SIMILARITY COEFFICIENTS OF NORMAL DISTRIBUTIONS IN SELECTING THE OPTIMAL TREATMENTS

Elisabeth Rakus-Andersson
Department of Mathematics and Natural Sciences, Blekinge Institute of Technology
371 79 Karlskrona, Sweden

Janusz Frey
Department of Urology, Örebro University Hospital
701 85 Örebro, Sweden

ABSTRACT

In the current research, we aim to define a new form of the similarity coefficient to compare the resemblance grade of two Gaussian density functions. We aim to assess the method utility on a theoretical model. The density functions are stated for a biological marker "*survival length*", observed in three groups of patients, suffering from a hypothetical disease. The first group consists of patients who are not treated, whereas we recommend 2 possible treatment methods for the second and the third group, respectively. All the "*survival length*" assumptions of the model (mean values and standard deviations) are made to exclude the equivocal conclusion, regarding a selection of the better treatment. At the first stage, we apply the measure of similarity to populations: survival among untreated patients contra survival among patients after Treatment 1. Another similarity coefficient estimates a relation between populations: survival among untreated patients versus survival among patients after Treatment 2. The lower value of the coefficient points out the more effective treatment. In order to simplify calculations, proposed in the definition of a similarity coefficient, we approximate the Gaussian curve by a specially designed polynomial, known as the π-function.

KEYWORDS

Gaussian density function, π-function, similarity coefficient, survival length, optimal treatment.

1. INTRODUCTION

In modern times, the progress of technology and biological sciences is an undisputable fact, which gives us the great opportunity to process large amounts of data, contributing to the development of, among others, medical sciences. There are many specific registers of diseases, allowing to pose the most important questions, concerning the diseases, when using the population-based observational methodology of studies. This is a cost-effective and useful compliment to the randomized clinical trials, recognized as a best source for data. Some of the important questions still cannot be answered, even after employing the modern epidemiological/bio-statistical methods.

Our purpose is to develop and to evaluate the method, simplifying normal distribution comparisons, as well as to test its usefulness, efficacy, and reliability. To accomplish this test, we evolve the theoretical model with the "*survival length*" as a biological parameter for the hypothetical disease, which can be managed with two different treatments. We would like to emphasize that the model data only have a theoretical character, and the data values are selected in order to make a differentiation between treatments more difficult. We have decided to build our model only on the basis of the most often measured parameters, such as mean values and standard deviations. We assume that in the biological sciences and medicine, most of data will have the normal (Gaussian) distribution of a density function, especially when a number of observations is high. The standard

parameters of the density functions, like the mean value and the standard deviation, are thus estimated on the basis of large samples (Gauss, 1809; Stigler, 1986; Krishnamoorthy, 2006 and many others).

In the medical problem to be solved, we focus on the selection of an efficacious therapy, recommended for patients who are touched by the hypothetical disease, which shortens the humans' lives. We also assume that there exist data, referring to the natural course of the disease without any treatment. Another assumption suggests that the data concerning "*survival length*" of patients, cured with Treatment 1 and 2, are available. The biological parameter "*survival length*" is normally distributed, as stated above. We assume that we, by means of collected data, can estimate the means and the standard deviations for groups of patients examined. The population of untreated patients will reveal the survival lengths constituting a reference set of the model. We confront the reference set of survival lengths with other sets of survival, found in the samples of patients, who have been treated by the therapies recommended.

For the purpose of the model, the following assumptions have been made: the patients with the non-treated disease are marked by the mean survival length of 3.0 years with the standard deviation of 0.9 years. The group of patients, with Treatment 1 proved, has the mean survival length of 4.2 years with the standard deviation of 1.2 years. Lastly, the group, where Treatment 2 has been used, is characterized by the mean survival length of 4.0 years with the standard deviation of 1.3 years. The choice of the superior treatment will be carried out to provide an unequivocal answer.

Generally, to compare how like or dislike the survival lengths are for pairs (reference survival without treatment, survival after Treatment i), where $i = 1,\ldots,n$, is a treatment number, we need a measure of similarity for two normal density functions.

