Nonparametric Comparison of Densities Based on Statistical Bootstrap

De Brabanter, K.[1], Sahhaf, S.[2], Karsmakers, P.[1,3], De Brabanter, J.[1,4], Suykens, J.A.K.[1], De Moor, B.[1]

January 19, 2010

[1] K.U. Leuven, Department of Electrical Engineering (ESAT)-SCD
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
{kris.debrabanter, johan.suykens, bart.demoor}@esat.kuleuven.be

[2] IMEC
Kapeldreef 75, 3001 Leuven, Belgium
sahar.sahhaf@imec.be

[3] K.H.Kempen (Associatie K.U.Leuven), Dep. IBW
Kleinhoefstraat 4, 2440 Geel, Belgium
peter.karsmakers@khk.be

[4] KaHo Sint Lieven (Associatie K.U.Leuven), Dep. Industrieel Ingenieur
G. Desmetstraat 1, 9000 Gent, Belgium
jos.debrabanter@kahosl.be

Abstract

In this paper we study which user specified parametric density best fits a given data set. It is quite common in the field of reliability analysis of MOSFETs to model failure times with a Weibull or a lognormal distribution. However, sometimes it can be quite hard to decide which density best explains the data. We develop a hypothesis test to determine which parametric density can be chosen. This test is based on the Kullback-Leibler divergence between a nonparametric kernel density estimator and the parametric densities. In this setting, the distribution of the test statistic is difficult to compute or cannot be determined analytically. Hence, we propose the use of statistical bootstrap combined with variance stabilization as an alternative approach.

Keywords: *Statistical Bootstrap, Variance Stabilization, Least Squares Support Vector Machines, Hypothesis Test, Kernel Density Estimation, Kullback-Leibler Divergence.*

1 Introduction

One of the key issues in the evolution of MOSFET technology is device scaling [12]. Thin oxide layers are an important part of most microelectronic devices. They serve as insulating gate material in transistors. When these very thin oxide layers are subjected to voltages of a few volts, due to the induced electrical field, a gradual degradation of the oxide properties occurs and finally device failure is followed.

Different phases before the final device failure are shown in Figure 1. Due to the stress, defects are generated in the oxide which can create conduction paths with multiple traps where current flows through. This phenomenon is observed as a small current jump and is referred to as soft-breakdown. In the following part, the wear-out phase, more defects are generated along the existing conduction paths resulting in the wear out of the paths. This phenomenon is observed as fluctuation of current until device failure occurs. At this final stage, called hard-breakdown, abrupt increase of the current (up to mA) is observed. The time-to-soft-breakdown (t_{SBD}), wear out time (t_{WO}) and the time-to-hard-breakdown (t_{HBD}) are random variables. Understanding the related distributions of oxide breakdown and its

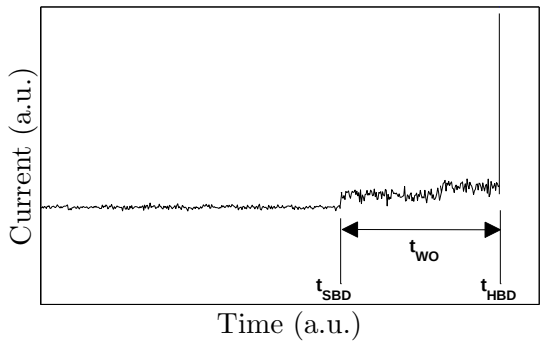


Figure 1: Illustration of the different phases of oxide degradation and breakdown in ultra-thin gate stacks (a.u. denotes arbitrary units).

preceding degradation phase are needed in order to predict the lifetime at operating conditions. SBD is well known to be Weibull distributed [27]. About the distribution of the wear-out phase however, contradictory reports can be found [14]. It is even claimed [28] that the shape of the wear-out distribution depends on the choice of the failure current criterion.

In this work we propose a hypothesis test, based on statistical bootstrap with variance stabilization and a nonparametric kernel density estimator, assisting the researcher to find the proper underlying distribution of random variables. Further, we illustrate the capability of this technique by means of toy examples.

This paper is organized as follows: Section 2 describes the Parzen-Rosenblatt density estimator and shows how a density problem can be converted to a nonparametric regression problem. The problem formulation and construction of the hypothesis test is illustrated in Section 3. Section 4 reports the results on artificial data sets. Finally, Section 5 states the conclusion.

