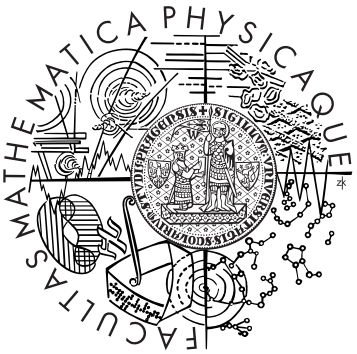
Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS



Peter Greškovič

Study of non-Gaussian light curves using Karhunen-Loève expansion

Astronomical Institute of Charles University in Prague

Supervisor: Mgr. Tomáš Pecháček, Ph.D.
Astronomical Institute,
Academy of Sciences of the Czech Republic

Study program: General Physics
Study plan: Astronomy and Astrophysics

2011

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

Prague 4. 8. 2011 Peter Greškovič

Title:	Study of non-Gaussian light curves using Karhunen-Loève expansion
Author:	Peter Grešković
Department:	Astronomical Institute of Charles University in Prague
Supervisor:	Mgr. Tomáš Pecháček, Ph.D., Astronomical Institute of the Academy of Sciences of the Czech Republic
Supervisor's e-mail:	pechacek_t@seznam.cz

Abstract: We present an innovative Bayesian method for estimation of statistical parameters of time series data. This method works by comparing coefficients of Karhunen-Loève expansion of observed and synthetic data with known parameters. We show one new method for generating synthetic data with prescribed properties and we demonstrate on a numerical example how this method can be used for estimation of physically interesting features in power spectra calculated from observed light curves of some X-ray sources.

Keywords: statistical methods, Bayesian statistics, time series data, X-ray light curves

Názov práce:	Štúdium negausovských svetelných kriviek s využitím Karhunen-Loèveho rozvoja
Autor:	Peter Grešković
Katedra (ústav):	Astronomický ústav Univerzity Karlovy
Vedúci diplomovej práce:	Mgr. Tomáš Pecháček, Ph.D., Astronomický ústav AV ČR
e-mail vedúceho:	pechacek_t@seznam.cz

Abstrakt: Predstavíme inovatívnu bayesovskú metódu na odhad štatistických parametrov časových rád. Táto metóda funguje na základe porovnávania koeficientov Karhunen-Loèveho rozvoja pozorovaných a umelo generovaných dát so známymi parametrami. Ukážeme novú metódu na generovanie svetelných kriviek s predpísanými parametrami a na numerickom príklade ukážeme, ako sa táto metóda dá v spojení s navrhovanou analýzou využiť na určovanie fyzikálne zaujímavých parametrov výkonového spektra pozorovaných svetelných kriviek od niektorých typov röntgenových zdrojov.

Kľúčové slová: štatistické metódy, bayesovská štatistika, časové rady, röntgenové svetelné krivky

Contents

Introduction

Some of the X-ray sources are amongst the most luminous objects in the known universe. They can serve us as unique natural laboratories for testing high-energy physics. They could help us to understand behavior of matter at extreme temperatures and pressures or be useful in testing various relativistic effects. The basic model of these objects (accretion discs) has been known for several decades. Despite great advances in knowledge of these objects, that have been made in past years, we still do not know many details of the physical processes that are going on in them.

Before we can speculate about further details of our physical models, we must find inspiration for them in observed data. Unfortunately, it would seem that we have reached our observational limits. The data, that is provided by our satellites often does not have the quality, that would be required for useful interpretation. The information, which is extracted from the data, is usually not even good enough to test existing hypotheses. Current situation (as of 2011) in X-ray astronomy does not look promising for new missions, which could improve situation with available instruments.

There is, however, another way to get more information. If we cannot get better data, we must improve our analysis. Advances in the field of statistical analysis and invention of new theoretical tools show promise to improve our situation. To that end, this work tries to contribute by offering new tool that is yet to be proved useful.

Chapter 1

Random processes

Variability of X-ray sources, amongst other complicated physical processes, is a result of some very complicated underlying physics. As in many other physical systems, one cannot trace every single event contributing to the observed data. Moreover, those measurements are burdened by observational errors, which are combined effect of many more influences impossible to trace. Therefore, the measured data seems random in its nature and we are left with statistical analysis of its properties.

When dealing with observations of complicated and seemingly random processes, sometimes we can reduce the observations to small number of parameters (e.g. cylinder of gas can be described by a few thermodynamic quantities). In other cases it is the features which stand out of the average behavior that give us the relevant information. Before we can look for one or the other, we must understand what random processes are and what properties they have.

This chapter contains review of mathematics used later. We give detailed discussion of random processes, their properties and ways to handle and represent them. The first section introduces most of the notation used throughout this work. In the second and third sections we give review of the common definitions and clarify and expand the notation introduced before. Most of the definitions in this chapter can be considered common knowledge. In one form or the other they can be found in various literature, e.g. [Bendat and Piersol, 2000]. The remaining sections describe ways of representing random functions that we use and discusses their properties.

1.1 Definition of random process

In physics, we always try to find some relations between some information that we already have (known parameters) and some other information that we are interested in (usually measured values of some physical quantity). Good example of this is the association of time to the observed brightness. Physical processes are any phenomena that produce the interesting information. We will designate the input parameters (usually the time) as t and the measured value (one possible outcome) as x.

If a clearly specified relation can be found that connects the available relevant information to some sort of output, then the phenomena is called *deterministic process*. As opposed to the deterministic process, an outcome of a random or *stochastic process* is subject to uncertainty. Hence, we cannot predict exact outcome based on available information. Random property of a physical process can be either consequence of lack of information or inherent property of the process (quantum physics).

Measurements of outcome of random process are called *random variables*. Let one possible outcome of the random process for some fixed values of the input parameters T_i is designated $x_i^{(j)} = X(\theta_j, t)|_{t=T_i}$. Another possible outcome of the same random process at the same values of the input parameters will be designated $x_i^{(k)} = X(\theta_k, t)|_{t=T_i}$. All possible outcomes for that values of the input parameters will be designated X_i (or $X(T_i)$). Since the set of all possible outcomes may depend on the values of the input parameters, we will designate the random process as $X(t)$ or just X, which means all possible outcomes for all possible values of the input parameters. Outcomes for one specific value of the input parameters will be labeled by subscript.

Function that connects values of input parameters to one possible outcome of the random process over some range of the input parameters is called random function. Any random function $X(\theta, t)|_{\theta=\theta_j}$ can be looked at as a single *realization of random process*. We will represent the random process as a collection of many (but usually some finite number m) of its realizations, i.e. functions $X^{(j)}(t) = X(\theta, t)|_{\theta=\theta_j}$ taken for many different values of $\theta_j \in \{\theta_1, \theta_2, \ldots, \theta_m\}$. In most cases we will omit the dependence on θ in our notation and we will just specify if we are talking about one specific realization or all of them. The realizations will be labeled by superscript in brackets.

Random functions associated with real physical processes are usually required to be continuous. However, we will most often work only with some finite collection of values of a random function sampled at specific values of input parameters $t \in \{T_1, T_2, \ldots, T_n\}$. Such a collection of random variables will be called *random vector*. Instead of $X^{(i)}(T_j)$ or $X^{(i)}(T_k)$ we will write just $x_j^{(i)}$ or $x_k^{(i)}$. One random vector will contain n random values from one specific realization (i.e. values of one random function for n different values of the input parameters) and it will be designated as $x_j^{(i)}$; $j \in \{1, 2, \ldots, n\}$. A random process, represented as a collection of random vectors, then will be $x_j^{(i)}$; $i \in \{1, 2, \ldots, m\}$, $j \in \{1, 2, \ldots, n\}$.

1.2 Probability

Although the exact numerical value of the outcome of a random process cannot be determined, we may know *probability* of realization of certain value (or probability of finding the value within certain interval). On the abstract mathematical level, the probability is defined using the following

Kolmogorov's axioms:

Let Ω be a non-empty set. An algebra on Ω is a set F of subsets of Ω that has Ω as its member and that is closed under union and complementation with respect to Ω. Let P be a function from F to the real numbers obeying:

- (Non-negativity) $P(A) \geq 0, \forall A \in F$
- (Normalization) $P(\Omega) = 1$
- (Finite additivity) $P(A \cup B) = P(A) + P(B) \quad \forall A, B \in F \quad A \cap B = \emptyset$

We call the function P probability function and (Ω, F, P) probability space. [Kolmogorov, 1933]

Many various concepts that meet certain simple criteria (see [Hájek, 2010] for details) can be considered an equivalent "interpretations" of probability. The two of the most common ones that are used in physics are *frequency interpretation* and Bayesian or *subjectivistic interpretation*.

Frequency probability of getting result x, given the input parameters, is defined as the limiting value of the ratio of number of measurements of that value $m(X = x)$ and the total number of measurements m:

$$P_X(x) = \lim_{m \to \infty} \frac{m(X = x)|_{t=T}}{m} \tag{1.1}$$

That is given that results of the experiment are drawn at random from a pool of all possible results. The existence of such a converging infinite sequence of possible draws must be postulated axiomatically. We will approximate this number by empirical probability

$$p_X(x) = \lim_{m \to \infty} \frac{m(X = x)|_{t=T}}{m}, \tag{1.2}$$

where m will be our number of realizations that we have and $m(X = x)|_{t=T}$ the number of realizations of the value x. From the law of large numbers it follows, that this number approaches $P_X(x)$ defined by (1.1).

