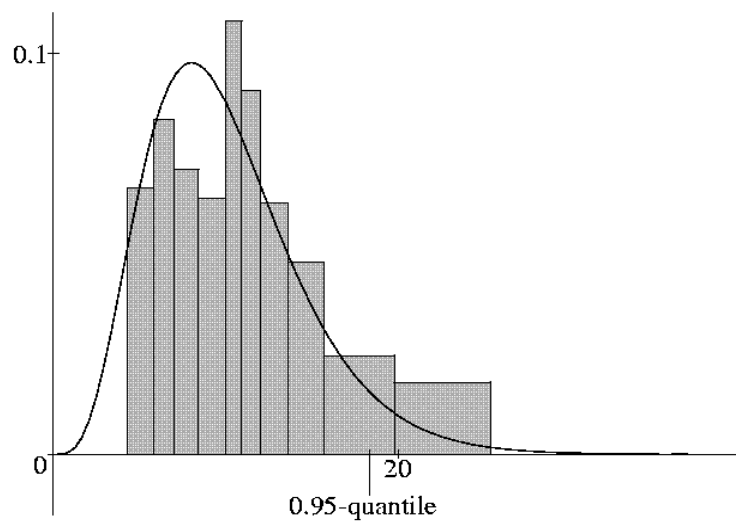




Goodness of Fit Tests of L_2 -Type

Hannelore Liero



Mathematische Statistik und
Wahrscheinlichkeitstheorie

Universität Potsdam – Institut für Mathematik

Mathematische Statistik und Wahrscheinlichkeitstheorie

Goodness of Fit Tests of L_2 -Type

Hannelore Liero

Department of Mathematics, University of Potsdam, 14469 Potsdam, Germany

e-mail: liero@rz.uni-potsdam.de

Preprint 2003/15

August 2003

Impressum

© Institut für Mathematik Potsdam, August 2003

Herausgeber: Mathematische Statistik und Wahrscheinlichkeitstheorie
am Institut für Mathematik

Adresse: Universität Potsdam
Am Neuen Palais 10
14469 Potsdam

Telefon:
Fax: +49-331-977 1500
E-mail: +49-331-977 1578
neisse@math.uni-potsdam.de

ISSN 1437-739X

Goodness of Fit Tests of L_2 -Type

Hannelore Liero

Institute of Mathematics, University of Potsdam

Abstract

We give a survey on procedures for testing functions which are based on quadratic deviation measures. The following problems are considered: Testing whether a density function lies in a parametric class of functions, whether continuous random variables are independent; testing cell probabilities and independence in sparse data sets; testing the parametric fit of a regression homoscedasticity in a regression model and testing the hazard rate in survival models with censoring and with and without covariates.

Keywords and phrases: Density tests, sparse data, testing independence, regression fit, homoscedasticity, hazard rate under censoring with covariates

AMS subject classification: Primary 62G10, 62G20

1 Introduction

We give a survey on procedures for testing functions in several nonparametric setups. The common idea of all considered tests is to express the deviation of the alternative from the hypothesis by a quadratic distance measure between a nonparametric kernel type estimator for the function of interest and a smoothed function characterizing the hypothetical function. Based on limit theorems stating that these quadratic functionals are asymptotically normally distributed we formulate asymptotic α -tests. Several aspects of the application of these test procedures are investigated.

So, in Section 2.1 after introducing a test statistic for checking whether a density function belongs to a parametric class we discuss the behavior of the power of the resulting test in detail. In Section 2.2 we apply similar ideas to test independence of two continuous random variables. Here the main point is to find good estimators for the standardizing terms in the limit theorem to avoid bias problems in the application of this limit statement for the formulation of the test.

In Section 3 the discrete analogues of the density test problems are considered. For estimating cell probabilities in sparse multinomial data sets Simonoff (1996) introduced local polynomial estimators. We use a special case of these kernel estimators to test hypothetical cell probabilities and compare our approach with the "classical" test procedure based on frequencies. Furthermore, the connection between testing in sparse data sets and testing a density is investigated. The case of testing independence in a sparse contingency table completes these considerations for sparse data.

In the following two sections we consider the nonparametric regression model. In Section 4 about testing whether a regression function has a parametric form we review the results of Härdle and Mammen (1993) to show, that bootstrap methods can be useful to apply tests of L_2 -type in practice.

Section 5 deals with testing homoscedasticity in a regression model. Here we show, how the conditional variance can be estimated nonparametrically. Further, we mention the problem of estimating the variance in a nonparametric homoscedastic regression model with random design.

In the last section tests for testing the hazard function in survival models for censored observations are given. Firstly the case without covariates is investigated; here the main point is to handle the maximum likelihood estimator for the unknown parameter in the hypothetical hazard function. In Section 6.2 the model with fixed covariates is studied. Here, following the approach of Van Keilegom and Veraverbeke (2001), we construct our test statistic on the basis of a weighted estimator for the hazard function, where the weights depend on the covariates.

2 Tests for densities

2.1 Testing whether a density has a parametric form

Let Z_1, \dots, Z_n be independent and identically distributed (i.i.d.) random variables with Lebesgue density f . We wish to test whether f lies in the parametric class

$$\mathcal{F} = \{ f_{\vartheta} = f(\cdot, \vartheta) : \vartheta \in \Theta \subseteq \mathbb{R}^d \}$$

against the alternative that f does not belong to \mathcal{F} , i.e.

$$\mathcal{H} : f \in \mathcal{F} \quad \text{against} \quad \mathcal{K} : f \notin \mathcal{F}.$$

The idea of the test procedure is to compare an estimator \hat{f}_n , which is "good for all possible densities f ", with the hypothetical one. It is well-known that the Rosenblatt-Parzen kernel estimator

$$\hat{f}_n(t) = \frac{1}{nb_n} \sum_{i=1}^n K\left(\frac{t - Z_i}{b_n}\right)$$

is such a good nonparametric estimator. Here K is the kernel function satisfying some regularity conditions and $\{b_n\}$ is a sequence of bandwidths tending to zero as n tends to infinity. As deviation measure we choose the weighted L_2 -distance. This approach was studied among others by Bickel and Rosenblatt (1973), Ghosh and Huang (1991), Liero, Läuter and Konakov (1998). Note, that the kernel estimator \hat{f}_n is not an unbiased estimator. Thus, deriving the limiting distribution of this distance one has to handle the bias. To avoid this problem it seems to be useful to take instead of the difference between \hat{f}_n and a hypothetical f_ϑ the difference of \hat{f}_n from its expectation under \mathcal{H} , that is from

$$\mathbb{E}_{\mathcal{H}} \hat{f}_n(t) = \frac{1}{b_n} \int K\left(\frac{t-x}{b_n}\right) f(x, \vartheta) dx = \int K(x) f(t - xb_n, \vartheta) dx.$$

In other words, we compare the smoothed data with a smoothed version of the hypothetical density. Since this expectation depends on the unknown parameter ϑ one has to replace it by a suitable estimator. Liero et al. (1998) propose to use the maximum likelihood estimator, say $\hat{\vartheta}_n$, which is \sqrt{n} -consistent under \mathcal{H} . Thus, finally we define the following test statistic:

$$\mathbb{Q}_n = \int \left(\hat{f}_n(t) - e_n(t, f_{\hat{\vartheta}_n}) \right)^2 a(t) dt$$

with $e_n(t, f_{\hat{\vartheta}_n}) = \int K(x) f(t - xb_n, \hat{\vartheta}_n) dx$ and a weight function a , which is introduced to control the region of integration and has to be chosen by the statistician. Before we formulate the basic limit statement let us introduce some notation, which are used also in the next sections: For $b > 0$ we write $K_b(t) = \frac{1}{b} K(t/b)$. Further, we define $\kappa^2 = \int K^2(x) dx$ and the convolution $\kappa^*(z) = \int K(u) K(z+u) du$. Throughout the paper we assume

- (K) The kernel K is a Lipschitz continuous density function with finite support.
- (W) The weight function a is nonnegative, piecewise continuous and bounded on \mathbb{R} ; (resp. on \mathbb{R}^2)

Theorem 2.1 *Suppose that (K), (W) and the following assumptions are satisfied: Any density $f \in \mathcal{F}$ is bounded on \mathbb{R} , Lipschitz continuous and partially differentiable w.r.t. ϑ ; $\nabla_{\vartheta} f(\cdot, \cdot)$, the vector of the partial derivatives, is bounded and uniformly continuous in both arguments. The estimator $\hat{\vartheta}_n$ is \sqrt{n} -consistent under \mathcal{H} . Further, $\int |\nabla_{\vartheta} f(t, \vartheta)| a(t) dt < \infty$ for each $\vartheta \in \Theta$ and $nb_n \rightarrow \infty$, $b_n \rightarrow 0$ and $b_n(\log n)^{\zeta} \rightarrow 0$ for some $\zeta > d/2$. Then under \mathcal{H}*

$$\frac{nb_n^{1/2}}{\sigma_{f_n}} (\mathbb{Q}_n - \mu_{f_n}) \xrightarrow{\mathcal{D}} \mathbf{N}(0, 1)$$

where

$$\mu_{f_n} = (nb_n)^{-1} \kappa^2 \int f(t, \hat{\vartheta}_n) a(t) dt \quad \text{and}$$

$$\sigma_{f_n}^2 = 2 \int f(t, \hat{\vartheta}_n)^2 a^2(t) dt \int (\kappa^*(z))^2 dz.$$

Applying this limit statement we obtain an asymptotic α -test of \mathcal{H} against \mathcal{K} by the rule: Reject \mathcal{H} if $\mathbb{Q}_n \geq \mu_{f_n} + z_\alpha \sigma_{f_n} / (nb_n^{1/2})$, where z_α is the $(1 - \alpha)$ -quantile of the standard normal distribution.

