Snowball Sampling Developments used in Marketing Research

Irina-Maria Drăgan
and
Alexandru Isaic-Maniu

Bucharest University of Economic Studies
Department of Statistics and Econometrics
Address: 15-17 Calea Dorobanți, District 1, room 2610,
Bucharest 010552, Romania

Corresponding author: Irina-Maria Drăgan
E-mail: irina.dragan@csie.ase.ro
Telephone number: +40723408598

Alexandru Isaic-Maniu
E-mail: alexandru.isaic@csie.ase.ro
Telephone number: +40722893783

Abstract
The snowball sampling as network research presents numerous advantages in registering the "hidden populations", but also a major disadvantage – the absence of objective, numeric criteria for the decision of stopping the investigation. There is also a certain ambiguity about the volume of the investigated sample, which cannot be previously calculated, but is at the choice of the investigator to stop the research when he considers a satisfactory volume of information was accumulated.
The purpose of this paper is to adopt the sequential method for the case of small-volume populations, a typical situation for the researches that require the snowball sampling method.

Key words: Snowball sampling, sequential sampling, decision criteria, decision risks.

1. Introduction

Snowball sampling is a research method using surveys and data registering, usually helpful in the case of sociological, psychological or management studies, and is recommended when:

- The population cannot be exactly delimited or enumerated (e.g. homeless people);
- The sample features are rare (drug addicts, people with rare diseases, unemployed young people, club members, elites, etc.);
- When the research refers to behaviors, perceptions, habits, in order to describe "typical" cases, that cannot be generalized at the level of the entire population (a typical rural family, dropping out of school, the issue of malnourished children, political leaders in a certain geographical area, members of an organization, virtual communities, ethnic communities, the elite classes in certain fields, intellectually gifted children, socially stigmatized children, the study of illegal immigrants, young people practicing exclusive sports (hockey, golf, skating), etc. The most usual cases, for which this non-probabilistic method has been applied, are: the research on drug-addicts, persons with rare illnesses or AIDS, etc).

The characteristic of this type of survey is that it is not used in estimating the characteristics of the general population, but in estimating the characteristics of a population network in “hidden” populations (rare, difficult to identify).

The term “hidden population”, synonym with “very rare” population, or “difficult to identify” population, generally refers to those populations which are short of official information, or that represent less than 2% of the national population. In this manner, due to their rareness, this type of population is difficult to identify, to study and engage in an investigation, most of the times because of its social stigma, legal status, and the lack of visible consequences of its members. Starting from the definitions presented, it is clear that for a “hidden population” there cannot be a pre-defined manner of sampling, and it cannot constitute the subject of neither a probabilistic sampling method, nor of most of the non-probabilistic surveys.

2. Conceptual and Theoretical Aspects

A few major topics have been the favorite subject of some authors concerned with the use of the “snowball” method in survey research; thus, we can summarize:

a) The general issue of social networks, studied by authors such as: Martino, F., Spoto, A. (2006); Kuhar, R., Švab, A. (2008); Atkinson, R., Flint, J. (2001); Hanneman, R., A. (2001); Dodds, P., S., Muhomad, R., Watts, D., J. (2003); Adams, S., Carter, N., Hadlock (2008);
b) Migration and human traffic, is the subject of studies published by Salt, J. and Hogarth,J. (2000); Paspalanova, M: (2006); Salt, J., Clarke, J., Smidt, S. (2000);
c) The issue and estimation of the homeless people: D`Onise, K., Wang, Y., McDermott, R. (2007); Skeldon, R. (2004);
d) Tobacco and drug use is a subject also widely analyzed by Spreen, M., Coumans, M. (2003); Avico, U., Kaplan, C., Korczak, D. ,Van Meter (1988); Hendricks, V., M., Blanken, Adriaans, N. (1992); Griffiths, P., Gossop, M., Powis, B., Strang, J., Duncan, D., F., White, J., B., Nichalson, T. (2003); Etter, J., F., Perneger, T., V. (2000);
e) Child abuse and the issue of teenagers, is the subject of several interesting studies, such as: Barlow, J., Davis, H., McIntosh, E., Jarett, P., Mockford, C., Stewart-Brown, S. (2007); Day, C.,Davis, H. (2006); Kirkpatrick, S., Barlow, J., Stewart-Brown, S., Davis, H. (2007);
f) The research of sexual minorities through the snowball method is reflected in the studies of several authors such as: Švab, A., Kuhar, R. (2008); Diamond, M. (1993); Standfort, T. (1997); Aaron, D., J., Chang, Y-F., Markovic, N., LaPorte, R., E. (2003);
g) The statistical approach on snowball sampling and the methodological issues, specific to the research can be found at: : Berg, S. (1988); Borgatti, S., P., Everett, M., G., Freeman, L., C. (2002); Johnson, T., P. (2005); Song Hoon Lee, Pan-Jum Kim, Hawoong Jeong (2006); Kolaczyk, E., D. (2009); Gabor, M., R. (2007); Goodman, L., A. (1961); Streeton, R., Cooke, M., Campbell, J. (2004), Kostinen, J., Snijders, T. (2007), Salganik ,M . (2006).

