Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 5, Issue Special (2012): 83-95

RECENT DEVELOPMENTS ON TESTING IN CANCER RISK: A FRACTAL AND STOCHASTIC GEOMETRY

M. Stehlík[1], T. Mrkvička[2], J. Filus[3] and L. Filus[4]

[1]Johannes Kepler University Linz, Austria,
[2]University of South Bohemia, Czech Republic,
[3]Oakton Community College, USA, [4]Northeastern Illinois University, USA
Email: Milan.Stehlik@jku.at

Abstract

The aim of this paper is to discuss recent development on testing in cancer risk. We consider both area of fractal and stochastic geometry based cancer. We introduce the exact distributions of the likelihood ratio tests of several recently used tests and discuss their properties. We also show possibility of testing for cancer using some stochastic geometry descriptors. Tests for some new stochastic models in cancer risk are also given.

Key words: bivariate pseudoexponential distribution, breast cancer, cancer testing, intrinsic volumes, likelihood ratio test, pathology, random closed sets, statistical testing, stochastic dependencies.

1. Introduction

In analytical cancer research we typically meet two types of cancer. One type is related with tumor growth and we can meet boundary problems of evolution, the second one is related to the structure of the tissue which can be observed from the 2D images obtained from histology. In this paper we discuss recent papers related to both areas.

When we consider fractal based cancer diagnostic, many times a statistical procedure to assess the fractal dimension is needed. We shall look for some analytical tools for discrimination between cancer and healthy ranges of fractal dimensions of tissues. Baish and Jain (2000) investigated planar tissue preparations in mice observing the remarkably consistent scaling exponents (fractal dimensions) for tumor vasculature even among tumor lines that have quite different vascular densities and growth characteristics. One potential problem is the reconstruction of fractal dimension from boundary which may have a practical importance for Wilms tumor (Giebel, 2008). The recent tests in this area are discussed in section 2.

On the other hand, in previous investigations, it has been shown that the texture of mammary tissue, as seen at low magnification, may be characterized quantitatively in terms of stochastic geometry (see Mattfeldt (2003) and references therein). In Mrkvička and Mattfeldt (2011) the images of the mammary cases were tested for compatibility with a Boolean model (20 cases of mastopathy and 20 cases of mammary cancer, each with 10 images). In section 3 we show possibility of testing for cancer using some stochastic geometry descriptors.

2. Testing in Pseudoexponential Model

Relatively new, stochastic models such as the bivariate pseudo generalized gamma, in particular the "Gumbelian type" of pseudoexponential bivariate survival functions of, say random vector (X_1, X_2), are applied in medical cancer investigations. We consider the medical trials in which the (random) sizes of tumor X_1, X_2 in two different patient populations are considered, first of all, with respect to their (in general, stochastic) dependence on a variety of circumstances. The typical such circumstances are the kind of the treatment, as well as some additional positive or negative stresses like a diet, smoking tobacco or excessive use of alcohol.

The "natural" stochastic independence and identity (referred to as 'homogeneity') of corresponding probability distributions of the tumor sizes X_1, X_2, as measured in two groups of independently chosen patients, "breaks up" for two parallel reasons that we consider in this paper.

The reason for the dependence may be use of a common medication for both the groups that changes the tumor's behavior in both comparing to untreated (independent) cases. The second reason, for non-homogeneity, may be viewed as, say, "stochastic reaction" on the stresses that only one of the groups has, possibly, experienced. Notice, the two corresponding experiments may either be performed separately or together.

The data, i.e., measurements of the tumor sizes taken in such medical trials from the two populations, does not, in general, show well fit to the common bivariate normal distribution. However, the marginal X_1, X_2 samples often indicate close algebraic relationship to the exponentials. If they are yet dependent, the problem of such a stochastic dependence modeling until almost recently was a big problem often only solved by "forcing" the real life data to "obey" always "the same" bivariate normal distribution's pattern.

To go around this "necessity" we have chosen the, relatively recently established in literature (Filus and Filus 2006) bivariate pseudoexponential distribution of say (X_1, X_2). The strength of this new class of distributions (as well as other, similar, models that recently were published (see Filus and Filus, 2007 and 2008) in stochastic modeling follow from the fact that the general way of the dependencies formal description obeys basically the same paradigm of conditioning (a parameter of a probability distribution of X_2 was set to be a function of x_1, when $X_1 = x_1$) as the Gaussian models, while the enhancement of flexibility in various applications is very significant.