Some Remarks. 1. This test may be regarded as an analogue of a modified Cramér-von Mises test for testing whether an unknown distribution function lies in a parametric family of distribution functions. In contrast to the test for densities the limit distribution under the null hypothesis of the Cramér-von Mises test statistic with estimated parameter depends on the error of the parameter estimation. This is due to the fact that the normalizing factor n in the Cramér-von Mises test statistic is of the same order as the square of the rate of consistency of the parameter estimation, while in the density case this factor is $nb_n^{1/2}$ which tends to infinity slower. Therefore the error of the parameter estimation can be neglected in the problem presented here.

2. One can show (see Liero (1999)) that the limit statement formulated in Theorem 2.1 holds true if the bandwidth b_n is replaced by an adaptively chosen bandwidth \hat{b}_n as long as $\hat{b}_n/b_n \xrightarrow{P} c$ for an arbitrary but fixed deterministic bandwidth b_n , satisfying the conditions of Theorem 2.1 and some positive constant c .

3. Theorem 2.1 says nothing about the order of convergence of the distribution of the standardized test statistic to its limit. Simulations show that the approximation of the critical values by those of the standard normal distribution may fail for moderate sample size n . Therefore this limit theorem should be considered more as a theoretical result which gives an insight into the behavior of the test statistic, but it is not recommended for the approximate calculation of the critical values. (See also Section 4.)

Power considerations. It is easy to show, that the proposed L_2 -test is consistent, that is, if the alternative holds then the probability for rejecting \mathcal{H} tends to one. Therefore, for a characterization of the test and the comparison with other tests it is useful to investigate the asymptotic behavior of the power under local alternatives. In the literature there are different approaches to that problem. Here we will follow the “classical” approach and consider local alternatives of the form

$$\mathcal{K}_n : f_n(\cdot) = f(\cdot, \vartheta) + \Delta_n(\cdot)$$

where $\{\Delta_n\}$ is a sequence of functions tending to zero and ϑ is arbitrarily fixed. The aim is to study how the power depends on the convergence behavior of the disturbing function Δ_n . Such investigations were done under different aspects by Bickel and Rosenblatt (1973), Rosenblatt (1975), Ghosh and Huang (1991) and Liero et al. (1998). They considered the following types of alternatives: The so-called *Pitman alternatives*, *sharp peak alternatives* and alternatives with *rapidly oscillating disturbing functions*. To derive the behavior of the power $\Pi(\Delta_n) = \mathbb{P}_{\mathcal{K}_n} \left(\mathbb{Q}_n \geq \mu_{f_n} + z_\alpha \sigma_{f_n} (nb_n^{1/2})^{-1} \right)$, where $\mathbb{P}_{\mathcal{K}_n}$ is

the probability measure with respect to the local alternative, one has to study the asymptotic properties of the parameter estimator $\hat{\vartheta}_n$ under local alternatives. This is done in the paper of Liero et al.(1998), where an asymptotic expansion of the maximum likelihood estimator is given.

Generally speaking, the results in the paper mentioned above say that the L_2 -test is sensitive against local alternatives \mathcal{K}_n , where the weighted L_2 -norm of the disturbing function Δ_n behaves asymptotically as $n^{-1/2}b_n^{-1/4}$, in other words $\Pi(\Delta_n)$ tends to a number between α and 1, if $n^{1/2}b_n^{1/4}\|\Delta_n a^{1/2}\|_2 \rightarrow c \neq 0$.

In more detail one can prove the following results: 1. The error of the parameter estimation has an influence on the value of the limit of the power under Pitman alternatives and rapidly oscillating disturbing terms. Under sharp peak alternatives the value of the power does not depend on that estimation error.

2. Measured in the L_2 -norm all three types of alternatives tend to the hypothesis at the same rate of convergence.

3. The highly oscillating disturbing function can be interpreted as a function with a growing number of peaks. But, here more sharpness of the peaks is compensated by a larger number of peaks. Thus, the L_2 -norm of the disturbing function does not depend on the sharpness of the peaks, and the asymptotic behavior of the power under highly oscillating alternatives and under Pitman alternatives does not differ qualitatively.

4. If we translate our problem of testing a density function into a problem of testing distribution functions we get the following results: Pitman alternatives remain Pitman type alternatives also in the context of distribution functions. Therefore, our L_2 -density test is worse than the Cramér-von Mises test, if we compare both with respect to this type of alternatives. The sharp peak disturbing function yields for distribution functions a disturbing function of sharp peak type, but with other "less sharp peaks". That means that there exist alternatives of sharp peak type which are detected by the test based on density estimators, but not by the classical Cramér - von Mises test. The reason is that integration of the alternative density smoothes the sharp peak away. Integration of the rapidly oscillating disturbing function leads to the following result: Although the behavior of the power of the L_2 -test under Pitman and highly oscillating alternatives is qualitatively the same, we can find highly oscillating disturbing functions where the Cramér-von Mises test fails, but the L_2 - density test does not. The explanation is, that also these 'infinitely many peaks' are smoothed away by the translation from density to distribution function, despite of their growing number.

5. The investigations show that a larger bandwidth improves the power. Heuristically speaking, this means, that the rate of convergence of the alternative, measured in the L_2 -norm, may increase if the variance of the kernel estimator tends to zero faster. This feature, incidentally, conflicts with the fact discussed before stating that the approximation of the distribution of the test statistic by the standard normal distribution improves if the bandwidth tends to zero faster.

6. A test based on the integrated difference of \hat{f}_n from the hypothetical $f_{\hat{\vartheta}_n}$ is discussed by Liero (1999). Here the additional bias term can lead to an increase or an decrease of the power.

7. In the paper of Liero et al. (1998) a so-called L_∞ -test, which is based on the (normalized) maximal deviation of \hat{f}_n from $E_{\mathcal{H}}\hat{f}_n(t)$ is studied. The power considerations carried out there show that with respect to Pitman alternatives the L_2 -test behaves better than the L_∞ -test. Further, it is proved that there exist local alternatives of sharp peak type

for which the L_∞ -test distinguishes between hypothesis and alternative, but the L_2 -test does not.

2.2 Testing independence

Let $(U_1, V_1), \dots, (U_n, V_n)$ be a sample of i.i.d. $(\mathbb{R} \times \mathbb{R})$ -valued random variables with density f . We wish to test whether U_i and V_i are independent, that is the test problem has the form

$$\mathcal{H} : f = g \cdot h \quad \text{against} \quad \mathcal{K} : f \neq g \cdot h,$$

where g and h are the marginal densities of U_i and V_i , respectively. Again we will use a kernel estimator for the construction of our test statistic. It is defined by

$$\hat{f}_n(s, t) = \frac{1}{nb_n^2} \sum_{i=1}^n K\left(\frac{U_i - s}{b_n}, \frac{V_i - t}{b_n}\right).$$

Here $K : \mathbb{R}^2 \rightarrow \mathbb{R}$ is the kernel function and b_n is the bandwidth sequence. We take a kernel of product form, i.e.

$$K(x, y) = K_1(x) K_2(y) \quad \text{with} \quad \int K_j(x) dx = 1 \quad \text{for} \quad j = 1, 2. \quad (1)$$

The estimators of the marginal densities are given by

$$\hat{g}_n(s) = \frac{1}{nb_n} \sum_{i=1}^n K_1\left(\frac{s - U_i}{b_n}\right) \quad \text{and} \quad \hat{h}_n(t) = \frac{1}{nb_n} \sum_{i=1}^n K_2\left(\frac{t - V_i}{b_n}\right).$$