The principle of this sampling method includes the identification, done by the researcher, based on specific reasoning, of a number of respondents to be interviewed, and which in their turn, shall indicate (recommend) other respondents which will make the object of the research.

Due to the methodology used, consisting in the initial subjects indicating additional subjects (random methods can be used to generate the first respondents), this method can decrease the searching costs very much, but can also become costly when introducing the systematic error, because the use itself reduces the probability for the survey to have an accurate representation of the population.

The procedure would gain additional precision if the sequential analysis method were to be adopted; this procedure is adapted for the case of population with reduced volume, specific to social networks, which is actually the objective of this paper.

3. Brief History of Sequential Analysis

According to Irving I. Burr (1953) a so-called sequential test is a procedure through which after each phase (measurement, determining, testing etc) a certain hypothesis can be accepted, rejected, or additional information (proof, phase) can be requested. Precisely for these reasons, the size of the examined sample is random and not known in advance.

In some cases however, this volume is very small or on the contrary, very large. The uncertainty state, that is when an additional phase is requested, can take more or less, depending on the supplementary information brought by each additional individual.

The theoretical basis of the sequential analysis was set in the '40s (the 20th century), independent research in this regard being performed in Great Britain (see George A. Barnard, 1946, "Sequential tests in industrial statistics", Journal of the Royal Statistical Society, Ser. B., No.8, p. 1-26) and in USA, leader from this point of view being Abraham Wald, the one who succeeded in proving the "critical aspects" of this methodology.

Abraham Wald (1902-1950), considered to be the father of sequential analysis was born in Transylvania, Romania, in a Jewish family, in a period when the Romanian region belonged to the Austro-Hungarian Empire. He studies at the University in Cluj, continuing in Vienna with his Doctorate's degree under Professor Karl Menger. He is the one, who introduces Wald to a well-known Austrian economist and banker, Karl Schlesinger, who succeeds in making Wald be interested in econometric research, a new and developing field at the time.

By 1938, the time he remained in Vienna, he succeeded in demonstrating in a rigorous manner, thanks to his scientific interest and results, a series of suppositions that had been the concern of many economists at that time. Wald wrote in German at that time, a series of his papers being republished in English after the Second World War (e.g. "On some Systems of equations of mathematical economics", 1936, translated in Econometrica).

In the '40s, at Columbia University in New York, the Statistical Research Group of Columbia University (also known as SRGCU) was being founded, which under the War Department name, attacked a series of issues of interest and military use, such as the optimal assignment of resources, ammunition quality verification, the way of distributing artillery hits around a fixed target, comparing the penetration resistance for various armors, etc. (the U.S.A as we know, goes to war against Japan in December 1941, after the Pearl Harbor episode).

This group included important personalities in the American scientific society such as Milton Friedman, Harold Freeman, W. Allen Wallis and others. Wald was adopted in this multi-disciplinary group as a consultant on statistic issues, also having the advantage of speaking other languages than English (German, Hungarian, French, and most probably Romanian). Friedman and Wallis requested Wald's help, who dedicated himself since, with all of his energy and talent, in the direction of finding solutions for the problems brought up by the "sequential experiment" - as he accurate called it ever since 1943.