Another definition of probability, that we are going to use, is setting probability equal to some kind of subjective degree of belief or confidence level. This can be further defined and constrained by some additional conditions. Different authors add various conditions. The least constraining conditions (by de Finetti) are conformance to the probability calculus and so called conditioning, i.e. constant change of the probability in the light of new evidence as it comes:

$$P_{\text{new}}(x) = P_{\text{old}}(x|E) \tag{1.3}$$

provided that the new probability is non-zero. $P(x|E)$ means the value given by the original probability function for x in case that E is true. Although this probability is defined very loosely, thanks to the conditioning property it will be very useful. We can declare our level of belief to be equivalent

to (1.2) and update it as new values of measurements come. For truly random values and big number of measurements, differences between those two probabilities tend to zero in this case.

We will be needing to determine the probability of finding measured value (actual outcome of a random function) within certain interval. To do this, we need to know its *probability density function* (PDF). The PDF $\rho_X(x;t)$ of the random process X is a function of the possible outcome of the process and the input parameters. Probability of finding the measured value (outcome of X) within any given interval Ω for some value of input parameters $t = T$ is equal to integral of the PDF over that interval.

$$p(x \in \Omega)|_{t=T} = \int_{\Omega} \rho_X(x;t)|_{t=T}\,\mathrm{d}x \tag{1.4}$$

This relation can be taken as definition of PDF. We say, that the measured values are random variables drawn from distribution with the corresponding PDF.

For real-valued outcomes, the integral of PDF from $-\infty$ to some possible outcome x is called *cumulative distribution function* (CDF).

$$F(x;t) = \int_{-\infty}^{x} \rho_X(y;t)\mathrm{d}y \tag{1.5}$$

Its significance is the probability of finding the value of the outcome lower than the x [$F(x;T) = p(X < x)|_{t=T}$].

For two different random variables A and B we can define their *joint probability* $p_{A,B}(a,b) = p(A = a, B = b)$ as probability of finding value of the variable $A = a$ and, simultaneously, the value of the second variable $B = b$.

In special case of statistically independent variables, their joint probability is equal to product of their individual probabilities.

$$p_{A,B}(a,b) = p_A(a) \cdot p_B(b) \qquad \text{for independent } A \text{ and } B \tag{1.6}$$

In general, this does not need to be true.

For variables that are not independent, the joint probability can be rewritten using so called *conditional probability* of the two variables.

$$p_{A,B}(a,b) = p_A(a) \cdot p_{B|A}(b|a) \tag{1.7}$$

The conditional probability $p_{B|A}(b|a)$ expresses the probability that $B = b$ given that $A = a$. For the variables, that are not entirely independent, that is not the same as $p_B(b)$.

We may also define *joint probability distribution function* for the two variables

$$\rho_{A,B}(a,b;t) = \rho_A(a;t) \cdot \rho_{B|A}(b|a;t), \tag{1.8}$$

where the $\rho_{B|A}(b|a)$ is so called *conditional probability distribution* e.g. the probability distribution of B when the value of A is held at a (in some time). The joint probability distribution can be integrated over some 2D interval of

possible values of the variables to find the probability of finding the values of both a and b within the interval at the same time.

In a manner similar to definition of the one-dimensional CDF we can define cumulative distribution function for two variables. Its significance will be the probability of finding values of both variables lower than given constraints: $F_{A,B}(a, b; t) = p(A < a, B < b|t)$.

We can marginalize the joint probability distributions by integration over all possible values of one of the variables to find *marginal distribution* for the other variable.

$$f_A(a, t) = \int_{\Omega_B} \rho_{A,B}(a, b; t)\,\mathrm{d}b \tag{1.9}$$

Using the marginal distribution, the conditional distribution of variable B can be written as

$$\rho_{B|A}(b|a; t) = \frac{\rho_{A,B}(a, b; t)}{f_B(b)} \tag{1.10}$$

Using the conditional distributions and the process of marginalization via other variables the concepts of joint probability distribution and joint probability can be easily extended to multiple variables.

$$\begin{aligned} \rho_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n; t) &= f_{X_1,X_2,\ldots,X_{n-1}}(x_1, x_2, \ldots, x_{n-1}; t) \\ &\quad \cdot f_{X_n|X_1,X_2,\ldots,X_{n-1}}(x_n|x_1, x_2, \ldots, x_{n-1}; t) \end{aligned} \tag{1.11}$$

1.3 Properties of random processes

For certain fixed values of the input parameters $t = T_i$ we may get many possible outcomes of random process $x \in \Omega_X(T_i)$, where the $\Omega_X(T_i)$ is domain of all possible outcomes of X for the values of parameters T_i. If we denote $\rho_X(x; t)|_{t=T_i}$ is the PDF of the random process X at that time, then the number

$$\mathrm{E}[X]|_{t=T_i} = \int_{\Omega_X(T_i)} x \cdot \rho_X(x; t)|_{t=T_i}\,\mathrm{d}x \tag{1.12}$$

is called *expected value* or mean of the process at $t = T_i$. We are going to use symbol μ_X for the mean (in specific time μ_{X_i}). From now on, we will omit the dependence of mean (and other variables and functions) on the value of the parameters (time) in our notation, unless we need to point it out.

In discrete case, if the time is the only parameter and we have m realizations of the possible outcome of the random process in the time T_i, we can approximate the mean (μ_{X_i}) as

$$\tilde{\mu}_{X_i} = \frac{1}{m} \sum_{j=1}^{m} x_i^{(j)} \tag{1.13}$$

Mean is the first moment of the random variable. Second moment about

the mean μ_X is defined like this:

$$\begin{aligned} \mathrm{Var}(X) &= \int_{\Omega_X} (x-\mu_X)^2 \cdot \rho_X(x)\,\mathrm{d}x \\ &= \mathrm{E}[(X-\mu_X)^2] \\ &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \\ &\equiv \sigma_X^2 \end{aligned} \tag{1.14}$$

and is called *variance*. Square root of the variance is denoted σ_X and is called *standard deviation* of the random variable X.

Higher moments are defined by replacing the square in the definition integral for variance by higher order power. The moments can be standardized by dividing them by corresponding power of the standard deviation. Standardized third moment is called skewness and standardized fourth moment is called kurtosis.

Again, in discrete case for m realizations of the random variable we can approximate the variance of the random variable by

$$\tilde{\sigma}_{X_i}^2 = \frac{1}{m-1}\sum_{j=1}(x_i^{(j)} - \mu_{X_i})^2. \tag{1.15}$$

For two random variables A and B we may define a quantity

$$\mathrm{Cov}(A,B) = \mathrm{E}[(A-\mathrm{E}[A])\cdot(B-\mathrm{E}[B])] \tag{1.16}$$

called *covariance*. The covariance can be standardized by dividing it by the standard deviations.

$$\mathrm{Corr}(A,B) = \frac{\mathrm{Cov}(A,B)}{\sigma_A \sigma_B} \tag{1.17}$$

This quantity is called *correlation* of variables A and B. The correlation always falls to the interval $[-1,1]$.

In discrete case, we can approximate the covariance of these numbers from available corresponding realizations $a_i^{(k)}$ and $b_j^{(k)}$ as

$$\begin{aligned} \mathrm{Cov}_{\mathrm{appx}}(A_i,B_j) &= \frac{1}{m}\sum_{k=1}^{m}[(a_i^{(k)}-\mu_{A_i})\cdot(b_j^{(k)}-\mu_{B_j})] \\ &= \frac{1}{m}\sum_{k=1}^{m} a_i^{(k)} b_j^{(k)} - \mu_{A_i}\mu_{B_j}. \end{aligned} \tag{1.18}$$

The later form is useful for incremental calculation.

Correlation and covariance are measures of strength of linear relationship between the variables. If the correlation coefficient of the two variables is 0 then the variables are said to be *uncorrelated*. Between uncorrelated variables there still may exist some strong, but other than linear relationship.

For example, if the variables A_i and B_j would be chosen as possible coordinates on a unit circle or a sine wave, their correlation coefficient would be zero, but there would be definitive dependence between $a_i^{(k)}$ and $b_j^{(k)}$.

If no relationship between the two variables can be found, then the variables are said to be *independent*.

Covariance and correlation coefficients calculated from values drawn from the same random process X (as opposed to two different random processes A and B) in two different times, say r and s, are called *autocovariance* $C_X(r,s)$ and *autocorrelation* $R_X(r,s)$ functions. In discrete case, when $X_i = X(T_i)$ are elements of a random vector, the autocovariance and autocorrelation functions are represented by *autocovariance and autocorrelation matrices* C_{ij} and R_{ij}. Elements of those matrices are defined as $c_{ij} = \mathrm{Cov}(X(T_i), X(T_j))$ and $r_{ij} = \mathrm{Corr}(X(T_i), X(T_j))$ respectively.