Since the dependencies between the two patient populations are "mutual" we have chosen the 'first Gumbel bivariate exponential distribution' that can be proven to be a separate version of the pseudoexponential class of distributions with the possibility as to comprise mutual "physical" (not only stochastic) influence of, say, quantities X_1 and X_2. This, however old, (Gumbel) distribution (see Gumbel, 1960), viewed as a kind of the pseudoexponential, gains a wide range of practical interpretations that until quite recently was lacking of. In the paper, it is shown briefly

that the considered Gumbel distribution may, as an applied model, be widely extended to the class of the distributions that, between others, contains its Weibullian version.

Getting back to the main stream of our investigation that, basically, is of statistical matter, we must at this point emphasize that in the applied here pseudoexponential model, the practical biomedical situations, mathematically, are reflected as follows. The existence and magnitude of the stochastic dependence is expressed by an "additional" parameter that we denoted throughout by 'A' (as the case A = 0, corresponds to independence), while the fact of homogeneity is described by equality of means $\mu_1 = \mu_2$.

For pseudoexponential dependence we introduce in Filus, Filus and Stehlík (2009) the following assumptions:

(A1) We have two groups of planar tissues.

(A2) We observe the norms $\| y_{1,i} \|, \| y_{2,i} \|, i = 1,..., N$ of planar points $y_{1,i}, y_{2,i}, i = 1,..., N$ from tissues 1,2.

(A3) We make a log transform to obtain exponentially distributed random variables $x_{j,i} = \alpha_j \log \| y_{j,i} \|, j = 1,2; i = 1,..., N$ with the joint density given by a pseudoexponential model with survival function

$$P(X_1 > x_1, X_2 > x_2) = \exp(-\theta_1 x_1 - \theta_2 x_2 - \theta_2 A\phi(x_1)x_2) \quad (1)$$

introduced by Filus and Filus (2007). Here θ_1, θ_2 denote the marginal scales and ϕ denotes the dependence function. Parameter A takes on nonnegative values only [otherwise, (unless some additional, in general artificial, conditions were imposed) the values of the survival function (i.e., the probabilities) might easily become greater than one or approach infinity] and possibly also, that the range of its values is (roughly) determined by the range of its initial estimator (maximum likelihood, for example).

We use the Pareto model for the norms of observed points, according to diffusion in fractals (see e.g. Filus, Filus and Stehlík, 2009). Pareto model $P(\lambda, \gamma)$ is a typical model for data with Pareto tails given by the cdf

$$F(x) = 1 - \left(\frac{\lambda}{x}\right)^{\gamma}, \; x > \lambda,$$

where $\gamma > 0$ is the shape parameter that characterizes the tail distribution and $\lambda > 0$ is the scale parameter. In our setup, $\lambda_j = 1$ and $\dim_1 = \theta_1 = \gamma_1$ and $\dim_2 = \theta_1(1 + A\phi(x_1))$.

(A4) We assume relatively small level of dependency given by

$$\sup_{x_{1,i}} | A\phi(x_{1,i}) | \leq 0.03 \quad (2)$$

which comes from the tolerance given by Baish and Jain (2000).

This assumption can be flexible changed for the new conditions coming from medical diagnostics. Under assumptions introduced above we derived an exact likelihood ratio test of Hausdorff dimension homogeneity hypotheses

$$H_0 : \theta_1 = \theta_2 \ \& \ A = A_1 \ \ vs. \ H_1 : \theta_1 \neq \theta_2 \ \& \ A = A_1, \qquad (3)$$

where $A_1 \neq 0$ is a dependency level, i.e. testing, whether the fractal dimensions of two issue samples are both in the tolerance given by Baish and Jain (2000). Theorem 1 in Filus, Filus & Stehlík (2009) gives the following: Let $\| y_{1,i} \|, \| y_{2,i} \|, i = 1,...,N$ be the sample and let $x_{j,i} = \gamma_j \log \| y_{j,i} \|, j = 1,2; i = 1,...,N$ be the dependent sample according to the dependence structure under the assumptions A1-A4. Then the likelihood ratio test statistics $-\ln \Lambda$ of the hypothesis (3) has the form

$$-2N \ln 2N + 2N \ln \sum_{i=1}^{N} \left(x_{1,i} + x_{2,i} + A\phi(x_{1,i}) x_{2,i} \right) +$$

$$-N \ln \sum_{i=1}^{N} x_{1,i} - N \ln \sum_{i=1}^{N} \left(x_{2,i} + A\phi(x_{1,i}) x_{2,i} \right).$$

The statistical properties of latter introduced test have been studied with extensive simulations in Stehlík et. al (2011a). It has been shown that statistical power is high (e.g. 0.9 for 15 observations) for dependence function of the form $\phi(x) = x^{\delta}, \delta < 1.1$ which brings new light on hazard function inference in this framework. Also a misspecification study has been conducted, where ML estimator of unknown parameters of pseudoexponential dependence has been evaluated. We may input the estimators of γ_j to the test. We can use the maximum likelihood estimator for estimation of parameter *A* in model (1). For $\phi(\gamma, u) = g(u)$ we get a Hausdorff dimension γ invariant test under the H_0. For instance, test is invariant on dimensions when we choose the appropriate $\phi(x) = C$. Here *C* is an appropriate value which should be estimated for a given data.

