

Use of ELVIS II platform for random process modelling and analysis of its probability density function

Yu.S. Maslennikova, I.S. Nugmanov

Kazan Federal University, Kremlevskaya str.18, Kazan 420018, Russia

E-mail: JSMaslennikova@kpfu.ru, nugml3@yandex.ru

Abstract. The problem of probability density function estimation for a random process is one of the most common in practice. There are several methods to solve this problem. Presented laboratory work uses methods of the mathematical statistics to detect patterns in the realization of random process. On the basis of ergodic theory, we construct algorithm for estimating univariate probability density distribution function for a random process. Correlational analysis of realizations is applied to estimate the necessary size of the sample and the time of observation. Hypothesis testing for two probability distributions (normal and Cauchy) is used on the experimental data, using χ^2 criterion. To facilitate understanding and clarity of the problem solved, we use ELVIS II platform and LabVIEW software package that allows us to make the necessary calculations, display results of the experiment and, most importantly, to control the experiment. At the same time students are introduced to a LabVIEW software package and its capabilities.

1. Introduction

Probability density function (PDF) estimation is one of the major concern in areas such as signal detection and recognition, machine learning, neural networks, digital signal processing and computer vision. On the one hand, it offers a flexible way to investigate the properties of a given data set and provides a solid basis for efficient data mining tools. On the other hand, it is crucial in unsupervised learning tasks and Bayesian inference and classification [1].

Probability density estimation for random variables and stochastic processes is one of the key sections in any course of statistics and statistical physics, in particular. Recently, many mathematical tools provide special libraries for the empirical evaluation of probability density function in different ways. But often, the use of ready-made functions does not allow students to fully understand the essence of the matter. The most effective method of development of educational material for this section is to perform laboratory practicum, when students independently perform the measurement of stochastic processes dynamics and then estimate various statistical parameters of these processes.

The National Instruments Educational Laboratory Virtual Instrumentation Suite (NI ELVIS) is an educational design and prototyping platform based on NI LabVIEW. The NI ELVIS teaching platform takes students from discovering engineering theory to practical hands-on experience with industry relevant technology and can be used in laboratory practicum of statistical radiophysics [2].

In this paper, we propose an example of a laboratory practicum for a probability density function estimation using the ELVIS II platform capabilities platform and LabVIEW software package. This laboratory practicum is successfully used during statistical radiophysics course at the Kazan Federal University for many years.



2. PDF function estimation for ergodic random processes

Below we describe theoretical backgrounds for the theory of ergodic stochastic processes and methods of probability density function estimations.

According to the ergodic theorem, mean and correlation function of the stationary process $\xi(t)$ can be calculated by averaging over time the values of a single realization of $v(t)$ for a continuous random process or a single realization $v(n)$ for a random sequence $\xi(n)$.

A sufficient condition for the applicability of the ergodic theorem is the lack of correlation between the two values a stationary random process (in broad sense) divided by an infinite length of time, i.e.,

$\lim_{T \rightarrow \infty} \frac{1}{T} R(\tau) = 0$, where $R(\tau)$ is autocorrelation function of the process $\xi(t)$. Another sufficient condition for the applicability of the ergodic theorem is the Slutsky equation [3]:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T R(\tau) d\tau = 0, \quad (1)$$

where T is the time of observation for the random process.

For ergodic process averaging value is equal to

$$\langle \xi(n) \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} v(n), \quad (2)$$

where N is the number of elements in the observed series. Practically, number of the elements N in the observed sequence is limited.

Proposed method is based on the measurements of the time of the realization $v(t)$ for an ergodic stochastic process $\xi(t)$ between two fixed levels v_i and $v_i + \Delta v$ over a sufficiently large time interval of observation T (figure 1b).

Let us estimate the probability density function of noise with sufficiently large T . First of all, we write a new random process $\eta(t)$ (figure 1), depending on $\xi(t)$, with values:

$$y(t) = \begin{cases} 1, & \text{if } v_i < \xi(t) \leq v_i + \Delta v, \\ 2, & \text{if } v_i > \xi(t) \text{ or } \xi(t) > v_i + \Delta v. \end{cases} \quad (3)$$

Mathematical expectation of the process $\eta(t)$ in an arbitrary moment in time:

$$M[\eta(t)] = P(v_i < \xi(t) \leq v_i + \Delta v). \quad (4)$$

For sufficiently small Δv , ignoring the change in the probability density function $w(v)$ in the interval $(v_i, v_i + \Delta v)$, we have $M[\eta(t)] \approx w(v) \Delta v$. The process $\eta(t)$ is also ergodic.