2 Kernel Density Estimation

In the following sections we describe two possible techniques to estimate a density function f. First, we state some notations and conventions.

Throughout this paper it will be assumed that we have a sample $X_1, \ldots, X_n$ of independent, identically distributed observations from a continuous univariate distribution with probability density function f. Also $\hat{f}$ will be the kernel estimator with kernel K and window width (or bandwidth) h. For each x, $\hat{f}(x)$ can be seen as a random variable, because of its dependence on the observations $X_1, \ldots, X_n$.

2.1 Parzen-Rosenblatt Kernel Density Estimator

Probably one of the most popular and known methods to estimate a density function is the Parzen-Rosenblatt estimator [17, 18]. This kernel estimator with kernel K is defined by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1}$$

where h is called the bandwidth or smoothing parameter of the kernel. Also, the kernel K satisfies the condition

$$\int_{-\infty}^{\infty} K(x)dx = 1.$$

The estimator (1) can be considered as a sum of "bumps" placed above each observation. The resulting estimator is then the sum of all these bumps. The kernel function K determines the shape of the bumps while the bandwidth h determines their width.

The problem of choosing how much to smooth (choice of bandwidth h) is of crucial importance in any kernel based method. Various methods have been proposed for the case of density estimation. A few examples are reference rules, plug-ins, least-squares cross-validation, bootstrap based methods,... We will not discuss these methods since they are beyond the scope of this paper. More information about these methods can be found in [21, 25]. In general, when the bandwidth becomes large no details will be visible because of oversmoothing. On the other hand when the bandwidth is too small, spurious fine structures become visible.

2.2 Regression View of Density Estimation

In this Section we establish a connection between density estimation and nonparametric regression, see e.g. [9, 26]. This connection can be seen by using a binning technique. Suppose we are interested in estimating the density function f on an interval $[a, b]$. Partition the interval $[a, b]$ into N subintervals $\{I_k, k = 1, \ldots, N\}$ of equal length $\Delta = (b-a)/N$. Let x_k be the center of I_k and y_k be the proportion of the data $\{X_i, i = 1, \ldots, n\}$ falling in the interval I_k, divided by the bin length Δ. The number of subintervals can be determined by $N = \lceil (b-a)/3.49\,\mathrm{MAD}(X) n^{-1/3} \rceil$. MAD stands for Median Absolute Deviation and is defined as $\mathrm{median}(|X - \mathrm{median}(X)|)$ and $\lceil \cdot \rceil$ denotes the largest integer. It is clear that the bin counts $n\Delta y_k$, i.e. the number of sample points falling in the kth bin, have a binomial distribution [13]:

$$n\Delta y_k \sim \mathrm{Bin}(n, p_k) \quad \text{with} \quad p_k = \int_{I_k} f(x)dx, \tag{2}$$

with p_k the probability content of the kth bin. For a fine enough partition, i.e. $N \to \infty$, it can be calculated from (2), by using a Taylor series expansion of the density in the kth bin around the midpoint of the bin, that

$$\mathrm{E}[y_k] \approx f(x_k), \quad \mathrm{Var}[y_k] \approx \frac{f(x_k)}{n\Delta}. \tag{3}$$

Consequently, we could regard the density estimation problem as a heteroscedastic, i.e. the random variables have different variances (see (3)), nonparametric regression problem based on the data $\{(x_k, y_k) : k = 1, \ldots, N\}$ which are approximately independent [8]. The nonparametric regression problem is defined as follows

$$y_k = m(x_k) + \varepsilon_k, \quad \varepsilon_k = e_k \eta(m(x_k), x_k),$$

where e_k are independent and identically distributed (i.i.d.), i.e. each random variable has the same probability distribution as the others and all are mutually independent. The function η expresses the heteroscedasticity and m is an unknown real-valued smooth function that we want to estimate. Although it is possible to design techniques able of coping with heteroscedastic data, often in practice homoscedastic data are preferred. The homoscedasticity can be accomplished via Anscombe's variance stabilizing transformation [1] to the bin count $c_k = n\Delta y_k$, i.e.

$$y_k^\star = 2\sqrt{c_k + \frac{3}{8}}.$$

The density estimator is then obtained by applying a nonparametric smoother to the transformed data set $\{(x_k, y_k^\star) : k = 1, \ldots, N\}$, and taking the inverse

of Anscombe's transformation. Let $\hat{m}^\star(x)$ be a regression smoother based on the transformed data. Then the density estimator is defined by

$$\hat{f}(x) = \mathcal{C}\left[\frac{\hat{m}^\star(x)^2}{4} - \frac{3}{8}\right]_+, \tag{4}$$

where $\mathcal{C}$ is a normalization constant such that $\hat{f}(x)$ integrates to 1 and $[z]_+ = \max(z, 0)$. The following example motivates the regression view of density estimation.