When we are using pseudoexponential model which leads to the non invariant test, we should estimate γ_1, γ_2. Notice, that estimation of γ's corresponds to the estimation of Pareto tail, which is a complicated problem (see Stehlík et. al, 2010)

2.1. Testing for the Range of Fractal Dimension

If tissues from both groups satisfy the homogeneity, i.e. the hypothesis of homogeneity $\theta_1 = \theta_2$ is not rejected at an appropriate size of the test, one can examine whether the common dimension, say θ, comes from the healthy range (e.g. in the case of Baish and Jain (2000) it means that θ varies in the range 1.7 ± 0.03) or cancer range (1.89 ± 0.04).

We have derived the exact LR test of the hypothesis

$$H_0 : \theta_1 = \gamma_0 \text{ versus } H_1 : \theta_1 \neq \gamma_0.$$

together with the exact power function. The original data points are sometimes measured in the big units, however, for our analysis smaller units could be better, e.g. centimeters, since our method is using the asymptotical stable law of continuous jump density of Lévy flight.

2.2. The Independent Case

If we deny dependence in pseudoexponential model, then we may consider observations to be a sample, which may vary at most in scale parameter, from gamma family. The basic model, which is both flexible and parametric, is a generalized gamma distribution. Let us consider a sample from a generalized gamma distribution (ggd) density, introduced by Stacy (1962), of the form

$$f(y_i \mid \vartheta_i) = \frac{\alpha}{\sigma_i \Gamma\left(\frac{1+\beta}{\alpha}\right)} \left(\frac{y_i}{\sigma_i}\right)^{\beta} \exp\left(-\left(\frac{y_i}{\sigma_i}\right)^{\alpha}\right), \quad y_i > 0, i = 1, 2, ..., N$$

and $\vartheta_i = (\alpha, \beta, \sigma_i)$, where $\alpha, \beta > 0$ are shape parameters and $\sigma_i > 0$ is a scale parameter. Special cases of generalized gamma are one-sided normal, χ^2_n, Weibull, gamma, lognormal distribution (limit case). We discuss an exact LR test of the homogeneity hypothesis

$$H_0 : \sigma_1 = \ldots = \sigma_N \text{ vs. not } H_0 .$$

In Stehlík, 2003, 2006 and 2008 the exact LR test for homogeneity has been derived

$$H_0 : \sigma = \sigma_0 \text{ vs. } H_1 : \sigma \neq \sigma_0 .$$

Under H_0 the r.v's $y_1, \ldots, y_N$ are iid according to generalized gamma distribution with an unknown common scale parameter σ. Then the LR homogeneity statistic $-\ln \Lambda_N$ has under H_0, the same distribution as the random variable

$$-\left(\frac{\beta+1}{\alpha}\right) \ln\{N^N u_1 \ldots u_{N-1}(1-u_1-\ldots-u_{N-1})\},$$

where the vector $(u_1, \ldots, u_{N-1})$ has a generalized Beta distribution $B\left(\frac{\beta+1}{\alpha}, \ldots, \frac{\beta+1}{\alpha}\right)$ on the simplex $\{u : 0 < u_1 < 1, \ldots, 0 < u_{N-1} < 1-u_1-\ldots-u_{N-2}\}$.

The distribution of LR is under the homogeneity independent on unknown scale parameter of sample (this is an advantage against some asymptotical tests).

The widely used approach (for instance see chapter 19.4 in Balakrishnan and Basu (1996) among others) is to use a LRT statistic $2\ln\Lambda = 2(L(\hat{\vartheta}_1) - L(\hat{\vartheta}_0))$ where $\hat{\vartheta}_0$ and $\hat{\vartheta}_1$ are the MLE's of the scales under the null and alternative. This plug-in LRT parameter estimation is usually accomplished by the EM algorithm, however the calculation of the test statistic and the Monte-Carlo simulation of its null distribution depend heavily on the particular implementation of the EM algorithm, e.g. starting and stopping strategies (see Seidel et al. (2000)). But this is definitely not needed in exponential case (despite the normal one) since we have a pivotal LR test statistics with known exact distribution (see Stehlík (2003) and (2006)). This test is even AOBS (see Rublík, 1989a,b; Stehlík, 2003). The small and mid sample properties of this test, also called ELRH and test against the 2 component mixture for exponential complete samples have been studied by Stehlík and Wagner (2011).