There were some attempts, already made to find this measure. The oldest trial involved the comparison of two confidence intervals, built for two distributions (Neyman, 1937 and many others). Recently, the method has awoken some critical remarks and, in spite of the great popularity, there are authors who avoid it (Hoekstra et al., 2014; Morey et al., 2015).

The Bhattacharyya coefficient is an approximate measurement of the amount of overlap between two statistical samples. The coefficient can be used to determine the relative closeness of the two samples being considered (Bhattacharyya, 1943). In 1989 Inman and Bradley proposed a measure of the agreement between two probability distributions by checking their dissimilarity index. Another proposal appeared in (Mishra et al., 1986), where the confidence interval was constructed for the overlapping coefficient.

In our concept, given two normal density functions, we calculate an area under the common part of the functions to divide it later by the total area under both functions. We produce, as a "*similarity coefficient*" a real ratio value equal to 0 in the case of the total dissimilarity, and a value equal to one, if the densities are identical.

We realize that the less area of the common part provides us with the lesser value of the coefficient. We thus choose the optimal Treatment i, whose "*survival length*" compared to reference survival, generates the least quantity of similarity.

Since Gaussian functions are difficult analytic tools of calculations, then we will try to approximate them by a particular π-function of the second degree, very well adapted to the Gaussian shape.

Section 2 constitutes a general description of the method. In Section 3, we introduce a concept of "*similarity coefficient"*. The approximation of a normal density function by a π-function is discussed in Section 4. By comparing similarity coefficients, computed for survivals after different therapies in relation to the reference survival, we finally state a hierarchy of therapies. Some concluding remarks are sampled in Section 6.

2. THE GENERAL MODEL OF A SELECTION OF THE BEST MEDICINE

In the paper prepared, we wish to provide a measure of similarity of two compared normal density functions to decide if two states of the same clinical feature are distributed in like manner or not.

Let us suppose that the clinical attribute "*survival length*" takes values x in space $X = [x_{\min}, x_{\max}]$. The attribute is one of the most important parameters, characteristic of a very invasive disease.

The patients sometimes are not aware of the existence of the disease; therefore there exists a group of patients who are not treated. The sampling of survival lengths for untreated patients creates interval $T \subset X$. We find the mean value μ_T and the standard deviation σ_T for the observations of "*surviving length*", belonging to

T. The density function $f_T(x, \mu_T, \sigma_T)$ is normally distributed. This function will be named the "*reference density function*".

To improve some prognoses of "*survival length*", therapies Treatment 1, Treatment 2,…,Treatment *n* are recommended. We sample the results of the survival in *n* groups of patients. The values of the survival are allocated in intervals $T_1 \subset X$, $T_2 \subset X$,…,$T_n \subset X$ after testing the results of Treatment 1, Treatment 2,…,Treatment *n*, respectively in each group. New density functions $f_{T_i}(x, \mu_{T_i}, \sigma_{T_i})$, $i = 1,\ldots,n$, are assigned to the effects of Treatment *i*. If Treatment *i*, $i = 1,\ldots,n$, has an expected positive effect for the prolongation of the survival length, then condition $\mu_T < \mu_{T_i}$ should be satisfied.

We aim to extract the most efficacious treatment bringing the most promising survival results, when relating them to survivals of patients who did not get any treatment. Therefore, we pairwise compute values of the similarity coefficients C_i between the reference function $f_T(x, \mu_T, \sigma_T)$ and functions $f_{T_i}(x, \mu_{T_i}, \sigma_{T_i})$, $i = 1,\ldots,n$.

Among *n* coefficients $C_i(\ f_T(x, \mu_T, \sigma_T)\ ,\ f_{T_i}(x, \mu_{T_i}, \sigma_{T_i})\ )$, $i = 1,\ldots,n$, we select the least value, which points out the most effective treatment in the aspect of prolonging the survival length.

The similarity coefficient is designed by authors as a complement in comparative methods.

3. DENSITY FUNCTIONS AND SIMILARITY COEFFICIENTS

Generally, the probability density function $f(x, \mu, \sigma)$ is given by

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{1}$$

for a mean value μ and a standard deviation σ, computed from the sample.