Example 1 *Given the failure times of an electric device (measured in days) [16]. We want to estimate the failure density by means of the Parzen-Rosenblatt estimator and the regression view of density estimation. Figure 2 shows the result of both estimates. It is clear that the Parzen-Rosenblatt estimator (thin line) suffers from a slight drawback when applied to data from long-tailed distributions. Because the bandwidth is fixed across the entire sample, there is a tendency for spurious noise to appear in the tails of the estimates. This example shows this behavior by disregarding the fact that failure times are naturally non-negative. However, this drawback can be overcome by adaptive bandwidths [21] and/or boundary kernels [20, 25]. On the other hand, the LS-SVM based estimate (bold line) for density estimation (see also Section 2.3) can deal with this difficulty.*

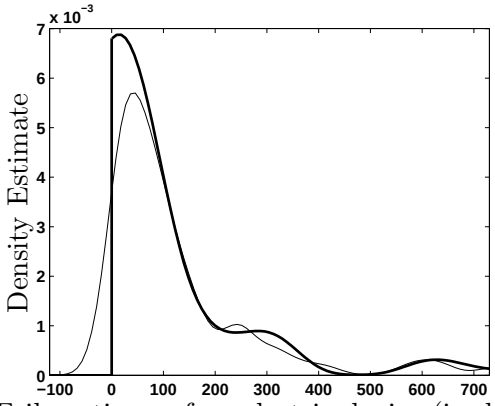


Figure 2: Density estimation of failure times of an electric device (measured in days) for the Parzen-Rosenblatt estimator (thin line) and the regression view of density estimation based on LS-SVM (bold line).

In the following section, we briefly sketch the basics of Least Squares Support Vector Machines (LS-SVM) for regression. This nonparametric technique can be used to estimate $\hat{m}^\star$.

2.3 Least Squares Support Vector Machines for Regression

We briefly illustrate the basic principles of LS-SVM. Further information regarding this topic can be found in Suykens et al. [22].

Consider the model

$$m(x) = w^T \varphi(x) + b,$$

with the feature map $\varphi : \mathbb{R}^d \to \mathbb{R}^{D_\varphi}$, d and D_φ the dimension of the input space and the feature space respectively, $w \in \mathbb{R}^{D_\varphi}$ and $b \in \mathbb{R}$. The following regularized least squares cost function can be formulated in the primal weight space

$$\begin{aligned} \min_{w,b,e} \mathcal{J}(w,e) &= \tfrac{1}{2} w^T w + \tfrac{\gamma}{2} \sum_{k=1}^{n} e_k^2 \\ \text{s.t.} \quad Y_k &= w^T \varphi(X_k) + b + e_k, \quad k = 1, \ldots, n. \end{aligned}$$

By using Lagrange multipliers [2], the dual solution is characterized by the following linear system in the dual variables α

$$\left(\begin{array}{c|c} 0 & 1_n^T \\ \hline 1_n & \Omega + \frac{1}{\gamma} I_n \end{array} \right) \left(\begin{array}{c} b \\ \hline \alpha \end{array} \right) = \left(\begin{array}{c} 0 \\ \hline Y \end{array} \right),$$

where $\Omega \in \mathbb{R}^{n \times n}$ with $\Omega_{ij} = \varphi(X_i)^T \varphi(X_j) = \mathcal{K}(X_i, X_j)$ for all $i, j = 1, \ldots, n$, $\mathcal{K}(\cdot, \cdot)$ a positive definite kernel and $Y = (Y_1, \ldots, Y_n)^T$. In the remaining of the paper we take $\mathcal{K}(X_i, X_j) = \exp(-\|X_i - X_j\|_2^2 / h)$. Based on Mercer's theorem, the resulting LS-SVM model for regression becomes

$$\hat{m}(x) = \sum_{k=1}^{n} \hat{\alpha}_k \mathcal{K}(x, X_k) + \hat{b}. \tag{5}$$

The tuning parameters h (kernel bandwidth) and γ (regularization parameter) can be found via cross-validation techniques [3].