If for any set of the input parameters $t \in \{T_1, T_2, \ldots, T_n\}$ and any value of some new variable τ, the joint probability distribution of random variables $X_i = X(T_i)$ and the joint probability distribution of variables $X_i' = X(T_i + \tau)$ are the same, then the random process $X(t)$ is said to be *strictly-stationary*.

The mean and covariance of realizations of strictly-stationary random process for any value of input parameters t are the same. Moreover, the autocovariance and autocorrelation functions are only functions of the difference of t: $C_X(r,s) = C_X(r-s) = C_X(\tau)$ and $R_X(r,s) = R_X(r-s) = R_X(\tau)$.

If $\mu_X(T) = \mu_X(T+\tau)$ and $\mathrm{Var}(X_T) = \mathrm{Var}(X_{T+\tau})$ holds for any τ, it is said to be *weak-sense stationary*. Weak-sense stationary process need not to be strictly stationary.

The random process $X(t)$ is said to be *ergodic*, if its properties can be deduced from single sufficiently long realization. Ergodicity implies stationarity, but the reverse need not to be true.

1.4 Karhunen-Loève expansion

An analytic function can be expressed as sum of its Taylor series. In general, if we have a complete system of orthogonal functions $f_i(x)$ on some interval closed $[a,b]$, i.e. functions for which following holds:

$$\delta_{ij} = \int_a^b f_i(x)^\star f_j(x) \cdot \mathrm{d}x,$$

then any continuous function $\varphi(x)$ on that interval can be expressed as the sum

$$\varphi(x) = \sum_{j=1}^{\infty} \xi_j \cdot f_j(x) \qquad x \in [a,b] \tag{1.19}$$

where the ξ_j are coefficients of the expansion that can be determined by

$$\xi_j = \int_a^b f_j(x)\varphi(x) \cdot \mathrm{d}x. \tag{1.20}$$

If we represent the function on the interval $[a, b]$ as some finite number of n samples $\varphi_i = \varphi(x_i),\ x_i \in [a, b]$ (function values in points distributed over that interval), then we do not need infinite sum to determine the expanded values. The n samples is fully determined by n coefficients.

Instead of infinite set of orthogonal functions, we can take finite set of n orthogonal vectors $\vec{f_j}$ (of dimension n), with elements $f_j^i,\ i, j \in \{1, 2, \ldots, n\}$. The expanded values can be expressed as

$$\varphi_i = \sum_{j=1}^{n} \xi_j f_j^i \tag{1.21}$$

where the coefficients ξ_j are

$$\xi_j = \sum_{i=1}^{n} \varphi_i f_j^i. \tag{1.22}$$

The choice of the set of the orthogonal vectors (functions) is arbitrary. The coefficients are determined by this orthogonal system and the expanded vector (function). For a random vector X_i these coefficients are some random numbers.

The Karhunen-Loève theorem (see e.g. [Stark and Woods, 1986]) states, that if we choose the system of orthogonal vectors as the system of eigenvectors of the autocorrelation matrix of that random vector, then the coefficients of the expansion are uncorrelated. Moreover, if we choose the coefficients of the expansion for its k-th realization like this:

$$\xi_j^{(k)} = \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_j}} f_j^i x_i^{(k)}, \tag{1.23}$$

where the coefficients λ_j are eigenvalues of the autocorrelation, then the coefficients $\xi_j^{(k)}$ have zero means and unit variances. We can reconstruct the original values of $x_i^{(k)}$ ($\vec{X}^{(k)}$) from the coefficients $\xi_j^{(k)}$ and the eigensystem like this:

$$x_i^{(k)} = \sum_{j=1}^{n} \sqrt{\lambda_j} \xi_j^{(k)} f_j^i. \tag{1.24}$$

This is called Karhunen-Loève expansion (KLE) of the vector X_i.

It follows, that *any two random vectors represented using KLE with the same basis (the set of orthogonal vectors f_i and values λ_i) have the same autocovariance functions*.

For a continuous random function we would replace the sum with integral. The set of orthogonal functions would be the system of eigenfunctions of its autocovariance function. To completely reconstruct all values of the random function everywhere in the interval $[a, b]$, we would have to calculate the infinite sum. In practice, however, we can often truncate the sum after some finite number of terms, since in convergent sum the first few terms are the most important ones and the error (difference between the sum and the value of the expanded function) decreases as we add more terms.

1.5 Fourier expansion and power spectrum

Another common choice of an orthogonal system for expansion of a function is system of functions defined as

$$f_j(x) = \exp\left(i2\pi\frac{n}{\tau}x\right), \tag{1.25}$$

where $\tau = |b-a|$ is length of the interval where we want to use the expansion (ideally period of the expanded function). The coefficients of the expansion for a function $y(x)$ are

$$c_j = \frac{1}{\tau}\int_a^b y(x)\exp\left(-i2\pi\frac{n}{\tau}x\right)\mathrm{d}x \tag{1.26}$$

and the function can be reconstructed with

$$y(x) = \sum_{j=-\infty}^{\infty} c_j f_j(x). \tag{1.27}$$

This is the well known Fourier series.

In discrete case, realization of n-dimensional random vector with components x_k can be reconstructed using coefficients

$$c_j = \frac{1}{n}\sum_{k=0}^{n-1} x_k \exp\left(-i2\pi\frac{jk}{n}\right) \tag{1.28}$$

and finite sum

$$x_k = \sum_{j=0}^{n-1} c_j \exp\left(i2\pi\frac{kj}{n}\right). \tag{1.29}$$

Vector, whose components c_j are defined by (1.28), is called *discrete Fourier transform* of $\vec{x}$. Relation (1.29) is called *inverse discrete Fourier transform*.

For a continuous function $y(x)$ we can calculate its *Fourier transform*

$$\widehat{y}(\omega) = \int_{-\infty}^{\infty} y(x)e^{-2\pi ix\omega}\mathrm{d}x \tag{1.30}$$

and its *energy spectral density*

$$S(\omega) = \left|\int_{-\infty}^{\infty} y(x)e^{-i\omega t}\mathrm{d}t\right|^2 = \frac{\widehat{y}(x)\cdot\widehat{y}^*(x)}{2\pi}. \tag{1.31}$$

The energy spectral density is useful tool for searching of periodicity, as it shows how energy is distributed with angular frequency ($\omega = 2\pi f$).

Wiener-Kinchin theorem states, that for a wide-sense stationary process the power spectral density is the Fourier transform of the autocorrelation function.

$$S(\omega) = \lim_{T\to\infty}\frac{1}{2T}\mathrm{E}\left[\left|\int_{-T}^{T} e^{-i\omega t}x(t)\mathrm{d}t\right|^2\right] \tag{1.32}$$

Chapter 2

Frameworks for analysis of physical measurements

Any acceptable physical theory must be formulated based on some observations and it must give predictions about values of the some physical properties. It must be possible to observe (measure) those properties and put them to test by experiment. The standard method of acquiring knowledge, referred to as the scientific method, relies on continuous cycle of gathering evidence and updating existing theories, which can be (partially or completely) refuted by experiments. This cycle has following stages: observation or measurement, interpretation of measurements and creation (correction) of hypothesis, prediction of values of physical properties and testing the hypothesis by new observations and experiments. Interpretation of the initial measurements and creation of the hypothesis using deductive approach is a key step in this cycle.

In this chapter we explain our notion of physical measurements and what they represent. We give an interpretation of measured values and we present the results of measurement as output of random processes described in the previous chapter. Then we show methods that are used to extract useful information in form of statistical inference out of such measurements.

First, we define the basic task of the statistical analysis. Then we describe two different approaches to accomplish that task. The two methods, that are used in modern physics, are the frequentist approach and the Bayesian approach. Basic principles of both of those approaches are briefly reviewed and their advantages and disadvantages are discussed.

2.1 Interpretation of measurements

To make the interpretation of measurements possible, we need to identify results of the observation. Measurement, or the measured value of some real physical quantity (observable), is always composed of two elements. The first element is the real value of the quantity of the interest (signal) and

second one is the observational error (noise).

$$\text{measurement} = \text{signal} + \text{noise}$$

The real value of the measured quantity, is somehow connected to the physical system. It depends on the physical system properties, and the system properties can be often (but not always) deduced from the real values of the observable quantities. The assumptions about how the physical system works, what properties it has and what values of the physical quantities should be accessible to observation are called model (or physical theory about the system). We can create model based on values of the physical quantities and we can test the model by comparing its predictions with observations.

The second component of the measured value, called noise, is considered to have unknown random value. All observations are burdened with these errors. By stating that observational errors are of random nature we are saying that we cannot, in principle, know the values of the errors. Entire measurement is therefore a random value. Statistical analysis of measurements tries to define relations between the results of measurements and the real values of the physical quantities.

Since any model we might have about the system is only an assumption, the best we can do is to estimate probabilities of finding the measured data for different assumptions and compare them. Then we can decide, based on these probabilities, between various models and accept or reject them as valid hypotheses for our physical system. In most cases, we can never be completely sure about validity of physical theories and we can only rule out some of the assumptions as "extremely unlikely" or accept some of them as "very likely".

