Two states are called communicating, if there exist such indices $i_1, \ldots, i_k$ as well as $j_1, \ldots, j_l$, that $\lambda_{i,i_1}, \lambda_{i_1,i_2}, \ldots, \lambda_{i_k,j} > 0$ and $\lambda_{j,j_1}, \lambda_{j_1,j_2}, \ldots, \lambda_{j_l,i} > 0$. We also assume that each state communicates with itself.

The communication relation is an equivalence relation and introduce the decomposition of the set of the states into the classes of communicating states $A_1,\dots,A_m$.

The class A_s follows the class A_r, if there exist the states $i \in A_r$ and $j \in A_s$ such that $\lambda_{i,j} > 0$.

A class is called ergodic, if it is not followed by any class. By compliance with or satisfaction of these conditions the proposition holds:

The stationary distribution of the process, $\lim_{t\to\infty} p_i(t) \xrightarrow{P} p_i$, which is independent of the initial distribution exists if and only if there exists exactly one ergodic class, whereas the stationary probabilities satisfy the system of equations:

$$\sum_{i=0}^{N} p_i \lambda_{i,j} = 0, \quad j = 0,1,2,\dots,N, \quad \sum_{i=1}^{n} p_i = 1 \tag{7.3.3}$$

4. Basic characteristics of the reliability function

Let A be a set of the states and let $i \in A$.

The transition time to the set A is defined as a random variable

$$\chi_i(A) = \inf\{t : \xi(t) \in A \,|\, \xi(0) = i\}.$$

Its probability distribution is obtained by means of Laplace transform

$$\varphi_i(z) = E[\exp(-z\,\chi_i(A))].$$

From the formula of the total probability for the expected value

$$\varphi_i(z) = \frac{-\lambda_{i,i}}{z-\lambda_{i,i}}\left(\sum_{j\in A}\frac{\lambda_{i,j}}{-\lambda_{i,i}} + \sum_{j\notin A, j\neq i}\frac{\lambda_{i,j}}{-\lambda_{i,i}}\varphi_j(z)\right)$$

is computed in the following way.

In the state i the process is for the time ξ_i which is the random variable with exponential distribution and Laplace transform $\frac{-\lambda_{i,i}}{(z-\lambda_{i,i})}$, next, from this state it passes with the probability $\frac{-\lambda_{i,j}}{(-\lambda_{i,i})}$ to the state j.

If $j \in A$, then $\chi_i(A) = \xi_i$, and if $j \notin A$, then $\chi_i(A) = \xi_i + \chi_j(A)$ and both components are independent.

The solutions of the equations system

$$\begin{cases} \sum_{j\notin A} (z\,\delta_{i,j} - \lambda_{i,j})\varphi_j(z) = \lambda_i(A) \\ \quad i \notin A, \lambda_i(A) = \sum_{j\in A} \lambda_{i,j} \end{cases} \tag{7.3.4}$$

are rational functions and after inversing them the probability distribution of the random variable $\chi_i(A)$ is obtained.
Differentiating the equation system (7.3.4) as regards to z and substituting $z=0$, the expected transition times

$$E(\chi_i(A)) = -\varphi_i'(0)$$

are obtained and they are satisfying the equations system

$$\sum_{j \notin A} \lambda_{i,j} E(\chi_j(A)) + 1 = 0, \; i \notin A .$$

Let $q_{i,j}(A)$ denote probability of an event, that in the moment of the first entrance of the process to the set *A,* the process will enter the state $j \in A$, on condition that at initial moment the process was in the state $i \notin A$.
In the analogical way as above are obtained the forward equations for these probabilities

$$q_{i,j}(z) = \frac{\lambda_{i,j}}{-\lambda_{i,i}} + \sum_{k \notin A, k \neq i} \frac{\lambda_{i,k}}{-\lambda_{i,i}} q_{k,j}(A) .$$

Transforming the equations we have

$$\sum_{k \notin A} \lambda_{i,k} q_{k,j}(A) = -\lambda_{i,j}, \; i \notin A, \, j \in A \tag{7.3.5}$$