The formulation of the test procedure goes back to Rosenblatt (1975). His idea was to compare a kernel estimator of f with the estimator of f under the hypothesis, that is with the product of kernel estimators of the marginal densities. This leads to the test statistic

$$\mathbb{I}_n = \int (\hat{f}_n(s, t) - \hat{g}_n(s) \cdot \hat{h}_n(t))^2 a(s, t) ds dt,$$

where a is again a suitable weight function and the integration is taken over \mathbb{R}^2 . Let us denote the expectations of \hat{f}_n , \hat{g}_n , and \hat{h}_n , by \bar{f}_n , \bar{g}_n and \bar{h}_n , respectively. Further, define

$$\xi_{f_n} = (nb_n^2)^{-1} D_{1n} - (nb_n)^{-1} D_{2n}$$

with

$$\begin{aligned} D_{1n} &= \int \int K_1^2(u) g(s - ub_n) du \int K_2^2(v) h(t - vb_n) dv a(s, t) ds dt \\ D_{2n} &= \int \left(\bar{g}_n^2(s) \int K_2^2(v) h(t - vb_n) dv \right. \\ &\quad \left. + \bar{h}_n^2(t) \int K_1^2(u) g(s - ub_n) du \right) a(s, t) ds dt, \end{aligned}$$

and

$$\tau_f^2 = 2 \int g^2(s) h^2(t) a^2(s, t) ds dt \int (K_1 * K_1)^2(r) dr \int (K_2 * K_2)^2(v) dv.$$

With these notations we can formulate:

Theorem 2.2 *Suppose that the marginal densities g and h are Lipschitz continuous and bounded. Let the kernel K be of product type (1). The K_j 's ($j = 1, 2$) satisfy condition (K) and the weight function a condition (W). If $f = g \cdot h$, then*

$$nb_n (\mathbb{I}_n - \xi_{fn}) \xrightarrow{\mathcal{D}} \mathbf{N}(0, \tau_f^2) \quad (2)$$

as $b_n \rightarrow 0$ and $nb_n^2 \rightarrow \infty$.

To apply limit statement (2) for the construction of the test procedure we have to replace the unknown terms D_{1n} and D_{2n} by estimators which are consistent with a certain rate of convergence. To avoid bias problems we do not follow the proposal of Rosenblatt (1975), who replaced the unknown functions g and h by the kernel estimators \hat{g}_n and \hat{h}_n . Observe, that

$$D_{1n} = b_n^{-2} \int \Omega_{1n}(s) \Omega_{2n}(t) a(s, t) ds dt \quad \text{and}$$

$$D_{2n} = b_n^{-1} \int \left([\bar{g}_n(s)]^2 \Omega_{2n}(t) + [\bar{h}_n(t)]^2 \Omega_{1n}(s) \right) a(s, t) ds dt$$

where $\Omega_{1n}(s) = \mathbf{E}K_1^2 \left(\frac{s-U_1}{b_n} \right)$ and $\Omega_{2n}(t) = \mathbf{E}K_2^2 \left(\frac{t-V_1}{b_n} \right)$.

We estimate these quantities by

$$\tilde{\Omega}_{1n}(s) = \frac{1}{n} \sum_{i=1}^n K_1^2 \left(\frac{s-U_i}{b_n} \right) \quad \text{and} \quad \tilde{\Omega}_{2n}(t) = \frac{1}{n} \sum_{i=1}^n K_2^2 \left(\frac{t-V_i}{b_n} \right).$$

and obtain as estimator for D_{1n}

$$\tilde{D}_{1n} = b_n^{-2} \int_A \tilde{\Omega}_{1n}(s) \tilde{\Omega}_{2n}(t) a(s, t) ds dt,$$

which is consistent and unbiased under \mathcal{H} . Estimators of D_{2n} and τ_f^2 are given by

$$\tilde{D}_{2n} = b_n^{-1} \int_A \left([\hat{g}_n(s)]^2 \tilde{\Omega}_{2n}(t) + [\hat{h}_n(t)]^2 \tilde{\Omega}_{1n}(s) \right) a(s, t) ds dt \quad \text{and}$$

$$\tilde{\tau}_{fn}^2 = 2 \int \hat{g}_n^2(s) \hat{h}_n^2(t) a^2(s, t) ds dt \int (K_1 * K_1)^2(u) du \int (K_2 * K_2)^2(v) dv.$$

Set $\tilde{\xi}_{fn} = (nb_n^2)^{-1} \tilde{D}_{1n} - (nb_n)^{-1} \tilde{D}_{2n}$. It is easy to verify that the limit statement formulated in Theorem 2.2 remains valid if the unknown terms ξ_{fn} and τ_{fn} are replaced by these estimators. Thus, an asymptotic α -test is provided by: Reject \mathcal{H} , if $\mathbb{I}_n \geq \tilde{\xi}_{fn} + z_\alpha \tilde{\tau}_{fn} / (nb_n)$.

Some remarks. 1.) In difference to the approach of Rosenblatt (1975) we propose another estimator of the standardizing terms in the limit theorem for \mathbb{I}_n . The advantage of our method is, that this estimator is unbiased. So we do not need additional assumptions on the smoothness of the underlying densities to ensure that the limit theorem remains valid with the estimated standardizing terms.

2.) The behavior of the power of this test is qualitatively very similar to that of the L_2 -test considered in Section 2.1. As there one can show that the power tends to a nontrivial limit, i.e. a number between α and 1, if $\sqrt{nb_n} \|\Delta_n a^{1/2}\|_2 \rightarrow c \neq 0$, where the disturbing function Δ_n describes the deviation from independence.

3 Tests for sparse data sets

In the classical case the number of cells, say k , in a multinomial distribution or a contingency table is assumed to be fixed. But there are data sets where the total number of observations is moderate in comparison to the total number of cells. Consequently the number of observations falling in each cell is rather small. We describe such sparseness mathematically by assuming $k = k_n \rightarrow \infty$ as n tends to infinity. It is known that for sparse data nonparametric smoothing techniques provide estimators of the cell probabilities, which have a better asymptotic performance than the frequency estimators, see for example Aerts et al. (1997) and Simonoff (1996). Here we use such smoothed estimators to define a test statistic of L_2 -type.

3.1 Testing cell probabilities in sparse multinomial data

Let $p_n = (p_{n1}, \dots, p_{nk_n})^t$ be the vector of cell probabilities of a k_n -cell multinomial distribution, where n is the total sample size. The simplest test problem is to test

$$\mathcal{H}: p_{ni} = \pi_{ni} \text{ for all } i = 1, \dots, k_n \text{ against } \mathcal{K}: p_{ni'} \neq \pi_{ni'} \text{ for some } i', \quad (3)$$

where $\pi_n = (\pi_{n1}, \dots, \pi_{nk_n})^t$ is a vector of given cell probabilities. To formulate the test procedure we start with the definition of the estimators of the cell probabilities p_{ni} , $i = 1, \dots, k_n$. As smoothed estimators we propose local constant estimators, which are the simplest local polynomial estimators introduced by Simonoff (1996). For the definition of these estimators let $x_{nj} = (j - \frac{1}{2})/k_n$ be equidistant design points on the interval $[0, 1]$ and denote the relative frequency of cell i by p_{ni}^* . The data (x_{nj}, p_{nj}^*) can be considered as regression type data. Following the idea of smoothing in the regression set-up we estimate the cell probability p_{ni} by

$$\hat{p}_{ni} = \frac{\sum_{\mu=1}^{k_n} K\left(\frac{x_{n\mu} - x_{ni}}{b_n}\right) p_{n\mu}^*}{\sum_{\mu=1}^{k_n} K\left(\frac{x_{n\mu} - x_{ni}}{b_n}\right)},$$

where K is a kernel function and b_n is a sequence of bandwidths introduced already in the previous section. For simplicity of writing we skip the subscript n in the notation of the cell probabilities and the design points; furthermore we write

$$\hat{p}_i = \frac{1}{k_n b_n} \sum_{\mu=1}^{k_n} L_i\left(\frac{x_{\mu} - x_i}{b_n}\right) p_{\mu}^* \quad \text{with} \quad L_i(u) = \frac{K(u)}{\frac{1}{k_n b_n} \sum_{\mu=1}^{k_n} K\left(\frac{x_{\mu} - x_i}{b_n}\right)}.$$

As test statistic we propose the sum of squared differences between the estimators and their expectations under the hypothesis:

$$\mathbb{T}_n = \sum_{i=1}^{k_n} (\hat{p}_i - \mathbb{E}_{\mathcal{H}} \hat{p}_i)^2 \quad \text{with} \quad \mathbb{E}_{\mathcal{H}} \hat{p}_i = \frac{1}{b_n k_n} \sum_{j=1}^{k_n} L_i\left(\frac{x_j - x_i}{b_n}\right) \pi_j.$$

Set

$$\mu_{\pi n} = \frac{1}{n k_n^2} \sum_{i=1}^{k_n} \left[\sum_{j=1}^{k_n} L_{i,b_n}^2(x_j - x_i) \pi_j - \left(\sum_{j=1}^{k_n} L_{i,b_n}(x_j - x_i) \pi_j \right)^2 \right]$$

where $L_{i,b}(u) = \frac{1}{b}L_i(\frac{u}{b})$, and

$$\sigma_{\pi n}^2 = 2k_n^2 b_n \sum_{l=1}^{k_n} \sum_{m=1}^{k_n} C_{lm}^2 \pi_l \pi_m \quad \text{with} \quad C_{lm} = k_n^{-2} \sum_{i=1}^{k_n} L_{i,b_n}(x_l - x_i) L_{i,b_n}(x_m - x_i).$$

The following limit theorem shows that \mathbb{T}_n is asymptotically normal under \mathcal{H} .