The domain of $f(x, \mu, \sigma)$, whose graph is known in statistics as the Gaussian curve, is allocated in $X = (-\infty, \infty)$. We suppose in this study that the graphs of $f_T(x, \mu_T, \sigma_T)$ and $f_{T_i}(x, \mu_{T_i}, \sigma_{T_i})$ have only one intersection point, due to the assumption $\mu_T < \mu_{T_i}$.

Example 1

We observe the normally distributed attribute "*survival length*" in patients, who suffer from a hypothetical disease (e.g., cancer). Suppose $X = [0, 20]$ (survival provided in years). After sampling the data of survival length in interval $T \subset X$ among patients who did not get any treatment, we estimate the reference density function $f_T(x,3.0,0.9)$, $x \in T \subset X$.

There are two types of recommended therapies, Treatment 1 and Treatment 2, which are expected to improve the length of survival. The survival lengths of patients, cured by Treatment 1, are collected in interval $T_1 \subset X$ and estimated by $f_{T_1}(x,4.2,1.2)$, $x \in T_1 \subset X$. Effects of Treatment 2, seen in interval T_2, are data creating the pattern of $f_{T_2}(x,4.0,1.3)$, $x \in T_2 \subset X$.

The graphs of density functions $f_T(x,3.0,0.9)$, $f_{T_1}(x,4.2,1.2)$ and $f_{T_2}(x,4.0,1.3)$ are sketched in Figure 1.

We note the clear resemblance between $f_{T_1}(x,4.2,1.2)$ and $f_{T_2}(x,4.0,1.3)$. To judge which treatment is more effective, we propose an implementation of our own similarity coefficient for two density functions. Let us thus compare a similarity coefficient between $f_T(x,3.0,0.9)$ and $f_{T_1}(x,4.2,1.2)$ to a similarity coefficient between $f_T(x,3.0,0.9)$ and $f_{T_2}(x,4.2,1.2)$. The lower value of the similarity coefficient will indicate the more efficacious therapy.

We propose a similarity coefficient $C(\ f_T(x, \mu_T, \sigma_T)\ ,\ f_{T_i}(x, \mu_{T_i}, \sigma_{T_i})\ ) = C(\ f_T(x)\ ,\ f_{T_i}(x)\ )$ stated by the formula

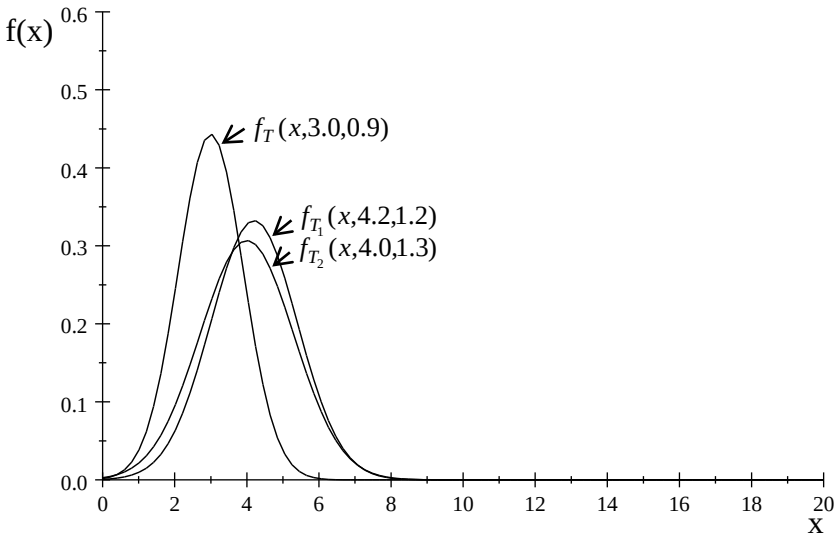

Figure 1. The density functions $f_T(x,3,0.9)$, $f_{T_1}(x,4.2,1.2)$ and $f_{T_2}(x,4.0,1.3)$

$$C(f_T(x), f_{T_i}(x)) = \frac{\text{area under } (f_T(x) \cap f_{T_i}(x))}{\text{area under } (f_T(x)) + \text{area under } (f_{T_i}(x)) - \text{area under } (f_T(x) \cap f_{T_i}(x))}, \tag{2}$$

where

$$f_T(x) \cap f_{T_i}(x) = \{f(x) : f(x) = \min(f_T(x), f_{T_i}(x)), x \in X\}. \tag{3}$$

Let us note that areas under $f_T(x)$ and under $f_{T_i}(x)$ are equal to 1 as sums of all densities.