3 Problem Formulation and Hypothesis Test

3.1 Problem Formulation

Given the observations $X_1, \ldots, X_n$, we are interested in the probability density function f. One way to proceed is to use nonparametric techniques (see Section 2) to estimate f. On the other hand, in some domains one is particularly interested in acquiring a parametric probability density function [19]. Often, one has some presumption about the form of the parametric density. For example in the field of reliability analysis of MOSFETs, Weibull, lognormal, exponential, gamma and Gumbel densities are widely used [16]. As an example, we can formulate the following problem statement: "Which of the given parametric probability density function, Weibull or lognormal, best fits our data?"

To answer this question we propose a methodology based on the comparison between a nonparametric density estimate, obtained by LS-SVM, and a parametric density (chosen by the user). Let us assume that the nonparametric density estimate is our best possible estimator of the unknown true underlying density. Simply put, we can then calculate some kind of distance measure between the nonparametric density estimate and the chosen parametric density. The parametric density corresponding to the smallest distance best fits the given data. Many types of distances between probability densities exist [24] e.g. Hellinger distance, Total Variation, Kullback-Leibler (KL) divergence,... In this paper we use the Kullback-Leibler divergence.

Definition 1 *The Kullback-Leibler divergence between two distributions P and Q is defined as*

$$\mathrm{KL}(P,Q) = \int p(x) \log \frac{p(x)}{q(x)}\, dx, \tag{6}$$

where p and q denote the densities of P and Q.

Typically P represents our best possible estimate of the "true" distribution of the data (LS-SVM estimate) and Q is the chosen parametric distribution e.g. Weibull, lognormal, etc.

Notice that we did not use the word distance but divergence for KL. This is because the axioms to be called "a distance", i.e. non-negativity, identity of indiscernibles, symmetry and subadditivity, are not all satisfied. However, links can be established between KL, Hellinger distance and Total Variation [24].

3.2 Hypothesis Test Based on Statistical Bootstrap

Let $\hat{F}$ be the LS-SVM estimate of the distribution, obtained by applying (4) and (5), and let $\hat{f}$ denote the corresponding estimated density. Also, denote by Q_1 and Q_2 the parametric distributions and by q_1 and q_2 their corresponding parametric densities. These parametric densities are chosen by the user. Consider a situation in which a random sample $\mathcal{X} = \{X_1, \ldots, X_n\}$ is observed from its unspecified probability distribution function G_θ, where θ, a characteristic of G, is unknown. We can formulate a hypothesis test as follows:

$$\begin{aligned} &\mathrm{H}_0 : \theta = \mathrm{KL}(\hat{F}, Q_1) - \mathrm{KL}(\hat{F}, Q_2) \leq \theta_0 \\ &\qquad\qquad \text{vs.} \\ &\mathrm{H}_1 : \theta = \mathrm{KL}(\hat{F}, Q_1) - \mathrm{KL}(\hat{F}, Q_2) > \theta_0 \end{aligned} \tag{7}$$

where θ_0 is some known bound, H_0 and H_1 represent the null hypothesis and the alternative hypothesis respectively. In other words, if H_0 is accepted, then the parametric density q_1 is closest to the nonparametric estimate and hence is more suited to model the density of the sample than the parametric density q_2 and vice versa.

Let $\hat{\theta}$ be an estimator of θ and $\hat{\sigma}^2$ an estimator of the variance σ^2 of $\hat{\theta}$. Define the "studentized" test statistic

$$T_n = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}}.$$

The inclusion of the scale factor $\hat{\sigma}$ ensures that T_n is asymptotically pivotal as $n \to \infty$ [10, 11]. A statistic T_n is called pivotal if it possesses a fixed probability distribution independent of θ [4, 15]. Using a pivotal statistic means that we only need to deal with the appropriate standard distribution rather than the whole family of distributions. Note, however, that pivoting often does not hold [7] unless a variance stabilizing transformation [23] for the parameter estimate of interest is applied first.