Mean value for the process $\eta(t)$:

$$\langle \eta(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T y(t) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{j=1}^n \tau_{ij}, \quad (5)$$

where $y(t)$ - realization of the random process, $\eta(t)$ - pulses with unit amplitude and a random duration of τ_{ij} , τ_{ij} - time spent in the interval $(v_i, v_i + \Delta v)$ for process $\xi(t)$, n - number of intervals of random length, when $v_i < \xi(t) \leq v_i + \Delta v$ for an observation time T .

Values of τ_{ij} and n are random and depend on the one-dimensional density $w(v)$. Then suppose that the measurement time T so great that we can write:

$$\langle \eta(t) \rangle \approx \frac{1}{T} \sum_{j=1}^n \tau_{ij}. \quad (6)$$

Using the ergodic property of the process $\eta(t)$:

$$\langle \eta(t) \rangle \approx E[\eta(t)] \Rightarrow w(v_i) \Delta v \approx \frac{1}{T} \sum_{j=1}^n \tau_{ij}. \quad (7)$$

Time of observation T , total duration of the pulses of unit amplitude $\sum_{j=1}^n \tau_{ij}$ voltage $(v_i, v_i + \Delta v)$ can registered using different devices. Thus, it is possible to experimentally estimate one-dimensional probability density function of the noise belonging to the interval $(v_i, v_i + \Delta v)$ and replace the exact density distribution $w(v_i)$ with its estimation $\hat{w}(v_i)$ in equation (7). Each level of quantization v_i corresponds estimation $\hat{w}(v_i)$, calculated according to the equation:

$$\hat{w}(v_i) \approx \frac{1}{\Delta v T} \sum_{j=1}^n \tau_{ij}. \quad (8)$$

The experimental estimations $\hat{w}(v_i)$ can be used to construct a histogram of the probability density function (figure 1a).

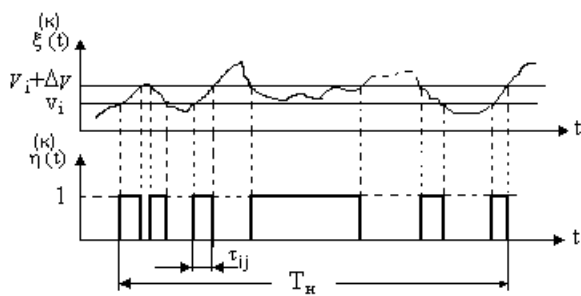


Figure 1a. Original stochastic process $\xi(t)$ and binary random process $\eta(t)$.

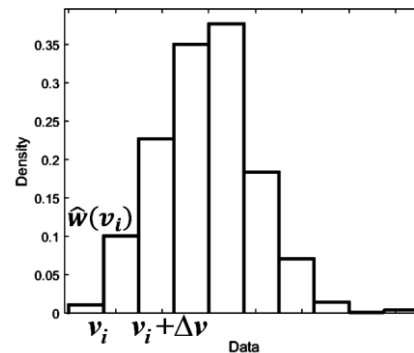


Figure 1b. Empirical estimation of a probability density function

For experimental evaluation of the density function $w(v_i)$ we must justify the choice of the observation time T (or the number of counts N during the observation T), the number of quantization levels of the process $\xi(t)$ and the size of the quantization interval Δv .

Statistical hypothesis testing of distribution law requires independent sampling in time. If the observations are dependent, then one-dimensional empirical distribution function $\hat{w}(v_i)$ will be distorted.

It is impossible to specify the exact value Δt for the test process used in this laboratory practicum. We need to select the time Δt from the condition of uncorrelated values of noise, separated by an interval of time Δt , though it is impossible for certain to state that these two values will be independent. However, the theory has proven that if the distribution of random variables is normal, then uncorrelated random variables are independent.