The central problem of the statistical analysis of measurements can be therefore formulated as follows: analysis of measurements aims to find the most likely possible value of the measured variable given values of the measurements (our evidence). It also tries to determine reliability of the result or express our level of trust in it.

Finding the reliability of the result for discrete values means finding probability, that the observed variable has that particular value. For continuous values it means finding the probability, that the value is contained in some small set. That is, for the continuous variables we are trying to determine

$$P(x \in I|M), \tag{2.1}$$

where the x is the actual value of the observed variable, I is the said set and M are measured values of the variable burdened with errors. The set I is referred to as *confidence interval*. It may (or may not) be chosen so, that it contains the most likely value. It is defined as such subset of all possible values that has certain probability of finding the the value of the observed variable within it and has smallest possible size. By the size S of a set I we mean

$$S(I) = \int_I 1 \mathrm{d}dI. \tag{2.2}$$

This set should be determined only by the probability (level of confidence) that the actual value of the observed variable falls within it and quality of our measurements. Measurements with smaller error typically give smaller confidence intervals. The smaller the confidence interval is, the more reliable result we have.

2.2 Frequentist approach

The most common approach to figure out the most likely value and confidence interval is called *frequentist approach*. It is based on the frequency interpretation of probability. From the frequency definition of probability and our interpretation of measurements it follows, that the most likely value of the observed variable is simply the value, that would yield the observed values of measurements the greatest number of times if we could repeat the measurement infinitely many times (under exactly the same conditions).

Analysis of measurements using this approach always assumes that we know something about probability distribution of errors. For example, if the observational errors are some random values with symmetric probability distribution around zero, then the most likely value of the observed variable is simply the arithmetical average of the measured values for many measurements. Probability, that the difference of the actual value of the measured variable and the average is significant, tends to zero with number of measurements going to infinity. That is true for any admissible symmetric probability distribution of errors. In other words, the value of average gets closer to the actual value as the number of measurements increases.

If we assume some specific shape of the distribution of the errors, we can calculate the probability that the arithmetic average μ of the measurements falls within certain interval around the actual value.

The value of the observational error consists of contributions from many different sources. If those contributions can be considered independent, which is often a good assumption, then the probability distribution of an average error tends to normal Gaussian distribution

$$\varrho_g x = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2.3}$$

as the number of the contributions goes to infinity. This is the statement of classical central limit theorem. Width of the distribution σ^2 can be determined from variance of measurements.

Gaussian distribution of errors simplifies many things. For example, we can determine the confidence interval for some level of confidence P by construction called χ^2 statistics. First we calculate value $\chi^2_{\min}$

$$\chi^2_{\min} = \sum_i \frac{(x_i - \mu)^2}{\sigma^2} \tag{2.4}$$

from values of measurement x_i. Any value b such that

$$k.\,|b-\mu| < \frac{\sigma}{\sqrt{\sum_i x_i^2}} \tag{2.5}$$

belongs into confidence interval with the level of confidence equal to

$$P = \int_{-k\sigma}^{k\sigma} \varrho_g(x)\mathrm{d}x \tag{2.6}$$

i.e. integrating the Gaussian normal distribution with variance σ^2 centered around the average value μ within boundaries given by χ^2 statistics gives probability equal to the level of confidence P (see e.g. [Muller and Madejsky, 2009]).

Calculating average of measurements with symmetrical monotonically decreasing distribution of errors and determining confidence interval by χ^2 statistics for measurements with Gaussian noise are the most basic examples of methods based on frequentist approach. For measurements with the Gaussian distribution of errors, there are numerous sophisticated methods to calculate confidence intervals, determine various indirect variables, search for tendencies in observed data and even process power spectra of noisy time series.

The ease of the interpretation of data using frequency probability and the central limit theorem, which guarantees that the distribution of errors is usually Gaussian, are major advantages.

However, the assumption about Gaussian distribution is usually at the cornerstone of all complicated statistical methods and they may not give good results when this assumption is not true. It has been shown, that if some of the contributions to the errors are significantly more important than the others and they are not entirely independent, the errors no more have the Gaussian distribution. The same is true also under somewhat more abstract assumption of fractal underlying geometry, which is widely observed. See e.g. [Metzler and Klafter, 2000] for details. That means, that non-Gaussian processes can be quite common in nature. Modification of existing methods to incorporate non-Gaussian distribution is very complicated, often completely impossible. When it is possible, it must be done on case-to-case basis.

Another important thing to notice is that to calculate the arithmetical average with sufficiently small confidence interval we actually need many measurements performed at the same (or very similar) conditions. This can be a problem, especially if the values of interest change with time. Methods to process measurements of such values can be very complicated.

2.3 Bayesian approach

Instead of assuming, that there is some fixed value of the observed variable and that the frequency of of certain values of measurements that are burdened with random errors with some known probability distribution can

tell us something about this value, Bayesian approach uses completely different method. For any possible value of the variable of interest, it tries to determine the probability density that corresponds to it.

To find out what is the Bayesian probability of some possible value we could start by assigning it some arbitrary non-zero probability and then correct our assumption based on observations. This is principally different approach then the frequentist one.

Calculating probabilities using Bayesian approach is done via Bayes theorem. The Bayes theorem follows from definitions of joint probability and conditional probabilities.

$$P_{AB}(a,b) = P_A(a) \cdot P_{B|A}(b|a) = P_b(b) \cdot P_{A|B}(a|b)$$

leads to

$$P_{A|B}(a|b) = P_A(a) \cdot P_{B|A}(b|a) \cdot \frac{1}{P_B(b)}. \tag{2.7}$$

Using the equation (1.4) it can be easily shown, that the same relation hold for probability densities. All the terms in the equation (2.7) have their established names. The term $P_{A|B}$ is called *posterior probability*. Its meaning is the probability of the some value of interest (a) (or the tested hypothesis) given the observed data (b). $P_A(a)$ is called *apriori probability* and it is our assumption about the probability of the tested value before observation. This assumption is to be corrected in the light of new evidence (observation). $P_{B|A}(b|a)$ is so called *likelihood function* which stands for probability of finding the measured values given the assumed value of data. This is where our assumption (distribution function) about errors come in. The last term $1/P_B(b)$ is *normalization factor*. Its meaning is probability to find the particular measured values out of all possible values there could be. The name comes from the fact, that it is thanks to this term, that the sum of probabilities for all possible values is 1. As long as we are only interested in comparing probabilities of different values, we can disregard this term (or normalize our posterior probabilities so that the sum is 1 later). If we assume that all the possible values are equally probable, we can disregard the apriori probability term too.

If we can find the probability density of finding the measured data as a function of the parameters (e.g. likelihood function), we can perform the Bayesian estimate. Then we can find the smallest set (of parameters) on which the integral of the found probability density reaches the required level of confidence (P).

If we have more than useful measurement, and we cannot calculate the likelihood directly, using all of the measurements, we can find the final confidence region using so called *Neyman construction*. We just find confidence region with desired level of confidence for each measurement separately and then find the confidence region as intersection of those confidence intervals. It is immediately obvious, that more measurements constraints our confidence interval better (result is usually smaller confidence interval).

One of the advantages of this approach is, that we were able to isolate the step, where we were are required to make assumption about the distribution of the data. For different distributions, We only need to change calculation of the likelihood function. This makes methods based on the Bayesian approach much more universal, because they can be readily adapted for handling measurements with any distribution of errors via simple change. The same modification usually requires extensive modifications of the standard (frequentist) methods for processing measurements.

Chapter 3

Variable luminosity of accretion flows

The primary motivation behind development of new statistical method, which is described in next chapter, is investigation of X-ray light curves from active galactic nuclei and compact objects. When we talk about X-ray light curves from compact objects, we mean mainly time domain observations of flux by satellites (like RXTE, XMM-Newton, Chandra…) of binary systems of white dwarfs, neutron stars, stellar BH's or active galactic nuclei supermassive BH's with accretion discs. Light curves of many of these objects show similar properties. This could lead to the conclusion, that basic mechanism responsible for of creation of the radiation and variability are the same (see e.g. [Uttley and McHardy, 2005]).

The radiation of these sources originates in accretion discs and shows strong variability on relatively short time scales. Source of the variability, as well as reliable models of the phenomena, remain largely unknown. Still, we can make elaborate guesses properties of these objects based on properties of the variability.

In the first paragraph of this chapter we give very brief review of some power spectral features, that are observed in these light curves. Then we discuss, how these features are connected to physical properties of those systems and what conclusions can be drawn from them. Finally, we point out the biggest complications in their observation.

3.1 Shape of the power spectrum

The discussed light curves show irregular variability on typical time scales that vary from fraction of seconds (stellar BH's, see [Axelsson et al., 2006]) to days (supermassive galactic BH's, see [Markowitz and Edelson, 2004]). Apart from the irregular variability, some objects show signs of periodic or quasi-periodic behavior. The typical time scales and periodic frequencies with highest energy in the power spectra of the time series observations can be attributed to some physical properties of the systems [van der Klis et al., 1985].