Since the distribution function of T_n under H_0 is not known or cannot be derived analytically, we can use the bootstrap principle [5, 6, 7] to test the hypothesis (7). The bootstrap principle is based on bootstrap samples. A bootstrap sample $\mathcal{X}^* = \{X_1^*, \ldots, X_n^*\}$ is an unordered collection of n sample points drawn randomly form $\mathcal{X}$ with replacement, so that each X_i^* has probability n^{-1} of being equal to any of the X_j's. Algorithm 1 states the bootstrap principle to test the hypothesis $\theta \leq \theta_0$ using variance stabilization.

Note that the constant θ_0 has been replaced in Step 2 of Algorithm 1 by the estimate of θ, i.e. $\hat{\theta}$, derived from $\mathcal{X}$. This is crucial if the test is to have good power properties. It is also important in the context of the accuracy level of the test [10, 11]. Also because we have used variance stabilization we do not need to divide by a standard error estimate in Step 2 and Step 4 since it is constant [7, 23]. Another advantage of variance stabilization is that the number of bootstrap resamples decreases significantly. Indeed, the number of bootstrap resamples in this case is $B_1 \cdot B_2 + B_3 = 100 \cdot 25 + 1000 = 3500$. In case we did not use this transformation we would have to calculate for each bootstrap resample an estimate of the standard error. Denote by B the number of bootstrap resamples to calculate the test statistic and B^* the number of resamples to calculate an estimate of the standard error, then the total number of resamples is $B \cdot B^* = 1000 \cdot 25 = 25000$.

4 Illustrative Examples

In all simulations $\hat{F}$ denotes the LS-SVM estimate of the distribution. As kernel function we used the RBF kernel and the tuning parameters of the LS-SVM (kernel bandwidth h and regularization parameter γ) are each bootstrap resample determined by leave-one-out cross-validation. The nominal level of significance α is set to 0.05, $B_1 = 100$, $B_2 = 25$ and $B_3 = 1000$ in Algorithm 1. For each bootstrap resample the parameter(s) of the parametric densities are estimated via Maximum Likelihood (ML).

Algorithm 1 Testing the hypothesis $\theta \leq \theta_0$ using variance stabilization

1: *Estimation of the variance stabilizing transformation*
   (a) Generate B_1 bootstrap samples $\mathcal{X}_i^*$ from $\mathcal{X}$ and calculate $\hat{\theta}_i^*, i = 1, \ldots, B_1$ for $B_1 = 100$.
   (b) Generate B_2 bootstrap samples from $\mathcal{X}_i^*, i = 1, \ldots, B_1$, and calculate $\hat{\sigma}_i^{*2}$, a bootstrap estimate for the variance of $\hat{\theta}_i^*, i = 1, \ldots, B_1$. Set $B_2 = 25$.
   (c) Estimate the variance function ψ by smoothing the values of $\hat{\sigma}_i^{*2}$ against $\hat{\theta}_i^*$, by using for example LS-SVM, see (5).
   (d) Estimate the variance stabilizing transformation Φ by

$$\Phi = \int_\theta [\psi(u)]^{-1/2}\, du.$$

2: *Calculation of the bootstrap statistic.* Generate B_3 bootstrap samples $\mathcal{X}_i^*$ from $\mathcal{X}$ and calculate $T_{n,i}^* = \Phi(\hat{\theta}^*) - \Phi(\hat{\theta}), i = 1, \ldots, B_3$ for $B_3 = 1000$.
3: *Ranking.* Rank the collection $T_{n,1}^*, \ldots, T_{n,B_3}^*$ into increasing order to obtain $T_{n,(1)}^* \leq \ldots \leq T_{n,(B_3)}^*$.
4: *Test.* Reject H_0 if $T_n = \Phi(\hat{\theta}) - \Phi(\theta_0) > T_{n,(q)}^*$, q determines the level of significance of the test and is given by $\alpha = \lfloor (B_3 + 1 - q)(B_3 + 1)^{-1} \rfloor$, with α the nominal level of significance [10] and $\lfloor \cdot \rfloor$ denotes the smallest integer.
5: *Back-transformation.* If required, back-transform $T_{n,i}^*$ and T_n by considering the inverse transformation of Φ, i.e. $\Phi^{-1}(\Phi(\hat{\theta}^*) - \Phi(\hat{\theta}))$ and $\Phi^{-1}(\Phi(\hat{\theta}) - \Phi(\theta_0))$ respectively.

