

ABSTRACT

Teza atinge diverse domenii ca teoria informației, analiza și prelucrarea semnalelor aleatoare, criptografie bazată pe sisteme haotice și prelucrarea limbajului natural, domenii ce fac parte din programul de învățământ și cercetare al Facultății de Electronică, Telecomunicații și Tehnologia Informației din cadrul Universității Politehnica din București și sunt studiate în cadrul colectivului de lucru coordonat de Prof. Dr. Ing. Adriana Vlad. Teza s-a concentrat pe metodele statistice care au realizat legătura dintre aceste domenii, urmărind: interpretarea datelor experimentale (implicit modalități de colectare a datelor), decizii privind o serie de parametri, legături existente între date (dependența/independența statistică, corelație/necorelație) și legi de repartiție. De asemenea, teza a urmărit și construirea de instrumente și procedee statistice ce pot fi folosite pentru interpretarea datelor, dar și pentru obținerea unor sisteme precum generatoare de numere aleatoare performante. O mare parte din conținutul tezei se concentrează pe investigații privind diversele semnale haotice, interesante datorită caracterului lor ergodic. Rezultatele cercetării s-au adăugat rezultatelor obținute anterior de colectivul de lucru și furnizează o serie de răspunsuri la diferite probleme statistice, multe dintre ele fiind cunoscute ca probleme dificil de explicat matematic sau chiar ca probleme deschise în literatură. Printre aceste răspunsuri putem găsi elemente de noutate absolută referitoare la independența statistică pe domeniul continuu atunci când legile de probabilitate sunt necunoscute. În acest sens, teza aduce unele rezultate noi, îmbogățind un instrument statistic existent (testul de independență) atât la nivel conceptual cât și practic. O altă problemă investigată în teză a fost analiza dependenței statistice dintre estimatorilor de medie și dispersie cu scopul de a găsi un test de corelație/necorelație a datelor și implicit de detectare a simetriei legilor de probabilitate. De asemenea, s-a conceput un ansamblu de unelte statistice pentru evaluarea calității generatoarelor de numere binare pseudoaleatoare. Aceste ansambluri de metode au fost folosite pentru verificarea unui generator de numere pseudoaleatoare obținut din semnale haotice folosind un concept nou dezvoltat în grupul de cercetare pornind de la procedura running-key. O altă problemă abordată este reprezentată de conceperea unui ansamblu de metode statistice cu care să se poată investiga noțiunea de sensibilitate la parametrul de control ale semnalelor haotice. Teza prezintă și o serie de analize și prelucrări statistice pe un corpus lingvistic care a vizat relevanța studiilor statistice în modelarea limbii române scrise.

Toate aceste studii și aplicații au fost dezvoltate folosind metode statistice uzuale (teste de probabilitate, metoda Monte-Carlo, teste de concordanță Hi-Pătrat, teste de concordanță Kolmogorov-Smirnov) grupate în ansambluri originale, la care s-au adăugat metode și instrumente de lucru mai speciale, originale, de stabilire a independenței statistice între variabile aleatoare de tip continuu și de analiză de simetrie/asimetrie a legilor de probabilitate.

The thesis touches various areas such as information theory, the random signals analysis and processing, cryptography based on chaotic systems and natural language processing, areas that are part of the teaching and research program of the Faculty of Electronics, Telecommunications and Information Technology, University Politehnica of Bucharest and are studied by the team coordinated by Prof. Dr. Ing. Adriana Vlad. The thesis is focused on statistical methods that have made the connections between these areas, following: interpretation of the experimental data (including how to collect data), decisions regarding a series of parameters, relations between data (statistical dependence / independence, correlation / uncorrelation) and distribution laws. The thesis is also focused on designing tools and statistical methods that can be used for data interpretation and for building systems such as statistically performant random number generators. A large part of the thesis focuses on the investigation of the different chaotic signals, due to their interesting ergodic characteristic. The obtained results were added to the previous results of the research team providing answers to various statistical problems, many of which are known to be difficult to explain mathematical problems or even open problems in the literature. Among these answers there are important new elements regarding the statistical independence of the continuous random variables when the probability laws are unknown. In this respect, the thesis brings some new results, enriching an existing statistical procedure (the independence test) at both conceptual and practical levels. Another problem investigated in this thesis was to analyse the statistical dependence between the variance and mean estimators aiming to obtain a data correlation / uncorrelation test and, implicitly, a skewness test. Also it was designed an ensembles of statistical tools to assess the quality of binary pseudorandom number generators. These combinations of methods have been used to analyse a pseudorandom number generator obtained from a chaotic signal using the running-key procedure which is a newly concept developed by the research group. Another problem treated in this thesis is the design of a set of statistical methods capable to investigate the impact of control parameter sensitivity on the statistical behaviour of the chaotic maps. The thesis presents a large statistical analysis on a literary corpus of printed Romanian language aiming to underline the relevance of statistical studies on the language modelling.

The overall mentioned studies and applications have been developed by means of usual statistical methods (probability tests, Monte - Carlo analyses, Chi-square concordance tests, Kolmogorov - Smirnov tests) grouped into original ensembles, to which was added special and original methods and tools for deciding upon the statistical independence between continuous random variables and also for deciding upon the skewness of the probability laws.