The more or less random light curve can be described as a realization of a random process. Its most basic properties are specified by its CDF and autocorrelation function. The autocorrelation function appears to have double exponential form:

$$C(t) = \sigma^2 \cdot e^{-\frac{|t|}{\tau}}, \tag{3.1}$$

whet the constant τ is the typical time scale of variations. According to the Wiener-Kinchin theorem, power spectrum of this process is given by Fourier transform of the autocorrelation function

$$S(\omega) = \widehat{C}(t)(\omega) = \frac{2\sigma}{\omega^2 + \frac{1}{\tau^2}} \tag{3.2}$$

i.e. it has form, which is approximately constant for $\omega^2 \ll \tau^{-2}$ and falls like ω^{-2} for big ω. Typical power spectrum of an X-ray light curve then looks like the one on figure 3.1. The break is not actual discontinuity, but rather artifact created by logarithmic scales. However, approximating the relations between logarithm of power spectrum and logarithm of frequency by linear functions for low and high frequencies and finding their meeting point seems to work well for determination of the typical time scale. In some cases, there is more than one typical time scale and there are more breaks in the power spectrum. Position of this break is an important feature, because from its frequency we can deduce time scale of variability which is linked to typical time scales of physical processes in the source.

Another important feature, that is probably linked to physical properties of the system, are quasi-periodic oscillations. Figure 3.1 shows typical appearance of the QPO in power spectrum. Although QPOs show as a peak on power spectrum, there is no coherent periodic process happening at their frequency.

3.2 Physical properties linked to the variability

Although the specific mechanism responsible for creation of radiation and variability is not known, we can draw some general conclusions from basic principles.

The most basic result, pointed out in very early stages of exploration of what we call today compact objects, is relativistic limit to size. If the radiated power is created within certain volume of space with edge-to-edge dimension of s, and we assume, that the power creation is approximately evenly distributed within the volume, then the total radiated power cannot significantly change faster than s/c (the c is speed of light). The reason is, that opposite ends of the volume could not be synchronized faster than with the speed of light. Sudden change in luminosity in all parts of the volume at once, that is not causally connected, would be very unlikely, given the nature of observed variability.

Another meaning of the typical time scale could be deduced from our other assumptions about the sources of the radiation. When we consider the

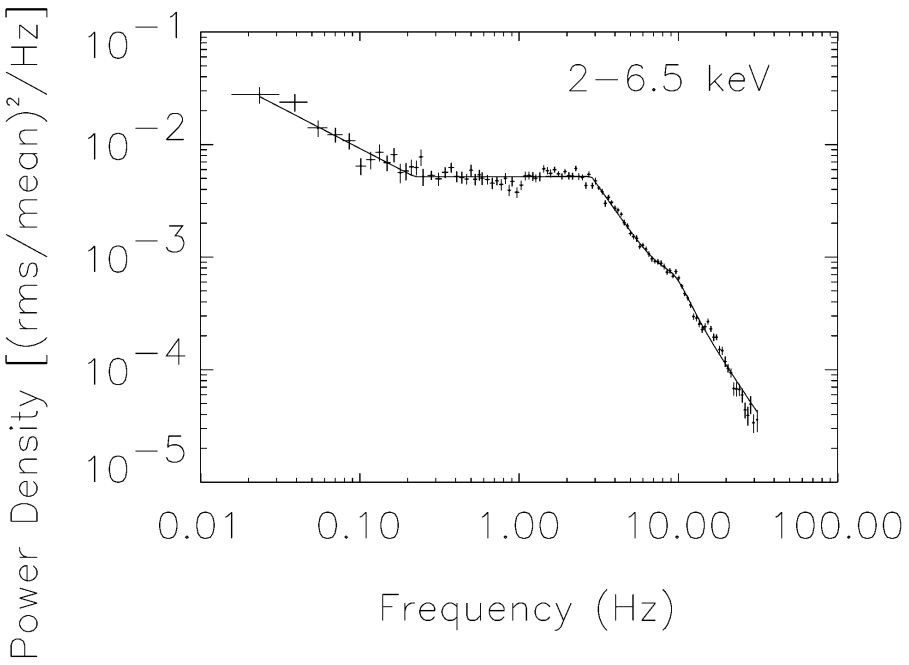


Figure 3.1: Observed power spectral density from X-ray light curve of Cygnus X-1 in transition between low and high energy state. The most obvious feature is break on frequency corresponding to typical time scale of variability. Break like this appears to be universal feature of power spectra of many sources. Other features (secondary breaks) can be sometimes observed. Picture taken from [Cui et al., 1997].

sources as accretion discs, (in some cases there is even direct evidence for that), we can expect the matter in them to orbit with certain orbital period, is given by mass of the central object M. This period can be for most of the time approximated by non-relativistic Kepler orbital period:

$$T^2 = \frac{2\pi^2}{GM} \cdot R^3 \tag{3.3}$$

Since this period depends on distance, orbiting alone is probably not the main source of variability. Although, according to our best physical models of accretion discs, most energy is created in the innermost regions, and hence, relatively small range of orbital distances. There are some theories, which consider e.g. orbiting luminous hot spots as the source of variability.

If the central object is black hole, according to general relativity there exists last stable orbit, which must be the last possible inner edge of any stable accretion disc. Radius of the last stable orbit r_L scales with Schwarzschild radius R_s, which scales linearly with mass.

$$r_L \propto R_s = \frac{2GM}{c^2} \tag{3.4}$$

There is a strong correlation between the (independently determined) mass of the central object and typical time scales (determined from the break in

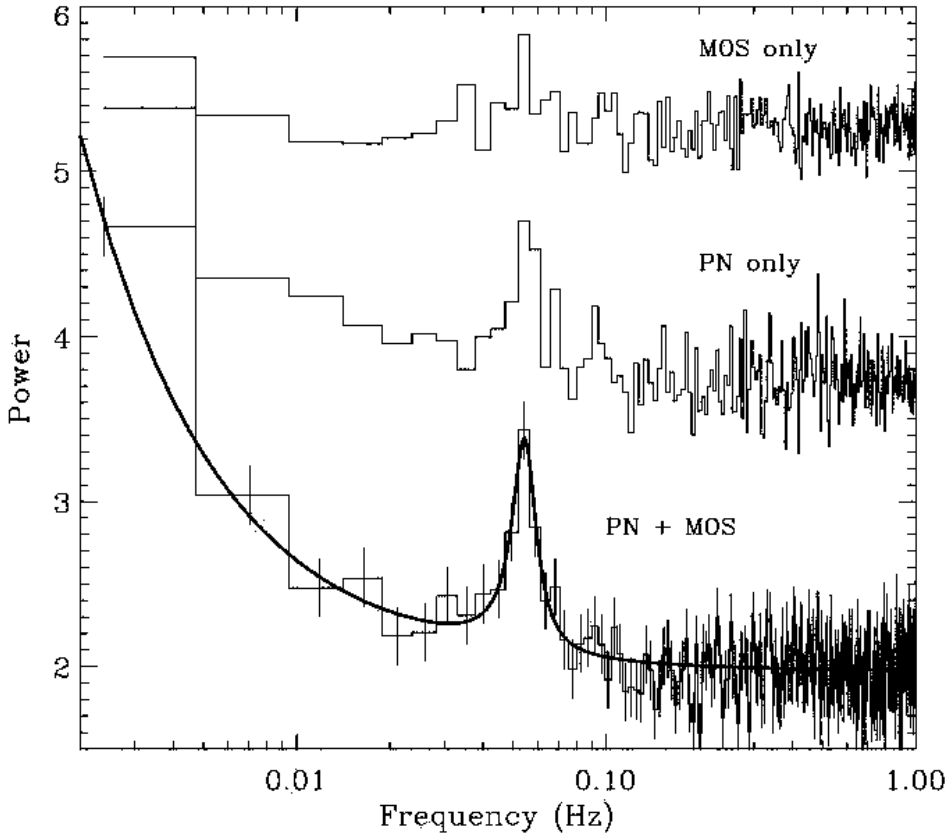


Figure 3.2: Quasi-periodic oscillations in power spectrum from an ULX in M82 (M82 X-1) in several energy bands. Image credit ESA, XMM-Newton, [Strohmayer and Mushotzky, 2003].

power spectrum). See figure 3.2 or work of [McHardy et al., 2004]. This could indicate, that the typical time scale could be linked to the dimension of the most luminous parts of the accretion disc.

Also the frequency of the quasi-periodic oscillations appears to be linked to the orbital frequency at the last stable orbit. If the QPOs are observed, this relation can serve as good measure of the mass of the central object.

