Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 49(63), Fascicola 2, 2004

The statistical behaviour of the chaotic signals: application to cryptography

Adriana Vlad[1,2], Adrian Luca[1]

Abstract – The paper is devoted to the analysis of the statistical behaviour of chaotic signals having in view their suitability for the cryptografic applications. The investigation is illustrated using the chaotic signals provided by the logistic function. The procedure of investigation combines information theory notions with statistical inferences based on one or two data sets. The following statistical tools are considered: probability estimations with multiple confidence intervals, test on probability and test on equality between probabilities. The type II statistical error plays a special role in the design of the experimental data size.

Key words: chaotic signals, ergodicity, statistical inferences on probability, type statistical error, noisy information channel.

I. INTRODUCTION

The chaotic behaviour [1], [2] has been noticed for many dynamic systems in discrete time. For such a system, the x_k state at time k depends on the previous state, $x_{k+1} = f(x_k)$. Several types of f functions can be considered in practice. Our paper considers the logistic function:

$$x_{k+1} = Rx_k(1-x_k) \qquad (1)$$

where the R parameter belongs to the (0;4) interval and the x_k value belongs to the (0; 1) interval. The chaotic behaviour is met when $R \geq 3.5699456$. Notice that all the numerical results presented in our paper correspond to the logistic function; however, the investigation can be applied to any chaotic systems.

Let us consider the equation (1). Fig. 1. shows x_k as a function of $k \in \{0,1,...,200\}$, corresponding to $x_0 = 0.113$ and to $R = 3.9$. When considering the same $R = 3.9$ value but different x_0 initials values, a set of different curves are obtained. We consider each of these curves as a sample of a random process. Hence, the sample space is the (0;1) interval in which x_0 is randomly selected [3], [8].

Section II presents the investigations for the first order statistical description (the probability distribution function), [3]-[8]. Some results are displayed in Table 1 and Fig. 2. With this purpose we considered the (0; 1) interval (the chaotic signal values) consisting of $Q = 27$ non-overlapping equal length intervals.

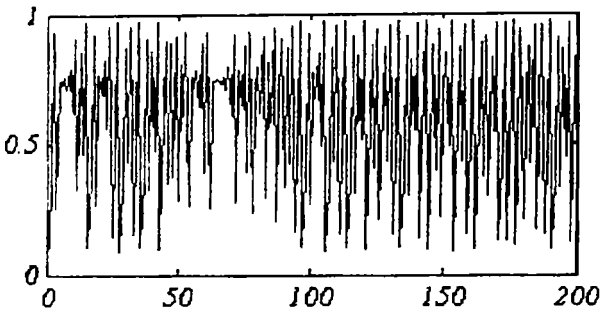

Fig. 1. A sample of the random process which models the chaotic system; $R = 3.9$ and $x_0 = 0.113$.

Section III investigates the conditional probabilities $p(I_j / I_i)$, i.e. the probability that a trajectory (chosen randomly from the ensemble) passes through Ij interval at k_2 iteration on the condition that at k_1 iteration the same trajectory passes through Ii interval. A problem was whether we can speak or not about the statistical independence of the two discrete random variables sampled at k_1 and k_2 iterations.

Section IV is devoted to the applications of chaotic systems to the cryptographic field.

II. FIRST ORDER STATISTICAL DESCRIPTION

Be the random sequence obtained from the chaotic signal when, instead of the continuous random variable sampled at each k iteration, we consider a discrete random variable with Q values. The Q discrete values are assigned to the Q non-overlapping equal length intervals covering the (0; 1) interval of values taken by the chaotic signal. We illustrate the procedure of investigation for $Q = 27$ intervals. We

[1] Faculty of Electronics and Telecommunications, POLITEHNICA University of Bucharest
[2] The Research Institute for Artificial Intelligence, Romanian Academy
corresponding address adriana_vlad@yahoo.com

started the study with $Q = 27$ intervals having in view some immediate applications in enciphering natural texts.

Table 1. First order statistical description of the random process modelling the chaotic systems.