4.1 Toy Example 1

In this toy example we generate a data set $X_i \sim \mathrm{Weib}(1, 2)$ for $i = 1, \ldots, 500$. Suppose we want to test the following hypothesis

$$\mathrm{H}_0 : \mathrm{KL}(\hat{F}, \mathrm{Weib}) - \mathrm{KL}(\hat{F}, \mathrm{Log}\mathcal{N}) \leq 0$$
$$\text{vs.}$$
$$\mathrm{H}_1 : \mathrm{KL}(\hat{F}, \mathrm{Weib}) - \mathrm{KL}(\hat{F}, \mathrm{Log}\mathcal{N}) > 0.$$

Figure 3(a) shows the lognormal (thin line) and Weibull (bold line) density for the data set with parameters estimated by ML. Figure 3(b) shows the histogram of the test statistic $T_{n,i}^*, i = 1 \ldots, B_3$. Based on $T_n < T_{n,(\lfloor (B_3+1)(1-\alpha) \rfloor)}^*$ we cannot reject the null hypothesis H_0 on a significance level $\alpha = 0.05$. This indicates that the Weibull distribution is best suited for the given data set, since its KL divergence is the smallest.

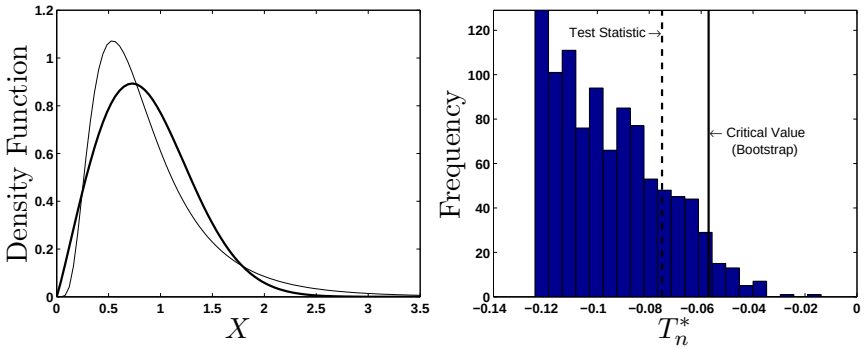


Figure 3: (a) lognormal (thin line) and Weibull (bold line) density for the generated data set with parameters estimated by ML; (b) Null distribution of the test statistic. The dotted line indicates the value of T_n and the solid line indicates the critical value based on the bootstrap samples T_n^*.

4.2 Toy Example 2

We generate a data set $X_i \sim \text{Exp}(1)$ for $i = 1, \ldots, 500$. We test the following hypothesis

$$\text{H}_0 : \text{KL}(\hat{F}, \text{Log}\mathcal{N}) - \text{KL}(\hat{F}, \text{Exp}) \leq 0$$

vs.

$$\text{H}_1 : \text{KL}(\hat{F}, \text{Log}\mathcal{N}) - \text{KL}(\hat{F}, \text{Exp}) > 0.$$

Figure 4(a) shows the lognormal (thin line) and exponential (bold line) density for the data set with parameters estimated by ML. Figure 4(b) shows the histogram of the test statistic $T_{n,i}^*, i = 1 \ldots, B_3$. Based on $T_n > T^*_{n,(\lfloor (B_3+1)(1-\alpha) \rfloor)}$ we reject the null hypothesis H_0 on a significance level $\alpha = 0.05$. This indicates that the exponential distribution best fits the given data set, since its KL divergence is the smallest.

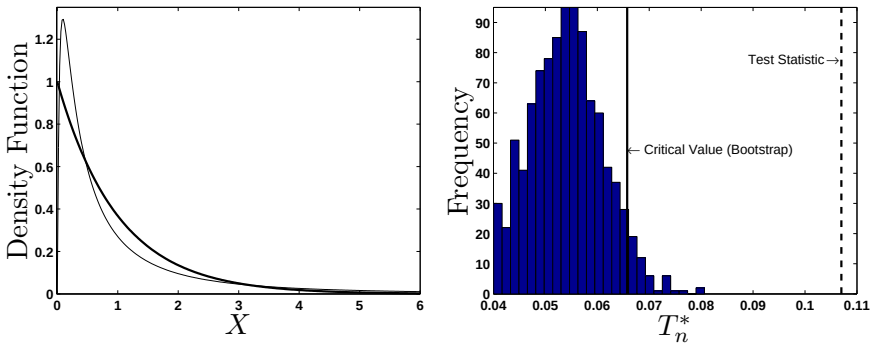


Figure 4: (a) lognormal (thin line) and exponential (bold line) density for the generated data set with parameters estimated by ML; (b) Null distribution of the test statistic. The dotted line indicates the value of T_n and the solid line indicates the critical value based on the bootstrap samples T_n^*.