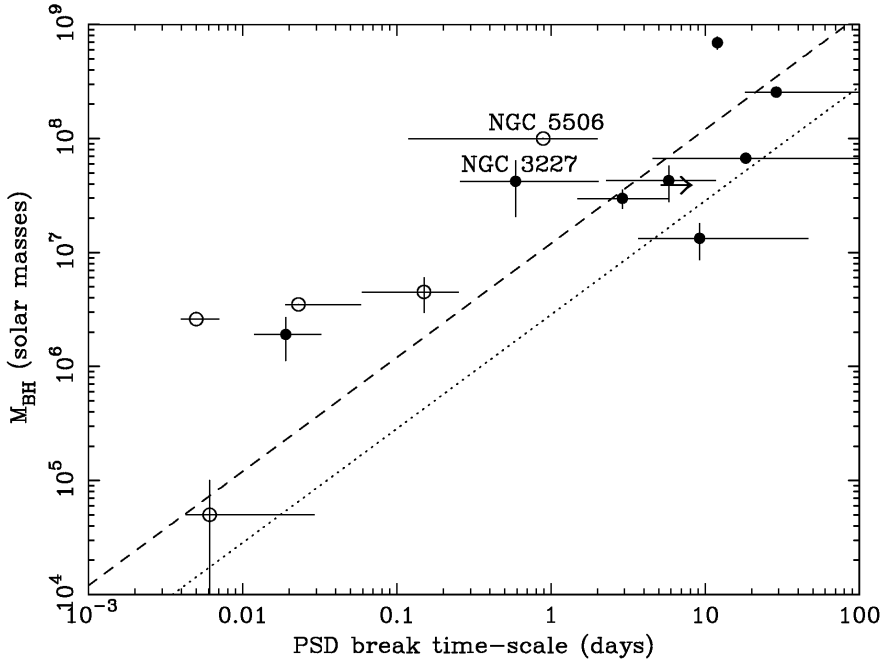


Figure 3.3: Correlation between independently measured mass of central objects and typical time scale of variability in accretion discs. Picture taken from [Uttley and McHardy, 2005].

Chapter 4

Bayesian analysis of light curves

In the last chapter we talked about how time-domain observations can help us to understand basic properties of compact objects. Before we can draw any conclusions from these observations, we must overcome obstacles that are connected with their interpretation. Most of them have very high level of noise. We cannot repeat measurement of flux at some particular moment to improve precision and the detectors are not advanced enough to give us data with the precision we would like. At short time measurements we often deal with individual photon counts. Moreover, the distribution of data is very non-standard. It has log-normal rather than Gaussian distribution. Most statistical method based on assumption of Gaussian distribution might give us imprecise results. That is why many authors reached out to Monte Carlo simulations and methods based on Bayesian statistics.

This chapter forms the core of the theoretical part of this work. We show how the Bayesian framework can be used for analysis of light curves and what kind of useful information can be extracted. We summarize the proposed method for analysis of light curves and discuss its advantages and disadvantages.

We compare this method to some other methods, that have been proposed by different authors. We also show, how these new method can be used specifically in case of X-ray light curves for exploration of some properties that are impossible to analyze otherwise.

The developed method is universal in that way, that it can be readily used for investigation of many other problems, that deal with series of measurements of constantly changing quantity. Examples of such data are seismogram, weather records and so on.

4.1 Description of the method

In our analysis, the light curves are modeled as random functions of time. That means, that the value of interest (brightness) is random variable. Rather

than determining the actual value of brightness at certain time, we are interested in statistical properties of the function as a whole.

In principle, this is done by comparing observed light curve with many simulated light curves with known properties. Since the light curves are random, highly variable and noisy, the comparison is non-trivial task. The randomness property is the reason, why we must express the results of comparison in terms of probability. The light curve can have, in principle, almost any shape, only some of them are more likely than the others.

We do not try to compare values of brightness in different light curves directly. Instead, we express the curves using Karhunen-Loève expansion (KLE). Since we consider both, the simulated light curves and the observed one, random, the coefficients of the KLE will be random too.

They will strongly depend on the basis functions, which are in this case eigenfunctions (eigenvectors) of the autocovariance (see Eq. (1.23) and definition of KLE).

The KLE coefficients are uncorrelated, but not independent. They are connected by strong non-linear relations. Their probability distribution must be described as joint probability distribution function for all coefficients (JPDF).

We can look at the KLE as transformation (in analogy to the discrete Fourier transform). It takes vectors from space vectors that contain values of brightness (direct space) and brings them to space of vectors that contain values of KLE coefficients. That transformation is non-singular (reverse transformation exists). Than if we have processes, that are somehow distinguished in direct space, e.g. processes with some distribution, they must be somehow distinguished also after KLE transform. Therefore, the precise form of the JPDF must depend not only on the basis functions, but also on those other properties of the random processes.

The corner stone of our method is to estimate the JPDF of the KLE coefficients. For a given autocovariance function, and other properties, we try to empirically estimate the JPDF from KLE coefficients of generated synthetic light curves. Then we find values of the KLE coefficients of the actual observed light curve and we estimate value of the probability density for that particular combination of KLE coefficients.

4.2 Application of Bayesian framework

The result of the estimate described in the previous paragraph is a value of probability density. That value is calculated for KLE coefficients of a particular observed light curve from JPDF given by synthetic light curves with specific values of parameters.

We can estimate the value of probability density (probability in infinitesimal interval) for the observed light curve for any value of the parameters. That means, that we can find the probability density function as a function of the investigated parameters, given the observed light curve. That is ex-

actly the definition of Bayesian likelihood function.

To complete the Bayesian estimate, we need to set some value of apriori probability and estimate probability (density) of finding our particular data for calculation of normalizing factor. In our case, that would mean finding the probability of finding the particular combination of values of KLE coefficients of observed data out of all possible combinations. We can avoid the later problem by doing normalization in simpler way. The choice of apriori probabilities is, in principle, unconstrained.

We choose some uniformly distributed grid of parameters in parameter space. This grid should cover some reasonable range such, that probability of finding the parameters outside of this grid should be negligible. If we don't have any other apriori knowledge about the probability distribution of the parameters, we can assign the same apriori probability value to all of them (i.e. $1/n$, where the n is the number of grid points, so that total apriori probability is 1). Now the purpose of the normalization factor in Bayesian estimate is such, that integral of the final probability distribution function over all possible values of parameters should be 1. We can calculate the normalization factor by summing up the values of probability density for all tested values of parameters. Final probability value then will be the value of the likelihood function, multiplied by $1/n$ (apriori probability) and divided by this sum (normalizing factor).

4.3 Joint probability distribution of ξ_i

The entire process of calculating likelihood function had two key steps: generation of the synthetic light curves and empirical estimation of JPDF from their KLE coefficients. The problem of generation of the synthetic light curves will be dealt with in the next chapter. Estimation of the JPDF is described here.

Simplest method, that could be considered, would be direct estimation of density at some point from density of coefficients found in small neighborhood. The KLE coefficients of the observed process would represent one point in n-dimensional space, where the n is the number of coefficients. We would like to estimate, how many points, representing KLE coefficients of random light curves generated for some value of parameters, falls into small neighborhood of the point from observed light curve. The probability, that the observed light curve has parameters of the synthetic light curves would be (in limit for number of points going to infinity and size of the neighborhood going to zero) equal to fraction of points that fall into the neighborhood and the total number of points. For one dimensional case, this is analogical to estimating probability distribution from histogram. Even for simple 1D histogram, too many values would be required to achieve good level of precision. Instead of one dimensional points, we would have points in n-dimensional space. To achieve the same level of precision, as we would achieve with 1000 points in one dimension, we would need 1000^2 points

in two dimensions. Typical dimension (number of coefficients) for our case would on the order of hundreds. This simple consideration shows, that this method is not usable, because the number of realizations we would have to generate would be too big.

For 1D histograms there are some methods, which allow to "smooth" the distribution and draw nice histograms (and provide good estimates of probability density) even from small number of realizations (points drawn from the distribution of interest). These methods are based on summing up some contributions to total density from every point. Each point $\vec{x_i}$ is approximated by some smooth function $K(\vec{x})$ called kernel (e.g. Gaussian). At any point of their vector space, the probability is approximated by normalized sum of the contributions:

$$P(\vec{x}) \doteq \sum_i \frac{K(\vec{x} - \vec{x_i})}{n} \tag{4.1}$$

More sophisticated methods use kernel variable with some other estimates of density. For details of multi variative kernel methods see [Terrel and Scott, 1992]. Unfortunately, for dimension 100 and 1500 points, we were not able to approximate the joint probability distribution even with those methods.

The method, that proved successful, was in fact the simplest one. We emphasized that the coefficients are uncorrelated, but not independent. If they were independent, we could calculate their joint probability density simply by multiplying their individual probability densities:

$$\varrho(\xi_1, \xi_2, \ldots, \xi_n) = \varrho(\xi_1) \cdot \varrho(\xi_2) \cdots \varrho(\xi_n) \tag{4.2}$$

Surprisingly, this value appears to give the best approximation even compared to multi variative kernel estimation. As will be shown by our results, the reliability of this estimate can be significantly improved by using only some of the coefficients.

Chapter 5

Generation of random data with prescribed properties

Important step in calculating likelihood function for Bayesian analysis of light curves, as described in last chapter, was generation of random synthetic light curves. These random light curves were modeled as random processes with prescribed properties. To investigate the power spectral features, we need to model random processes with known autocorrelation function and CDF. The generation of realizations of such random processes is non-trivial task, which accounted for much work in this thesis.