Wald understood the two essential issues must be resolved:

a) Demonstrating that the sequential procedure is ended (finishes) after a finite number of steps – meaning that it cannot continue indefinitely and
b) Building a quantitative criterion for the decisional process.

Unlike the classical case of the statistic survey, the sequential analysis involves three decisions:

(i) To continue the research
(ii) To accept the initial hypothesis

(iii) To reject the hypothesis (and automatically accept the alternative of that hypothesis).

In 1943, Wald writes a secret technical report entitled "Statistical Analysis of Statistical Data: Theory", which SRGCU sent to the Defense Research Committee of War Department.

In this document, Wald strongly founded the sequential analysis, by creating the so-called SPRT (Sequential Probability Ratio Test). This SPRT practically becomes the work instrument in sequentially testing statistical hypotheses.

After the war ended – precisely in May 1945, SRGCU receives the approval from the War Department of eliminating the "secret" classification of its reports and papers, and thus ever since Wald's and the other's results began to be known by the international scientific world, being published in various specialty journals – mainly American (The Annals of Statistics, Journal of the American Statistical Association, Industrial Quality Control etc).

The main volume in the field appeared in 1947 at John Wiley Publishing house, volume which is Sequential Analysis by Wald.

The examples given by Wald in his book belong to qualitology domain, proving the importance that the famous savant gave to the issues imposed by this developing domain.

Also, Wald included in his monograph, a brief historical research, in the attempt to identify similar preoccupations in the area, in other regions of the world. Thus, he mentions the sequential experiments performed in Bengal (India) by Prasanta Chandra Mahalanobis (1893-1972) regarding the census of the jute plantations, as well as the theoretical results obtained in certain particular sequential analysis issues by the British statisticians C.M. Stokman and G.A. Barnard in 1944.

4. The Sequential Test

Let us consider a measurable variable *X* – a continuous random variable, in statistics, characterized by its density f(x;θ), where θ is an unknown parameter (or a parameter vector) on which we concentrate our deduction. Wald's construction is also valid for discrete random variables, but we shall operate with the continuous ones to demonstrate the method.

Let be a hypothesis

$$H_0 : \theta = \theta_0 \tag{1}$$

With the alternative:

$$H_1 : \theta = \theta_1, (\theta_0 < \theta_1) \tag{2}$$

Here, θ_0 and θ_1 may have various interpretations – for example an average value.

Then, $f(x;\theta_0)$ represents the density of *X* when H_0 is true, while $f(x;\theta_1)$ is the density of *X* when the alternative H_1 is just. We make successive measurements on variable *X*, $x_1, x_2, x_3 \cdots$ and we write the likelihood functions associated to the two hypotheses:

$$P_{o,n} = f(x_1;\theta_0)\cdot f(x_2;\theta_0)\ldots f(x_n;\theta_0) \tag{3}$$

$$P_{l,n} = f(x_1;\theta_1)\cdot f(x_2;\theta_1)\ldots f(x_n;\theta_1) \tag{4}$$

The quantity

$$R_n = P_{ln}/P_{on} \tag{5}$$

is called likelihood ratio (term introduced by Sir Roland A. Fisher, 1890-1962).

The sequential test is built as follows: for starters, we choose two constants A and B (depending on the associated risks α and β, corresponding to the two hypotheses), both A and B being positive, and A>B.

At each step of the experiment (interview, measurement, examination, etc) we calculate R_n; if B< R_n <A, the experiment continues, extracting a new sample for analysis. If $R_n \leq A$, then the process is completed, by accepting the alternative hypothesis (H_1) and automatically rejecting the nil one (H_0). However if $R_n \geq B$ then the experiment is completed by accepting the H_0 hypothesis while rejecting H_1.

Wald (see his book in 1947, and Barsanu-Pipu and others, 2002) showed that it is difficult to obtain exact values for A and B but the approximations

$A \approx (1-\beta)/\alpha$ and $B \approx \beta(1-\alpha)$

fully satisfy the theoretical requests of the sequential method.