5 Conclusion

We have developed a fully automated method, based on bootstrap with variance stabilization, to test which user defined parametric distribution best fits a given data set. The test statistic is based on the difference of the Kullback-Leibler divergences between a kernel density estimator and the pre-specified parametric distributions. In order to use LS-SVM to estimate the density, we transformed the density estimation problem into a nonparametric regression problem via a binning technique. The proposed method was successfully tested on two toy examples confirming the strength of the hypothesis test.

Acknowledgements

Research supported by: Research Council KUL: GOA AMBioRICS, CoE EF/05/006 Optimization in Engineering(OPTEC), IOF-SCORES4CHEM, several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects G.0452.04, G.0499.04, G.0211.05, G.0226.06 l, G.0321.06, G.0302.07, G.0320.08, G.0558.08, G.0557.08, research communities (ICCoS, ANMMM, MLDM); IWT: PhD Grants, McKnow-E, Eureka-Flite+; Helmholtz: viCERP. Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, 2007-2011); EU: ERNSI.

References

[1] F. J. Anscombe. The transformation of Poisson, Binomial and Negative-Binomial data. *Biometrika*, 35:246–254, 1948.

[2] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.

[3] P. Burman. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503–514, 1989.

[4] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, N.J., 1999.

[5] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press (Reprinted with corrections), 2003.

[6] B. Efron. Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7(1): 1–26, 1979.

[7] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, 1993.

[8] J. Fan. Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.*, 91:674–688, 1996.

[9] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, 1996.

[10] P. Hall and D. M. Titterington. The effect of simulation order on level accuracy and power of Monte Carlo tests. *J. Royal Statist. Soc. Series B*, 51:459–467, 1989.

[11] P. Hall and S. R. Wilson. Two guidelines for bootstrap hyphothesis testing. *Biometrics*, 47:757–762, 1991.

[12] C. Hu. Future CMOS scaling and reliability. *Proc. of the IEEE*, 81(5):682–689, 1993.

[13] N. Johnson, S. Kotz, and N. Balakrishnan. *Discrete Multivariate Distributions*. Wiley & Sons, 1997.

[14] A. Kerber, T. Pople, M. Röhner, K. Mosig, and M. Kerber. Impact of failure criteria on the reliability prediction of CMOS devices with ultrathin gate oxides based on voltage ramp stress. *IEEE Electron Device Letters*, 27(7):609–611, 2006.

[15] E. L. Lehmann and J .P. Romano. *Testing Statistical Hypothesis*. Springer, New York, 3rd edition, 2005.

[16] P. D. T. O'Connor. *Practical Reliability Engineering*. Wiley & Sons, 1985.

[17] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33(3):1065–1076, 1962.

[18] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.*, 27(3):832–837, 1956.

[19] S. Sahhaf, R. Degraeve, Ph. J. Roussel, B. Kaczer, T. Kauerauf, and G. Groeseneken. A new TDDB reliability prediction methodology accounting for multiple SBD and wear out. *IEEE Transactions on Electron Devices*, 56(7):1424–1432, July 2009.

[20] D. W. Scott. *Multivariate Density Estimation: Theory, Practice and Visualization*. Wiley & Sons, 1992.

[21] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapmann & Hall, 1st edition, (reprinted) 1992.

[22] J. A. K Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.

[23] R. J. Tibshirani. Variance stabilization and the bootstrap. *Biometrika*, 75(3):433–444, 1988.

[24] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

[25] M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapmann & Hall, 1995.

[26] L. Wasserman. *All of Nonparametric Statistics*. Springer, 2006.

[27] E. Y. Wu, J. Aitken, E. Nowak, A. Vayshenkera, P. Varekampa, G. Hueckel, J. McKenna, D. Harmon, L. K. Han, C. Montrose, and R. Dufresne. Voltage-dependent voltage-acceleration of oxide breakdown for ultra-thin oxides. *IEDM Techn. Dig.*, pages 541–544, 2000.

[28] E. Y. Wu, S. Tous, and J. Suñé. On the progressive breakdown statistical distribution and its voltage acceleration. *IEDM Techn. Dig.*, pages 493–496, 2007.