Numerical Computation of the Small Balls Probability for Random Functions with Normal Components

JIŘÍ ZELINKA
Department of Mathematics and Statistics, Faculty of Science
Masaryk University
Kotlářská 2, 611 37 Brno
CZECH REPUBLIC

Abstract: Statistical methods are often based on the properties of the distribution of random variables or random vectors. In functional data analysis (FDA) we do not work with random observation containing a finite random vector, but the whole function is one observation. We call it *the functional random variable* or *the random function*, in short. This paper offers the possibility to generate random functions with normal components. In this case, the probability of small balls can be calculated numerically using the characteristic function. This tool can be very useful in simulations and testing various kinds of estimates.

Key-Words: Numerical integration, characteristic function, normal distribution

Received: April 3, 2020. Revised: August 2, 2020. Accepted: September 2, 2020. Published: October 10, 2020.

1 Introduction

In the previous paper (see [6]), the author showed how it is possible to express the random functional variable (shortly random function or random curve) using the Fourier series. A brief summary of these results follows.

Let the random function $\boldsymbol{\chi}$ be a measurable mapping of a probability space $(\Omega, \mathcal{A}, P)$ into a separable Hilbert space H with the orthonormal base $\Psi = (\psi_n)_{n=0}^{\infty}$. Then $\boldsymbol{\chi}$ can be expressed in the form

$$\boldsymbol{\chi} = \sum_{n=0}^{\infty} X_n \psi_n, \tag{1}$$

where X_n are one-dimensional random variables called *components of the random function* $\boldsymbol{\chi}$ *with respect to the orthonormal base* Ψ. These components are defined by the scalar product

$$X_n(\omega) = (\boldsymbol{\chi}(\omega), \psi_n). \tag{2}$$

For fixed $\omega \in \Omega$ the relationship

$$\|\boldsymbol{\chi}(\omega)\|^2 = \sum_{n=0}^{\infty} X_n^2(\omega)$$

holds. Then the norm of random function $\boldsymbol{\chi}$ is a random variable and we can express it in the form

$$\|\boldsymbol{\chi}\|^2 = \sum_{n=0}^{\infty} X_n^2. \tag{3}$$

It is also possible to define the stochastic independence of the random function's components in the following way. We can say that *the components* X_n, $n = 0, 1, 2, \ldots$ *of the random function* $\boldsymbol{\chi}$ *with respect to the orthonormal system* Ψ *are (mutually) independent if every finite subset of the components is (mutually) independent*.

The sum of squares of the coefficients of the Fourier series must be finite. The analogous condition has to hold for the components of a random function:

$$P\left(\sum_{n=0}^{\infty} X_n^2 < \infty\right) = 1. \tag{4}$$

The sufficient condition to fulfill this property for independent components $X_0, X_1, X_2, \ldots$ is

$$\sum_{n=0}^{\infty} E(X_n^2) < \infty, \tag{5}$$

which can be derived using Markov's inequality.

Now it is very easy to simulate random functions with given distribution. For instance, let X_n, $n = 0, 1, \ldots$, be the normally distributed random variables, $X_n \sim N(\mu_n, \sigma_n^2)$. As $E(X_n^2) = \mu_n^2 + \sigma_n^2$ the condition (5) is equivalent to the conditions

$$\sum_{n=0}^{\infty} \mu_n^2 < \infty, \quad \sum_{n=0}^{\infty} \sigma_n^2 < \infty. \tag{6}$$

The following figure (Fig. 1) presents the collection of 10 random functions defined from $L^2(0,1)$ with components $X_0 \sim N((0,1)$, $X_n \sim N(\frac{1}{2^n}, \frac{1}{10n^2})$, for $n > 0$. The functions

$$\psi_n(x) = \begin{cases} \quad 1 & \text{for} \quad n = 0, \\ \sqrt{2}\cos(n\pi x) & \text{for} \quad n > 0 \end{cases} \tag{7}$$

were used as the orthonormal base.