		$k = 200$		$k = 300$		$k = 500$		$k = 2000$	
		$\hat{p}$	$\hat{\varepsilon}_r$	$\hat{p}$	$\hat{\varepsilon}_r$	$\hat{p}$	$\hat{\varepsilon}_r$	$\hat{p}$	$\hat{\varepsilon}_r$
I_1	0.0000; 0.0370	0	0	0	0	0	0	0	0
I_2	0.0370; 0.0741	0	0	0	0	0	0	0	0
I_3	0.0741; 0.1111	0.05000	0.08543	0.05380	0.08219	0.05493	0.08130	0.05160	0.0840
I_4	0.1111; 0.1481	0.03980	0.09626	0.04330	0.09212	0.04450	0.09082	0.04590	0.08935
I_5	0.1481; 0.1852	0.03500	0.10291	0.03540	0.10231	0.03590	0.10156	0.03510	0.10276
I_6	0.1852; 0.2222	0.02810	0.11526	0.02770	0.11612	0.03030	0.11087	0.02450	0.12367
I_7	0.2222; 0.2593	0.02500	0.12116	0.02320	0.12717	0.02600	0.11996	0.02290	0.12802
I_8	0.2593; 0.2963	0.02020	0.13650	0.02240	0.12948	0.02200	0.13067	0.02250	0.12918
I_9	0.2963; 0.3333	0.02170	0.13160	0.02190	0.13098	0.02140	0.13253	0.02090	0.13414
I_{10}	**0.3333; 0.3704**	**0.07290**	**0.06989**	**0.06990**	**0.07149**	**0.07000**	**0.0750**	**0.07110**	**0.07084**
I_{11}	0.3704; 0.4074	0.04290	0.09257	0.04380	0.09155	0.03910	0.09716	0.04250	0.09303
I_{12}	**0.4074; 0.4444**	**0.04990**	**0.08552**	**0.04750**	**0.08776**	**0.04840**	**0.08690**	**0.04800**	**0.08728**
I_{13}	0.4444; 0.4814	0.02500	0.12240	0.02800	0.11164	0.02740	0.11677	0.02410	0.12472
I_{14}	0.4815; 0.5185	0.02670	0.11833	0.02990	0.11164	0.03000	0.11145	0.02900	0.11341
I_{15}	0.5185; 0.5556	0.02420	0.12445	0.02650	0.11870	0.02500	0.12240	0.02720	0.11721
I_{16}	0.5556; 0.5926	0.02380	0.12552	0.02310	0.12745	0.02200	0.13067	0.02330	0.12689
I_{17}	0.5926; 0.6296	0.02350	0.12634	0.02400	0.12498	0.02360	0.12606	0.02830	0.11484
I_{18}	**0.6296; 0.6667**	**0.02580**	**0.12043**	**0.02310**	**0.12745**	**0.02350**	**0.12634**	**0.02650**	**0.11879**
I_{19}	0.6667; 0.7037	0.02620	0.11949	0.02460	0.12341	0.02310	0.12745	0.02290	0.12802
I_{20}	0.7037; 0.7407	0.02680	0.11810	0.02430	0.12419	0.02360	0.12606	0.02300	0.12774
I_{21}	0.7407; 0.7778	0.02665	0.11879	0.02370	0.12579	0.02470	0.12326	0.02680	0.11810
I_{22}	**0.7778; 0.8148**	**0.02940**	**0.11261**	**0.02630**	**0.11925**	**0.02885**	**0.11443**	**0.02870**	**0.11402**
I_{23}	0.8148; 0.8519	0.03260	0.10676	0.03080	0.10994	0.02850	0.11443	0.03000	0.11144
I_{24}	0.8519; 0.8889	0.06710	0.07308	0.07170	0.07052	0.06680	0.07325	0.06360	0.07520
I_{25}	0.8889; 0.9259	0.07270	0.06999	0.07040	0.07122	0.06780	0.07267	0.07210	0.07031
I_{26}	0.9259; 0.9630	0.10590	0.05690	0.11020	0.05569	0.11000	0.05575	0.11090	0.05549
I_{27}	0.9630; 1.0000	0.07780	0.06747	0.07450	0.06908	0.08300	0.06514	0.07860	0.06710