Example 2

To compute an area of $f_T(x) \cap f_{T_1}(x)$ from Ex. 1, we should at first find the intersection point between $f_T(x)$ and $f_{T_1}(x)$. This generates an equation $\frac{1}{0.9\sqrt{2\pi}} e^{\frac{-(x-3.0)^2}{2(0.9)^2}} = \frac{1}{1.2\sqrt{2\pi}} e^{\frac{-(x-4.2)^2}{2(1.2)^2}}$, which has no exact solution over $X = (-\infty, \infty)$. The domain $X = (-\infty, \infty)$ should be replaced by a continuous but closed interval, and the exponential function should be approximated by a polynomial possessing almost the same shape.

4. APPROXIMATIONS OF *F*-DENSITY FUNCTIONS BY π-FUNCTIONS

To fulfill the suggestions from Ex. 2, we prepare the model of approximation of the Gaussian function $f(x,\mu,\sigma) = f(x)$ by the π-function expanded as an equation

$$f(x) \approx \pi(x, \alpha_1, \beta_1, \gamma_1 = \alpha_2, \beta_2, \gamma_2, \varepsilon) = \begin{cases} 2\varepsilon\left(\frac{x-\alpha_1}{\gamma_1-\alpha_1}\right)^2 & \text{for} \quad \alpha_1 \le x \le \beta_1, \\ \varepsilon\left(1-2\left(\frac{x-\gamma_1}{\gamma_1-\alpha_1}\right)^2\right) & \text{for} \quad \beta_1 \le x \le \gamma_1, \\ \varepsilon\left(1-2\left(\frac{x-\alpha_2}{\gamma_2-\alpha_2}\right)^2\right) & \text{for} \quad \alpha_2 \le x \le \beta_2, \\ 2\varepsilon\left(\frac{x-\gamma_2}{\gamma_2-\alpha_2}\right)^2 & \text{for} \quad \beta_2 \le x \le \gamma_2, \end{cases} \tag{4}$$

in which α_1, β_1, γ_1, α_2, β_2, γ_2 and ε are parameters and $\beta_1 = \frac{\alpha_1+\gamma_1}{2}, \beta_2 = \frac{\alpha_2+\gamma_2}{2}$.

The graph of the π-function looks like a bell and has intersection points with the x-axis in (α_1, 0) and (γ_2, 0). After adjusting parameters of the π-function to the parameters of $f(x,\mu,\sigma)$, the π-function is a very efficient tool of approximating the density function $f(x)$ with a little error.

We suggest that $\alpha_2 = \gamma_1 = \mu$. The value of ε is provided by $\varepsilon = f(\mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\mu-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}$.

$\pi(x)$ should approximate $f(x)$ with a large precision. Thus, a multiplier k is added to σ. The product $k\sigma$ is a factor of the domain of the π-function, decided as $(\mu - k\sigma, \mu + k\sigma)$. We also calculate $\alpha_1 = \mu - k\sigma$, $\gamma_1 = \alpha_2 = \mu$, $\beta_1 = \frac{(\mu-k\sigma)+\mu}{2} = \frac{2\mu-k\sigma}{2}$, $\gamma_2 = \mu + k\sigma$ and $\beta_2 = \frac{(\mu+k\sigma)+\mu}{2} = \frac{2\mu+k\sigma}{2}$. Equation (4) takes a new shape, which is expressed in parameters of μ, σ and k, as (Rakus-Andersson and Frey, 2014)