Theorem 3.1 *Assume (K) and $|\pi_{ni} - \pi_{nj}| \leq L k_n^{-1} |x_{ni} - x_{nj}|$ for all i, j and some constant L . If $b_n \rightarrow 0$, $b_n k_n \rightarrow \infty$ and $nb_n \rightarrow \infty$ as $n \rightarrow \infty$, then under \mathcal{H}*

$$\frac{nk_n \sqrt{b_n}}{\sigma_{\pi n}} (\mathbb{T}_n - \mu_{\pi n}) \xrightarrow{\mathcal{D}} \mathbf{N}(0, 1).$$

Under the assumptions of Theorem 3.1 we get the test: Reject the hypothesis \mathcal{H} if $\mathbb{T}_n \geq \mu_{\pi n} + z_\alpha \sigma_{\pi n} / (nk_n \sqrt{b_n})$.

Comparison to the "classical" approach. Another possibility to test (3) is the quadratic deviation of the unsmoothed estimators of the cell probabilities, that is to use the test statistic

$$\mathbb{S}_n = \sum_{i=1}^{k_n} (p_i^* - \pi_i)^2.$$

Applying results proved by Holst (1972) and Burman (1987) one can show that this statistic, properly standardized, is also asymptotically normally distributed. The asymptotic α -test based on this limit statement has the following form: Reject the hypothesis \mathcal{H} , if $\mathbb{S}_n \geq \mathbf{E}_{\mathcal{H}} \mathbb{S}_n + \rho_{\pi n} z_\alpha / (n\sqrt{k_n})$, where $\mathbf{E}_{\mathcal{H}} \mathbb{S}_n = \frac{1}{n} \left(1 - \sum_{i=1}^{k_n} \pi_i^2\right)$, and $\rho_{\pi n}^2 = 2k_n \sum_{j=1}^{k_n} \pi_j^2$ is a sequence of positive numbers tending to a positive constant.

To compare both test procedures we consider the behavior of the power under local alternatives of the form: $\mathcal{K}_n : p_i := \pi_i + \delta_i$ with $\sum_{i=1}^k \delta_i = 0$.

Let us denote the power of the test based on the (unsmoothed) frequency estimators by β_{1n} , and that of the test based on the \hat{p}_{ni} 's by β_{2n} . Then under mild conditions on the disturbing terms δ_i in Liero (2001) it is proved that

$$\lim_{n \rightarrow \infty} \frac{\beta_{1n}}{\beta_{2n}} \leq 1,$$

where "=" holds if $nk_n \sqrt{b_n} \sum_{i=1}^{k_n} \delta_i^2 \rightarrow 0$ or $n\sqrt{k_n} \sum_{i=1}^{k_n} \delta_i^2 \rightarrow \infty$.

That means, roughly speaking, the test based on the quadratic deviation of the local polynomial estimator is better than the test based on the frequency estimators. Only in the case that the square of the L_2 -norm of the disturbing terms, i. e. $\sum_i \delta_i^2$, is very large, the power of both tests tends to one; and in the case that the bandwidth b_n is very small, i. e. we smooth only "a little bit", both tests behave poorly.

Connection to the goodness-of-fit test using densities. Suppose that the cell probabilities p_i and π_i are generated by latent densities f and f_0 , respectively, which are defined on $[0, 1]$:

$$p_i = \int_{I_i} f(x) dx, \quad \pi_i = \int_{I_i} f_0(x) dx \quad \text{where} \quad I_i = [(i-1)/k_n, i/k_n].$$

Then the test problem (3) corresponds to the simple problem

$$\mathcal{H} : f = f_0 \quad \text{against} \quad \mathcal{K} : f \neq f_0. \quad (4)$$

Furthermore it follows from results proved by Augustyns (1997) that

$$\mu_{\pi n} = (nk_n b_n)^{-1}(\kappa^2 + o(1)) \quad \text{and} \quad \lim_{n \rightarrow \infty} \sigma_{\pi n}^2 = 2 \int f_0^2(x) dx \int (\kappa^*(z))^2 dz.$$

Thus we have the following correspondence between the test for the cell probabilities and the modified test for testing (4) (where the weight function a is the indicator of $[0, 1]$):

$$\begin{aligned} k_n \mathbb{T}_n = k_n \sum_{i=1}^{k_n} (\hat{p}_i - \mathbb{E}_{\mathcal{H}} \hat{p}_i)^2 & \text{ corresponds to } \mathbb{Q}_n = \int (\hat{f}_n(t) - \mathbb{E}_{\mathcal{H}} \hat{f}_n(t))^2 dt, \\ k_n \mu_{\pi n} & \text{ to } \mu_{f_n} = (nb_n)^{-1} \kappa^2 \int_0^1 f_0(t) dt \quad \text{and the variance term } \sigma_{\pi n}^2 \text{ to } \sigma_{f_n}^2 = \\ & 2 \int_0^1 f_0^2(t) dt \int (\kappa^*(z))^2 dz. \end{aligned}$$

Moreover, let us consider the behavior of the power from the viewpoint of the existence of a latent density. For that purpose we write the local alternative in the form $\mathcal{K}'_n : f_n := f_0 + \Delta_n$. Suppose that $\sum_{i=1}^{k_n} \delta_i^2 \sim \frac{1}{k_n} \int_0^1 \Delta_n^2(u) du$. Then, expressed in terms of densities, the power of the second test tends to a nontrivial limit, if the square of the L_2 -norm of the disturbing function Δ_n is asymptotically equivalent to $(n\sqrt{b_n})^{-1}$. Note, that this is the same rate of convergence as in the problem of testing a density obtained before.

3.2 Testing independence in sparse contingency tables

We consider a two-dimensional contingency table with $k_n = l_n \cdot m_n$ cells, where $l_n \rightarrow \infty$ and $m_n \rightarrow \infty$. The (joint) cell probabilities are denoted by p_{nij} , the marginal cell probabilities by q_{ni} and r_{nj} . To test independence we have to check the hypothesis

$$\mathcal{H} : p_{nij} = q_{ni} r_{nj} \quad \text{for all } (i, j) \quad \text{against} \quad \mathcal{K} : p_{ni'j'} \neq q_{ni'} r_{nj'} \quad \text{for some } (i', j'). \quad (5)$$

For testing (5) we will use

$$\mathbb{M}_n = \sum_{i=1}^{m_n} \sum_{j=1}^{l_n} \left(\hat{p}_{nij} - \hat{q}_{ni} \cdot \hat{r}_{nj} \right)^2,$$

where \hat{p}_{nij} , \hat{q}_{ni} and \hat{r}_{nj} are the local constant estimators of p_{nij} , q_{ni} and r_{nj} . Following the ideas presented in the previous sections these estimators have the following form (we skip the n in the subscript if appropriate):

$$\hat{p}_{ij} = \frac{\sum_{\mu=1}^{m_n} \sum_{\nu=1}^{l_n} K \left(\frac{x_{\mu} - x_i}{b_n}, \frac{y_{\nu} - y_j}{b_n} \right) p_{\mu\nu}^*}{\sum_{\mu=1}^{m_n} \sum_{\nu=1}^{l_n} K \left(\frac{x_{\mu} - x_i}{b_n}, \frac{y_{\nu} - y_j}{b_n} \right)} = \frac{1}{m_n l_n b_n^2} \sum_{\mu=1}^{m_n} \sum_{\nu=1}^{l_n} L_{ij} \left(\frac{x_{\mu} - x_i}{b_n}, \frac{y_{\nu} - y_j}{b_n} \right) p_{\mu\nu}^*,$$

where $x_\mu = (\mu - \frac{1}{2})/m_n$ and $y_\nu = (\nu - \frac{1}{2})/l_n$ are equidistant design points on the interval $[0, 1] \times [0, 1]$, p_{ij}^* is the relative frequency of cell (i, j) and

$$L_{ij}(u, v) = \frac{K(u, v)}{\frac{1}{m_n l_n b_n^2} \sum_{\mu=1}^{m_n} \sum_{\nu=1}^{l_n} K\left(\frac{x_\mu - x_i}{b_n}, \frac{y_\nu - y_j}{b_n}\right)}.$$