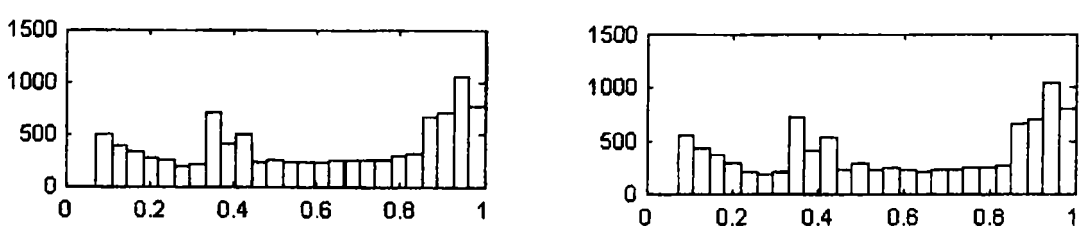

Fig. 2 Histograms for the frequencies distribution of Table 1: $Q = 27$ intervals, at $k = 200$ (left) and at $k = 500$ (right); on the vertical axis– the occurrences of the intervals

We started the investigation with verifying the first order stationarity of the chaotic signal by considering the Q discret intervals. This implies a comparative study of the discrete random variables sampled at different iterations.

We determine the probablity that at the k iteration the chaotic signal passes through a certain Ii interval (chosen from the Q possible intervals). Then, we try to verify if this probability depends or not on the k iteration (the sampling time) while preserving the same Ii interval. For the statistical inferences used the experimental data should comply with the *i.i.d.* model (*i.e.* observations coming up from independent and identically distributed random variables). Moreover, for the statistical tests we compared independent data sets.

The first experimental results are presented in Table 1 where we considered four different iterations: $k = 200$, $k = 300$, $k = 500$, $k = 2000$ and $N = 10000$ trajectories for each sampling time. Note that for each k sampling time we used $N = 10000$ different trajectories, generated by different initial conditions (randomly chosen from $(0; 1)$ interval)). Hence, for the four iterations $k = 200$, $k = 300$, $k = 500$, $k = 2000$ we had at our disposal four independed *i.i.d.* data sets (that means we generated 40000 different trajectories of the chaotic signal).

For example for the $I_{10} = (0.3333; 0.3704)$ interval, at $k = 200$, the estimated value of the p probability that the chaotic signal passes through I_{10} is $\hat{p} = m / N = 0.07290$ (m is the occurrence number of the investigated interval in the considered *i.i.d.* data set).

Each time we experimentally checked-up the de Moivre-Laplace conditions in the form $\sqrt{N\hat{p}(1-\hat{p})} \geq 14$, [4]-[7]. As a consequence we can say that the p true (unknown) probability lies inside the $(\hat{p}*(1-\hat{\varepsilon}_r); \hat{p}*(1+\hat{\varepsilon}_r)) = (0.06780; 0.07799)$ interval computed with $1-\alpha = 0.95$ statistical confidence level;

$\hat{\varepsilon}_r = z_{\alpha/2} * \sqrt{\bar{p}(1-\bar{p})/N} = 0.06989$ is the relative experimental error, where $z_{\alpha/2} = 1.96$ is the $\alpha/2$ point value corresponding to the standard Gaussian law (of 0 mean and 1 variance).

For the same $I_{10} = (0.3333; 0.3704)$ interval, but at $k = 500$ iteration, the estimated value is $\hat{p} = 0.07000$, the 95% confidence interval is $(0.06499; 0.07500)$ and the relative experimental error $\hat{\varepsilon}_r = 0.0750$. It can be noticed that the two confidence intervals for the probability overlap; this brings some evidence in the favor of the stationarity assumption.

The fact that the confidence intervals overlap encouraged us to a more detailed investigation.

Thus, we continued the study with applying the test on the equality between two probabilities (see Appendix).

We succesively compared the two data sets (one for $k = 200$ and another one for $k = 500$) for each Ii interval ($i = 1 \div 27$ and $j = 1 \div 27$) in Table 1.