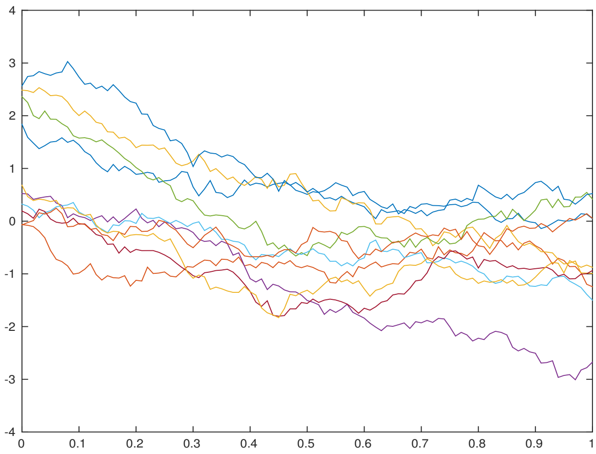

Figure 1: Simulated random functions

This paper focuses on the small balls $B(\chi, h)$ defined as

$$B(\chi, h) = \{\eta \in H; \|\eta - \chi\| < h\}.$$

and the probability $P(\boldsymbol{\chi} \in B(\chi, h))$ for random function $\boldsymbol{\chi}$ and the fixed (non-random) $\chi \in H$. If

$$\chi = \sum_{n=0}^{\infty} x_n \psi_n$$

then the probability of the ball $B(\chi, h)$ is given by the probability of the set of all $\omega \in \Omega$ for which

$$\sum_{n=0}^{\infty} (X_n(\omega) - x_n)^2 < h^2. \tag{8}$$

It is clear that the information about the distributions of the components X_n can help us to evaluate the probability of small balls. On the other had we can made simulations using components of the random function with predetermined distributions and obtain the exact values of some estimated quantities by this way.

2 Normally distributed components

Now let us suppose independent and normally distributed components X_n of the random function $\boldsymbol{\chi}$, $X_n \sim N(\mu_n, \sigma_n^2)$, $\sigma_n > 0$. It is clear that for the random function $\boldsymbol{\chi}$ and for the non-random element $\chi \in H$, $\chi = \sum_{n=0}^{\infty} x_n \psi_n$, and the difference $\boldsymbol{\chi} - \chi$ has also normal components with the same variances and means $\mu_n - x_n$. For this reason the below derived properties of the random functions are also valid for the differences of this kind.

As it follows from the relationships (3) the probability $P(\|\boldsymbol{\chi}\| < h)$ is equal to the probability $P\left(\sum_{n=0}^{\infty} X_n^2 < h^2\right)$. Due to this fact we focus on properties of the distribution of the random variable $Y = \sum_{n=0}^{\infty} X_n^2$.

The random value X_n^2 has chi-squared distribution with 1 degree of freedom. Its characteristic function is

$$\varphi_{X_n^2}(t) = \frac{1}{\sqrt{1 - 2\sigma_n^2 t\, i}} e^{\frac{\mu_n^2 t\, i}{1 - 2\sigma_n^2 t\, i}} \tag{9}$$

and $E(X_n^2) = \sigma_n^2 + \mu_n^2$. So we suppose that the conditions in (6) are fulfilled. Now it is seen that the characteristic function of $Y = \sum_{n=0}^{\infty} X_n^2$ is

$$\varphi_Y(t) = \prod_{n=0}^{\infty} \frac{1}{\sqrt{1 - 2\sigma_n^2 t i}} e^{\frac{\mu_n^2 t\, i}{1 - 2\sigma_n^2 t\, i}}$$

i.e.

$$\varphi_Y(t) = \frac{1}{\sqrt{\prod_{n=0}^{\infty} (1 - 2\sigma_n^2 t i)}} e^{\sum_{n=0}^{\infty} \frac{\mu_n^2 t\, i}{1 - 2\sigma_n^2 t\, i}}. \tag{10}$$

The question of the convergence of the infinite product in (10) is solved in the 15-th chapter of [4] where the following theorem can be found:

Theorem 1 *Suppose $\{u_n\}$ is a sequence of bounded complex functions on a set S, such that $\sum_{n=0}^{\infty} |u_n(t)|$ converges uniformly on S. Then the product $\prod_{n=0}^{\infty} (1 + u_n(t))$ converges uniformly on S.*

Let us put $u_n(t) = -i2\sigma_n^2 t$. Evidently, $\sum_{n=0}^{\infty} |u_n(t)|$ converges uniformly on any compact set S because $\sum_{n=0}^{\infty} |u_n(t)| = 2|t| \sum_{n=0}^{\infty} \sigma_n^2$. So the infinite product $\prod_{n=0}^{\infty} (1 - 2\sigma_n^2 t i)$ in (10) converges uniformly on any compact set S.