5. Transcription of the characteristics from the point 4 by means of $\chi_i(A)$ and the probability $q_{i,j}(A)$.
Let at the initial moment $\xi(0)=0$, that is at the initial moment all the elements of the system are functioning (all the elements of the population are in the state of life).
Then

$$Z_0' = \chi_0(E_-) .$$

The probability $q_{0,j}(E_-)$ is computed from the equations system of (7.3.5) and from the formula of the total probability is derived

$$P\{Z_i'' \le t\} = \sum_{j \in E_-} q_{0,j}(E_-) P\{\chi_j(E_+) \le t\} .$$

Probability of the event that in the moment of the first renewal the process enters the state $k \in E_+$ from the formula of the total probability is equal to

$$\sum_{j \in E_-} q_{0,j}(E_-) q_{j,k}(E_+) .$$

If these probabilities are known, then the probability distribution Z_1' and the probabilities of entrance to given states at the moment of the second failure (at the moment of the second change of the population state) can be computed. This in turn makes it possible to compute the probability distribution $Z_2^{'}$ and so on.

To compute the stationary distributions of the random variables $Z^{'}$ and $Z^{''}$ let us introduce the following probabilities:

- $q_i(E_+)$ - probability of an event that in the moment of the renewal of the population (as the system) in stationary conditions the process enters the state $i \in E_+$,
- $q_i(E_-)$ - probability of an event that in the moment of the death of the population (the failure of the system) in stationary conditions the process enters the state $i \in E_-$.

$q_i(E_+)$ is the probability of the event that in the time dt the process will pass from the set E_- to the state $i \in E_+$ on condition that in this time the process passed from the set E_- to the set E_+.

Formally, it is transcribed by formulas

$$q_i(E_+) = \frac{\sum_{j \in E_-} p_j \lambda_{j,i}}{\sum_{i \in E_+} \sum_{j \in E_-} p_j \lambda_{j,i}} \tag{7.3.6}$$

and analogically

$$q_i(E_-) = \frac{\sum_{j \in E_+} p_j \lambda_{j,i}}{\sum_{i \in E_-} \sum_{j \in E_+} p_j \lambda_{j,i}} \tag{7.3.6}$$

where p_j are the stationary probabilities of the process.

The probability distributions of the random variables $Z^{'}$, $Z^{''}$ and Z - residual time of the system life are given by

$$P\{Z^{'} \le t\} = \sum_{i \in E_+} q_i(E_+) P\{X_i(E_-) < t\} \tag{7.3.7}$$

$$P\{Z^{''} \le t\} = \sum_{i \in E_-} q_i(E_-) P\{X_i(E_+) < t\} \tag{7.3.8}$$

$$P\{Z > t\} = \sum_{i \in E_+} p_i P\{X_i(E_+) > t\} \tag{7.3.9}$$

The expected values of the basic reliability characteristics are given by

$$E(Z_0^{'}) = E(X_0(E_-)) ,$$

$$E(Z^{'}) = \sum_{i \in E_+} q_i(E_+) E(X_i(E_-)) ,$$

$$E(Z^{''}) = \sum_{i \in E_-} q_i(E_-) E(X_i(E_+)) ,$$

$$E(Z) = \sum_{i \in E_+} p_i E(X_i(E_-)) ,$$

8. The pure death process

The pure death process is a homogenous Markov process $\xi(t)$ with the space of states 0,1,2,...,n,..., for which

$$\lambda_{i,j}=\lambda_i=0 \text{ as well as } \lambda_{i,j}=0 \text{ for } j\neq i+1 .$$

That is, the pure death process passes successively through the states $0\rightarrow 1\rightarrow 2\rightarrow ...n\rightarrow ...$ and in the state k it is for the time t_k with exponential distribution

$$P\{t_k>t\}=e^{-\lambda_k t}$$

and next it passes to the state $k+1$ with the probability one.

Let $E_+=\{0,1,2,...,n\}$ be a set of the functioning states of the system (the set of the states of the life of the population), $E_-=\{n+1,n+2,...\}$ is a set of the failed states of the system (the set of the states of death of the population), and $\xi(0)=0$.

Because the pure death process describes behaviour of the non- renewable system, we assume that a basic characteristic of the pure death process is the time until the first failure of the system (the time until the first change of the state of the population, the time until the first of death of the population)

$$Z_0'=X_0(E_-).$$

And as before

$$X_i(A)=\inf\{t:\xi(t)\in A|\ \xi(0)=i\} \quad \text{and} \quad E[\exp(-zX_i(E_-))]=\varphi_i(z)$$

are introduced.