3. Random process modelling and PDF function estimation using ELVIS II platform and LabVIEW software package

In this laboratory practicum use a personal computer, NI ELVIS II laboratory platform and a test breadboard. For numerical analysis of the probability density function we use virtual apparatus, created with a universal programming system LabVIEW. The main view of this program is shown in the figure 2.

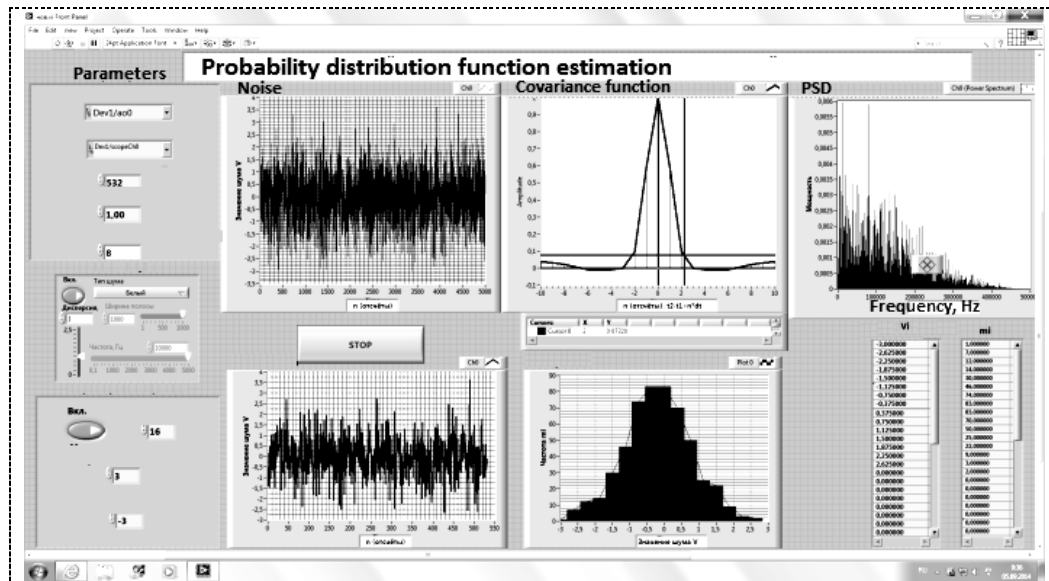


Figure 2. The main view of laboratory practicum for the PDF function estimation (in LabVIEW software package).

We use the noise generation in the LabVIEW environment and we output the noise from digital to analog devices (pin A0 0 connector) to the first channel oscilloscope CH 0 (Channel 0 SCOPE). The virtual apparatus has the following tasks:

- Generate a random process (“Noise” on the figure 2);
- Show the covariance function of the generated data (“Covariance function” on the figure 2);
- Show the power spectral density (“PSD” on the figure 2);
- Visualizes a data sample.
- Display a histogram of the distribution density for the noise values.

At the beginning of this work we need to calculate the observation time. Let’s describe basics of this procedure.

If the generated set of noisy data consists of N samples from the general population with the distribution $w(v)$, then the difference between the probability density \hat{p}_i for v_i interval and theoretical probability p_i should be smaller than ε and would be determined only by random errors. The probability of such deviation is written as:

$$P(|\hat{p}_i - p_i| > \varepsilon) = \alpha, \quad (9)$$

where ε is infinitesimal value.

According to the law of large numbers [2] the distribution of random variable $\psi = \hat{p}_i - p_i$ when $N \rightarrow \infty$ corresponds to the normal distribution law with a mean value $E[\Psi] = 0$ and variance $D[\Psi] = p_i(1 - p_i)/N$.

The probability of the case, that a random variable $\psi = \hat{p}_i - p_i$ is in range of $(-\varepsilon, \varepsilon)$, is equal to

$$P(-\varepsilon < \hat{p}_i - p_i < \varepsilon) = \frac{2}{\sqrt{2\pi}} \int_0^{\frac{\varepsilon}{\sigma}} \exp\left(-\frac{x^2}{2}\right) dx = 1 - \alpha, \quad (10)$$

where $z_\alpha = \varepsilon \sqrt{\frac{N}{p_i(1 - p_i)}}$.