$$f(x) \approx \pi(x,\mu,\sigma,k) = \begin{cases} 2\frac{1}{\sigma\sqrt{2\pi}}\left(\frac{x-(\mu-k\sigma)}{k\sigma}\right)^2 & \text{for} \quad \mu - k\sigma \le x \le \frac{2\mu-k\sigma}{2}, \\ \frac{1}{\sigma\sqrt{2\pi}}\left(1-2\left(\frac{x-\mu}{k\sigma}\right)^2\right) & \text{for} \quad \frac{2\mu-k\sigma}{2} \le x \le \mu, \\ \frac{1}{\sigma\sqrt{2\pi}}\left(1-2\left(\frac{x-\mu}{k\sigma}\right)^2\right) & \text{for} \quad \mu \le x \le \frac{2\mu+k\sigma}{2}, \\ 2\frac{1}{\sigma\sqrt{2\pi}}\left(\frac{x-(\mu+k\sigma)}{k\sigma}\right)^2 & \text{for} \quad \frac{2\mu+k\sigma}{2} \le x \le \mu + k\sigma. \end{cases} \tag{5}$$

To find the value of k, we first adopt the property $\int_{-\infty}^{\infty} f(x)dx = 1$ (the sum of all densities in the distribution is equal to 1). This leads to the conclusion that $\int_{-\infty}^{\mu} f(x)dx = 0.5$. Hence, the integral with a half of the π-function as the integrand satisfies the same condition. This means that the equality

$$\int_{\mu-k\sigma}^{\frac{2\mu-k\sigma}{2}} 2\frac{1}{\sigma\sqrt{2\pi}}\left(\frac{x-(\mu-k\sigma)}{k\sigma)}\right)^2 dx + \int_{\frac{2\mu-k\sigma}{2}}^{\mu} \frac{1}{\sigma\sqrt{2\pi}}\left(1-2\left(\frac{x-\mu}{k\sigma}\right)^2\right)dx = 0.5 \tag{6}$$

is satisfied with a minimal error.

Equation (6) has, as an unknown, parameter k. The equation is solved exactly by means of the package Maple 9.

The error of approximating the density function $f(x)$ by $\pi(x)$ is rated as

$$Error(f(x),\pi(x)) = \left|1 - \int_{\mu-k\sigma}^{\mu+k\sigma} \pi(x)dx\right|, \tag{7}$$

since the structure $\int_{-\infty}^{\infty} f(x)dx = 1$ is an estimation of the area under $f(x)$, equal to 1. Expression $\int_{\mu-k\sigma}^{\mu+k\sigma} \pi(x)dx$ stands for the area under $\pi(x)$. The area consists of a sum of four integrals in compliance with Eq. (5).

Example 3

Let us find π-functions fitted for $f_T(x,3.0,0.9)$, $f_{T_1}(x,4.2,1.2)$ and $f_{T_2}(x,4.0,1.3)$. We prepare the following data for $\pi_T(x)$, which fits for $f_T(x)$: $\varepsilon = \frac{1}{0.9\sqrt{2\pi}} = 0.443$, $k = 2.508$, $\alpha_1 = 0.743$, $\gamma_1 = \alpha_2 = 3.0$, $\beta_1 = 1.871$, $\gamma_2 = 5.257$, and $\beta_2 = 4.129$. The domain of $\pi_T(x)$ is interval $[3 - 2.508 \cdot 0.9, 3 + 2.508 \cdot 0.9] = [0.743, 5.257]$.

The equation of $\pi_T(x)$ is provided by

$$\pi_T(x,3.0,0.9,2.508) = \begin{matrix}\pi_{T,1}(x)\\ \pi_{T,2}(x)\\ \pi_{T,3}(x)\\ \pi_{T,4}(x)\end{matrix}\begin{cases}0.886\left(\frac{x-0.743)}{2.257}\right)^2 & \text{for } 0.743 \le x \le 1.871,\\ 0.443\left(1-2\left(\frac{x-3.0}{2.257}\right)^2\right) & \text{for } 1.871 \le x \le 3.0,\\ 0.443\left(1-2\left(\frac{x-3.0}{2.257}\right)^2\right) & \text{for } 3.0 \le x \le 4.129,\\ 0.886\left(\frac{x-5.257}{2.257}\right)^2 & \text{for } 4.129 \le x \le 5.257.\end{cases} \tag{8}$$