Then the equations system (7.3.4) takes on the form

$$\begin{cases} -\lambda_i\varphi_{i+1}(z)+(z+\lambda_i)\varphi_i(z)=0,\ i<n \\ (z+\lambda_n)\varphi_n(z)=\lambda_n \end{cases}$$

from which it follows that

$$\varphi_0(z)=E\left(e^{-z\tau_0'}\right)=\frac{\lambda_0\lambda_1\cdots\lambda_n}{(z+\lambda_0)...(z+\lambda_n)} ,$$

$$\overline{F}(t)=P\{Z_0'>t\}=\lambda_0\lambda_1\cdots\lambda_n\sum_{k=0}^{n}\frac{e^{-\lambda_k t}}{\lambda_k w'(-\lambda_k)} \qquad (8.1)$$

where $w(x)=(x+\lambda_0)(x+\lambda_1)...(x+\lambda_n)$.

Because the time until the first moment of the failure of the system (the time until the first change of the state of the population, the time until the first death of the population) is a sum of the random variables

$$Z_0'=\xi_0+\xi_1+...+\xi_n$$

hence the expected time until the first failure of the system (the time until the first change of the state of the population, the time until the first death of the population) is equal to

$$E(Z_0') = \frac{1}{\lambda_0} + \frac{1}{\lambda_1} + \ldots + \frac{1}{\lambda_n} \tag{8.2}$$

9. The general pure death process

The general pure death process is a homogeneous Markov process $\xi(t)$ with the number of states $E = \{\underline{Y}: \ \underline{Y}: Y^n \rightarrow \{0,1\}^n\}$, where $\underline{Y} = (Y_1,\ldots,Y_n)^T$, $Y_i = 0$ or $Y_i = 1$, for which only immediate transitions in the form

$$\underline{Y} = (Y_1,\ldots,Y_{i-1},0,Y_{i+1},\ldots,Y_n) \rightarrow \underline{Y}' = (Y_1,\ldots,Y_{i-1},1,Y_{i+1},\ldots,Y_n), \quad i = 1,\ldots,n,$$

with the transition rates $\lambda_i(\underline{Y})$ are possible.

Such a transition denotes a change of the state of one element of the population and can be interpreted in the research of the population lifetime as a failure, an arrival or a death of the i-th element.

By the previous denotations E_+ is a set of the states of life of the population (a set of the functioning states of the system), E_- is a set of the states of death of the population (a set of the failed states of the system).

We assume that in the initial moment all the elements are in the state of life that is $\underline{Y}(0) = \underline{0}$.

The notion of the way π is defined as class of the realisation of the process $\underline{Y}(t)$ with an assumed sequence of the states, through which the process passes. This sequence begins from the state at the initial moment zero and ends with the state at the moment of the change of the state of the system (the failure of the system, the death of the population)

$$\pi = \left(\underline{Y}^{(0)}, Y^{(1)},\ldots,Y^{(m)}\right),$$

$$\underline{Y}^{(0)} = (0,0,\ldots,0), \ \underline{Y}^{(k)} \in E_+ \ \text{for} \ k < m, \ Y^{(m)} \in E_-.$$

Each transition from a state to a state described by the way $\pi = \left(\underline{Y}^{(0)}, Y^{(1)},\ldots,Y^{(m)}\right)$ is the change of the state of the system element (which is the failure of the system element or arrival or death of the element of the population of the size n, $n \geq 1$).

Summary rate is defined then as

$$\lambda(\underline{Y}) = \sum_{i=1}^{n} \lambda_i(\underline{Y}),$$

where $\lambda_i(\underline{Y}) = 0$, if in the state $\underline{Y}$ i-th element is already failed (i-th element of the population has arrived).