Scale testing has been developed in Stehlík 2003, 2006 and 2008. The exact cumulative distribution function of the Wilks statistics $-2\ln\Lambda = 2G_N(\sum_{i=1}^{N}(\frac{y_i}{\sigma_0})^{\alpha}) - 2G_N(N)$ of the LR test of the hypothesis

$$H_0 : \sigma = \sigma_0 \text{ vs. } H_1 : \sigma \neq \sigma_0.$$

has under H_0 the form

$$F_{\nu N}^{\Gamma}\left\{-\nu N W_{-1}(-e^{-1-\frac{\tau}{2\nu N}})\right\} - F_{\nu N}^{\Gamma}\left\{-\nu N W_0(-e^{-1-\frac{\tau}{2\nu N}})\right\}, \tau > 0$$

where $\nu = (\beta+1)/\alpha$. The Lambert W function is defined to be the multivalued inverse of the complex function $f(y) = ye^y$. As the equation $ye^y = z$ has an infinite number of solutions for each (non-zero) value of $z \in \mathbf{C}$, the Lambert W has an infinite number of branches. Here W_0 & W_{-1} denote the real valued branches.

2.2.1 I-divergence Decomposition

Let y_i are iid $\Gamma(\nu_i, \gamma_i)$. We define the *I*-divergence of the observed vector y in the sense of Pázman (1993) as

$$I_N(y,\gamma) := I(\hat{\gamma}_y, \gamma) = -\sum_{i=1}^{N}\{\nu_i - \nu_i \ln(\nu_i)\} + \sum_{i=1}^{N}\{y_i\gamma_i - \nu_i \ln(y_i\gamma_i)\}.$$

In Stehlík (2003) is proved that in distribution

$$I_N(y, \gamma(0)) = -\ln\Lambda_s \oplus (-\ln\Lambda_H \mid \gamma_1 = \ldots = \gamma_N)$$

where $\gamma_i = 1/\sigma_i, \gamma(0) = (\gamma_0, \ldots, \gamma_0)$ and $\oplus$ denotes that the summed variables are independent.

Thus, in the case of KL divergence and gamma family, we have an interesting decomposition of KL divergence from observed vector to the canonical hypothesized parameter to the LR statistics of scale and homogeneity discussed above. In Karagrigoriou and Mattheou (2010) a generalized family of measures of divergence are investigated and applied successfully in statistical inference. Similar deconvolution ideas will be of further interest also for such families of divergences.

2.2.2 Exact LR Tests with Missing and Censored Data

In biomedical experiments, for many reasons, missing and censored samples are typical. In Stehlík 2007 and 2009 we constructed exact LR test for the case of known missing function. Let $y = (y_1, .., y_n)$ be the complete sample of time-to-failures. Let $\Phi(y) = \left(\Phi_1(y), \Phi_2(y), ..., \Phi_k(y)\right)$ be the smooth missing function, $k < n$ and such that $\Phi_j(y)$ being independent, not necessarily identically distributed according to the generalized gamma with known shape parameters α, β and unknown scale parameter. In Stehlík 2007 and 2009 we constructed the exact LR test for the $\Phi(y)$ affine with the data, e.g. for the Pareto we have $\Phi_j(y) = \ln(y_{j1}) + ... + \ln(y_{jr_j})$.

In Balakrishnan and Stehlík (2008) we derived the exact LR test for scale parameter under Type II, progressively Type II and Type I inference.

2.2.3 Application to Wilms Tumors

In Stehlík et. al (2011b) we illustrate the application of fractal geometry in medicine, especially testing for range of Hausdorff dimension. The studies recently done in medicine show fractals can be applied for oncology and the description of pathological architecture of tumors. This fact is not surprising, as due to the irregular structure, cancerous cells can be interpreted as fractals. Cancer diagnosis can be done in special cases via determination of fractal dimension.

3. Testing Cancer from Tissue Images

Here we assume that we have two samples of 2D images tissues. One sample is taken from malignant tumor tissues and second one is taken from masthopathic tumor tissues. The aim here is to construct a statistical test, which is able to distinguish between the two groups and decide for a possibly new image if it belongs to masthopathic group or not. The important point here is to find a good descriptor of the image which differs for the two group as much as it is possible. Mattfeldt (2003) described several descriptors of the image. First order descriptors, like volume fraction of epithelium and surface area of epithelium; Second order descriptors, like pair correlation function, contact density function and statistic of the Laslett test; Descriptors based on correlation dimension. In this Section we describe two other descriptors and illustrate the possibility of using them for testing the cancer.