The error of approximation of $f_T(x)$ by $\pi_T(x)$ is evaluated as 0.0000604.
Function $f_{T_1}(x)$ is approximated, due to Eq. (5), by $\pi_{T_1}(x)$ yielded as

$$\pi_{T_1}(x,4.2,1.2,2.502) = \begin{matrix}\pi_{T_1,1}(x)\\ \pi_{T_1,2}(x)\\ \pi_{T_1,3}(x)\\ \pi_{T_1,4}(x)\end{matrix}\begin{cases}0.666\left(\frac{x-1.198}{3.002}\right)^2 & \text{for } 1.198 \le x \le 2.669,\\ 0.333\left(1-2\left(\frac{x-4.2}{3.002}\right)^2\right) & \text{for } 2.699 \le x \le 4.2,\\ 0.333\left(1-2\left(\frac{x-4.2}{3.002}\right)^2\right) & \text{for } 4.2 \le x \le 5.701,\\ 0.666\left(\frac{x-7.202}{3.002}\right)^2 & \text{for } 5.701 \le x \le 7.202.\end{cases} \tag{9}$$

The approximation error is computed to be equal to 0.000193.
The third function $\pi_{T_2}(x)$ is fitted for $f_{T_2}(x)$ by equation

$$\pi_{T_2}(x,4.0,1.3,2.506) = \begin{matrix}\pi_{T_2,1}(x)\\ \pi_{T_2,2}(x)\\ \pi_{T_2,3}(x)\\ \pi_{T_2,4}(x)\end{matrix}\begin{cases}0.614\left(\frac{x-0.742}{3.258}\right)^2 & \text{for } 0.742 \le x \le 2.371,\\ 0.307\left(1-2\left(\frac{x-4.0}{3.258}\right)^2\right) & \text{for } 2.371 \le x \le 4.0,\\ 0.307\left(1-2\left(\frac{x-4.0}{3.258}\right)^2\right) & \text{for } 4.0 \le x \le 5.629,\\ 0.614\left(\frac{x-7.258}{3.258}\right)^2 & \text{for } 5.629 \le x \le 7.258.\end{cases} \tag{10}$$

The difference error between $\pi_{T_2}(x)$ and $f_{T_2}(x)$ is 0.000144.

Figure 2 contains functions $\pi_T(x)$, $\pi_{T_1}(x)$, $\pi_{T_2}(x)$ (thick lines) and $f_T(x)$, $f_{T_1}(x)$, $f_{T_2}(x)$ (thin lines).

5. THE SELECTION OF THE OPTIMAL MEDICINE

A revised version of the similarity coefficient $C(\pi_T(x), \pi_{T_i}(x))$, provided that $\pi_T(x)$ approximates $f_T(x)$ and $\pi_{T_i}(x)$ fits for $f_{T_i}(x)$, $i = 1,\ldots,n$, is proposed as

$$C(\pi_T(x), \pi_{T_i}(x)) = \frac{\text{area under } (\pi_T(x) \cap \pi_{T_i}(x))}{\text{area under } (\pi_T(x)) + \text{area under } (\pi_{T_i}(x)) - \text{area under } (\pi_T(x) \cap \pi_{T_i}(x))}, \tag{11}$$

where

$$\pi_T(x) \cap \pi_{T_i}(x) = \left\{\pi(x) : \pi(x) = \min(\pi_T(x), \pi_{T_i}(x)), x \in X\right\}. \tag{12}$$

We note that $C(\pi_T(x), \pi_{T_i}(x)) = 1$, if both density distributions will be identical (the same π-functions), and $C(\pi_T(x), \pi_{T_i}(x)) = 0$ for disjoint domains of $\pi_T(x)$ and $\pi_{T_i}(x)$ (the intersection of domains constitutes an empty set). Thus