The infinite sum in the exponential part is also absolutely convergent on any compact set S:

$$\sum_{n=0}^{\infty} \left| \frac{\mu_n^2 t\, i}{1 - 2\sigma_n^2 t\, i} \right| = \sum_{n=0}^{\infty} \frac{\mu_n^2 |t|}{\sqrt{1 + 4\sigma_n^4 t^2}} \le |t| \sum_{n=0}^{\infty} \mu_n^2.$$

Together it is seen the the characteristic function $\varphi_Y(t)$ is defined for any real t. Now, let's look at its properties.

Theorem 2 *The real part of $\varphi_Y(t)$ is an even function and its imaginary part is an odd function of the variable t.*

Proof: At first, we have to identify uniquely the square root of the expression $1-2\sigma_n^2 ti$ in the denominator of $\varphi_{X_n^2}(t)$. The trigonometric form of $1-2\sigma_n^2 ti$ is

$$1-2\sigma_n^2 ti = r_n(\cos\alpha_n + i\sin\alpha_n)$$

where

$$\begin{aligned} r_n &= \sqrt{1+4\sigma_n^4 t^2}, \\ \cos\alpha_n &= \frac{1}{r_n}, \\ \sin\alpha_n &= -\frac{2\sigma_n^2 t}{r_n}, \\ \alpha_n &\in (-\frac{pi}{2}, \frac{pi}{2}). \end{aligned}$$

As r_n is the even function of the variable t it can be easily seen that the real part of $1-2\sigma_n^2 ti$ is also an even function while the imaginary part is an odd function of the variable t.

Let us choose the square root of $1-2\sigma_n^2 ti$ as $\sqrt{r_n}(\cos\frac{\alpha_n}{2} + i\sin\frac{\alpha_n}{2})$. Then

$$\frac{1}{\sqrt{1-2\sigma_n^2 t\, i}} = \frac{1}{\sqrt{r_n}}(\cos\frac{\alpha_n}{2} - i\sin\frac{\alpha_n}{2}).$$

Next, for $\alpha_n \in (-\frac{pi}{2}, \frac{pi}{2})$ we have

$$\cos\frac{\alpha_n}{2} = \sqrt{\frac{1+\cos\alpha_n}{2}} = \sqrt{\frac{r_n+1}{2r_n}}$$

and

$$\sin\frac{\alpha_n}{2} = \frac{\sin\alpha_n}{2\cos\frac{\alpha_n}{2}} = \sqrt{\frac{2}{r_n(r_n+1)}}\sigma_n^2 t.$$

These formulae yield the fact that the real part of $\frac{1}{\sqrt{1-2\sigma_n^2 t\, i}}$ is an even function of the variable t and the imaginary part is an odd function of the variable t.

The exponential part of (9) can be expressed as

$$\exp\left(\frac{\mu_n^2 t\, i}{1-2\sigma_n^2 t\, i}\right) =$$

$$= \exp\left(-\frac{2\sigma_n^2\mu_n^2 t^2}{1+4\sigma_n^4 t^2} + \frac{\mu_n^2 t}{1+4\sigma_n^4 t^2} i\right) =$$

$$= \exp\left(-\frac{2\sigma_n^2\mu_n^2 t^2}{1+4\sigma_n^4 t^2}\right)\cdot$$

$$\cdot\left(\cos\frac{\mu_n^2 t}{1+4\sigma_n^4 t^2} + i\sin\frac{\mu_n^2 t}{1+4\sigma_n^4 t^2}\right).$$

We can see again that the real part of this expression is en even function while the imaginary part is an odd function of the variable t. As this property is preserved by multiplying complex functions it is valid for $\varphi_{X_n^2}(t)$ as well as for the characteristic function $\varphi_Y(t)$. □

Theorem 3 *The characteristic function $\varphi_Y(t)$ is absolutely integrable.*

Proof: We have several possibilities how to express the exponential part of $\varphi_{X_n^2}(t)$. One of them is

$$\exp\left(-\frac{\mu_n^2}{2\sigma_n^2}\left(1-\frac{1}{1+4\sigma_n^4 t^2}\right) + \frac{t\mu_n^2}{1+4\sigma_n^4 t^2} i\right)$$

so its absulute value is

$$\exp\left(-\frac{\mu_n^2}{2\sigma_n^2}\left(1-\frac{1}{1+4\sigma_n^4 t^2}\right)\right).$$