Table 2 presents the results only for I_{10}, I_{12}, I_{18} and I_{22} intervals. All the four tests were passed; the test values and the decisions are shown in Table 2. As a conclusion, the stationarity assumption is again sustained.

Table 2. Experimental values for the test on the equality between two probabilities

Test: T_1	I_{10}	I_{12}	I_{18}	I_{22}
$\|z\|$	0.0819	1.2248	1.2910	1.4523
H_0 / H_1	H_0	H_0	H_0	H_0
β	$2*10^{-11}$	$9.8*10^{-6}$	0.1532	0.0328

Because all tests are passed, the β probability of type II statistical error (that means H_0 accepted, although the two compared probabilities are not equal) is important. It was computed according to (4) (see Appendix).

Fig. 3 shows the β values as a function of p_1.

There are three plots for δ values: $\delta = 0.1$, $\delta = 0.15$, $\delta = 0.2$, $N_1 = N2 = N = 10000$, $\alpha = 0.05$.

Table 2 presents the β values for the corresponding intervals, when $\delta = 0.20$. For $\delta = 0.10$, β values are much larger. For a better accuracy (low values for β while $\delta < 0.15$) we need to resume the experiment generating much more trajectories of the chaotic signal.

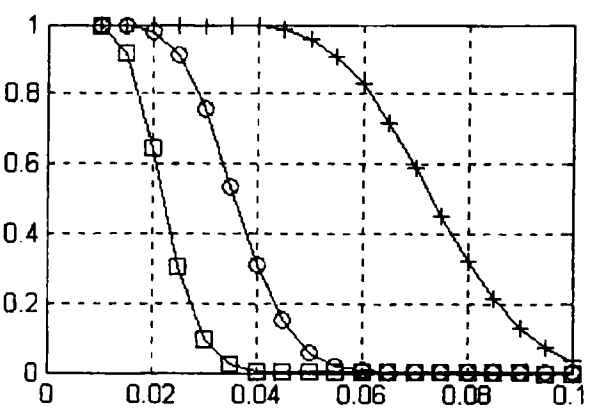

Fig. 3 The type II error size for the test on equality between probabilities. On the horizontal - p_1 probablity ; on vertical - β values. The curves corresponding to: $\delta = 0.1$ - "+", $\delta = 0.15$ - "o" and $\delta = 0.2$ - "□"

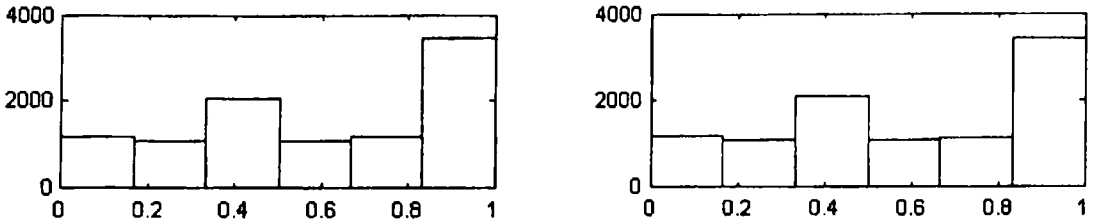

Fig 4. Frequency distribution representation: the histograms if we "discretize" in $Q = 6$ intervals at $k = 200$ (left) and at $k = 500$ (right), on the vertical axis – the occurrences of the intervals

Fig. 2 shows histograms coresponding to the frequency distribution from Table 1. Instead of the relative frequencies of the intervals, the histograms are constructed on the basis of intervals occurences.

The study was resumed for Q=6 intervals. This number of Q=6 intervals could be of some interest in the cryptographic field when two iterations are simultaneously considered and assigned to an alphabet character of the natural language (for exemple letters, punctuation marks). Fig. 4 presents histograms for the random process modelling the chaotic signal.