This chapter deals with the methods for generating random data. We describe the *spectral synthesis*, that is can be used for simulating light curves, since it can produce data with satisfactory autocorrelation. However, the distribution of the generated data will be Gaussian. Many other methods to accomplish this task were proposed by other authors. Some of them try to solve the flaw with distribution, but our investigation of some of them has shown, that they are usually not suitable, because they are severely constrained to very specific type of data they can generate (like the method proposed in [Phoon et al., 2005]) or they are relatively cumbersome to implement (method by [Ferrante et al., 2005]).

For this reasons, we have proposed a new method, which is rather computationally inefficient, but it is very universal and has good properties that make it more suitable for our purposes. This method is described in second paragraph.

5.1 Spectral synthesis

Spectral synthesis is one of the standard methods for generating random data with an arbitrary power spectrum. Design of this method is based on analysis of Fourier transform and power spectrum of stochastic process.

In its basic form, the distribution of generated data is Gaussian. For log-normal distribution it can be adapted (translation process). It also generates light curves only from very small subset of all possible light curves with the desired properties. The basic idea is, that one can get data with known

power spectrum like this:

$$x(t) \propto \sum_{\omega} \sqrt{S(\omega)} \cos[\omega t - \varphi(\omega)], \tag{5.1}$$

where the $\varphi(\omega) \in [0, 2\pi]$ is random. Every angular frequency $\omega = 2\pi\nu$ has therefore fixed amplitude and only phases are randomized between data points.

The method, as described in [Timmer and Koenig, 1995] does following:

1. Pick the power spectrum $S(\omega)$.
2. For every frequency ω_i draw two random numbers from Gaussian distribution and multiply those numbers by $\sqrt{\frac{1}{2}S(\omega_i)}$.
3. Use those numbers as real and imaginary parts of coefficients of Fourier transform. For coefficients with negative frequencies, use their complex conjugates.
4. Calculate the time series by inverse (discrete) Fourier transform.

Using this method it is possible to generate time series that resemble light curves, that can be used for certain purposes. Of course, it is obvious, that such data cannot be simply used if we are looking for spectral features that are not consistent with simple power law (e.g. break in the power spectrum).

5.2 Brutal force algorithm

To generate data with arbitrary admissible distribution and autocorrelation we proposed simple method, that is based on random swapping. This method could be entitled as brutal-force matching of autocorrelation function on data with prescribed distribution. Described version of the method only works for discrete representations of stationary processes.

It is relatively easy to generate a set of random numbers with prescribed distribution. One easy method (that we are actually using) is called *inverse sampling method*. The basic idea of this method is, that CDF F_ϱ for any distribution ϱ has the meaning of probability and the result of operation $F_\varrho(x)$ for any x always falls within interval $[0, 1]$. Moreover, if we draw some random variable x from this distribution, then $F_\varrho(x)$ will be uniformly distributed over the interval $[0, 1]$. But then the inverse function to the CDF, which we denote F_ϱ^{-1}, applied on a random numbers drawn from uniform distribution on $[0, 1]$ will result in variables x, which are random numbers drawn from the corresponding distribution (see [Press et al., 1992] for proof).

$$x = F_\varrho^{-1}(y); \qquad y \in U[0, 1]; \qquad P(a < x \leq b) = \int_a^b \varrho(z)\mathrm{d}z \tag{5.2}$$

We can therefore easily generate random vector of dimension n: $\vec{X} = [x_1, x_2, \ldots, x_n]$ with desired distribution. The vector with components generated via inverse transform sampling method from independent, uniformly distributed numbers y_i, may not have the desired autocovariance.

The following discussion is limited to random vectors that can be considered as samplings of stationary random processes (i.e. every element is value of one specific realization of some stationary random process X (with desired distribution) in some time: $x_i = X^{(k)}(T_i)$).

If it is, at least in principle, possible to have a random process with our desired target distribution and target autocovariance, then we call this combination of autocovariance and distribution admissible. We can match only admissible combinations of autocovariance and distributions.

For one realization of such random vector, we will approximate the empirical autocovariance by this formula:

$$R_{\text{appx}}(t) = \sum_{i=1}^{n-t} (x_i - \bar{x}) \cdot (x_{i+t} - \bar{x}), \qquad t = 1, 2, \ldots, n-3 \tag{5.3}$$

This formula approximates the autocovariance well for most of t (except for $t \doteq n$) if n is big enough. This can be seen from the formula (1.19), definition of autocovariance and the condition of stationarity.

We can reorder elements within the random vector without changing the distribution. For every possible ordering of the elements, we can calculate the empirical autocorrelation using formula (5.3).

If we define distance of two autocovariance functions as

$$D^2(R_1, R_2) = \sum_t [R_1(t) - R_2(t)]^2 \tag{5.4}$$

then we can say, that for every possible ordering, the distance of the autocovariance approximated using (5.3) for that ordering and the target autocovariance $R(t)$ is some number. There is only $n!$ possible orderings. Amongst those orderings, there must exist one, for which the distance between approximated autocovariance of this ordering and the target autocovariance is minimum. We call this autocovariance $R_{\min}(t)$.

We can assign each possible ordering some number d, which is equal to distance of approximated autocovariance for that ordering from target autocovariance. Given all possible sets of n random numbers drawn from the desired distribution, we could, at least in principle, find a probability distribution of finding some value of d. We call this distribution ϱ_d. The bigger n we have, the more possible orderings we could realize and the more numbers d we draw from that probability distribution for one particular set of numbers x_i.

For some set of numbers x_i with target distribution ϱ, that are picked at random, we have the bigger chance of finding d close to zero the more possible orderings we can make. That means, that average value of $D(R_{\min}, R)$ is a decreasing function of n. The bigger n we have, the closer to target autocovariance we can get by reordering.

If we swap some two values $x_j \leftrightarrow x_k, j < k$, we do not need to recalculate approximated autocovariance entirely. Instead, we can only calculate change in R_{appx} for every t:

$$\begin{aligned} &\Delta R_{\text{appx}}(t) = 0 \qquad \text{if} \qquad t > k \\ &\Delta R_{\text{appx}}(t) = 0 \qquad \text{if} \qquad t > (k-j),\ j < t,\ k > (n-t) \\ &\Delta R_{\text{appx}}(t) = 0 \qquad \text{if} \qquad t = (k-j),\ j < t,\ k > (n-t) \end{aligned} \tag{5.5}$$

and for other values of t,

$$\begin{aligned} \Delta R_{\text{appx}}(t) = R_{\text{appx}}(t) & \\ - \ & \{[(x_{i-t} - \bar{x}) \cdot (x_i - \bar{x})] + [(x_k - \bar{x}) \cdot (x_{i-t} - \bar{x})]\}_1 \\ - \ & \{[(x_{k-t} - \bar{x}) \cdot (x_k - \bar{x})] + [(x_i - \bar{x}) \cdot (x_{k-t} - \bar{x})]\}_2 \\ - \ & \{[(x_{i+t} - \bar{x}) \cdot (x_i - \bar{x})] + [(x_k - \bar{x}) \cdot (x_{i+t} - \bar{x})]\}_3 \\ - \ & \{[(x_{k+t} - \bar{x}) \cdot (x_k - \bar{x})] + [(x_i - \bar{x}) \cdot (x_{k+t} - \bar{x})]\}_4 \end{aligned} \tag{5.6}$$

where we count only those terms in $\{\}_i$, which make sense. For example, we cannot count $\{\}_1$ for any t for which $i < t$ and so on. If we calculate ΔR_{appx} for every t, we can calculate $\Delta D(R_{\text{appx}}, R)$.

We start by calculating approximated autocovariance for initial ordering and storing that information somewhere (we keep record for every t). Now we can pick i, j at random, calculate possible $\Delta D(R_{\text{appx}}, R)$ that we would get, if we swapped elements at those two positions, and if the distance of the approximated autocovariance of the new ordering would be closer to the target autocovariance, then we perform the swap and update our stored information about the approximated autocovariance. If that change would increase the distance, we pick new pair of i and j.

Numerical experiments has shown, that the target autocovariance can be matched quite well by our approximated autocovariance after approximately n^2 tries. More tries increases statistical probability, that we get better precision, but is time consuming.

Chapter 6

Numerical experiment

In this chapter we describe simple test of the proposed method that we designed to demonstrate usability of the method for analysis of X-ray light curves. This experiment shows basic usage of the method and reliability of its results.

For this test, we generated one testing random light curve, which was approximated as a series of 100 values (random vector). The autocovariance of was given by

$$C(\tau) = \exp\left(-\frac{\tau}{b_0}\right). \tag{6.1}$$

This form was chosen because it is somewhat similar to the observed autocovariance of X-ray light curves. The curve also had log-normal distribution given by PDF

$$\varrho(x) = \frac{1}{x \cdot \sqrt{2\pi a_0^2}} \cdot \exp\left[\frac{(\ln x - \mu)^2}{2a_0^2}\right] \tag{6.2}$$

Since this was only a toy example, we artificially prescribed one parameter $\mu = 0$ for all curves. The values of other parameters were $a_0 = 1$ and $b_0 = 30$ for the testing curve.