In previous investigations, it has been shown that the texture of tissue, as seen at low magnification, may be characterized quantitatively in terms of stereology (Mattfeldt 2003). Basically, glandular tissue may be subdivided into three phases, namely the epithelial cells (the tumor cells), the lumina, and the stroma, which together account for 100% of the tumor tissue. These three phases may be understood as random closed sets (RACS) with positive volume fraction (volume processes). Applying methods of spatial statistics to digitized images, or by simple manual counting methods, it is possible to characterize these three phases quantitatively in terms of area fraction A_A, boundary length density L_A and Euler number density χ_A (see e.g. Mattfeldt et al., 2007). The Euler number can be calculated as the number of isolated components minus the number of holes. The three aforementioned specific intrinsic volumes have a clear stereological interpretation; hence they can be used for the estimation of stereological model parameters:

$$\hat{V}_V = A_A,\ \hat{S}_V = \frac{4}{\pi} L_A,\ \hat{M}_V = 2\pi\chi_A$$

where V_V is the volume fraction, S_V is the mean surface area per unit reference volume, and M_V is the curvature density (integral of mean curvature per unit volume); by $\hat{V}_V$, $\hat{S}_V$ and $\hat{M}_V$ we denote the estimators of these quantities.

However, a RACS is not uniquely characterized by the specific intrinsic volumes. They inform basically about the amount of the phases per unit volume, but not about the pattern in which the features are arranged ('histological texture, architecture'). A useful nonparametric way to describe the tissue texture in this sense consists in the estimation of second-order statistics of the RACS (Mattfeldt, 2003). These summary statistics provide a quantitative characterization of the inner order of the structure in terms of attraction (clustering) and repulsion as a function of distance, without

specifying any particular stochastic model. The specific intrinsic volumes computed from parallel sets of RACS (e.g. the Minkowski sum of the RACS and the disk with radius r) with varying radius of the parallel set r (Mrkvička, 2009) reflect also the inner order of the structure and furthermore these statistics are based on the specific intrinsic volumes. Therefore the two new described descriptors are based on the specific intrinsic volumes computed from parallel sets of RACS.

The descriptor, which was founded to be the most promising, is Euler number density, but not its single value but the standard deviation of Euler number densities computed from the set of images belonged to one tumor case. Since the images for masthopathic case are more structured and organized, the Euler number densities for masthopathic case are concentrated closely to zero. On the other side the Euler number densities for malignant case reveal much bigger variability.

The second most promising descriptor was observed to be area fraction of parallel set of the original image. Here we used averages of area fractions per case. Since the images for masthopathic case are more structured and organized, the area fraction of parallel set does not change much. On the other side the area fraction of parallel set for malignant case increases a lot, because the disordered structure is almost filled by parallel set. It is important to use a big parallel set to distinguish the two cases.

3.1 Mammary Cancer

Forty cases of human mammary tumors submitted for histopathological diagnosis were investigated. Twenty cases were fibrous mastopathies, i.e. benign lesions where the glandular architecture of the mammary tissue within the lobules was fully preserved and the sole change consisted in an increase of fibrous tissue between the lobules. These were compared to 20 cases of invasive ductal mammary cancer, the most frequent type of breast cancer in humans. One paraffin section per case with a nominal thickness of 4 μm from the center of the lesion was stained with hematoxylin and eosin. Ten visual fields per case from the lobular parenchyma were evaluated in the group of mastopathies at $10\times$ primary magnification at the level of the objective of the light microscope by systematic random sampling. Ten visual fields per case from non-necrotic invasive tumor tissue were evaluated in the group of carcinomas at the same magnification by the same sampling strategy, i.e. systematic sampling with a random start. The selected visual fields were transmitted to the image analysis system Kontron IBAS 2000 with a black-and-white CCD camera. The result was a gray level image with a resolution of 512×512 pixels at a final magnification of $430\times$ on the screen. By segmentation a binary image was produced, which consisted of two phases only (Fig. 1, 2). All images were interactively segmented by the same person by tracing the epithelial formations. The epithelial component --- the union of the primary grains --- was shown as white, whereas the whole non-epithelial remainder of the tissue --- the pore space, consisting of fibrous stroma, blood vessels, nerves, gland lumina, etc. --- was shown as black. These data were also studied previously in Mattfeldt (2003) and Mrkvička and Mattfeldt (2011).