Again we take a kernel of product type. Then as estimators of the marginal cell probabilities $q_i = \sum_j p_{ij}$ and $r_j = \sum_i p_{ij}$ we obtain straightforward

$$\hat{q}_i = \frac{1}{k_n b_n} \sum_{\mu=1}^{m_n} L_{1i}\left(\frac{x_\mu - x_i}{b_n}\right) q_\mu^* \quad \text{and} \quad \hat{r}_j = \frac{1}{l_n b_n} \sum_{\nu=1}^{l_n} L_{2j}\left(\frac{y_\nu - y_j}{b_n}\right) r_\nu^*$$

with weight functions

$$L_{1i}(u) = \frac{K_1(u)}{\frac{1}{m_n b_n} \sum_{\mu=1}^{m_n} K_1\left(\frac{x_\mu - x_i}{b_n}\right)} \quad \text{and} \quad L_{2j}(v) = \frac{K_2(v)}{\frac{1}{l_n b_n} \sum_{\nu=1}^{l_n} K_2\left(\frac{y_\nu - y_j}{b_n}\right)},$$

and marginal frequencies q_μ^* and r_ν^* .

To formulate the asymptotic normality of \mathbb{M}_n under \mathcal{H} we make use of the following notation: \bar{p}_{ij}, \bar{q}_i , and \bar{r}_j , are the expectations of \hat{p}_{ij}, \hat{q}_i , and \hat{r}_j , respectively. Define

$$\xi_{pn} = (nm_n l_n b_n^2)^{-1} d_{1n} - (nm_n l_n b_n)^{-1} d_{2n}$$

with

$$\begin{aligned} d_{1n} &= (m_n l_n)^{-1} b_n^2 \sum_{i=1}^{m_n} \sum_{\mu=1}^{m_n} K_{1i, b_n}^2(x_\mu - x_i) q_\mu \sum_{j=1}^{l_n} \sum_{\nu=1}^{l_n} K_{2j, b_n}^2(y_\nu - y_j) r_\nu \\ d_{2n} &= b_n \left(m_n^{-1} l_n \sum_{j=1}^{l_n} \bar{r}_j^2 \sum_{i=1}^{m_n} \sum_{\mu=1}^{m_n} K_{1i, b_n}^2(x_\mu - x_i) q_\mu \right. \\ &\quad \left. + l_n^{-1} m_n \sum_{i=1}^{m_n} \bar{q}_i^2 \sum_{j=1}^{l_n} \sum_{\nu=1}^{l_n} K_{2j, b_n}^2(y_\nu - y_j) r_\nu \right) \end{aligned}$$

and

$$\tau_{pn}^2 = 2 m_n l_n \sum_{i=1}^{m_n} \sum_{j=1}^{l_n} \bar{q}_i^2 \bar{r}_j^2 \iint (K * K)^2(x, y) dx dy.$$

The basis of our test is the following theorem:

Theorem 3.2 *Suppose that the marginal probabilities satisfy for all i, μ, j and ν , and some constants L_1 and L_2*

$$|q_i - q_\mu| \leq L_1 m_n^{-1} |x_i - x_\mu| \quad \text{and} \quad |r_j - r_\nu| \leq L_2 l_n^{-1} |y_j - y_\nu|.$$

Further, the kernel K is of product type (1), and the K_j 's ($j = 1, 2$) satisfy (K). If the hypothesis \mathcal{H} holds, then

$$\frac{nm_n l_n b_n}{\tau_{pn}} (\mathbb{M}_n - \xi_{pn}) \xrightarrow{\mathcal{D}} \mathbf{N}(0, 1)$$

as $b_n \rightarrow 0$, $m_n l_n b_n^2 \rightarrow \infty$ and $nb_n^2 \rightarrow \infty$.

To apply this limit statement for the construction of an asymptotic α -test we have to replace the unknown terms d_{1n} and d_{2n} in ξ_{pn} by suitable estimators. To avoid bias problems we choose

$$\tilde{d}_{1n} = (m_n l_n)^{-1} b_n^2 \sum_{i=1}^{m_n} \sum_{\mu=1}^{m_n} K_{1i, b_n}^2(x_\mu - x_i) q_\mu^* \sum_{j=1}^{l_n} \sum_{\nu=1}^{l_n} K_{2j, b_n}^2(y_\nu - y_j) r_\nu^*,$$

which is an unbiased \sqrt{n} -consistent estimator of d_{1n} . The second term is estimated by the consistent estimator

$$\begin{aligned} \tilde{d}_{2n} = b_n \left(m_n^{-1} l_n \sum_{j=1}^{l_n} r_j^{*2} \sum_{i=1}^{m_n} \sum_{\mu=1}^{m_n} K_{1i, b_n}^2(x_\mu - x_i) q_\mu^* \right. \\ \left. + l_n^{-1} m_n \sum_{i=1}^{m_n} q_i^{*2} \sum_{j=1}^{l_n} \sum_{\nu=1}^{l_n} K_{2j, b_n}^2(y_\nu - y_j) r_\nu^* \right), \end{aligned}$$

and the variance term can be replaced by

$$\tilde{\tau}_{pn}^2 = 2m_n l_n \sum_{i=1}^{m_n} \sum_{j=1}^{l_n} q_i^{*2} r_j^{*2} \int \int (K * K)^2(x, y) dx dy.$$

It is easy to verify that under \mathcal{H}

$$b_n^{-1} (\tilde{d}_{1n} - d_{1n}) \xrightarrow{P} 0, \quad \tilde{d}_{2n} - d_{2n} \xrightarrow{P} 0 \quad \text{and} \quad \tilde{\tau}_{pn}^2 - \tau_{pn}^2 \xrightarrow{P} 0.$$

Thus, with $\tilde{\xi}_{pn} = (nm_n l_n b_n^2)^{-1} \tilde{d}_{1n} - (nm_n l_n b_n)^{-1} \tilde{d}_{2n}$ an asymptotic α -test is provided by: Reject \mathcal{H} , if $\mathbb{M}_n \geq \tilde{\xi}_{pn} + z_\alpha \tilde{\tau}_{pn} / (nm_n l_n b_n)$.

4 Parametric versus nonparametric regression fit

Härdle and Mammen (1993) investigated the problem of testing whether a regression function has a parametric form. Let us shortly review their results. We have the following model: The pairs (X_i, Y_i) , $i = 1, \dots, n$, are i.i.d. $(\mathbb{R} \times \mathbb{R})$ -valued random variables satisfying

$$Y_i = r(X_i) + \sqrt{v(X_i)} \varepsilon_i, \quad i = 1, \dots, n, \quad (6)$$

with the unknown regression function $r(\cdot) = \mathbf{E}(Y_1 | X_1 = \cdot)$ and the conditional variance $v(\cdot) = \mathbf{E}((Y_1 - r(X_1))^2 | X_1 = \cdot)$. Conditionally on X_1, \dots, X_n the errors ε_i are independent and identically distributed with expectation zero and variance one. The test problem is:

$$\mathcal{H} : r \in \{r(\cdot, \vartheta) : \vartheta \in \Theta \subseteq \mathbb{R}^d\} \quad \text{against} \quad \mathcal{K} : r \notin \{r(\cdot, \vartheta) : \vartheta \in \Theta \subseteq \mathbb{R}^d\},$$

and as test statistic Härdle and Mammen propose the L_2 -distance

$$\mathbb{R}_n(r_{\hat{\vartheta}_n}) = \int (\hat{r}_n(t) - \mathcal{K}_n r_{\hat{\vartheta}_n}(t))^2 a(t) dt.$$

Here \hat{r}_n is the Nadaraya-Watson kernel estimator with bandwidth b_n , and $\mathcal{K}_n r_{\hat{\vartheta}_n}$ denotes its smoothed version under \mathcal{H} , $\hat{\vartheta}_n$ is a suitable parameter estimator. It is proved that the properly standardized test statistic is asymptotically normal. The main point of the paper is to investigate different *bootstrap procedures* for the approximation of the critical values of the test. As already pointed out, the convergence of the distribution of an ISE-type statistic to the normal distribution is very slow (see Remark 3 in Section 2), therefore quantiles of the normal distribution are not appropriate for testing in practice. Since this problem arises not only in the context of regression testing it seems to be useful to think about, whether it is possible to apply similar bootstrap approaches also in other setups.