For a temporal description we generated several individual trajectories of the chaotic signal for $L = 10000$ iterations. We measured how many times the investigated trajectory (randomly chosen from the ensamble) passes through a certain interval of values; be $\tilde{m}$ the occurrence number. The relative occurrence number $\tilde{p} = \tilde{m} / L$ was computed for each I_i interval of values (I_i is the same from Table 1 where we "discretized" the (0; 1) interval in $Q = 27$ non-overlapping intervals of equal length).

Another issue was if $\tilde{p} = \tilde{m} / L$ (the temporal relative frequency of the investigated interval) lies inside of the confidence interval for the probability corresponding to the same I_i investigated interval (at k iteration).

As an illustration we used three curves with initial condition: $x_0 = 0.31$, $x_0 = 0.456$ si $x_0 = 0.758$ and four investigated intervals: I_{10}, I_{12}, I_{18} and I_{22} (see Table 3). We computed the temporal relative frequency $\tilde{p} = \tilde{m} / L$ of the investigated interval for each trajectory. For example for the trajectory with $x_0 = 0.31$ the temporal relative frequency corresponding to I_{10} interval is $\tilde{p} = 0.0725$.

Looking at Table 1, the 95% confidence interval for the probability assigned to I_{10} interval at $k = 500$ was $(0.06499; 0.07500)$. We can see that $\tilde{p}$ lies inside this confidence interval for the probability.

We resumed this type of investigation for each I_i interval and several trajectories; all the numerical results sustained the ergodicity assumption of the first order distribution function.

We continued the verify this type of ergodicity by using a test of probability [4], [6], [7]. In this test the *i.i.d.* data sets is the same we used in Table 1 for a fixed k (the considered iteration) and the theoretical probability of the test was the temporal value for the corresponding I_i interval in Table 3.

Table 3. Temporal description

	$x_0 = 0.31$	$x_0 = 0.456$	$x_0 = 0.758$
I_{10}	0.0725	0.0711	0.0708
I_{12}	0.0481	0.0487	0.0483
I_{18}	0.0214	0.0219	0.0213
I_{22}	0.0253	0.0273	0.0234

Thus, the null hypothesis H_0 has the form $H_0 : p = p_0$ where p_0 denotes the temporal probability obtained for a certain trajectory.

We successively applied this test (Table 4) for I_{10}, I_{12}, I_{18} and I_{22} interval considering the *i.i.d.* data sets obtained at $k = 500$. The theoretical p_0 probabilities are those from Table 3 and the trajectory with initial condition $x_0 = 0.31$. All the tests were passed, thus sustaining again the ergodicity assumption.

Table 4. Test of probability

Test	I_{10}	I_{12}	I_{18}	I_{22}
p_0	0.0725	0.0481	0.0214	0.0253
$\|z\|$	0.9641	0.1402	0.8983	0.7642
H_0/H_1	H_0	H_0	H_0	H_0

III. SECOND ORDER STATISTICAL DESCRIPTION

Here, we again consider the (0; 1) interval of values of the chaotic signal discretized in Q non-overlapping intervals of equal length. For the second order statistical description we shall consider simultaneously two iterations (k_1 and k_2).

This leads to the noisy information channel shown in Fig. 5. Fig. 5 illustrates our procedure of investigation considering $Q = 6$ intervals. As a consequence, the input space $X = \{x_1, ..., x_i, ..., x_6\}$ corresponds to the I_i interval at k_1 iteration and the output space $Y = \{y_1, ..., y_j, ..., y_6\}$ corresponds to the I_i interval at the $k_2 = k_1 + k$ iteration.

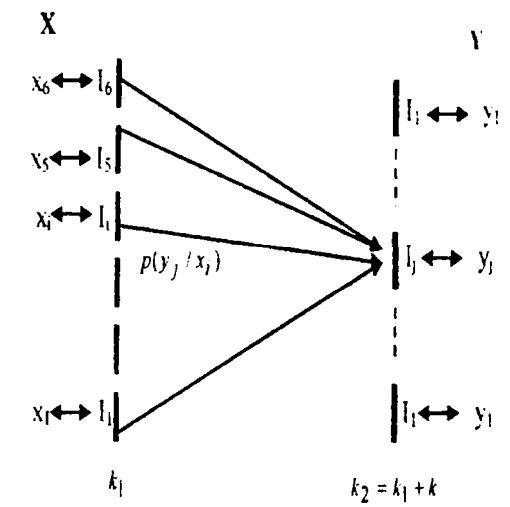

Fig. 5. The channel diagram (transition graph)