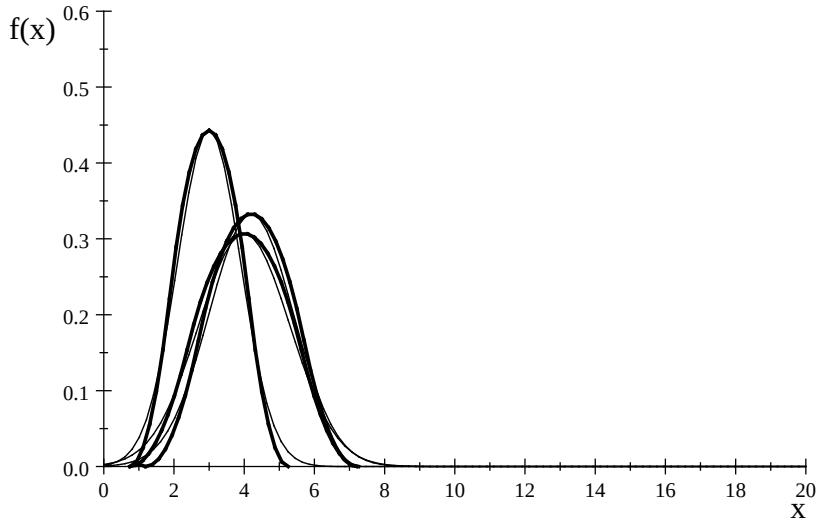

Figure 2. Functions $\pi_T(x)$, $\pi_{T_1}(x)$, $\pi_{T_2}(x)$ (thick lines) and $f_T(x)$, $f_{T_1}(x)$, $f_{T_2}(x)$ (thin lines)

$$0 \le C(\pi_T(x), \pi_{T_i}(x)) \le 1 . \tag{13}$$

It means that the lower value of *C* is followed by the greater dissimilarity of two compared normal density distributions of different states, assigned to a clinical marker with values in space *X*.

Example 4

We first evaluate $C(\pi_T(x), \pi_{T_1}(x))$. The process starts with finding an intersection point between $\pi_T(x)$ and $\pi_{T_1}(x)$. Since we have suggested the existence of only one common point, then we will solve the following equations: $\pi_{T,3}(x) = \pi_{T_1,1}(x)$ - no solution in [3.0, 4.129]∩[1.198, 2.669] = 0, $\pi_{T,3}(x) = \pi_{T_1,2}(x)$ - solution is 3.83, $\pi_{T,4}(x) = \pi_{T_1,1}(x)$ - no solution in the common domain, $\pi_{T,4}(x) = \pi_{T_1,2}(x)$ - no solution in the common domain. Hence,

$$\text{area under } (\pi_T(x) \cap \pi_{T_1}(x)) = \int_{1.198}^{2.669} \pi_{T_{1,1}}(x)dx + \int_{2.669}^{3.83} \pi_{T_{1,2}}(x)dx + \int_{3.83}^{4.129} \pi_{T,3}(x)dx + \int_{4.129}^{5.257} \pi_{T,4}(x)dx = 0.543 .$$

The integrands have been prepared by the Matlab program.

We note that $\text{area under } (\pi_T(x)) = 1 - error(f_T(x), \pi_T(x)) = 0.9999396$ and $\text{area under } (\pi_{T_1}(x)) = 0.999807$ due to Eq. (7). Finally, $C(\pi_T(x), \pi_{T_1}(x)) = \frac{0.543}{0.9999396+0.999807-0.543} = 0.3727$.

By following the same procedure, we obtain $C(\pi_T(x), \pi_{T_2}(x)) = 0.4419$.

Since the lower value of the similarity coefficient assists the more effective treatment, then we will decide that Treatment 1 has the better influence on the survival length in patients who were cured by this therapy.

6. CONCLUSION

The authors' research field is the application of computational intelligence to medicine. Especially, medical decisions have been made in medication, where a hierarchy of therapies, used in a morbid unit, has been stated (Rakus-Andersson, 2008, 2012; Rakus-Andersson and Frey, 2014). The practical examples contained clinical data in these models.