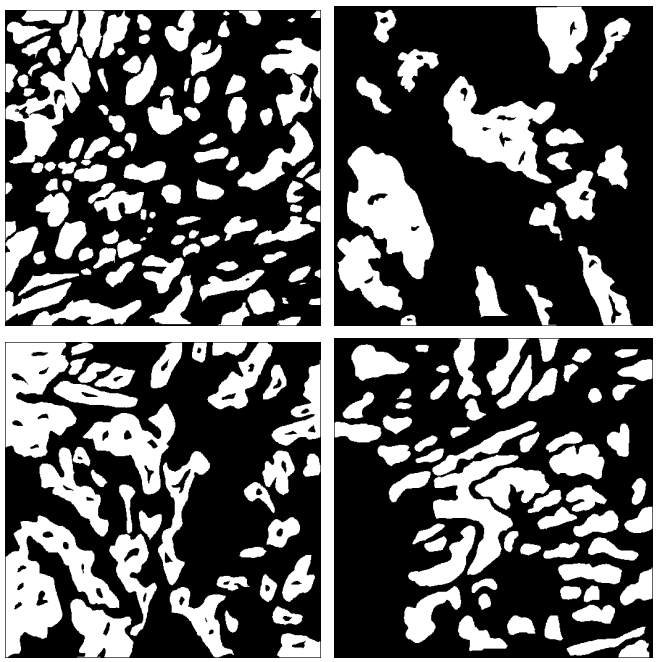

Figure 1:Four images of malignant tumor scanned in the resolution 512× 512 pixels

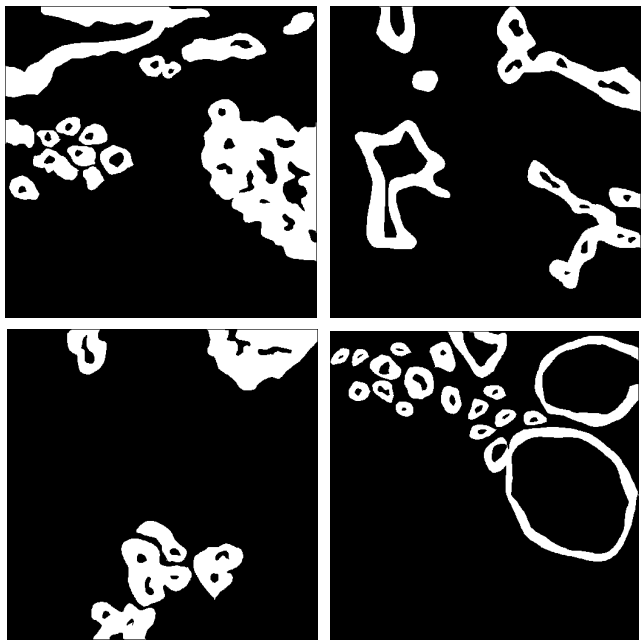

Figure 2: Four images of mastopathic tumor scanned in the resolution 512 × 512 pixels

Figure 3 left shows the standard deviations of Euler number densities computed for each masthopathic and malignant case. Figure 3 right shows the average area fractions computed from parallel sets of original images for each masthopathic and malignant case. The parallel set is made by dilation of the original image by disc with radii 24 pixels.

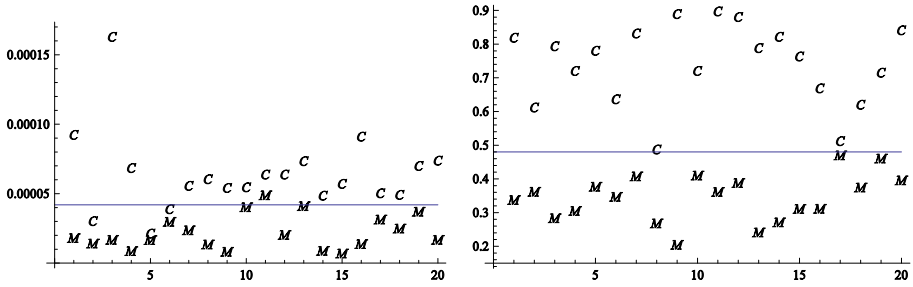

Figure 3: Left: Standard deviations of Euler number densities per case. Right: Average area fractions of big parallel sets per case. The mark C corresponds to malignant cases (Cancer) and the mark M corresponds to masthopathic cases. The horizontal lines correspond to the optimal distinguishing of the tumors.

3.2 Testing the Hypothesis of Masthopathic Case

A possibility of testing that the given case is masthopathic or alternatively malignant is to approximate the distribution of one descriptor by normal distribution. There is also a possibility to approximate the variance of Euler numbers by χ^2 distribution, but it seems that the studied variances do not follow χ^2. This is probably caused by the fact that the Euler numbers for one case do not come from normal distribution with same parameters but rather from various normal distributions.

If we denote the chosen descriptor *C*, then under the hypothesis H_0, that the case is masthopathic, we approximate the distribution of *C* by a normal distribution with mean value and standard deviation estimated from the descriptors *C* measured on already observed cases of masthopathic tumors. Denote the distribution function of this normal distribution *F*. Using both our descriptors, we can write hypothesis H_0 with one sided alternative

$$H_0 : c \leq \mathbb{E}C \text{ vs. } H_1 : c > \mathbb{E}C.$$

The *p-value* of such test then corresponds to:

$$p = 1 - F(\hat{c}),$$

where $\hat{c}$ is descriptor measured in tested tumor case.