Suppose that (X_i^*, Y_i^*) , $i = 1, \dots, n$, is a bootstrap sample, then create $\mathbb{R}_n^*(r_{\hat{\vartheta}_n^*})$ like $\mathbb{R}_n(r_{\hat{\vartheta}_n})$ by the squared deviation between the parametric fit $r_{\hat{\vartheta}_n^*}$ and the nonparametric fit \hat{r}_n^* (both computed from the bootstrap sample). The conditional distribution of $\mathbb{R}_n^*(r_{\hat{\vartheta}_n^*})$ under the (X_i, Y_i) can be approximated by Monte Carlo simulations. From this Monte Carlo approximation $(1 - \alpha)$ quantile \hat{q}_α is defined, and one rejects \mathcal{H} if $nb_n^{1/2}\mathbb{R}_n(r_{\hat{\vartheta}_n}) > \hat{q}_\alpha$. Härdle and Mammen show that the *naive resampling* does not work. The same is true for the so-called *adjusted residual bootstrap*. As an alternative they propose the *wild bootstrap*. The idea is to construct a bootstrap sample (X_i^*, Y_i^*) , $i = 1, \dots, n$, such that $\mathbf{E}^*(Y_i^*|X_i^*) = r_{\hat{\vartheta}_n}(X_i^*)$, where \mathbf{E}^* denotes the conditional expectation $\mathbf{E}(\cdot|(X_i, Y_i), i = 1, \dots, n)$. In simulation studies Härdle and Mammen consider parametric models of polynomials of different degree. It turns out that in all cases wild bootstrap estimates the distribution of $nb_n^{1/2}\mathbb{R}_n(r_{\hat{\vartheta}_n})$ quite well. The normal approximation with estimated standardizing terms is totally misleading. The inaccuracy of the normal approximation increases with the dimension of the parametric model. Moreover, the authors give Monte Carlo estimates for the power of the test with bootstrapped quantiles and consider the influence of the bandwidth on the level of the test.

5 Testing homoscedasticity in nonparametric regression

Again, assume model (6). Now, we wish to check whether the model is heteroscedastic, that is, we wish to test the hypothesis

$$\begin{aligned} \mathcal{H} : v(t) &= v && \text{for some } v > 0 \text{ and all } t \in [0, 1] \quad \text{against} \\ \mathcal{K} : v(t) &\neq v && \text{for all } v > 0. \end{aligned}$$

In the paper of Liero (2003a) the following approach is proposed: As test statistic we take the L_2 -distance between a nonparametric kernel estimator of v in the underlying heteroscedastic model (6) and an estimator of the conditional variance in the hypothetical homoscedastic model

$$Y_i = r(X_i) + \sqrt{v} \varepsilon_i, \quad i = 1, \dots, n, \quad (7)$$

with $v \in \mathbb{R}_+$. To avoid bias problems in the limit theorem we modify this difference and use the following statistic:

$$\mathbb{V}_n = \int (\hat{v}_n(t) - \eta_{nv}(t))^2 a(t) dt$$

The estimator \hat{v}_n has the form

$$\hat{v}_n(t) = \sum_{i=1}^n \sum_{j=1, i \neq j}^n \frac{1}{2} (Y_i - Y_j)^2 W_{nij}(t, X_1, \dots, X_n)$$

with weights

$$W_{nij}(t, X_1, \dots, X_n) = \frac{K\left(\frac{t-X_i}{b_n}\right) K\left(\frac{t-X_j}{b_n}\right)}{\sum_{\mu=1}^n \sum_{\nu=1, \nu \neq \mu}^n K\left(\frac{t-X_\mu}{b_n}\right) K\left(\frac{t-X_\nu}{b_n}\right)},$$

and the term η_{nv} is defined by

$$\eta_{nv}(t) = v + \sum_{i=1}^n \sum_{j=1, i \neq j}^n \frac{1}{2} (r(X_i) - r(X_j))^2 W_{nij}(t, X_1, \dots, X_n).$$

Liero (1999) showed that \mathbb{V}_n (properly standardized) is asymptotically normally distributed. But to apply this limit result for the construction of a test it is necessary to replace the unknown terms r and $v > 0$ by suitable estimators. This leads to the problem of estimating the variance in a homoscedastic regression model (7). For the fixed design model this problem is investigated by several authors. For the present random design model three different estimators for v are given by Liero (1999). It turns out that these estimators are \sqrt{n} -consistent under the hypothesis, which is sufficient to use them for the construction of the desired asymptotic α -test. Furthermore, it seems to be useful to replace the unknown regression function r in the term η_{nv} by a Nadaraya-Watson kernel estimator with a suitable bandwidth. The question of an appropriate choice of this bandwidth is discussed in Liero (2003a). Moreover, power considerations with respect to different types of local alternatives complete the approach presented there.

6 Testing the hazard function under censoring

6.1 Survival model without covariates

Firstly we consider a survival model without covariates, that is: Let Y_1, \dots, Y_n be a sequence of i.i.d. survival times with absolutely continuous distribution function F . As often occurs in applications the Y_i 's are subject to random right censoring, i.e. the observations are

$$T_i = \min(Y_i, C_i) \quad \text{and} \quad \delta_i = 1(Y_i \leq C_i)$$

where C_1, \dots, C_n are i.i.d. continuous random censoring times which are independent of the Y -sequence. The δ_i indicates whether Y_i has been censored or not. The function of interest is the hazard rate λ which is defined by

$$\lambda(t) = \lim_{s \downarrow 0} \frac{1}{s} \mathbb{P}(t \leq Y_i \leq t + s | Y_i \geq t).$$

We wish to test whether λ lies in a parametric class of functions, i.e.

$$\mathcal{H} : \lambda \in \mathcal{L} = \{\lambda(\cdot, \vartheta) \mid \vartheta \in \Theta \subseteq \mathbb{R}^d\} \quad \text{against} \quad \mathcal{K} : \lambda \notin \mathcal{L}.$$

Since no parametric form of the alternative is assumed we will use a nonparametric estimator of λ for testing \mathcal{H} against \mathcal{K} . The idea for the construction of such a nonparametric estimator goes back to the paper of Watson and Leadbetter (1964), who considered the case without censoring. The censored case was investigated, for example by Lo and Singh (1986) and by Diehl and Stute (1988). To describe the estimation procedure we introduce the distribution function of the observations T_i and the subdistribution function of the uncensored observations:

$$H(t) := \mathbf{P}(T_i \leq t) \quad \text{and} \quad H^U(t) := \mathbf{P}(T_i \leq t, \delta_i = 1).$$

Since

$$1 - H(t) = (1 - G(t))(1 - F(t)) \quad \text{and} \quad H^U(t) = \int_0^t (1 - G(s)) dF(s),$$

where G is the distribution function of the censoring times C_i , the cumulative hazard function $\Lambda(t) := \int_0^t \lambda(s) ds$ can be written as

$$\Lambda(t) = \int_0^t \frac{dF(s)}{1 - F(s_-)} = \int_0^t \frac{dH^U(s)}{1 - H(s)}.$$

Now, for estimating Λ we replace H^U and H by their empirical versions, that is by

$$\hat{H}_n^U(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(T_i \leq t, \delta_i = 1) \quad \text{and} \quad \hat{H}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(T_i \leq t). \quad (8)$$

The resulting estimator

$$\hat{\Lambda}_n(t) := \int_0^t \frac{d\hat{H}_n^U(s)}{1 - \hat{H}_n(s_-)} = \sum_{i=1}^n \frac{\mathbf{1}(T_{(i)} \leq t) \delta_{[i]}}{n - i + 1}$$

is the Nelson-Aalen estimator of Λ . Here $T_{(1)} \leq \dots \leq T_{(n)}$ are the ordered observations and $\delta_{[i]} = \delta_j$ if $T_j = T_{(i)}$. As estimator of the derivative of Λ we define the kernel smoothed Nelson-Aalen estimator

$$\hat{\lambda}_n(t) := \frac{1}{b_n} \int K\left(\frac{t-s}{b_n}\right) d\hat{\Lambda}_n(s) = \frac{1}{b_n} \sum_{i=1}^n \frac{K\left(\frac{t-T_{(i)}}{b_n}\right) \delta_{[i]}}{n - i + 1}, \quad (9)$$

where K is a kernel function and $\{b_n\}$ is a sequence of bandwidths tending to zero at an appropriate rate. As before we choose as test statistic the L_2 -distance of $\hat{\lambda}_n$ from the "smoothed version of the hypothesis"

$$e_n(t, \lambda_\vartheta) := \int K_{b_n}(t-s) \lambda(s, \vartheta) ds = \int K_{b_n}(t-s) d\Lambda(s, \vartheta),$$

where $\Lambda(t, \vartheta) = \int_0^t \lambda(s, \vartheta) ds$. Since the parameter ϑ is unknown we have to replace it by a suitable estimator. We propose to take the maximum likelihood estimator. The likelihood function is given by

$$L_n(\vartheta, T_1, \delta_1, \dots, T_n, \delta_n) = \prod_{i=1}^n L(\vartheta, T_i, \delta_i)$$

with

$$\begin{aligned} L(\vartheta, T_i, \delta_i) &= (1 - G(T_i))^{\delta_i} (1 - F(T_i, \vartheta))^{1-\delta_i} f(T_i, \vartheta)^{\delta_i} g(T_i)^{1-\delta_i} \\ &= \lambda(T_i, \vartheta)^{\delta_i} \exp(-\Lambda(T_i, \vartheta)) (1 - G(T_i))^{\delta_i} g(T_i)^{1-\delta_i}, \end{aligned} \quad (10)$$

where g is the density of the censoring times. Thus, the maximum likelihood estimator $\hat{\vartheta}_n$ is a (measurable) maximizer of

$$l_n(\vartheta) = \sum_{i=1}^n (\delta_i \log \lambda(T_i, \vartheta) - \Lambda(T_i, \vartheta)).$$