This is the even and unimodal function with maximal value equal to 1 (for $t=0$) and its limit for $t\to\pm\infty$ is equal to $e^{-\frac{\mu_n^2}{2\sigma_n^2}}$. We can use the estimate

$$\left|\varphi_{X_n^2}(t)\right| = \frac{1}{\left|\sqrt{1-2\sigma_n^2 t\, i}\right|}\left|e^{\frac{\mu_n^2 t\, i}{1-2\sigma_n^2 t\, i}}\right| \le \frac{1}{\sqrt[4]{1+4\sigma_n^4 t^2}}$$

and

$$|\varphi_Y(t)| =\le \prod_{n=0}^{\infty}\frac{1}{\sqrt[4]{1+4\sigma_n^4 t^2}}$$

Let's denote as $\varphi_k(t)$ the partial product of the previous expression, i.e.

$$\varphi_k(t) = \prod_{n=0}^{k}\frac{1}{\sqrt[4]{1+4\sigma_n^4 t^2}}.$$

We see that the sequence of the functions $\varphi_k(t)$ is non-increasing in k for every t. So

$$|\varphi_Y(t)| \le \prod_{n=0}^{\infty}\frac{1}{\sqrt[4]{1+4\sigma_n^4 t^2}} \le \varphi_k(t),\ \forall k.$$

Then

$$|\varphi_Y(t)| \le \varphi_3(t) \le \frac{1}{1+4s^4 t^2}$$

where $s = \min\{\sigma_0,\sigma_1,\sigma_2,\sigma_3\}$. But the function $\frac{1}{1+4s^4t^2}$ is absolutely integrable, $\int_{-\infty}^{\infty}\frac{1}{1+4s^4t^2}dt = \frac{\pi}{2s^2}$. Therefore $\varphi_Y(t)$ is also absolutely integrable. □

Now, we know that the random variable Y is absolutely continuous with the probability density function f_Y and the cumulative distribution function F_Y

which can be evaluated using the formulae (see [5])

$$f_Y(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_Y(t) dt \tag{11}$$

and

$$F_Y(x) = \frac{1}{2} + \frac{i}{2\pi} p.v. \int_{-\infty}^{\infty} e^{-itx} \frac{\varphi_Y(t)}{t} dt \tag{12}$$

where $p.v$, denotes the Cauchy principal value of the integral. Then the probability of the small ball can be evaluated as

$$P(\|\boldsymbol{\chi}\| < h) = P\left(\sum_{n=0}^{\infty} X_n^2 < h^2\right) = F_Y(h^2).$$

3 Numerical computations and simulations

The same parameters as for the example from the Figure 1 were used, i.e., $\mu_0 = 0$, $\sigma_0^2 = 1$, $\mu_n = \frac{1}{2^n}$, $\sigma_n^2 = \frac{1}{10n^2}$, for $n > 0$. The partial product and sum in (10) were calculated until the absolute value of the difference of two successive members was less then 10^{-6}, so 280 members were used for the approximation of the characteristic function φ_Y. The real and the imaginary parts of φ_Y are shown in Figure 2.

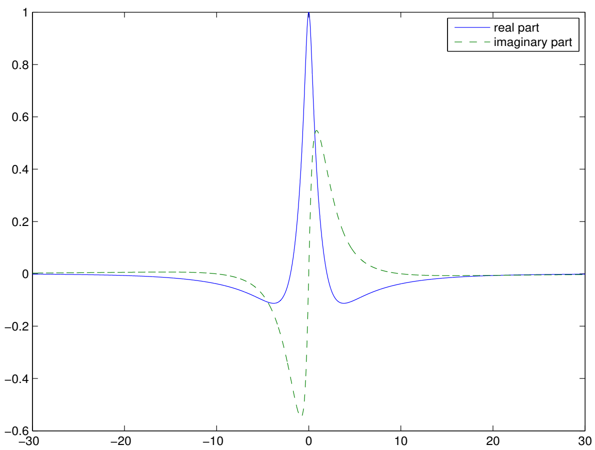

Figure 2: Characteristic function of $\sum_{n=0}^{\infty} X_n^2$.