For the first time we have studied the behavior of biological parameter "*survival length*" in the context of its density of appearance in some patient groups. The density function, indicating survival in patients without

provided treatment, has been set together experimentally with two other density functions. These have been formed for survivals among patients cured by two alternative treatments.

To judge similarity of two density functions, we have designed "*similarity coefficient*" as a proportion between the area under the common part of functions and the total area under both of them. We have not been able to computationally work with infinite domains of Gaussian functions; therefore we have adjusted π-functions as their thorough approximations with domains, defined in the form of close and continuous intervals. In our experiment, a computer program easily solves all issues, sketched for polynomials. The similarity coefficient is located in interval [0, 1], which agrees with the full dissimilarity of functions for value of 0 and full likelihood for value of 1.

We think that the similarity coefficient, proposed by us, has some advantages. We can create it for an arbitrary amount of pairs, the computations are simple and, most of all, we do not introduce any hypothetical uncertainty. The obtained values of the coefficients are clearly interpretable without being burdened by confidence thresholds.

We believe that the model can help patients and physicians to make a consent and proper therapeutic decision. It could be a valuable tool for patients, e.g., at the same age who will be allowed comparing prognosis of the disease in the current stage with treatment results, achieved in other groups. In this way we may avoid referring to the disease itself, as it is done today. The graphical and computational form of earlier analysed data can also facilitate the reception of treatment effects expected.

In this paper, we have used theoretical data, since our practical experiment contained too little number of patients.

In the future works, we intend to name exactly a disease and treatments after inspecting larger data records. Another task is to suggest a value of the similarity coefficient in order to reconstruct an unknown survival density function after a certain treatment, when comparing it to the reference density function. This allows prognosticating dynamically some demands made for treatments.

REFERENCES

Bhattacharyya, A., 1943. On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distributions. *In Bulletin of the Calcutta Mathematical Society*, Vol. 35, pp. 99-109.

Davis, C. H, 1857. *Gauss's "Teoria Motus"*. Little, Brown and Company, Boston, USA.

Hoekstra, R. et al., 2014. Robust Misinterpretation of Confidence Intervals. *In Psychonomic Bulletin Review*, in press.

Inman H. and Bradley E., 1989. The Overlapping Coefficient as a Measure of Agreement between Probability Distributions and Point Estimation of the Overlap of Two Normal Densities. *In Communications in Statistics - Theory and Methods* Vol. 18, Issue 10, pp. 3851-3874.

Krishnamoorthy, K., 2006. *Handbook of Statistical Distributions with Applications.* Chapman & Hall/CRC., London, England, ISBN: 1-58488-635-8.

Mishra, S. N. et al., 1986. Overlapping Coefficient: the Generalized t Approach. *In Communications in Statistics - Theory and Methods*, Vol. 15, Issue 1, pp. 123-128.

Morey, R. D. et al., 2015. *The Fallacy of Placing Confidence in Confidence Intervals.* In press.

Neyman, J., 1937. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *In Philosophical Transactions of the Royal Society*, Vol. 236, pp. 333-380.

Rakus-Andersson, E., 2008. Decision-making Techniques in Ranking of Medicine Effectiveness. *In Advanced Computational Intelligence Paradigms in Healthcare 3*, Springer-verlag, Berlin Heidelberg, pp. 51-73.

Rakus-Andersson, E., 2012. The Parametric *s*-functions and the Perceptron in Gastric Cancer Surgery Decision Making. *In Proceedings of WCCI 2012 World Congress*, Brisbane, Australia, IEEE Computational Intelligence Society, pp. 1852-1859.

Rakus-Andersson, E., 2014. Complex Control Models with Parametric Families of Fuzzy Constrains in Evaluation of Resort Management System. *In Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 18, Issue 3, fujipress, pp. 271-279.

Rakus-Andersson, E. and Frey, J., 2014. The Choquet Integral Applied to Ranking Therapies in Radiation Cystitis. *In Proceedings of IEEE IS 2014*, Warszawa, Poland, Springer, Berlin-Heidelberg, pp. 443-452.

Stigler, S. M., 1986. *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard University Press, Cambridge, Massachusetts, USA, ISBN: 0-674-40340-1.