Table 6. Noise matrix estimation (proportions)

$p(y_j/x_i)$	y_1	y_2	y_3	y_4	y_5	y_6
x_1	0.1020	0.0977	0.2239	0.1011	0.1236	0.3518
x_2	0.1229	0.1066	0.2051	0.1192	0.1048	0.3415
x_3	0.1184	0.1055	0.2129	0.[illegible]	[illegible]	[illegible]
x_4	0.1020	0.1067	0.2229	0.1199	0.1265	0.3220
x_5	0.1108	0.1092	0.2024	0.1117	0.1209	0.3451
x_6	0.1142	0.1128	0.1939	0.1053	0.1205	0.3533
$p(y_j)$	**0.1157**	**0.1107**	**0.2010**	**0.1059**	**0.1191**	**0.3476**

The problem is whether the random discrete variables sampled at two iteration are statistically independent or not; in the affirmative case, for what $k = k_2 - k_1$ distance we can think of independence.

With this purpose we verify if the following relation $p(y_j / x_1) = ... = p(y_j / x_6) = p(y_j)$ is valid or not. $p(y_j)$ is the probability that the chaotic signal passes through I_j interval at k_2 iteration and $p(y_j / x_i)$ is the probability that a trajectory (chosen randomly from the ensemble) passes through Ij interval at k_2 iteration on the condition that at k_1 iteration the same trajectory passes through Ii interval.

For a quick decision concerning their independence, we computed the mutual information, (2).

$$I(X;Y) = H(X) - H(X/Y) = H(Y) - H(Y/X) \quad (2)$$

The Table 6 shows the conditional probabilities $p(y_j / x_i)$ for $k_1 = 200$ and $k_2 = k_1 + 50 = 250$.

The mutual information corresponding to Table 6 is very low: $I(X;Y) = 0.001642$. This suggest the independence between the input and the output (also revealed in Table 6 by the equality between probabilities estimates $p(y_j / x_i) \equiv p(x_i)$, $i = 1 \div 6$ and $j = 1 \div 6$).

We resumed this procedure of verifying the statistical independence. Table 7 shows some results that indicate a $k = k_2 - k_1$ distance for which we can speak about independence; this happens for $k \geq 30$. The investigation was carried out on $N = 10000$ trajectories.

This procedure based on the noisy information channel assigned to "discretized" chaotic signal was further resumed for $Q = 27$ intervals. Some results are presented in Table 8.

Table 7. The mutual information of the information channel

	k_1	k_2	$I(X;Y)$
$k = 10$	200	210	0.016035
	300	310	0.013822
$k = 30$	200	230	0.001164
	300	330	0.002171
$k = 50$	200	250	0.001642
	300	350	0.001812

This time the mutual information was computed considering a large number of trajectories: $N = 50000$. We can notice that when we can speak about independence the k distance is larger than for $Q = 6$ (Table 7).

Table 8. The mutual information of the channel

	k_1	k_2	$I(X;Y)$
$k = 10$	300	310	0.136066
$k = 20$	300	320	0.011214
$k = 50$	300	350	0.008182
$k = 500$	300	800	0.008017
$k = 1000$	300	1300	0.007521

We also computed the conditional probability and the mutual information $I(X;Y)$ for different k_1 and $k = k_2 - k_1$. All results sustain the second order stationarity for the discrete random process assigned to the chaotic signal.

IV. CONCLUSION AND OPEN PROBLEMS

This paper suggests how to obtain from the chaotic signal a stationary discrete information source (and according to case practically zero-memory) having the same symbols as a printed natural language. For example we can generate an information source with $Q = 27$ symbols that may correspond to printed Romanian (the alphabet whitout blank and punctuation marks), where we omit some very low frequency characters. The message generated by this information source (provided by the chaotic signal) may be a key in a various enciphering methods.
An immediate example that can be further used in different variants is to make a summation modulo Q (successively for each character) between the plaintext and the key. On the basis of the entropy (redundancy) of the information source corresponding to the key and also using some knowledge about the entropy of natural language (the plaintext) we can evaluate the performance of the cipher.