Wald also makes the following remark (idem op. cit. page 38):

"For reasons related to the easiness of calculus, it is much more convenient to work with the logarithm of fraction $P_{1,n}/P_{0,n}$ rather than the fraction itself. The reason is that log ($P_{1,n}$ /$P_{0,n}$) can also be written as a sum of n terms."

Indeed, we have:

$$\log R_n = \log\frac{f(x_1;\theta_1)}{f(x_1;\theta_0)} + \log\frac{f(x_2;\theta_1)}{f(x_2;\theta_0)} + \ldots + \log\frac{f(x_n;\theta_1)}{f(x_n;\theta_0)} \tag{6}$$

And putting

$$z_i = \log\frac{f(x_1;\theta_1)}{f(x_1;\theta_0)}, \quad i = 1,2,\ldots,n \tag{7}$$

We can write the decision rules as:

(1) If

$$\log B < \sum z_i < \log A \tag{8}$$

the research continues;

(2) If

$$\sum z_i \geq \log A \tag{9}$$

the H_1 hypothesis is accepted and H_0 rejected;

(3) If

$$\sum z_i \leq \log B \tag{10}$$

the H_0 hypothesis is accepted, and H_1 rejected.

By "log" we usually refer to the natural logarithm, because in the continuous case, many densities are exponential and applying a natural logarithm leads to obtaining convenient forms for z_i.

5. Sequential Alternative for the "Hidden Population" Research

Usually, "hidden populations" have reduced volume (from the statistical point of view). In general, specialists agree that "small" populations imply N<5000 units.

Under probabilistic perspective, sampling the individuals and including them in the research, are no longer independent events, so the binomial or the Poisson model cannot be used (Isaic-Maniu, Al., Voda, V. (2009)) for building concrete schemes, and it is more indicated the hypergeometric model. If a population of volume N contains a number of D_1 individuals, the probability for the sample of volume $n: x_1, x_2, \ldots, x_n$ to contain d suspect individuals is established as follows:

$$P_{1,n} = \frac{C_{D_1}^{d} C_{N-D_1}^{n-d}}{C_N^n} \tag{11}$$

Similarly, for the calculus of the probability that the researched group to contain the accepted number of rejected items D_2, we use the relation:

$$P_{2,n} = \frac{C_{D_2}^{d} C_{N-D_2}^{n-d}}{C_N^n} \tag{12}$$

Similar to the sequential plan of large populations, performing the probability ratio according to the Wald test: $P_{2,n} / P_{1n}$, fraction that we compare to the numbers of acceptance and rejection, in order to take a decision. The three conditions in order to take the decision are:

- The condition of accepting hypothesis $H_1 : P \leq P_1$ and closing the investigation concluding that the number of persons in the researched category is under the admissible level of the investigation:

$$\frac{C_{D_2}^{d} C_{N-D_2}^{n-d}}{C_{D_1}^{d} C_{N-D_1}^{n-d}} \leq \frac{\beta}{1-\alpha} \tag{13}$$

- The condition of accepting hypothesis $H_2 : P \geq P_2$ and implicitly of closing the research concluding that the researched population contains a percentage of individuals in the researched category above the tolerated level:

$$\frac{C_{D_2}^{d} C_{N-D_2}^{n-d}}{C_{D_1}^{d} C_{N-D_1}^{n-d}} \geq \frac{1\text{-}\beta}{\alpha} \tag{14}$$

- The condition of continuing the research:

$$\frac{\beta}{1-\alpha} < \frac{P_{2,n}}{P_{1,n}} < \frac{1-\beta}{\alpha} \tag{15}$$

In a situation such as this one, it is necessary to extract a new subject and to recalculate the probability ratio.
To bring the test at the level of practical operating possibility, we proceed in linearizing the probability ratios expressions and elaborating some lines forming a diagram in a plane, in order to use the investigation diagram as a practical work instrument in the research process.
After arithmetic operations, we obtain the equations of the lines describing the acceptance and rejection as follows:

- The acceptance equation:

$$A_c = N\left(\frac{\beta}{1-\alpha}\right)^{\frac{1}{D_2-D_1}} \cdot \left[1-\left(\frac{D_1}{D_2}\right)^{\frac{1}{D_2-D_1}}\right] d + N\left[1-\left(\frac{\beta}{1-\alpha}\right)^{\frac{1}{D_2-D_1}}\right] \tag{16}$$

The rejection equation:

$$R_p = N\left(\frac{1\text{-}\beta}{\alpha}\right)^{\frac{1}{D_2-D_1}} \cdot \left[1-\left(\frac{D_1}{D_2}\right)^{\frac{1}{D_2-D_1}}\right] d + +N\left[1-\left(\frac{1\text{-}\beta}{\alpha}\right)^{\frac{1}{D_2-D_1}}\right] \tag{17}$$

The lines intersect in the *(n,d)* plane, because the angular coefficients are not equal.
The ordinate of the intersection point of the two lines is:

$$d_n = \frac{1}{1-\left(\frac{D_1}{D_2}\right)^{\frac{1}{D_2-D_1}}} \tag{18}$$

Taking into account that $D_1 = NP_1$ and $D_2 = NP_2$, the relation becomes:

$$d_n = \frac{1}{1-\left(\frac{P_1}{P_2}\right)^{\frac{1}{(P_2-P_1)N}}} \tag{19}$$

relation which can be calculated depending on the input variables of the control scheme.

If the volume of the researched population is $N > 100$, and the probable factions $P_1 \le 0{,}1,\ P_2 \le 0{,}2,$ than the previous relation can be approximated through:

$$d_n = \frac{P_2 - P_1}{\ln\frac{P_2}{P_1}} n + \frac{1}{2} \tag{20}$$

So, the intersection point of the acceptance or rejection lines has the ordinate d_n, and abscissa N (figure 1). The graphic representation through cuts is achieved starting from the acceptance and rejection equations and equaling $d_n = 0$. Thus, the intersections with the axis are (Pârțaschi, I., Isaic-Maniu, Al., Vodă, V. - 2008):

$$n_a = N\left[1-\left(\frac{\beta}{1-\alpha}\right)^{\frac{1}{N(P_2-P_1)}}\right] \tag{21}$$

and

$$n_r = N\left[1-\left(\frac{1-\beta}{\alpha}\right)^{\frac{1}{N(P_2-P_1)}}\right] \tag{22}$$

We can see that for $d_n > D_1$ is assured satisfying the condition $\frac{P_{2,n}}{P_{1,n}} \ge \frac{1-\beta}{\alpha}$. In other words, the line $d_n = D_1 = NP_1$ is another rejection limit, that truncates control, accelerating the decision in the case where the maximum admissible level of the characteristic is reached.

6. Numerical Example and Conclusion

In order to perform an investigation among the students of a college regarding tobacco use, we chose the snowball method, completed by the sequential statistical test. The researched characteristic is “tobacco use”. According to the established rule, we considered as “smokers” those students who consume more than two cigarettes a month. The number of students from the targeted upper classes of the college: $N = 450$, a reduced volume population, and therefore we apply the sequential control alternative. The statistical investigation parameters are: the students proportion smaller or equal to 3% (P_1), leads to concluding that the students are”non-smokers”; while the exceeding of the level 5% (P_2) smokers leads to conclude that the students are “smokers” and to begin a special anti-smoking programme within that college; therefore P_1=3%, P_2=5% and the decisional risks are α = 7%, respectively β = 10%.

First we shall establish the acceptance and rejection lines cuts:

$$n_a = 450\left[1-\left(\frac{0{,}10}{0{,}09}\right)^{\frac{1}{450(0{,}05-0{,}05)}}\right] = 98{,}76$$

$$n_r = 450\left[1-\left(\frac{1-0{,}10}{0{,}07}\right)^{\frac{1}{450(0{,}05-0{,}03)}}\right] = -147{,}655$$

The ordinate for the intersection point between the decisional lines is:

$$d_n = \frac{0{,}05 - 0{,}03}{\ln \frac{0{,}05}{0{,}03}} 450 + 0{,}5 = 18{,}19$$

The closing of the research through the sequential method shall be achieved by the line representing the maximum number of "smokers" in the students' population, when the level is considered to be acceptable, resulting that: $D_1 = N*P_1$ = 450 * 0,03 = 13,5.