The test statistic is given by

$$\mathbb{L}_n := \int \left(\hat{\lambda}_n(t) - e_n(t, \lambda_{\hat{\vartheta}_n}) \right)^2 a(t) dt.$$

To formulate the test procedure we state the following limit theorem, proved in Liero (2003b). Let T_H be the right end point of the distribution H and fix an arbitrary point $T' < T_H$. Further set

$$\begin{aligned} \mu_n(\lambda_\vartheta) &= (nb_n)^{-1} \kappa^2 \int \frac{\lambda(t, \vartheta)}{1 - H(t)} a(t) dt, \\ \sigma^2(\lambda_\vartheta) &= 2 \int \left(\frac{\lambda(t, \vartheta)}{1 - H(t)} \right)^2 a^2(t) dt \int (\kappa^*(z))^2 dz. \end{aligned}$$

Theorem 6.1 *Suppose that the kernel satisfies (K), that the weight function a fulfills (W) and vanishes outside $[0, T']$ and that the distribution function H is Lipschitz continuous. Further, let any hazard rate $\lambda \in \mathcal{L}$ be bounded, Lipschitz continuous and partially differentiable w.r.t. ϑ ; $\nabla_\vartheta \lambda(\cdot, \cdot)$ is bounded and uniformly continuous in both arguments. If $b_n \rightarrow 0$ and $nb_n^2 \rightarrow \infty$, then under \mathcal{H} we have for all $\lambda \in \mathcal{L}$*

$$nb_n^{1/2} \left(\int \left(\hat{\lambda}_n(t) - e_n(t, \lambda_\vartheta) \right)^2 a(t) dt - \mu_n(\lambda_\vartheta) \right) \xrightarrow{\mathcal{D}} \mathbf{N}(0, \sigma^2(\lambda_\vartheta)). \quad (11)$$

To conclude from Theorem 6.1 to the asymptotic normality of our test statistic we use the consistency of the maximum likelihood estimator. For that purpose we formulate the following regularity conditions:

- (i) For all $t \in [0, \infty)$ and all $i, j = 1, \dots, k$ the second derivatives $\nabla_i \nabla_j \lambda(t, \vartheta)$ and $\nabla_i \nabla_j \Lambda(t, \vartheta)$ exist and are continuous on Θ° , the open kernel of Θ .
- (ii) For all $\vartheta \in \Theta^\circ$ and all $i, j = 1, \dots, k$

$$\begin{aligned} \nabla_i \int \lambda(t, \vartheta) dt &= \int \nabla_i \lambda(t, \vartheta) dt, \\ \nabla_i \nabla_j \int \lambda(t, \vartheta) dt &= \int \nabla_i \nabla_j \lambda(t, \vartheta) dt. \end{aligned}$$

- (iii) For any $\vartheta \in \Theta^\circ$ there exist a ν_ϑ -neighborhood $U(\vartheta, \nu) \subset \Theta^\circ$ of ϑ , and a measurable function $M(\cdot, \cdot, \vartheta)$ with $\mathbf{E}M(T_1, \delta_1, \vartheta) < \infty$ such that

$$|\nabla_i \nabla_j \log L(\vartheta', \cdot, \cdot)| \leq M(\cdot, \cdot, \vartheta) \quad \text{for all } \vartheta' \in U(\vartheta, \nu)$$

for all $i, j = 1, \dots, k$

(iv) The determinant of the Fisher information $I(\vartheta) = (I_{ij}(\vartheta))_{i,j=1,\dots,k}$ with

$$I_{ij}(\vartheta) = \int \nabla_i \lambda(t, \vartheta) \nabla_j \lambda(t, \vartheta) \frac{(1 - H(t, \vartheta))}{\lambda(t, \vartheta)} dt$$

is nonzero for all $\vartheta \in \Theta^o$.

Under these conditions we have: Suppose that $\hat{\vartheta}_n$ is consistent under \mathcal{H} , then

$$\sqrt{n} \left(\hat{\vartheta}_n - \vartheta \right) \xrightarrow{\mathcal{D}} \mathbf{N}(0, I(\vartheta)^{-1})$$

for any $\vartheta \in \Theta^o$. That is, the maximum likelihood estimator $\hat{\vartheta}_n$ is \sqrt{n} -consistent. Therefore the limit statement (11) remains true for \mathbb{L}_n . Furthermore, in the standardizing terms the unknown distribution function H can be replaced by \hat{H}_n . Thus, finally we obtain an asymptotic α -test by the rule: Reject \mathcal{H} , if $\mathbb{L}_n \geq \mu_{\lambda n} + z_\alpha \sigma_{\lambda n} / (nb_n^{1/2})$, where $\mu_{\lambda n} = (nb_n)^{-1} \kappa^2 \int \frac{\lambda(t, \hat{\vartheta}_n)}{1 - \hat{H}_n(t)} a(t) dt$ and $\sigma_{\lambda n}^2 = 2 \int \left(\frac{\lambda(t, \hat{\vartheta}_n)}{1 - \hat{H}_n(t)} \right)^2 a^2(t) dt \int (\kappa^*(z))^2 dz$.

6.2 Survival model with fixed covariates

Now, let us extend the previous approach to survival models with covariates. That is, at fixed design points $x_1 \leq x_2 \leq \dots \leq x_n$ we have nonnegative survival times Y_1, \dots, Y_n . For simplicity we assume that the support of the covariates x_i is the interval $[0, 1]$. Consequently, from the mathematical point of view, the Y_i 's are no longer identically distributed. We define analogously to section 6.1

$$F_{x_i}(t) = \mathbf{P}(Y_i \leq t), \quad H_{x_i}(t) := \mathbf{P}(T_i \leq t), \quad H_{x_i}^U(t) := \mathbf{P}(T_i \leq t, \delta_i = 1) \quad \text{and}$$

$$\Lambda_{x_i}(t) := \int_0^t \lambda_{x_i}(s) ds = \int_0^t \frac{dH_{x_i}^U(s)}{1 - H_{x_i}(s)}.$$

The problem of nonparametric estimation of λ , the survival function $1 - F$ and the hazard λ has been studied by several authors. We mention here: Gonzalez-Manteiga and Cadarso-Suarez (1994) and Van Keilegom and Veraverbeke (1997, 2001, 2002). Roughly speaking, the main aim of these papers is to approximate the distance between the function of interest and its nonparametric estimator by a sum of independent random variables. Based on such an approximation consistency properties are established and asymptotic normality at fixed points t and x is derived. A modification of a result proved by Van Keilegom and Veraverbeke (2001) leads to a limit statement for the quadratic deviation. First, let us define the estimators. The idea is the same as before - Λ is estimated by a Nelson-Aalen type estimator. But to take into account the covariates we take instead of the empirical distribution functions (8) *weighted* empirical distribution functions:

$$\hat{H}_{xn}(t) = \sum_{j=1}^n w_{nj}(x) \mathbf{1}(T_j \leq t), \quad \hat{H}_{xn}^U(t) = \sum_{j=1}^n w_{nj}(x) \mathbf{1}(T_j \leq t, \delta_j = 1).$$

Following Van Keilegom and Veraverbeke (2001) we will use Gasser-Müller type kernel weights $w_{nj}(x)$. They are defined as

$$w_{nj}(x) = \frac{1}{c_n(x)} \int_{x_{j-1}}^{x_j} \frac{1}{a_n} W \left(\frac{x-z}{a_n} \right) dz \quad \text{with} \quad c_n(x) = \int_0^{x_n} \frac{1}{a_n} W \left(\frac{x-z}{a_n} \right) dz.$$

Here $x_0 = 0$, W is a symmetric kernel function and a_n is a sequence of bandwidths. Then a nonparametric estimator of $\Lambda_x(t)$ is given by

$$\hat{\Lambda}_{xn}(t) := \int_0^t \frac{d\hat{H}_{xn}^U(s)}{1 - \hat{H}_{xn}(s_-)} = \sum_{i=1}^n \frac{\mathbf{1}(T_{(i)} \leq t) \delta_{[i]} w_{n[i]}(x)}{1 - \sum_{k=1}^{i-1} w_{n[k]}(x)}.$$