Then from the properties of the Markov process

$$p_i(\underline{Y}) = \frac{\lambda_i(\underline{Y})}{\lambda(\underline{Y})}$$

is the probability of the event that the process passes from the state $\underline{Y} = (Y_1, ..., Y_{i-1}, 0, Y_{i+1}, ..., Y_n)$ to the state $\underline{Y}' = (Y_1, ..., Y_{i-1}, 1, Y_{i+1}, ..., Y_n)$, that is i -th element changes its state (fails, arrives or dies).
Let i_k denote the number of the element changing the state (failing, arriving or dying) when it passes from the state $Y^{(k-1)}$ to the state $Y^{(k)}$ during the passage of the way π.
Then from the formula of the total probability the probability of the change of the system state (failure of the system, death of the population) equals

$$F(t) = P\{Z'_0 \le t\} = \sum_{\pi} F(t|\pi)p(\pi) \tag{9.1}$$

where sum is calculated for all the possible ways π,

$$p(\pi) = p_{i_1}\left(e^{(0)}\right)p_{i_2}\left(e^{(1)}\right)...p_{i_m}\left(e^{(m)}\right)$$

is the passage probability of the way π, and $F(t|\pi)$ is the conditional probability of the change of state of the system (failure of the system, death of the population) on the condition of the passage of the way π.
Since in each state $\underline{Y}^{(k)}$ the process remains for the time with the exponential distribution with the parameter $\lambda\left(\underline{Y}^{(k)}\right)$, hence the conditional process $\underline{Y}(t)$ on the condition of the passage of the way π is the pure death process with transition rates being given by $\lambda\left(\underline{Y}^{(k)}\right)$.
Substituting to the form (9.1) the expression (8.1) the probability of the first death of the population (the first failure of the system) is obtained.
The expected time of the first death of the population (the first failure of the system) from the equality (8.2) is derived and equals

$$E\left(Z'_0\right) = \sum_{\pi} p(\pi)\left[\frac{1}{\lambda\left(\underline{Y}^{(0)}\right)} + ... + \frac{1}{\lambda\left(\underline{Y}^{(m-1)}\right)}\right].$$

10. Conclusions

In this study the process of Internet data collection is interpreted and analysed as a random experiment or the life test of population surveyed by using the notions and methods of the probability and reliability theories. A random set of respondents who participate in Internet survey is called the uncontrolled sample and defined as the counting process by using Poisson processes. The proposed approach allows to study some stochastic properties of the process of the Internet data collection, the calculation and the estimate of the basic characteristics by the assumed assumptions. Moreover, the Markov methods are applied to description of a life testing experiment in which the basic characteristics of a reliability of the length of the population lifetime are derived, when the finite population is interpreted as a system with the monotonic structure function and the changes of the population states are described through the death and the general death processes.

11. Acknowledgments

The research was partially supported by the Ministry of Science and Higher Education programs 03/S/0063/04, 03/S/0059/07 and 03/E/0015/08, Warsaw School of Economics.

12. References

Barlow, R.E. & Proschan, F. (1965, 1996). *Mathematical theory of reliability*. J. Wiley and Sons, Inc. New York or Philadelphia SIAM, ISBN 0-89971-369-2

Barlow, R.E. & Proschan, F. (1975). *Statistical theory of reliability and life testing*, Holt, Reinhart and Winston INC. New York

Biffignandi, S. & Pratesi, M. (2000). Modelling firm response and contact probabilities in Web surveys, *Proceedings of the "Second International Conference on establishment surveys"*, USA, June 2000, Buffalo

Bracha, Cz. (1996). *Theoretical foundation of survey sampling* (in Polish), Polish Scientific Publisher, ISBN 83-01-12123-8, Warsaw

Callegaro, M. & Disogra, CH. (2008). Computing response metrics for online panels, *Public Opinion Quarterly*, Vol. 72, No. 5, 2008, Special Issue, pp. 1008-1032, ISSN 0033-362X

Couper, M.P. & Miller, P.V. (2008). Web survey methods, *Public Opinion Quarterly*, Vol. 72, No. 5, 2008, Special Issue, pp. 831-835, ISSN 0033-362X

Dilman, Don A.; Smyth, J. & Christian, L.M. (2008*). Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method, 3rd ed.,* Hoboken, NJ: John Wiley Co., ISBN 978-0-471-69868-5, New York

Galesic, M. (2006). Dropout on the Web: effects of interest and burden experiences during an online survey, *Journal of Official Statistics*, Vol. 22, No. 2, June 2006, pp. 313-328, ISSN 0282-423X

Getka-Wilczyńska, E. (2003). Using WWW for data collection in a national survey on scholarship granted to young researches, *Statistics in Transition*, *Journal of the Polish Statistical Association,* Vol. 6, No. 1, June 2003, pp. 33-55, ISSN 1234-7655