We used $I = [-500, 500]$ as the interval for the numerical calculation of the integrals in (11) and (12) because the values of the characteristic in the boundary points are small enough: $|\varphi_Y(\pm 500)| < 10^{-7}$. This interval was split into 20,000 subintervals and the composite trapezoidal rule was used for the numerical integration together with the five-step Romberg integration to achieve better accuracy. The probability density function and the cumulative distribution function are displayed in Figure 3.

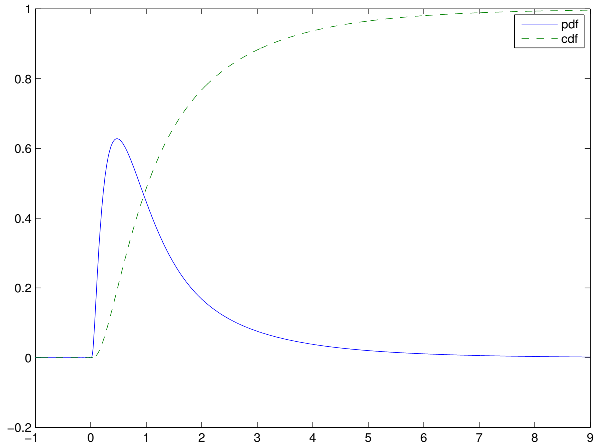

Figure 3: Probability density function and cumulative distribution function of $\sum_{n=0}^{\infty} X_n^2$.

We did a simple simulation to test the correctness of the theoretical calculations. We generated N random functions $\boldsymbol{\chi}$ with the same distribution as before for $N \in \{100, 500, 1000, 5000, 10000\}$. Then we estimated the probability $P(\|\boldsymbol{\chi} - \chi\| < h)$ for $h \in \{0.5, 1, 1.5, 2\}$ by the approximation $P_n(h)$ in the form of the ratio of the number of favorable outcomes and the total number of possible outcomes. The norm $\|\boldsymbol{\chi} - \chi\| = \left(\int_0^1 (\boldsymbol{\chi}(x) - \chi(x))^2 dx\right)^{1/2}$ was also computed using the composite trapezoidal rule with the Romberg integration.

The function $\chi(x) = 1 - x$ was used for the simulations. The Fourier coefficients of this function in the orthonormal base (7) are

$$x_0 = \frac{1}{2}, \quad x_n = \begin{cases} \frac{2\sqrt{2}}{n^2\pi^2}, & n \text{ odd.} \\ 0, & n > 0, \text{ even} \end{cases}$$

Then the exact probability $P(\|\boldsymbol{\chi} - \chi\| < h)$ is given by $F_Y(h^2)$ for $Y = \sum_{n=0}^{\infty} X_n^2$ where

$$X_n \sim N(\mu_n - x_n, \sigma_n^2).$$

The following table (Table 1) summarizes the results.

4 Conclusion

The methods presented in this paper can be used for the simulations of random functions and numerical

Table 1: Estimates of probabilities of the small balls and numerical values.

h	$P_{100}(h)$	$P_{500}(h)$	$P_{1000}(h)$	$P_{5000}(h)$	$P_{10000}(h)$	$F_Y(h^2)$
0.5	0.0800	0.1300	0.1240	0.1238	0.1250	0.1173
1.0	0.5400	0.5920	0.5290	0.5420	0.5379	0.5346
1.5	0.7700	0.8100	0.7830	0.7950	0.7865	0.7856
2.0	0.9100	0.9260	0.9210	0.9166	0.9185	0.9150

calculation of some probabilities concerning the distribution of these random functions. This fact offers a wide range of applications, for example, to test various estimation methods in FDA.

References:

[1] Ferraty, F. ed.:Recent Advances in Functional Data Analysis and Related Topics. Springer Science & Business Media, ISBN 3-790-82736-3 (2011).

[2] Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Springer, New York. ISBN 0-387-30369-3 (2006).

[3] NPFDA web page: `https://www.math.univ-toulouse.fr/staph/npfda/`

[4] Rudin, W.: Real and Complex Analysis. McGraw Hill (1970)

[5] Ushakov, N.G.: Selected Topics in Characteristic Functions, De Gruyter, ISBN 9-067-64307-6 (1999)

[6] Zelinka J.: Random Functional Variable and Fourier Series. In: Aneiros G., Bongiorno E., Cao R., Vieu P. (eds) Functional Statistics and Related Fields. Contributions to Statistics. Springer, (2017)