With the help of the control diagram, we can proceed to the effective investigation of the students. The first investigated group of students let to the following results: after a succession of 56 subjects consuming two cigarettes a month, a "smoker" student has been identified (consuming more than two cigarettes a month), than 55 "non-smokers" (consuming two cigarettes a month), thus reaching the acceptance category, the investigation closing with the final decision of considering the students in the senior college classes as "non-smokers" (statistically, accepting H_0), the proportion of "smokers" in the last two grades if the college being under 3% (P_1=3%).

In conclusion, we can appreciate that statistical methods can complete the sociologic snowball research, through the additional precision that is combined with the intuition of the field researcher, in order to continue or stop an investigation.

REFERENCES

Aaron, D., J., Chang, Y., F., Markovic, N., La Porte, R., E. (2003): *Estimating the lesbian population: a capture-recapture approach*, JECH online, BMJ Publishing, 57, page 207 - 209.

Adams, S., Carter, N., Hadlock, Ch., (2008): *Change in Connectivity in a Social Network over Time: A. Bayesian Perspective*, CONNECTIONS, Vol. 28, Issue 2, page 17 - 27.

Atkinson, R., Flint, J., (2001): *Accessing Hidden and Hard-to-Reach Populations: Snowball Research Strategies*. Social Research Update, No. 33, http://www.soc.surrey.ac.uk/sru/SRU33.html (19. 3. 2007).

Avico, U., Kaplan, C., Korczak, D. and Van Meter, K., (1988): *Cocaine epidemiology in three European Community cities: a pilot study using a snowball sampling methodology*. Brussels, European Communities Health Directorate.

Barlow, J., Davis, H., McIntosh, E., Jarrett, P., Mockford, C.,Stewart-Brown, S., (2007): *Role of home visiting in improving parenting and health in families at risk of abuse and neglect: results of a multicentre randomized controlled trial and economic evaluation*. Archives of Diseases in Childhood, 92, page 229 - 233.

Bârsan-Pipu, N., Isaic-Maniu, Al. Vodă, V., (2002): *Decision procedures in multivariate statistical process control*. Economic Computation and Economic Cybernetics Studies and Research, vol. 36, no. 1 - 4, page 77 - 85.

Berg, S., (1988): *Snowball sampling*, in Kotz, S. and Johnson, N., L., (Eds): Encyclopedia of Statistical Sciences, Vol. 8.

Borgatti, S., P., Everett, M., G. and Freeman, L., C., (2002): *Ucinet - Software of Social Network Analysis User`s Guide*. Harvard: Analytic Technologies, Inc.

Bur, Iv., (1953): *Fundamental Principles of Sequential Analysis*. Industrial Quality Control, vol. 9;

D`Onise, K., Wang, Y., McDermott, R., (2007): *The Importance of Numbers: Using Capture-recapture to make the Homeless Count in Adelaide*: „Australian Journal of Primary Health, 13(1), page 89 - 96.

Day, C. & Davis, H., (2006): *The effectiveness and quality of routine child and adolescent mental health care outreach clinics*. British Journal of Clinical Psychology, 45, page 439 - 452.

Diamond, M., (1993): *Homosexsuality and Bisexuality in Different Populations*. Archives of Sexual Behaviour, 22 (4), page 291 - 310.

Dodds, P., S., Muhamad, R. ,Watts, D., J., (2003): *An experimental study of search in a global social networks*. Science, 301, page 827 - 829.

Duncan, DF. ,White , JB .,Nicholson, T., (2003): *Using internet-based surveys to reach hidden populations*: *Case of nonabusive illicit drug users*`, American Journal of Health Behavior, 27(3), page 208-218.http://www.duncan-associates.com/hiddenpop

Etter, J., F. Perneger, T., V., (2000): *Snowball sampling by mail: application to a survey of smokers in the general population*. International Journal of Epidemiology, 29 (1), page 43 - 45.