Getka-Wilczyńska, E. (2004). About random properties of an uncontrolled sample (in Polish, O losowych własnościach próby niekontrolowanej), *Annals of College of Economic Analyses*, Vol. 13, 2004, pp. 59-69, ISSN 1232-4671

Getka-Wilczyńska, E. (2005). Stochastic properties of Internet data collection process, *Proceedings of the 55th Session of the International Statistical Institute*, CD-ROM, paper No. 998, ISBN 1-877040-28-2, Australia, April 2005, International Statistical Institute, Sydney

Getka-Wilczyńska, E. (2007). Markov methods in test of the population lifetime, *Proceedings of the 56th Session International Statistical Institute 56th Session,* CD-ROM, papers No. 1708, ISBN 978-972-8859-71-8, Portugal, July 2007, Centro de Estatística e Aplicacões (CEAUL), Instituto Nacional de Estatística (INE) & International Statistical Institute(ISI), Lisbon

Getka-Wilczyńska, E. (2008). The Internet sample, *Statistics in Transition – new series, Journal of Polish Statistical Association*, Vol. 8, No. 3, December 2008, pp. 553-560

Hardy, G.H.; Littlewood, J.E. & Polya, G. (1934). *Inequalities.* Cambridge University Press, Cambridge Internet and Innovative Data Collection (2001). *Bulletin of the International Statistical Institute*, 53rd *Session*, *Proceedings, Tome LIX, Three books, Book 1*, Topic 16, pp. 229-253, ISBN 90-73592-20-8, South Korea, August 2001, International Statistical Institute, Seoul

Kingman, J.F.C. (2002): *Poisson' processes* (in Polish), Polish Scientific Publishers, ISBN 83-01-13534-4, Warsaw

Kopociński, B. (1973). *Outline of the renewal and reliability theory,* (in Polish), Polish Scientific Publishers, Warsaw

Kordos, J. (1982). *Quality of statistical data* (in Polish), Polish Economic Publisher, Warsaw

Koutra, M. V.; Tsitmidelis, S. & Zissmopoulos, V. (2003). Evaluation of reliability bounds by set covering models, *Statistics and Probability Letters*, 61, 2003, pp. 163-175, ISSN 0167-7152

Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel Web surveys, *Journal of Official Statistics*, Vol. 22, No. 2, June 2006, pp. 329-349, ISSN 0282-423X

Särndal, C.E.; Swensson, B. & Wretman, J. (1992). *Model assisted Survey Sampling,* Springer-Verlag New York Inc., ISBN 0-387-97528-4, New York

Sołowiew, A.D. (1983). *Analytical methods in reliability theory* (in Polish), Technical – Scientific Publisher, ISBN 83-204-0473-8, Warsaw

Steczkowski, J. (1988). *Application of survey sampling in social-economics research* (in Polish), Polish Economic Publisher, Warsaw

Tillĕ, Y. (2006). *Sampling algorithms,* Springer Science-Business Media, Inc., ISBN 10:0-387-30814-8; ISBN 13:978-0387-30814-2, New York

Toepoel, V.; Das, M. & Van Soes, A. (2008). Effects of design in Web surveys, *Public Opinion Quarterly*, Vol. 72, No. 5, 2008, Special Issue, pp. 985-1007, ISSN 0033-362X

Vehovar, V. (2007). Workshop on Internet Survey Methodology, Ullehammer, September 2007, http://vv.ris.org

www.WebSM.org

Wywiał, J. (1992). *Statistical survey sampling in economics research* (in Polish), Economic Academy, Katowice

Engineering the Computer Science and IT

Edited by Safeeullah Soomro

ISBN 978-953-307-012-4
Hard cover, 506 pages
Publisher InTech
Published online 01, October, 2009
Published in print edition October, 2009

It has been many decades, since Computer Science has been able to achieve tremendous recognition and has been applied in various fields, mainly computer programming and software engineering. Many efforts have been taken to improve knowledge of researchers, educationists and others in the field of computer science and engineering. This book provides a further insight in this direction. It provides innovative ideas in the field of computer science and engineering with a view to face new challenges of the current and future centuries. This book comprises of 25 chapters focusing on the basic and applied research in the field of computer science and information technology. It increases knowledge in the topics such as web programming, logic programming, software debugging, real-time systems, statistical modeling, networking, program analysis, mathematical models and natural language processing.