Probability integral (10) is tabulated [4]. With known confidence level $(1-\alpha)$ and using tabulated values of the normal distribution or by numerical integration, we can find the value of z_α . Using z_α , we

$$\text{get } N = \frac{z_\alpha^2 p_i (1 - p_i)}{\varepsilon^2} \leq \frac{z_\alpha^2}{\varepsilon^2}.$$

The value of ε is usually set based on practical needs.

After estimation of the number observation points N , we generate the random set of the white noise data. While generating we need to control the covariance function of the noise and its probability distribution function. It is needed to be sure that our measurements are not correlated therefore independent. After that, we can see the histogram of probability distribution function. If everything is good, we export generated values for the further analysis.

4. Testing hypothesis of distribution law

Experimentally found PDF function must be identified with one of the distribution law. We can make a hypothesis H_0 of the type of distribution law and the proposed distribution of $w(v)$ can be either true and false. We introduce a metrics D of the difference between the empirical distribution function and the hypothetical distribution.

Because of the empirical distribution function $w(v)$ corresponds to random variables then the metrics D is a random variable too. In this laboratory work we chosen a measure of deviation of the hypothetical distribution from the empirical as:

$$D = \chi^2 = \sum_{i=1}^h \frac{(m_i - Np_i)^2}{Np_i}, \quad (11)$$

where h is the number of quantization intervals for a noise, p_i - the hypothetical probability of the interval v_i , m_i - empirical mean value in the interval v_i .

From the literature [5] it is known that the measure of the metrics $D = \chi^2$ has χ^2 distribution with $h - 1$ degrees of freedom, where h is the number of quantization intervals for a noise. If we have number L of parameters of distribution function $w(v|H_0)$, then the number of degrees of freedom of χ^2 distribution is reduced by L . Therefore, χ^2 distribution degrees of freedom will be equal to $k = h - L - 1$.

If the confidence level q is specified, the boundary of the critical region χ^2_q is determined from the equation $P(\chi^2 > \chi^2_q | H_0) = q$, where H_0 is hypothesis that the sample belongs to the expected distribution law $w(v)$. The function $P(\chi^2 > \chi^2_q | H_0)$ is tabulated. [5]

Practice has shown that for the χ^2 test it is enough for the number of degrees of freedom to be in the range of (11–20). If during processing of the experiment results it occurs that $m_i < 10$, you need to combine intervals and add up m_i in corresponding united intervals. The resulting probability is equal to the sum of probabilities of united intervals. The number of degrees of freedom for χ^2 distribution is reduced by the number of intervals j merged and will be equal to $h = h - L - j - 1$.

Table of χ^2 distribution [1] makes it possible to determine the critical divergence (threshold) χ^2_q with the chosen permissible error probability q with accepted hypothesis H_0 about the law of distribution.

The procedure for the hypothesis testing is then reduced to the calculation of χ^2 and checking the following inequalities: if $\chi^2 > \chi^2_q$, H_0 hypothesis is rejected, if $\chi^2 \leq \chi^2_q$, H_0 hypothesis is not rejected.

In the laboratory practicum we propose to check hypotheses about the normal distribution law and the Cauchy distribution. During hypothesis testing, students show empirical PDF function and compare it with hypothetical distributions on the graphics. After that they make a conclusion about hypothesis testing.

5. Conclusion

In this paper we proposed an example of laboratory practicum for the probability density function estimation and a random processes analysis. On the basis of ergodic theory, we realized the algorithm for estimating of probability distribution function for a random process. Use of ELVIS II platform and LabVIEW software package provides wide possibilities for students. They make necessary

calculations, display results of the experiment and, most importantly, to control the experiment. At the same time students are introduced to a LabVIEW software package and its capabilities. Proposed laboratory practicum has been successfully used for many years at Kazan Federal University for statistical radiophysics studies.

References

- [1] Parzen E 1962 On estimation of a probability density function and mode *Ann. Math. Stat.* **33** 1065–1076
- [2] <http://www.ni.com/ni-elvis/>
- [3] Nicholson W 2005 Microeconomic Theory (10th ed.) *Mason, Ohio: Thomson Higher Education*
- [4] Moll Victor Hugo 2014 Special Integrals of Gradshteyn and Ryzhik: the Proofs *Series: Monographs and Research Notes in Mathematics I(1 ed.) Vol I*
- [5] Box, Hunter and Hunter 1978 Statistics for experimenters *Wiley*