Now, further smoothing with a kernel K and bandwidth b_n leads to the estimator of the hazard function:

$$\begin{aligned} \hat{\lambda}_{xn}(t) &= \frac{1}{b_n} \int K\left(\frac{t-s}{b_n}\right) d\hat{\Lambda}_{xn}(s) \\ &= \frac{1}{b_n} \sum_{i=1}^n K\left(\frac{t-T_{(i)}}{b_n}\right) \delta_{[i]} \frac{w_{n[i]}(x)}{1 - \sum_{k=1}^{i-1} w_{n[k]}(x)}. \end{aligned}$$

Note that if we take the weights all equal to n^{-1} then the estimator becomes the estimator defined in (9) for the case without covariates. Now, consider the problem of testing the simple hypothesis

$$\mathcal{H} : \lambda_x(t) = \lambda_x^o(t) \text{ for all } t, x \text{ against } \mathcal{K} : \lambda_{x'}(t') \neq \lambda_{x'}^o(t') \text{ for some } t', x'$$

As test statistic we propose the following quadratic deviation:

$$\mathbb{W}_n = \frac{1}{n} \sum_{i=1}^n \int \left(\hat{\lambda}_{x_i n}(t) - e_n(t, \lambda_{x_i}^o) \right)^2 a(t) dt,$$

where

$$e_n(t, \lambda_{x_i}^o) = \int K_{b_n}(t-s) \lambda_{x_i}^o(s) ds.$$

To formulate the limit theorem for this functional we introduce the following quantities.

$$\begin{aligned} \zeta_n(\lambda) &= (n^2 a_n b_n)^{-1} \kappa^2 \omega^2 \sum_{i=1}^n \int \frac{\lambda_{x_i}(t)}{1 - H_{x_i}(t)} a(t) dt, \quad (12) \\ \rho_n^2(\lambda) &= 2n^{-1} \sum_{i=1}^n \int \left(\frac{\lambda_{x_i}(t)}{1 - H_{x_i}(t)} \right)^2 a^2(t) dt \int (\kappa^*(z))^2 dz \int (\omega^*(z))^2 dz, \end{aligned}$$

where ω^* denotes the convolution of the kernel W . For the design points set $\underline{s}_n = \min_{1 \leq i \leq n} (x_i - x_{i-1})$ and $\bar{s}_n = \max_{1 \leq i \leq n} (x_i - x_{i-1})$. Further define $r_n = (n a_n b_n)^{-1} \log n + (b_n^{1/2} + a_n^2 b_n^{-1}) (n a_n)^{-1/2} (\log n)^{1/2} + a_n^4 b_n^{-1}$.

The following assumptions are used:

- (i) $x_n \rightarrow 1$, $\bar{s}_n = O(n^{-1})$, $\bar{s}_n - \underline{s}_n = o(n^{-1})$
- (ii) The derivatives $\frac{\partial^2 H}{\partial x^2}$, $\frac{\partial^2 H}{\partial t^2}$, $\frac{\partial^2 H}{\partial x \partial t}$ and $\frac{\partial^3 H}{\partial x^2 \partial t}$ exist and are continuous in the interval $[0, 1] \times [0, T']$.
- (iii) The derivatives $\frac{\partial^2 H^U}{\partial x^2}$, $\frac{\partial^2 H^U}{\partial t^2}$ and $\frac{\partial^2 H^U}{\partial x \partial t}$ exist and are continuous in the interval $[0, 1] \times [0, T']$.

With these assumptions we can state the following theorem:

Theorem 6.2 Suppose that the kernels K and W satisfy (K). If assumptions (i)-(iii) and (W) are fulfilled and $r_n^2 n(a_n b_n)^{1/2} \rightarrow 0$, $r_n(a_n b_n)^{-1/2} \rightarrow 0$, then under \mathcal{H}

$$\frac{n(a_n b_n)^{1/2}}{\rho_n(\lambda^o)} (\mathbb{W}_n - \zeta_n(\lambda^o)) \xrightarrow{\mathcal{D}} \mathbf{N}(0, 1).$$

The only unknown function in the standardizing terms is H_x . We replace H_x by its consistent estimator \hat{H}_{xn} . If $((na_n)^{-1/2}(\log n)^{1/2} + a_n^2)(a_n b_n)^{-1/2} \rightarrow 0$, then the estimation error tends to zero fast enough such that the limit statement remains valid with the estimated distribution function. So, finally we get the rule:

Reject \mathcal{H} , if $\mathbb{W}_n \geq \tilde{\zeta}_n(\lambda^o) + z_\alpha \tilde{\rho}_n(\lambda^o)/(n(a_n b_n)^{1/2})$. Here $\tilde{\zeta}_n(\lambda^o)$ and $\tilde{\rho}_n(\lambda^o)$ are defined as in (12), where H_x is replaced by \hat{H}_{xn} .

Some remarks. 1.) The investigation of the power of this test requires more technical effort than that carried out in Section 2, but roughly speaking it leads to the same conclusion. Namely that the power tends to a nontrivial limit, if the squared L_2 -norm of the disturbing function tends to a nonnegative constant with a rate $n(a_n b_n)^{1/2}$.

2.) It seems to be not very difficult to extend the presented approach to the problem of testing, whether the unknown hazard rate λ_x lies in a parametric class. A more complicated problem is to test a semiparametric hypothesis. For example, suppose that the hypothetical class of hazard functions is the class of proportional hazard functions with unknown baseline hazard function $\alpha(\cdot)$ and a parametric function describing the influence of the covariates. For the construction of the test statistic one has to estimate both functions. Using the partial likelihood method one obtains a suitable estimator for the parametric part. To estimate the baseline function it seems to be useful to apply an approach via the Breslow estimator for the cumulative baseline hazard. But for this estimator rates of convergence are not derived. That means, it is not clear whether the estimation error tends to zero fast enough such that the limit theorem remains valid with the estimated hypothetical hazard function.

References

1. Aerts, M. and Augustyns, I. and Janssen, P. (1997), Smoothing sparse multinomial data using local polynomial fitting, *J. Nonparametric Statist.*, **8**, 127–147
2. Augustyns, I. (1997), Local polynomial smoothing of sparse multinomial data, *Thesis, Limburgs Universitar Centrum*
3. Bickel, P. and Rosenblatt, M. (1973), On some global measures of deviation of density function estimators, *Annals of Mathematical Statistics*, **1**, 1071–1095
4. Burman, P. (1987). Central limit theorem for quadratic forms for sparse tables, *J. Multivar. Anal.*, **22**, 258–277
5. Diehl, S. and Stute, W. (1988), Kernel density and hazard function estimation in the presence of censoring, *J. Multivariate Anal.*, **25**, 299–310
6. Ghosh, B. and Huang, Wei-Min (1991). The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit, *Annals of Statistics*, **19**, 999–1009

7. González-Manteiga, W. and Cadarso-Suarez, C. (1996). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications, *J. Nonparametric Statistics*, **4**, 65–78
8. Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *Annals of Statistics*, **21**, 1926–1947
9. Holst, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit tests, *Biometrika*, **59**, 137–145
10. Liero, H. and Läuter, H. and Konakov, V. D. (1998). Nonparametric versus parametric goodness of fit, *Statistics*, **31**, 115–149
11. Liero, H. (1999). Global deviation measures in nonparametric curve estimation, *Habilitationsschrift an der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Potsdam*
12. Liero, H. (2001). L_2 -tests for sparse multinomials, *Statistics & Probability Letters*, **55**, 147-158
13. Liero, H. (2003a). Testing homoscedasticity in nonparametric regression, *Nonparametric Statistics*, **15**, 31-51
14. Liero, H. (2003b). Testing the hazard rate, *Preprint, Institut für Mathematik, Universität Potsdam*
15. Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations, *Probab. Theory Related Fields*, **71**, 455–465
16. Rosenblatt, M. (1975). A quadratic measure of derivation of two-dimensional density estimates and a test of independence, *Annals of Statistics*, **3**, 1–14
17. Simonoff, J. S. (1996) *Smoothing Methods in Statistics*, Springer
18. Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression, *Ann. Inst. Statist. Math.*, **49**, 467–401
19. Van Keilegom, I. and Veraverbeke, N. (2001). Hazard rate estimation in nonparametric regression with censored data, *Ann. Inst. Statist. Math.*, **53**, 730–745
20. Van Keilegom, I. and Veraverbeke, N. (2002). Density and hazard estimation in censored regression models, *Bernoulli*, **8**, 607–625