Then we chose a 20×20 grid in plane of parameters a and b with values of the parameter a uniformly distributed in the interval $[0.1, 1.9]$ and $b \in [5, 55]$ around the parameters of the testing curve.

For each of the 400 points (all combinations of the parameters) we generated 1500 random light curves with those parameters and performed their KLE.

Then we took all coefficients of the same order at every point (e.g. all 1500 realizations of the coefficient ξ_1 from light curves generated for parameters $a = 0.1$, $b = 5$) and we approximated their cumulative distribution by function

$$F_{\text{appx}}(y) = \frac{N(\xi_i^{a,b} \leq y)}{N}, \tag{6.3}$$

where the $N(\xi_i^{a,b} \leq y)$ is the number of values of the coefficients lower than y. This way we approximated marginal cumulative distributions F_{ξ_i} at every point. With 1500 points it was possible to approximate these functions relatively well.

Since the cumulative distribution is integral of the probability density, slope (derivation) of this function at any value of y approximates the corresponding probability density at that value. Hence, we were able to determine the marginal distributions ρ_{ξ_i} as slopes of the approximated cumulative densities.

At last, we used formula (6.2) to get the final probability density function. As stated before, this density depends on how many terms are used in the calculation. It has significance of likelihood function. To get correct probability, we should normalize this function so that sum of all values is 1.

Results are shown in figure 6. It can be seen, that the area with highest probability density is located near real value. Value of the parameter a was determined much better than value of b. Some bias was to be expected, since the generated curve was random (and its KLE coefficients need not to lie exactly in the center of the region of highest probability). There were several false areas of high probability, but the point with highest probability was very near the correct value (red dot).

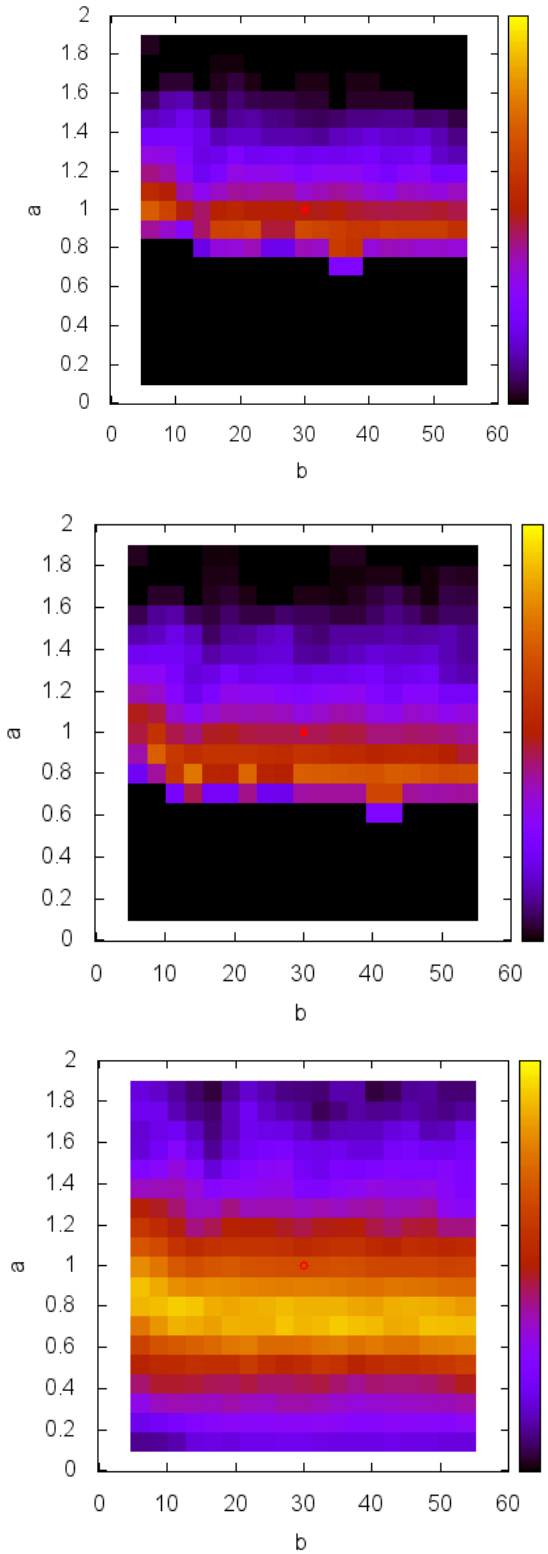


Figure 6.1: Probability density calculated from (6.2) using only the first 30 (up), 50 (middle) and all 100 coefficients. The point with the highest value of the probability for the most precise result lies right under the correct value. Estimated values of parameters are $a = 0.8,\ b = 30$.

Conclusions

Methods similar to ours were proposed by some other authors. Each of those methods has certain disadvantages when compared to method proposed here. For example [Vaughan, 2010] proposed Bayesian method for processing power spectra, but it was based on an assumption of Gaussian distribution of data. Another method, described by [Muller and Madejsky, 2009], used better Bayesian approach but still used χ^2 statistics for estimates of likelihood probability. It would be interesting to compare results of these methods on the same problem.

Also, more investigation should be done with regards to the sensitivity of the proposed method to various influences, like different distributions of the data. Although the method appears to be working on numerical examples, an application to real world data could provide some useful pointers. Unfortunately, processing of real X-ray data proved to be time consuming task mostly outside of the scope of this work.

Even with those things missing, we were still able to provide description of new statistical method and provide the most basic test of its functionality.

References

Axelsson, M., Borgonovo, L. and Larsson, S. [2006], 'Temporal variability of Cygnus X-1 (Axelsson+, 2006)', *VizieR Online Data Catalog* **345**, 20975–+.

Bendat, J. and Piersol, A. [2000], *Random Data: Analysis and Measurement Procedures*, Wiley.

Cui, W., Zhang, S. N., Focke, W. and Swank, J. H. [1997], 'Temporal properties of cygnus x-1 during the spectral transitions', *The Astrophysical Journal* **484**(1), 383.

Ferrante, F. J., Arwade, S. R. and Graham-Brady, L. L. [2005], 'A translation model for non-stationary, non-Gaussian random processes', *Probab Eng Mech* **20**(3), 215–228.
URL: *http://dx.doi.org/10.1016/j.probengmech.2005.05.003*

Hájek, A. [2010], 'Interpretations of probability', *The Stanford Encyclopedia of Physlosophy* . Spring 2010 Edition.
URL: *http://plato.stanford.edu/archives/spr2010/entries/probability-interpret/*

Kolmogorov, A. N. [1933], 'Grundbegriffe der wahrscheinlichkeitrechnung', *Ergebnisse Der Mathematik* . Translated as Foundations of Probability.

Markowitz, A. and Edelson, R. [2004], 'An Expanded Rossi X-Ray Timing Explorer Survey of X-Ray Variability in Seyfert 1 Galaxies', *ApJ* **617**, 939–965.

McHardy, I. M., Papadakis, I. E., Uttley, P., Mason, K. O. and Page, M. J. [2004], 'Combined long and short timescale X-ray variability of NGC 4051 with RXTE and XMM-Newton', *Nuclear Physics B Proceedings Supplements* **132**, 122–128.

Metzler, R. and Klafter, J. [2000], 'The random walk's guide to anomalous diffusion: a fractional dynamics approach', *Physics reports* **339**(2000), 1–77.

Muller, M. and Madejsky, G. [2009], 'Parameter estimation and confidence regions in the method of light curve simulations for the analysis of power density spectra', *ApJ* **700**, 243–250.

Phoon, K., Huang, H. and Quek, S. [2005], 'Simulation of strongly non-gaussian processes using karhunen-loeve expansion', *Probabilistic Engineering Mechanics* **20**(2), 188 – 198.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. [1992], *Numerical recipes in C (2nd ed.): the art of scientific computing*, Cambridge University Press, New York, NY, USA.

Stark, H. and Woods, J. [1986], *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice-Hall, Inc.

Strohmayer, T. E. and Mushotzky, R. F. [2003], 'Discovery of X-Ray Quasi-periodic Oscillations from an Ultraluminous X-Ray Source in M82: Evidence against Beaming', *ApJ* **586**, L61–L64.

Terrel, G. and Scott, D. [1992], 'Variable kernel density estimation', *Annals of Statistics* **20**, 1236–1265.

Timmer, J. and Koenig, M. [1995], 'On generating power law noise.', *Astronomy and Astrophysics* **300**, 707–+.

Uttley, P. and McHardy, I. M. [2005], 'X-ray variability of NGC 3227 and 5506 and the nature of active galactic nucleus "states"', *MNRAS* **363**, 586–596.

van der Klis, M., Jansen, F., van Paradijs, J., Lewin, W. H. G., van den Heuvel, E. P. J., Trumper, J. E. and Szatjno, M. [1985], 'Intensity-dependent quasi-periodic oscillations in the X-ray flux of GX5 - 1', *Nature* **316**, 225–230.

Vaughan, S. [2010], 'A bayesian test for periodic signals in red noise', *MNRAS* **402**, 307–320.

And as a bonus, here is a picture of kitten:

. . . because everybody likes kittens.