Gabor, M., R., (2007): *Types of non-probabilistic sampling used in marketing research*. *"Snowball Sampling"*, Management & Marketing , Nr. 3, page 76 - 91.

Goodman, L., A., *Snowball sampling*. Annals of Mathematical Statistics, 32, page 148 - 170.

Griffiths, P., Gossop, M., Powis, B. Strang, J., (1993): *Reaching hidden populations of drug user by privileged access interviewers: methodological and practical issues*, Addiction, vol. 88, page 1617 - 1626.

Hanneman, R., A., (2001): *Introduction to Social Network Methods*,http://faculty.ucr.edu/hanneman/SOC157/

Hendricks, V., M., Blanken, P. and Adriaans, N., (1992): *Snowball Sampling: A Pilot Study on Cocaine Use*. Rotterdam, IVO.

Isaic-Maniu, Al. Vodă, V., (2009): *The Poisson property in the case of some continuous distributions*. Romanian Statistical Review, No. 2, page 56 - 64.

Johnson, T., P., (2005): *Snowball Sampling*. Encyclopedia of Biostatistics, John Wiley et Sons.

Kirkpatrik, S., Barlow, J., Stewart-Brown, S., Davis, H., (2007): *Working in partnership: user perceptions of intensive home visiting*. Child Abuse Review, 16, page 32 - 46.

Kolaczyk, E., D., (2009): *Statistical Analysis of Network Data*., ISBN: 978-0-387-88145-4.

Koskinen, J., Snijders, T., (2007): *Bayesian inference for dynamic social network data*. "Journal of Statistical Planning and Inference", 137 (2007), page 3930 - 3938.

Kuhar, R., Švab, A., (2008): *Researching hard-to-reach social groups: the example of gay and lesbian population in Slovenia*. Social Research - Journal for General Social Issues, 3, page 529 - 552.

Martino, F., Spoto, A., (2006): *Social network analysis: a brief theoretical review and further perspectives in the study of information technology*. University of Padova, Psychology Journal, 4 (1), page 53 - 86.

Pârțachi, I., Isaic-Maniu, Al., Vodă, V., (2008): *Controlul statistic al calității*. ASE-Bucuresti, ASEM, Chişinău.

Paspalanova, M., (2006): *Identifying and assessing the development of populations of undocumented migrants: the case of undocumented Poles and Bulgarian in Brussels*, Catholic University of Leuven, Belgia.

Peterson, E., (2005): *Evaluation of digital libraries using snowball sampling*, First Monday, volume 10, number 5.

Salganik, M., (2006): *Variance Estimation, Design and Sample Size Calculations for Respondent - Driven Sampling*. Journal of Urban Health, , Nov. 83 (Suppl. 1), page 98 - 112.

Salt, J. ,Hogarth, J., (2000): *Migrant trafficking and human smuggling in Europe* - A review of the evidence in IOM Migrant Traffiking and Human Smuggling in Europe, IOM, Geneva.

Skeldon, R., (2004): *Migration, the Asian financial crisis and its aftermath*, "International Migration and the New Millenium". Aldershat, Ashgate, page 57 - 74 (http://books.google co.uk/books).

Song Hoon Lee, Pan-Jum Kim, Hawoong Jeong, (2006): *Statistical properties of sampled networks*. Physical Review.

Spreen, M., Coumans, M., (2003): *A note on network sampling in drug abuse research*. Maastricht University & Addiction Research Institute Rotterdam, INSNA.
Standfort, T., (1997): *Sampling male homosexuality*. “Researching sexual behavior: Methodological issuer”, page 261 - 275, Bloomington: Indiana University Press.
Streeton, R., Cooke, M. & Campbell, J., (2004): *Researching the researchers: using a snowballing technique*, Nurse Researcher, 12 (1), page 35-47.
Wald, A., (1936): *On some systems of equations of mathematical economics*. Economica, vol. 19, October, page 368 – 463- (translation by Otto Eckstein).
Wald, A., (1947): *Sequential Analysis*, John Wiley, New York.

Figure 1. The sequential control diagram for reduced volume population