ACKNOWLEDGEMENT

The autors wish to thank to Professor Jean-Pierre Barbot, the head of the ECS research group from ENSEA, Cergy, France where this study was started, for his assistance and support.

APPENDIX

In the appendix we briefly present the test on the equality between two probabilities:
Be there two samples each complying with i.i.d. statistical model with the sample size $N_1 = N_2 = N = 10000$. Denoting by m_1 the number of successes of the event in the first data sample, the probability estimate is $\hat{p}_1 = m_1 / N_1$. Similarly, in the second data sample, the probability estimate is $\hat{p}_2 = m_2 / N_2$. The two statistical hypotheses (null hypotheses H_0/alternative hypotheses H_1) are: $H_0 : p_1 = p_2$ and $H_1 : p_1 \neq p_2$. We have to verify whether the two estimates $\hat{p}_1$ and $\hat{p}_2$ derive from the same theoretical probability. We apply the test based on the z test value defined in (3):

$$z = (\hat{p}_1 - \hat{p}_2) / \sqrt{p_1(1-p_1)/N_1 + p_2(1-p_2)/N_2}\ ,$$
$$\text{where } p_1 = p_2 = p \cong (m_1 + m_2)/(N_1 + N_2) \qquad (3)$$

If $|z| \leq z_{\alpha/2}$ (where the $z_{\alpha/2}$ is $\alpha/2$ point value corresponding to the standard Gaussian law of 0 mean and 1 variance) then we shall consider that the two probabilities are equal. Otherwise, i.e. when $|z| > z_{\alpha/2}$, we reject the equality hypothesis at an α significance level.
Type II error means not to reject H_0 although it is false. This happens when the test value passes the test, however $p_1 \neq p_2$. The probability of this situation depends on the p_1 and $p_2 = p_1(1-\delta)$ value for fixed α, N_1 and N_2. It is denoted by $\beta(p_1, p_2)$ and is computed according to equation (4):

$$\beta(p_1, p_2) = \int_{-\hat{\sigma}^* z_{\alpha/2}}^{-\hat{\sigma}^* z_{\alpha/2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-(p_1-p_2))^2}{2\sigma^2}\right)dx \qquad (4)$$

where:

$$\hat{\sigma} = \sqrt{p(1-p)(1/N_1 + 1/N_2)} \quad ;$$

$$\sigma = \sqrt{p_1(1-p_1)/N_1 + p_2(1-p_2)/N_2}\ .$$

REFERENCES

[1] Al. Serbanescu, *Aplicatii ale sistemelor haotice in comunicatii*, Editura Academiei Tehnice Militare, Bucuresti, 2004
[2] M S. Baptista, "Cryptography with chaos", *Physics Letters*, A 240, (1998), pp. 50-54
[3] Al. Spataru, *Fondaments de la theorie de la transmission de l'information*, Presse Polytechniques Romandes, 1987
[4] J. Devore, *Probability and Statistics for Engineering and the Sciences*, second edition, Brooks/Cole Publishing Company, Monterey, California, 1987
[5] A. Vlad, B. Badea, *Metode statistice in prelucrarea datelor Estimarea statistica*, Editura Paideia, Bucuresti, 2002
[6] A. Vlad, B. Badea, M. Mitrea, *Metode Statistice in Prelucrarea Informatie. Compendiu si Aplicatii*, Editura Metropol, Bucresti, 1999
[7] V. Craiu, *Verificarea Ipotezelor Statistice*, Editura Didactica si Pedagogica, Bucuresti, 1972
[8] A. Vlad, M. Ferecatu, M. Mitrea, *Teoria transmisiunii informatiei II: elemente teoretice ilustrate in MathCAD*. Ed. Paideia, Bucuresti 2002