3.2.1 Test Using the Standard Deviation of the Euler Number Density of a Case

If C is the standard deviation of the Euler number density of a case, then estimated parameters of the normal distribution under H_0 from our 20 cases of masthopathic tumor are $\mathbb{E}C = 0.0000218664$ and $\sigma_C = 0.000012309$.

This test was applied for all our cases of masthopathic and malignant tumors. The resulting *p-values* are recorded in Table 1.

Cancer - Malignant Tumor	$5.39 * 10^{-9}$	0.2498	0	0.000078	0.5220
	0.084	0.0030	0.000885	0.0044	0.0040
	0.00034	0.00034	0.000014	0.0152	0.00218
	$9.89 * 10^{-9}$	0.0107	0.013	0.000048	0.000012
Mastopathic Tumor	0.6251	0.7395	0.6633	0.86098	0.6686
	0.2637	0.4443	0.7553	0.8690	0.0687
	0.0144	0.5530	0.06069	0.8561	0.8889
	0.7476	0.2227	0.4056	0.1099	0.6627

Table 1: p-values of tests with use of the standard deviation of the Euler number density of a case

Only one *p-value* out of 20 is under 0.05 for masthopathic cases, thus the significance level is estimated to be 0.05. 3 cases of cancer have *p-value* over 0.05, thus estimated power of the test is 0.85 on the base of our studied data.

3.2.2 Test Using Average Area Fraction of Big Parallel Set of a Case

If C is average area fraction of big parallel set of a case, then estimated parameters of the normal distribution under H_0 from our 20 cases of masthopathic tumor are $\mathbb{E}C = 0.343888$ and $\sigma_C = 0.070184$.

This test was also applied for all our cases of masthopathic and malignant tumors. The resulting *p-values* are recorded in Table 2.

Cancer - Malignant Tumor	$6.60 * 10^{-12}$	0.000068	$7.51 * 10^{-11}$	$4.20 * 10^{-8}$	$2.59 * 10^{-10}$
	0.000015	$1.92 * 10^{-12}$	0.0206	$4.10 * 10^{-15}$	$4.18 * 10^{-8}$
	$1.55 * 10^{-15}$	$1.08 * 10^{-14}$	$1.27 * 10^{-10}$	$4.99 * 10^{-12}$	$1.15 * 10^{-9}$
	$1.88 * 10^{-6}$	0.0083	0.000042	$6.27 * 10^{-8}$	$7.18* 10^{-13}$
Mastopathic Tumor	0.538	0.400	0.807	0.712	0.320
	0.485	0.183	0.861	0.976	0.175
	0.402	0.267	0.929	0.851	0.682
	0.680	0.036	0.333	0.049	0.231

Table 2: *p-values* of tests with use of the average area fraction of big parallel set of a case

Two *p-values* are under 0.05 for masthopathic cases, thus the significance level is estimated to be 0.1. But one of these is just under the significance level. None of the cancer cases have *p-value* over 0.05, thus estimated power of the test is 1.00 on the base of our studied data.

4. Conclusion

There are many important issues which should be addressed in a future research.

1. The important issue is that we may estimate the Hausdorff dimension *HD* of the underlying fractal set. We know that Box Counting dimension *BCD*, typically routinely estimated, satisfies $BCD \geq HD$. Moreover, HD provides a nice relationship between HD dimensions of the slices (typically observed by pathologists) and HD of the original 3D tissue. This relationship is not known for the *BCD*. Further research should be done in model validation via real data (e.g. on the Sierpinski carpet, various ramifications), D-optimal design issues (correlated observations, for some issues see Müller and Stehlík, 2009).

2. In the particular case shown in section 3 the given descriptors showed a good possibility for distinguishing the mastopathic and malignant tissue in mammary cancer. However it would be practical to study also other types of cancer. Further possibility of testing can be obtained by employing the random closed sets model of 2D image.

Acknowledgement: We acknowledge partial support by Project 58p5 "Aktion Czech Republic-Austria:" "Testing and modelling of cancer." The authors are thankful for constructive comments of the anonymous Referees and the Editors & Guest Editors of the journal, which improved the quality of the paper. We also acknowledge comments from Stefan Giebel**.**

References

1. Baish, J.W. and Jain R.K. (2000). Fractals and cancer, Cancer Research, 60, p. 3683-3688.
2. Balakrishnan, N. and Basu, A.P. (1996). The Exponential Distribution: Theory, Methods, and Applications, Gordon and Breach, New York.
3. Balakrishnan, N. and Stehlík, M. (2008). Exact likelihood ratio test of the scale for censored Weibull sample, IFAS res. report Nr. 35.
4. Fabián, Z. and Stehlík, M. (2009). On robust and distribution sensitive Hill like method, IFAS Res. Report Nr. 43.
5. Filus, L. and Filus, J. (2006). On Some New Classes of Multivariate Probability Distributions, Pakistan J. Statist., 22, p. 21-42.
6. Filus, J. and Filus, L. (2007). On new multivariate probability distributions and stochastic processes with system reliability and maintenance applications, Method Comput Appl Prob, 9, p. 425-446.
7. Filus, J. and Filus, L. (2008). On Multicomponent System Reliability with Microshocks-Microdamades Type of Components' Interaction, Lecture Notes in Engineering and Computer Science, IMECS 2008, ISBN: 978-988-17012-1-3, Vol. II, p. 1946-1951.
8. Filus, J., Filus, L. and Stehlík, M. (2009). Pseudoexponential modelling of cancer diagnostic testing, Biometrie und Medizinische Informatik, Greifswalder Seminarberichte, Heft 15, p. 41-54.
9. Giebel, S.M. (2008). Statistical Analysis of Renal tumors with Infants, Bull Soc Sci Med Grand Duche Luxemb, 1, p. 121-30.
10. Gumbel, E.J. (1960). Bivariate exponential distribution, J. Amer. Statist. Assoc., 50, p. 698-707.

11. Karagrigoriou, A. and Mattheou, K. (2010). On Distributional Properties and Goodness-of-Fit Tests for Generalized Measures of Divergence, Comm. Statist. Theory Methods, 39, p. 472-482.
12. Mattfeldt, T. (2003). Classification of binary spatial textures using stochastic geometry, nonlinear deterministic analysis and artificial neural networks, Int. J. Pattern Recogn. Artif. Intell., 17, p. 275-300.
13. Mattfeldt, T., Meschenmoser, D., Pantle, U. and Schmidt, V. (2007). Characterization of mammary gland tissue using joint estimators of Minkowski functionals, Image Anal. Stereol, 26, p. 13-22.
14. Mrkvička, T. (2009). On testing of general random closed set model hypothesis, Kybernetika, 45/2, p. 293-308.
15. Mrkvička, T. and Mattfeldt, T. (2011). Testing histological images of mammary tissues on compatibility with the boolean model of random sets, Image Anal. Stereol, 30, p. 11-18.
16. Müller, W.G. and Stehlík, M. (2009). Issues in the optimal design of computer simulation experiments, Appl. Stochastic Models Bus. Ind., 25, p. 163-177.
17. Pázman, A. (1993). Nonlinear statistical Models, Chapters 9.1 and 9.2. Kluwer Acad. Publ., Dordrecht.
18. Rublík, F. (1989a). On optimality of the LR tests in the sense of exact slopes, Part 1, General case, Kybernetika 25, p. 13-25.
19. Rublk, F. (1989b). On optimality of the LR tests in the sense of exact slopes, Part 2, Application to Individual Distributions, Kybernetika, 25, p. 117-135.
20. Seidel, W., Mosler, K. and Alker, M. (2000). A cautionary note on likelihood ratio tests in mixture models, Ann. Inst. Statist. Math., 52, p. 481-487.
21. Stacy, E.W. (1962). A generalization of the gamma distribution, Ann. Math. Statist., 33, p. 1187-92.
22. Stehlík, M. (2003). Distributions of exact tests in the exponential family, Metrika 57, p. 145-164.
23. Stehlík, M. and Ososkov, G.A. (2003). Efficient testing of the homogeneity, scale parameters and number of components in the Rayleigh mixture, JINR Rapid Communications E-11-2003-116.
24. Stehlík, M. (2006). Exact likelihood ratio scale and homogeneity testing of some loss processes, Statist. Probab. Lett., 76, p. 19-26.
25. Stehlík, M. (2008). Homogeneity and Scale Testing of Generalized Gamma Distribution, Reliability Engineering & System Safety, 93, p. 1809-1813.
26. Stehlík, M. Potocký, R. Waldl, H. and Fabián, Z. (2010), On the favourable estimation of fitting heavy tailed data, Comput. Statist. 25, p. 485-503.
27. Stehlík, M. and Wagner, H. (2011). Exact likelihood ratio testing for homogeneity of exponential distribution, Comm. Statist. Simulation Comput., 40, p. 663-684.
28. Stehlík, M., Filus, J., Filus, L. and Korycinski, P. (2011a). Pseudo-dependence modeling in medical trials: scale homogeneity testing, Technical report.
29. Stehlík, M., Prostakova, J., Giebel, S. and Schenk, J. (2011b). Statistical inference on fractals for cancer risk assessment, Technical report.